

Innovative Approaches to Smoke Detection in Satellite Imagery

by

Liang Zhao

Bachelor of Information Technology (Honours)

UniSA STEM



**University of
South Australia**

November 2024

A thesis submitted for the degree of

PhD in Computer and Information Science

Contents

List of Figures	iv
List of Tables	vi
List of Abbreviations	vii
Abstract	xi
Declaration	xiv
Acknowledgements	xv
List of Publications	xvii
1 Introduction	1
1.1 Background and Rationale of the Research	2
1.1.1 Importance of Early Fire (EF) Smoke Detection	2
1.1.2 Challenges in Smoke Detection	5
1.1.3 Limitations in Previous Research	7
1.2 Research Questions	10
1.3 Overview of the Contributions to the Field	12
1.4 Structure of the Thesis	13
2 Technical and Theoretical Background	15
2.1 Satellites and Sensors	16
2.1.1 Satellites	16
2.1.2 Sensors	20
2.1.3 Fire and Smoke Detection Using Satellites	24
2.2 Spectral Patterns	30
2.3 Transfer Learning	32
3 Literature Review	34
3.1 Smoke Detection Using Non-satellite Imagery	39
3.1.1 Pixel-level Approaches	39

3.1.2	Scene-level Approaches	42
3.2	Smoke Detection Using Satellite Imagery	43
3.2.1	Pixel-level Approaches	43
3.2.2	Scene-level Approaches	48
3.3	Cross-sensor Transfer Learning	50
3.3.1	Transfer Learning in Computer Vision Using Conventional RGB Images	50
3.3.2	Transfer Learning Applied to Remote Sensing Tasks	52
3.4	Identified Research Gaps	55
4	A Lightweight Model: VIB_SD	58
4.1	Introduction	60
4.2	Materials and Methods	64
4.2.1	RGB USTC_SmokeRS Dataset	64
4.2.2	Multispectral Landsat Imagery Dataset	65
4.2.3	VIB_SD	69
4.3	Experimental Settings and Evaluation Metrics	76
4.3.1	Experimental Settings	76
4.3.2	Evaluation Metrics	78
4.4	Results	79
4.4.1	Results of the First Experiment	79
4.4.2	Results of the Second Experiment	80
4.5	Case Studies	82
4.6	Discussion	89
4.7	Conclusion	90
5	IA: Learning Spectral Patterns	92
5.1	Introduction	94
5.2	IA	96
5.3	Experimental Settings	101
5.3.1	Datasets	101
5.3.2	Training Settings	102
5.3.3	Evaluation Metrics	103
5.4	Experimental Results	103
5.4.1	Model Performance with and without IA	104
5.4.2	Visualisation of IA-extracted Spectral Patterns	105
5.4.3	Ablation Studies and Parameter Selection	108
5.5	Discussion	109
5.6	Conclusion	111

CONTENTS

6	Cross-sensor Transfer Learning	112
6.1	Introduction	115
6.2	Materials and Methods	119
6.2.1	Datasets	119
6.2.2	Methods	121
6.3	Experimental Settings and Evaluation Metrics	124
6.3.1	Training Environment	124
6.3.2	Pre-training IA_VIB_SD on Landsat6c	124
6.3.3	Training the Benchmark and Transferred Models on Sentinel7c	125
6.3.4	Evaluation Metrics	125
6.4	Results and Discussion	126
6.5	Conclusion	131
7	Conclusions	133
7.1	Summary of Contributions	134
7.2	Practical Implications	136
7.3	Future Directions	137
	Appendix A: Statements of Authorship	140

List of Figures

1.1	Variants of smoke in Landsat 8 Operational Land Imager (OLI) true-colour imagery. (a) Dark grey smoke plumes under cirrus clouds. (b) Long slim smoke plume in bright colour. (c) Dispersed smoke on the edge of the image. (d) Brown-coloured dense smoke in the whole image. (e) Wide, dispersed smoke in light blue colour covering the whole image. (f) Spread dense smoke in dark grey colour under altocumulus clouds.	6
1.2	Cloud, haze, dust, and smoke captured in Moderate Resolution Imaging Spectroradiometer (MODIS) true-colour imagery are difficult to visually differentiate.	6
2.1	Spectral signatures of vegetation, water, and soil across different wavelengths, adapted from [149].	31
3.1	Comparison of smoke scenes in a typical surveillance camera image and Landsat 8 OLI imagery.	37
3.2	Previous transfer learning approaches require the imagery data from both satellites to have equal channel counts or to align the channel counts through adjustments (e.g., dimensionality reduction).	55
4.1	Two smoke scenes are visualised in different bands. (a) RGB. (b) short-wave infrared (SWIR) ₂ , near-infrared (NIR), and blue.	62
4.2	Overlapped tiling.	68
4.3	Structure of Variant Input Bands for Smoke Detection (VIB _{SD}).	70
4.4	Attention modules. (a) Spatial Attention. (b) Channel Attention.	72
4.5	Inception-Residual module.	74
4.6	Inception-Attention module.	75
4.7	Boxplots based on the 10 results. (a) Accuracy. (b) Kappa.	81
4.8	Smoke scenes. (a) Smoke mixed with thin clouds above the seaside. (b) Diffused smoke at multiple sites under altocumulus clouds. (c) Cloud-free smoke (in the red circle) over a very small geographic extent. (d) Cloud-free smoke plumes of varying scales at two different sites.	83
4.9	Prediction results for scene (a) in Figure 4.8.	85

LIST OF FIGURES

4.10	Prediction results for scene (b) in Figure 4.8.	86
4.11	Prediction results for scene (c) in Figure 4.8.	87
4.12	Prediction results for the top right smoke area of scene (d) in Figure 4.8.	88
4.13	Prediction results for the middle smoke area of scene (d) in Figure 4.8.	88
5.1	The Structure of the Input Amplification (IA) Module. (a) IA (b) Spatial Attention (c) Channel Attention.	97
5.2	Process of spectral pattern extraction by IA. (a) Raw spectral pattern extraction is achieved through band attention by using 1×1 filters with non-linear activation (b) Spatial attention emphasises the pixels that are spatially related while ignoring none-related pixels, hence refines the class-oriented spectral patterns (c) Channel attention weights the spectral patterns according to their importance in the classification task	98
5.3	Original MODIS imagery samples (left) vs. two samples of deep-pseudo bands extracted by IA	106
5.4	Original Landsat imagery samples (left) vs. two samples of deep-pseudo bands extracted by IA	107
6.1	The proposed transfer learning approach increases the input dimension by utilising IA. Notably, although IA was originally designed to extract spectral patterns to improve VIB_SD performance, it also serves as a mechanism for mapping diverse input bands to a consistent number of spectral patterns, which are used as input to VIB_SD.	118
6.2	Schematic of model training and transfer learning approaches. (a) IA_VIB_SD trained purely on Landsat6c (as the pretrained model to be transferred) or Sentinel7c (as the benchmark model to evaluate the performance of the transferred models). (b) Conventional approach to transferring Landsat6c-trained IA_VIB_SD to Sentinel7c: “HT” and “HTNL” map seven input bands in Sentinel7c data to six bands, aligning with the Landsat6c-pretrained model, using linear and non-linear mapping, respectively. (c) Proposed approach to transferring Landsat6c-trained IA_VIB_SD to Sentinel7c: “FineTune” involves replacing the IA Module in the pretrained IA_VIB_SD, freezing the VIB_SD, and updating its weights through fine-tuning; “Init” entails training IA and VIB_SD together with random IA weights and pretrained VIB_SD weights. . . .	122
6.3	Boxplots of the results.	129
6.4	Line charts of the results.	130

List of Tables

2.1	Summary of low Earth orbit (LEO), medium Earth orbit (MEO), and geostationary orbit (GEO) orbit types and their characteristics.	17
2.2	Comparison of some attributes in different satellite sensors	24
2.3	Examples of sensors and their hot-spot detection bands	25
2.4	Comparison of satellite sensors suitable for fire detection [28, Table 1] .	26
2.5	Some spectral indices used in remote sensing	31
3.1	Approaches used in smoke detection across different platforms.	38
3.2	Summary of the YOLO Model family	43
4.1	Number of images in USTC_SmokeRS.	64
4.2	A sample record in the South Australia (SA) historical fires dataset. .	66
4.3	Wavelengths of the selected bands of Landsat 5 Thematic Mapper (TM) and Landsat 8 OLI.	67
4.4	Class distribution of the Landsat6c dataset.	69
4.5	Formulas for accuracy and Kappa.	79
4.6	Model performance comparison.	79
4.7	Performance of models using variant bands based on 10 results.	80
4.8	Performance of models using variant bands with the same samples. . .	82
5.1	A summary of datasets used in this paper	101
5.2	Results of using the USTC_SmokeRS dataset	104
5.3	Results of using the Landsat6c dataset	104
5.4	Ablation study about attention mechanism	108
5.5	Ablation study about the number of 1×1 Conv2D Layers	108
5.6	Ablation study about the output channel number of the IA module . .	109
6.1	Accuracy of Landsat6c-trained IA_VIB_SD predicting Sentinel7c imagery	115
6.2	Specifications of selected bands from Landsat 5 TM, Landsat 8 OLI, and Sentinel-2 Multispectral Instrument (MSI) used in the Landsat6c and Sentinel7c	120
6.3	The class distributions of Landsat6c and Sentinel7c	120
6.4	Parameter counts of the models when trained using Sentinel7c.	126

LIST OF TABLES

6.5 Results of all transferred models versus the benchmark model trained
with 10% - 50% Sentinel7c data 127

List of Abbreviations

- AHI** Advanced Himawari Imager. 4, 8, 22, 23, 25, 26, 61, 65
- AVHRR** Advanced Very-High-Resolution Radiometer. 9, 44–46
- BT** brightness temperature. 44, 46, 47, 93, 95
- CALIPSO** Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation. 45
- CNN** convolutional neural network. xii, 8, 28, 35, 36, 38, 42, 43, 48, 49, 51, 52, 59, 61–63, 90, 94–96, 102, 105, 109–111, 134
- CubeSat** cube satellite. 13, 18, 19, 137
- DEA** Digital Earth Australia. 65, 68
- DL** deep learning. xi, xii, 4, 7, 8, 10–14, 19, 20, 27–29, 35, 36, 38, 41, 42, 47, 50, 51, 55–57, 59, 61, 63, 64, 76, 93–96, 100, 109–111, 113, 115, 118, 121, 131, 134, 135, 137, 138
- EF** early fire. i, xi, xii, 2–5, 7, 9, 10, 12, 13, 35, 49, 56, 57, 59–62, 64, 65, 73, 90, 94, 113, 134, 136, 137
- FCN** fully convolutional network. 8, 36, 38, 41, 47
- FN** false negative. 27, 28, 56, 89
- FNR** false negative rate. 91, 95, 103–105, 108, 109, 125, 127, 128
- FRP** Fire Radiative Power. 27, 47
- GAN** generative adversarial network. 38, 41
- GEO** geostationary orbit. vi, 16, 17, 22

List of Abbreviations

- IA** Input Amplification. v, vi, xii, xiii, 13, 14, 94–111, 113–115, 117–119, 121–123, 125, 126, 128, 131, 134–138
- IR** infrared. xii, xiii, 9–12, 14, 21, 28, 29, 36, 38–40, 44, 46, 49, 51, 53, 55–57, 59, 61, 63–65, 70, 76, 82, 90, 93, 95, 101, 113, 134
- LEO** low Earth orbit. vi, 16–18, 22
- MEO** medium Earth orbit. vi, 16, 17
- MLP** multi-layer perceptron. 35, 36, 38, 41, 42, 46, 47
- MODIS** Moderate Resolution Imaging Spectroradiometer. iv, v, xii, 6, 8, 9, 12, 16, 19, 22–26, 44, 46, 48, 56, 59, 61, 64–66, 93, 94, 96, 101, 106
- MSI** Multispectral Instrument. vi, 23, 24, 114, 119, 120, 136
- MWIR** mid-wave infrared. 3, 4, 10, 24–28, 44, 46
- NBART** Nadir Corrected Bi-directional Reflectance Distribution Function Adjusted Reflectance Coupled with a Terrain Illumination Correction. 65
- NBR** Normalised Burn Ratio. 30, 31
- NDBI** Normalised Difference Built-Up Index. 30, 31
- NDVI** Normalised Difference Vegetation Index. 27, 30, 31
- NIR** near-infrared. iv, xi, xii, 3, 10, 12, 13, 19, 21, 22, 28, 31, 43, 44, 46, 47, 57, 60–64, 67, 69, 76, 80, 82, 96, 101, 111, 119, 120, 136
- NRT** near real-time. 3, 8, 12, 27, 55, 56, 61, 62, 115, 116, 136, 137
- NSW** New South Wales. 65, 67, 68, 101
- OLI** Operational Land Imager. iv, vi, 3, 5, 6, 16, 22–24, 36, 37, 59, 63, 65, 67, 68, 83, 90, 113, 119, 120, 136
- SA** South Australia. vi, 13, 65–68, 91, 101
- SmallSat** small satellite. xi, xiii, 7, 8, 10, 12, 18, 20, 55, 56, 59, 61–63, 69, 80, 90, 91, 117, 134, 135, 137, 138
- SVM** support vector machine. 35, 40, 42

- SWIR** shortwave infrared. iv, xi, xii, 3, 10, 12, 13, 21, 31, 43, 44, 46, 47, 57, 60–64, 67, 69, 76, 80, 82, 96, 101, 111, 119, 120, 136
- TIR** themal infrared. 3, 4, 10, 22, 24, 25, 27–29, 44, 46
- TIRS** Thermal Infrared Sensor. 16, 26
- TM** Thematic Mapper. vi, 59, 63, 65, 67, 68, 90, 113, 119, 120, 136
- TOA** top of atmosphere. 46, 47
- UAV** unmanned aerial vehicle. 2, 14, 35, 39, 41, 42, 52
- USD** US dollars. 7, 17, 18, 26
- VIB_SD** Variant Input Bands for Smoke Detection. iv–vi, xii, xiii, xviii, 12–14, 59, 60, 63, 64, 69–71, 74, 76, 77, 79, 80, 90, 91, 93, 95, 96, 102, 104, 105, 108, 113–115, 117–119, 121–123, 125, 126, 128, 131, 134–138
- VIIRS** Visible Infrared Imaging Radiometer Suite. 19, 25, 26, 47
- VIT** vision transformer. 95, 96, 110

Abstract

Early detection of wildfires is crucial to mitigate their catastrophic effects on lives and natural and built environments. Smoke (referring to *fire smoke* in this thesis) detection is considered a promising approach for early fire (EF) detection since smoke plumes are usually the first indicators when wildfires occur. Using satellite imagery for smoke detection stands out because it provides cost-effective monitoring that covers large scales and remote areas. Scene-level detection classifies imagery into smoke or other classes based on whether smoke is in the imagery. It offers faster inference and better scalability than pixel-level detection, which aims to identify all smoke pixels individually. This makes scene-level detection ideal for fire disaster mitigation. In this thesis, smoke detection refers to satellite-based scene-level smoke detection unless specified.

Despite significant advancements in smoke detection enabled by deep learning (DL), several limitations in previous research still need to be addressed. Existing DL models are often complex and excessively demanding in terms of power supply, memory usage, and computing resources. Such models are difficult to deploy onboard satellites, particularly small satellites (SmallSats). Additionally, previous DL-based smoke detection research focused on RGB imagery and did not pay enough attention to spectral information that could potentially improve detection accuracy. Furthermore, an effective mechanism for fast model development for multiple satellites is yet to be investigated to enhance the timely detection of EF smoke. This needs to consider that new sensors usually lack observational data and that imagery data from different sensors often present significant disparities.

This thesis develops three innovative approaches to address these limitations progressively. To facilitate the study, two multispectral imagery training datasets, **Landsat6c** and **Sentinel17c** with a medium spatial resolution of 30 metres and 10 metres respectively, are created. **Landsat6c** additionally incorporates one near-infrared (NIR) and two shortwave infrared (SWIR) bands and **Sentinel17c** includes four additional

infrared (IR) bands (two NIR and two SWIR), compared to the sole publicly available training dataset `USTC_SmokeRS`, derived from Moderate Resolution Imaging Spectroradiometer (MODIS) three-band (RGB) imagery with a low spatial resolution (1 km).

The first approach proposes a lightweight convolutional neural network (CNN) model called `Variant Input Bands for Smoke Detection (VIB_SD)`. `VIB_SD` contains less than 2% of the parameters of the state-of-the-art model `SAFA` (1.66 million versus 84.2 million) but demonstrates competitive accuracy (93.57% versus 96.22%) when trained using `USTC_SmokeRS`. This approach proceeds to train `VIB_SD` using `Landsat6c` with five different band combinations to investigate the contributions of IR bands to detection accuracy. Results show that incorporating the NIR band enhances accuracy compared to using solely RGB bands (84.82% versus 83.20%) and integrating both SWIR bands leads to further improvements compared to using just one SWIR band (86.45% versus 85.64%). Case studies illustrate `VIB_SD`'s effectiveness in detecting EF smoke amidst cloud cover when trained with `Landsat6c`.

The second approach further investigates how to effectively explore useful spectral information in IR bands to improve smoke detection accuracy. Specifically, this approach introduces a DL module named `Input Amplification (IA)` which enables DL models to automatically learn class-oriented spectral patterns. `IA` amplifies the input band dimension (e.g., three or six) to 32 (determined through experiments), with the learned spectral patterns added as pseudo bands. This allows for simultaneous learning of multiple spectral patterns and integrating them with the original bands. Functioning as an input pre-processing block, `IA` facilitates seamless integration with various DL architectures. The effectiveness of `IA` is demonstrated through its integration with different CNN architectures (i.e., `ResNet50`, `InceptionResNetV2`, `MobileNetV2`, and `VIB_SD`) and testing on both `USTC_SmokeRS` and `Landsat6c`. Significant accuracy improvements were observed for the examined CNN models after integration with `IA`, showcasing `IA`'s great potential in advancing smoke detection. Notably, higher accuracy increments were obtained when using `Landsat6c` featuring additional IR bands (4.61%, 1.08%, 1.9%, and 3.54% respectively for `ResNet50`, `InceptionResNetV2`, `MobileNetV2`, and `VIB_SD`).

The third approach introduces a novel cross-sensor transfer learning method, notably aided by the `IA` module. This aims to facilitate fast model development for multiple satellites, considering data disparities across different sensors and the limited observational data available from new sensors. `Landsat6c`, originally containing

1836 images (about 600 images per class), was expanded to 2770 images (more than 900 images per class) and used as the source domain. **Sentinel17c**, consisting of only 351 images (about 120 images per class), served as the target domain. The model incorporating IA and VIB_SD, named IA_VIB_SD, was employed. The transferability of the **Landsat6c**-trained IA_VIB_SD model to **Sentinel17c** was investigated using various transfer learning techniques and compared to the performance of the benchmark IA_VIB_SD model exclusively trained on pure **Sentinel17c** data. The proposed transfer learning method resulted in a transferred model with an average accuracy 5% higher than the benchmark model. Notably, the proposed transfer learning method outperformed conventional transfer learning methods by more than 1% in terms of accuracy, even when trained on only 10% of the **Sentinel17c** dataset.

In summary, the core achievement of this thesis is the IA_VIB_SD model. This model significantly enhances smoke detection accuracy by employing both RGB and IR bands and learning class-oriented spectral patterns from these bands. It has great potential to facilitate onboard satellite smoke detection, particularly on SmallSats and SmallSat constellations, due to its lightweight design. Aided by the IA module's capability of adapting learned spectral patterns, IA_VIB_SD pretrained on one sensor demonstrates high accuracy when transferred to a new sensor using minimal training data from the target sensor.

Declaration

I declare that this thesis presents work carried out by myself and does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any material previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, are clearly acknowledged.

Liang Zhao
November 2024

Acknowledgements

I am profoundly grateful to my supervisory panel for their invaluable guidance, exceptional advice, and unwavering support throughout my PhD journey. My deepest thanks go to my principal supervisor, Associate Professor Jixue Liu, whose expertise and insight have been pivotal in shaping my research since my Honour's degree. Despite the various challenges along the way, my academic journey has always been pleasant under his heartfelt and supportive mentorship.

I am equally indebted to my co-supervisors, Professor Jiuyong Li and Dr. Stefan Peters, both of whom also guided me during my Honour's degree. The constructive advice and feedback they offered significantly enhanced the quality and depth of my research. Their astute guidance and continuous encouragement have seen my dramatic growth from an inexperienced undergraduate to a competent researcher.

I extend my deep appreciation to my industrial advisors, Mr. Norman Mueller and Mr. Simon Oliver, whose practical perspectives and assistance have greatly enriched my academic experience. The invaluable industrial resources they shared and the experiences they imparted greatly boosted the efficiency of my work, particularly during data collection and processing, which posed substantial challenges.

I gratefully acknowledge the support provided by the Australian Government Research Training Program (RTP) Scholarship, which has enabled me to pursue this research. I am also thankful for the ongoing support from SmartSat CRC, whose provision of a top-up scholarship has been instrumental in my research endeavours. The opportunities they provided me to engage with industry partners and present my project have been invaluable in broadening my professional horizons and deepening my understanding of the field.

I also want to express my gratitude to my colleagues in the Data Analytics Group. To Dr. Sha Lu and Xiongren Chen, who have provided precious inspiration through

in-depth discussions and brainstorming. To Dr. Xiaomei Li, Dr. Ziqi Xu, Dr. Andres Mauricio Cifuentes Bernal, Dr. Ji-Young Park, Dr. Oscar Blessed Deho, and Dr. Ha Xuan Tran for their encouragement, expertise, and resource sharing. To all other members of the group who have enriched my knowledge through presentations and discussions in group meetings. My PhD journey has been enjoyable and memorable because of you.

Furthermore, I am deeply thankful to my family—my parents, my partner, and particularly my two sons, Aaron and Archer, born during my PhD study. It is their unconditional love and support that have enabled me to come this far. The immense joy brought by the birth of Aaron and Archer has made my PhD journey unique and unforgettable.

Once again, I must express my heartfelt gratitude to my supervisors and the STEM academic unit at the University of South Australia. Their support has been particularly meaningful since the birth of my sons during my PhD studies. The flexibility and understanding they have shown have enabled me to balance my academic commitments with my responsibilities as a parent, for which I am deeply thankful.

Lastly, I extend my appreciation to all who have supported me in this endeavour, both seen and unseen. Your contributions have not gone unnoticed, and I am deeply grateful to each one of you.

Liang Zhao

November 2024

List of Publications

The main content of this thesis builds upon three first-authored, peer-reviewed journal papers, which present the core findings of my PhD research:

1. **L. Zhao**, J. Liu, S. Peters, J. Li, S. Oliver, and N. Mueller, “Investigating the impact of using IR bands on early fire smoke detection from Landsat imagery with a lightweight CNN model”, *Remote Sensing*, vol. 14, no. 13, p. 3047, 2022. *Chapter 4 of this thesis is based on this paper*

2. **L. Zhao**, J. Liu, S. Peters, J. Li, N. Mueller, and S. Oliver, “Learning class-specific spectral patterns to improve deep learning-based scene-level fire smoke detection from multi-spectral satellite imagery”, *Remote Sensing Applications: Society and Environment*, vol. 34, 2024, 101152. *Chapter 5 of this thesis is based on this paper*

3. **L. Zhao**, J. Liu, S. Peters, J. Li, N. Mueller, and S. Oliver, “Cross-sensor transfer learning for fire smoke scene detection using variable-bands multi-spectral satellite imagery aided by spectral patterns”, *International Journal of Remote Sensing*, 2024. *Chapter 6 of this thesis is based on this paper*

Other co-authored publications not included in the content of this thesis:

4. S. Lu, E. Jones, **L. Zhao** *et al.*, “Onboard AI for fire smoke detection using hyperspectral imagery: an emulation for the upcoming Kanyini Hyperscout-2 mission”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024. *The model prototype used for the emulation in this paper is the VIB_SD model proposed in publication 1*

5. J. Liu, J. Li, S. Peters, **L. Zhao**, “A transformer boosted UNet for smoke segmentation in complex backgrounds in multispectral LandSat imagery”, Remote Sensing Applications: Society and Environment, vol. 36, 2024, 101283. *Collaborated work with my supervisors*

Chapter 1

Introduction

1.1 Background and Rationale of the Research

1.1.1 Importance of Early Fire (EF) Smoke Detection

In recent years, the world has witnessed a surge in devastating fire disasters, exacerbated by the far-reaching impacts of global climate change [1]. These catastrophic events have inflicted tremendous losses on communities and ecosystems [2, 3, 4]. For example, two wildfires in California, USA, in August 2020 burned 682,135 acres, destroyed 3,843 structures, and caused 21 deaths collectively; another wildfire in 2018 burned a smaller area but destroyed 18,804 structures and caused 85 deaths [5]. In Australia, an unprecedented bushfire season lasted almost half a year from 2019 to 2020, with hotspots burning simultaneously in multiple regions across the country. The fires burned 24 to 40 million hectares, emitted 434 million tons of carbon dioxide, killed 33 people and more than 3 billion animals [6], and caused 1.95 billion Australian dollars in smoke-related health costs [7].

EF detection is crucial in mitigating the destructive impact of wildfires. The earlier fires can be detected, the more lives, property, and natural resources can potentially be saved [8]. Various methods have been employed with this aim:

- **Volunteer Geographic Information:** This method involves people reporting fires via phone apps or calls. While effective in populated areas, it relies heavily on human presence and prompt reporting [9].
- **Fire Towers:** Traditionally manned by lookouts who visually scan for smoke, these towers are now gradually equipped with surveillance cameras and other advanced monitoring technologies. This provides continuous monitoring even when personnel are absent, enhancing detection capabilities under various visibility conditions [10, 11]. However, surveillance cameras often cover only a limited area, potentially missing fires outside their field of view, and can fail due to harsh weather conditions or technical issues [12]. In addition, fire towers are often located in remote areas, resulting in costly maintenance and operation, and delays in response time [13, 14].
- **Unmanned aerial vehicles (UAVs) or Aerial Imagery:** UAVs or planes equipped with cameras can detect fires from the air. This method allows for high-resolution imagery and can cover large areas quickly, but it is resource-intensive and dependent on flight operations [15, 16].

- **Satellite Remote Sensing:** Satellite-based detection offers comprehensive coverage, especially beneficial in remote areas where ground-based methods are impractical. Satellites equipped with thermal and infrared sensors can detect heat signatures from fires, providing early warnings and continuous monitoring capabilities [17, 18].

Satellite-based fire detection has gained significant attention due to the wide availability of satellites monitoring the Earth seamlessly at decreasing costs. According to the Union of Concerned Scientists, more than 6,700 operational satellites are orbiting the Earth [19]. Many of these satellites are equipped with mid-wave infrared (MWIR) and thermal infrared (TIR) bands, which are sensitive to temperatures, making them suitable for active fire detection.

However, using satellites for the timely detection of EFs, which typically burn at small geographical extents and lower temperatures, remains challenging due to several factors:

- EFs can be easily obscured by thick canopies, clouds, haze, or the smoke plumes they emit.
- EFs can be masked by the heated background in hot weather, and false alarms can frequently be caused by other hot sources (e.g., heated bare soils or deserts, highly reflective regions, gas flames, geysers, hot springs) when using the MWIR or TIR bands for detection [20, 21, 22].
- Satellites suitable for detecting smaller fires are typically sun-synchronous satellites in near-polar orbits, which are often equipped with sensors of higher spatial resolutions. However, these satellites tend to have lengthy revisit times (or temporal resolution when referring to sensors), leading to delayed detection of burning fires. For example, the Landsat 8 Operational Land Imager (OLI) sensor has a spatial resolution of 30 metres in its near-infrared (NIR) and shortwave infrared (SWIR) bands, meaning that the smallest fire it can detect is 900 m². However, its temporal resolution is 16 days, meaning that a fire starting right after the satellite has passed the fire location can keep burning for 16 days without being detected by the satellite.
- In contrast, sensors suitable for near real-time (NRT) fire detection, typically those on geostationary satellites, tend to have coarse spatial resolutions, leading

to the overlooking of smaller fires. For example, the Advanced Himawari Imager (AHI) on the geostationary satellite Himawari 8 has a temporal resolution of 10 minutes, meaning it can capture an image of the same location on Earth every 10 minutes. However, the spatial resolution of its MWIR band, used for hotspot detection, is 2 km, meaning the smallest active fire it can detect is 4 km².

Smoke (referring to *fire smoke* in this thesis) detection emerges as a promising alternative to direct fire detection due to the following advantages:

- Smoke plumes are less likely to be obscured as they can rise quickly into the sky, often becoming the first indicator seen from space when fires occur.
- Smoke often presents distinctive colours against the background, particularly vegetation, making it highly observable from space.
- Smoke plumes disperse quickly over larger areas than the actual spread of fires, allowing smaller fires to be detected more effectively.
- Smoke plumes have significantly lower temperatures than burning fires and other hot sources, thus reducing the likelihood of false positive alarms compared to fire hot-spot detection using MWIR or TIR bands.

These advantages position smoke detection as an effective and proactive approach to EF detection. In particular, utilising satellite imagery and advanced deep learning (DL) models has emerged as a promising and cost-effective strategy, since well-trained DL models have demonstrated significantly higher scene classification accuracy than traditional algorithms.

Smoke detection can be either at the pixel level, which segments all smoke pixels from others, or at the scene level, which classifies if the scene in the imagery contains smoke. Compared to pixel-level detection, scene-level detection (or scene detection) has the following advantages:

- Classification models are generally less complex than segmentation models. They focus on determining whether an image contains smoke, which involves a single decision per image [23].

- Classification models require labelled images indicating the presence or absence of smoke. These labels are simpler and quicker to generate than pixel-level annotations needed for segmentation. Scene-level labelling is usually more accurate compared to pixel annotation [24].
- Classification models typically process images faster during inference since they only need to output a single label per image. This can be crucial for real-time monitoring and early warning systems [25].
- Due to their simplicity, classification models can be more easily scaled to analyse a large number of images, covering extensive geographical areas without requiring significant computational resources [26].

This thesis focuses on smoke detection at the scene level using satellite imagery. Unless specified, smoke detection refers to *satellite-based scene-level smoke detection* in this thesis.

1.1.2 Challenges in Smoke Detection

To prevent fire disasters, smoke plumes must be detected quickly and accurately while they are still early and small. Higher spatial resolution is required for satellite sensors to detect smaller smoke plumes and higher temporal resolution is needed for quicker detection. Achieving higher accuracy typically involves more complex detection models.

However, the accurate and timely detection of EF smoke faces several challenges:

1. Challenges related to accurate smoke detection:

- The diverse shapes and colours of smoke plumes, as shown in Figure 1.1, captured by the Landsat 8 OLI sensor.
- The presence of confounding aerosol phenomena, such as clouds, haze, and dust, overlap with smoke in spectral and structural characteristics and often intermingle, making it harder to distinguish them visually, as seen in Figure 1.2.

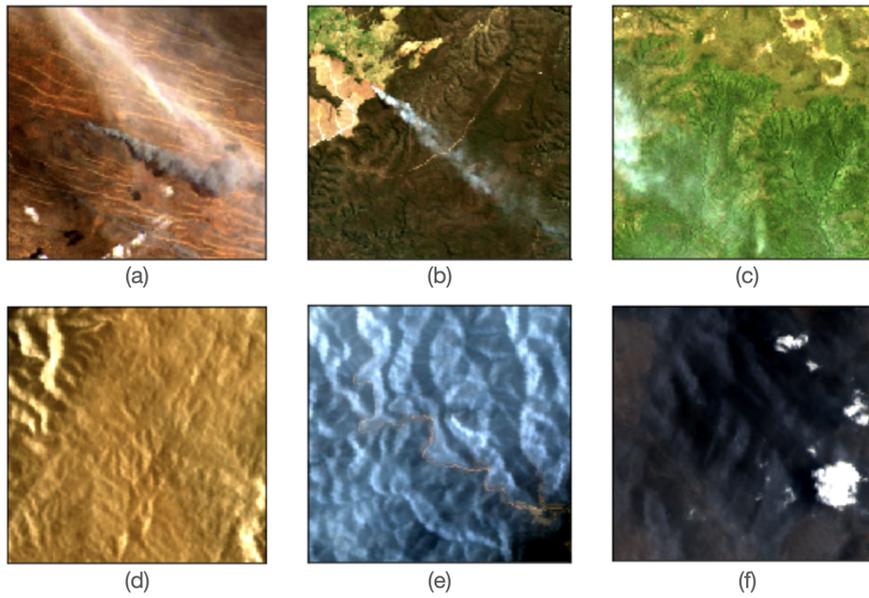


Figure 1.1: Variants of smoke in Landsat 8 OLI true-colour imagery. (a) Dark grey smoke plumes under cirrus clouds. (b) Long slim smoke plume in bright colour. (c) Dispersed smoke on the edge of the image. (d) Brown-coloured dense smoke in the whole image. (e) Wide, dispersed smoke in light blue colour covering most of the image. (f) Dense smoke in dark grey colour under altocumulus clouds. Adopted from [27, Figure 1]

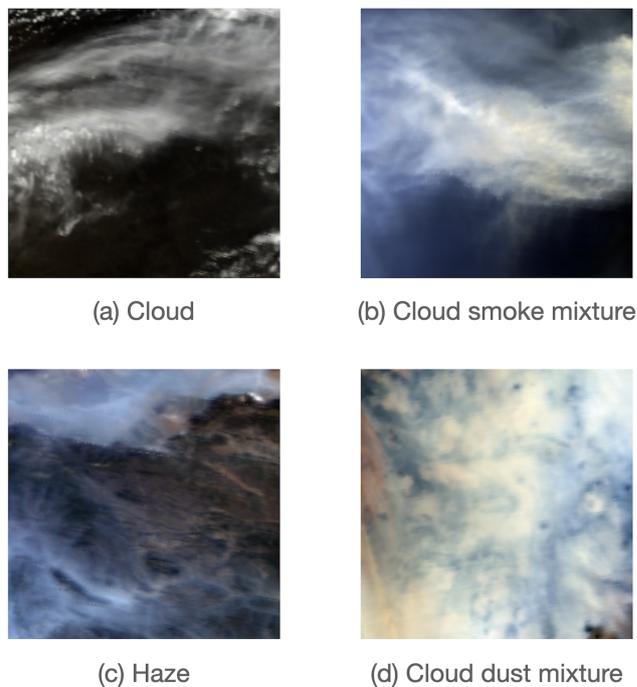


Figure 1.2: Cloud, haze, dust, and smoke captured in Moderate Resolution Imaging Spectroradiometer (MODIS) true-colour imagery are difficult to visually differentiate.

2. Challenges related to timely smoke detection:

- Delays due to latency caused by data downlink, image processing, and analysis in ground-based approaches, especially DL methods. For example, the latency of Sentinel-2 A/B is 2-4 hours [28], in addition to a revisit time of 3-5 days. During this period, small fires can develop uncontrollably.
- Limited power supply, memory, and computing resources for timely onboard satellite detection, often requiring detection models to trade accuracy for suitability [29, 30, 31]. As such, despite generally producing higher accuracy, complex DL models are often excluded from onboard satellite applications. This particularly affects small satellites (SmallSats), which are increasingly considered the future of satellite-based fire detection [32, 33].

3. Challenges related to achieving both high temporal resolution and spatial resolution:

- Sun-synchronous satellite sensors generally have lower temporal resolutions but higher spatial resolutions, while geostationary satellites typically offer higher temporal resolutions but coarser spatial resolutions. Although some geostationary satellite sensors achieve spatial resolutions below 50 metres, such satellites are rare, and the cost to develop, launch, and operate them is often prohibitive. Known geostationary satellites with spatial resolutions below 50 metres, such as Gaofen-4, Gaofen-13, Gaofen-13-02, Ludi Tance-4, and the recently launched Yaogan-41, are all owned by the Chinese government, with highly restricted public data access [34, 35, 36]. Estimates suggest that developing a single satellite of this type can cost from hundreds of millions to over a billion US dollars (USD), with annual operational costs running into several million to tens of millions of USD [37, 38]. Instead of relying solely on geostationary satellites with high spatial resolutions, a more feasible and cost-effective alternative is required to achieve the necessary temporal resolution and spatial resolution for the timely detection of EF smoke.

1.1.3 Limitations in Previous Research

Using satellite imagery for smoke detection has been studied for a long time. However, previous research primarily focused on pixel-level detection, which, as discussed earlier, is less advantageous for fire disaster mitigation compared to scene-level detection. Moreover, pixel-level detection predominantly relied on non-DL methods and

often involved manual feature engineering. For instance, multiple threshold values were typically derived manually from selected spectral or pseudo bands to help distinguish smoke from clouds and other confounding pixels [39, 40]. These threshold values can be significantly influenced by the physical and chemical characteristics of smoke particles and the spectral radiometric characteristics of the sensors.

One noteworthy study on pixel-level smoke detection is the work in [41], which proposed a fully convolutional network (FCN) to segment smoke pixels in Himawari 8 AHI imagery. This approach avoided cumbersome feature engineering by leveraging a DL method, which also learns spatial features to enhance pixel-level detection. However, the smoke masks (the labels in the training dataset) were annotated using an algorithm [42] rather than ground truth data. This highlights a common issue in using satellite imagery for pixel-level smoke detection with DL methods. A reliable pixel-level satellite imagery training dataset has not yet been made publicly available to support future research efforts in this field.

Satellite-based scene-level smoke detection, as a relatively new research area and the focus of this thesis, was first investigated in 2019 within the work of Ba et al. [43]. In this study, the authors proposed SmokeNet, a convolutional neural network (CNN) model specifically designed for smoke detection, and introduced USTC_SmokeRS, the first satellite imagery training dataset derived from MODIS RGB imagery and labelled at the scene level. SmokeNet outperformed other renowned DL models (e.g., ResNet [44], DenseNet [45], SE-ResNet [46]) when trained and tested on USTC_SmokeRS. Following this work, Chen et al. [47] proposed the current state-of-the-art model SAFA, which improved accuracy from 92.75% with SmokeNet to 96.22% using USTC_SmokeRS. SAFA also outperformed other advanced DL models that were developed more recently, including D-CNN [48], RSSC-ETDL [49], KFBNet [50], and LPDCMEN [51], among others.

However, several limitations are yet to be addressed in satellite-based scene-level smoke detection for fire disaster mitigation, beyond improving the accuracy.

Firstly, the DL models specifically designed for smoke detection are complex and have a substantial number of parameters. SmokeNet comprises 53.5 million parameters, while SAFA contains 84.2 million parameters. The models can hardly meet the growing demand for using SmallSats or SmallSat constellations for NRT fire disaster monitoring, due to their complexity and high requirements for power supply, memory space, and

computational resources.

Secondly, previous research did not effectively address the accurate detection of EF smoke. On the one hand, studies using the USTC_SmokeRS MODIS RGB imagery did not focus on spectral information, particularly in the infrared (IR) bands, which is crucial for distinguishing smoke from other confounding aerosols. On the other hand, although MODIS is onboard the sun-synchronous satellites Aqua and Terra, the USTC_SmokeRS dataset has a low spatial resolution of 1 km. This means that small smoke plumes, which are critical for early alerts, are likely to be overlooked.

Thirdly, previous research did not effectively address the timely detection of smoke, primarily due to the low temporal resolution (0.5 days) of MODIS, even when using both Aqua and Terra for detection. While multiple satellites may collectively improve temporal resolution, this approach requires customising models for each satellite due to significant disparities in their imagery data. However, an effective mechanism for fast model development across multiple satellites, considering the challenges posed by limited training data or even a lack of observational data from new satellites, is yet available. Although the Hazard Mapping System of the NOAA serves as a commendable example of a multi-satellite fire and smoke detection program, it primarily relies on traditional remote sensing algorithms to identify thermal anomalies indicative of active fires. Additionally, the Hazard Mapping System depends on manual processing and validation to review and integrate results from different sensors for the final prediction and dissemination [52, 53].

Although outside the primary focus of this thesis, it is worth noting that the work in [54] incorporated spatiotemporal changes for smoke detection in satellite imagery. This study presents a rare example of detecting spatiotemporal changes in sun-synchronous satellite imagery, which does not naturally form a time series. The changes were derived from two Advanced Very-High-Resolution Radiometer (AVHRR) images—one before the fire event and one during the event—capturing the same location during two separate satellite passes. The two passes are at least 12 hours apart due to the temporal resolution of AVHRR. Manual geometric correction is required to align the images, allowing for the identification of spatiotemporal changes. In addition, this method is more practical for smoke detection with geostationary satellites, which produce time series imagery at short intervals with less need for geometric correction. However, compared to video imagery, detecting spatiotemporal changes in satellite imagery is more complex and is easily influenced by other dynamic phenomena such as cloud

cover. This is primarily due to differing observation patterns and the longer intervals between image captures. Furthermore, exploring dynamic spatiotemporal information for smoke detection requires additional data processing and more complex methods, especially with sun-synchronous satellites. Therefore, smoke detection in static satellite imagery is more feasible and remains the focus of this thesis.

1.2 Research Questions

Motivated by the urgent need to enhance fire disaster mitigation, this thesis aims to develop innovative approaches for the accurate and timely detection of EF smoke. Consequently, this thesis is structured around three interrelated research questions, which address the limitations of previous research and work collectively towards achieving this goal:

- ***RQ-1 How to design a lightweight DL model that can potentially be deployed onboard SmallSats for accurate EF smoke detection using multispectral satellite imagery?***

The first research question can be further divided into two sub-questions:

- *RQ-1.1 How to design such a lightweight DL model to achieve competitive accuracy with state-of-the-art models?*
- *RQ-1.2 Can the inclusion of IR bands in satellite imagery effectively improve smoke detection accuracy?*

The first sub-question seeks to investigate and employ advanced DL techniques to create a lightweight model without sacrificing accuracy. The model’s performance will be fairly compared to state-of-the-art models using the existing USTC.SmokeRS training dataset. The second sub-question aims to investigate the effectiveness of multispectral satellite imagery with additional IR bands in enhancing smoke detection accuracy. Such investigation has yet to be conducted with solid evidence in the literature. To support this study, a multispectral satellite imagery training dataset with medium to high spatial resolution must be created to ensure effectiveness in detecting EF smoke. In terms of the spectral bands, the NIR and SWIR regions, which are less sensitive to temperatures compared to MWIR and TIR, are the primary focus due to the rapid cooling of smoke plumes.

- ***RQ-2 How can smoke detection accuracy be improved more effectively using multispectral satellite imagery, beyond simply including additional IR bands in model training?***

This research question can be resolved by addressing two sub-questions:

- *RQ-2.1: What essential information in multispectral satellite imagery is likely to improve smoke detection accuracy more effectively?*
- *RQ-2.2: How can such information be precisely extracted and used to improve the accuracy of DL models for smoke detection?*

The second research question focuses on developing an innovative approach to enhance the accuracy of DL models by identifying and utilising essential information for smoke detection from multispectral imagery. Compared to simply training DL models with multispectral satellite imagery data, this approach explores more refined techniques for customised feature extraction tailored to the task. Specifically, the aim is to boost accuracy more effectively.

- ***RQ-3 How to leverage transfer learning for fast model development across multiple satellites for timely smoke detection, given insufficient labelled training data from some satellites?***

Transfer learning has been widely employed to update DL models when training data is insufficient in the target domain. Adapting an already-trained model from one satellite to another saves the time and effort required for designing and training models from scratch for the target satellites. While transfer learning seems an ideal solution to fast model development across multiple satellites, its feasibility depends on resolving the following sub-questions:

- *RQ-3.1 How to address variant spectral bands and numbers of spectral bands in the imagery data from different satellite sensors?*
- *RQ-3.2 How to handle data distribution disparities in the imagery data from different sensors?*
- *RQ-3.3 How to achieve good accuracy with limited training data from the new sensors?*

This research question aims to propose an innovative transfer learning approach to address all three sub-questions. After resolving the first and second research questions, a multispectral satellite imagery training dataset will be available to serve as the source domain. A pretrained lightweight DL model on this dataset

will also be ready for transfer. Additionally, a new labelled multispectral satellite imagery training dataset must be constructed to serve as the target domain. This dataset must be derived from a sensor different from the source sensor, featuring different spectral radiometric characteristics, a varying number of spectral bands, and minimal training samples.

