Synergizing Deep Learning and Representation Learning for Hyperspectral Image Analysis

Ph.D. Thesis

By
Aditya Dixit



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE

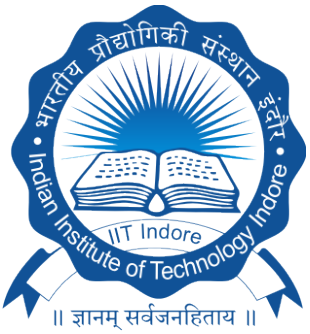
June 2026

Synergizing Deep Learning and Representation Learning for Hyperspectral Image Analysis

A Thesis

Submitted in partial fulfillment of the
requirements for the award of the degrees
of
Doctor of Philosophy

by
Aditya Dixit



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE

June 2026

INDIAN INSTITUTE OF TECHNOLOGY INDORE

I hereby certify that the work which is being presented in the thesis entitled **Synergizing Deep Learning and Representation Learning for Hyperspectral Image Analysis** in the partial fulfilment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **Department of Computer Science & Engineering**, **Indian Institute of Technology Indore,** is an authentic record of my own work carried out during the time period from **August 2022** to **June 2026** under the supervision of **Dr. Puneet Gupta**, **Associate Professor, Computer Science and Engineering.**

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Aditya Dixit 01-07-2026

Signature of the student with date

(Aditya Dixit)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

01-07-2026

Signature of Thesis Supervisor with date

(Dr. Puneet Gupta)

Aditya Dixit has successfully given his Ph.D. Oral Examination held on 19-08-2026

19-08-2026

Signature of Thesis Supervisor with date

(Dr. Puneet Gupta)

Acknowledgments

First and foremost, I bow my head in humble gratitude to Shri Radha Rani and Lord Krishna for their infinite blessings, guidance, and grace throughout this journey. Their divine presence has given me the strength to overcome challenges, the patience to persevere, and the faith to pursue my goals with dedication. Every achievement presented in this thesis is a reflection of their blessings, and the words “Radhe Radhe" and “Raam Raam" have remained a constant source of inspiration and inner peace, reminding me that sincere effort, coupled with divine grace, can make even the most difficult paths achievable.

I express my deepest respect and sincere gratitude to my supervisor, Dr. Puneet Gupta, for his invaluable guidance, constant encouragement, and unwavering support throughout my Ph.D. journey. His mentorship has been one of the most significant influences on both my academic and personal growth. Under his guidance, I have learned not only how to conduct meaningful research but also how to think critically, write scientific papers, present ideas effectively, and approach challenging problems with confidence. His insightful discussions, constructive suggestions, and high academic standards have continually motivated me to improve my work. His patience, trust, and confidence in my abilities encouraged me to explore challenging research directions and strive for excellence at every stage of this thesis. I have always admired the clarity and determination with which he approaches complex research problems, and I have learned invaluable lessons from his dedication and professionalism. I feel truly fortunate to have had the opportunity to work under such an inspiring and supportive mentor.

I sincerely thank the members of my Postgraduate Student Progress Committee (PSPC), Prof. Surya Prakash and Prof. Ruchi Sharma, for their insightful suggestions, constructive feedback, and continuous evaluation of my research work. I am also grateful to the former Head of the Department, Prof. Somnath Dey, and the current Head of the Department, Dr. Ranveer Singh, Department of Computer Science and Engineering, for their support and encouragement. I further extend my heartfelt thanks to Mr. Shailendra Verma, Mrs. Ujavala Gorakh Langhi, Mr. Jagat Singh, and Mr. Lalit Jain for their academic assistance and for always being supportive with administrative and official matters.

I am deeply thankful to the Indian Institute of Technology Indore for offering a stimulating academic environment and excellent research facilities. I express my sincere gratitude to Prof. Suhas Joshi, Director, IIT Indore, for fostering a culture of curiosity, academic independence, and research excellence.

I feel privileged to be a part of the Deep Intelligence Lab and to have the support of its wonderful members. I sincerely appreciate my seniors, Dr. Lokendra Birla, Dr. Sneha Shukla, and Dr. Anup Kumar Gupta, along with my colleagues-turned-friends Rupesh Kumar, Anirban Nath, Trishna Saikia, Ashutosh Dhamaniya, Aravind Ramagiri, Rajesh Kumar, Vipendra Singh, Apurv Anand, Adit Srivastava, Vaidehi Agarwal, Debashish Utpal Biswas, Rahul Kumar Rathore, Utkarsh Singh Rathore, and Nischit Hosamani, for their encouragement, collaboration, unwavering support, and above all, friendship. I am especially grateful to Dr. Anup Kumar Gupta for his invaluable guidance, insightful advice, and constant encouragement throughout my research journey.