1.3 Overview of the Contributions to the Field

This thesis has conducted three studies to answer the research questions presented in Section 1.2. These studies have made the following contributions to the field of satellite-based smoke detection:

1. Contributions of study one (based on [27]):
 - Development of a lightweight DL model called **Variant Input Bands for Smoke Detection (VIB_SD)**, which has the potential for onboard deployment on SmallSats for NRT smoke detection. **VIB_SD** achieves competitive accuracy (93.57% versus 96.22%) with the state-of-the-art model SAFA, while using less than 2% of its parameters (1.66 million versus 84.2 million) when trained on the MODIS RGB imagery dataset USTC_SmokeRS. (For a fair comparison, the USTC_SmokeRS imagery data was resized to $224 \times 224 \times 3$ to align with the training settings of SmokeNet and SAFA.)
 - Creation of a multispectral satellite imagery smoke detection training dataset named **Landsat6c**, consisting of 1836 six-band (i.e., RGB, NIR, SWIR_1, SWIR_2) 256×256 images. **Landsat6c** is the first multispectral satellite imagery dataset designed for scene-level smoke detection. With a 30-metre spatial resolution, **Landsat6c** ensures that EF smoke plumes are captured in the imagery and can be effectively detected by trained DL models. **Landsat6c** was further expanded for study three to provide additional training samples for each class in the source domain.
 - Demonstration of the significance of using multispectral satellite imagery with additional IR bands for smoke detection through comprehensive experiments employing **VIB_SD** and **Landsat6c**.
 - The effectiveness of deploying **VIB_SD** onboard SmallSats has been experimentally verified. **VIB_SD** served as the model prototype for a SmartSat CRC-funded project, which simulated NRT smoke scene detection onboard

Kanyini [55], a cube satellite (CubeSat) co-designed by SmartSat CRC and the South Australia (SA) government.

2. Contributions of study two (based on [56]):

- Invention of a DL module named **Input Amplification (IA)**, which automatically learns class-oriented spectral patterns from multispectral imagery for smoke detection. IA can be seamlessly integrated with DL models, enabling them to extract and use these class-oriented spectral patterns to effectively improve smoke detection accuracy.
- Creation of a new lightweight DL model named **IA_VIB_SD**, which integrates IA with **VIB_SD**. When trained on **Landsat6c**, **IA_VIB_SD** achieves significantly improved accuracy compared to **VIB_SD** (85.33% versus 81.79%), which simply takes the multispectral imagery training data as input.

3. Contributions of study three (based on [57]):

- Introduction of a novel transfer learning approach aided by IA. This approach robustly adapts a pretrained model using imagery from one satellite to another, regardless of variations in spectral radiometric characteristics or the number of spectral bands. It is the first transfer learning approach in the literature to address variant multispectral imagery in both source and target satellites without incorporating new DL models or intermediate domains. This method significantly facilitates the use of multiple existing satellites or satellite constellations for the timely and accurate detection of EF smoke.
- Creation of another multispectral satellite imagery training dataset named **Sentinel7c** for smoke detection. **Sentinel7c** comprises 351 seven-band (i.e., RGB, NIR, NIR_2, SWIR_2, and SWIR_3) 256×256 images. With a 10 metre spatial resolution, **Sentinel7c** further enhances the capabilities of trained models in detecting EF smoke.

1.4 Structure of the Thesis

This thesis continues with the following chapters:

- Chapter 2 introduces the essential technical concepts and theoretical background relevant to the work presented in this thesis.

- Chapter 3 provides a comprehensive review of smoke detection approaches, focusing primarily on satellite-based methods and covering non-satellite platforms, including UAVs and surveillance cameras. Additionally, it reviews transfer learning techniques, with an emphasis on cross-sensor transfer learning, where data distribution disparity is a significant concern.
- Chapter 4 proposes the lightweight DL model `VIB.SD` and introduces the multi-spectral satellite imagery training dataset `Landsat6c` for scene-level smoke detection. It also presents the results of investigating the impact of using IR bands on smoke detection by employing `Landsat6c` and `VIB.SD` [27].
- Chapter 5 presents the `IA` module, which enables DL models to automatically learn class-oriented spectral patterns from multispectral satellite imagery to enhance smoke detection accuracy [56].
- Chapter 6 presents a novel cross-sensor transfer learning approach to facilitate the use of multiple satellites for smoke detection. This transfer learning approach adapts class-oriented spectral patterns rather than the data distribution. The `IA` module aids in both learning and adapting these spectral patterns [57].
- Chapter 7 concludes the thesis, discussing the practical implications, limitations of the research, and future research directions.

Chapter 2

Technical and Theoretical Background

This chapter addresses the technical and theoretical concepts related to the research of the thesis, including satellites and sensors, spectral patterns, and transfer learning.

2.1 Satellites and Sensors

Satellites and sensors are distinct yet complementary concepts often employed together in the realms of space and remote sensing. Satellites typically carry multiple sensors as part of their payload. For instance, the Terra satellite is equipped with MODIS, Multi-angle Imaging SpectroRadiometer, and Clouds and the Earth's Radiant Energy System [58]. Similarly, Landsat 8 carries OLI and the Thermal Infrared Sensor (TIRS) [59].

These sensors collect data about Earth and other celestial bodies. The data collected by these satellite-borne sensors is transmitted back to Earth for analysis and utilisation in various applications, including weather forecasting, environmental monitoring, and scientific research [60, 61].

2.1.1 Satellites

Satellites are artificial objects that are intentionally placed into orbit around Earth or other celestial bodies. Satellites serve various purposes, including communication, weather monitoring, navigation, Earth observation, scientific research, and military applications. A satellite typically consists of a power source (such as solar panels), communication systems, propulsion systems, control systems, and various payloads, which can include sensors. There are different types of satellites based on their function, such as communication satellites, weather satellites, navigation satellites (e.g., GPS), and Earth observation satellites. Satellites operate in various orbits, including low Earth orbit (LEO), medium Earth orbit (MEO), and geostationary orbit (GEO), depending on their mission requirements. Table 2.1 shows that each orbit type serves specific purposes based on altitude and characteristics [62].

It is noteworthy that sun-synchronous satellites, as a specific type of LEO satellites, have orbits designed to ensure that they pass over the same part of the Earth at roughly the same local solar time. Such orbits, typically between 600 km to 800 km above the Earth's surface with inclinations around 98 degrees, revolve around the Earth at the same rate that the Earth orbits the Sun, maintaining a consistent relationship with the position of the Sun. This orbital characteristic ensures consistent illumination

Table 2.1: Summary of LEO, MEO, and GEO orbit types and their characteristics.

Orbit Type	Altitude Range (km)	Orbital Period (hours)	Applications	Examples
LEO	160 - 2,000	1.5 - 2	Earth observation, scientific missions, space stations, communication	ISS ¹ , Hubble Space Telescope, Dove Satellites, Iridium Satellites
MEO	2,000 - 35,786	2 - 12	Navigation, communication, scientific missions	GPS Satellites, Galileo Satellites, GLONASS Satellites, O3b Satellites
GEO	35,786	24	Communication, weather monitoring, broadcasting	GOES, SES Satellites, Intelsat Satellites, Himawari Satellites

¹ International Space Station.

conditions in imagery captured in a specific time frame over successive years or a particular area over a sequence of days. This makes sun-synchronous satellites ideal for Earth observation, environmental monitoring, and reconnaissance. Some well-known sun-synchronous satellites include the NOAA sun-synchronous satellite family (e.g., NOAA 6/15/18/19, Suomi NPP), Terra, Aqua, Landsat 7/8/9, and Sentinel-1/2 [63].

Most Earth observation satellites, including sun-synchronous satellites, tend to be medium (typically between 500 kg to 1,000 kg) to large (typically more than 1,000 kg) satellites because they require substantial instrumentation and power to perform their missions effectively. Despite being powerful and capable, their complexity necessitates extensive testing and stringent quality control measures, leading to higher launching costs due to complicated launch logistics [64] and overall mission costs [65]. For instance, the Terra satellite (with a launch mass of approximately 4,864 kg), a key component of the Earth Observing System of NASA, had a total mission cost estimated at around 1.3 billion USD, which included development, launch, and operations [66, 67]. In addition, these larger satellites often require longer development and construction times, which can delay deployment and limit technological innovation [68]. Once in orbit, large satellites are less adaptable to changes in mission requirements and can be challenging to upgrade or repair, often resulting in a reliance on outdated technology over their operational lifespan [69].

In contrast, SmallSats offer numerous advantages that address these disadvantages. SmallSats refer to satellites weighing lower than 500 kg. They can be further divided into nanosats (less than 10 kg), microsattelites (10-100 kg), and minisattelites (100-500 kg). The reduced complexity of SmallSats simplifies testing and integration processes, decreasing overall mission expenses [70]. For instance, CubeSats, a standardised type of nanosats typically measured in multiples of 10x10x10 cm units called “U” (weighing about 1.33 kg per unit), can be built and launched for as little as 50,000 USD to 500,000 USD depending on their complexity and the launch provider [71]. The lower cost and quicker development cycles of SmallSats enable faster deployment and more frequent technological updates [72]. They are also more flexible in terms of launch options, as their smaller size allows them to be launched as secondary payloads or in large numbers within a single mission [73]. This flexibility extends to their operational adaptability, as SmallSats can be easily reconfigured, replaced, or augmented with newer technology [74]. Consequently, SmallSats provide a cost-effective, agile, and innovative alternative to their larger counterparts, making space more accessible and versatile for a wider range of applications and stakeholders [75, 76]. SmallSats are typically deployed in LEO due to several factors, including lower launch costs, reduced latency for communications, and the suitability of LEO for many SmallSat missions such as Earth observation, scientific research, and technology demonstrations.[77, 78, 79, 80, 81].

A single satellite often suffers from limited coverage, lengthy revisit times (the period between a satellite visiting the same location on Earth twice), and potential single points of failure [82]. Satellite constellations, consisting of multiple satellites working together in coordinated orbits, were developed to provide enhanced coverage, improved revisit time, and greater reliability [83]. These satellites typically employ identical or near-identical sensors to ensure consistent data quality, simplify calibration and operation, reduce costs, and enhance data integration and long-term mission reliability. The earliest satellite constellations can be traced back to the Iridium constellation launched in the late 1990s and the GPS constellation operational since the 1990s [84, 85]. More and more satellite constellations, notably many CubeSat constellations, have been launched or planned to be launched into space to accomplish more complex tasks. For example, the recent mega-constellations like OneWeb and Starlink aim to provide global broadband internet, demonstrating the growing capabilities and applications of satellite constellations [86, 87]. Planet Labs, an American private Earth imaging company, operates a CubeSat constellation called Doves, consisting of more than 430 Dove

and SuperDove CubeSats in sun-synchronous orbits. These satellites are equipped with sensors, including a NIR band with a 3.7-meter spatial resolution, suitable for fire and smoke detection [88, 89]. PlanetScope, the data product from these satellites, has been used in various environmental monitoring applications, including fire detection. A European company called OroraTech has planned to launch 100 CubeSats equipped with high-resolution thermal imagers by 2026, jointly achieving an average 30-minute revisit time [90]. These advancements mark a significant evolution in satellite technology, enabling continuous global coverage, real-time data collection, and enhanced communication networks, which are crucial for various applications such as disaster management, environmental monitoring, and global navigation systems [91, 92].

As an alternative to satellite constellations, multiple satellites with varied sensors can be employed collectively to enhance temporal and geographical coverage, combine complementary data, improve accuracy, and ensure consistent monitoring, particularly during sensor downtimes or outages. For instance, NOAA uses the Hazard Mapping System to monitor wildfires across north and central America by integrating data from multiple satellites such as NASA’s Aqua and Terra with the MODIS sensor, Suomi-NPP and NOAA-20 with Visible Infrared Imaging Radiometer Suite (VIIRS), and geostationary satellites like GOES equipped with the Advanced Baseline Imager [93]. Although using multiple satellites with different sensors typically requires significant manual data engineering due to disparities in spectral and radiometric characteristics, it offers several advantages compared to satellite constellations:

- **Data Availability:** Existing satellites like Landsat 8 and Sentinel-2 provide a rich repository of observational data spanning many years, enabling the collection and labelling of data from historical fire smoke events [94, 95]. Moreover, these satellites can capture diverse fire smoke types, which can contribute to training more robust DL models. In contrast, satellite constellations typically have shorter operational histories, limiting their ability to provide comprehensive datasets [96]. Furthermore, most constellations are privately owned and restrict public access to their data, reducing their utility for research purposes [97].
- **Cost Efficiency:** Utilising the infrastructure of existing satellites is far more cost-effective than launching and maintaining new satellite constellations [98]. The expense of deploying and operating constellations is significantly higher than using satellites already in orbit, making the latter a more resource-efficient solution for fire monitoring [99].
- **Technological Contribution:** Tackling the challenges associated with integrat-

ing data from multiple existing satellites can lead to technological advancements that enhance data processing and utilisation. For example, developing methods to adapt DL models to inconsistently distributed data from various sensors could improve fire detection systems and support faster deployment across constellations [100].

- **Immediate Implementation:** Since existing satellites are already operational, fire detection strategies can be implemented immediately. This is vital for timely responses to wildfires, potentially reducing damage and saving lives [101]. The urgency of fire detection, combined with the availability of satellite data, underscores the importance of leveraging these assets [102].

One objective of this thesis is to investigate mechanisms for fast model development that leverage data from multiple existing satellites—such as Landsat 5/8 and Sentinel-2 A/B—for accurate and timely smoke detection, while minimising data requirements and manual engineering efforts. Such mechanisms can also be adapted for satellite constellations, enabling new SmallSat constellations to rapidly and efficiently perform smoke detection without the need for complex integration or concerns about the spectral radiometric characteristics of the sensors.

2.1.2 Sensors

Sensors are devices or instruments that detect and measure physical properties, such as light, heat, motion, or sound, and convert these measurements into data. Sensors can be deployed on various platforms, including satellites, aircraft, ground-based stations, and handheld devices. When used on satellites, sensors collect data from space for remote sensing applications [103]. Sensors gather information about the environment or the object being observed. In the context of satellites, sensors collect data for various applications, including imaging, temperature measurement, atmospheric analysis, and more. A sensor typically includes a detector (which senses the physical property), electronics to process the signal, and sometimes systems for data storage or transmission. Sensors can be divided into active sensors and passive sensors based on how they gather information and the type of energy they rely on:

- **Active Sensors** have an internal source of energy, which emits energy (usually in the form of electromagnetic waves) towards the target, then measures the reflected or backscattered energy and the time it takes for the energy to return

after interacting with the target. This information can be used to determine various properties of the target, such as distance, speed, and material characteristics [104]. Examples of active sensors include *Radar* and *SAR* (Synthetic Aperture Radar), which uses radio waves to detect objects and measure their distance and speed, and *lidar*, which uses laser pulses to measure distances and create high-resolution maps of surfaces. Active sensors can operate day or night since they do not rely on external light sources. Their energy pulses can usually penetrate through clouds, smoke, and certain materials, providing data in various weather conditions. Active sensors are widely applied in the areas of weather forecasting (e.g., weather radars), topographic mapping (e.g., lidar), and military and defence (e.g., radar systems).

- **Passive Sensors** do not have their own energy source and rely on external sources of energy, primarily sunlight. These sensors measure the naturally occurring energy that is reflected from or emitted by the Earth's surface or atmosphere. This includes visible light, IR radiation, and thermal radiation. Examples of passive sensors include *optical sensors*, which capture visible light and create RGB images similar to photographs, *thermal sensors* that detect IR radiation (usually used to measure temperature) emitted by objects, and *multispectral and hyperspectral sensors*, which capture data across multiple wavelengths, providing detailed information about the composition and properties of objects. Passive sensors are typically simpler and less expensive than active sensors and can provide a wide range of information about the environment, including vegetation health, land use, and atmospheric conditions (e.g., the presence of smoke) [104]. However, the applications of passive sensors are limited by the availability of natural light or thermal emissions.

Sensors have the following key attributes:

- **Spectral Bands** refer to specific ranges of wavelengths in the electromagnetic spectrum that sensors on satellites or other imaging devices can detect. Each spectral band corresponds to a particular range of wavelengths and can capture unique information about the Earth's surface and atmosphere based on how different materials reflect or emit electromagnetic energy at those wavelengths. *Multispectral bands*, in particular, are a subset of spectral bands, covering various parts of the spectrum, typically including visible, NIR, and sometimes SWIR regions [105].
- **Spectral Resolution** is the ability of a sensor to distinguish between different wavelengths of electromagnetic radiation. It determines how many spectral bands

the sensor can capture and how narrow these bands are. For example, the spectral resolution of multispectral sensors, which capture data in a few broad spectral bands, is lower compared to hyperspectral sensors, which capture data in many narrow spectral bands [106].

- **Spatial Resolution** refers to the smallest distinguishable distance between two objects on the Earth’s surface that can be resolved by a sensor, typically measured in meters. Higher spatial resolution yields more detailed images, allowing for finer distinctions between objects. Different bands of a sensor may have different spatial resolutions. For example, panchromatic bands are often designed with the highest spatial resolution for detailed imaging; multispectral bands typically have lower spatial resolutions compared to panchromatic bands; TIR bands often have the lowest spatial resolutions due to the longer wavelengths and the need for larger detector elements to capture sufficient thermal energy. The altitude of the satellite’s orbit plays a predominant role in determining a sensor’s overall spatial resolution. Satellite sensors in lower orbits like LEO generally provide more detailed imagery compared to those in higher orbits like GEO. For instance, the WorldView-3 satellite, operating in low Earth orbit at approximately 617 km, can capture images with a high spatial resolution of 1.24 meters in its visible and NIR bands, and 0.31 meters in its panchromatic band [107]. The OLI sensor on Landsat 8, orbiting at 705 km, offers a spatial resolution of 30 metres in its multispectral bands [108]. In contrast, the Himawari-8 AHI sensor, operating in GEO at approximately 35,786 km, provides a spatial resolution of 500-2000 metres across its bands [109]. The design of the sensors also plays an important role in affecting their spatial resolutions. For example, the MODIS sensors on Terra and Aqua, operating at the same altitude as OLI, have a lower spatial resolution of 250 metres in some of their visible and NIR bands due to their broader swath width and different mission objectives [110]. Notably, the High Resolution Optical Imager, carried by the GEO satellite Gaofen-4 and enhanced by state-of-the-art technology, can provide a spatial resolution of 50 metres in its visible bands [111]. Such a high spatial resolution is phenomenal and rare among GEO sensors. However, the data access to Gaofen-4 is restricted to Chinese governmental agencies and research institutions within China, with limited and controlled access to their commercial partners or international collaborators.
- **Swath Width** is the width of the ground area that a satellite sensor can capture in a single pass. It is typically measured in kilometres. A wider swath width allows a sensor to cover a larger area of the Earth’s surface in a single pass, which

is useful for applications such as land cover mapping and disaster monitoring. For example, the Landsat 8 OLI sensor has a swath width of 185 km, the Sentinel-2 Multispectral Instrument (MSI) has a swath width of 290 km, while the MODIS sensor on Terra and Aqua has a swath width of 2,330 km [112].

- **Temporal Resolution** refers to the time interval at which a sensor can capture data over the same location. It is associated with, but different from, the satellite’s revisit time, which implies how quickly a satellite passes over the same point on the Earth’s surface. For example, the MODIS sensors on Terra and Aqua have a temporal resolution of about half a day (considering both satellites working together), whereas each satellite has a revisit time of 1-2 days. The Landsat 8 OLI’s temporal resolution is 16 days, determined by the satellite’s revisit time of 16 days. Himawari-8 AHI has a temporal resolution of 10 minutes for full-disk images and 2.5 minutes for specific areas, despite the satellite remaining static over the same location on Earth. Temporal resolution is crucial for monitoring dynamic processes, as it determines how often data is collected for the same location [113]. This attribute is essential for applications such as weather monitoring, crop growth analysis, and environmental change detection. The temporal resolution of a sensor or the revisit time of a satellite is predominantly affected by the orbit altitude and the swath width of the sensor [114, 92, 115].
- **Radiometric Resolution** is the ability of a sensor to distinguish between different levels of radiance or reflectance. It determines how finely a sensor can quantify the detected energy, typically measured in bits. Higher radiometric resolution allows for more precise detection of subtle differences in energy levels. For example, an 8-bit sensor can distinguish 256 different radiant levels, while a 12-bit sensor can distinguish 4,096 levels. The Landsat 8 OLI sensor has a radiometric resolution of 12 bits, enabling it to detect 4,096 different levels of radiance. In comparison, the Sentinel-2 sensor has a radiometric resolution of 12 bits for most of its bands, but it also includes bands with 10-bit radiometric resolution, distinguishing 1,024 levels. This variation in radiometric resolution allows Sentinel-2 to balance between precision and data volume, making it suitable for detailed surface analysis as well as broader environmental monitoring [116].

Table 2.2 compares some attributes of three different sensors onboard different sun-synchronous satellites.

Table 2.2: Comparison of some attributes in different satellite sensors

Attribute	MODIS Terra/Aqua	OLI Landsat 8	MSI Sentinel-2 A/B
Orbit Altitude (km)	705	705	786
Swath Width (km)	2330	185	290
Temporal Resolution (day)	0.5	16	3-5
Number of Spectral Bands	36	11	13
Blue Bandwidth (nm)	459-479	452-512	458-523
Blue Spatial Resolution (m)	250	30	10

In summary, satellites are platforms that carry sensors into space, while sensors are the instruments that collect data. The combination of satellites and sensors enables remote sensing and valuable information collection from space. Often, the names of the sensors are used to refer to the satellites when the context is clear. In this thesis, the sensor names and satellite names are used interchangeably. For example, MODIS may refer to the satellite missions Aqua or Terra.

2.1.3 Fire and Smoke Detection Using Satellites

Satellites have been used to monitor fire-related events, including pre-fire observations (e.g., fire risk monitoring), active fire detection (e.g., hot-spot detection, flaming or smouldering fire detection, smoke detection), and post-fire analysis (e.g., Burned areas detection) [28].

Hot-spot detection using satellite sensors is the predominant method of active fire detection, and primarily relies on MWIR and TIR bands due to their sensitivity to high temperatures [117, 118]. The bandwidths of MWIR typically cover 3.55 to 3.93 μm , while the TIR bands range from 10.5 to 12.5 μm .

Many sensors are equipped with onboard hot-spot detection modules. Table 2.3 presents examples of these sensors along with the specific bands used in their hot-spot detection modules [119, 120].

Table 2.4 summarises satellite sensors suitable for fire detection [28].

Table 2.3: Examples of sensors and their hot-spot detection bands

Sensor	Band	Wavelength Range (μm)
MODIS	Band 21 (MWIR)	3.929–3.989
MODIS	Band 31 (TIR)	10.78–11.28
VIIRS	I4 Band (MWIR)	3.55–3.93
VIIRS	I5 Band (TIR)	10.5–12.4
GOES-16/17	MWIR	~3.9
GOES-16/17	TIR	~10.7
AHI	Band 7 (MWIR)	3.8
AHI	Band 13 (TIR)	10.4
AHI	Band 14 (TIR)	11.2

Table 2.4: Comparison of satellite sensors suitable for fire detection [28, Table 1]

Satellite/ Sensor	Spatial Resolution*	Temporal Resolution	Latency	Typical Detection Target	Cost
Himawari 8-AHI	2 km	10 min	17 min**	Hot-spots	Free
MODIS	500 m/1 km	1 day	10-30 min	Hot-spots	Free
AVHRR	1.09 km	12 h	10-30 min	Hot-spots (day and night)	Free
Suomi-VIIRS	375/750 m	12 h	10-30 min	Hot-spots (day and night)	Free
Landsat-7/8	30/60/100 m	8-16 days	Initial TIRS: <12h Tier1/Tier2: 14-26 days	Burned areas	Free
Sentinel-1 A/B	10 m	6 days	2 - 12 h (average 5-6 h)	Cloud-free burned areas	Free
Sentinel-2 A/B	10/20 m	3-5 days	2 - 12 h (average 5-6 h)	Burned areas	Free
Sentinel-3 A/B	1 km	1 day	2 - 12 h (average 5-6 h)	Hot-spots (day and night)	Free
RadarSAT	3 m	1 day	4 h	Cloud-free burned areas	12 USD/km ²
WorldView-3	0.37/1.24/3.7 m	< 1 day	n/a	Burned areas	58 USD/km ²

* Spatial resolutions of the respective relative bands (e.g., MWIR)

** Himawari 8/9 hot-spots are generally loaded onto the Sentinel Hot-spots platform approx. 17 minutes after acquisition time (in rare cases up to 30 min)

As demonstrated in Table 2.3 and Table 2.4, the sensors used for active fire detection are typically passive sensors, either sun-synchronous or geostationary. The relatively higher spatial resolutions of sun-synchronous sensors enable them to detect smaller fires more effectively. However, their low temporal resolutions (generally more than 12 hours) implies that they may fail to detect burning fires promptly. In contrast, the high temporal resolutions (normally less than 30 minutes) of geostationary sensors allows them for NRT active fires detection. Nevertheless, their coarser spatial resolutions, particularly in the MWIR and TIR bands, indicate that they are prone to overlook nascent fires, which typically have smaller sizes [28, 117, 118].

Apart from the contradiction between temporal resolution and spatial resolution, the accuracy of onboard hot-spot detection is also affected by the following factors: atmospheric interference, sensor limitations, and data processing constraints. Atmospheric conditions such as clouds and aerosols can obstruct the sensor’s view, leading to false negatives (FNs) or inaccuracies in hotspot identification [121]. Moreover, the sensitivity and resolution of the sensors onboard satellites are often limited by their design and the operational constraints of the satellite platform, which can result in false positives or missed detections [122]. Specific factors causing false positives include hot backgrounds such as deserts or urban areas during summer, which can be mistakenly identified as fires. On the other hand, nascent fires burning at lower temperatures may not be detected, resulting in FNs [123].

Algorithms such as Fire Radiative Power (FRP) and spectral indices like Normalised Difference Vegetation Index (NDVI) are commonly used in hotspot detection. While FRP estimates the energy released by fires and helps in quantifying their intensity, it can be misled by non-fire heat sources, thus generating false positives. Spectral indices like NDVI can help distinguish vegetation from burned areas but may fail in heterogeneous landscapes where different types of ground cover exist [124]. Additionally, these algorithms often require complex calibration and validation processes, and their accuracy can be compromised by the resolution of the satellite imagery and the presence of noise in the data [122].

Furthermore, onboard satellite detection systems are constrained by limited power supply, storage space, and computational capability. These limitations often necessitate the exclusion of complicated methods such as DL models, which are typically resource-intensive [25, 125]. Although some large satellites may have adequate resources to support such onboard DL models, many current systems do not have the capacity

for these demanding computational tasks [126]. These factors collectively reduce the reliability of onboard satellite hotspot detection systems and highlight the need for complementary ground-based processing methods.

Free from the constraints on power supply, storage space, and computational capacity, ground-based approaches offer much more flexibility, allowing more fire-related objects (e.g., fire flames, smoke plumes) to be detected with a wider range of methods. For example:

- **False colour composition** enhances the visibility of fire and smoke by using different spectral bands to create composite images that highlight these features. By combining IR and visible bands, false colour compositions can distinguish between hot surfaces, active fires, and smoke plumes. This technique is particularly effective in identifying fire locations and extents, providing clear visual contrasts that are easier to interpret than single-band images [127, 128].
- **Multi-threshold methods** apply multiple threshold values to different spectral bands and pseudo bands to detect fire and smoke. For instance, a combination of TIR and MWIR thresholds can be used to identify active fires, while visible and NIR thresholds can help detect smoke. Multi-threshold methods can reduce false positives by cross-referencing multiple criteria, thus improving detection accuracy [129, 130].
- **DL classification methods** such as CNNs, can classify satellite images to detect fire and smoke. These models are trained on large datasets of labelled images, learning to identify features associated with fire and smoke. DL classification methods offer high accuracy and can adapt to different types of landscapes and atmospheric conditions. They can distinguish between fire, smoke, clouds, and other objects, reducing false positives and FNs [131, 132].
- **DL segmentation methods** such as U-Net and Mask R-CNN, can identify and delineate the boundaries of fire and smoke within an image. These models provide pixel-level accuracy, allowing for precise mapping of fire extents and smoke plumes. Segmentation methods are particularly useful for detailed analysis and monitoring of fire progression and its impact on the environment [133, 134].

Ground-based approaches have significant advantages over onboard systems. They allow for the use of more complex and computationally intensive algorithms, such as DL

models, which can achieve higher accuracy and reliability in detecting and classifying fire events. These approaches can leverage extensive computational resources and large labelled datasets, facilitating the development and refinement of advanced detection algorithms [135, 136]. Moreover, the flexibility to update and improve these models continuously ensures that the detection systems can adapt to new types of data and evolving environmental conditions.

However, one drawback of ground-based approaches is the latency caused by the time needed for data downlink, processing and dissemination, which can lead to significant delays in detecting and responding to fires. Additionally, effectively training DL models requires large amounts of labelled data, which can be challenging to obtain and maintain. This reliance on extensive datasets can limit the scalability and adaptability of these models in different regions or under varying conditions. Furthermore, the computational resources needed for processing and analysing satellite data on the ground can be substantial, necessitating powerful infrastructure and continuous maintenance [136].

Whether using onboard satellite or ground-based methods, it is important to note that smoke detection is influenced by several factors.

Firstly, smoke detection is heavily influenced by the chemical characteristics of the smoke, which ultimately determine its spectral signature in satellite imagery pixels. Smoke is composed of various gases, particulate matter, and aerosols, with common constituents including carbon dioxide (CO_2), carbon monoxide (CO), methane (CH_4), volatile organic compounds, and particulate matter ($\text{PM}_{2.5}$ and PM_{10}) [137]. The spectral properties of these components affect how smoke is detected in different bands of satellite sensors, especially in the visible, IR, and TIR spectra [138].

Additionally, different types of smoke arise from various fuel types, moisture levels, and environmental conditions, including wind. For instance, smoke from burning peatlands, forests, and grasslands can vary significantly in composition and density, affecting the visibility and detectability of smoke in satellite images [139]. Moisture content in the fuel also influences the amount and type of smoke produced; wetter fuels produce more smoke due to incomplete combustion [140]. Wind conditions can further disperse smoke, impacting its concentration and the area it covers, affecting satellite detection capabilities [141].

Furthermore, the altitude to which smoke elevates is another critical factor in its detectability. Smoke can rise to various altitudes depending on the intensity of the fire and atmospheric conditions. Smoke plumes can ascend to the lower stratosphere during intense wildfires, but most smoke resides in the troposphere [142]. Over time, smoke particles disperse and settle, eventually becoming less visible in satellite imagery as they mix with the surrounding atmosphere and undergo chemical transformations [143].

2.2 Spectral Patterns

A “Spectral pattern” can be generally described as a pattern in the pixel values across the spectral bands of remotely sensed imagery, indicative of certain characteristics of the object, material, or phenomenon within the pixel. Although the term “Spectral pattern” has been frequently mentioned in the fields of remote sensing and computer vision [see 144, 145, 146, 147], its definition remains vague in the literature.

Spectral patterns are usually related to but distinct from spectral signatures. The latter refers to the characteristic ways in which objects, materials, or phenomena reflect, absorb, or emit electromagnetic radiation across various wavelengths [148]. In contrast, spectral patterns are potential rules that can be uncovered, often based on the spectral signatures of these objects, materials, or phenomena, to differentiate them. Figure 2.1 demonstrates the spectral signatures of vegetation, water, and soil, highlighting how each reflects electromagnetic radiation differently across various wavelengths.

Spectral indices, such as NDVI, Normalised Burn Ratio (NBR), and Normalised Difference Built-Up Index (NDBI), can be considered a special group of spectral patterns. Spectral indices are calculated using designated spectral bands and formulas, and their values present different patterns against the original bands, indicating the existence or occurrence of certain objects or events.

Table 2.5 presents several common spectral indices utilised in remote sensing literature. It is important to note that spectral patterns should indicate the type of a pixel and are typically not simple linear combinations of the selected band values.

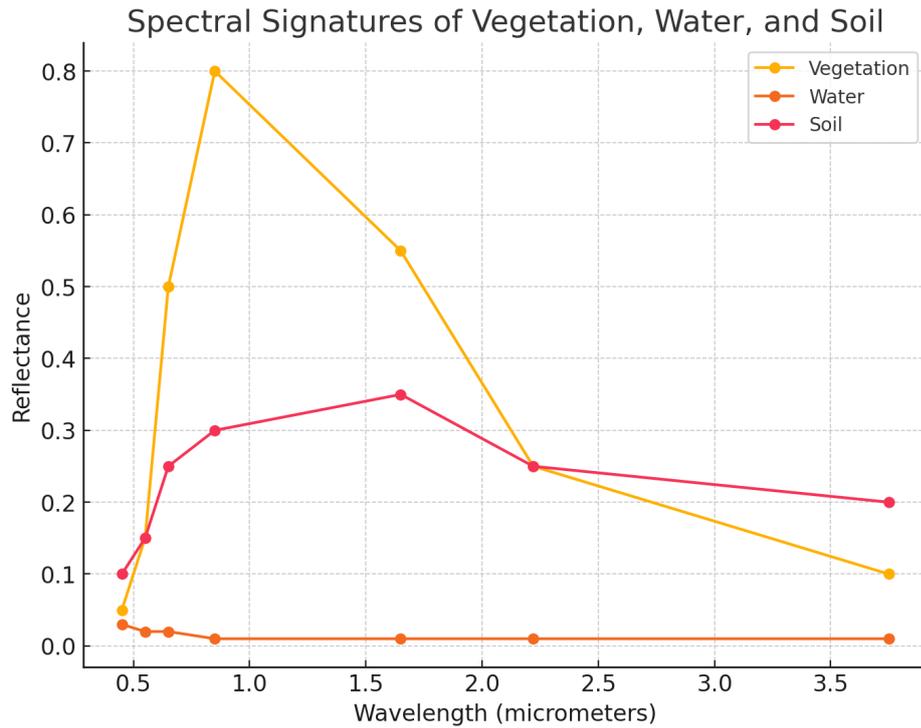


Figure 2.1: Spectral signatures of vegetation, water, and soil across different wavelengths, adapted from [149].

Table 2.5: Some spectral indices used in remote sensing

Index	Formula	Objective	References
			[150]
NDVI	$\frac{NIR-Red}{NIR+Red}$	Highlight vegetation	[151]
			[152]
			[153]
			[152]
NBR	$\frac{NIR-SWIR}{NIR+SWIR}$	Highlight burnt areas	[154]
			[155]
			[156]
NDBI	$\frac{SWIR-NIR}{SWIR+NIR}$	Highlight urban areas	[157]
			[158]

Similarly, the multi-threshold values derived from multiple bands to distinguish fire or smoke pixels from pixels belonging to other objects can be considered as spectral patterns.

To make discussions precise, this thesis gives a formal definition of “spectral pattern” as follows:

Definition 2.2.1. Given an input image $X \in \mathbb{R}^{W \times H \times C}$, where $W \in \mathbb{N}$, $H \in \mathbb{N}$, and $C \in \mathbb{N}$ represent the width, height, and number of spectral channels of X , for any pixel $P_{(i,j)} = (p_{(i,j)}^1, \dots, p_{(i,j)}^k, \dots, p_{(i,j)}^C)$ in X , where $i \in [0, W) \cap \mathbb{N}$ and $j \in [0, H) \cap \mathbb{N}$ are the indices of the pixel, $p_{(i,j)}^k \in \mathbb{R}$ is the value of the pixel $P_{(i,j)}$ in the k -th channel, and $k \in [1, C] \cap \mathbb{N}$, a spectral pattern refers to a semantic mapping SP , which maps $P_{(i,j)}$ to a new value $p_{(i,j)}^{SP} \in \mathbb{R}$ that is indicative to certain physical/chemical properties of $P_{(i,j)}$:

$$SP: P_{(i,j)} \mapsto p_{(i,j)}^{SP} \quad (2.2.1)$$

For example, a spectral pattern indicative of smoke may map smoke pixels to values close to 1 and other pixels close to 0, assuming all original pixel values are normalised to the range $[0, 1]$. The aforementioned spectral indices and multi-threshold values align with Definition 2.2.1.

2.3 Transfer Learning

Transfer learning is a machine learning technique where a model developed for a particular task is reused as the starting point for a model on a second task. Formally, transfer learning can be defined as follows:

Definition 2.3.1. A *domain* is a set $\mathcal{D} = \{\mathcal{X}, P(X)\}$ consisting of a feature space \mathcal{X} and a marginal probability distribution $P(X)$, where $X = \{x_1, x_2, \dots, x_n\} \in \mathcal{X}$.

Definition 2.3.2. A *task* is a set $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ consisting of a label space \mathcal{Y} and a predictive function $f(\cdot)$, which is typically learned from the training data consisting of pairs $\{(x_i, y_i)\}$, where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.

Definition 2.3.3. Given a source domain $\mathcal{D}_S = \{\mathcal{X}_S, P_S(X)\}$ and a source learning task $\mathcal{T}_S = \{\mathcal{Y}_S, f_S(\cdot)\}$, and a target domain $\mathcal{D}_T = \{\mathcal{X}_T, P_T(X)\}$ and a target learning

task $\mathcal{T}_T = \{\mathcal{Y}_T, f_T(\cdot)\}$, *transfer learning* aims to enhance the learning of the target predictive function $f_T(\cdot)$ in \mathcal{T}_T using knowledge from $f_S(\cdot)$, which encompasses information from \mathcal{D}_S and \mathcal{T}_S , where $\mathcal{D}_S \neq \mathcal{D}_T$ and/or $\mathcal{T}_S \neq \mathcal{T}_T$ [159, 125].

Based on the above definition, transfer learning typically applies to the following scenarios:

- $\mathcal{D}_S \neq \mathcal{D}_T$ but $\mathcal{T}_S = \mathcal{T}_T$: This scenario is typically encountered in **transductive transfer learning**.
 - Addressing data distribution differences: Techniques such as domain adaptation can be used to align the feature spaces \mathcal{X}_S and \mathcal{X}_T or to adapt the model to work in the target domain [160, 133].
 - Example: Using a model trained on English text for sentiment analysis to perform sentiment analysis on Chinese text.
- $\mathcal{D}_S = \mathcal{D}_T$ but $\mathcal{T}_S \neq \mathcal{T}_T$: This scenario is typically encountered in **inductive transfer learning** [159].
 - Addressing different tasks: Techniques such as fine-tuning the model on the new task-specific data or using multi-task learning approaches can help [161, 162].
 - Example: Using a model trained for image classification to help with object detection in the same set of images.
- $\mathcal{D}_S \neq \mathcal{D}_T$ and $\mathcal{T}_S \neq \mathcal{T}_T$: This scenario is typically encountered in **unsupervised transfer learning**.
 - Addressing both differences: Techniques such as unsupervised domain adaptation and transfer learning algorithms that can handle both domain and task discrepancies are applied [163, 164].
 - Example: Using a model trained on labelled image data to help with clustering of unlabelled text data.

Chapter 6 of this thesis focuses on using a smoke detection model trained on Landsat imagery data to perform smoke detection on Sentinel-2 imagery data, where $\mathcal{D}_S \neq \mathcal{D}_T$ but $\mathcal{T}_S = \mathcal{T}_T$. This is an example of **transductive transfer learning**.

Chapter 3

Literature Review

This chapter reviews the literature closely related to the research aim of this thesis: developing innovative approaches to smoke detection using satellite imagery for fire disaster mitigation. Achieving this aim requires addressing accurate and timely smoke detection, as outlined in Chapter 1. Accordingly, the literature review concentrates on two main areas:

1. **Smoke detection approaches using remotely sensed imagery.** The review focuses on satellite-based approaches but extends to approaches for non-satellite platforms (e.g., surveillance cameras, UAVs, and aircraft), covering both pixel-level and scene-level detection. This provides a comprehensive context about various methods, and how common methods (e.g., CNN models) are used differently for satellite and non-satellite imagery. Despite unique challenges related to satellite-based smoke detection, approaches using non-satellite imagery can provide valuable information for accurate smoke detection using satellite imagery.
2. **DL-based transfer learning approaches in remote sensing, focusing on cross-sensor transfer learning.** The review aims to provide insights about leveraging transfer learning for fast model development for multiple satellites to enhance timely smoke detection collectively.

In addition, the review examines if the satellite-based approaches can effectively detect EF smoke, which is determined by the spatial resolution of the satellite imagery and is crucial for fire disaster mitigation.

The content of this Chapter is organised as follows. Section 3.1 reviews approaches using non-satellite imagery. Section 3.2 reviews approaches using satellite imagery. Section 3.3 reviews DL-based transfer learning in remote sensing. Section 3.4 summarises identified gaps in the previous research that this thesis aims to address.

The approaches discussed in Section 3.1 and Section 3.2 are further categorised into pixel-level and scene-level detection.

Pixel-level detection, which aims to identify smoke pixels individually, has been extensively studied using both non-satellite and satellite imagery. Approaches range from traditional methods like image processing and support vector machines (SVMs) to more advanced artificial neural networks, including multi-layer perceptrons (MLPs) (also known as fully connected neural networks) and more complex architectures such

as CNNs and FCNs. Consequently, traditional methods and artificial neural networks are reviewed separately in Section 3.1.1 and Section 3.2.1 in the context of pixel-level detection. It is important to note that DL models generally refer to artificial neural networks with more than three hidden layers. In this thesis, while foundational MLPs are discussed separately, other types of artificial neural networks, such as CNNs and FCNs, are considered DL models since they typically employ more than three hidden layers.

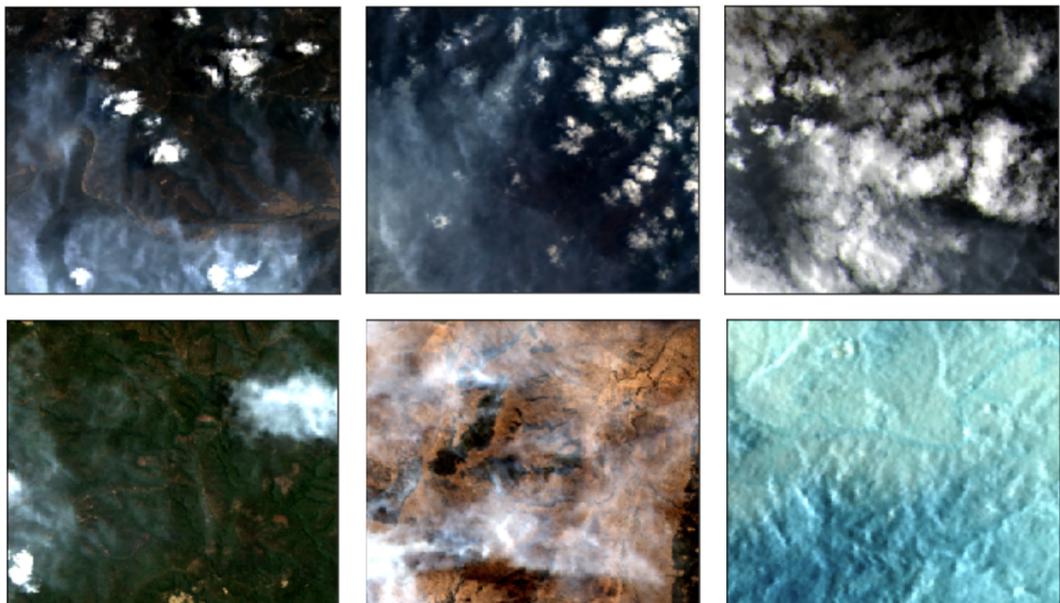
In contrast, scene-level detection, relying exclusively on DL, has emerged as a new area of study in the last decade, driven by the rapid development of image classification and object detection using CNNs and other DL techniques. Image classification methods predict the entire image as smoke or other scene classes. Object detection methods are more complex, aiming to detect possible smoke regions within the image to infer the scene class (e.g., non-smoke scene if smoke is not detected), with a predicted bounding box around each detected smoke region. Object detection is often adopted for smoke detection using surveillance camera imagery, where smoke plumes typically appear vertically with definable boundaries and without interference from clouds. In contrast, smoke plumes in satellite imagery generally diffuse horizontally over larger areas with less distinct boundaries, and are often mixed with or obscured by clouds, making object detection using bounding boxes less practical. Additionally, DL models require fixed input shapes, constraining satellite imagery-based smoke detection to smaller imagery tiles, in which smoke plumes can cover the entire tile, rendering bounding boxes unnecessary. Figure 3.1 compares a typical smoke scene in surveillance camera imagery with various smoke scenes in Landsat 8 OLI imagery, suggesting that image classification is a more suitable option for smoke detection using satellite imagery.

Since DL-based scene-level smoke detection depends heavily on labelled imagery datasets, the availability of such satellite imagery datasets is discussed in Section 3.2.2. It is worth noting that the imagery data used in previous studies, particularly those utilising surveillance cameras, primarily contain the RGB bands. Only a few studies, specifically those focused on satellite-based smoke detection, have incorporated both RGB and IR bands.