I also express my heartfelt gratitude to my seniors, Dr. Anuj Rai, Dr. Rupendra Pratap Singh, and Dr. Aditya Anshul for guiding me as mentors and elder brothers. I feel fortunate to have a wonderful circle of friends whose companionship, encouragement, and unwavering support made this demanding journey both meaning and memorable. My heartfelt thanks go to Priyanka Choudhary, Mayank Ramnani, Trishna Saikia, and many others for standing by me through every high and low. I am equally grateful to Anoop KR, Bharath, Deepak Prasad, Mithun Singh, Priyanshu Singh, and Atharva Tendulkar for their warmth and friendship.

Above all, I am profoundly grateful to my parents, Mr. Kaushal Kishor Dixit and Mrs. Baby Dixit, whose unwavering belief in me has been my greatest source of strength. Their resilience, hard work, integrity, and independent spirit have constantly inspired me to dream bigger and work harder. Their unconditional love, endless prayers, and countless sacrifices have been the foundation of my PhD journey, providing me with the courage and determination to persevere through every challenge. I am equally grateful to my beloved brother, Mr. Ayush Dixit, for encouraging me at every step and for his constant love and support throughout my Ph.D. jouney.

Finally, I extend my sincere appreciation to everyone who, directly or indirectly, contributed to this research and supported me during my doctoral journey. Your encouragement,

kindness, and goodwill have played an invaluable role in bringing this work to fruition.

Aditya Dixit

Dedicated to My Family, Friends
and Respected Teachers

Abstract

Hyperspectral imaging has emerged as an important remote sensing technology due to its ability to capture detailed spectral information across hundreds of contiguous wavelength bands. This rich spatial-spectral information has enabled a wide range of real-world applications, including environmental monitoring, precision agriculture, mineral exploration, land-cover mapping, and the assessment of greenhouse gas emissions. However, in practice, hyperspectral data acquisition is often affected by sensor limitations, atmospheric disturbances, illumination changes, and transmission errors. These factors introduce different types of noise that degrade image quality and make subsequent analysis more challenging. In addition, methane (CH_4) plume detection and segmentation is difficult due to weak spectral signatures, strong inter-band redundancy, complex backgrounds, and limited annotated data. These issues motivate the need for learning-based approaches that can improve hyperspectral image denoising (HSID) and CH_4 monitoring performance. This work focuses on developing deep learning (DL) frameworks for HSID, CH_4 plume segmentation, and spatial-spectral representation (SSR) learning. The main objective is to improve hyperspectral image (HSI) analysis by making better use of spatial, spectral, contextual, and frequency-domain information. The contributions are organized into three main directions.

The first contribution introduces `IDEATION` (hybr**I**d 3**D** cnn-transform**E**r **A**nd spa**T**ial-spectral attent**I**on framework for hyperspectral image den**O**isi**N**g), a hybrid framework for HSID. The proposed model integrates 3D convolutional neural networks (CNN), 3D Transformer-based feature extraction, and 3D U-Net architecture to simultaneously capture local spatial-spectral features and long-range contextual dependencies. A spatial-spectral transformer (SST) block is designed to sequentially model spatial, spectral, and joint spatial-spectral relationships, while a Global Feature Attention (GFA) module captures correlations both within individual HSIs and across the training dataset. By combining convolutional and attention-based learning, `IDEATION` effectively suppresses complex and mixed noise while preserving spectral and structural information, leading to improved denoising performance across multiple benchmark datasets.

The second contribution presents `ALKANE` (**A**ttention-based **L**earning and **K**nowledge

Aggregation **N**etwork with **E**nhanced skip connections), an attention-based framework for CH_4 plume segmentation from hyperspectral imagery. The model integrates spatial and spatial-frequency representations within a unified encoder-decoder structure to improve plume localization and boundary delineation. A frequency-aware Transformer is proposed to jointly analyze magnitude and phase information, enabling the capture of weak and diffuse CH_4 plume regions. Additionally, an Efficient Global Channel Attention (EGCA) mechanism is used in the decoder to retain informative features while reducing redundant spectral responses. Experiments on multiple hyperspectral CH_4 datasets acquired by different sensors show that `ALKANE` outperforms existing methods in all cases.