Prior to the detailed review, Table 3.1 summarises the smoke detection approaches found in the literature, organised by detection levels, platforms, bands used, approach types, and techniques employed.



(a) Smoke plumes fit within bounding boxes in a surveillance camera image.



(b) In Landsat 8 OLI imagery tiles, dense smoke plumes are often surrounded by diffuse, lighter plumes without clear boundaries. Smoke and clouds are difficult to separate using bounding boxes, and bounding boxes become unnecessary when smoke covers entire tiles.

Figure 3.1: Comparison of smoke scenes in a typical surveillance camera image and Landsat 8 OLI imagery.

Table 3.1: Approaches used in smoke detection across different platforms.

Detection Level	Platform	Bands Used	Approach Type	Techniques Employed
Pixel level	Non-satellites	RGB and/or IR	Traditional	Image processing techniques (e.g., histograms, wavelets, segmentation, image enhancement, optical flow); machine learning techniques (e.g., bag of features, k-means clustering, robust adaboost); other approaches such as fuzzy logic and extended Kalman filter.
			Artificial neural networks	MLP, DL segmentation models (e.g., FCNs, Deeplabv3+), GAN; hybrid methods incorporating traditional techniques such as optical flow.
	Satellites	RGB and/or IR	Traditional	Deterministic methods (e.g., false colour composite, multi-threshold values derived from spectral bands and pseudo bands, including spectral indices); traditional machine learning techniques (e.g., k-means clustering, Fisher's linear discriminant, random forest).
			Artificial neural networks	MLP, DL segmentation models.
Scene level	Non-satellites	RGB	Artificial neural networks	MLP, CNN (e.g., DNCNN, MobileNetV2, YOLO family, 3D CNN, Faster-R-CNN).
	Satellites	RGB	Artificial neural networks	CNN (e.g., SmokeNet, SAFA).

3.1 Smoke Detection Using Non-satellite Imagery

The primary non-satellite platforms used for smoke detection are surveillance cameras and UAVs. These platforms operate in environments significantly different from satellites, with distinct observation patterns. For instance, surveillance cameras are typically fixed on the ground, facing a specific direction, and produce RGB imagery in a time series, often in video format. Unless otherwise specified, smoke detection in video refers to the use of RGB imagery data in the following reviews. In contrast, UAVs are frequently equipped with IR sensors, flying at low altitudes and observing smaller areas, whereas satellites operate in distant orbits and commonly employ sensors with IR bands. Smoke plumes usually disperse vertically in surveillance camera imagery but appear horizontally in UAV and satellite imagery. Compared to satellites, smoke plumes observed from surveillance cameras and UAVs are less likely to intermingle with or be obscured by other aerosols such as clouds, dust, and haze. Additionally, the shapes and colours of smoke plumes tend to show less variability in surveillance and UAV imagery than in satellite imagery. These differences arise primarily due to variations in observation angles, distances from the source, and the scales at which smoke plumes disperse [165, 166, 167, 168].

Since the primary focus of this thesis is on smoke detection using satellite imagery, the review of non-satellite approaches in this section concentrates on the methods employed. Discussions about the advancements and limitations are kept minimal.

3.1.1 Pixel-level Approaches

Approaches Using Traditional Methods

The traditional approaches to smoke detection using non-satellite imagery can be grouped into image processing techniques and traditional machine learning techniques [169], depending on whether machine learning techniques were employed.

Image processing techniques commonly employ deterministic methods to extract useful information or features from different bands (or channels). These features can then be transformed into discriminative values for smoke detection.

Frequently used image processing techniques in non-satellite applications include histograms, wavelets, segmentation, and image enhancement [169]. Histograms were often utilised to analyse the distribution of pixel intensities, which can be altered by

smoke pixels. For example, Yuan et al. [170] combined histograms with dynamic texture analysis to detect smoke in video sequences. Wavelet transforms, used for multi-resolution analysis, extract smoke features at different scales, as demonstrated by Liu and Ahuja [171]. Segmentation approaches typically relied on colour and texture. Yuan [172] developed a model combining motion orientation and colour analysis for smoke identification. Image enhancement techniques were employed to improve the visibility of smoke. Çelik et al. [173] used statistical colour models and contrast enhancement for this purpose. Notably, Ye et al. [174] developed an algorithm that integrates multiple image processing techniques to detect moving fire and smoke blobs from videos, combining colour segmentation and wavelet analysis.

Traditional image processing methods often make several assumptions: that smoke colour is homogeneous, that smoke and non-smoke regions exhibit distinct reflected solar radiations, and that the background remains relatively stable [175]. However, these assumptions may not accurately reflect real-world conditions, potentially leading to a high false alarm rate. As a result, image processing techniques are frequently combined with machine learning techniques to extract features from images and reduce reliance on these assumptions.

Frequently used traditional machine learning approaches include SVMs [176], bag of features [177], k-means clustering [178], and robust adaboost [179, 169]. For example, Gubbi et al. [180] combined wavelet analysis with SVM for smoke detection in video. Zhang et al. [181] utilised a bag of features model combined with SVM to detect smoke in videos. Wu et al. [182] proposed a robust adaboost classifier using both static and dynamic features extracted from image processing for video-based smoke detection. Ajith and Martinez-Ramon [183] explored unsupervised segmentation techniques for detecting fire and smoke in IR videos, comparing k-means clustering, Gaussian mixture models [184], Markov random fields [185], and Gaussian Markov random fields [186]. The authors concluded that Markov random fields provided the best performance.

Fuzzy logic [187] was also employed to derive possible smoke pixels in combination with an extended Kalman filter [188], as demonstrated by Yuan et al. [189]. By adopting the extended Kalman filter, the fuzzy logic module was able to dynamically refine the rules, leveraging the learning capability of the filter.

Approaches Using Artificial Neural Networks

MLPs, the earliest and simplest type of artificial neural networks, were initially applied to smoke detection using non-satellite imagery. For example, Yu et al. [190] employed an MLP model for smoke detection in videos, using motion and colour features extracted from video frames as input. In another study [191], the same authors proposed employing MLP alongside more complex motion features derived from the Lucas-Kanade optical flow algorithm [192].

In the last decade, DL techniques have become mainstream. Yuan et al. [193] designed an FCN with an encoder-decoder structure using skip connections for video-based smoke segmentation. The model was trained on synthetic smoke images and outperformed other smoke detection approaches. Similarly, Cheng et al. [194] utilised Deeplabv3+ [195] and DenseCRF [196] for smoke detection and optimisation. Deeplabv3+ is an DL segmentation model with an encoder-decoder structure that employs atrous convolution [197], offering a more flexible receptive field compared to traditional convolution in regular FCNs. DenseCRF is a fully connected conditional random field that enhances pixel-wise classification using contextual information. After segmentation, they developed a heatmap model based on brightness values in the HSV colour space to indicate smoke thickness. Additionally, they trained a GAN model using sequential information to predict the trend of smoke motion, providing useful information to assist with evacuation or rescue efforts. Alternatively, Barmpoutis et al. [198] used two Deeplabv3+ networks for fire and smoke segmentation, employing 360-degree RGB cameras mounted on UAVs.

Aslan et al. [199] trained a GAN model in two stages for smoke detection in video. In the first stage, the GAN model was trained using background video images (without smoke) and random noise perturbations. The discriminator learnt to robustly distinguish the normal, smoke-free scenes from the altered scenes generated by the noise vectors. In the second stage, the discriminator was retrained using both the background images and real smoke images. This enabled the network to classify real smoke as an anomaly in the scene, leading to successful detection. The retrained discriminator was then adopted as the final classifier.

More recently, Yuan et al. [200] proposed a lightweight network incorporating an attention mechanism to reduce complexity in smoke segmentation. Another study [201], by the same authors, further refined smoke segmentation by employing multi-scale residual paths and weighted surveillance.

3.1.2 Scene-level Approaches

Scene-level smoke detection using non-satellite imagery is a relatively new area of study, relying exclusively on DL methods, as previously mentioned. An early example is the work of Frizzi et al. [202], which proposed a simple nine-layer CNN for fire and smoke detection in video. The authors employed a sliding window on the final feature map to locate fire and smoke areas. Similarly, Yin et al. [203] introduced a 14-layer CNN, named DNCNN, for scene-level smoke detection using images captured by surveillance cameras.

CNN models with more complex structures typically extract features that are more representative of underlying patterns, especially in challenging tasks like smoke detection. However, increased complexity may limit their use on small UAVs. Researchers have worked to balance this trade-off. For instance, Chen et al. [204] proposed a nine-layer CNN for detecting smoke and fire from UAVs. Jiao et al. [205] introduced a YOLOv3-tiny model, modified from YOLOv3 [206], to detect fire, smoke, and fire plus smoke on small UAVs. Zeng et al. [207] modified RefineNet [208] using depthwise separable convolutions [209] and channel pruning based on scale and shift factors, resulting in a lighter model with comparable accuracy. Similarly, Hossain et al. [210] employed an MLP to detect smoke using colour and texture features extracted with Local Binary Pattern [211] from 16×16 image tiles. Their method processed 19 frames per second and outperformed other techniques, including SVM, random forest, and YOLOv3.

It is noteworthy that the YOLO model family [212, 213] has been frequently utilised for fire and smoke detection in UAV and surveillance camera-based systems, further exemplified by works such as [214, 215, 216, 217]. This can primarily be attributed to their relatively lightweight architectures and strong performance in object detection. However, the review of the YOLO model family’s use in smoke detection on non-satellite platforms will not be further expanded. As aforementioned, image classification techniques are preferred over object detection methods for satellite-based smoke detection, which is the focus of this thesis. Table 3.2 provides a brief summary of the YOLO model family.

In contrast, some researchers utilised more complex networks to handle spatiotemporal features in videos. Hohberg [225] trained a 3D CNN [226] to extract spatiotemporal features for wildfire smoke detection in video. The 3D CNN demonstrated improved

Table 3.2: Summary of the YOLO Model family

Model	Released	Key Characteristics
YOLOv1 [212]	2016	Single bounding box per grid cell
YOLOv2 (YOLO9000) [218]	2017	Input image random resizing, anchor boxes, batch normalisation
YOLOv3 [206]	2018	Multi-scale detection, Darknet-53 backbone [206]
YOLOv4 [219]	2020	Balances high accuracy and speed with CSPDarknet backbone [220]
YOLOv5 [221]	2020	Lightweight, optimised for PyTorch implementation
YOLOv6 [222]	2022	Designed for high efficiency in industrial applications
YOLOv7 [223]	2022	Improved speed and accuracy with scalable architecture
YOLOv8 [224]	2023	Flexible architecture supporting segmentation tasks

performance over traditional CNNs with a lower false positive rate. Alternatively, Zhang et al. [227] trained a Faster R-CNN [228] model using synthetic smoke images for simultaneous smoke detection and localisation in video, supported by the region proposal network within the Faster R-CNN architecture.

3.2 Smoke Detection Using Satellite Imagery

3.2.1 Pixel-level Approaches

Approaches Using Traditional Methods

Early approaches to smoke detection in satellite imagery primarily relied on false colour composite, statistical or mathematical analyses, and traditional machine learning techniques.

False colour composite techniques usually employ three distinct bands from satellite sensors to enhance visual interpretation of smoke pixels, as demonstrated in [229, 230]. The bands typically include one visible, one NIR, and one SWIR, as smoke and cloud pixels exhibit varying radiometric characteristics in these bands. Derived

pseudo-bands, such as the Smoke Aerosol Reflectance Index and Water Index, were later introduced to enhance this approach, as seen in [231]. However, one limitation of this approach is that other aerosol phenomena prone to confusion with smoke, such as dust and haze, were not adequately addressed. Additionally, this method is not well-suited for automated workflows when processing large-scale satellite datasets [232, 43].

Given these limitations, approaches based on deriving multiple threshold values using selected spectral bands and/or pseudo bands, referred to as multi-threshold approaches, have gained popularity over false colour composite. The reflectance and brightness temperature (BT) values in certain spectral bands were typically used to derive the threshold values. This is because the spectral signatures of smoke often exhibit distinctive patterns from clouds and other confounding phenomena in these bands. One or more bands could be used in both the visible and IR regions (including NIR, SWIR, MWIR, and TIR), depending on the sensors used. An initial multi-threshold approach was proposed by Christopher et al. [39], who used multi-thresholds on several spectral combinations to extract texture features with the grey level difference vector [233]. Smoke detection was then performed based on texture analysis. Following this study, Baum and Trepte [234] proposed a grouped threshold approach to discern smoke, snow, clouds, fire, and clear sky. Similarly, Asakuma et al. [40] defined multi-threshold values to extract texture features for smoke detection using imagery from the GMS-5 satellite and AVHRR. In preparation for the launch of the Insat-3D satellite, Shukla and Pal [235] applied multi-threshold values for smoke detection over India using MODIS imagery with similar spectral bands. These threshold values were derived from spectral bands in the visible, MWIR, and TIR regions, along with a CLD pseudo band [236].

Notably, Chrysoulakis et al. [237, 54] proposed using multi-temporal and multi-spectral changes in four derived pseudo-bands with experimental thresholds for smoke detection in AVHRR imagery. The authors collected two images—one before and one during the smoke event. After aligning the images through geometric correction, multi-spectral thresholds were applied to mask water and cloud pixels. Multi-temporal thresholds were then used to detect the smoke core, based on the absence of smoke in the earlier image. Finally, texture information was used to expand the detected smoke area from the core, assuming homogeneity of smoke texture.

This method is a rare example of employing spatiotemporal changes for smoke detection using sun-synchronous satellite imagery. It is suitable for post hoc analyses and

is more practical when applied to smoke detection using geostationary satellites, which produce time series imagery at short intervals with less frequent need for geometric corrections. However, it is less practical for real-time detection due to the need to repeatedly acquire two images, with smoke absent in the earlier one—a condition that often requires manual verification. Furthermore, other dynamic events, such as clouds, can influence spatiotemporal changes, particularly given the long intervals between image captures when using sun-synchronous satellites (e.g., 12 hours for AVHRR).

A common problem with multi-threshold approaches is that the threshold values are often customised and hard to generalise. These values can be influenced by factors including atmospheric conditions, terrain, solar zenith angles at the time of image acquisition, and the spectral-radiometric characteristics of the sensors [238, 43]. Defining the threshold values requires consideration of these factors, often demanding substantial experience and domain knowledge.

More recent research employed various traditional machine learning methods, primarily supervised techniques, for automatic smoke detection. Notably, false colour composite and multi-threshold approaches were frequently adopted to extract and label pixel samples for training and verifying these models. For instance, Asakuma et al. [40] developed a supervised Euclidean classification model for smoke detection, using training pixels obtained through false colour analysis. Similarly, a supervised classification tree model was utilised in [239], where Himawari-8 sample data pixels were collected using multi-threshold approaches. In contrast, Li et al. [240] employed unsupervised k-means clustering to classify smoke, vegetation, water, and cloud pixels, followed by Fisher’s linear discriminant method [241] to separate smoke from cloud pixels, which were not adequately clustered by k-means. Although the sample pixels were unlabelled, they were initially extracted using multi-threshold methods. Despite the use of empirical threshold values to differentiate smoke from cloud pixels, this approach faced difficulties in distinguishing smoke from warm clouds and detecting smoke accurately in the downwind direction.

Although it deviates from the focus of this thesis, it is worth mentioning that Yao et al. [242] proposed a random forest [243] model to predict the minimum smoke height caused by forest fires using Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO). This was possible because CALIPSO is a sun-synchronous active satellite that emits lidar signals for remote sensing. Their research offers valuable insights into potential data fusion using lidar data in the future.

Very recently, [244] employed an improved sub-pixel mapping method based on the random forest algorithm for smoke detection in Himawari-8 imagery. In another recent study, [245] applied Mahalanobis distance [246] to measure the similarity between smoke pixels and a smoke concentration centre identified using the Laplace operator.

It is noteworthy that traditional machine learning techniques typically focused on distinguishing smoke from clouds by utilising specific spectral features (e.g., differences in BT values in some IR bands), which often required hand-crafted feature engineering.

Approaches Using Artificial Neural Networks

Early applications of artificial neural networks in pixel-level smoke detection using satellite imagery primarily relied on simpler MLPs, constrained by limited computing capabilities and the relatively early development of neural network techniques. For instance, Li [232] proposed an MLP with a single hidden layer containing 10 neurons to classify pixels in AVHRR imagery into “Smoke”, “Cloud”, and “Land”. The model’s input consisted of top of atmosphere (TOA) reflectance values from band 1 (Red) and band 2 (NIR), along with BT values from bands 3–5 (in the TIR region). The training pixel samples were manually selected from representative polygons of the aforementioned classes.

Similarly, in a more recent study, Li et al. [238] proposed another shallow MLP consisting of one hidden layer with 20 neurons to classify pixels in MODIS imagery into “Smoke”, “Cloud”, and “Underlying Surface”. The model’s input vector comprises six values: TOA reflectance values in bands 3 (Red), 7 (SWIR), 8 (NIR), 26 (NIR), the BT value in band 11 (NIR), and the BT difference between band 20 (MWIR) and 32 (TIR). Notably, the training pixel samples were obtained using the multi-threshold approach, incorporating smoke events from different seasons to address the spectral variance of smoke caused by seasonal impacts.

Both networks produced good results but failed to precisely detect smoke pixels with low density or in the downwind direction. The reasons behind these failures are multifaceted. As the authors noted, one possible cause could be related to the locations where training samples were collected, which often results in spectral variance of smoke due to different types of fuels. Another possible issue, though rarely mentioned, is the varying sun–sensor–surface geometry. This often causes differences in reflectance or

BT values between identical surface coverages captured at the same date and time but at different latitudes and longitudes. Such discrepancies may further affect detection accuracy.

One common issue with the two MLP models is that spatial information could not be utilised, as the methods focused on individual pixels. Additionally, both approaches examined only a fixed set of input features determined by domain knowledge. However, other potentially useful features were not explored or compared.

To incorporate spatial information for pixel-level smoke detection, Larsen et al. [41] proposed a more advanced DL model: an FCN for smoke segmentation in Himawari-8 imagery. The authors created a smoke segmentation training dataset (unpublished) with 975 imagery files, each containing a binary smoke mask serving as the label. A cloud-masking algorithm [42] was employed to create the smoke masks for the selected 975 imagery files, which had a spatial resolution of 2 km. Notably, the imagery data comprised five spectral bands (i.e., RGB, NIR, and SWIR) and two derived pseudo-bands (TOA temperature and FRP), allowing for spectral information to be explored across a broader spectrum. The input imagery was fed into an encoding block followed by a decoding block with skip connections, resulting in a binary image that classified pixels as either smoke or background. The model achieved an overall accuracy of 99.5% on the test dataset, which comprised 30% of the original dataset.

This study combined spatial and spectral features for pixel-level smoke detection using advanced DL techniques, which is notable in the literature. However, while FRP is considered a reasonable spectral indicator for smoke detection, the authors did not explore its specific contribution to the model's accuracy. Additionally, the labels of the training dataset—the smoke masks—were derived using an algorithm rather than real ground truths. This highlights a common issue in pixel-level smoke detection, as mentioned earlier, raising concerns about the reliability of the reported performance due to the limitations of the labelling process.

More recently, Shen et al. [247] employed combined DL models to reconstruct low-resolution VIIRS imagery by adapting it to super-resolution using Landsat imagery, thereby enhancing smoke detection accuracy in the reconstructed VIIRS imagery. This study primarily focuses on DL-based image reconstruction rather than on designing DL models specifically for smoke detection.

3.2.2 Scene-level Approaches

The rapid development of computer vision techniques in the last decade, particularly CNNs since AlexNet [25], has greatly advanced the accuracy of image classification. Various customised CNN models have been designed to tackle specific challenges in scene classification and object detection in remotely sensed imagery, including satellite imagery. For instance, one recent model, MGSNet, proposed by Wang et al. [248], uses multi-scale features and spatial context-aware mechanisms to enhance the separation of target and background information. Similarly, LSCNet, also proposed by Wang et al. [249], employs large kernels and multi-frequency attention to improve the learning of discriminative information, thereby increasing object detection accuracy. However, both models were designed to classify land use and land cover (e.g., urban areas, forests, harbours) or detect specific objects (e.g., aircraft, ships, vehicles) rather than smoke detection.

While CNN applications have been frequently studied for land use management, as further demonstrated in the works [250, 251, 252], few studies have focused on smoke detection using satellite imagery. This may be largely because land use and land cover ground truth data are relatively easy to collect, and several open-access datasets are available, such as the UC-Merced Dataset [253], RSSCN7 Dataset [254], and SAT4 and SAT6 datasets [255]. In contrast, collecting and labelling training data for smoke detection is much more challenging, and the sole publicly available training dataset for satellite-based smoke detection was published only a few years ago.

The first publication using CNNs for smoke detection in satellite imagery was the work by Ba et al. [43]. In this work, the authors proposed SmokeNet, a CNN model customised to address specific challenges in smoke detection (e.g., similarities between smoke, clouds, haze, and dust) using satellite imagery. Notably, a scene-level imagery training dataset for smoke detection, named USTC_SmokeRS, was created in this work, based on MODIS RGB imagery with a spatial resolution of 1 km.

The USTC_SmokeRS dataset, the first and only publicly available satellite imagery training dataset for smoke detection, comprises 6,225 MODIS RGB images collected over nearly 20 years, covering six continents. The dataset includes six classes: “Smoke”, “Cloud”, “Haze”, “Dust”, “Land”, and “Seaside”, each containing more than 1,000 training samples. The data collection and labelling process was meticulously supported by various methods, including geometric correction, radiometric calibration, analysis of thermal anomalies, and visual interpretation of true-colour and false-colour composi-

tions. Google Earth was also employed to identify land cover types. It is important to note that “Smoke”, “Dust”, and “Haze” images were purposely selected to exclude the presence of one another, although clouds may be present in these images and do not affect the image labels. In contrast, “Cloud” images are free from all other confounding aerosols. While this strategy aids trained models in distinguishing smoke from clouds, the inclusion of smoke-free dust and haze scenes contributes less to the accurate detection of smoke when mixed with dust or haze. Moreover, EF smoke plumes are likely to be overlooked due to the coarse spatial resolution of the imagery. Additionally, the absence of IR bands in the dataset limits the ability to explore useful spectral information that could improve smoke detection accuracy. No multispectral satellite imagery training dataset for smoke detection was available prior to the research conducted in this thesis, to the best of the author’s knowledge.

SmokeNet integrated the residual attention module [256] and the channel attention module [46] to help extract fine-grained features for distinguishing smoke from other aerosols. Additionally, a spatial attention module was developed to enhance the extraction of distinct spatial features related to different scene classes. Trained on the USTC_SmokeRS dataset, SmokeNet achieved an overall classification accuracy of 92.75%, outperforming previous state-of-the-art CNN models, including VGGNet [257], ResNet [44], DenseNet [45], AttentionNet [258], and SE-ResNet [46]. However, the classification accuracy specific to the “Smoke” class was 87.68% due to a higher number of misclassifications among smoke, dust, and haze images.

Building on the work in [43], Chen et al. [47] further proposed the current state-of-the-art scene-level smoke detection model, SAFA. In addition to incorporating the attention mechanism [259] and residual learning [260], SAFA was designed with a Global Information Extraction Path to capture global features, encompassing both background and aerosol phenomena. Furthermore, a Salient Feature Extraction Path was developed to learn salient features that highlight the key characteristics of different scenes. Specifically, a module called Mutual Activation Interim was introduced to smooth the fusion of features between different levels in the Global Information Extraction Path. The model makes its final predictions by combining the outputs of the Salient Feature Extraction Path and Global Information Extraction Path, weighted by two learnable coefficients.

Trained on USTC_SmokeRS, both SAFA and SmokeNet used 64% of the dataset for training, 16% for validation, and the remaining 20% for testing. The same train-

ing and testing settings, as well as evaluation metrics, were applied in both studies. SAFA achieved a testing accuracy of 96.22%, surpassing SmokeNet’s 92.75%. This accuracy also outperformed that of other advanced DL models developed more recently, including D-CNN [48], RSSC-ETDL [49], KFBNNet [50], HRNet [261], BoCF [262], and LPDCMEN [51], among others.

It is noteworthy that both SmokeNet and SAFA incorporated the attention mechanism [259] to extract salient features. The attention mechanism has been widely used in scene classification models, such as in [256, 263, 264, 265, 266, 267]. However, the implementations vary. SmokeNet adopted the channel attention implementation from Hu et al. [46] and implemented the spatial attention module based on a similar algorithm. In contrast, SAFA implemented its own spatial attention module and channel attention module in a more complex manner, incorporating parallel average pooling and max-pooling, feature map transformation with dual kernel sizes, and learnable coefficients.

Additionally, both SmokeNet and SAFA employed residual learning. SmokeNet adopted the residual attention module proposed in [256], which allowed the learning of fine-grained features relevant to the classification tasks. SAFA, on the other hand, used the residual blocks from He et al. [260] as the backbone blocks and further integrated the residual blocks with the spatial attention and channel attention modules to extract salient features in its Salient Feature Extraction Path.

3.3 Cross-sensor Transfer Learning

In this section, a brief introduction to transfer learning in computer vision tasks, typically based on conventional RGB images, is presented. Subsequently, transfer learning approaches specifically applied to remote sensing tasks, with a particular focus on the utilisation of multispectral imagery, are reviewed.

3.3.1 Transfer Learning in Computer Vision Using Conventional RGB Images

DL models for typical computer vision tasks, such as image classification, segmentation, and object detection, have predominantly been designed to process imagery data

containing three channels (RGB bands by default). This aligns with how human vision perceives colour through red, green, and blue channels. Consequently, most publicly available imagery training datasets, including those derived from satellite imagery, consist of only RGB bands. For example, ImageNet [268], CIFAR-10 / CIFAR-100 [269], COCO [24], PASCAL VOC [270], as well as the aforementioned satellite imagery datasets UC-Merced [253], RSSCN7 [254], and SAT4/SAT6 [255], all comprise RGB images.

Transfer learning has been widely leveraged in computer vision tasks to adapt DL models pretrained on larger datasets for new tasks using smaller, domain-specific datasets (e.g., datasets for smoke detection). For instance, many renowned CNN models have been pretrained on ImageNet. These models have learned to extract essential features for general image classification tasks and can be readily transferred and adapted to datasets such as UC-Merced for land-use classification.

The process typically involves two steps when using new imagery data from the target domain. First, the classification head is retrained while freezing the convolutional base of the pretrained CNN model. Second, the entire model is fine-tuned with a very low learning rate. In the first step, the transferred model directly utilises its capabilities learned from the source domain for feature extraction while learning a new prediction mechanism for the target domain using the extracted features. In the second step, the model gradually updates both its feature extraction capabilities and prediction mechanism simultaneously, adapting itself to the target domain through controlled fine-tuning.

The above transfer learning strategy generally performs well when both the source and target domains utilise conventional RGB images. However, it becomes less practical when the number of channels in the imagery data differs between the source and target domains. In remote sensing tasks, for instance, the acquired imagery typically includes IR bands, which provide significant spectral information beyond the visible range. Some sensors, such as hyperspectral sensors, may even divide the visible spectrum into more than three bands (i.e., standard RGB bands) to capture finer spectral details. This richer spectral information in remotely sensed multispectral imagery is often crucial for tasks like smoke detection. However, transferring a CNN model pretrained on conventional RGB images to multispectral imagery is significantly more challenging. The transfer learning strategy must not only account for the variation in input channels but also handle feature extraction across diverse spectral bands. Sim-

ilarly, transferring a model across different satellite sensors faces the same challenges, as these sensors often have varying numbers of spectral bands, which may also differ in their spectral regions and bandwidths.

Compared to transferring a CNN model from an RGB source domain to a multi-spectral target domain, studies on transfer learning where both the source and target domains utilise multispectral satellite imagery are even rarer, primarily due to the scarcity of labelled multispectral satellite training data.

3.3.2 Transfer Learning Applied to Remote Sensing Tasks

Transfer learning has been extensively investigated for various remote sensing tasks. For instance, Verma et al. [271] employed transfer learning to map urban slums using satellite imagery, while Shabbir et al. [272] adapted ImageNet-trained ResNet50 [260] for land cover scene classification. Nowakowski et al. [273] and Patel et al. [274] explored transfer learning for crop type classification using hyperspectral imagery and benchmark CNN models. Wu et al. [275] investigated transferring pretrained MobileNetV2 [276] for wildfire detection from UAV imagery, and Agrawal et al. [277] utilised CNNs with transfer learning for natural disaster classification in satellite and drone images.

However, while transfer learning has been widely studied for land use and crop identification tasks in satellite imagery, as demonstrated in the works [271, 272, 278, 274, 273, 279, 280], no study has focused on smoke scene detection due to the limited availability of suitable datasets. In addition, previous studies on transfer learning in remote sensing primarily concentrated on transferring models trained using non-satellite imagery to satellite imagery (e.g., Sharma et al. [281], Xiao et al. [282], Verma et al. [271]).

Cross-sensor transfer learning, involving both the source and target domains using satellite imagery, had not been thoroughly explored until 2021 when Wang et al. [283] introduced a method using knowledge distillation [284] and two distinct CNN network structures for knowledge transfer between different sensors. This method involves fine-tuning a pretrained model from the source domain incorporating RGB images to generate pseudo labels for multispectral and hyperspectral data within two different target domains. Subsequently, a distinct model is selected and trained using the data featuring pseudo labels from the target domains.

The underlying premise of this approach is rooted in the belief that the fine-tuned source model is proficient in accurately labelling data within the target domains. However, it is essential to note that no mechanism is currently in place to validate whether these pseudo labels accurately represent the true classes of data in the target domain. Furthermore, it is noteworthy that this approach entails training a new model recursively using data with pseudo labels from the target domain rather than transferring a pretrained model from the source domain.

In a recent study, Tao et al. [285] proposed a transitive transfer learning framework as an innovative solution to simultaneously address data distribution gaps and task differences between source and target domains. This framework operates by strategically constructing multiple intermediate domains, allowing the transfer learning task to be divided into manageable subtasks. Each subtask focuses on one aspect—either data adaptation or task adaptation, progressively transitioning from the source domain to the target domain through these intermediate domains.

For instance, consider a case where the source domain uses RGB imagery with 10 scene classes, while the target domain employs multispectral imagery with 30 scene classes. Two intermediate domains can be constructed from the target domain by grouping the bands in the imagery data: RGB and IR. The first intermediate domain contains only the RGB bands, while the second intermediate domain includes the IR bands. Consequently, the first subtask uses RGB imagery (same as the source domain), but the task shifts to 30 classes. The second subtask then shifts to IR imagery while keeping the task (30 classes) unchanged. Ultimately, this framework facilitates the transition from RGB imagery with 10 classes to multispectral imagery with 30 classes.

This structured approach supports cross-sensor transfer learning tasks that involve varying spectral bands, numbers of spectral bands, and classification tasks. However, each subtask requires a specific strategy to handle either data distribution or task differences, leading to substantial data pre-processing and model customisation. Moreover, the overall success of this framework heavily depends on intensive model training for all transfer learning subtasks.

The transfer learning strategies presented in [283] and [285] both fall under heterogeneous-network transfer learning, where the model used for the target domain is independent of the model pretrained in the source domain.

In contrast, homogeneous-network transfer learning, where the model pretrained from the source domain is updated for the target domain, is more straightforward. However, it primarily relies on bridging data distribution gaps, which becomes more challenging when dealing with multispectral imagery that contains varying spectral bands and numbers of bands. Traditionally, previous research has required that the channel counts of the imagery data in the target domain either match or be transformable (e.g., through band selection or mapping) to align with the channel counts of the source domain imagery. This ensures that the convolutional base of the pretrained model can be transferred smoothly. These methods can be categorised into different groups based on how they align the channel counts of input imagery between the source and target domains:

1. Transferring a model pretrained with a grayscale image dataset to multispectral satellite imagery, where the transferred model processes each band individually, using a subset of bands specifically selected from the multispectral satellite imagery. This method was demonstrated in Sharma et al. [286, 281].
2. Transferring a model pretrained with RGB image datasets (e.g., ImageNet) to satellite imagery with three input channels (e.g., RGB bands) or three channels derived through dimensionality reduction techniques, such as principal component analysis. This approach is exemplified in Sharma et al. [281], Jian et al. [287], Gadiraju et al. [280], and Patel et al. [274].
3. Transferring a model pretrained with RGB image datasets (e.g., ImageNet) to multispectral satellite imagery that has been mapped to three channels, as demonstrated in He et al. [288] and Singhal et al. [289].
4. Transferring a model pretrained for one task to a new task (e.g., transferring a model pretrained for crop type classification to land use classification) using multispectral imagery with the same number of bands from the same satellite, as explored in Cong et al. [290].

Figure 3.2 illustrates the above transfer learning approaches.

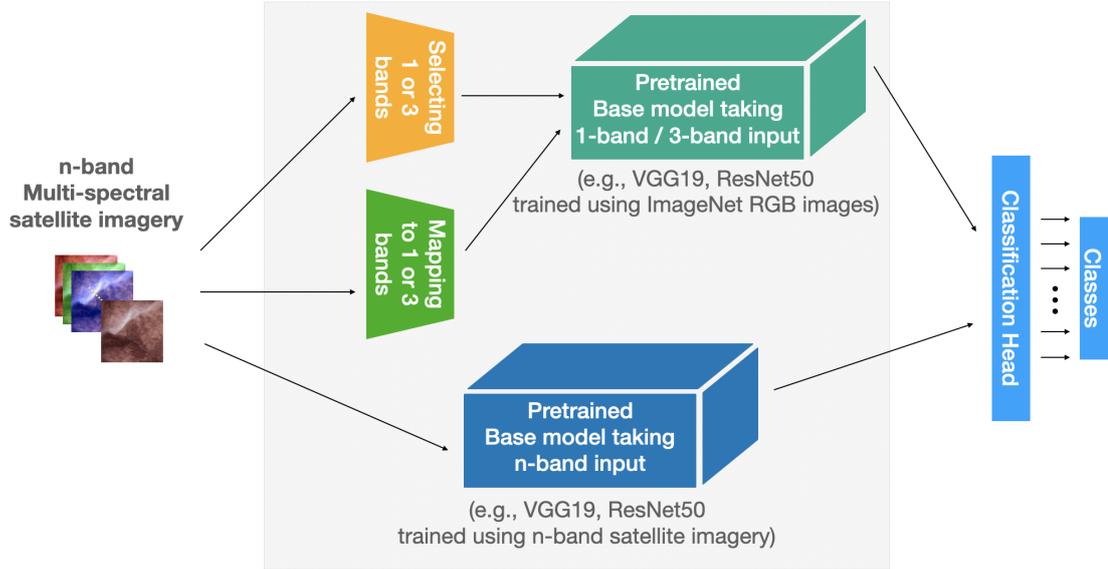


Figure 3.2: Previous transfer learning approaches require the imagery data from both satellites to have equal channel counts or to align the channel counts through adjustments (e.g., dimensionality reduction).

3.4 Identified Research Gaps

This thesis focuses on research gaps related to DL-based approaches, given their superior scalability and ability to automatically learn fine-grained semantic features compared to traditional methods. Based on a comprehensive literature review, the identified research gaps in the previous research are summarised as follows:

1. **Research gaps related to DL models, particularly those designed specifically for scene-level smoke detection:**
 - Limited suitability for deployment onboard SmallSats, which are increasingly used in satellite constellations for NRT fire/smoke detection, due to the models' heavyweight architectures (often with tens of millions of parameters) and their excessive demands on power, memory, and computational resources.
 - Insufficient exploration of spectral information, especially regarding IR bands, which are essential for distinguishing smoke from similar aerosol phenomena (e.g., clouds, haze, dust). The lack of IR bands in the training data and the absence of effective methods to fully leverage these spectral bands remain underexplored.

2. Research gaps in terms of effectiveness for fire disaster mitigation:

- Potentially high FNs in detection due to overlooking EF smoke, partly caused by the coarse spatial resolution of MODIS RGB imagery used for training.
- Delayed or missed detections due to the low temporal resolution (12 hours) of MODIS.
- Lack of research on using multiple satellites with higher spatial resolutions to collectively achieve higher temporal resolution for timely EF smoke detection.

3. Other research gaps related to the above issues:

- Inadequate annotated satellite imagery training datasets, particularly those with IR bands and from sensors with higher spatial resolutions. Before this thesis, the only publicly available dataset was USTC_SmokeRS, derived from MODIS RGB imagery with a spatial resolution of 1 km.
- Lack of research on adapting pretrained models to new satellite sensors for faster model development, especially when observational data from the new sensors is limited. Adapting models from one satellite sensor to another can facilitate rapid development; however, models trained on one sensor's data may perform suboptimally on another sensor's imagery due to differences in spectral bands and radiometric characteristics.

To address the above research gaps, the following objectives are pursued:

1. Design a DL architecture that is lightweight (in terms of the number of parameters) yet accurate for satellite-based smoke detection. This architecture must use multispectral satellite imagery containing IR bands so that essential spectral information in these bands can be explored. In the meantime, the contribution of IR bands on smoke detection can be also investigated and verified. Ideally, this architecture should be suitable for onboard SmallSats deployment, paving the way to employ satellite constellations equipped with onboard DL models for accurate NRT scene-level smoke detection.
2. Construct new satellite imagery training datasets meeting the following requirements to effectively facilitate fire disaster mitigation:

- Using satellites equipped with good spatial resolutions so EFs smoke plumes that cover small geographic extents can be effectively captured in their imagery and detected by the DL model.
 - Including essential IR bands (e.g., NIR, SWIR) that provide significant spectral information to help distinguish smoke and other confounding aerosols more effectively.
 - Reflecting variant types of smoke in terms of their colours, shapes, density, scales, etc., to improve the model's robustness in detecting variant smoke plumes.
3. Develop a mechanism that enables the DL model to automatically learn useful spectral patterns hidden in the IR bands that can improve smoke detection accuracy.
 4. Investigate effective approaches to robustly adapting a pretrained DL model to new sensors with limited training data, facilitating fast model development to use multiple satellites for accurate and timely smoke detection.

Chapter 4

VIB_SD: A Lightweight DL Model for Smoke Detection Using Satellite Imagery with IR Bands

As discussed in Chapter 1, existing DL models customised for smoke detection have limited suitability for onboard satellite applications, particularly for SmallSats. This is mainly due to their heavyweights and excessive demands for power supply, memory space, and computational resources. Applying a DL model for onboard SmallSat smoke detection often requires trading the model’s accuracy for suitability.

While multispectral satellite imagery provides more essential information for smoke detection than RGB satellite imagery, previous DL-based research only used the RGB bands. In addition, the sole publicly available imagery training dataset (i.e., USTC_SmokeRS [43]) used for the studies of satellite-based smoke detection was derived from MODIS with a coarse spatial resolution of 1 km. EF smoke plumes are likely to be overlooked in such lower spatial resolution imagery. Consequently, the models trained using this dataset tend to be ineffective for EF smoke detection.

This chapter addresses the above limitations relating to the first research question posed in Chapter 1:

RQ-1 How to design a lightweight DL model that can potentially be deployed onboard SmallSats for accurate EF smoke detection using multispectral satellite imagery?

To prepare for potential onboard-SmallSat smoke detection, this chapter proposes the aforementioned lightweight CNN model VIB_SD, incorporating residual learning and attention mechanism with customised convolutional layers for feature extraction in variant scales. Despite its substantially lighter weight, with less than 2% of the state-of-the-art model SAFA’s parameter count (1.66 million versus 84.2 million), VIB_SD achieved competitive accuracy (93.57% versus SAFA’s 96.22%) when trained on the RGB USTC_SmokeRS dataset.

To further contribute towards EF smoke detection using multispectral satellite imagery with higher spatial resolutions, this chapter presents a new training dataset, Landsat6c. As the first multispectral satellite imagery dataset for smoke detection, Landsat6c was constructed using imagery from Landsat 5 Thematic Mapper (TM) and Landsat 8 OLI, with a spatial resolution of 30 metres. It contains 1836 images, labelled into three classes: “Smoke”, “Clear”, and “Other_aerosol”.

Additionally, to investigate the impact of using extra IR bands on the accuracy

of smoke detection, this chapter compares the VIB_SD model trained on Landsat6c using five different band combinations. The results demonstrate that adding the NIR band improved prediction accuracy over the RGB bands. Adding both SWIR bands improved the model’s performance further over adding only one SWIR band. The case studies show that the model trained with multispectral bands could effectively detect EF smoke plumes over small geographic extents, even if they are undercovered or mixed with clouds.

This chapter is organised as follows: Section 4.1 provides the background of the work; Section 4.2 introduces the satellite imagery datasets used, the structure of VIB_SD, and the implementation details of its key modules; Section 4.3 outlines the experimental settings and evaluation metrics; Section 4.4 presents the experimental results and findings; Section 4.5 demonstrates the effectiveness of the model trained with multispectral bands in detecting smoke mixed with clouds or over small geographic extents, through case studies; Section 4.6 discusses the results in further detail and explores future possibilities; and Section 4.7 concludes the chapter.

The work presented in this chapter is based on my paper *“Investigating the Impact of Using IR Bands on Early Fire Smoke Detection from Landsat Imagery with a Lightweight CNN Model”* [27], which was published in the journal “Remote Sensing” in June 2022.

4.1 Introduction

Aggravated by climate change, wildfires are occurring more frequently and often develop rapidly, leading to significant societal, ecological, and economic consequences [2, 3, 4]. Detecting fires in their early stages can help prevent the disastrous impacts of wildfires. One cost-effective approach is using satellite imagery for EF detection, as an increasing number of satellites are being launched to monitor the Earth. However, EF detection from satellite imagery is challenging, as EFs can be easily obscured by thick forest canopies, clouds, or the smoke plumes they emit. Even with thermal bands, fires can be masked by heated backgrounds during hot weather, and false alarms are often triggered by heated bare soils, deserts, or other highly reflective regions [20, 21, 22].

Detecting smoke to infer fires presents a valuable alternative to direct fire detection. Smoke has the following characteristics: (1) it can rise above forest canopies in a short

time and typically has distinct colours compared to vegetation; (2) smoke disperses more quickly and over a larger area than the fire itself, making it easier for satellites to capture; (3) the temperature of smoke is significantly lower than the hot background.

Nevertheless, as discussed in Chapter 1 and demonstrated in Figure 1.1, smoke detection from satellites is challenged by other factors: (1) the varying characteristics of smoke plumes, including their shapes, colours, and scales; (2) the similarity and overlap in spectral signatures between smoke and other objects, such as snow, clouds, and dust [232, 291, 238, 43].

Early research attempted to distinguish smoke in satellite imagery from other confounding objects (e.g., water, snow, clouds) using handcrafted features derived from deterministic rules at the pixel level [39, 234, 40, 237, 54, 235, 239, 240]. However, these features are often closely tied to specific local conditions, making it difficult for them to generalise effectively.

The development of DL techniques, especially CNNs, has shifted the focus of smoke detection in satellite imagery to the scene level in recent years [43, 47]. DL models can automatically extract semantic features to determine whether the satellite imagery contains smoke, regardless of the shapes and positions of the smoke plumes, even when other confounding objects or aerosols are present. However, several gaps remain:

- Existing satellite-based scene-level smoke detection DL models are complex and struggle to meet the growing demands for onboard-SmallSat NRT smoke detection.
- These DL models were trained on the USTC_SmokeRS dataset [43], derived from MODIS RGB imagery, leaving valuable information in the IR bands unexamined. Such information could improve detection accuracy, particularly when fires are obscured. For instance, in Figure 4.1, visualising the smoke scenes using SWIR_2, NIR, and blue bands reveals active fires in vivid red, burnt scars in dark red (bottom-left image), and smoke in light blue in both bottom images. These properties are not clear in the RGB images in the top row.
- Furthermore, previous DL-based smoke detection studies were conducted using satellite imagery with coarse spatial resolutions (e.g., 0.25–1 km in MODIS imagery and 0.5–2 km in Himawari-8 AHI imagery), where EF smoke over small geographic extents could easily be overlooked. Effective detection of EF smoke over small geographic areas has yet to be investigated, primarily due to the lack of

suitable imagery training data with higher spatial resolutions that can accurately capture EF smoke.

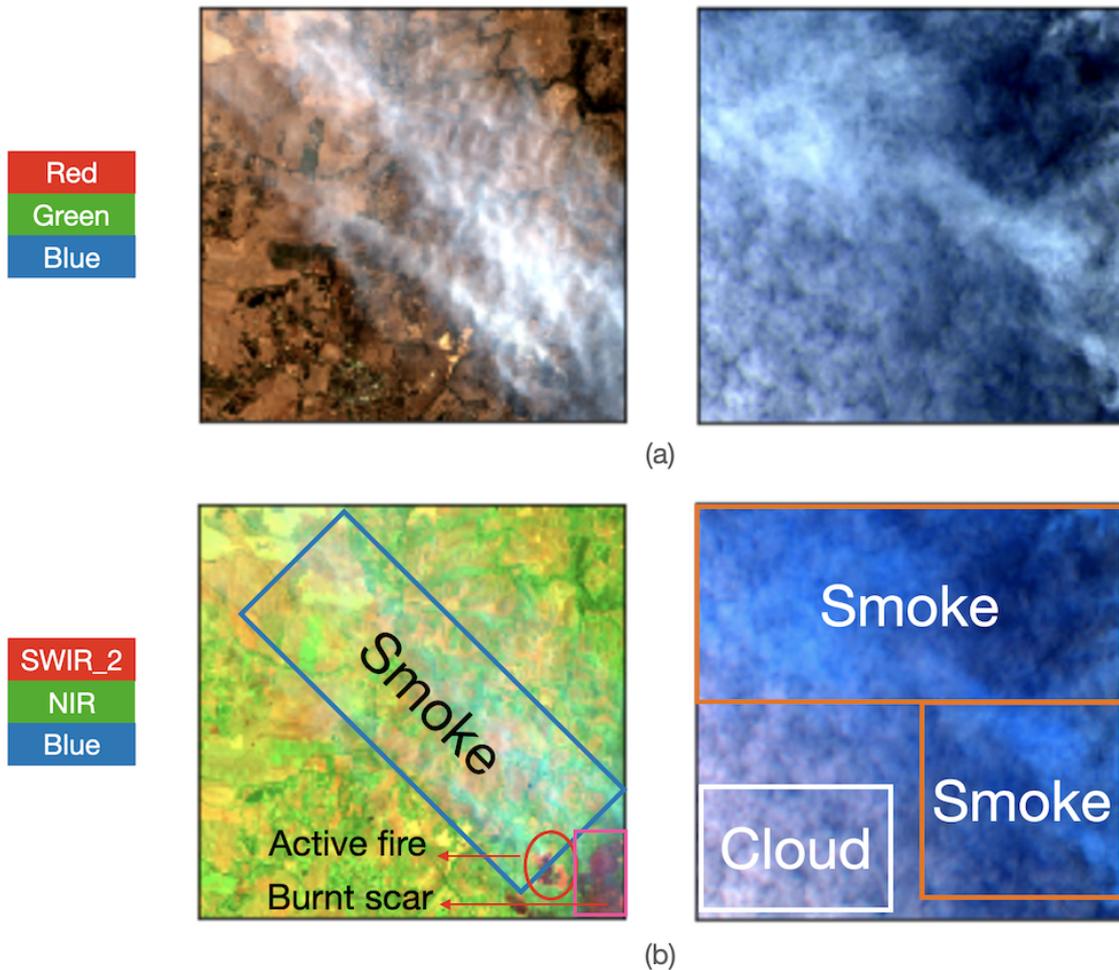


Figure 4.1: Two smoke scenes are visualised in different bands. (a) RGB. (b) SWIR_2, NIR, and blue.

To address the above gaps, this work specifically aims to:

- Design a CNN model that is both accurate and lightweight, with the potential to be deployed onboard SmallSats for NRT smoke detection.
- Construct a labelled multispectral imagery training dataset with higher spatial resolution, supporting future studies on EF smoke detection and advancing fire disaster mitigation efforts.

- Investigate the impact of using additional IR bands in DL-based smoke detection, validating the nuanced contributions of individual IR bands.

For the first aim, this work proposes the lightweight CNN model VIB_SD, which integrates the latest DL techniques proven effective for smoke detection to achieve optimal detection accuracy. By taking multispectral satellite imagery with varying numbers of bands as input, VIB_SD enables the investigation of the nuanced contributions of individual IR bands using various band combinations. VIB_SD will be introduced in Section 4.2.3.

For the second aim, this work creates a multispectral imagery training dataset, Landsat6c, for smoke detection, based on Landsat 5 TM and Landsat 8 OLI imagery data. Landsat6c contains 1836 $256 \times 256 \times 6$ imagery tiles, featuring six spectral bands: RGB, the NIR band, and two SWIR bands (i.e., SWIR_1 and SWIR_2), all with a 30-m spatial resolution. Details about Landsat6c will be introduced in Section 4.2.2.

For the third aim, this work conducts comprehensive experiments by training VIB_SD using Landsat6c with five different band combinations: RGB, RGB plus NIR, RGB plus NIR and SWIR_1, RGB plus NIR and SWIR_2, and all six bands. The nuanced contributions of the individual IR bands are analysed by comparing the model performances across these combinations. The experimental details will be introduced in Section 4.3.1. Results show that adding the NIR band significantly improved model prediction accuracy, while both SWIR bands further enhanced accuracy.

The major contributions of the work presented in this chapter are:

- The development of the lightweight CNN model VIB_SD, which achieved competitive accuracy compared to the state-of-the-art SAFA model (93.57% versus SAFA’s 96.22%) when trained on USTC_SmokeRS. More importantly, this performance was achieved with less than 2% of the parameters used by SAFA. VIB_SD demonstrates significant potential for onboard-SmallSat applications due to its lightweight architecture and high accuracy.
- The creation of a multispectral satellite imagery smoke detection training dataset, Landsat6c, derived from moderate spatial resolution (30-metre) Landsat imagery. As the first dataset of its kind in the literature, the six-band Landsat6c is labelled into three smoke-related scene classes (i.e., “Smoke”, “Clear”, and “Other_aerosol”). This dataset serves as a valuable resource for future research

on satellite-based smoke detection, particularly studies utilising IR bands for EF smoke detection.

- A comprehensive investigation into the impact of IR bands on smoke detection, marking the first time in the literature. The findings suggest that while individual IR bands (i.e., NIR, SWIR_1, and SWIR_2) contribute to the effective improvement of smoke detection accuracy, using all three IR bands together yields the highest accuracy. These results offer valuable guidance for band selection strategies in multispectral or hyperspectral satellite imagery smoke detection applications.

4.2 Materials and Methods

This section introduces the two satellite imagery datasets used in this work, and the VIB_SD model along with its key modules.

4.2.1 RGB USTC_SmokeRS Dataset

The USTC_SmokeRS dataset was used to evaluate VIB_SD by comparing it with existing models developed and trained on this dataset.

The USTC_SmokeRS dataset consists of 6225 $256 \times 256 \times 3$ RGB images collected from MODIS (Level-1B), which has a spatial resolution of 1 km. The dataset contains six smoke-related scene classes: “Smoke”, “Cloud”, “Dust”, “Haze”, “Land”, and “Seaside”. The number of images in each class of the USTC_SmokeRS dataset is shown in Table 4.1. Further details on the dataset can be found in [43]. To the best of the author’s knowledge, the USTC_SmokeRS dataset was the only labelled satellite imagery dataset for DL-based smoke scene detection prior to this research.

Table 4.1: Number of images in USTC_SmokeRS.

Smoke	Cloud	Haze	Dust	Land	Seaside	Total
1016	1164	1002	1009	1027	1007	6225

4.2.2 Multispectral Landsat Imagery Dataset

One of the key contributions of this work is the construction of a labelled multispectral moderate spatial resolution satellite imagery dataset, named Landsat6c, specifically for EF smoke detection. The dataset comprises three smoke-related scene classes: “Smoke”, “Clear”, and “Other_aerosol”. This dataset is utilised to explore the contribution of additional IR bands to the accuracy of smoke detection. This section outlines the processes of data collection and labelling.

Data Source

The multispectral Landsat imagery data were collected based on historical wildfires in Australia. Although the dataset is geographically specific to Australia, the methods developed are not limited to applications in other regions globally.

The Landsat series was selected as the target satellite platform due to its significantly higher spatial resolution (30 m) compared to MODIS and Himawari-8 AHI. The primary data source is Landsat 8 OLI, launched on 11 February 2013, and in continuous operation since then. To increase the likelihood of capturing smoke-related imagery, the query period was extended back to 2010, allowing Landsat 5 TM (decommissioned on 5 June 2013) to be used as a minor data source. Landsat 7 was excluded due to the black stripes in its imagery, caused by the failure of its scan line corrector on 31 May 2003. Landsat 9, launched on 27 September 2021, was not included as its data were not publicly available at the time the data collection process commenced.

The surface reflectance Landsat imagery data, processed using the Nadir Corrected Bi-directional Reflectance Distribution Function Adjusted Reflectance Coupled with a Terrain Illumination Correction (NBART) algorithm [292], were queried and downloaded from the Digital Earth Australia (DEA) Sandbox platform [293]. The NBART Landsat imagery data are indexed by DEA with open access to the public and can be queried and downloaded based on specified location coordinates, time range, and band requirements.

The time windows and spatial ranges required for querying the data were extracted from historical fire datasets in SA and New South Wales (NSW), hosted on Data SA [294] and Data NSW [295], respectively. Table 4.2 presents a sample record from the SA historical fires dataset.

Table 4.2: A sample record in the SA historical fires dataset.

Attributes	Values
FID	5801
INCIDENTNU	202011011
INCIDENTNA	Overland Corner/Calperum
INCIDENTTY	Bushfire
FIREDATE	2020-11-15
FINANCIALY	2020/2021
FIREYEAR	2020
SEASON	SPRING
DATERELIAB	1
IMAGEINFOR	Landsat 8 17/11/2020
FEATURESOU	33
CAPTURESOU	4
HECTARES	1447.32
SHAPE_Leng	0.91332
SHAPE_Area	0.00141367
min_longi	140.372
max_longi	140.613
min_lati	-33.7796
max_lati	-33.6953
Sensor	MODIS
geometry	(POLYGON (((140.4123221680001 -33.6969758549999 ...)))

Data Collection Strategy and Processes

The construction of Landsat6c consists of two phases: imagery collection, followed by tiling and labelling.

1. Imagery Collection:

In this phase, scripts were developed to extract time and location information

from the historical fire datasets on Data SA and Data NSW, and to download Landsat 5 TM and Landsat 8 OLI imagery in bulk based on the derived temporal and spatial data. The query time range was extended to cover 16 days before and after the recorded fire date, which may not precisely indicate the ignition date due to recording discrepancies. This extension also allowed imagery from the same area to be collected at different times and under varying weather conditions. A 5 km buffer along both the longitude and latitude was applied to the polygon coordinates of the burnt scars. The returned imagery was then visually inspected to select those successfully capturing smoke.

Six spectral bands in the imagery data were retained, including the RGB bands, the NIR band, the SWIR_1 band, and the SWIR_2 band. As shown in Table 4.3, the wavelengths of the six selected bands for Landsat 5 TM and Landsat 8 OLI differ slightly, though each corresponding band falls within a similar range. The majority of the imagery data were collected from Landsat 8 OLI. The thermal band was excluded because the initial aim was to construct a mixed imagery dataset from both Landsat and Sentinel-2 (A and B), which could be used to train a model adaptable to different sensors. Since Sentinel-2 does not include a thermal band, it was considered preferable to exclude the thermal band from the Landsat imagery. However, incorporating the thermal band will be considered in future data collection, where applicable.

Table 4.3: Wavelengths of the selected bands of Landsat 5 TM and Landsat 8 OLI.

Band	Wavelength (μm)	
	Landsat 5 TM	Landsat 8 OLI
Red	0.63–0.69	0.64–0.67
Green	0.52–0.60	0.53–0.59
Blue	0.45–0.52	0.45–0.51
NIR	0.76–0.90	0.85–0.88
SWIR_1	1.55–1.75	1.57–1.65
SWIR_2	2.08–2.35	2.11–2.29

A total of 477 imagery files covering fire sites across a wide range of locations were downloaded from DEA Sandbox. Of these, eight files were from SA, with the remainder from NSW. Fifteen imagery files were sourced from Landsat 5 TM, while the rest were from Landsat 8 OLI. The areas covered by these files varied significantly, depending on the scale of the fires.

2. Tiling and Labelling:

In this phase, the imagery files were tiled into $256 \times 256 \times 6$ patches with a 50% overlap both horizontally and vertically. The overlap between patches aids the model in learning to recognise smoke, regardless of its position within the patches. The tiling process is illustrated in Figure 4.2.

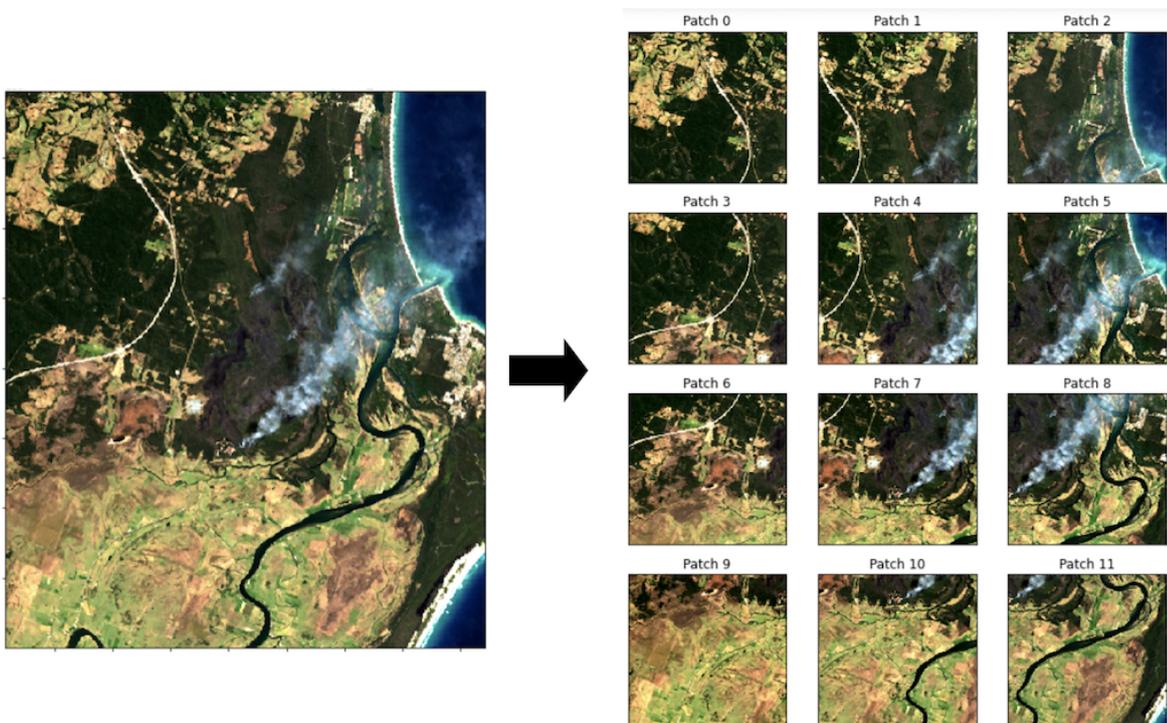


Figure 4.2: Overlapped tiling.

The patches were labelled into three classes: “Smoke”, “Clear”, and “Other_aerosol”.

“Other_aerosol” refers to non-smoke scenes that are not classified as “Clear”, such as scenes with cloud, dust, haze, or other aerosol mixtures. Patches were labelled as “Clear” if no visible aerosol was present, and as “Smoke” if they contained any visible smoke.

Identifying smoke in the patches is not always straightforward, as shown in the bottom images of Figure 4.1. To improve labelling accuracy, the patches were visually examined in false colour using the SWIR_2 band, the NIR band, and the blue band. Additional imagery files containing only clear backgrounds or clouds were also downloaded during this phase to balance the number of images in the non-smoke classes.

After tiling 36 imagery files from various fire sites and labelling the resulting patches, Landsat6c is created, containing 1836 $256 \times 256 \times 6$ imagery patches with more than 600 patches in each class. Landsat6c captures a wide range of smoke scenes (e.g., smoke in different shapes, sizes, colours, and densities; smoke over diverse backgrounds; smoke mixed with various types of clouds), reflecting the complexity of wildfire events and the challenges that smoke detection faces.

Table 4.4 presents the class distribution of the Landsat6c dataset.

Table 4.4: Class distribution of the Landsat6c dataset.

Smoke	Other_aerosol	Clear	Total
615	605	616	1836

4.2.3 VIB_SD

As mentioned in Section 4.1, the design of VIB_SD needs to meet two key requirements:

- Be lightweight (in terms of parameters) to improve efficiency for potential onboard-SmallSat applications;

- Maintain the highest possible detection accuracy despite the lightweight design.

To achieve good accuracy, attention mechanism [259], residual learning [260], and the inception structure [296, 297] were integrated to assist in extracting smoke-related features. Module stacking was minimised to control the model’s size. To support a smooth investigation into the impact of IR bands, the implementation ensures flexibility in training VIB_SD with input imagery containing varying numbers of bands.

Figure 4.3 displays the main structure of VIB_SD on the left, alongside the structures of the stem block, reduction block, and classification head on the right.

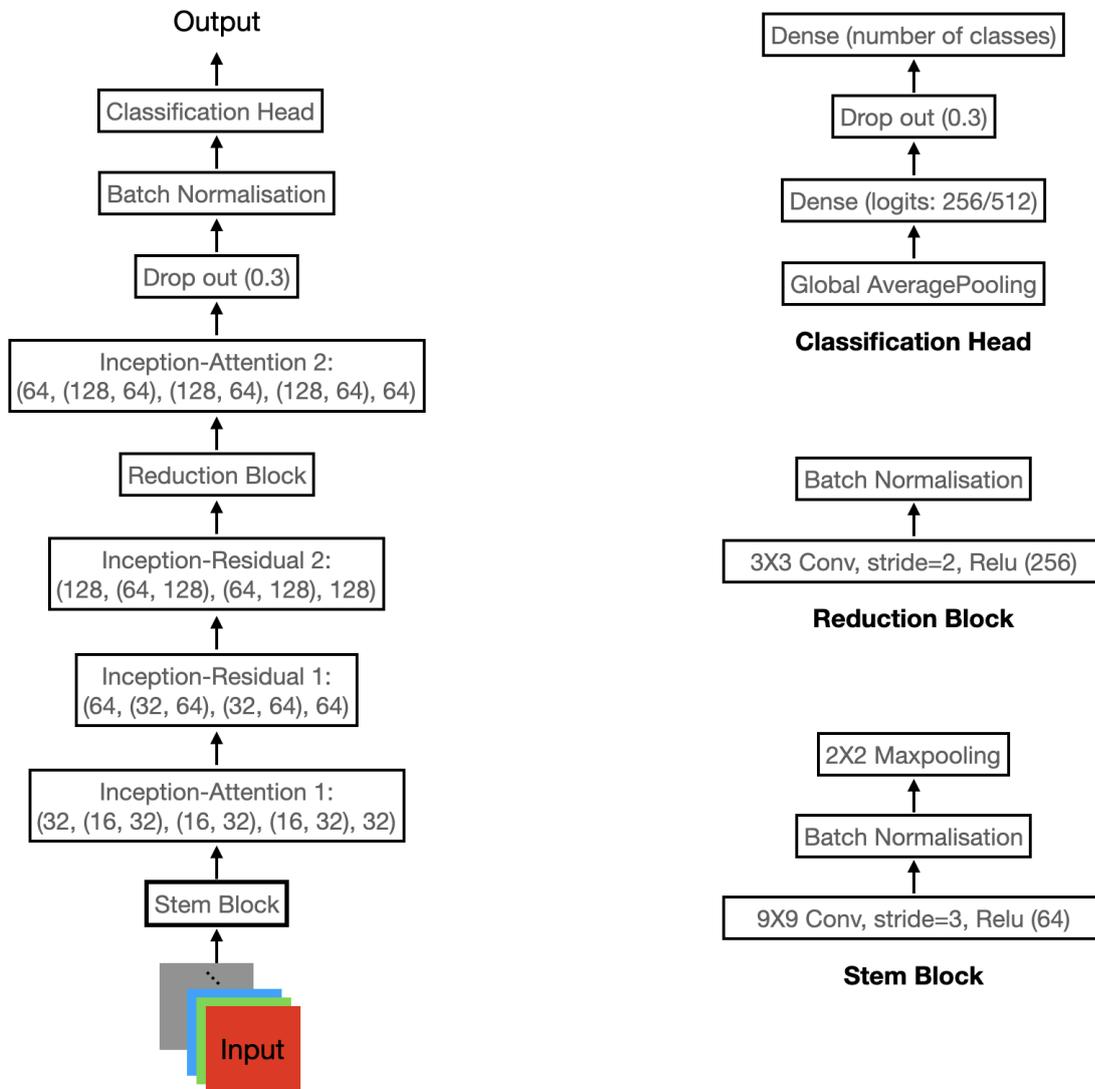


Figure 4.3: Structure of VIB_SD.

As demonstrated in Figure 4.3, VIB_SD contains two major modules: Inception-Residual (shown in Figure 4.5) and Inception-Attention (shown in Figure 4.6), both of which adopt multi-path architectures inspired by Inception-ResNet-V2 [297]. The multi-path architectures enable the two modules to extract features across multiple scales simultaneously, enhancing the robustness of VIB_SD in detecting smoke plumes of varying sizes. Unlike the residual block in Inception-ResNet-V2, the Inception-Residual module in VIB_SD employs more paths and larger kernels. This design helps the model learn more fine-grained features, accounting for the diverse characteristics of smoke plumes. Additionally, the Inception-Attention module enhances multi-scale feature learning by incorporating both Spatial Attention and Channel Attention, the other two key modules of VIB_SD.

Below is a more detailed description of the Spatial Attention, Channel Attention, Inception-Residual, and Inception-Attention modules. The Spatial Attention and Channel Attention modules are introduced first, as they form components of the Inception-Attention module.

1. Spatial Attention:

The Spatial Attention module is designed to learn the weight of each pixel within each channel of a feature map. These weights are learnt simultaneously, with the variations among them helping to deduce spatial associations between pixels, ultimately improving the model’s predictive performance.

As depicted in Figure 4.4 (a), the Spatial Attention module first reshapes the input feature map $F = [f_1, f_2, \dots, f_c] \in \mathbb{R}^{W \times H \times C}$ into a 2D vector $V = [v_1, v_2, \dots, v_l]$, where $l = W \times H$. Each $v_i = [p_1^i, p_2^i, \dots, p_C^i]$ is a 1D vector representing the pixel values at position i across all channels in F after flattening, with p_j^i denoting the pixel value at the j -th channel.

Next, V is passed through two fully connected layers, both activated using the sigmoid function. The intermediate output’s dimensionality is reduced by a ratio of $r = 16$ [43], minimising computational complexity. The result is then reshaped into the spatial attention distribution $S = [s_1, s_2, \dots, s_C]$, where $s_j \in \mathbb{R}^{W \times H}$ represents the spatial attention distribution of the channel f_j .

The final output of the spatial attention module, $O^s = [o_1^s, o_2^s, \dots, o_C^s]$, is obtained by multiplying the spatial attention distribution S with the feature map F , where $o_j^s = s_j \times f_j$. Further details can be found in [43].

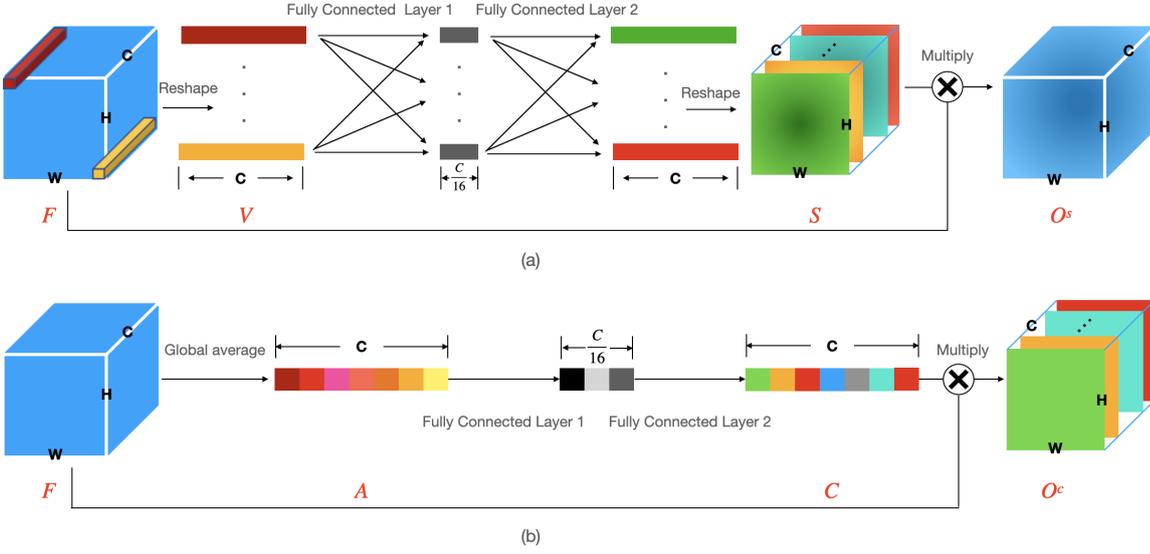


Figure 4.4: Attention modules. (a) Spatial Attention. (b) Channel Attention.

2. Channel Attention:

The Channel Attention module aims to learn the weight of each channel in a feature map. This weight indicates the importance of the channel in predicting the class of the image.

As illustrated in Figure 4.4 (b), for any input feature map $F = [f_1, f_2, \dots, f_c] \in \mathbb{R}^{W \times H \times C}$, the Channel Attention module first applies a global average pooling to generate a vector $A = [a_1, a_2, \dots, a_C]$, where $a_j \in \mathbb{R}$.

A is then transformed using two fully connected layers with a dimension reduction ratio $r = 16$ [43], activated by a ReLU function and a sigmoid function, respectively. The transformed output is the channel attention distribution $C = [c_1, c_2, \dots, c_C]$, where $c_j \in \mathbb{R}$ is the weight of channel f_j .

The final output of the Channel Attention module $O^c = [o_1^c, o_2^c, \dots, o_C^c]$ is then obtained by multiplying the channel attention distribution C by F , where $o_j^c = c_j \times f_j$. Readers can refer to [43, 46] for more details.

3. Inception-Residual:

The Inception-Residual module aims to learn residuals associated with spatial features in various scopes since information in the residuals may be important for detecting EF smoke that usually presents in a small area in the satellite imagery.

As displayed in Figure 4.5, the Inception-Residual module uses a four-path inception block with kernels of different sizes (1, 3, and 5, respectively, with another path employing max pooling) to achieve this purpose. Notably, depthwise-separable convolution [298, 276] was used to reduce computational complexity by using coupled $n \times 1$ and $1 \times n$ convolutional layers instead of $n \times n$ convolutional layers.

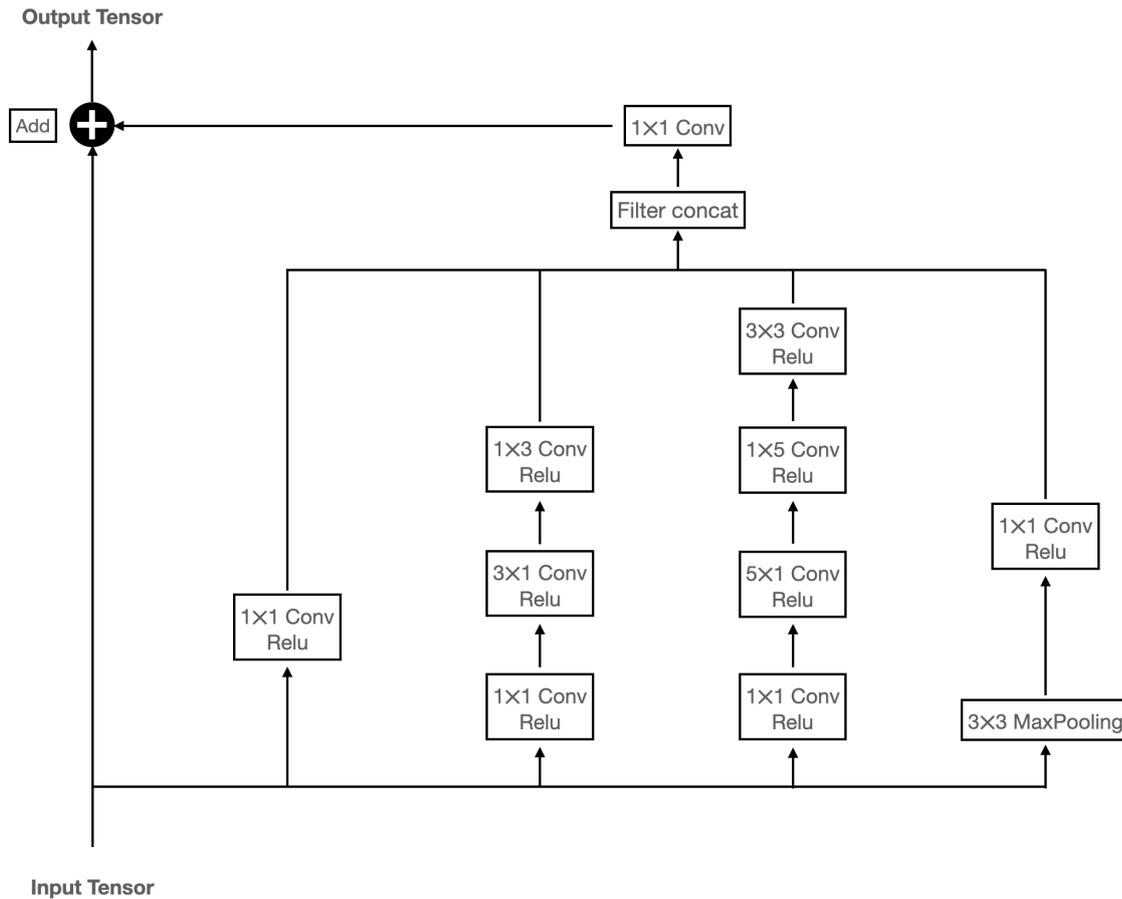


Figure 4.5: Inception-Residual module.

4. Inception-Attention:

The Inception-Attention module in VIB_SD aims to apply the attention mechanism to spatial features in various scopes extracted using kernels of different sizes.

As shown in Figure 4.6, the Inception-Attention module incorporates five paths. Specifically, spatial features were extracted in three different scopes through three paths. One path uses a kernel size of 3 to extract spatial features in small scopes. Another path uses a kernel size of 7 to extract spatial features in medium scopes. The third path adopts a kernel size of 11 to extract spatial features in large scopes. The three paths are each followed by a Spatial Attention module.

The remaining two paths, respectively employing a 1×1 convolutional layer and a 3×3 max pooling layer followed by a 1×1 convolutional layer, focus more on the channel-wise connections in the input tensor.

The feature maps generated from the five paths are concatenated, after which a Channel Attention module is used to allocate weights to the channels in the new feature map. This helps the model make predictions based on the importance of the extracted spatial features.

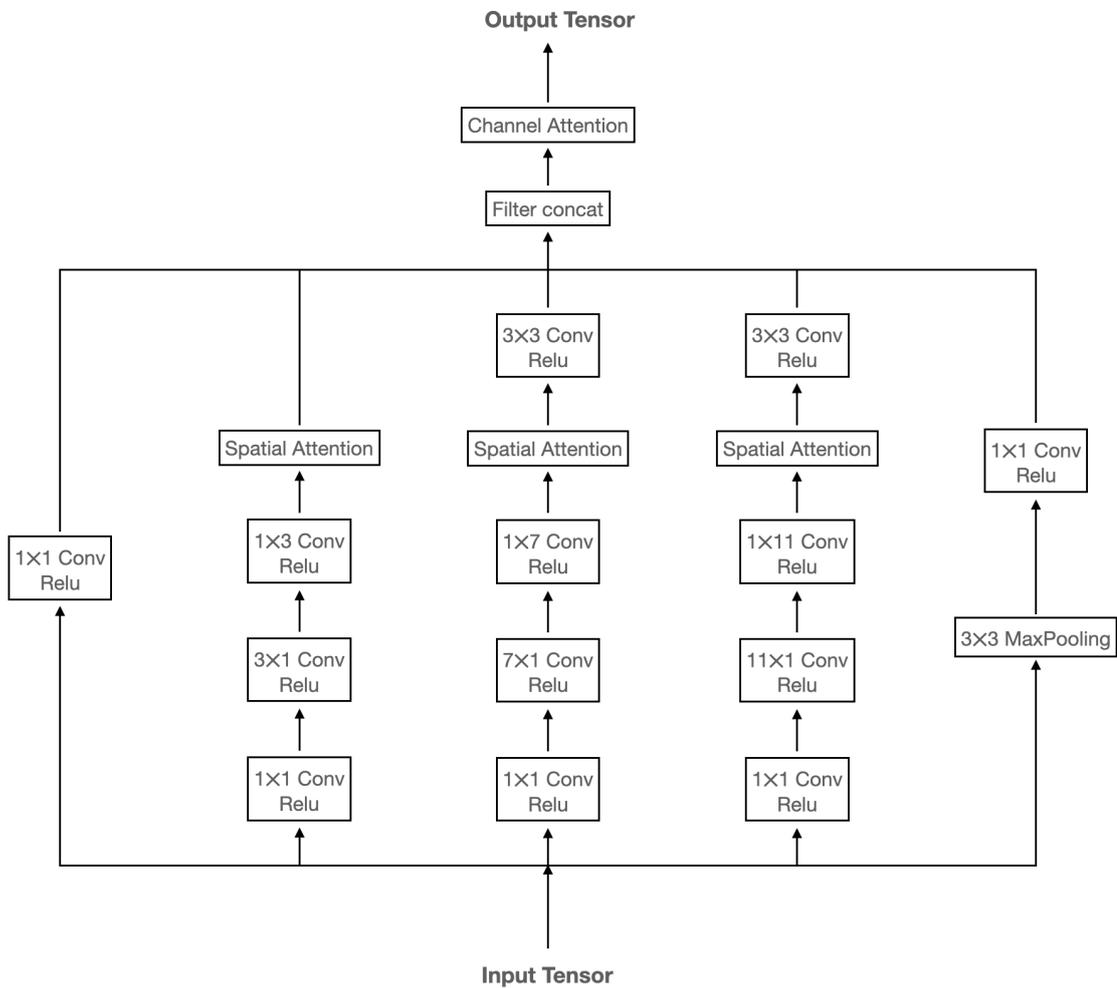


Figure 4.6: Inception-Attention module.

4.3 Experimental Settings and Evaluation Metrics

4.3.1 Experimental Settings

The work in this Chapter includes two separate experiments. The first experiment evaluates the performance of VIB_SD using the USTC_SmokeRS dataset. The second experiment investigates the impact of IR bands in satellite-based smoke detection, employing VIB_SD alongside the multispectral satellite imagery dataset Landsat6c.

In the first experiment, VIB_SD is trained and compared with SmokeNet [43], SAFA [47], and Inception-ResNet-V2 [297] under identical training settings. The USTC_SmokeRS dataset is used as SmokeNet and SAFA were originally trained on this dataset. As previously noted, SmokeNet was the best-performing model before SAFA, which is now considered state-of-the-art. Inception-ResNet-V2 features a similar key component to VIB_SD: a multi-path residual learning block.

The model comparison in this work does not extend to other potential DL models that could be applied for smoke detection, as the comparison against the previous and current state-of-the-art models is sufficient to evaluate VIB_SD's performance. Moreover, comprehensive comparisons, including models such as BoCF [299], RSSC-ETDL [49], LPDCMEN [51], D-CNN [48], and KFBNNet [50], among others, have already been conducted with SmokeNet and SAFA in [43, 47].

In the second experiment, VIB_SD is trained using Landsat6c with five different band combinations: RGB, RGBN, RGBNS1, RGBNS2, and RGBNS1S2. Here, N refers to the NIR band, S1 to the SWIR_1 band, and S2 to the SWIR_2 band. In subsequent sections of this Chapter, the five band combinations also refer to the five VIB_SD models derived from these combinations, where the context is clear.

The contribution of different IR bands can be evaluated with these five models as follows:

- The contribution of NIR is assessed by comparing the RGBN model to the RGB model.
- The contribution of the SWIR bands is assessed by comparing RGBNS1 or RGBNS2 to RGBN.
- The individual contribution of each SWIR band is assessed by comparing RGBNS1S2 to RGBNS1 and RGBNS2.

The two experiments share some common settings:

- In both experiments, the datasets are split into training (64%), validation (16%), and testing (20%) sets, with a batch size of 32, consistent with [43, 47].
- The same optimisation methods, regularisation strategies, and loss function are adopted in both experiments.
 - Adam [300] is used for optimisation, with the learning rate dynamically reduced from 0.01 by a factor of 0.2 when the validation loss fails to decrease after 20 epochs. The maximum number of epochs is increased to 500, compared to 200 in [43, 47], since the training accuracy shows potential for further improvement after 200 epochs. Early stopping is applied when the validation accuracy fails to improve within 90 epochs to avoid redundant training while ensuring optimal performance.
 - The loss function is sparse categorical cross-entropy, defined by the following formula:

$$\text{Loss} = - \sum_{i=1}^C y_i \cdot \log \hat{y}_i, \quad (4.3.1)$$

where C is the number of classes, \hat{y}_i is the predicted probability of an instance being the i -th class, and y_i is 1 if the ground truth label is the i -th class, and 0 otherwise.

However, the number of logits in VIB_SD differs between the two experiments, as shown in Figure 4.3. This number is set to 512 when using USTC_SmokeRS, as the dataset contains six classes. It is set to 256 when training VIB_SD with Landsat6c, as this dataset only includes three classes.

Other experimental settings are customised for each experiment as follows:

1. Customised settings for the first experiment:

To ensure fair model comparison, the same data pre-processing steps as in [43, 47] are followed.

The input images for training are resized to $230 \times 230 \times 3$ before random cropping to $224 \times 224 \times 3$. Training samples are augmented with random horizontal and vertical flipping. Validation and testing samples are resized directly to $224 \times 224 \times 3$. All samples are standardised using the “per_image_standardization” function provided by TensorFlow.

2. Customised settings for the second experiment:

All imagery files in the training data are augmented with random horizontal and vertical flipping, maintaining the original dimensions of $256 \times 256 \times 6$. Standardisation is not applied, as better training performance is observed without it. Augmentation is not applied to imagery files used for validation and testing.

The five models generated using the five band combinations are compared in two ways:

- (a) The training process is repeated 10 times, with samples generated by random splitting using the split ratios mentioned above. The samples obtained differ for each split. The overall performance of the models is compared. Variations in training data distributions ensure objective analysis of training performance.
- (b) All five models are trained using the same training and testing samples in a single random split. The models’ performance is compared, providing a direct comparison under identical conditions.

4.3.2 Evaluation Metrics

The work in this Chapter uses accuracy and Cohen’s Kappa coefficient (referred to as Kappa for brevity in the following content) as the evaluation metrics, following the approach in [43, 47] to facilitate model evaluation and comparison.

The formulas for calculating accuracy and the Kappa are defined in Table 4.5. N denotes the total number of images; i refers to a specific class; N_{ii} represents the number of true positive predictions for class i ; N_{i-} denotes the number of images from class i that were classified as other classes; N_{-i} refers to the number of images from other classes that were classified as class i .

Table 4.5: Formulas for accuracy and Kappa.

	Predicted Class 1	\dots	Predicted Class t
Actual Class 1	N_{11}	\dots	N_{1t}
\vdots	\vdots	\dots	\vdots
Actual Class t	N_{t1}	\dots	N_{tt}
Accuracy	$\frac{\sum_1^t N_{ii}}{N}$		
Kappa	$\frac{N \sum_1^t N_{ii} - \sum_1^t (N_{i-} N_{-i})}{N^2 - \sum_1^t (N_{i-} N_{-i})}$		

4.4 Results

4.4.1 Results of the First Experiment

Table 4.6 compares the parameter count, accuracy, and Kappa of the four models. The results show that VIB_SD significantly reduces the parameter count, with only a minor reduction in accuracy compared to the state-of-the-art model, SAFA. However, VIB_SD achieves higher accuracy and a better Kappa than both SmokeNet and Inception-ResNet-V2. Notably, VIB_SD uses less than 2% of SAFA’s parameters.

Table 4.6: Model performance comparison.

Model	Parameter Count	Accuracy	Kappa
SmokeNet	53.5M	92.75%	0.9130
SAFA	84.2M	96.22%	0.9546
Inception-ResNet-V2	54.4M	91.33%	0.8958
VIB_SD	1.66M	93.57%	0.9227

The results indicate VIB_SD’s great potential for onboard-SmallSat applications in accurate smoke detection.

4.4.2 Results of the Second Experiment

Table 4.7 presents the parameter count, the accuracy and Kappa ranges within the 95% confidence interval, along with the best accuracy and Kappa for each model, based on 10 results.

Table 4.7: Performance of models using variant bands based on 10 results.

Model	Parameter Count	Accuracy	Best Accuracy	Kappa	Best Kappa
RGB	1.660 M	$83.28 \pm 1.57\%$	86.45%	0.7488 ± 0.0234	0.7964
RGBN	1.666 M	$87.78 \pm 1.38\%$	92.41%	0.8164 ± 0.0207	0.8861
RGBNS1	1.671 M	$87.78 \pm 1.12\%$	89.97%	0.8164 ± 0.0168	0.8491
RGBNS2	1.671 M	$86.4 \pm 0.09\%$	89.43%	0.7956 ± 0.0135	0.8413
RGBNS1S2	1.68 M	$86.21 \pm 1.18\%$	89.16%	0.7929 ± 0.0178	0.8373

According to Table 4.7, the RGB model shows the lowest accuracy and Kappa, while the RGBN model achieves the highest accuracy and Kappa under both metrics. This suggests that adding the NIR band improves model performance. However, accuracy and Kappa unexpectedly decreased when the SWIR bands were added in combination with the NIR band. Specifically, a larger decrease in both metrics was observed when both SWIR bands were added compared to when only one SWIR band was included. Potential reasons for these results will be discussed in Section 4.6.

Figure 4.7 presents boxplots of the accuracy and Kappa derived from the 10 results of the five models.

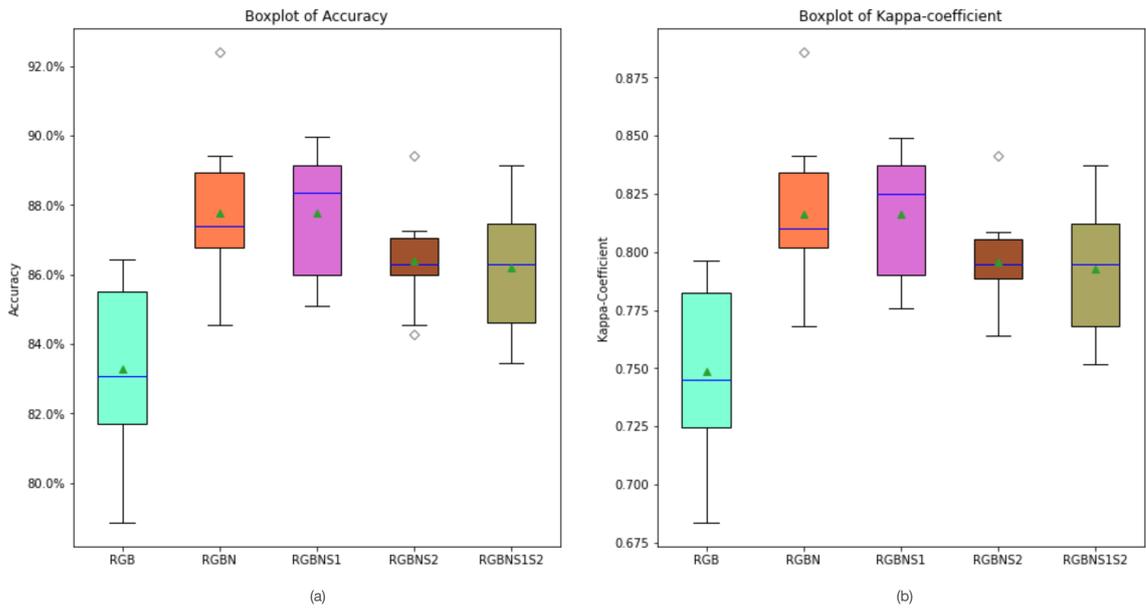


Figure 4.7: Boxplots based on the 10 results. (a) Accuracy. (b) Kappa.

The boxplots in Figure 4.7 indicate that the RGB model is less effective than the other four models. In contrast to what Table 4.7 suggests by evaluating the models using mean values, Figure 4.7 indicates that RGBNS1 has the highest median accuracy and Kappa.