The third contribution introduces `AUGMENT` (supervised contr**A**stive learning and inpainting for hyperspectral denoising and methane pl**U**me se**GMENT**ation), a unified framework that combines supervised contrastive learning (SCL) with hyperspectral inpainting for denoising and CH_4 plume segmentation. The method generates multiple structured versions of HSI data using geometric transformations and masking-based inpainting, enabling the model to learn more meaningful SSR. By enforcing consistency across these transformed views, `AUGMENT` improves representation learning (RL), especially when training data is limited, and thereby improves performance in both denoising and segmentation tasks. It is also designed to integrate with existing CNN and Transformer-based HSI pipelines.

Together, these contributions form a complete framework for HSI analysis through improved denoising, CH_4 plume segmentation, and RL. By addressing challenges related to noise corruption, weak plume visibility, and limited annotated data, this work supports more effective use of DL methods in hyperspectral remote sensing applications.

List of Publications

(A) Publications from the PhD Thesis Work

A1. Journal Articles Published:

J1. **Aditya Dixit**, Anup Kumar Gupta, Puneet Gupta, Saurabh Srivastava, and Ankur Garg. "UNFOLD: 3D U-Net, 3D CNN and 3D Transformer based Hyperspectral Image Denoising." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-10, 2023, Art no. 5529710, DOI: 10.1109/TGRS.2023.3328922, (**IF: 9.4**).

J2. **Aditya Dixit**, and Puneet Gupta. "FUMESNet: Exploring Frequency-based Transformer and Improving Skip Connection for Hyperspectral Methane Plume Segmentation." *IEEE Transactions on Instrumentation and Measurement*, vol. 75, pp. 1-11, 2026, Art no. 5004711, DOI: 10.1109/TIM.2026.3667330, (**IF: 7.0**).

J3. **Aditya Dixit**, Anup Kumar Gupta, Puneet Gupta, and Ankur Garg. "RECREATE: Supervised contrastive learning and inpainting based hyperspectral image denoising." *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.233, pp. 14-24, 2026, DOI: 10.1016/j.isprsjprs.2026.01.022, (**IF: 12.9**).

A2. Conference Proceedings Published:

C1. **Aditya Dixit**, Nischit Hosamani, Puneet Gupta, and Ankur Garg. "VISIONARY: Novel Spatial-Spectral Attention Mechanism for Hyperspectral Image Denoising." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2736-2745, 2025, DOI: 10.1109/WACV61041.2025.00271, (**Core A**).

(B) Other Publications during PhD

B1. Journal Articles Published:

J1. Rajesh Kumar, Abhay Bagla, **Aditya Dixit**, Adit Srivastava, Puneet Gupta, Vivek Gupta, and Jatin Kharbanda. "IRIS: Automated mobile microscope instrument for white blood cell classification" *Measurement*, vol. 283, 2026, Art no. 122096. DOI: 10.1016/j.measurement.2026.122096, (**IF: 6.1**).

Under Review/Revision:

J1. **Aditya Dixit**, Abhishek Christopher, Sunil Surendra K G, Puneet Gupta, and S. Esakkirajan. "GRADIENT: Frequency-Guided Transformer Framework for Circular Knitted Fabrics Defect Segmentation", *IEEE Transactions on Industrial Informatics* (2026), (under review)

J2. Pulkit Garg, **Aditya Dixit**, P C Uma Mahesh, Puneet Gupta, and Indrasen Singh. "METRICATE: Deep Learning-Based von Mises Stress Prediction using a Hybrid CNN–Transformer Architecture", *Engineering Applications of Artificial Intelligence* (2026), (under review)

J3. Devanshi Chhatbar, **Aditya Dixit**, Rajesh Kumar, Manju Purohit, and Puneet Gupta. "STAINER: Frequency-Aware Attention Learning for Virtual Staining of Unstained Bright-Field Histopathology Images", *ACM Transactions on Intelligent Systems and Technology* (2026), (under review)

J4. **Aditya Dixit**, NKK Sabaridasan, Puneet Gupta, and S. Esakkirajan. "AMOUNT: Semantic Segmentation of multiple Indian Coins by Synergizing CNN and Transformer", *IEEE Transactions on Human-Machine Systems* (2025), (Rebuttal submitted)

B2. Conference Proceedings Published:

C1. Hani Shanableh, Mohamed Barakat A. Gibril, Ahmed Mansour, **Aditya Dixit,** Rami Al-Ruzouq, Nezar Hammouri, Fouad Lamghari, Safa M. Ahmed, Ratiranjan Jena,