Table 4.8 presents the results of the five models trained and tested using the same samples obtained from a single split. Since all models were trained on the same training samples and tested on the same testing samples, the models can be compared more fairly, although these results may not represent their best possible accuracy and Kappa. Boldface font highlights the best accuracy and Kappa.

Table 4.8: Performance of models using variant bands with the same samples.

Model	Testing Accuracy	Kappa
RGB	83.20%	0.7483
RGBN	84.82%	0.7723
RGBNS1	85.64%	0.7842
RGBNS2	85.64%	0.7843
RGBNS1S2	86.45%	0.7966

Table 4.8 shows a slightly different trend: the more bands a model includes, the better its performance. It is also worth noting that RGBNS1 and RGBNS2 achieved the same prediction accuracy, with minor differences in their Kappa. This indicates that although both models correctly predicted the same number of images, the distribution of true positive predictions across classes varied.

Tables 4.7 and 4.8, along with Figure 4.7, all suggest that using additional IR bands can effectively improve model prediction accuracy. In particular, adding the NIR band greatly enhances accuracy compared to using only the RGB bands.

Based on the fair comparison results in Table 4.8, the following inferences can be made:

- Both the SWIR_1 and SWIR_2 bands contain valuable information for smoke detection; adding either contributes similarly to improving prediction accuracy.
- The SWIR_1 and SWIR_2 bands provide distinctive information, and adding both further enhances prediction accuracy.

4.5 Case Studies

To examine the effectiveness of using multispectral moderate spatial resolution imagery for smoke detection, the RGBNS1S2 model was used to make predictions on four

different smoke scenes captured by Landsat 8 OLI, as shown in Figure 4.8. None of the four scenes were included in the training dataset. The best-performing weights of RGBNS1S2, which achieved the highest accuracy of 89.16%, were selected for these predictions.

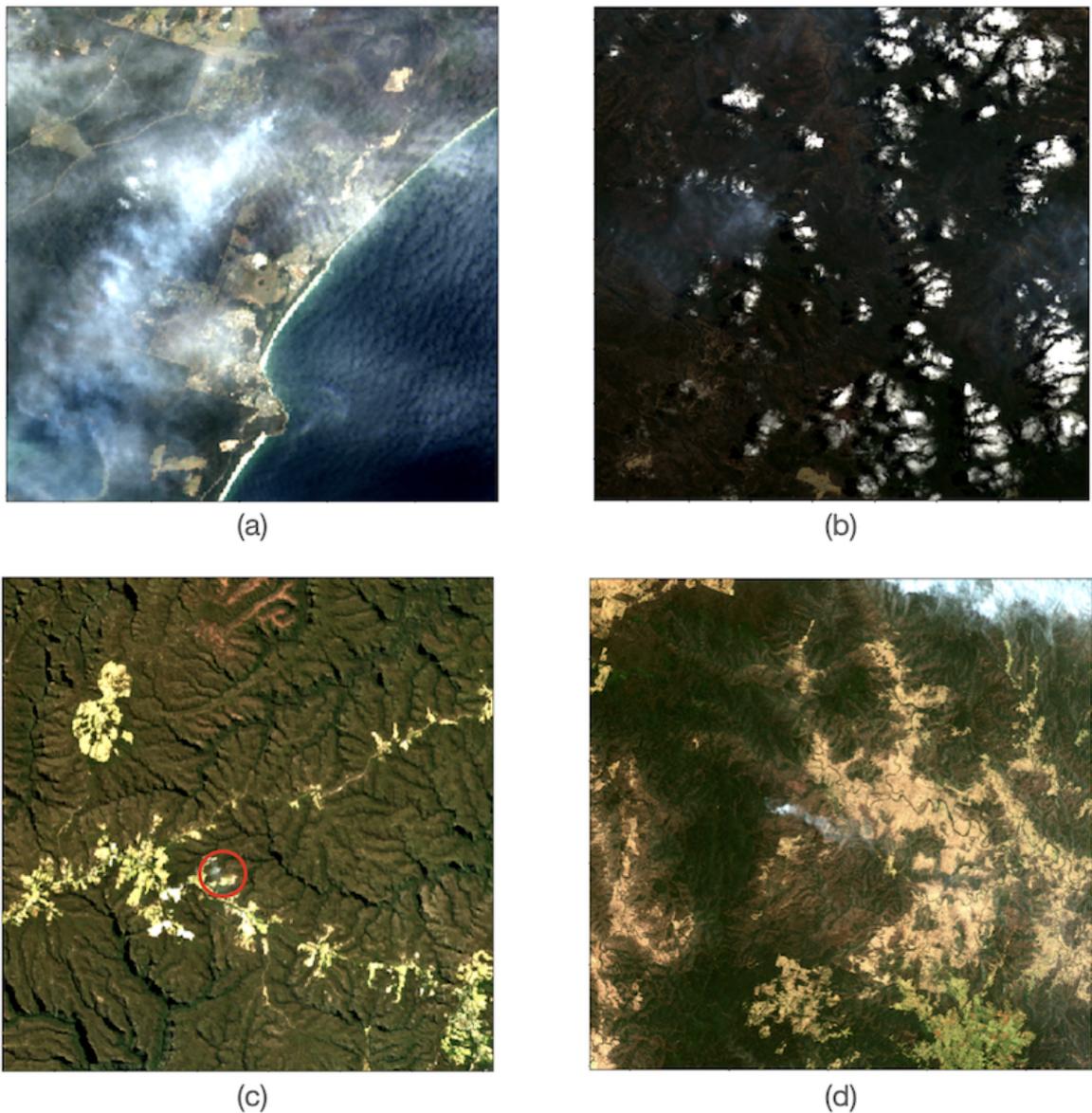


Figure 4.8: Smoke scenes. (a) Smoke mixed with thin clouds above the seaside. (b) Diffused smoke at multiple sites under altocumulus clouds. (c) Cloud-free smoke (in the red circle) over a very small geographic extent. (d) Cloud-free smoke plumes of varying scales at two different sites.

The smoke scenes were tiled first, and predictions were then conducted on the

resulting patches. Since the area covered in scene (a) was too small to be properly tiled with a 50% overlap, the overlap rate was increased to 75% for scene (a) during prediction. For the other three scenes, the overlap rate remained at 50%.

The prediction results for scene (a) are shown in Figure 4.9. In the results, the text above each patch displays the patch ID, the predicted class (where CLR refers to “Clear”, SMK refers to “Smoke”, and OA refers to “Other_aerosol”), and the probability of the predicted class. The text below each patch shows the probabilities of the patch being “Clear”, “Other_aerosol”, or “Smoke” from left to right.

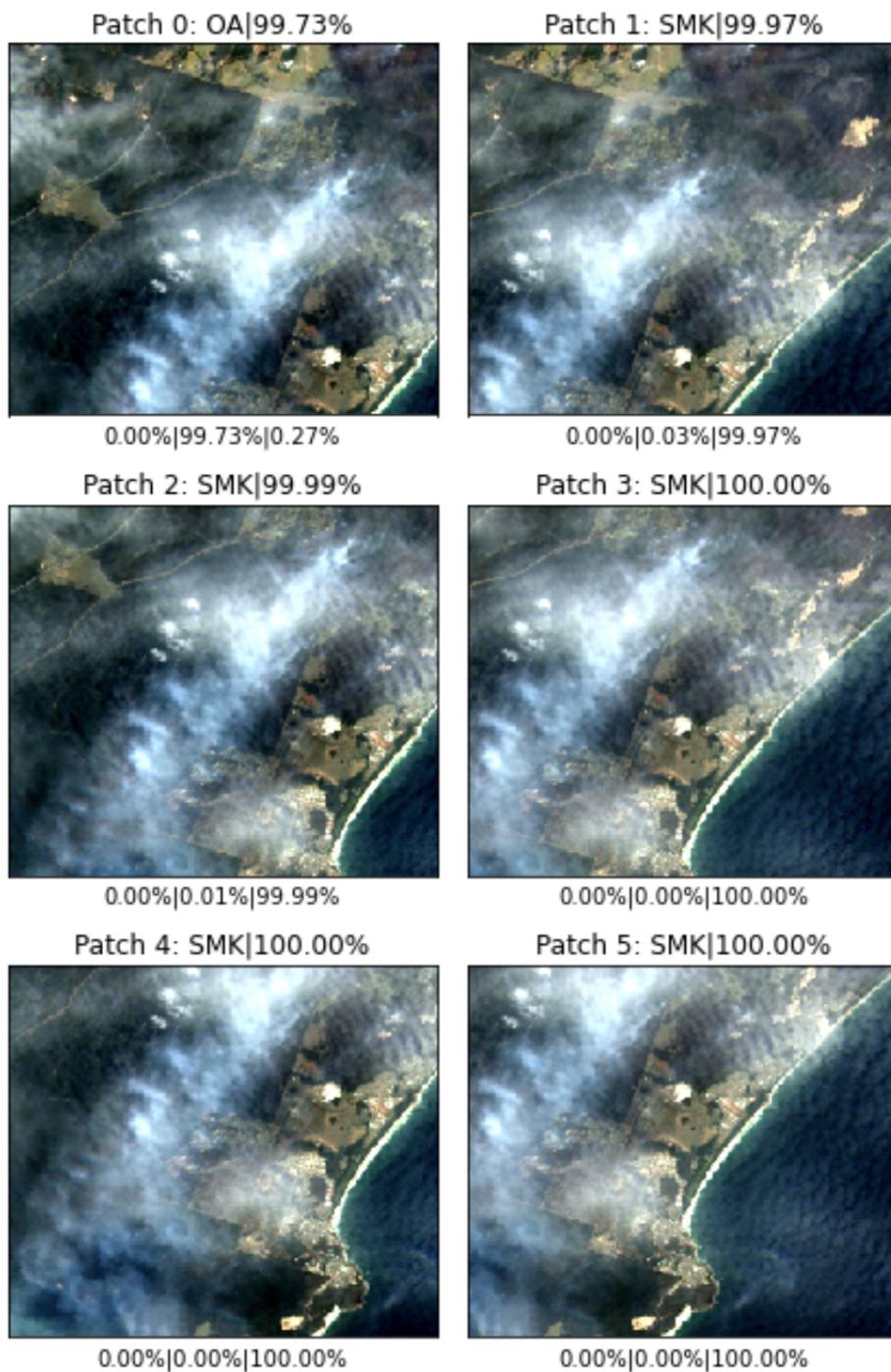


Figure 4.9: Prediction results for scene (a) in Figure 4.8.

In Figure 4.9, all patches were correctly predicted except for patch 0, which was incorrectly classified as “Other_aerosol”.

The prediction results for scene (b) are shown in Figure 4.10.

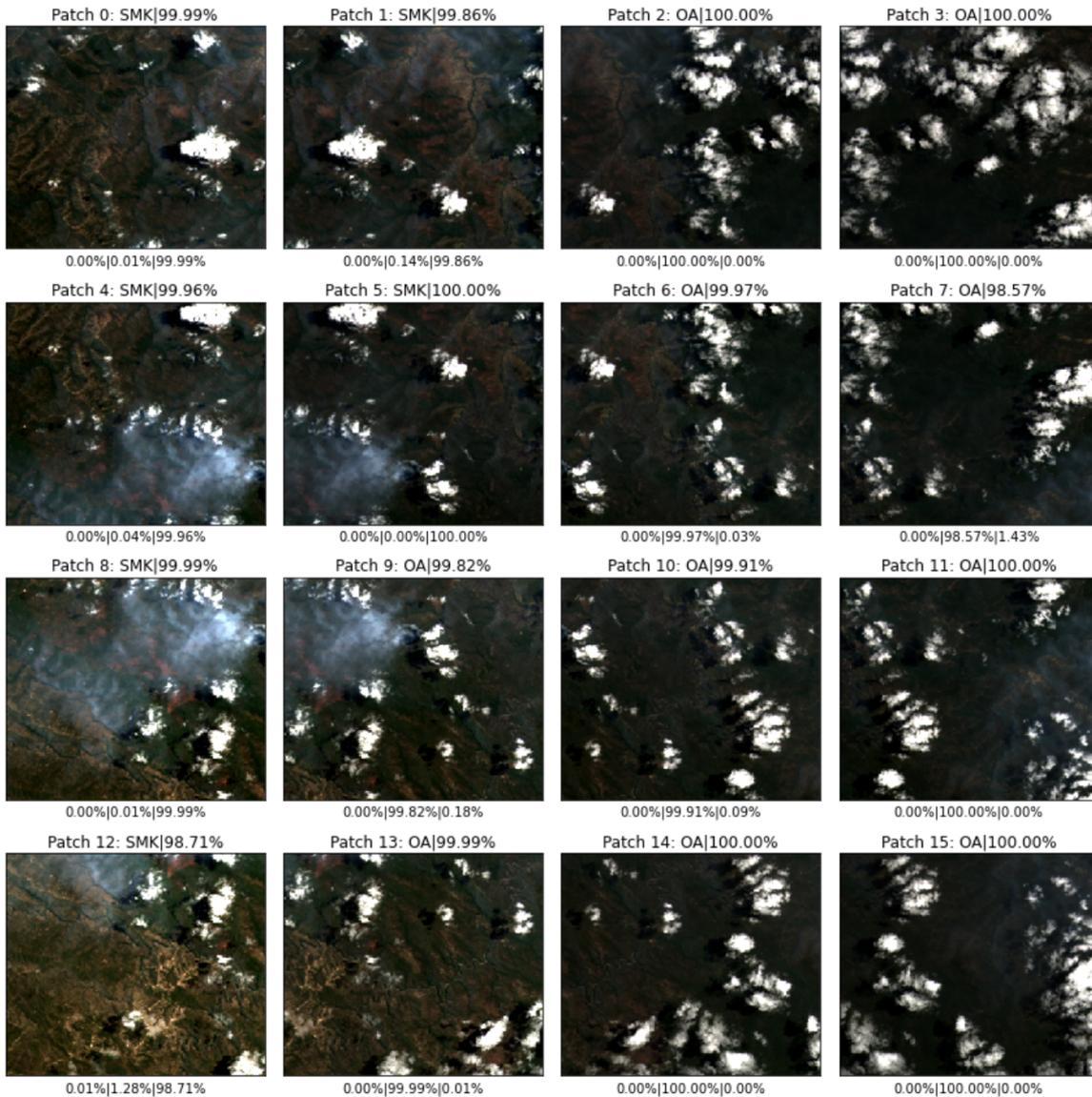


Figure 4.10: Prediction results for scene (b) in Figure 4.8.

In Figure 4.10, patches 0, 1, 4, 5, 8, and 12 were correctly predicted as “Smoke”. The remaining patches were also “Smoke” but were incorrectly classified as “Other_aerosol”.

The prediction results for scene (c) are shown in Figure 4.11.

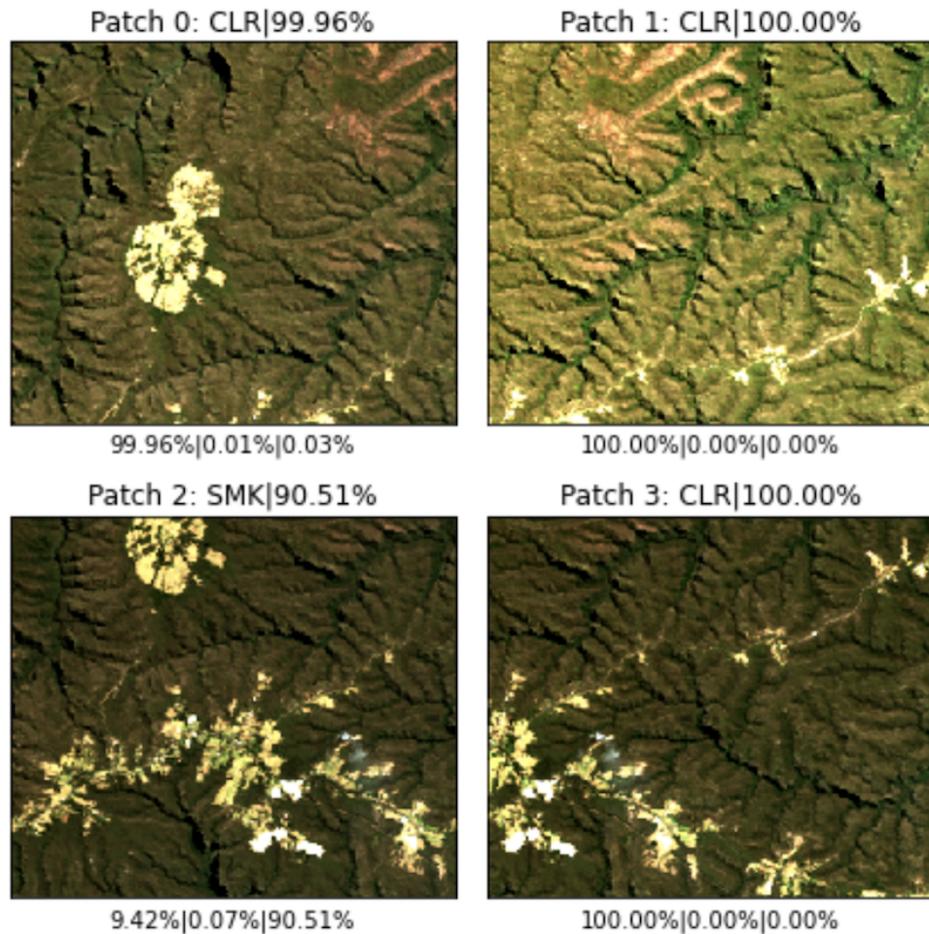


Figure 4.11: Prediction results for scene (c) in Figure 4.8.

In Figure 4.11, the two patches at the top were both correctly predicted as “Clear” with high probability. The two patches at the bottom both contain smoke on a very small scale; however, only the left patch was correctly predicted, while the right patch was incorrectly classified as “Clear”.

Since scene (d) is very large, displaying the results for all patches is impractical. Therefore, only the patches containing smoke were selected to verify the prediction performance. The prediction results for the smoke area in the top right corner of scene (d) are shown in Figure 4.12, while the prediction results for the smoke area in the middle of scene (d) are shown in Figure 4.13.

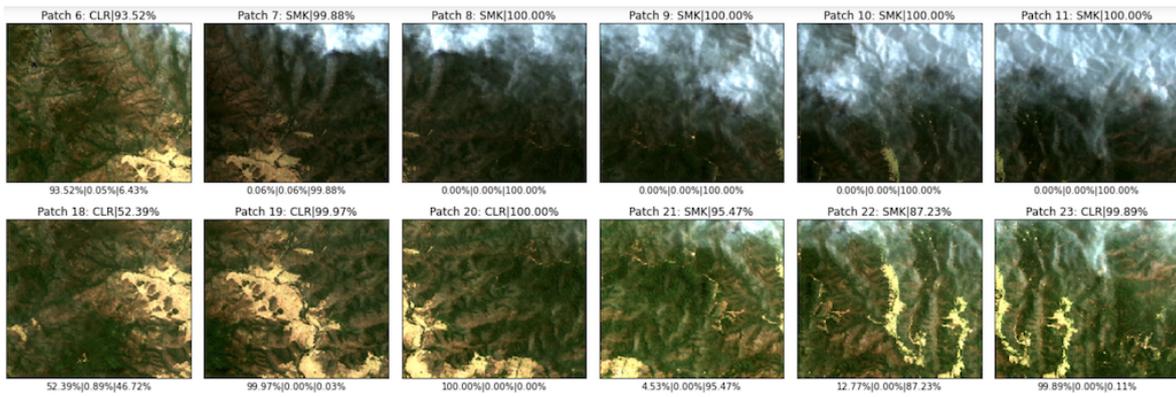


Figure 4.12: Prediction results for the top right smoke area of scene (d) in Figure 4.8.

In Figure 4.12, almost all patches were correctly predicted, except for patches 6 and 23, which both contain smoke but were incorrectly classified as “Clear”.

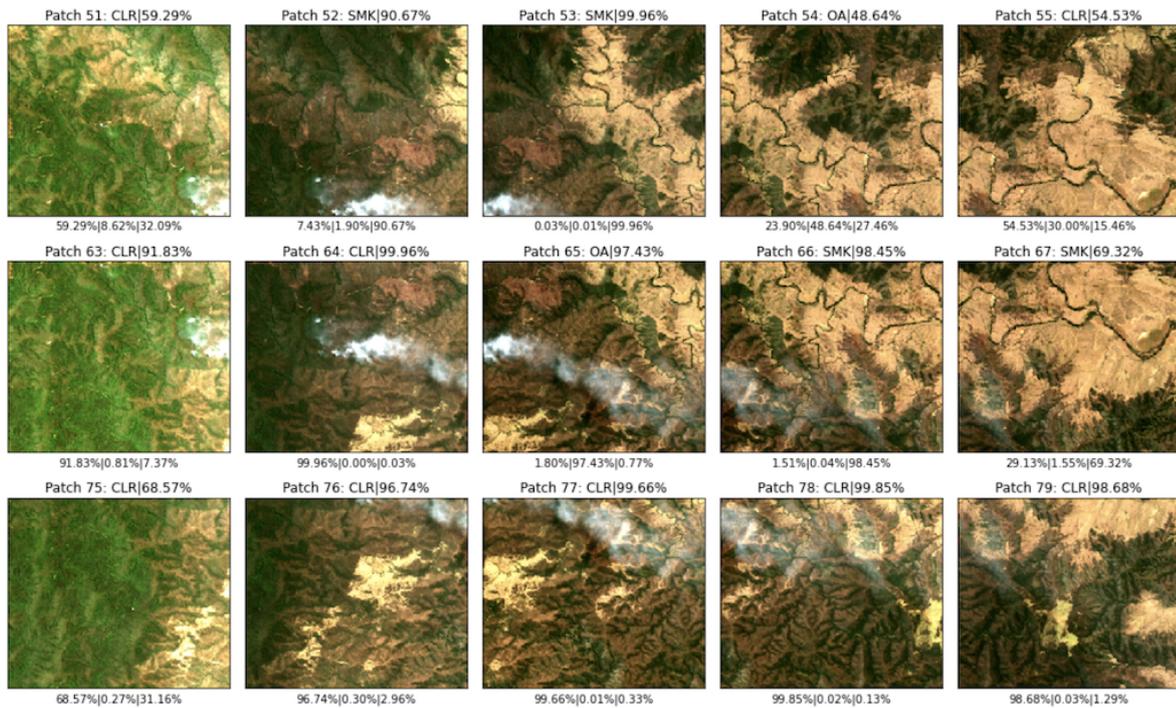


Figure 4.13: Prediction results for the middle smoke area of scene (d) in Figure 4.8.

In Figure 4.13, patches 52, 53, 66, and 67 were correctly predicted as “Smoke,” while patches 51, 63, 64, 65, 76, 77, 78, and 79 were incorrectly classified as either “Clear” or “Other_aerosol.”

The above results demonstrate that:

- The RGBNS1S2 model achieves good overall accuracy;
- Smoke in the Landsat imagery can be detected, although some FNs are present;
- FN detections are more likely to occur where the smoke is small or mixed with clouds.

4.6 Discussion

The results shown in Table 4.8 aligned with the expectation that using more spectral bands can effectively improve model prediction accuracy. However, the results shown in Table 4.7 and Figure 4.7 raised a question: why did both the RGBNS2 and RGBNS1S2 models yield worse overall performance compared to RGBN and RGBNS1?

Several factors could contribute to this issue. One factor might be linked to the imbalance of smoke scenes in the dataset. This may be inferred from the skewed distribution of the accuracy and Kappa of the models in Figure 4.7. Smoke could be obscured by clouds, in dark colours hidden within a dark background, located at the corners or edges of the images, or be too small or thin. When samples are randomly split, the training set may contain predominantly one type of smoke scene, while the testing set contains others, leading to compromised training and testing performance. In contrast, when the testing samples are more similar to the training samples, the models are likely to perform better. In this case, the overall performance of the models might change if they are trained multiple times. However, repeating the training process is time-consuming and may not guarantee different results. Expanding the dataset to provide more evenly distributed training and testing samples during the random splitting process could be a more effective solution.

Another factor could be whether the model is effectively learning from the additional spectral bands. During backpropagation, the model needs to update the weights for all input bands and extract useful features for prediction. If the weights for certain spectral bands are not properly updated, these bands may be treated as noise, negatively impacting performance. This issue could be related to the size of the training dataset or the model design. To address this, training with a larger dataset or further

refining the model design may be necessary. Future work will consider expanding the current dataset and fine-tuning the model structure to improve performance.

The case studies demonstrated the effectiveness of VIB_SD in detecting EF smoke when trained using Landsat6c with all six bands. Although the model’s prediction stability can be further improved, it shows potential for EF detection. Notably, while some “Smoke” patches were misclassified as “Clear” or “Other_aerosol,” “Clear” or “Other_aerosol” patches were rarely misclassified as “Smoke.” This implies a low false positive rate for predicting “Smoke,” making positive smoke predictions trustworthy. Since predictions are conducted on overlapping patches, even if some smoke patches are misclassified, the fire alarm could still be triggered as long as one smoke patch is correctly predicted.

Furthermore, using multi-source satellite imagery to achieve timely detection of EF smoke will be explored. The temporal resolution of Landsat 8 OLI is 16 days, which is insufficient for timely detection. However, combining imagery from multiple low-temporal-resolution satellites can collectively provide a much higher temporal resolution. Future research will aim to collect imagery datasets from additional satellites (e.g., Sentinel-2) to enable timely detection of EF smoke using multi-satellite imagery.

4.7 Conclusion

To facilitate satellite-based smoke detection, a multispectral imagery dataset, Landsat6c, was constructed using data from satellite sensors with moderate spatial resolution: Landsat 5 TM and Landsat 8 OLI. A lightweight CNN model, VIB_SD, was developed that could potentially be adopted for onboard-SmallSat applications, featuring significantly reduced parameters with only minor compromises in accuracy. Five models based on VIB_SD were trained with Landsat6c using different band combinations to evaluate the effectiveness of multispectral moderate spatial resolution imagery in EF smoke detection. The experimental results demonstrated that incorporating all three additional IR bands can effectively improve detection accuracy. The RGBS1S2 model was applied to real smoke scenes, with results indicating that the model can effectively detect EF smoke in various scenarios, though prediction stability requires further investigation. Future work will focus on refining the VIB_SD structure, expanding the dataset, collecting new data from other satellites (e.g., Sentinel-2), and integrating multiple data sources for timely EF smoke detection.

It is worth noting that, although beyond the scope of this thesis, the efficacy of deploying VIB_SD onboard SmallSats for smoke detection has been validated through the work in the fourth publication [55] listed at the beginning of this thesis. Specifically, VIB_SD was used as a model prototype for simulating smoke detection onboard Kanyini, a co-designed SmallSat by SmartSat CRC and the SA government.

The imagery data used for this simulation were synthetic HS2 (the sensor of Kanyini) 49-band hyperspectral imagery tiles. VIB_SD demonstrated optimal resource usage when performing onboard inference on 84 such 256×256 imagery tiles: processing time of 1.6 seconds, average memory usage of 29 MB with a peak memory usage of 54 MB, and power consumption of 1.31 W. VIB_SD also achieved high prediction accuracy (95.7%) with a low false negative rate (FNR) (2.3%).

Chapter 5

IA: Learning Class-oriented Spectral Patterns for smoke detection

One significant challenge of satellite-based smoke detection is the similarity in spatial and spectral characteristics between smoke and other confounding aerosols (e.g., clouds, haze, dust), as discussed in Chapter 1 and demonstrated in Figure 1.2. Early studies, primarily focusing on pixel-level detection, sought to derive various threshold values from reflectance and BT values in certain spectral bands to differentiate smoke and cloud pixels, based on their distinct spectral characteristics. These distinct spectral characteristics, represented by pixel-level threshold values or other potential spectral patterns (see Definition 2.2.1 in Chapter 2), provide an effective means of accurately detecting smoke among confounding aerosols.

However, these threshold values were derived using deterministic methods based on domain knowledge (e.g., selecting specific bands and determining thresholds), making them difficult to generalise or directly apply in DL models. Current DL models, including those customised for smoke detection, are not designed to extract and utilise these important spectral patterns effectively. DL models typically employ 3×3 or larger kernels in their initial layers, focusing more on exploring spatial features within a 3×3 or larger window, rather than discovering spectral patterns at the pixel level. Additionally, previous studies on satellite-based smoke detection were conducted using the MODIS RGB imagery dataset USTC_SmokeRS [43], limiting the exploration of spectral patterns due to the absence of IR bands.

No study has yet investigated how to extract useful spectral patterns to aid DL-based remote sensing classification tasks, specifically for smoke detection. Although the VIB_SD model proposed in Chapter 4 demonstrated significantly improved accuracy when trained using the multispectral dataset Landsat6c, it is not equipped with an effective mechanism to explicitly and proactively extract useful spectral patterns. Like previous DL models, VIB_SD uses larger kernels in its initial layers and merely takes the multispectral imagery data as input.

Further study is required to develop such a mechanism and examine whether extracting spectral patterns in the initial layers of a DL model can effectively improve its performance for satellite-based smoke detection. This chapter responds to this requirement and addresses the second research question posed in Chapter 1:

RQ-2 How can smoke detection accuracy be improved more effectively using multispectral satellite imagery, beyond simply including additional IR bands in model training?

Consequently, this chapter introduces the DL module IA, designed to enable DL models to automatically learn class-oriented spectral patterns from multispectral satellite imagery, thereby improving smoke detection accuracy. IA can be conveniently integrated with different DL architectures. The effectiveness of the IA module is evaluated on different CNN architectures using two satellite imagery datasets: USTC_SmokeRS, derived from MODIS with three spectral bands, and Landsat6c, derived from Landsat 5/8 with six spectral bands. Experimental results demonstrate that the IA module enhances smoke detection accuracy in CNN models. Additionally, spectral patterns extracted by the IA module are visualised using test imagery, showing that the IA module can effectively extract class-oriented spectral patterns.

The work presented in this chapter is based on my paper *“Learning Class-Specific Spectral Patterns to Improve Deep Learning-Based Scene-Level Fire Smoke Detection from multi-spectral Satellite Imagery”* [56], published in the journal *“Remote Sensing Applications: Society and Environment”* in February 2024.

The content of this paper is organised as follows: Section 5.1 introduces the background. Section 5.2 presents the proposed IA module. Section 5.3 describes the experimental settings, including datasets, training settings, and evaluation metrics. Section 5.4 interprets the results, including ablation studies and parameter selection for the IA module. Section 5.5 discusses potential applications of IA and future work. Section 5.6 presents the conclusion.

5.1 Introduction

Detecting EF smoke from satellite imagery is recognised as an effective and timely approach to preventing fire disasters, as smoke plumes are typically the first visible indicators of wildfires from space. By detecting smoke, small fires burning at lower temperatures, such as some early-stage grass fires, can be more easily identified. However, as demonstrated in Figure 1.2, smoke shares similar spatial and spectral characteristics with other aerosols (e.g. cloud, fog, haze, and dust) and often intermingles with these aerosols in satellite imagery, making accurate detection of smoke amongst them extremely challenging.

Early research on smoke detection from satellite imagery primarily focused on the

pixel level. Researchers used mathematical and statistical methods to derive threshold values from the reflectance and BT values in multiple spectral bands for each pixel. Smoke pixels were then distinguished from others based on the differences in these spectral-band threshold values [39, 234, 40, 237, 54]. These approaches demonstrated that specific spectral patterns are valuable for accurate smoke detection using satellite imagery. However, these threshold values heavily rely on domain knowledge and may be influenced by local conditions. For example, fuel types, which often vary with geographic location, may change the physical and chemical characteristics of smoke, making it difficult to generalise these thresholds.

DL models like CNNs and vision transformers (ViTs) can automatically extract highly abstract features without the need for cumbersome feature engineering and have been proven to greatly improve scene-level smoke detection accuracy. However, while current CNNs and ViTs excel in learning spatial patterns, they are not designed to effectively learn spectral patterns, which provide valuable information at the pixel level. Notably, CNNs employing 3×3 or larger kernels in the initial layers do not explicitly extract pixel-level spectral features from the input. In contrast, ViTs, incorporating the self-attention mechanism on divided subregions of the input imagery, focus more on spatial correlations of these subregions.

Take the CNN model VIB_SD, proposed in Chapter 4 and specifically designed for scene-level smoke detection, as an example. VIB_SD employs 9×9 kernels for spatial feature extraction right after the input layer. It demonstrated improved accuracy when trained with additional IR bands compared to when trained with only the RGB bands. Nonetheless, a relatively high FNR was still observed when smoke and other confounding aerosols co-exist, implying that even spatial patterns learned from multispectral imagery may not be adequate to differentiate these aerosols effectively. Additionally, the varying shapes and colours of smoke, as demonstrated in Figure 1.1, further indicate that spatial patterns alone are insufficient for accurate smoke detection from satellite imagery.

One reasonable hypothesis is that the accuracy of DL models for smoke detection can be improved by explicitly extracting spectral patterns, relating to individual pixels, at the very beginning of the model, particularly when using multispectral satellite imagery.

To validate this hypothesis, this chapter presents the IA module, designed to en-

able a DL model to automatically learn class-oriented spectral patterns that are useful for distinguishing smoke from the background and other visually similar objects. Unlike previous methods deriving threshold values based on domain knowledge, which are hard to generalise, IA automatically learns spectral patterns through supervised training without human intervention. After integrating with IA, DL models take the IA-extracted spectral patterns as input, thereby achieving enhanced performance.

The study further evaluates IA’s effectiveness using multiple baseline CNN models with variant architectures, namely ResNet50 [44], InceptionResnetV2 [297], MobileNetV2 [209], and VIB_SD. The models are trained both with and without IA, utilising two datasets: the three-band (RGB) USTC_SmokeRS MODIS dataset and the six-band (i.e., RGB, NIR, SWIR_1, and SWIR_2) Landsat6c dataset proposed in Chapter 4. Results show that incorporating the IA module effectively improved the prediction accuracy of the baseline CNN models for both datasets.

The IA module’s novelty and contributions are summarised as follows:

- IA uniquely facilitates class-oriented spectral pattern learning in DL models, a pioneering achievement in the literature. It accomplishes this by synergistically employing band, spatial, and channel attention directly on the input imagery.
- Designed to be lightweight as a pre-processing module, IA integrates seamlessly with various DL models, such as CNNs or VITs, adding minimal computational overhead. This feature is crucial for onboard satellite applications, where resource constraints are a significant concern.
- Beyond smoke detection, IA’s potential extends to other remote sensing domains like water observation and vegetation disease detection, where spectral patterns are pivotal (this hypothesis is left for verification by researchers studying in these domains). Additionally, the class-oriented spectral patterns learned by IA could facilitate DL model interpretation and cross-sensor transfer learning, given their informativeness and reduced variability compared to individual spectral bands.

5.2 IA

Figure 5.1 (a) depicts the structure of the IA module which has been devised to serve the following objectives:

1. Automatic extraction of pixel-level spectral patterns;

2. Extraction of multiple spectral patterns concurrently, with a focus on those containing valuable class-oriented information;
3. Integration of spectral patterns and spatial patterns to enhance the accuracy of scene-level smoke detection.

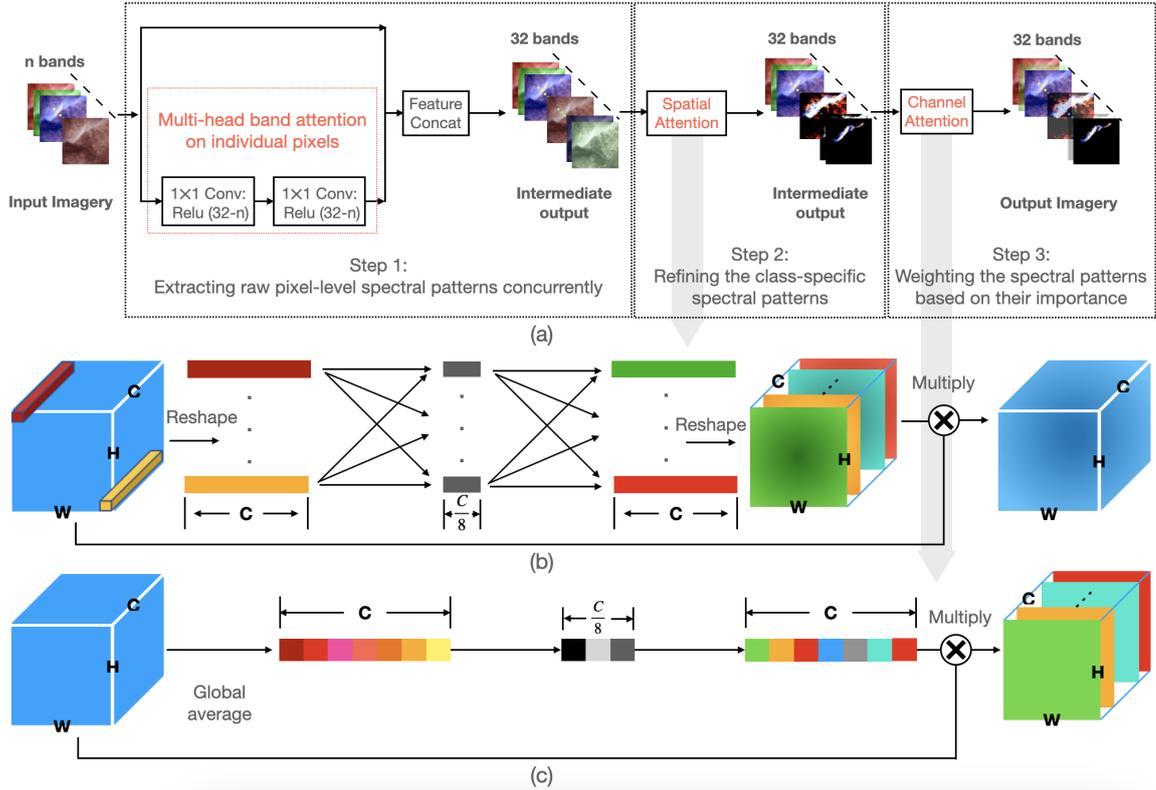


Figure 5.1: The architecture of the IA module. (a) IA (b) Spatial Attention* (c) Channel Attention**

* Modified based on Figure 4.4 (a), with the channel reduction ratio changed from 1/16 to 1/8;

** Modified based on Figure 4.4 (b), with the channel reduction ratio changed from 1/16 to 1/8.

The IA module accomplishes the above objectives through three successive steps that employ different types of attention:

1. Band attention is applied to individual pixels across the input spectral bands, which produces raw spectral patterns.
2. Spatial attention is employed on pixels within each spectral band and spectral pattern, leading to the refinement of raw spectral patterns.

3. Channel attention is utilised on the combined spectral bands and spectral patterns, focusing on significant spectral patterns.

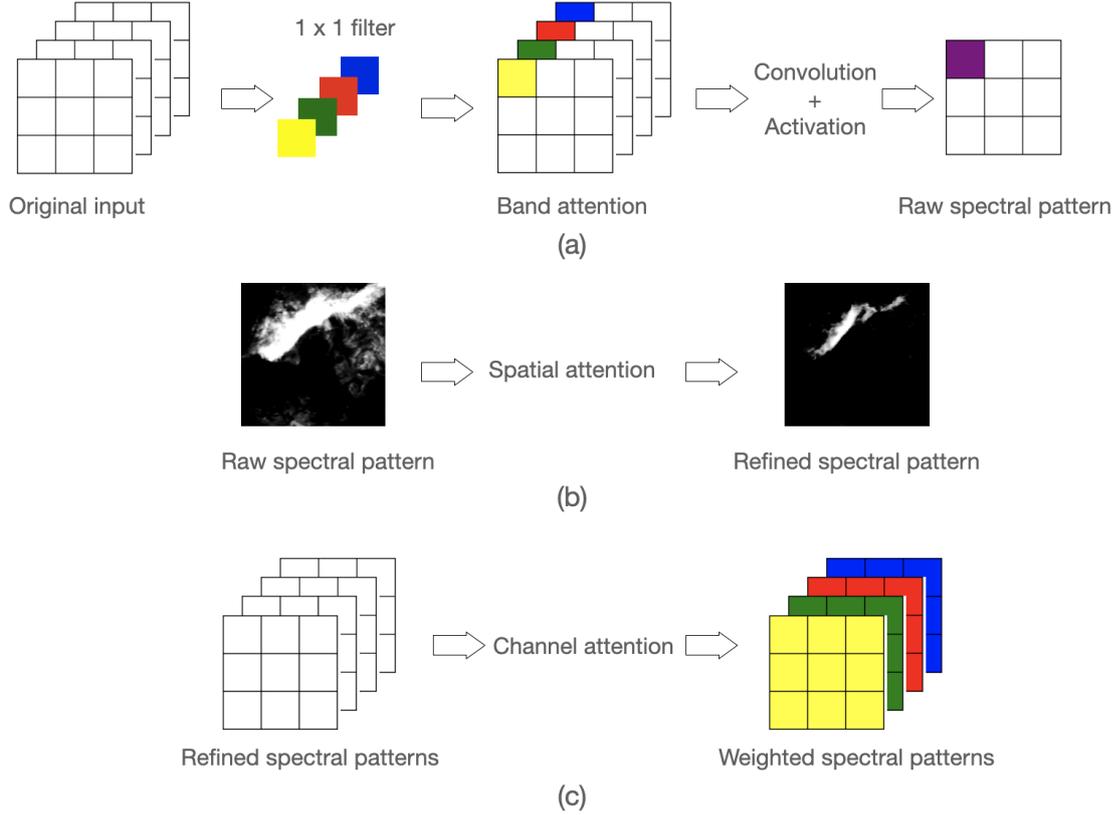


Figure 5.2: Process of spectral pattern extraction by IA. (a) Raw spectral pattern extraction is achieved through band attention by using 1×1 filters with non-linear activation (b) Spatial attention emphasises the pixels that are spatially related while suppressing other pixels, hence refines the raw spectral patterns (c) Channel attention weights the spectral patterns according to their importance in the classification task.

In the first step, the IA module employs 1×1 filters to learn the attention weights of a pixel's spectral bands. A 1×1 filter $F = [f^1, \dots, f^k, \dots, f^C]$ ($f^k \in \mathbb{R}$) linearly maps any pixel $P_{(i,j)} = [p_{(i,j)}^1, \dots, p_{(i,j)}^k, \dots, p_{(i,j)}^C]$ ($p_{(i,j)}^k \in \mathbb{R}$) to a new value $p_{(i,j)}^f = \sum_{k=1}^C f^k p_{(i,j)}^k$. The ReLU function then applies a non-linear transformation and produces a new value $p_{(i,j)}^{\text{ReLU}(f)}$. This process aims to extract spectral patterns, as defined in Definition 2.2.1. Since a 1×1 filter applies weights across all bands to a pixel, it acts as a band attention mechanism, highlighting the bands that more effectively differentiate various pixels. Figure 5.2 (a) illustrates this process.

By employing Conv2D layers with N such 1×1 filters and the ReLU activation function, IA achieves multi-head band attention and extracts N spectral patterns concurrently. The resulting output feature map of the Conv2D layer retains the same dimensions as the original spectral bands in the input imagery and can thus be considered as pseudo bands. During backpropagation, the pixel values in the extracted spectral patterns are adjusted based on their contributions to the class labels of the input imagery. For convenience, these extracted spectral patterns are also referred to as deep-pseudo bands hereinafter.

Two such 1×1 Conv2D layers are stacked to extract more fine-grained spectral patterns, which are then concatenated with the original spectral bands. This design ensures that the spatial and spectral information in the original imagery is retained in subsequent steps to guide the further refinement of the extracted spectral patterns. The intermediate output of the first step is referred to as raw spectral patterns, as they will be refined in later steps. According to Definition 2.2.1, each of the original spectral bands carried into the intermediate output in this step can be considered a special type of spectral pattern generated with a specific mapping: $P_{(i,j)} \mapsto p_{(i,j)}^k$, where $k \in [1, C] \cap \mathbb{N}$ refers to the k -th spectral band in the input imagery.

For ease of implementation, the number of output channels of IA is set to 32. Accordingly, its Conv2D layers use $N = 32 - n$ 1×1 filters, where n is the number of spectral bands in the input imagery. The choice of 32 was informed by ablation study results, considering computational complexity.

In the second step, the IA module incorporates a Spatial Attention module to refine the raw spectral patterns generated in the first step. As illustrated in Figure 5.1 (b), during backpropagation, the Spatial Attention module learns an attention distribution for each channel of the input feature map in the form of a $W \times H$ matrix, where $W \in \mathbb{N}$ and $H \in \mathbb{N}$ are the width and height of the input imagery. These attention distributions represent the importance levels of individual pixels.

Guided by supervised learning, the importance of each pixel is typically determined by its association with the target classes. Consequently, within the 32 refined spectral patterns—each associated with one spatial attention distribution—some may emphasise pixels belonging to one class, while others may highlight pixels associated with different classes. For example, certain spectral patterns may enhance smoke pixels

while suppressing those of other classes, or vice versa.

Figure 5.2 (b) illustrates this process. Notably, the intermediate output after this step no longer contains the original spectral bands as they are also weighted by specific spatial attention distributions.

In the third step, the IA module applies a Channel Attention module to weigh the importance of the extracted spectral patterns before they are used as input to a DL model. As depicted in Figure 5.1 (c) and Figure 5.2 (c), this module assigns higher weights to the spectral patterns that have greater representative power for the target classes. Unlike the band attention that captures the most representative bands for a pixel in the first step (Figure 5.2 (a)), the channel attention here weights the extracted features based on their significance in the prediction.

The Channel Attention module, together with the Spation Attention module in step two, guides the model to effectively learn class-oriented spectral patterns.

The implementation of spatial attention and channel attention is common and typified by [43] and [27] in smoke detection. In the work of this chapter, spatial attention and channel attention are applied from a novel perspective so that, together with 1×1 filters, they produce refined pixel-level spectral patterns while preserving spatial features and emphasising class-oriented information. In contrast, previous applications of spatial and channel attention were primarily focused on enhancing the extraction of spatial patterns.

The IA module learns class-oriented spectral patterns automatically, without requiring human expertise or intervention. This contrasts with traditional methods, such as spectral indices and threshold values, which rely on predefined formulas based on experience and domain knowledge. Additionally, the IA-extracted spectral patterns are used as new input to DL models when integrated with them. This process automates the link between the extracted spectral patterns and the scene-level classification, streamlining the learning pipeline for a more effective exploration of both spectral and spatial patterns.

5.3 Experimental Settings

This section introduces the datasets, training settings, and evaluation metrics used in this study.

5.3.1 Datasets

To evaluate the effectiveness of the IA module, this study employs both the USTC_SmokeRS dataset [43] and the Landsat6c dataset, with the latter proposed in Chapter 4 of this thesis.

The former, constructed using MODIS RGB images, spans nearly twenty years and covers six continents, excluding Antarctica [43]. The latter contains three additional IR bands, namely NIR, SWIR_1, and SWIR_2, covering historical fire events between 2010 and 2020 in SA and NSW, Australia. The wide range of global locations and temporal conditions ensures that IA’s effectiveness is evaluated across varying locations and times.

Table 5.1 presents a summary of the two datasets. For more detailed information about each dataset, refer to [43] and Section 4.2.2 of the thesis, respectively.

Table 5.1: A summary of datasets used in this paper

Dataset	Classes	Number of Images	
		Per Class	Total
USTC_SmokeRS	Cloud	1164	6225
	Dust	1009	
	Haze	1002	
	Land	1027	
	Seaside	1007	
	Smoke	1016	
Landsat6c	Clear	616	1836
	Other_aerosol	605	
	Smoke	615	

5.3.2 Training Settings

Four baseline models were selected, namely ResNet50 [44], InceptionResNetV2 [297], MobileNetV2 [209], and VIB_SD [27]. VIB_SD is a lightweight model specifically designed for smoke detection, as described in Chapter 4, while the other models are widely recognised in the literature. These baseline models represent various CNN architectures with differing depths and parameter counts. The objective is to verify whether integrating IA can effectively improve their prediction performance, irrespective of structural differences.

All baseline models were initially trained using the two datasets separately. Next, the models were integrated with the IA module, and the training process was repeated under the same settings. The integration of IA is straightforward: it is inserted directly after the input layer of the baseline models. The 32-channel feature map, produced by IA and having the same width and height as the original input imagery, is then fed to the baseline models as their new input.

For both datasets, 64% of the data was used for training, 16% for validation, and 20% for testing, consistent with the settings in Chapter 4. The test results were used for comparison. To minimise the risk of overfitting, all input imagery in the training data was augmented with random horizontal and vertical flipping. The augmentation was kept simple to avoid introducing noise.

All models were trained using an input size of 256×256 for the width and height dimensions, which corresponds to the original size of the imagery files. It is important to note that ResNet50 and InceptionResNetV2 have default input sizes of 224×224 and 299×299 , respectively. Using these default sizes would require resizing the input imagery, which could introduce interpolated pixel values across all input bands. This may cause the learned spectral patterns to deviate significantly from the authentic spectral patterns. To avoid this issue, the input size for ResNet50 and InceptionResNetV2 was adjusted to 256×256 . This change allowed the models to learn effectively without introducing interpolated pixels and did not affect the comparison.

The batch size was set to 32, and the number of epochs was set to 300. To avoid redundant training, early stopping was applied if the validation accuracy did not improve within 60 epochs. The initial learning rate was set to 0.01 and reduced by a factor of

0.8 if the validation loss did not decrease within 20 epochs. The Adam optimiser [300] was used for optimisation.

All algorithms were implemented using TensorFlow and trained under the Ubuntu 16.4 operating system. A global random seed was used, along with necessary local seeds for dataset preparation and model training, utilising the mirror strategy with two NVIDIA GeForce GTX 1080 Ti GPUs. This ensured the models were more comparable, considering the many random processes involved during training, such as random parameter initialisation, dataset splitting and shuffling, and job assignment.

The results of each baseline model were compared with those of the same model incorporating the IA module.

5.3.3 Evaluation Metrics

This study includes both accuracy (%) and the Kappa in the evaluation metrics, consistent with the settings in Chapter 4 of this thesis and the studies in [43, 47]. Additionally, the FNR of the target class “Smoke” is introduced as a new evaluation metric in this study, due to its importance in natural disaster detection.

The formulas for calculating accuracy and the Kappa were provided in Table 4.5. Using the same notations, the class-wise FNR is defined by the following formula:

$$FNR = \frac{N_{i-}}{N_{ii} + N_{i-}} \quad (5.3.1)$$

5.4 Experimental Results

This section presents and compares the test results of all baseline models, with and without the IA module, in Section 5.4.1. Following this, Section 5.4.2 visualises and analyses some of the spectral patterns in the deep-pseudo bands extracted by the IA module. Finally, Section 5.4.3 showcases the results of the ablation studies.

5.4.1 Model Performance with and without IA

The test results of the baseline models, both with and without the IA module, using the USTC_SmokeRS dataset and the Landsat6c dataset are shown in Table 5.2 and Table 5.3, respectively.

Table 5.2: Results of using the USTC_SmokeRS dataset

Model	IA	Parameter Count	Accuracy	Kappa	FNR
ResNet50	No	23.600M	86.43%	83.71%	21.60%
	Yes	23.693M	88.67%	86.41%	18.31%
InceptionResnetV2	No	54.35M	88.27%	85.92%	16.43%
	Yes	54.36M	90.92%	89.10%	12.21%
MobileNetV2	No	2.266M	89.88%	87.86%	18.31%
	Yes	2.276M	84.18%	81.04%	33.33%
VIB_SD	No	1.745M	92.85%	91.42%	15.50%
	Yes	1.897M	94.14%	92.96%	13.15%

Table 5.3: Results of using the Landsat6c dataset

Model	IA	Parameter Count	Accuracy	Kappa	FNR
ResNet50	No	23.603M	75.82%	63.68%	26.47%
	Yes	23.686M	80.43%	70.70%	29.41%
InceptionResnetV2	No	54.34M	83.97%	75.89%	24.77%
	Yes	54.35M	85.05%	77.52%	18.38%
MobileNetV2	No	2.263M	76.90%	65.11%	22.06%
	Yes	2.272M	78.80%	68.01%	21.32%
VIB_SD	No	1.676M	81.79%	72.61%	24.26%
	Yes	1.812M	85.33%	77.87%	13.97%

The results indicate that adding the IA module only slightly increased the parameter counts compared to the original models.

When training with the USTC_SmokeRS dataset, the IA module effectively improved all three evaluation metrics for ResNet50, InceptionResNetV2, and VIB_SD. However, the original MobileNetV2 achieved better results across all three metrics.

When training with the Landsat6c dataset, all four baseline models showed significant improvements in terms of accuracy and the Kappa. Additionally, except for ResNet50, the other three models also demonstrated improvements in the FNR for the class “Smoke”.

The results demonstrate that the IA module effectively enhances CNN-based smoke detection from satellite imagery. The reduced performance of MobileNetV2 with IA, observed when trained on the USTC_SmokeRS dataset, may be related to the extensive use of 1×1 filters in its depth-wise separable convolutions and inverted residual blocks. This aspect will be further discussed in Section 5.5.

5.4.2 Visualisation of IA-extracted Spectral Patterns

To better understand the spectral patterns extracted by the IA module, some of the class-oriented spectral patterns extracted by the VIB_SD model from imagery samples in both the USTC_SmokeRS and Landsat6c datasets were visualised.

In Figure 5.3, the five imagery samples in the leftmost column are labelled as “Cloud,” “Dust,” “Haze,” “Seaside,” and “Smoke” from the USTC_SmokeRS dataset. The grey-scale images in the two columns on the right are two visualised corresponding deep-pseudo bands produced by the IA module. It is evident that the spectral patterns exhibit class-related attributes. The pixels representing the target class and those from other classes, are highlighted respectively in each deep-pseudo band. Notably, Figure 5.3 shows that the IA module can accurately identify smoke pixels even in the absence of pixel-level ground truth in the training data.

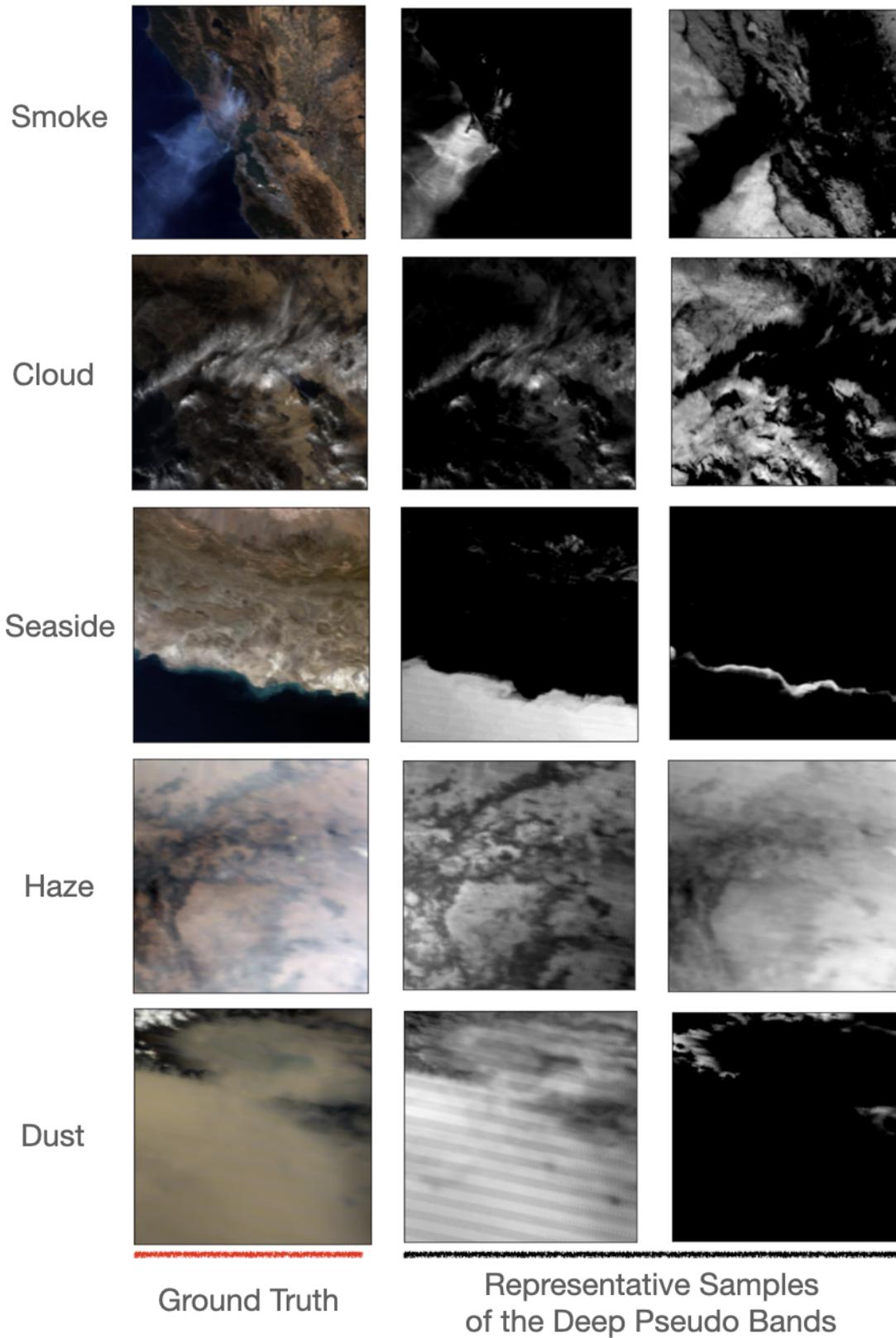


Figure 5.3: Original MODIS imagery samples (left) vs. two samples of deep-pseudo bands extracted by IA

In Figure 5.4, on the left are three ground truth imagery samples labelled as “Smoke,” “Other_aerosol,” and “Clears” from the Landsat6c dataset, visualised with the RGB bands; on the right is the visualisation of two corresponding spectral patterns containing class-oriented information. It can be observed that the spectral patterns capture class-oriented pixels very well for both smoke and other aerosols.

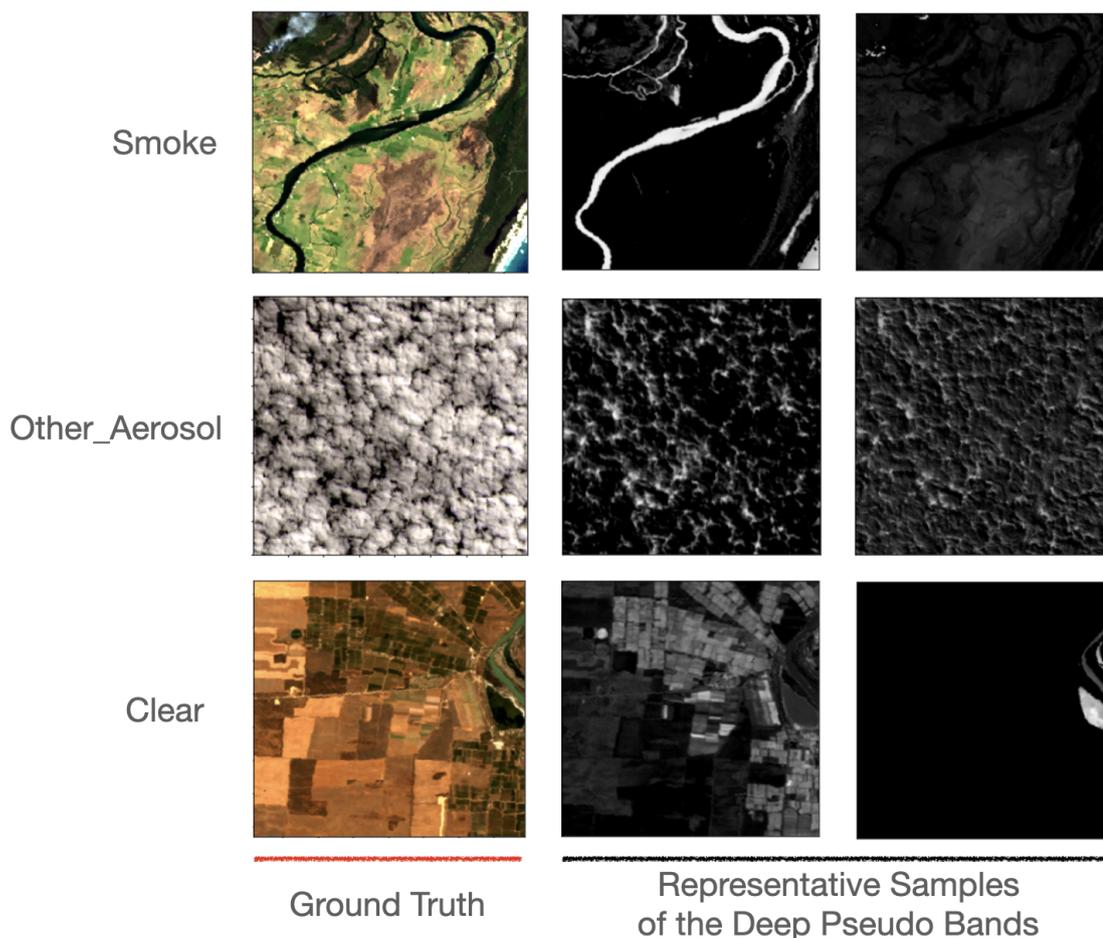


Figure 5.4: Original Landsat imagery samples (left) vs. two samples of deep-pseudo bands extracted by IA

Figures 5.3 and 5.4 also demonstrate that the IA module can enhance the model’s explainability. The visualisation of the deep-pseudo bands clearly illustrates the spectral patterns learned and utilised by the model in making predictions, providing tangible insight into the decision-making process.

5.4.3 Ablation Studies and Parameter Selection

Ablation studies were conducted to determine the optimal way to integrate the attention mechanism in the IA module, how many 1×1 Conv2D layers the IA module should include, and the number of output channels for the IA module.

The VIB_SD model was used exclusively for the ablation studies. For the studies on the attention mechanism and the number of 1×1 convolution layers, only the USTC_SmokeRS dataset was utilised. Both datasets were used for the study on the number of output channels of the IA module.

The results regarding the attention mechanism are shown in Table 5.4, which suggest that the VIB_SD model employing the IA module with both spatial and channel attention achieved the highest accuracy.

Table 5.4: Ablation study about attention mechanism

Attention Modules	Accuracy	Kappa	FNR
None	92.85%	91.42%	15.50%
Channel Attention	91.57%	89.88%	16.43%
Spatial Attention	92.29%	90.74%	18.31%
Both	94.14%	92.96%	13.15%

The results regarding the number of 1×1 Conv2D layers ($\#1 \times 1$ Conv2D layers) are shown in Table 5.5, which suggests that using two 1×1 Conv2D layers achieved the highest accuracy.

Table 5.5: Ablation study about the number of 1×1 Conv2D Layers

$\#1 \times 1$ Conv2D Layers	Accuracy	Kappa	FNR
1	92.37%	90.84%	16.43%
2	94.14%	92.96%	13.15%
3	93.98%	92.77%	11.27%
4	91.89%	90.26%	18.31%

The results regarding the number of output channels ($\#$ Output Channels) of the IA module are shown in Table 5.6. Based on the ablation study results from both datasets and considering computational complexity, 32 output channels were adopted in the proposed IA module.

Table 5.6: Ablation study about the output channel number of the IA module

Dataset	$\#$Output Channels	Accuracy	Kappa	FNR
USTC_SmokeRS	16	92.37%	90.84%	11.74%
	24	93.33%	92.00%	15.02%
	32	94.14%	92.96%	13.15%
	40	94.14%	92.96%	10.33%
	48	93.33%	92.00%	12.21%
Landsat6c	16	80.43%	70.53%	22.06%
	24	82.88%	74.21%	16.91%
	32	85.33%	77.87%	13.97%
	40	82.07%	73.08%	18.38%
	48	80.43%	70.41%	16.91%

5.5 Discussion

Smoke shares some similar spatial patterns with clouds, haze, fog, and other aerosols. This makes it challenging to discern smoke from these aerosols using DL models that primarily rely on spatial patterns. However, the particles in different aerosols exhibit distinct reflection characteristics across various bands, making spectral patterns useful for smoke detection. This is believed to be the primary reason why the IA module improves the accuracy of CNN models for smoke detection in satellite imagery.

The performance of integrating the IA module with various CNN architectures appears to be influenced by their specific designs. Notably, no performance improvement was observed when MobileNetV2 was trained with the IA module using the USTC_SmokeRS dataset. This outcome might be attributed to factors related to both

the USTC_SmokeRS dataset and the MobileNetV2 model. Firstly, the USTC_SmokeRS dataset, comprising only visible bands, may limit the representativeness of the spectral patterns learned, in contrast to the Landsat6c dataset, which includes additional spectral bands. Consequently, the spectral patterns from USTC_SmokeRS may be more susceptible to distortion by the CNN architecture. Secondly, MobileNetV2’s extensive use of depth-wise separable convolutions and inverted residual blocks, which feature 1×1 Conv2D layers, might alter the IA-learned spectral patterns due to the presence of 1×1 Conv2D layers. This alteration could lead to compromised performance. Readers are referred to [209] for a comprehensive understanding of the MobileNetV2 architecture. Further research is needed to identify the exact causes of this performance issue and to investigate possible solutions.

Since the IA module serves as an input pre-processing block, it is easily applicable to VITs as well as CNNs. Nonetheless, its efficacy on VITs was not evaluated in this study due to the difficulty in identifying, implementing, and training appropriate benchmark VITs within the constraints of this research project. Additional research is planned to explore this avenue in the future.

The IA module is not limited to smoke detection from satellite imagery and can be applied to a broader range of classification tasks, including those using non-satellite imagery. It would be particularly useful for tasks where the reflection characteristics of a class in different bands are distinct from those of other classes. For example, detecting water pollution, vegetation diseases, or diagnosing human diseases such as polyps or skin cancer may benefit from the IA module’s ability to extract class-oriented spectral patterns. Future work could investigate the effectiveness of the IA module for such tasks, as this is beyond the scope of this thesis.

In addition, it was observed that some of the deep-pseudo bands extracted by the IA module can successfully mark pixels belonging to certain target classes, even when no binary ground truths at the pixel level were provided in the training data. This suggests the potential of using the IA module for tasks involving pixel-level labelling or segmentation. Besides, since the deep-pseudo bands extracted by the IA module can be easily visualised, they can enhance the interpretability of DL models.

Furthermore, the IA module can be leveraged to facilitate transfer learning for training a smoke detection CNN model using imagery from multiple satellite sensors or updating a trained model with a few labelled images from a new satellite sensor. For

instance, one can first train a CNN model on a dataset with fewer spectral bands, and then add the IA module in front of the trained model, setting the input channels to match the number of bands in the new dataset and the output channels to the number of bands in the original dataset. This approach allows the new model to be fine-tuned through transfer learning using only a small number of images from the new dataset.

5.6 Conclusion

In conclusion, this study demonstrates the limitations of current DL models in exploring pixel-level spectral patterns that are critical for smoke detection in satellite imagery. To address this, a novel DL module called IA was proposed, enabling DL models to extract class-oriented pixel-level spectral patterns alongside spatial patterns. The IA module incorporates 1×1 filters and attention mechanisms and can be seamlessly integrated with existing DL models with minimal computational overhead.

The IA module was evaluated on two smoke satellite imagery datasets: USTC_SmokeRS, which includes only the RGB bands, and Landsat6c, which comprises six spectral bands (i.e., RGB, NIR, SWIR_1, and SWIR_2). The experimental results demonstrate that integrating the IA module with an existing CNN model effectively improves the model’s prediction accuracy.

The deep-pseudo bands extracted by the IA module were visualised, demonstrating that these bands successfully segmented pixels belonging to specific target classes. This suggests the potential of using the IA module for pixel-level labelling or segmentation tasks.

Moreover, the IA module learns to extract class-oriented pixel-level spectral patterns during the learning process based on the classification tasks, indicating its potential to be applied to a broader range of classification tasks in various domains. Overall, the IA module is a promising approach to improving the accuracy and interpretability of DL models, particularly in multispectral satellite imagery analysis.

Chapter 6

Cross-sensor Transfer Learning for Smoke Detection

Satellite-based smoke detection (referring to scene-level smoke detection in this chapter) for fire disaster prevention requires accurate and timely detection of EF smoke plumes, which often cover small geographic extents.

Chapter 5 proposed the IA module and integrated it with the VIB_SD model from Chapter 4, resulting in a more accurate smoke detection model, IA_VIB_SD. Additionally, Chapter 4 introduced the multispectral satellite imagery dataset Landsat6c, derived from Landsat 5 TM and Landsat 8 OLI imagery. Featuring extra IR bands, Landsat6c facilitates the IA module in learning class-oriented spectral patterns, thereby improving accuracy more effectively. Moreover, its 30-metre spatial resolution enables the capture of EF smoke plumes, which further ensures the capability of the trained IA_VIB_SD to detect such EF smoke effectively.

Nonetheless, the lengthy temporal resolution of 16 days for Landsat 5 TM and Landsat 8 OLI can hardly satisfy the demand for timely detection of EF smoke. Combining multiple satellites, such as the Landsat series providing medium to high spatial resolutions, could improve the detection of EF smoke while collectively delivering better temporal resolution. However, no investigation has yet been conducted in the literature on how to quickly equip these satellites with accurate DL models, especially given the lack of labelled training datasets, particularly for new satellites that have not accumulated sufficient observational data. Furthermore, different satellite sensors typically have varying spectral bands with distinct spectral and radiometric characteristics, which also pose significant challenges.

This chapter investigates feasible transfer learning strategies for rapid model development in smoke detection using multiple satellite sensors, addressing the third research question posed in Chapter 1:

RQ-3 How to leverage transfer learning for fast model development across multiple satellites for timely smoke detection, given insufficient labelled training data from some satellites?

Consequently, this chapter proposes a novel cross-sensor transfer learning approach through the learning and adaptation of spectral patterns, facilitated by the IA_VIB_SD model. The IA_VIB_SD model first learns to extract class-oriented spectral patterns for smoke detection from the multispectral training data in the source domain. When the model is transferred to the target domain, its ability to learn and utilise these

spectral patterns is carried over. This capability is then fine-tuned and updated for the new data in the target domain using a small amount of labelled training samples. As a result, the new IA_VIB_SD model achieves enhanced smoke detection accuracy in the target domain.

The adaptation of the pretrained IA_VIB_SD to the target domain data, which features a different number of spectral bands, is achieved by replacing the IA module for the target domain. It is important to note that, the VIB_SD component of the pretrained model is retained to carry over the knowledge learned from the source domain. This contrasts with the conventional transfer learning approach, which incorporates an additional mapping layer to align the input dimensions between the source and target domains, allowing the same model to be used in both domains.

The study employs the Landsat6c dataset as the source domain. Notably, Landsat6c is expanded from 1836 to 2770 imagery files in this study, ensuring that each class contains more than 900 training samples to enhance the robustness of the pretrained IA_VIB_SD. Additionally, this chapter introduces Sentinel7c, a new multispectral imagery smoke detection training dataset created from Sentinel-2 MSI imagery, to serve as the target domain. Sentinel7c consists of 351 imagery files with seven spectral bands and a 10-metre spatial resolution, compared to the six bands and 30-metre spatial resolution of the source domain dataset, Landsat6c.

The study compared the performance of the proposed transfer learning approach against the conventional approach (as illustrated in Figure 3.2 and demonstrated by the work in [288]). In addition, the performance of the transferred IA_VIB_SD models was compared to the IA_VIB_SD model trained purely on Sentinel7c data in the target domain. The results demonstrate that the proposed approach outperformed the conventional approach, and the transferred models surpassed the model trained purely on the target domain, even when trained on a small portion of the Sentinel7c dataset.

This chapter is based on my paper *“Cross-sensor transfer learning for smoke detection using variable-bands multi-spectral satellite imagery aided by spectral patterns”* [57], published in the International Journal of Remote Sensing in April 2024.

6.1 Introduction

Extreme wildfires, exacerbated by climate change, have occurred frequently in recent years and caused tremendous losses worldwide. There is an urgent need for feasible and effective solutions for NRT fire detection to mitigate such disasters. Smoke detection using advanced DL models onboard satellites presents great potential due to the proven accuracy of DL models and the reduced latency of onboard detection. Employing multiple satellite platforms collaboratively can further increase the likelihood of promptly identifying the initiation of fires.

However, using multiple satellites for smoke detection typically necessitates customised models for different satellites, which could significantly delay real-world applications.

On the one hand, models trained exclusively on imagery data from one specific satellite often exhibit suboptimal performance when directly applied to data from another satellite, even if the imagery from both satellites contains identical spectral bands. For instance, Table 6.1 presents the results of directly applying the Landsat6c-trained IA_VIB_SD model to make predictions on Sentinel7c imagery using the same set of six corresponding bands. The results reveal that the model’s predictive accuracy on Sentinel7c imagery was notably close to random guessing. More information about Landsat6c, IA_VIB_SD, and Sentinel7c can be found in Section 4.2 of Chapter 4, Section 5.2 of Chapter 5, and Section 6.2.1 of this chapter.

Table 6.1: Accuracy of Landsat6c-trained IA_VIB_SD predicting Sentinel7c imagery

Model	Training data	Accuracy	
		Landsat6c	Sentinel7c
IA_VIB_SD	Landsat6c	85.08%	54.13%

On the other hand, when a new satellite sensor is involved in onboard smoke detection, it must accumulate adequate observational data before a reliable training dataset with abundant samples can be created. This process can take a long time, with additional time needed for model training and deployment.

Several studies have investigated techniques to enhance DL model robustness for remote sensing classification tasks involving multiple imagery sources, though not fo-

cusing on smoke detection. For instance, [301] introduced the Multiscale Interactive Fusion Network, which incorporates a Multiscale Interactive Information Extraction block and a Global Dependence Fusion Module for classifying multi-source remote sensing data. Subsequently, [302] proposed the Representation-enhanced Status Replay Network, which addresses representation and classifier bias, as well as feature fusion imbalances, in multi-source remote sensing image classification.

However, these approaches require that the imagery from different sensors have identical or similar spectral bands with the same band dimensions. Additionally, such methods may compromise accuracy on individual sensors in favour of achieving overall accuracy across all source sensors.

To expedite real-world applications of using multiple satellites for NRT smoke detection, there is a pressing need to investigate effective approaches that enable rapid model development for multiple satellites, including newly launched ones, without compromising accuracy. Additionally, such approaches must account for variations in both the spectral bands and the number of bands in imagery from different satellites. Leveraging transfer learning emerges as a promising approach to achieve these aims.

Prior research on transfer learning in remote sensing has primarily focused on adapting models trained on RGB images, which does not align well with satellite imagery containing multiple spectral bands for detecting smoke and aerosols. To better preserve relevant features, this study focuses on exploring cross-sensor transfer learning for smoke detection using multispectral imagery from both the source and target sensors.

This introduces more complex challenges due to the involvement of multiple spectral bands, which can vary significantly between different sensors. For example, some spectral bands used in the training data from one sensor may be absent in data from another sensor, and the bands present across different sensors may differ in terms of spectral, spatial, and radiometric resolution. Additionally, unlike the abundance of large, readily available RGB image training datasets, labelled multispectral satellite imagery datasets are scarce in the field of remote sensing [280, 289, 303]. Finding a labelled multispectral satellite imagery dataset for smoke detection is exceptionally challenging, especially when considering multiple sensors.

This chapter presents an innovative cross-sensor transfer learning approach to address the above challenges, involving the creation of the Sentinel7c dataset as the

multispectral target domain to support the study.

The innovation of the proposed transfer learning approach is closely tied to the employment of the IA_VIB_SD model, specifically the IA module within it, for this investigation. This is due to the model’s ability to extract class-oriented spectral patterns with the assistance of the IA module. It is assumed that such spectral patterns, extracted from multispectral imagery from different sensors, are less variable compared to the spectral characteristics of the bands between sensors. Provided this assumption holds, adapting the spectral patterns learned by the IA module from both the source and target domains becomes a better alternative than addressing the data distribution discrepancies between the two domains. Furthermore, the lightweight design of IA_VIB_SD holds great potential for onboard SmallSats smoke detection, making the cross-sensor transfer learning study based on IA_VIB_SD even more valuable.

As mentioned, two labelled multispectral satellite imagery datasets, namely Landsat6c and Sentinel7c, are used as the source and target domains, respectively. Both datasets are labelled into three classes: “Clean,” “Other_aerosols,” and “Smoke.” The former was introduced in Chapter 4, originally comprising 1836 $6 \times 256 \times 256$ Landsat images, but has been expanded to 2770 images to provide more training samples in each class for this study. The latter, newly constructed to support this study, comprises 351 images sourced from Sentinel-2, each with dimensions of $7 \times 256 \times 256$. Sentinel7c possesses an additional spectral band compared to Landsat6c and maintains a 10-metre spatial resolution across all bands, distinct from the 30-metre spatial resolution of Landsat6c. Further details about the two datasets are available in Section 6.2.1.

The transfer learning approaches were evaluated based on IA_VIB_SD, using the Landsat6c and Sentinel7c datasets. The results reveal that all transferred models outperformed IA_VIB_SD trained solely on the Sentinel7c dataset when using 10% to 50% of the dataset for training. Notably, the highest accuracy was achieved with the transfer learning approach that replaces the IA module in the pretrained IA_VIB_SD and updates the model with the new imagery data.

In summary, the study in this chapter makes a two-fold contribution:

1. Proposed a novel cross-sensor transfer learning approach that accommodates variable spectral bands in multispectral imagery from different sensors by learning and adapting class-oriented spectral patterns. This approach presents an innovative strategy for few-shot learning in remote sensing and offers a promising

solution for smoke detection using multi-sensor multispectral satellite imagery, particularly for fire disaster prevention.

2. Introduced a labelled multispectral satellite imagery training dataset, Sentinel7c. This dataset, together with Landsat6c, holds significant potential for advancing research in DL-based smoke detection and cross-sensor transfer learning.

In contrast to conventional transfer learning approaches, as demonstrated in Figure 3.2, this study focuses on transferring a DL model pretrained on multispectral imagery from one sensor to a new sensor with potentially different spectral bands, as illustrated in Figure 6.1. The proposed transfer learning approach does not involve dimensionality reduction (in terms of channel count) in the new satellite imagery, which could result in information loss. Instead, dimensionality increment is leveraged by extracting class-oriented spectral patterns with the assistance of the IA module. Such a transfer learning strategy has not yet been explored in the existing literature.

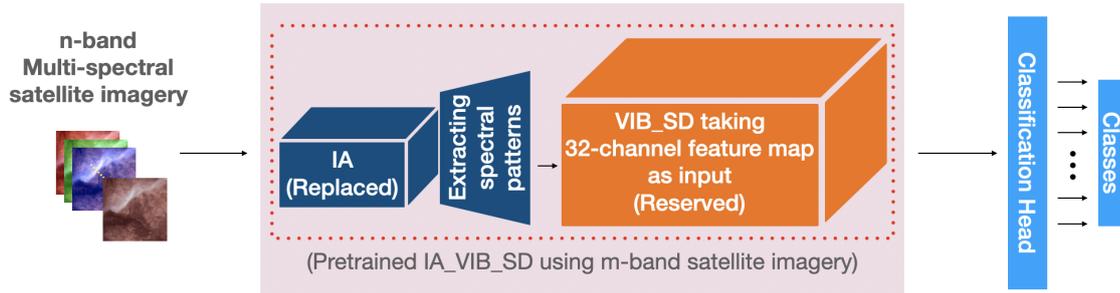


Figure 6.1: The proposed transfer learning approach increases the input dimension by utilising IA. Notably, although IA was originally designed to extract spectral patterns to improve VIB_SD performance, it also serves as a mechanism for mapping diverse input bands to a consistent number of spectral patterns, which are used as input to VIB_SD.

Furthermore, compared to the cross-sensor transfer learning approaches proposed in [283] and [285], this approach is more straightforward. It utilises a single DL model structure and does not involve defining subdomains to bridge the source and target domains, which would require additional cumbersome data pre-processing.

The remainder of this chapter is organised as follows:

- Section 6.2 introduces the materials and methods employed in this study.

- Section 6.3 details the experimental settings and evaluation metrics.
- The experimental results are discussed in Section 6.4.
- Finally, Section 6.5 presents the conclusions drawn from this study.

6.2 Materials and Methods

This section presents additional information about the two datasets used in this investigation, along with the transfer learning approaches examined in this study and the evaluation metrics. The architecture of the VIB_SD model and the construction of Landsat6c were detailed in Chapter 4. The design of the IA module, as well as the IA_VIB_SD model resulting from its integration with VIB_SD, was described in Chapter 5.

6.2.1 Datasets

As previously mentioned, the source domain dataset Landsat6c comprises six spectral bands: RGB, NIR, SWIR_1, and SWIR_2. In contrast, the target domain dataset Sentinel7c, newly created from Sentinel-2 MSI imagery, comprises seven bands, namely RGB, NIR, NIR_2, SWIR_2, and SWIR_3. The construction of Sentinel7c follows the same procedure as that of Landsat6c, as described in Chapter 4.

Aside from the difference in the number of bands, Landsat6c and Sentinel7c differ across other properties associated with the characteristics of their sensors. Table 6.2 provides detailed specifications for Landsat 5 TM, Landsat 8 OLI, and Sentinel-2 MSI, demonstrating the disparities in their bands, bandwidths, and spatial resolutions. These differences highlight the challenges that cross-sensor transfer learning must address when using multispectral satellite imagery from multiple sensors for smoke detection.

The imagery files in Landsat6c and Sentinel7c both have dimensions of 256×256 pixels in terms of width and height. Table 6.3 shows the class names and the respective number of images in each class for the two datasets.

Table 6.2: Specifications of selected bands from Landsat 5 TM, Landsat 8 OLI, and Sentinel-2 MSI used in the Landsat6c and Sentinel7c

Band	Wavelength (μm)			Spatial Resolution (m)		
	TM	OLI	MSI	TM	OLI	MSI
Red	0.63–0.69	0.636–0.673	0.65–0.68	30	30	10
Green	0.52–0.60	0.53–0.59	0.543–0.578	30	30	10
Blue	0.45–0.52	0.45–0.51	0.458–0.523	30	30	10
NIR	0.76–0.90	0.851–0.879	0.785–0.899	30	30	10
NIR_2	–	–	0.855–0.875	–	–	20 (Resampled to 10)
SWIR_1	1.55–1.75	1.566–1.651	–	30	30	–
SWIR_2	2.08–2.35	2.107–2.294	1.565–1.655	30	30	20 (Resampled to 10)
SWIR_3	–	–	2.1–2.28	–	–	20 (Resampled to 10)

Table 6.3: The class distributions of Landsat6c and Sentinel7c

Dataset	Number of Bands	Classes	Number of Images	
			Per class	Total
Landsat6c	6	Clear	944	2770
		Other_aerosol	916	
		Smoke	910	
Sentinel7c	7	Clear	112	351
		Other_aerosol	116	
		Smoke	123	

6.2.2 Methods

The study examines the performance of models derived from various transfer learning approaches against a benchmark model trained exclusively on the target domain dataset. The transfer learning approaches investigated in this chapter are based on the IA_VIB_SD model, which is first trained on the source domain dataset and then updated on the target domain dataset using the pretrained weights. This type of transfer learning is commonly referred to as homogeneous-network transfer learning, in contrast to heterogeneous-network transfer learning, which employs a different model for the target domain. The latter has been demonstrated in the work of [283] and [285], both reviewed in Chapter 3.

Notably, homogeneous-network transfer learning often requires modifying the input layer of the DL model when the imagery data differs in dimensions (e.g., the number of spectral bands) between the source and target domains. Similarly, the classification head must be replaced if the number of classes varies across domains. Despite these structural adjustments, the model transferred to the target domain is still considered a version of the source domain model.

Figure 6.2 provides a schematic representation of the model training and transfer learning methods used in this study. First, the IA_VIB_SD model is trained separately on Landsat6c (source domain) and Sentinel7c (target domain), as shown in Figure 6.2 (a). The latter serves as the benchmark model to evaluate the performance of the transferred models. Subsequently, the Landsat6c-pretrained IA_VIB_SD model is transferred to the target domain using four different transfer learning methods: two methods using the conventional approach, depicted in Figure 6.2 (b), and two methods employing the proposed approach, shown in Figure 6.2 (c).

In the context of the study in this chapter, the source domain dataset Landsat6c and the target domain dataset Sentinel7c share the same classes, eliminating the need to replace the classification head of IA_VIB_SD. However, since Sentinel7c contains seven spectral bands, one more than Landsat6c, this discrepancy must be addressed. Two options can be considered to handle this difference:

1. Replacing the IA module: This involves modifying the IA module to accept seven bands as input for the Sentinel7c dataset.

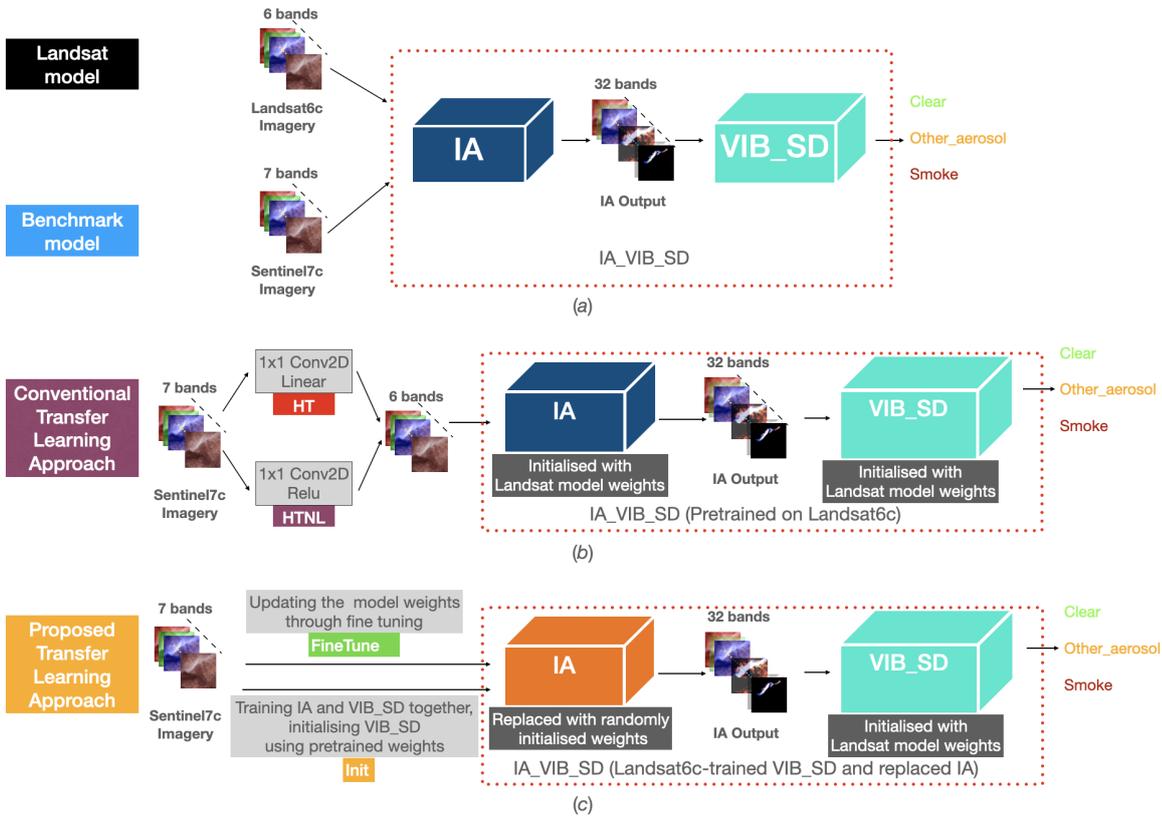


Figure 6.2: Schematic of model training and transfer learning approaches. (a) IA_VIB_SD trained purely on Landsat6c (as the pretrained model to be transferred) or Sentinel7c (as the benchmark model to evaluate the performance of the transferred models). (b) Conventional approach to transferring Landsat6c-trained IA_VIB_SD to Sentinel7c: “HT” and “HTNL” map seven input bands in Sentinel7c data to six bands, aligning with the Landsat6c-pretrained model, using linear and non-linear mapping, respectively. (c) Proposed approach to transferring Landsat6c-trained IA_VIB_SD to Sentinel7c: “FineTune” involves replacing the IA Module in the pretrained IA_VIB_SD, freezing the VIB_SD, and updating its weights through fine-tuning; “Init” entails training IA and VIB_SD together with random IA weights and pretrained VIB_SD weights.

2. Adding a mapping layer: This approach adds a mapping layer before IA_VIB_SD, mapping the seven-band input imagery to a six-band feature map, which can then be fed into IA_VIB_SD.