Tilal Mohamed, Mohammed Abdulraheem Almarzouqi, Nedal Salem Alafayfeh, Simon Zerisenay Ghebremeskel, "A Comparative Analysis of Deep Learning Methods for Ghaf Tree Detection and Segmentation from UAV-based Images." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 805-811, 2025, DOI: 10.5194/isprs-annals-X-G-2025-805-2025

Table of Contents

Abstract **i**

List of Figures **x**

List of Tables **xiv**

List of Abbreviations **xv**

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Thesis Contributions 3

1.2.1 Contribution 1: Hybrid 3D CNN-Transformer and Spatial–Spectral Attention Framework for Hyperspectral Image Denoising 3

1.2.2 Contribution 2: Frequency-Aware Transformer and Enhanced Skip Connections for Hyperspectral Methane Plume Segmentation . . . 5

1.2.3 Contribution 3: Supervised Contrastive Learning and Inpainting for Hyperspectral Denoising and Methane Plume Segmentation 7

1.3 Thesis Organization 8

2 Literature Survey **10**

2.1 Hyperspectral Image Denoising 10

2.1.1 Traditional Methods for Hyperspectral Image Denoising 10

2.1.2 CNN Methods for Hyperspectral Image Denoising 11

2.1.3 Transformer Methods for Hyperspectral Image Denoising 12

2.2 Methane Plume Segmentation 13
2.2.1 Traditional Methods for Methane Plume Segmentation 13
2.2.2 CNN-Based Methods for Methane Plume Segmentation 13
2.2.3 Transformer-Based Methods for Methane Plume Segmentation . . . 14
2.2.4 Frequency-based Segmentation Methods 14
2.3 Supervised Contrastive Learning and Inpainting-based Hyperspectral Image Denoising and Segmentation 16
2.3.1 Hyperspectral Image Denoising 16
2.3.2 Methane Plume Segmentation 17
2.3.3 Contrastive Learning 18
2.3.4 Inpainting 19

3 Hybrid 3D CNN–Transformer and Spatial–Spectral Attention Framework for Hyperspectral Image Denoising 21
3.1 Introduction 21
3.2 Proposed Methodology 23
3.2.1 Spatial-Spectral Cubic Transformer Block 26
3.2.2 Global Feature Attention 27
3.2.3 Training Loss Function 29
3.3 Experimental Settings 30
3.3.1 Dataset 30
3.3.2 Implementation Details 33
3.4 Experimental Results 33
3.4.1 Competing Methods 33
3.4.2 Simulated Experimental Results 34
3.4.3 Real-Data Experimental Results 36
3.4.4 Model Complexity 38
3.4.5 Ablation Study 38
3.5 Summary 40

4 Frequency-Aware Transformer and Enhanced Skip Connections for Hyperspectral Methane Plume Segmentation 42

4.1 Introduction . . . 42

4.2 Proposed Methodology . . . 44

4.2.1 Spatial Frequency Transformer (SFT) . . . 46

4.2.2 Adaptive Spatial-Frequency Integration (ASFI) . . . 48

4.2.3 Efficient Global Channel Attention (EGCA) . . . 50

4.3 Experimental Results . . . 51

4.3.1 Experimental Settings . . . 51

4.3.2 Comparative Analysis . . . 55

4.3.3 Model Complexity . . . 58

4.3.4 Ablation Study . . . 59

4.4 Summary . . . 60

5 Supervised Contrastive Learning and Inpainting for Hyperspectral Denoising and Methane Plume Segmentation 61

5.1 Introduction . . . 61

5.2 Proposed Methodology . . . 63

5.2.1 Contrastive Loss . . . 64

5.2.2 Data Fidelity Loss . . . 68

5.2.3 Total Loss . . . 69

5.3 Experimental Results . . . 69

5.3.1 Hyperspectral Image Denoising . . . 70

5.3.2 Hyperspectral Methane Plume Segmentation . . . 81

5.3.3 Comparative Analysis . . . 85

5.3.4 Model Complexity . . . 90

5.4 Ablation Studies . . . 91

5.5 Summary . . . 93

6 Conclusion and Future Work 95

List of Figures

1.1 Overview of the key contributions of the thesis. 3

3.1 Overview of our proposed method, IDEATION. 23

3.2 Illustration of the proposed SSCformer block incorporating the GFA module. The input feature representations are sequentially reordered, normalized, and processed through GFA modules, in which learnable weighting parameters (α_1 and α_2) govern feature aggregation. The GFA module employs multi-head attention to model spatial, spectral, and joint spatial-spectral relationships, thereby enhancing feature representation. 24

3.3 HSIs are divided into $16 \times 16 \times 16$ patches. The red patch is the focus, with green patches assessed for spectral correlation and orange patches for spatial correlation. Weights α_1 and α_2 combine spatial and spectral correlations respectively to enhance the red patch, which is then examined with blue patches for inter-domain correlations. 25

3.4 Visualization of SA vs GFA: In SA, attention is focused between patches within an image, X_i. In GFA, attention is focused between all patches and a common weight matrix, $\mathbf{W}$. This way, similarities among different image samples are captured by $\mathbf{W}$. 27

3.5 Visual denoising results for the ICVL dataset with Gaussian noise at a noise level of $\sigma = 50$. 31

3.6 Visual comparison of the denoised images for different cases in the ICVL dataset 33

3.7 Real denoised results on IP and PU datasets for bands (3, 103, 203) and (5, 13, 95), respectively. 34
3.8 Classification outcomes for IP and PU datasets. 36
3.9 Spectral curves of the ICVL dataset at coordinates (110, 110) under Gaussian noise (σ = 50). 37
3.10 Visual outcomes for real-world noise reduction on the Urban (1) and Real (2) Datasets. 37
3.11 Classification outcomes for Urban (1) dataset. 38

4.1 Overview of our proposed method ALKANE. 44
4.2 Architecture of the proposed SFT block. 47
4.3 Architecture of the proposed ASFI block 49
4.4 Visual comparison of methane plume segmentation results on strong plumes from the AVIRIS-NG dataset. 52
4.5 Visual comparison of methane plume segmentation results on weak plumes from the AVIRIS-NG dataset. 52
4.6 Visual comparison of methane plume segmentation results on strong plumes from the WV-3 dataset. 53
4.7 Visual comparison of methane plume segmentation results on weak plumes from the WV-3 dataset. 54
4.8 Visual comparison of methane plume segmentation results on strong plumes from the EMIT dataset. 56
4.9 Visual comparison of methane plume segmentation results on weak plumes from the EMIT dataset. 57

5.1 Overview of the proposed AUGMENT framework. 63
5.2 Visual comparison of denoising results on the ICVL dataset under Gaussian noise with σ = 50. 73
5.3 Comparison of reconstructed spectral signatures obtained by AUGMENT-integrated baseline methods on the ICVL dataset under Gaussian noise (σ = 50) at pixel location (120, 120). 75

5.4 Visual comparison of denoising results on the ICVL dataset under different complex noise. 76
5.5 Residual error maps of baseline and AUGMENT-enhanced denoising methods on the ICVL dataset under Gaussian noise with $\sigma = 50$. 77
5.6 False-color visualization of denoising results produced by baseline and AUGMENT-enhanced methods on the ICVL dataset under Gaussian noise ($\sigma = 50$), using bands 4, 11, and 19 for RGB composition. 78
5.7 Visual comparison of denoising results on the real-world IP and PU datasets. The displayed bands correspond to (2, 101, 205) for IP and (3, 11, 98) for PU. 79
5.8 Comparison of SVM-based classification outcomes on the PU and IP datasets, highlighting the impact of AUGMENT on downstream HSI analysis. 80
5.9 Visual comparison of methane plume segmentation results on strong plumes from the AVIRIS-NG dataset. 84
5.10 Visual comparison of methane plume segmentation results on weak plumes from the AVIRIS-NG dataset. 84
5.11 Visual comparison of methane plume segmentation results on strong plumes from the WV-3 dataset. 84
5.12 Visual comparison of methane plume segmentation results on weak plumes from the WV-3 dataset. 87
5.13 Visual comparison of methane plume segmentation results on strong plumes from the EMIT dataset. 88
5.14 Visual comparison of methane plume segmentation results on weak plumes from the EMIT dataset. 88

List of Tables

3.1 Denoising Results of various methods under different Gaussian Noise Levels on the ICVL Dataset 31
3.2 Denoising performance of various methods across five complex noise scenarios on the ICVL Dataset 32
3.3 Classification performance on IP, PU, and Urban datasets before and after hyperspectral image denoising 35
3.4 Ablation study of different architectural variants of the proposed IDEATION framework on the ICVL dataset. 39
3.5 Ablation study of the proposed method with the SSCformer block on the ICVL dataset. 40