The conventional approach in the second option, serving as the transfer learning benchmark, can be implemented through a linear mapping, similar to the method in [288], using a 1×1 convolutional layer without an activation function, referred to as the “HT” method. It can also be implemented using a non-linear mapping by introducing an activation function in the 1×1 convolutional layer, referred to as the “HTNL” method. Notably, both methods in this approach utilise all spectral bands from both

the source and target domains, which is advantageous compared to methods requiring band selection or dimensionality reduction techniques that may result in significant information loss.

The proposed approach, which is the first option, offers several advantages:

- The IA module can learn to extract class-oriented spectral patterns more effectively by leveraging all seven bands.
- The primary feature extraction base, the VIB_SD model, which processes a feature map consisting of 32 spectral patterns, remains unchanged. This ensures a smoother transition in transfer learning compared to scenarios where the major feature extraction base processes the original bands directly.

During the transfer learning process, both the new IA module or the mapping layer can initially be trained separately while VIB_SD is frozen. Subsequently, VIB_SD is unfrozen, and the entire model is fine-tuned with a low learning rate. This method, where the VIB_SD is fine-tuned while leveraging the new IA module, is referred to as “FineTune.”

Alternatively, the new IA module or the mapping layer can be trained together with the transferred VIB_SD model, initialised with pretrained parameters. This method, training the integrated model from scratch with the new IA module and pretrained VIB_SD, is labelled as “Init.”

It is important to highlight that the “HT” and “HTNL” methods modify the input imagery by mapping it to the same dimensionality as the imagery used to pretrain the models. Although the IA module aids “HT” and “HTNL” in learning useful spectral patterns, these patterns are not derived from the actual spectral bands. In contrast, the “FineTune” and “Init” approaches maintain the integrity of the original input imagery by learning genuine spectral patterns directly from the actual bands. Findings demonstrate that “FineTune” and “Init” deliver superior performance by preserving the original input imagery. Preliminary results indicate that “Init” outperforms “FineTune,” leading to the decision to train both “HT” and “HTNL” from scratch without fine-tuning.

In the following content, where the context is clear, “HT”, “HTNL”, “FineTune”, and “Init” also refer to the corresponding models resulting from the four transfer learning methods.

6.3 Experimental Settings and Evaluation Metrics

6.3.1 Training Environment

All algorithms were implemented in TensorFlow and trained on the Ubuntu 16.04 operating system. The training was conducted using two Nvidia GeForce 1080 GPUs, utilising the mirror strategy provided in the TensorFlow standard library. The Adam optimiser [300] was used for all models.

To ensure consistency in comparison, a combination of global and local random seeds were employed during training to control various random processes, including parameter initialisation, dataset splitting, shuffling, and the distribution of training jobs between the GPUs. This approach guarantees that data splits remain consistent across different models when using the same split percentage, enhancing comparability between the models.

6.3.2 Pre-training IA_VIB_SD on Landsat6c

For the Landsat6c dataset, 64% was allocated for training, 16% for validation, and the remaining 20% for testing. To enhance the training dataset, random horizontal and vertical flipping were applied as data augmentation techniques. A batch size of 32 was used, with a maximum of 300 training epochs. Early stopping was implemented to prevent redundant training, terminating the process if validation accuracy did not improve over 60 epochs. The initial learning rate was set to 0.01, with a reduction factor of 0.8 applied if the validation loss did not decrease within 20 epochs.

The model was trained using multiple seed combinations, and the weights yielding the highest testing accuracy of 85.08% were selected for further investigation into cross-sensor transfer learning using the Sentinel7c dataset.

6.3.3 Training the Benchmark and Transferred Models on Sentinel7c

The training task associated with the target domain involves training the benchmark IA_VIB_SD purely on Sentinel7c and updating the Landsat6c-pretrained IA_VIB_SD. This aims to verify the following:

1. Can the transferred models achieve higher accuracy than the benchmark model, considering the effects of differing spectral bands and the number of bands between the source and target domains?
2. Can the proposed transfer learning approach (which includes the “FineTune” and “Init” methods) outperform the conventional approach (which includes the “HT” and “HTNL” methods)?
3. Can the proposed approach maintain superior performance even in scenarios where labelled training data in the target domain are extremely limited?

To address these questions through concrete experimental evidence, all models were trained with 10%, 20%, and up to 50% of the Sentinel7c dataset, with their performance assessed using the remaining data for validation. Validation metrics were used for comparison. For each percentage of training data, the global seed was altered 10 times, and the overall model performance was compared across 10 sets of results for each training data percentage.

Due to the small size of the Sentinel7c dataset, a batch size of 8 was used. The maximum number of epochs was set to 300, and the early stopping strategy mentioned earlier was applied to avoid redundant training. For training the benchmark model and using the “Init”, “HT”, and “HTNL” methods, the initial learning rate was set to 0.01 and reduced by a factor of 0.8 if the validation loss did not decrease within 20 epochs. For the “FineTune” method, the initial learning rate was set to 0.0005 and similarly reduced by a factor of 0.8 if the validation loss did not decrease within 20 epochs.

6.3.4 Evaluation Metrics

The study in this chapter adopts the same evaluation metrics as in Chapter 5: accuracy (%), Kappa, and FNR for the target class, “Smoke.”

The evaluation of both the transferred models and the original model was based on their performance during validation, using 10%, 20%, and up to 50% of the Sentinel7c dataset for training. This corresponds to 90%, 80%, and down to 50% of the dataset, respectively, being utilised for validation.

6.4 Results and Discussion

In this section, the results of the transferred IA_VIB_SD obtained using different transfer learning methods are compared against the results of the benchmark IA_VIB_SD trained solely with the Sentinel7c dataset. The comparisons are based on the validation results using 10% up to 50% of the samples for training, with the rest allocated for validation.

Table 6.4 presents the parameter counts of all the transferred models in comparison to the benchmark IA_VIB_SD trained on the Sentinel7c dataset. It demonstrates that the proposed "FineTune" and "Init" methods maintain the same complexity as the original model, whereas "HT" and "HTNL" only introduce negligible additional parameters to the benchmark model.

Table 6.4: Parameter counts of the models when trained using Sentinel7c.

Model	Benchmark	FineTune	Init	HT	HTNL
Parameter Count	1,812,285	1,812,285	1,812,285	1,812,367	1,812,367

Table 6.5 presents results, including mean accuracies and Kappa, along with 95% confidence intervals. The best mean values for all metrics are highlighted in bold. Notably, the proposed approach, involving the replacement of the IA module, consistently achieved the best performance.

As indicated in Table 6.5, training a new IA_VIB_SD with only 10% of the Sentinel7c data significantly increased accuracy to 66.2% when used for predictions on the Sentinel7c dataset, surpassing the 54.13% accuracy obtained by the Landsat6c-pretrained IA_VIB_SD, as presented in Table 6.1. Particularly noteworthy is that the transferred model, utilising the proposed "FineTune" method, further improved the accuracy to 71.99%. Moreover, when trained with 50% of the Sentinel7c data, the

proposed “Init” method yielded an accuracy of 90.17%, which is 5.69% higher than the benchmark model trained exclusively on Sentinel7c data.

Table 6.5: Results of all transferred models versus the benchmark model trained with 10% - 50% Sentinel7c data

Data%	Model	Accuracy	Kappa	FNR
10%	Benchmark	66.20 ± 8.38%	49.36 ± 12.42%	39.62 ± 24.61%
	FineTune	71.99 ± 6.28%	57.99 ± 9.22%	30.95 ± 21.73%
	Init	71.87 ± 6.77%	57.83 ± 10.33%	33.10 ± 21.65%
	HT	69.97 ± 10.40%	54.99 ± 15.52%	36.63 ± 26.19%
	HTNL	70.82 ± 8.48%	56.33 ± 12.35%	36.36 ± 30.07%
20%	Benchmark	75.87 ± 6.63%	63.84 ± 9.77%	28.03 ± 16.69%
	FineTune	80.43 ± 4.34%	70.60 ± 6.47%	19.11 ± 14.09%
	Init	81.25 ± 5.36%	71.84 ± 8.10%	22.11 ± 11.42%
	HT	79.22 ± 4.66%	68.81 ± 6.96%	22.73 ± 18.25%
	HTNL	80.11 ± 5.85%	70.15 ± 8.81%	24.40 ± 12.36%
30%	Benchmark	80.49 ± 10.24%	70.72 ± 15.42%	26.24 ± 21.24%
	FineTune	82.69 ± 6.19%	74.00 ± 9.41%	18.29 ± 13.83%
	Init	85.55 ± 4.83%	78.32 ± 7.22%	19.65 ± 6.81%
	HT	82.04 ± 9.14%	73.02 ± 13.73%	21.00 ± 13.69%
	HTNL	83.76 ± 3.99%	75.61 ± 6.07%	20.29 ± 10.88%
40%	Benchmark	83.05 ± 3.94%	74.51 ± 5.95%	20.69 ± 12.10%
	FineTune	85.14 ± 3.54%	77.66 ± 5.34%	16.63 ± 14.61%
	Init	86.10 ± 3.83%	79.06 ± 5.87%	14.64 ± 9.89%
	HT	85.14 ± 6.06%	77.64 ± 9.10%	17.59 ± 15.36%
	HTNL	86.14 ± 5.31%	79.15 ± 8.05%	18.30 ± 13.84%
50%	Benchmark	84.48 ± 4.08%	76.64 ± 6.09%	20.19 ± 9.89%
	FineTune	85.63 ± 3.79%	78.39 ± 5.76%	15.62 ± 15.45%
	Init	90.17 ± 4.23%	85.25 ± 6.32%	10.73 ± 7.78%
	HT	87.01 ± 6.15%	80.48 ± 9.21%	14.82 ± 16.68%
	HTNL	89.48 ± 3.44%	84.20 ± 5.19%	16.35 ± 7.35%

It can be observed that the “FineTune” model excelled in terms of overall accuracy and Kappa when using 10% of the Sentinel7c dataset for training, and it maintained the lowest FNR for the target class “Smoke” when training data were at 10%, 20%, and 30%. However, the “Init” model began to exhibit higher overall accuracy and Kappa when using 20% of the training data and achieved the lowest FNR for the target class ‘Smoke’ when the training data exceeded 40%.

The observed trend suggests that the “FineTune” method, with its careful updates to pretrained model weights, effectively leveraged the spectral patterns learned from the source domain when the training data were extremely limited. In contrast, the “Init” method adapted better to the new spectral patterns in the target domain when more training data were available.

Figure 6.3 shows the boxplots of the results for all transferred models and the benchmark model. The figure reveals consistent trends when evaluating the models based on their median values.

Figure 6.4 presents a line chart depicting the performance of all models trained with 10% to 50% of the Sentinel7c dataset. The chart clearly illustrates that the proposed approach, particularly the “Init” method, demonstrates superior performance compared to the benchmark model and the conventional transfer learning approach.

In summary, the findings highlight the effectiveness of the proposed transfer learning approach for cross-sensor transfer learning, particularly in scenarios with limited training data from the new sensor. Notably, the approach is model-agnostic, allowing the replacement of the base model VIB_SD with any known model structure, provided that the input shape is modified to integrate with the IA module seamlessly.

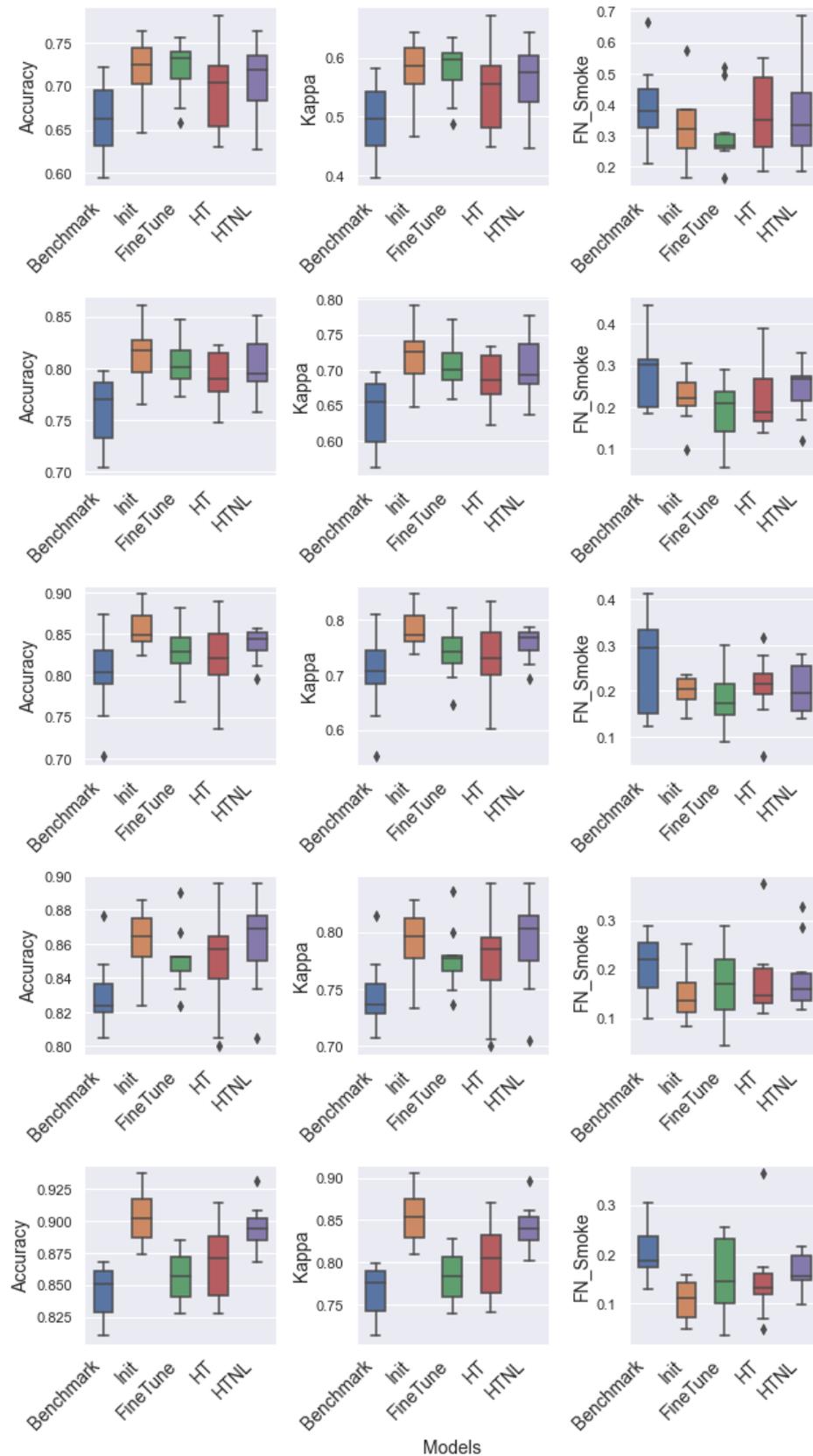


Figure 6.3: Boxplots of the results.

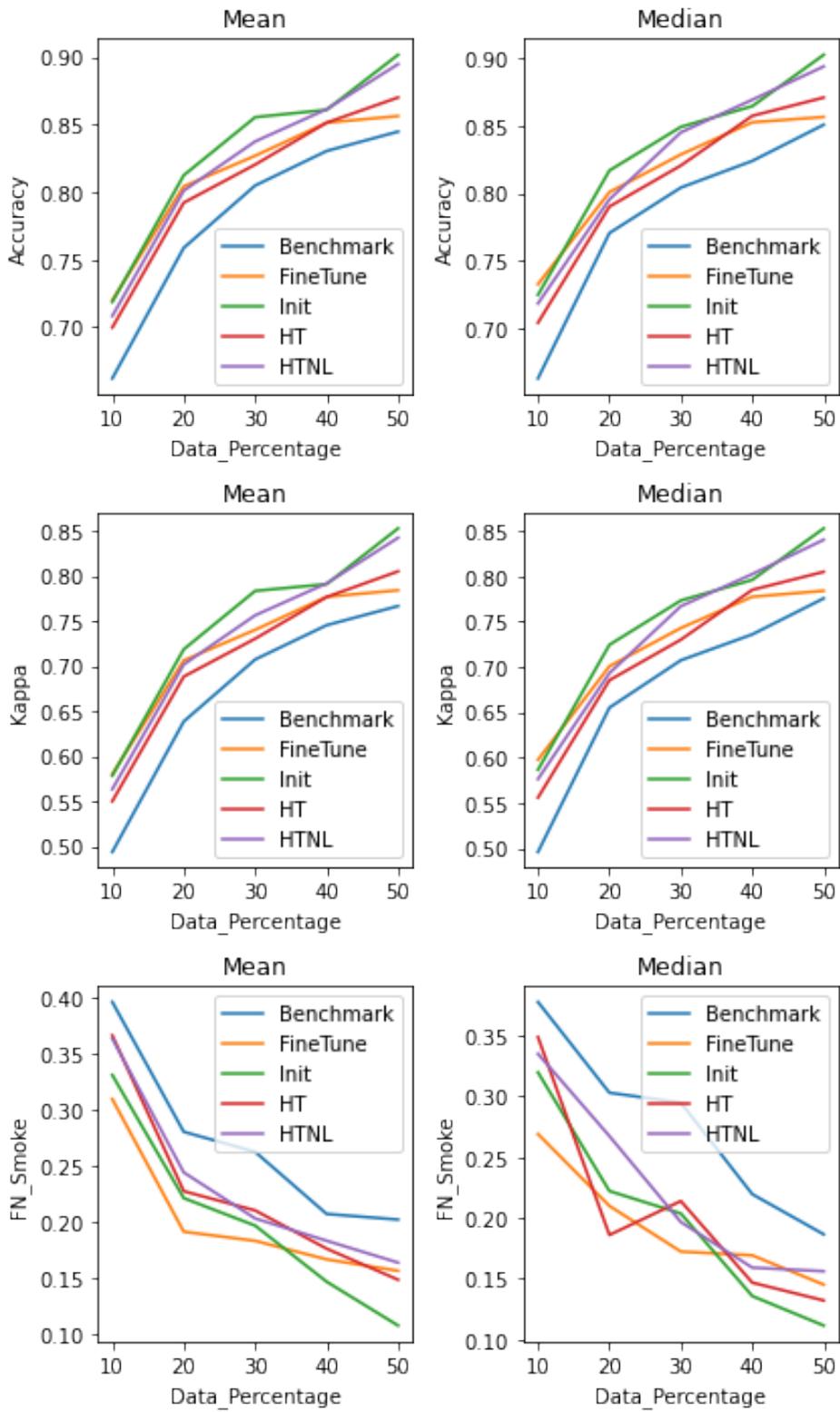


Figure 6.4: Line charts of the results.

Furthermore, the proposed approach stands out from the method presented in [283], which trained a brand new model for the target domain using data featuring pseudo labels generated by a pretrained model from the source domain. Similarly, in contrast to the approach outlined in [285], which involves substantial data pre-processing based on domain knowledge, the proposed approach is straightforward, eliminating the need for intensive model training, feature engineering or domain-specific knowledge.

Nevertheless, given that the proposed approach depends on the IA module, existing DL models must be integrated with the IA module prior to applying this approach. Consequently, it is not possible to directly transfer models that have already been trained on large datasets. Instead, these models must undergo retraining after integration before they can be utilised. Moreover, additional research may be required to ascertain the efficacy of the proposed method with specific base models, as the effectiveness of the IA module could be influenced by the unique structures of these models, as discussed in Chapter 5.

6.5 Conclusion

In this chapter, it is demonstrated that adapting class-oriented spectral patterns learned by the IA module can facilitate cross-sensor transfer learning for smoke detection using multispectral satellite imagery. The results show that, with the assistance of the IA module, transferring the pretrained IA_VIB_SD model from Landsat6c and updating it with a small amount of Sentinel7c data consistently resulted in significantly higher accuracy compared to training the model exclusively with an equivalent amount of Sentinel7c data. Remarkably, the best results were obtained using the proposed “Fine-Tune” and “Init” methods. These two methods replace the IA module to accept seven bands as input from the Sentinel7c dataset before updating the model. This indicates that the IA module autonomously adapts to new data and effectively retains critical spectral features learned from the Landsat6c dataset for accurate prediction.

It is worth noting that the number of input channels for VIB_SD is set to 32 to integrate with IA in IA_VIB_SD. This is because the IA module outputs a feature map with 32 channels, representing the spectral patterns it extracts by default. This strategy can be applied to any DL model structure. Therefore, this approach offers a model-agnostic, easy-to-implement solution for early smoke detection using multi-sensor, multispectral imagery, enabled by IA-supported cross-sensor transfer learning

that accounts for varying spectral bands across different sensors.

Chapter 7

Conclusions

7.1 Summary of Contributions

This thesis introduces three innovative approaches to enhancing timely and accurate scene-level EF smoke detection using multispectral satellite imagery, thereby facilitating fire disaster prevention.

The first approach is the lightweight CNN model VIB_SD, presented in Chapter 4, designed specifically for smoke detection with potential for onboard SmallSat applications. Additionally, the Landsat6c dataset, a multispectral satellite imagery smoke detection training dataset, was created. As the first of its kind in the literature, Landsat6c enables the exploration of IR bands to improve VIB_SD’s accuracy in smoke detection. A comprehensive investigation using VIB_SD and Landsat6c showed that each IR band contributes individually to enhancing VIB_SD’s detection accuracy.

The second approach, proposed in Chapter 5, is the IA module which enables DL models to learn class-oriented spectral patterns and enhance smoke detection accuracy, particularly with multispectral satellite imagery. IA’s effectiveness was demonstrated through its integration with DL models of various architectures, including widely-recognised CNN models and the VIB_SD model. Notably, the IA_VIB_SD model, combining IA and VIB_SD, outperformed other models when trained on both the RGB USTC_SmokeRS dataset and the multispectral Landsat6c dataset. Class-oriented characteristics can be clearly observed in the IA-learned spectral patterns through visualisation.

The third approach, introduced in Chapter 6, leverages cross-sensor transfer learning to facilitate fast model development, thereby achieving more timely smoke detection with a higher temporal resolution gained through combining multiple satellites for the task. The proposed transfer learning approach innovatively adapts class-oriented spectral patterns learned from the source domain to the target domain, facilitated by the IA_VIB_SD model. Specifically, by retaining the VIB_SD part within the source domain-pretrained IA_VIB_SD during the transfer, the knowledge learned from the source domain can be carried over to the target domain, while the IA module is replaced to adapt to the different band number in the target domain imagery. This approach achieved superior accuracy compared to training IA_VIB_SD purely on the target domain, as well as the conventional transfer learning approach, which employs a mapping layer to align the band dimensions between the source and target domains.

The major contributions of this thesis, rooted in the three innovative approaches, are summarised as follows:

1. Invention of the lightweight VIB_SD model, the IA module, and the integrated lightweight IA_VIB_SD model, contributing to accurate smoke detection utilising multispectral satellite imagery:

- VIB_SD achieved competitive accuracy with the state-of-the-art model, SAFA, while using less than 2% of its parameters, demonstrating great potential for onboard-SmallSat applications.
- The IA module, novelly incorporating band, spatial, and channel attention mechanisms, enables DL models to automatically learn class-oriented spectral patterns. It substantially enhances DL-based smoke detection using multispectral satellite imagery, marking a pioneering achievement in the literature.
- The IA_VIB_SD model, integrating IA and VIB_SD, further improves smoke detection accuracy with only a marginal increase in parameters compared to VIB_SD.

2. Development of an innovative cross-sensor transfer learning approach aided by the IA module, facilitating rapid model development, contributes towards solutions for timely smoke detection using multiple satellites:

- This approach is straightforward to implement, avoiding the need for complex data engineering, which typically requires domain expertise and is time-consuming.
- It offers flexibility in the choice of satellites and spectral band combinations in their data, as it accommodates varying band numbers between source and target domains.
- By preserving the original spectral bands, it ensures that the learning and adaptation of spectral patterns are based on authentic data, leading to robust model adaptation across different sensors.
- This approach also presents a promising few-shot learning strategy in remote sensing. Notably, the transferred IA_VIB_SD model significantly outperformed the benchmark IA_VIB_SD model, even when trained with only 10% of the target domain data, with as few as 10 training samples per class.

3. Creation of two multispectral satellite imagery training datasets from different sensors, contributing to future research:

- **Landsat6c:** Derived from Landsat 5 TM and Landsat 8 OLI, this dataset comprises six spectral bands (i.e., RGB, NIR, SWIR_1, and SWIR_2) with a spatial resolution of 30 metres. It is labelled into three scene classes (“Smoke”, “Clear”, and “Other_aerosol”), each containing over 900 training samples. This dataset can support future research and industrial applications in satellite-based EF smoke detection. In addition, it can be further used to generate pixel-level segmentation training datasets, facilitating future research in pixel-level smoke detection.
- **Sentinel17c:** Derived from Sentinel-2 MSI, featuring seven spectral bands (i.e., RGB, NIR, NIR_2, SWIR_2, and SWIR_3) with a spatial resolution of 10 metres and the same scene classes as **Landsat6c**. It contains 351 imagery files with over 100 samples per class. It serves as an ideal target domain for cross-sensor transfer learning. Additionally, it can be further expanded to serve as an independent training dataset for smoke detection studies based on Sentinel-2 MSI imagery.

7.2 Practical Implications

The findings and contributions of this thesis hold significant practical implications for fire disaster mitigation and remote sensing applications:

1. Potential for NRT EF smoke detection using multiple existing satellites with higher spatial resolution:

- The lightweight IA_VIB_SD model, already trained on Landsat6c and successfully transferred to Sentinel7c with fewer than 200 training samples in total, can be adapted to other satellites in a similar manner. The reduced need for labelled training samples allows for rapid model transfer.
- Once the models achieve the required detection accuracy through ground-based simulations, they can potentially be deployed onboard these satellites for timely and reliable identification of EF smoke scenes, which is crucial for mitigating extreme fire disasters.

2. Rapid development of DL models for CubeSat constellations and other resource-constrained platforms enhances the operational capabilities of satellite-based fire monitoring systems, supporting fast response and mitigation efforts:

- The Landsat6c-trained IA_VIB_SD model can be transferred to CubeSat sensors with sufficient observational data that have successfully captured smoke events.
- This approach can also be applied to CubeSat sensors yet to be launched using synthetic data for simulation, provided the sensor specifications are known, as demonstrated in [55].
- The success of IA_VIB_SD on onboard-SmallSats may inspire the development of models specifically designed for CubeSat constellations, expanding their capability for NRT smoke detection and other remote sensing tasks.

3. Applications in other remote sensing domains:

- The IA module or the IA_VIB_SD model could be applied in domains such as water observation, vegetation disease detection, or other remote sensing tasks where spectral information is crucial.
- The IA-aided transfer learning approach may benefit other areas of remote sensing where labelled imagery is scarce.
- The IA module has shown potential to improve the interpretability of DL models, aiding decision-making and analysis in environmental monitoring and disaster management applications.

7.3 Future Directions

Building on the contributions and findings of this research, several avenues for future work are identified:

1. Refinement and optimisation of IA_VIB_SD:

- Further refine and optimise the IA_VIB_SD model structure to enhance its accuracy and stability in detecting EF smoke scenes.

- Expand the current datasets and collect new datasets from different sensors, particularly hyperspectral sensors on SmallSats, to improve model training and validation.

2. Exploration of IA module applications:

- Investigate the application of the IA module in other remote sensing domains to verify its effectiveness in learning class-oriented spectral patterns for various environmental monitoring tasks.
- Develop methods to leverage the IA module for pixel-level labelling or segmentation tasks, enhancing its utility in detailed image analysis.

3. Advancement of cross-sensor transfer learning:

- Further explore and optimise the proposed transfer learning approach to improve its robustness and efficiency in adapting DL models to new sensors with limited training data.
- Extend the transfer learning framework to incorporate data from additional satellite sensors and constellations, supporting a more comprehensive and integrated approach to smoke detection and other remote sensing applications.

4. Integration of advanced techniques and additional information:

- Investigate state-of-the-art techniques such as self-attention mechanisms and lightweight transformers to improve the spatial and spectral feature extraction capabilities of DL models.
- Explore the integration of additional information, such as weather and environmental data, with satellite imagery to predict fire risk and enhance proactive fire disaster prevention.

5. Broader applications in remote sensing:

- Apply the IA module or IA_VIB_SD model in other remote sensing domains, such as water observation and vegetation disease detection, where spectral patterns can provide significant contributions.
- Explore cross-sensor applications in other domains, aided by the IA module, to enhance the versatility and applicability of DL models in various remote sensing tasks.

CHAPTER 7. CONCLUSIONS

This chapter concludes the thesis by summarising the significant contributions, practical implications, and future research directions, providing a comprehensive overview of the advancements achieved in developing innovative approaches to smoke detection using multispectral satellite imagery.

Appendix A: Statements of Authorship

Title of Paper	Investigating the impact of using IR bands on early fire smoke detection from Landsat imagery with a lightweight CNN model	
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style	
Publication Details Relevant reference details	Liang Zhao, Jixue Liu, Stefan Peters, Jiuyong Li, Simon Oliver, and Norman Mueller, "Investigating the impact of using IR bands on early fire smoke detection from Landsat imagery with a lightweight CNN model", Remote Sensing, volume. 14, issue 13, page. 3047, 2022, https://doi.org/10.3390/rs14133047	
PRINCIPAL AUTHOR		
Name of Principal Author (Candidate)	Liang Zhao	
Contribution to the Paper Brief description of your work in this publication	I designed the method under the guidance and supervision of my supervisors, implemented the method, constructed the dataset, and conducted the experiments and analyses. I primarily contributed to the drafting and revision of the manuscript.	
Overall Percentage (%)	60%	
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	 <p>Liang Zhao</p> <small>Digitally signed by Liang Zhao Date: 2024.10.10 15:34:14 +10'30'</small>	Date 10/10/2024
CO-AUTHOR CONTRIBUTIONS		
<p>By signing the Statement of Authorship, each author certifies that:</p> <ol style="list-style-type: none"> I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate to include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution. 		
Name of Co-Author 1	Jixue Liu	

% Contribution to the Paper	12%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	liuj  <small>Digitally signed by liuj DN: CN=liuj Reason: I am the author of this document Location: mawson lakes Date: 2024-10-11 11:48:09 Foxit Reader Version: 9.7.0</small>	11-Oct-24 Date
Name of Co-Author 2	Stefan Peters	
% Contribution to the Paper	8%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Stefan Peters  <small>Digitally signed by Stefan Peters Date: 2024.10.11 15:30:00 +10'30'</small>	11-Oct-24 Date
Name of Co-Author 3	Jiuyong Li	
% Contribution to the Paper	8%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Jiuyong Li  <small>Digitally signed by Jiuyong Li Date: 2024.10.11 15:05:12 +10'30'</small>	11/10/24 Date
Name of Co-Author 4	Simon Oliver	
% Contribution to the Paper	6%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Simon Oliver  <small>Digitally signed by Simon Oliver Date: 2024.11.05 10:29:22 -06'00'</small>	5-Nov-24 Date
Name of Co-Author 5	Norman Mueller	
% Contribution to the Paper	6%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Norman Mueller  <small>Digitally signed by Norman Mueller Date: 2024.11.05 15:22:11 +11'00'</small>	5-Nov-24 Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 7		
% Contribution to the Paper		

Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 8		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 9		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 10		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date

Title of Paper	Learning classs-pecific spectral patterns to improve deep learning-based scene-level fire smoke detection from multi-spectral satellite imagery	
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style	
Publication Details Relevant reference details	Liang Zhao, Jixue Liu, Stefan Peters, Jiuyong Li, Norman Mueller, and Simon Oliver, "Learning classs-pecific spectral patterns to improve deep learning-based scene-level fire smoke detection from multi-spectral satellite imagery", Remote Sensing Applications: Society and Environment, volume. 34, 2024, 101152. https://doi.org/10.1016/j.rsase.2024.101152	
PRINCIPAL AUTHOR		
Name of Principal Author (Candidate)	Liang Zhao	
Contribution to the Paper Brief description of your work in this publication	I designed and implemented the method, conducted the experiments and analyses. I primarily contributed to the drafting and revision of the manuscript.	
Overall Percentage (%)	60%	
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Liang Zhao  Digitally signed by Liang Zhao Date: 2024.10.10 16:07:44 +10'30'	Date 10/10/2024
CO-AUTHOR CONTRIBUTIONS		
By signing the Statement of Authorship, each author certifies that: <ol style="list-style-type: none"> I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution. 		
Name of Co-Author 1	Jixue Liu	

% Contribution to the Paper	10%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	liuj  Digitally signed by liuj DN: CN=liuj Reason: I am the author of this document Location: mawson lakes Date: 2024-10-11 11:49:08 Foxit Reader Version: 9.7.0	11-Oct-24 Date
Name of Co-Author 2	Stefan Peters	
% Contribution to the Paper	10%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Stefan Peters  Digitally signed by Stefan Peters Date: 2024.10.11 15:27:17 +10'30'	11-Oct-24 Date
Name of Co-Author 3	Jiuyong Li	
% Contribution to the Paper	10%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Jiuyong Li  Digitally signed by Jiuyong Li Date: 2024.10.11 15:06:28 +10'30'	11/10/24 Date
Name of Co-Author 4	Norman Mueller	
% Contribution to the Paper	5%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Norman Mueller  Digitally signed by Norman Mueller Date: 2024.11.05 15:23:06 +11'00'	5-Nov-24 Date
Name of Co-Author 5	Simon Oliver	
% Contribution to the Paper	5%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Simon Oliver  Digitally signed by Simon Oliver Date: 2024.11.05 10:29:58 -06'00'	5-Nov-24 Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 7		
% Contribution to the Paper		

Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 8		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 9		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 10		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date

Title of Paper	Cross-sensor transfer learning for fire smoke scene detection using variable-bands multi-spectral satellite imagery aided by spectral patterns	
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style	
Publication Details Relevant reference details	Liang Zhao, Jixue Liu, Stefan Peters, Jiuyong Li, Norman Mueller, and Simon Oliver, "Cross-sensor transfer learning for fire smoke scene detection using variable-bands multi-spectral satellite imagery aided by spectral patterns", International Journal of Remote Sensing, 45(10), 3332–3348. https://doi.org/10.1080/01431161.2024.2343430	
PRINCIPAL AUTHOR		
Name of Principal Author (Candidate)	Liang Zhao	
Contribution to the Paper Brief description of your work in this publication	I designed and implemented the method, constructed the dataset, conducted the experiments and analyses, and primarily contributed to the drafting and revision of the manuscript.	
Overall Percentage (%)	60%	
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Liang Zhao  Digitally signed by Liang Zhao Date: 2024.10.10 16:13:51 +10'30'	Date 10/10/2024
CO-AUTHOR CONTRIBUTIONS		
By signing the Statement of Authorship, each author certifies that: <ol style="list-style-type: none"> I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate to include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution. 		
Name of Co-Author 1	Jixue Liu	

% Contribution to the Paper	10%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	liuj  Digitally signed by liuj DN: CN=liuj Reason: I am the author of this document Location: mawson lakes Date: 2024-10-11 11:50:09 Foxit Reader Version: 9.7.0	11-Oct-24 Date
Name of Co-Author 2	Stefan Peters	
% Contribution to the Paper	10%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Stefan Peters  Digitally signed by Stefan Peters Date: 2024.10.11 15:23:16 +10'30'	11-Oct-24 Date
Name of Co-Author 3	Jiuyong Li	
% Contribution to the Paper	10%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Jiuyong Li  Digitally signed by Jiuyong Li Date: 2024.10.11 15:07:35 +10'30'	11/10/24 Date
Name of Co-Author 4	Norman Mueller	
% Contribution to the Paper	5%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Norman Mueller  Digitally signed by Norman Mueller Date: 2024.11.05 15:23:36 +11'00'	5-Nov-24 Date
Name of Co-Author 5	Simon Oliver	
% Contribution to the Paper	5%	
Signature <i>Signatures will be redacted by the Library at the time of publication</i>	Simon Oliver  Digitally signed by Simon Oliver Date: 2024.11.05 10:30:45 -06'00'	5-Nov-24 Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 7		
% Contribution to the Paper		

Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 8		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 9		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date
Name of Co-Author 10		
% Contribution to the Paper		
Signature <i>Signatures will be redacted by the Library at the time of publication</i>		Date

References

- [1] Dimitra Angra and Kalliopi Sapountzaki. “Climate change affecting forest fire and flood risk—Facts, predictions, and perceptions in central and south Greece”. In: *Sustainability* 14.20 (2022). ISSN: 2071-1050. URL: <https://www.mdpi.com/2071-1050/14/20/13395>.
- [2] Fantina Tedim et al. “Defining extreme wildfire events: difficulties, challenges, and impacts”. In: *Fire* 1.1 (2018), p. 9.
- [3] Jason A Vargo. “Time series of potential US wildland fire smoke exposures”. In: *Frontiers in public health* 8 (2020), p. 126.
- [4] Nicolas Borchers Arriagada et al. “Unprecedented smoke-related health burden associated with the 2019–20 bushfires in eastern Australia”. In: *Medical Journal of Australia* 213.6 (2020), pp. 282–283.
- [5] Ashley Kendell, Alison Galloway, and Colleen Milligan. “Wildfires: Introduction, Impact, Policy, and Planning”. In: *The Path of Flames*. CRC Press, 2024, pp. 3–18.
- [6] Jamie Ranse et al. “Impact of fine particulate matter (PM_{2.5}) smoke during the 2019/2020 Australian bushfire disaster on emergency department patient presentations”. In: *The Journal of Climate Change and Health* 6 (2022), p. 100113.
- [7] Fay H Johnston et al. “Unprecedented health costs of smoke-related PM_{2.5} from the 2019–20 Australian megafires”. In: *Nature Sustainability* 4.1 (2021), pp. 42–47.
- [8] J. McDaniel. “Fire Detection and Response: An Overview”. In: *International Journal of Wildland Fire* 30.5 (2021), pp. 350–365.
- [9] M. F. Goodchild. “Citizens as sensors: the world of volunteered geography”. In: *GeoJournal* 69.4 (2007), pp. 211–221.
- [10] A. H. Perera, L. Buse, and M. G. Weber. “Detection of wildfires”. In: *Emulating Natural Forest Landscape Disturbances: Concepts and Applications*. New York: Columbia University Press, 2007, pp. 105–123.

References

- [11] A. Roberts and P. Jones. “Fire detection advancements in modern fire towers”. In: *International Journal of Wildland Fire* 29.5 (2020), pp. 453–460.
- [12] R. Williams and T. Harris. “Limitations of surveillance camera coverage in fire detection”. In: *Fire Safety Journal* 55.3 (2022), pp. 204–210.
- [13] M. Johnson and R. White. “The cost of maintaining fire lookout towers”. In: *Fire Management Today* 78.2 (2018), pp. 15–20.
- [14] J. Smith and K. Lee. “Challenges of remote fire detection systems”. In: *Remote Sensing Applications* 12.1 (2019), pp. 34–42.
- [15] A. Restas. “Drone applications for supporting disaster management”. In: *World Journal of Engineering and Technology* 3.3 (2015), pp. 316–321.
- [16] V. G. Ambrosia, S. S. Wegener, and D. V. Sullivan. “Integrated UAV-based imaging and sensor systems for fire assessment and monitoring”. In: *International Journal of Remote Sensing* 32.24 (2011), pp. 7575–7587.
- [17] W. Schroeder. “Validation of GOES-R fire detection products using VIIRS”. In: *Remote Sensing of Environment* 173 (2016), pp. 224–234.
- [18] L. Giglio. “Active fire detection using the MODIS fire product”. In: *Remote Sensing of Environment* 83.1 (2003), pp. 195–203.
- [19] *UCS Satellite Database*. <https://www.ucsusa.org/resources/satellite-database>. Accessed on 14.11.2023.
- [20] Wanting Wang et al. “An improved algorithm for small and cool fire detection using MODIS data: A preliminary study in the southeastern United States”. In: *Remote sensing of Environment* 108.2 (2007), pp. 163–170.
- [21] Guang Xu and Xu Zhong. “Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8”. In: *Remote Sensing Letters* 8.11 (2017), pp. 1052–1061.
- [22] Zixi Xie et al. “A spatiotemporal contextual model for forest fire detection using Himawari-8 satellite data”. In: *Remote Sensing* 10.12 (2018), p. 1992.
- [23] Waseem Rawat and Zenghui Wang. “Deep convolutional neural networks for image classification: A comprehensive review”. In: *Neural computation* 29.9 (2017), pp. 2352–2449.
- [24] Tsung-Yi Lin et al. “Microsoft COCO: Common objects in context”. In: *European conference on computer vision* (2014), pp. 740–755.

- [25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in Neural Information Processing Systems* 25 (2012).
- [26] Jason Yosinski et al. “How transferable are features in deep neural networks?”. In: *Advances in neural information processing systems* 27 (2014).
- [27] Liang Zhao et al. “Investigating the impact of using IR bands on early fire smoke detection from Landsat imagery with a lightweight CNN model”. In: *Remote Sensing* 14.13 (2022), p. 3047.
- [28] Stefan Peters. “Remote fire detection: satellite-based early fire detection”. English. In: *The Forester* February/March (2021). Copyright 2021 The Institute of Foresters of Australia & Australian Forest Growers, pp. 32–34. URL: <https://hdl.handle.net/11541.2/29037>.
- [29] Youngwook Kang et al. “Neurosurgeon: Collaborative intelligence between the cloud and mobile edge”. In: *ACM SIGARCH Computer Architecture News* 45.1 (2017), pp. 615–629.
- [30] Baskara Ardianto, Mikail Abdullah, and Ari Iriawan. “Optimizing complexity of CNN models for resource constrained devices: QRS detection case study”. In: *Proceedings of the International Conference on Electrical Engineering, Informatics, and Its Education (CEIE)*. 2021, pp. 1–5.
- [31] Huayi Wu et al. “Enabling deep learning at the IoT edge: An empowered heterogeneous computing perspective”. In: *IEEE Computer* 53.7 (2020), pp. 24–33.
- [32] W. Schroeder et al. “The new VIIRS 375m active fire detection data product: algorithm description and initial assessment”. In: *Remote Sens Environ* 143 (2014), pp. 85–96. DOI: 10.1016/j.rse.2013.12.008.
- [33] G. Curzi, D. Modenini, and P. Tortora. “Large Constellations of Small Satellites: A Survey of Near Future Challenges and Missions”. In: *Aerospace* 7.9 (2020), p. 133. DOI: 10.3390/aerospace7090133.
- [34] Corey Johnston. “China’s Gaofen Satellite Program: Overview and Strategic Implications”. In: *Journal of Strategic Studies* 10.3 (2020), pp. 45–60.
- [35] Liang Gao and Hong Zhang. “Overview of China’s Earth Observation System Development”. In: *Remote Sensing* 9.6 (2017), p. 520.

References

- [36] Elsa Kania and John Costello. *No Place to Hide: A Look at China's Geosynchronous Surveillance Capabilities*. <https://www.csis.org/analysis/no-place-hide-look-chinas-geosynchronous-surveillance-capabilities>. 2024.
- [37] Jonathan Smith and Kimberly Lee. “The High Cost of Satellite Development: A Detailed Analysis”. In: *Aerospace Economics* 15.4 (2017), pp. 78–95.
- [38] Michael Thompson and Alice Roberts. “NASA’s Budget for Satellite Operations: An In-depth Review”. In: *Space Policy* 34 (2018), pp. 12–25.
- [39] Sundar A Christopher et al. “First estimates of the radiative forcing of aerosols generated from biomass burning using satellite data”. In: *Journal of Geophysical Research: Atmospheres* 101.D16 (1996), pp. 21265–21273.
- [40] Koji Asakuma et al. “Detection of biomass burning smoke in satellite images using texture analysis”. In: *Atmospheric Environment* 36.9 (2002), pp. 1531–1542.
- [41] Alexandra Larsen et al. “A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication”. In: *Journal of exposure science & environmental epidemiology* 31.1 (2021), pp. 170–176.
- [42] Yi Qin, Ross Mitchell, and Bruce W Forgan. “Characterizing the aerosol and surface reflectance over Australia using AATSR”. In: *IEEE Transactions on Geoscience and Remote Sensing* 53.11 (2015), pp. 6163–6182.
- [43] Rui Ba et al. “Smokenet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention”. In: *Remote Sensing* 11.14 (2019), p. 1702.
- [44] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [45] Gao Huang et al. “Densely connected convolutional networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 4700–4708.
- [46] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-excitation networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.
- [47] Shikun Chen et al. “Global2Salient: Self-adaptive feature aggregation for remote sensing smoke detection”. In: *Neurocomputing* 466 (2021), pp. 202–220.

- [48] Gong Cheng et al. “When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs”. In: *IEEE transactions on geoscience and remote sensing* 56.5 (2018), pp. 2811–2821.
- [49] Yansheng Li, Yongjun Zhang, and Zhihui Zhu. “Error-tolerant deep learning for remote sensing image scene classification”. In: *IEEE transactions on cybernetics* 51.4 (2020), pp. 1756–1768.
- [50] Fengpeng Li et al. “High-resolution remote sensing image scene classification via key filter bank based on convolutional neural network”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.11 (2020), pp. 8077–8092.
- [51] Yansheng Li et al. “Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 59.12 (2021), pp. 10590–10603.
- [52] Michael Ruminski et al. “Monitoring Wildfire Activity in North America with the Hazard Mapping System”. In: *Fire* 1.1 (2008), pp. 56–77.
- [53] NASA Applied Sciences. *Utilization of Multi-Sensor Active Fire Detections to Map Fires in the US. The Future of Monitoring Trends in Burn Severity (MTBS)*. 2024. URL: <https://appliedsciences.nasa.gov/what-we-do/projects/utilization-multi-sensor-active-fire-detections-map-fires-us-future-monitoring>.
- [54] N Chrysoulakis et al. “An improved algorithm for the detection of plumes caused by natural or technological hazards using AVHRR imagery”. In: *Remote sensing of environment* 108.4 (2007), pp. 393–406.
- [55] Sha Lu et al. “Onboard AI for Fire Smoke Detection Using Hyperspectral Imagery: An Emulation for the Upcoming Kanyini Hyperscout-2 Mission”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), pp. 9629–9640. DOI: 10.1109/JSTARS.2024.3394574.
- [56] Liang Zhao et al. “Learning class-specific spectral patterns to improve deep learning-based scene-level fire smoke detection from multi-spectral satellite imagery”. In: *Remote Sensing Applications: Society and Environment* (2024), p. 101152. ISSN: 2352-9385. DOI: <https://doi.org/10.1016/j.rsase.2024.101152>. URL: <https://www.sciencedirect.com/science/article/pii/S2352938524000168>.
- [57] Liang Zhao et al. “Cross-sensor transfer learning for fire smoke scene detection using variable-bands multi-spectral satellite imagery aided by spectral patterns”. In: *International Journal of Remote Sensing* 45.10 (2024), pp. 3332–3348.

References

- [58] NASA. *Terra Satellite Overview*. Accessed: 2024-06-13. 2024. URL: <https://terra.nasa.gov/about/terra-instruments>.
- [59] USGS. *Landsat 8 Overview*. Accessed: 2024-06-13. 2024. URL: <https://www.usgs.gov/landsat-missions/landsat-8>.
- [60] John A. Richards. *Remote Sensing Digital Image Analysis*. Springer, 2013.
- [61] National Oceanic and Atmospheric Administration (NOAA). *Remote Sensing and Satellites*. Accessed: 2024-06-13. 2024. URL: <https://www.noaa.gov/education/resource-collections/satellites/remote-sensing>.
- [62] National Aeronautics and Space Administration (NASA). *Overview of Satellite Types and Orbits*. Accessed: 2024-06-13. 2024. URL: <https://www.nasa.gov/satellites/overview>.
- [63] European Space Agency (ESA). *Sun-Synchronous Orbits Explained*. Accessed: 2024-06-13. 2024. URL: https://www.esa.int/Applications/Observing_the_Earth/Sun-synchronous_orbits.
- [64] Emily Johnson. “Launch Logistics for Large Satellites”. In: *Journal of Spacecraft and Rockets* 45 (2018), pp. 34–40.
- [65] John Smith. *Complexity in Satellite Development*. Space Tech Publishing, 2019.
- [66] Inmarsat. *Inmarsat-5 F4 Launch*. Accessed: 2024-06-13. 2023. URL: <https://www.inmarsat.com/newsroom/press-releases/inmarsat-5-f4-launch>.
- [67] NASA. *Terra Satellite Mission*. Accessed: 2024-06-13. 2023. URL: <https://terra.nasa.gov/about/terra>.
- [68] NASA. “Development Times for Large Satellites”. In: *NASA Reports* 10 (2020), pp. 15–20.
- [69] Jane Doe. “Challenges in Upgrading Large Satellites”. In: *International Journal of Space Science* 55 (2021), pp. 100–110.
- [70] George Wilson. *Small Satellite Integration*. Satellite Publishing House, 2017.
- [71] NASA. *CubeSat Cost Breakdown*. Accessed: 2024-06-13. 2023. URL: <https://www.nasa.gov/smallsat-institute/cubesat-cost>.
- [72] Michael Brown. “Advantages of Small Satellites”. In: *Space Innovations* 20 (2020), pp. 50–55.
- [73] Laura Garcia. “Flexible Launch Options for Small Satellites”. In: *Journal of Space Launch* 30 (2019), pp. 70–75.

- [74] Anna Miller. “Operational Adaptability of Small Satellites”. In: *Journal of Space Technology* 60 (2022), pp. 25–30.
- [75] Robert Taylor. *The Role of Small Satellites in Space Exploration*. Exploration Press, 2018.
- [76] MIT Technology Review. *Small Satellites and Innovation*. Accessed: 2024-06-13. 2023. URL: <https://www.technologyreview.com/2023/06/13/small-satellites-and-innovation/>.
- [77] NASA. *Small Satellites*. Accessed: 2024-06-13. 2023. URL: <https://www.nasa.gov/smallsat-institute>.
- [78] European Space Agency. *Small Satellites at ESA*. Accessed: 2024-06-13. 2023. URL: https://www.esa.int/Applications/Telecommunications_Integrated_Applications/Small_Satellites.
- [79] J. Smith and A. Doe. “The Growing Role of Small Satellites in Earth Observation”. In: *Journal of Spacecraft and Rockets* 58.4 (2023), pp. 1234–1245.
- [80] C. Hevner et al. “CubeSats: A new generation of picosatellite for education and industry low-cost space experimentation”. In: *Journal of Small Satellites* 1.1 (2011), pp. 3–16.
- [81] K. Woellert et al. “CubeSats: Cost-effective science and technology platforms for emerging and developing nations”. In: *Advances in Space Research* 47.4 (2011), pp. 663–684.
- [82] NASA. *How Satellites Work*. Accessed: 2024-06-13. 2023. URL: <https://www.nasa.gov/smallsat-institute>.
- [83] European Space Agency. *Satellite Constellations: Global Connectivity from Space*. Accessed: 2024-06-13. 2023. URL: https://www.esa.int/Applications/Telecommunications_Integrated_Applications/Small_Satellites.
- [84] Iridium Communications Inc. *Company History*. Accessed: 2024-06-13. 2023. URL: <https://www.iridium.com/company-history/>.
- [85] U.S. Government. *Global Positioning System (GPS)*. Accessed: 2024-06-13. 2023. URL: <https://www.gps.gov/systems/gps/>.
- [86] OneWeb. *OneWeb*. Accessed: 2024-06-13. 2023. URL: <https://www.oneweb.world/>.
- [87] SpaceX. *Starlink*. Accessed: 2024-06-13. 2023. URL: <https://www.spacex.com/what-is-starlink/>.

References

- [88] European Space Agency. “PlanetScope”. In: *Earth Online* (2024). URL: <https://earth.esa.int/eogateway/missions/planetscope>.
- [89] Planet. “Supporting Australia Initiatives”. In: *Planet* (2024). URL: <https://www.planet.com>.
- [90] OroraTech. *Technology - Our Satellites Change The Game*. Accessed: 2024-06-13. 2023. URL: <https://ororatech.com/technology/>.
- [91] NASA Earth Observatory. *Satellite Constellations*. Accessed: 2024-06-13. 2023. URL: <https://earthobservatory.nasa.gov/features/OrbitsCatalog>.
- [92] European Space Agency. *Satellite Orbits - An Overview*. Accessed: 2024-06-13. 2023. URL: https://www.esa.int/Applications/Observing_the_Earth/Overview4.
- [93] National Oceanic and Atmospheric Administration (NOAA). “Wildfire Monitoring Using HMS”. In: *NOAA Wildfire Monitoring Reports* (2020). URL: <https://www.noaa.gov/wildfire-monitoring>.
- [94] U.S. Geological Survey (USGS). *Landsat Satellite History and Data*. 2020. URL: <https://landsat.usgs.gov/>.
- [95] European Space Agency (ESA). *Sentinel-2 Mission Overview*. 2021. URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2.
- [96] Space News. “Data Integration Challenges in Satellite Constellations”. In: *Space News* (2024). URL: <https://spacenews.com/data-integration-constellations>.
- [97] Private Data Providers. *Data Access Issues in Satellite Constellations*. 2020. URL: <https://www.privatedata.com/satellite-access>.
- [98] W. Schroeder et al. “Cost Considerations in Satellite Fire Detection”. In: *Remote Sensing* 6 (2014), pp. 2440–2460. DOI: 10.3390/rs6032440.
- [99] S. Curzi et al. “Small Satellites for Earth Observation: Challenges and Future”. In: *Journal of Spacecraft and Rockets* 57 (2020), pp. 291–300. DOI: 10.2514/1.A34555.
- [100] CORDIS. *Technological Advancements in DL Models for Satellite Fire Detection*. 2024. URL: <https://cordis.europa.eu/project/id/101003456>.
- [101] USGS. *Satellite Data and Immediate Implementation for Fire Detection*. 2024. URL: <https://www.usgs.gov/fire-detection>.
- [102] European Space Agency (ESA). *Urgency of Satellite Data for Fire Detection*. 2024. URL: <https://www.esa.int/fire-detection-urgency>.

- [103] National Aeronautics and Space Administration (NASA). *Overview of Sensors and Their Applications*. Accessed: 2024-06-13. 2024. URL: <https://www.nasa.gov/sensors/overview>.
- [104] European Space Agency (ESA). *Active and Passive Sensors Explained*. Accessed: 2024-06-13. 2024. URL: https://www.esa.int/Applications/Observing_the_Earth/Active_and_Passive_Sensors.
- [105] European Space Agency (ESA). *Understanding Spectral Bands*. Accessed: 2024-06-13. 2024. URL: https://www.esa.int/Applications/Observing_the_Earth/Spectral_Bands.
- [106] National Aeronautics and Space Administration (NASA). *Spectral Resolution in Remote Sensing*. Accessed: 2024-06-13. 2024. URL: <https://www.nasa.gov/sensors/spectral-resolution>.
- [107] DigitalGlobe. *WorldView-3*. Accessed: 2024-06-13. 2023. URL: <https://www.digitalglobe.com/resources/satellite-information/worldview-3>.
- [108] NASA. *Landsat 8*. Accessed: 2024-06-13. 2023. URL: <https://landsat.gsfc.nasa.gov/satellites/landsat-8>.
- [109] Japan Meteorological Agency. *Himawari-8/9*. Accessed: 2024-06-13. 2023. URL: <https://www.jma.go.jp/jma/jma-eng/satellite/himawari8.html>.
- [110] NASA. *MODIS*. Accessed: 2024-06-13. 2023. URL: <https://modis.gsfc.nasa.gov/about/>.
- [111] China National Space Administration. *Gaofen-4*. Accessed: 2024-06-13. 2023. URL: <http://www.cnsa.gov.cn/n6758823/n6758838/c6772960/content.html>.
- [112] European Space Agency (ESA). *Swath Width and its Importance*. Accessed: 2024-06-13. 2024. URL: https://www.esa.int/Applications/Observing_the_Earth/Swath_Width.
- [113] National Aeronautics and Space Administration (NASA). *Temporal Resolution in Remote Sensing*. Accessed: 2024-06-13. 2024. URL: <https://www.nasa.gov/sensors/temporal-resolution>.
- [114] NASA Earth Observatory. *Orbits and Observations*. Accessed: 2024-06-13. 2023. URL: <https://earthobservatory.nasa.gov/features/OrbitsCatalog>.
- [115] J. Smith and A. Doe. “The Growing Role of Small Satellites in Earth Observation”. In: *Journal of Spacecraft and Rockets* 58.4 (2023), pp. 1234–1245.

References

- [116] European Space Agency (ESA). *Radiometric Resolution in Remote Sensing*. Accessed: 2024-06-13. 2024. URL: https://www.esa.int/Applications/Observing_the_Earth/Radiometric_Resolution.
- [117] Louis Giglio et al. “An enhanced contextual fire detection algorithm for MODIS”. In: *Remote sensing of environment* 87.2-3 (2003), pp. 273–282.
- [118] Wilfrid Schroeder et al. “The new VIIRS 375 m active fire detection data product: Algorithm description and initial assessment”. In: *Remote Sensing of Environment* 143 (2014), pp. 85–96.
- [119] Japan Meteorological Agency. *Himawari-8/9 Products*. 2016. URL: <https://www.data.jma.go.jp/mscweb/en/product/fire.html>.
- [120] NOAA. *Himawari-8/9 Overview*. 2016. URL: <https://www.ospo.noaa.gov/Operations/GOES-R/overview.html>.
- [121] Z. Li et al. “Challenges and advancements in the accuracy of satellite-based fire detection”. In: *Remote Sensing of Environment* 125 (2013), pp. 184–197.
- [122] A. M. S. Smith et al. “The limits of satellite remote sensing for detecting fire and its environmental effects”. In: *Ecological Applications* 25.7 (2015), pp. 1982–1998.
- [123] J. Brown et al. “Evaluating the performance of satellite-based fire detection algorithms”. In: *International Journal of Wildland Fire* 23.2 (2014), pp. 215–224.
- [124] M. W. Jones et al. “Improving the accuracy of satellite-based fire detection through enhanced algorithm development”. In: *International Journal of Remote Sensing* 38.14 (2017), pp. 4033–4055.
- [125] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [126] X. Lu et al. “A survey of deep learning approaches for remote sensing and geosciences”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.6 (2019), pp. 1882–1896.
- [127] Emilio Chuvieco et al. “Global burned-land estimation in Latin America using MODIS composite data”. In: *Ecological Applications* 20.3 (2010), pp. 870–883.
- [128] YJ Kaufman et al. “Potential global fire monitoring from EOS-MODIS”. In: *Journal of Geophysical Research: Atmospheres* 103.D24 (1998), pp. 32215–32238.
- [129] Louis Giglio et al. “An enhanced contextual fire detection algorithm for MODIS”. In: *Remote Sensing of Environment* 87.2-3 (2003), pp. 273–282.

- [130] Chris O Justice et al. “The MODIS fire products”. In: *Remote Sensing of Environment* 83.1-2 (2002), pp. 244–262.
- [131] Li Zhang, Liangpei Zhang, and V Kumar. “Deep learning for remote sensing data”. In: *IEEE Geoscience and Remote Sensing Magazine* 4.2 (2016), pp. 22–40.
- [132] X Hu et al. “Smoke detection in satellite images using deep learning”. In: *Remote Sensing* 11.2 (2019), p. 177.
- [133] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully convolutional networks for semantic segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 3431–3440.
- [134] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention* (2015), pp. 234–241.
- [135] Yi Chen et al. “Big Earth Data: a new frontier in Earth and information sciences”. In: *Big Earth Data* 4.1 (2020), pp. 1–16.
- [136] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *Computer Vision and Pattern Recognition* (2018).
- [137] Jeffrey S. Reid et al. “A global model for particulate matter emissions from vegetation fires”. In: *Atmospheric Chemistry and Physics* 5.3 (2005), pp. 993–1012.
- [138] David A. Peterson et al. “Quantifying the potential for high-altitude smoke injection in North American boreal wildfires using a chemistry-transport model”. In: *Atmospheric Chemistry and Physics* 18.8 (2018), pp. 5551–5568.
- [139] Sheryl K. Akagi et al. “Emission factors for open and domestic biomass burning for use in atmospheric models”. In: *Atmospheric Chemistry and Physics* 11.9 (2011), pp. 4039–4072.
- [140] Clay E. Stockwell et al. “Characterization of fuels and fire behavior for backward-inclined combustion of peat in Indonesia’s mega-rice project region”. In: *Journal of Geophysical Research: Atmospheres* 119.5 (2014), pp. 2928–2944.
- [141] Pablo E. Saide et al. “Evaluating WRFChem aerosol indirect effects in Southeast Asia during 7-SEAS/BASELInE using MODIS and VOCALS-REx data”. In: *Atmospheric Environment* 112 (2015), pp. 273–283.

References

- [142] Michael D. Fromm and Richard Servranckx. “Pyro-cumulonimbus injection of smoke to the stratosphere: Observations and impact of a super blowup in north-western Canada on 3–4 August 1998”. In: *Journal of Geophysical Research: Atmospheres* 108.D17 (2000), p. 4009.
- [143] Alma Hodzic et al. “Wildfire particulate matter in the western US and its impact on air quality in urban areas”. In: *Journal of Geophysical Research: Atmospheres* 112.D22 (2007), p. D23206.
- [144] Barry K Lavine, CE Davidson, and Anthony J Moores. “Genetic algorithms for spectral pattern recognition”. In: *Vibrational Spectroscopy* 28.1 (2002), pp. 83–95.
- [145] Alexander A Fingelkurts et al. “Systematic rules underlying spectral pattern variability: Experimental results and a review of the evidence”. In: *International Journal of Neuroscience* 113.10 (2003), pp. 1447–1473.
- [146] Nguyen Dinh Duong. “Water body extraction from multi spectral image by spectral pattern analysis”. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39 (2012), B8.
- [147] György Bázár et al. “NIR detection of honey adulteration reveals differences in water spectral pattern”. In: *Food Chemistry* 194 (2016), pp. 873–880.
- [148] Thomas M. Lillesand, Ralph W. Kiefer, and Jonathan W. Chipman. *Remote Sensing and Image Interpretation*. John Wiley & Sons, 2015.
- [149] John A Richards and Xiuping Jia. *Remote Sensing Digital Image Analysis: An Introduction*. Springer, 2016.
- [150] Nathalie Pettorelli et al. “Using the satellite-derived NDVI to assess ecological responses to environmental change”. In: *Trends in Ecology & Evolution* 20.9 (2005), pp. 503–510.
- [151] Rasmus Fensholt et al. “Evaluation of earth observation based long term vegetation trends—Intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data”. In: *Remote Sensing of Environment* 113.9 (2009), pp. 1886–1898.
- [152] S Escuin et al. “Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images”. In: *International Journal of Remote Sensing* 29.4 (2008), pp. 1053–1073.

- [153] Sha Huang et al. “A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing”. In: *Journal of Forestry Research* 32.1 (2021), pp. 1–6.
- [154] Emily Berndt et al. “Towards the Development of Real-time Normalized Burn Ratio (NBR) and delta NBR Imagery from GOES-16/17 and S-NPP”. In: *National Weather Association (NWA) Annual Meeting*. MSFC-E-DAA-TN73176. 2019.
- [155] Valentino Kevin Sitanayah Que et al. “Analisis Perbedaan Indeks Vegetasi Normalized Difference Vegetation Index (NDVI) dan Normalized Burn Ratio (NBR) Kabupaten Pelalawan Menggunakan Citra Satelit Landsat 8”. In: *Indonesian Journal of Computing and Modeling* 2.1 (2019), pp. 1–7.
- [156] Muhammad Ichsana Ali et al. “Monitoring the built-up area transformation using urban index and normalized difference built-up index analysis”. In: *International Journal of Engineering Transactions B: Applications* 32.5 (2019), pp. 647–653.
- [157] W Prasomsup et al. “Extraction technic for built-up area classification in Landsat 8 imagery”. In: *International Journal of Environmental Science and Development* 11.1 (2020), pp. 15–20.
- [158] Yuanmao Zheng et al. “An improved approach for monitoring urban built-up areas by combining NPP-VIIRS nighttime light, NDVI, NDWI, and NDBI”. In: *Journal of Cleaner Production* 328 (2021), p. 129488.
- [159] Sinno Jialin Pan and Qiang Yang. “A survey on transfer learning”. In: *IEEE Transactions on knowledge and data engineering* 22.10 (2010), pp. 1345–1359.
- [160] Shai Ben-David et al. “Analysis of Representations for Domain Adaptation”. In: *Advances in Neural Information Processing Systems* 19 (2007), pp. 137–144.
- [161] Ross Girshick et al. “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2014), pp. 580–587.
- [162] Sebastian Ruder. “An Overview of Multi-Task Learning in Deep Neural Networks”. In: *arXiv preprint arXiv:1706.05098* (2017).
- [163] Yaroslav Ganin et al. “Domain-Adversarial Training of Neural Networks”. In: *Journal of Machine Learning Research* 17.59 (2016), pp. 1–35.
- [164] Eric Tzeng et al. “Adversarial Discriminative Domain Adaptation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 7167–7176.

References

- [165] J. Smith and R. Doe. “Comparison of Satellite and UAV Remote Sensing for Wildfire Detection”. In: *Journal of Environmental Monitoring* 45.3 (2020), pp. 567–589.
- [166] A. Jones and S. Brown. “Atmospheric Effects on Fire Smoke Detection from Satellite Imagery”. In: *Remote Sensing of Environment* 123.7 (2019), pp. 1234–1250.
- [167] P. Williams and M. Green. “Modeling Fire Smoke Dispersion and Its Detection Using UAV Platforms”. In: *Geospatial Modeling & Fire Ecology* 67.1 (2018), pp. 89–102.
- [168] L. Anderson and C. Lee. “Impact of Observational Distance and Angle on Smoke Detection in Satellite and UAV Imagery”. In: *Journal of Remote Sensing Applications* 38.5 (2021), pp. 432–450.
- [169] Dinesh Komarasamy et al. “Review for Detecting Smoke and Fire in Forest using Different Technologies”. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 993. 1. IOP Publishing, 2020, p. 012056.
- [170] Fei Yuan, Wenqing Zhang, and Xudong Zheng. “Automatic smoke detection in video sequences using dynamic texture”. In: *Pattern Recognition* 42.9 (2008), pp. 2195–2206.
- [171] Zheng Liu and Narendra Ahuja. “Vision-based fire detection”. In: *IEEE Transactions on Image Processing* 16.5 (2004), pp. 1576–1589.
- [172] Fei Yuan. “A fast accumulative motion orientation model based on integral image for video smoke detection”. In: *Pattern Recognition* 41.1 (2008), pp. 39–53.
- [173] Taner Celik et al. “Fire detection using statistical color model in video sequences”. In: *Journal of Visual Communication and Image Representation* 18.2 (2007), pp. 176–185.
- [174] Shiping Ye et al. “An effective algorithm to detect both smoke and flame using color and wavelet analysis”. In: *Pattern Recognition and Image Analysis* 27 (2017), pp. 131–138.
- [175] Panagiotis Barmpoutis et al. “A review on early forest fire detection systems using optical remote sensing”. In: *Sensors* 20.22 (2020), p. 6442.
- [176] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine Learning* 20.3 (1995), pp. 273–297.

- [177] Josef Sivic and Andrew Zisserman. “Video Google: A text retrieval approach to object matching in videos”. In: *Proceedings Ninth IEEE International Conference on Computer Vision*. IEEE, 2003, pp. 1470–1477.
- [178] J. B. MacQueen. “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1 (1967), pp. 281–297.
- [179] Yoav Freund and Robert E. Schapire. “Experiments with a new boosting algorithm”. In: *Proceedings of the Thirteenth International Conference on Machine Learning (ICML)*. 1996, pp. 148–156.
- [180] Jayavardhana Gubbi, Slaven Marusic, and Marimuthu Palaniswami. “Smoke detection in video using wavelets and support vector machines”. In: *Fire Safety Journal* 44.8 (2009), pp. 1110–1115.
- [181] Xiangyang Zhang, Li Zhang, and Jun Liu. “A novel video-based smoke detection method using Bag of Features model and support vector machine”. In: *IEEE Transactions on Industrial Informatics* 12.6 (2016), pp. 2336–2346.
- [182] Xuehui Wu, Xiaobo Lu, and Henry Leung. “A video based fire smoke detection using robust AdaBoost”. In: *Sensors* 18.11 (2018), p. 3780.
- [183] Meenu Ajith and Manel Martínez-Ramón. “Unsupervised segmentation of fire and smoke from infra-red videos”. In: *IEEE Access* 7 (2019), pp. 182381–182394.
- [184] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the Royal Statistical Society* 39.1 (1977), pp. 1–38.
- [185] Stan Z. Li. “A Markov random field model for segmentation of multispectral remote sensing images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.6 (2014), pp. 3603–3615.
- [186] S. Ioffe and D. A. Forsyth. “Probabilistic models for segmentation of 3D scenes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.9 (2009), pp. 1576–1589.
- [187] Timothy J. Ross. *Fuzzy logic with engineering applications*. Wiley, 1974.
- [188] R. E. Kalman. “A new approach to linear filtering and prediction problems”. In: *Journal of Basic Engineering* 82 (1960), pp. 35–45.
- [189] Chi Yuan, Zhixiang Liu, and Youmin Zhang. “Learning-based smoke detection for unmanned aerial vehicles applied to forest fire surveillance”. In: *Journal of Intelligent & Robotic Systems* 93 (2019), pp. 337–349.

References

- [190] Yu Chunyu et al. “Video fire smoke detection using motion and color features”. In: *Fire technology* 46 (2010), pp. 651–663.
- [191] Yu Chunyu et al. “Video smoke recognition based on optical flow”. In: *2010 2nd International Conference on Advanced Computer Control*. Vol. 2. IEEE. 2010, pp. 16–21.
- [192] Bruce D Lucas and Takeo Kanade. “An iterative image registration technique with an application to stereo vision”. In: *IJCAI’81: 7th international joint conference on Artificial intelligence*. Vol. 2. 1981, pp. 674–679.
- [193] Feiniu Yuan et al. “Deep smoke segmentation”. In: *Neurocomputing* 357 (2019), pp. 248–260.
- [194] Shuhong Cheng, Jiyong Ma, and Shijun Zhang. “Smoke detection and trend prediction method based on Deeplabv3+ and generative adversarial network”. In: *Journal of Electronic Imaging* 28.3 (2019), pp. 033006–033006.
- [195] Liang-Chieh Chen et al. “Encoder-decoder with atrous separable convolution for semantic image segmentation”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 801–818.
- [196] Philipp Krähenbühl and Vladlen Koltun. “Efficient inference in fully connected crfs with gaussian edge potentials”. In: *Advances in neural information processing systems* 24 (2011).
- [197] Liang-Chieh Chen et al. “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017), pp. 834–848.
- [198] Panagiotis Barmpoutis et al. “Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures”. In: *Remote Sensing* 12.19 (2020), p. 3177.
- [199] Süleyman Aslan et al. “Early wildfire smoke detection based on motion-based geometric image transformation and deep convolutional generative adversarial networks”. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 8315–8319.
- [200] Feiniu Yuan et al. “A Lightweight Network for Smoke Semantic Segmentation”. In: *Pattern Recognition* 137 (May 2023), p. 109289. DOI: 10.1016/j.patcog.2022.109289.

- [201] Feiniu Yuan, Lin Zhang, and Xue Xia. “Smoke Semantic Segmentation with Multi-Scale Residual Paths and Weighted Middle Surveillances”. In: *Multimedia Tools and Applications* (Oct. 2023). ISSN: 1573-7721. DOI: 10.1007/s11042-023-17260-2.
- [202] Sebastien Frizzi et al. “Convolutional neural network for video fire and smoke detection”. In: *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2016, pp. 877–882.
- [203] Zhijian Yin et al. “A deep normalization and convolutional neural network for image smoke detection”. In: *Ieee Access* 5 (2017), pp. 18429–18438.
- [204] Yanhong Chen et al. “A UAV-based forest fire detection algorithm using convolutional neural network”. In: *2018 37th Chinese Control Conference (CCC)*. IEEE. 2018, pp. 10305–10310.
- [205] Zhentian Jiao et al. “A deep learning based forest fire detection approach using UAV and YOLOv3”. In: *2019 1st International Conference on Industrial Artificial Intelligence (IAI)*. IEEE. 2019, pp. 1–5.
- [206] Joseph Redmon and Ali Farhadi. “Yolov3: An incremental improvement”. In: *arXiv preprint arXiv:1804.02767* (2018).
- [207] Lei Zeng et al. “Fast smoke and flame detection based on lightweight deep neural network”. In: *2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. Vol. 1. IEEE. 2020, pp. 232–235.
- [208] Guosheng Lin et al. “Refinenet: Multi-path refinement networks for high-resolution semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1925–1934.
- [209] Mark Sandler et al. “Mobilenetv2: Inverted residuals and linear bottlenecks”. In: *Proceedings of the IEEE Conference on Computer Vision and Ppattern Recognition*. 2018, pp. 4510–4520.
- [210] FM Anim Hossain, Youmin M Zhang, and Masuda Akter Tonima. “Forest fire flame and smoke detection from UAV-captured images using fire-specific color features and multi-color space local binary pattern”. In: *Journal of Unmanned Vehicle Systems* 8.4 (2020), pp. 285–309.
- [211] Timo Ojala, Matti Pietikainen, and David Harwood. “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions”. In: *Proceedings of 12th international conference on pattern recognition*. Vol. 1. IEEE. 1994, pp. 582–585.

References

- [212] Joseph Redmon et al. “You only look once: Unified, real-time object detection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 779–788.
- [213] Juan Terven and Diana Cordova-Esparza. “A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond”. In: *arXiv preprint arXiv:2304.00501* (2023).
- [214] Rabia Bayraktar, Ahmet Emin Yetkin, and Merve Elmas Erdem. “Real Time Fire and Smoke Detection on RGB Images Using YOLOv7 with Optimized Hyper-Parameters”. In: *2023 IEEE East-West Design & Test Symposium (EWDTS)*. 2023, pp. 1–6. DOI: 10.1109/EWDTS59469.2023.10297085.
- [215] Zhan Dou et al. “An Improved YOLOv5s Fire Detection Model”. In: *Fire Technology* (Oct. 2023). ISSN: 1572-8099. DOI: 10.1007/s10694-023-01492-7.
- [216] Bowei Xiao and Chunman Yan. “A Lightweight Global Awareness Deep Network Model for Flame and Smoke Detection”. In: *Optoelectronics Letters* 19.10 (Oct. 2023), pp. 614–622. ISSN: 1993-5013. DOI: 10.1007/s11801-023-3041-x.
- [217] Yazan Al-Smadi et al. “Early Wildfire Smoke Detection Using Different YOLO Models”. In: *Machines* 11.246 (2023). ISSN: 2075-1702. DOI: 10.3390/machines11020246.
- [218] Joseph Redmon and Ali Farhadi. “YOLO9000: better, faster, stronger”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7263–7271.
- [219] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. “YOLOv4: Optimal speed and accuracy of object detection”. In: *arXiv preprint arXiv:2004.10934*. 2020.
- [220] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. “CSPNet: A new backbone that can enhance learning capability of CNN”. In: *arXiv preprint arXiv:1911.11929* (2020).
- [221] Glenn Jocher. *YOLOv5: An improved real-time object detection model*. <https://github.com/ultralytics/yolov5>. 2020.
- [222] Cheng Li, Yuan Lin, Weiqiang Liu, et al. “YOLOv6: A single-stage object detection framework for industrial applications”. In: *arXiv preprint arXiv:2209.02976* (2022).
- [223] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”. In: *arXiv preprint arXiv:2207.02696* (2022).

- [224] Glenn Jocher et al. *YOLOv8: A state-of-the-art real-time object detection model*. <https://github.com/ultralytics/yolov8>. 2023.
- [225] Simon Philipp Hohberg. “Wildfire smoke detection using convolutional neural networks”. PhD thesis. 2015.
- [226] Shuiwang Ji et al. “3D convolutional neural networks for human action recognition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 35. 1. 2013, pp. 221–231.
- [227] Qi-xing Zhang et al. “Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images”. In: *Procedia Engineering* 211 (2018), pp. 441–446.
- [228] Shaoqing Ren et al. “Faster R-CNN: Towards real-time object detection with region proposal networks”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 91–99.
- [229] Y-S Chung and HV Le. “Detection of forest-fire smoke plumes by satellite imagery”. In: *Atmospheric Environment (1967)* 18.10 (1984), pp. 2143–2151.
- [230] YS Chung and HS Kim. “Satellite monitoring of forest fires and associated smoke plumes occurring in Korea”. In: *Air Quality, Atmosphere & Health* 1.2 (2008), pp. 111–118.
- [231] Izumi NAGATANI and Jun-ichi KUDOH. “A False-color Composite Method for Wildfire Smoke Plume Identification Using MODIS Data”. In: *Journal of the remote sensing society of Japan* 33.1 (2013), pp. 38–47.
- [232] Zhanqing Li et al. “Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to AVHRR imagery”. In: *IEEE Transactions on geoscience and remote sensing* 39.9 (2001), pp. 1859–1870.
- [233] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. “A Comparative Study of Texture Measures for Terrain Classification”. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-6.4 (1976), pp. 269–285. DOI: 10.1109/TSMC.1976.4309459.
- [234] Bryan A Baum and Qing Trepte. “A grouped threshold approach for scene identification in AVHRR imagery”. In: *Journal of Atmospheric and Oceanic Technology* 16.6 (1999), pp. 793–800.
- [235] Bipasha Paul Shukla and PK Pal. “Automatic smoke detection using satellite imagery: preparatory to smoke detection from Insat-3D”. In: *International Journal of Remote Sensing* 30.1 (2009), pp. 9–22.

References

- [236] N Chrysoulakis and C Cartalis. “A new approach for the detection of major fires caused by industrial accidents, using NOAA/AVHRR imagery”. In: *International Journal of Remote Sensing* 21.8 (2000), pp. 1743–1748.
- [237] N Chrysoulakis and C Cartalis. “A new algorithm for the detection of plumes caused by industrial accidents, based on NOAA/AVHRR imagery”. In: *International Journal of Remote Sensing* 24.17 (2003), pp. 3353–3368.
- [238] Xiaolian Li et al. “Forest fire smoke detection using back-propagation neural network based on MODIS data”. In: *Remote Sensing* 7.4 (2015), pp. 4473–4498.
- [239] H Ismanto, H Hartono, et al. “Classification tree analysis (Gini-Index) smoke detection using Himawari.8 satellite data over Sumatera-Borneo maritime continent South East Asia”. In: *IOP Conference Series: Earth and Environmental Science*. Vol. 256. 1. IOP Publishing. 2019, p. 012043.
- [240] Xiaolian Li et al. “Automatic smoke detection in modis satellite data based on k-means clustering and fisher linear discrimination”. In: *Photogrammetric Engineering & Remote Sensing* 80.10 (2014), pp. 971–982.
- [241] R. A. Fisher. “The use of multiple measurements in taxonomic problems”. In: *Annals of Eugenics* 7.2 (1936), pp. 179–188.
- [242] Jiayun Yao et al. “Predicting the minimum height of forest fire smoke within the atmosphere using machine learning and data from the CALIPSO satellite”. In: *Remote sensing of environment* 206 (2018), pp. 98–106.
- [243] Leo Breiman. “Random forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.
- [244] Xihao Li et al. “Forest Fire Smoke Detection Research Based on the Random Forest Algorithm and Sub-Pixel Mapping Method”. In: *Forests* 14.3 (2023). DOI: 10.3390/f14030485.
- [245] Yehan Sun et al. “A Satellite Imagery Smoke Detection Framework Based on the Mahalanobis Distance for Early Fire Identification and Positioning”. In: *International Journal of Applied Earth Observation and Geoinformation* 118 (Apr. 2023), p. 103257. ISSN: 15698432. DOI: 10.1016/j.jag.2023.103257. (Visited on 11/20/2023).
- [246] P. C. Mahalanobis. “On the generalized distance in statistics”. In: *Proceedings of the National Institute of Sciences of India* 2.1 (1936), pp. 49–55.
- [247] Haotian Liang et al. “Super-Resolution Reconstruction of Remote Sensing Data Based on Multiple Satellite Sources for Forest Fire Smoke Segmentation”. In: *Remote Sensing* 15.4180 (2023). ISSN: 2072-4292. DOI: 10.3390/rs15174180.

- [248] Junjie Wang et al. “Remote sensing scene classification via multi-stage self-guided separation network”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2023).
- [249] Junjie Wang et al. “Large kernel sparse convNet weighted by multi-frequency attention for remote sensing scene understanding”. In: *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), pp. 1–12.
- [250] A Rahman et al. “Combining Landsat-8 spectral bands with ancillary variables for land cover classification in mountainous terrains of northern Pakistan”. In: *Journal of Mountain Science* 18.9 (2021), pp. 2388–2401.
- [251] MS Boori et al. “Supervised and unsupervised classification for obtaining land use/cover classes from hyperspectral and multispectral imagery”. In: *Sixth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2018)*. Vol. 10773. International Society for Optics and Photonics. 2018, p. 107730L.
- [252] AK Rai et al. “Landsat 8 OLI satellite image classification using convolutional neural network”. In: *Procedia Computer Science* 167 (2020), pp. 987–993.
- [253] Yi Yang and Shawn Newsam. “Bag-of-visual-words and spatial extensions for land-use classification”. In: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM. 2010, pp. 270–279.
- [254] Qingquan Zou et al. “Deep learning based feature selection for remote sensing scene classification”. In: *IEEE Geoscience and Remote Sensing Letters* 12.11 (2015), pp. 2321–2325.
- [255] Suliman Basaeed et al. “Lightweight convolutional neural network for remote sensing image classification”. In: *Journal of Computer Networks and Communications* 2016 (2016), pp. 1–9.
- [256] Fei Wang et al. “Residual attention network for image classification”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3156–3164.
- [257] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *ArXiv Preprint arXiv:1409.1556* (2014).
- [258] Donggeun Yoo et al. “Attentionnet: Aggregating weak directions for accurate object detection”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2659–2667.

References

- [259] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. “Neural machine translation by jointly learning to align and translate”. In: *arXiv preprint arXiv:1409.0473* (2014).
- [260] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [261] Ke Sun et al. “Deep high-resolution representation learning for visual recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)*, pp. 5199–5208.
- [262] Wen Liao et al. “Remote sensing image scene classification using bag of convolutional features”. In: *IEEE Geoscience and Remote Sensing Letters* 16.3 (2019), pp. 394–398.
- [263] Jinxi Guo et al. “Attention Based CLDNNs for Short-Duration Acoustic Scene Classification.” In: *Interspeech*. 2017, pp. 469–473.
- [264] Qi Wang et al. “Scene classification with recurrent attention of VHR remote sensing images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.2 (2018), pp. 1155–1167.
- [265] Wei Tong et al. “Channel-attention-based DenseNet network for remote sensing image scene classification”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 4121–4132.
- [266] Zhicheng Zhao et al. “Remote sensing image scene classification based on an enhanced attention module”. In: *IEEE Geoscience and Remote Sensing Letters* 18.11 (2020), pp. 1926–1930.
- [267] Xu Tang et al. “Attention consistent network for remote sensing scene classification”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 2030–2045.
- [268] Jia Deng et al. “ImageNet: A Large-Scale Hierarchical Image Database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Ieee. 2009, pp. 248–255.
- [269] Alex Krizhevsky and Geoffrey Hinton. *Learning multiple layers of features from tiny images*. Tech. rep. University of Toronto, 2009.
- [270] Mark Everingham et al. “The Pascal Visual Object Classes (VOC) challenge”. In: *International journal of computer vision* 88.2 (2010), pp. 303–338.

- [271] Deepank Verma, Arnab Jana, and Krithi Ramamritham. “Transfer learning approach to map urban slums using high and medium resolution satellite imagery”. In: *Habitat International* 88 (2019), p. 101981.
- [272] Amsa Shabbir et al. “Satellite and scene image classification based on transfer learning and fine tuning of ResNet50”. In: *Mathematical Problems in Engineering* 2021 (2021), pp. 1–18.
- [273] Artur Nowakowski et al. “Crop type mapping by using transfer learning”. In: *International Journal of Applied Earth Observation and Geoinformation* 98 (2021), p. 102313.
- [274] Usha Patel et al. “Crop type classification with hyperspectral images using deep learning: a transfer learning approach”. In: *Modeling Earth Systems and Environment* 9.2 (2023), pp. 1977–1987.
- [275] Haiyu Wu et al. “Transfer learning for wildfire identification in UAV imagery”. In: *2020 54th Annual Conference on Information Sciences and Systems (CISS)*. IEEE. 2020, pp. 1–6.
- [276] Andrew G Howard et al. “Mobilenets: Efficient convolutional neural networks for mobile vision applications”. In: *ArXiv Preprint arXiv:1704.04861* (2017).
- [277] Tushar Agrawal, Merin Meleet, et al. “Classification of natural disaster using satellite & drone images with CNN using transfer learning”. In: *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. IEEE. 2021, pp. 1–5.
- [278] TR Vijaya Lakshmi et al. “Classification of multi-spectral data with fine-tuning variants of representative models”. In: *Multimedia Tools and Applications* (2023), pp. 1–23.
- [279] Borja Espejo-Garcia et al. “Towards weeds identification assistance through transfer learning”. In: *Computers and Electronics in Agriculture* 171 (2020), p. 105306.
- [280] Krishna Karthik Gadiraju and Ranga Raju Vatsavai. “Remote sensing based crop type classification via deep transfer learning”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), pp. 4699–4712.
- [281] Yuvraj Sharma. “Improving transfer learning for use in multi-spectral data”. In: (2020).

References

- [282] Rong Xiao, Yuze Wang, and Chao Tao. “Fine-grained road scene understanding from aerial images based on semisupervised semantic segmentation networks”. In: *IEEE Geoscience and Remote Sensing Letters* 19 (2021), pp. 1–5.
- [283] Yuze Wang et al. “Cross-sensor remote-sensing images scene understanding based on transfer learning between heterogeneous networks”. In: *IEEE Geoscience and Remote Sensing Letters* 19 (2021), pp. 1–5.
- [284] Jianping Gou et al. “Knowledge distillation: A survey”. In: *International Journal of Computer Vision* 129 (2021), pp. 1789–1819.
- [285] Chao Tao et al. “A general transitive transfer learning framework for cross-optical sensor remote sensing image scene understanding”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2023).
- [286] Yuvraj Sharma and Robert Ross. “Less is more when applying transfer learning to multi-spectral data.” In: *AICS*. 2020, pp. 301–312.
- [287] Wang Jian et al. “Transfer-learning-based cloud detection for Zhuhai-1 satellite hyperspectral imagery”. In: *Frontiers in Environmental Science* 10 (2022), p. 1039249.
- [288] Xin He, Yushi Chen, and Pedram Ghamisi. “Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.5 (2019), pp. 3246–3263.
- [289] Utkarsh Singhal et al. “Multi-spectral image classification with ultra-lean complex-valued models”. In: *arXiv preprint arXiv:2211.11797* (2022).
- [290] Yezhen Cong et al. “Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 197–211.
- [291] Tom X-P Zhao, Steve Ackerman, and Wei Guo. “Dust and smoke detection for multi-channel imagers”. In: *Remote Sensing* 2.10 (2010), pp. 2347–2368.
- [292] Fuqin Li et al. “A physics-based atmospheric and BRDF correction for Landsat data over mountainous terrain”. In: *Remote Sensing of Environment* 124 (2012), pp. 756–770.
- [293] *DEA Sandbox*. <https://docs.dea.ga.gov.au/setup/Sandbox/sandbox.html>.
- [294] *Last Bushfire and Prescribed Burn Boundaries*. <https://data.sa.gov.au/data/dataset/fire-history>.

- [295] *NPWS Fire History - Wildfires and Prescribed Burns*. <https://data.nsw.gov.au/search/dataset/ds-nsw-ckan-82b910c2-a704-4344-b897-f4a31221ae5a/distribution/dist-nsw-ckan-5f4fc2e9-4879-47a0-a805-f652c4cf1928/details?q=fire>.
- [296] Christian Szegedy et al. “Going deeper with convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.
- [297] Christian Szegedy et al. “Inception-v4, inception-resnet and the impact of residual connections on learning”. In: *Thirty-first AAAI conference on artificial intelligence*. 2017.
- [298] François Chollet. “Xception: Deep learning with depthwise separable convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1251–1258.
- [299] Gong Cheng et al. “Remote sensing image scene classification using bag of convolutional features”. In: *IEEE Geoscience and Remote Sensing Letters* 14.10 (2017), pp. 1735–1739.
- [300] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *ArXiv Preprint arXiv:1412.6980* (2014).
- [301] Junjie Wang et al. “Hyperspectral and SAR image classification via multiscale interactive fusion network”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [302] Junjie Wang et al. “Representation-enhanced status replay network for multi-source remote-sensing image classification”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2023).
- [303] Min Chen et al. “Augmented associative learning-based domain adaptation for classification of hyperspectral remote sensing images”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 6236–6248. DOI: 10.1109/JSTARS.2020.3030304.