4.1 Quantitative comparison on the AVIRIS-NG dataset for strong and weak methane plumes 52
4.2 Quantitative comparison on the WV-3 dataset for strong and weak methane plumes 53
4.3 Quantitative comparison on the EMIT dataset for strong and weak methane plumes 55
4.4 Ablation study on EGCA kernel-size scaling parameters on the AVIRIS-NG dataset (Strong Plumes). 58
4.5 Ablation study of key components on AVIRIS-NG dataset (Strong Plumes). 59
4.6 Impact of frequency-domain components in ASFI (Strong Plumes). 59

5.1 Quantitative evaluation of denoising methods on the ICVL dataset in the presence of Gaussian noise. 71

5.2 Quantitative evaluation of denoising methods on the ICVL dataset under complex noise. 74
5.3 SVM classification accuracy before and after HSI denoising. 75
5.4 Sensitivity analysis of AUGMENT integrated with IDEATION on the ICVL dataset under Gaussian noise (σ = 50). Bold values indicate the parameter configuration used in the main experiments. 77
5.5 Ablation analysis of supervised contrastive learning within the proposed AUGMENT framework under Gaussian noise on the ICVL dataset. 80
5.6 Quantitative comparison of methane plume segmentation performance on the AVIRIS-NG dataset for strong and weak plumes. 83
5.7 Quantitative comparison of methane plume segmentation performance on the WV-3 dataset for strong and weak plumes. 86
5.8 Quantitative comparison of methane plume segmentation performance on the EMIT dataset for strong and weak plumes. 89
5.9 Ablation analysis of different AUGMENT components under Gaussian noise on the ICVL dataset. 90

List of Abbreviations

3D	Three-dimensional
ALKANE	**A**ttention-based **L**earning and **K**nowledge **A**ggregation **N**etwork with **E**nhanced skip connections
ASFI	Adaptive Spatial Frequency Integration
AUGMENT	supervised contr**A**stive learning and inpainting for hyperspectral denoising and methane pl**U**me se**GMENT**ation
BCE	Binary Cross Entropy
CH_4	Methane
CL	Contrastive Learning
CNN	Convolution Neural Network
CV	Computer Vision
DL	Deep Learning
DNN	Deep Neural Network
EGCA	Efficient Global Channel Attention
FFT	Fast Fourier Transform
FLOPS	Floating Point Operations
GAP	Global Average Pooling

GFA	Global Feature Attention
HSI	Hyperspectral Image
HSID	Hyperspectral Image Denoising
IDEATION	hybr**I**d 3**D** cnn–transform**E**r **A**nd spa**T**ial-spectral attent**I**on framework for hyperspectral image den**O**isi**N**g
IP	Indian Pines
LRD	Long Range Dependency
LTD	Limited Training Data
MF1	Mean F1 Score
MFPR	Mean False Positive Rate
MIoU	Mean Intersection Over Union
MLP	Multilayer Perceptron
MP	Mean Precision
MR	Mean Recall
NLP	Natural Language Processing
OA	Overall Accuracy
PSNR	Peak Signal-to-Noise Ratio
PU	University of Pavia
RL	Representation Learning
RS	Remote Sensing
SA	Self-Attention

SAM	Spectral Angle Mapper
SBMP	Single Band Multipass
SCL	Supervised Contrastive Learning
SFT	Spatial Frequency Transformer
SOTA	State-of-the-art
SSIM	Structural Similarity Index Measure
SSL	Self-Supervised Learning
SSR	Spatial Spectral Representation
SST	Spatial Spectral Transformer
SVM	Support Vector Machine
ViT	Vision Transformer
WV-3	World View-3

Chapter 1

Introduction

1.1 Background and Motivation

Hyperspectral imaging is an advanced sensing technology that captures information across hundreds of contiguous spectral bands, typically spanning wavelengths from the ultraviolet to the infrared (3). Unlike conventional RGB images, which capture information in only three color channels, an HSI provides a detailed spectral signature for every pixel in a scene. Consequently, HSIs preserve both spatial and spectral information, enabling accurate characterization and discrimination of materials with similar visual appearances but distinct spectral properties. Due to this capability, HSI has been widely adopted in numerous applications, including remote sensing (RS) (4), environmental monitoring (5), agriculture (6), mineral exploration (7), disaster assessment (8), medical imaging (9), and military surveillance (10).

Despite its advantages, acquiring HSI data is highly susceptible to various sources of degradation. Sensor imperfections, atmospheric interference, calibration errors, limited illumination, and transmission disturbances often introduce various types of noise, including Gaussian noise, impulse noise, stripes, dead pixels, and mixed noise (11). These degradations significantly distort the intrinsic spectral-spatial structure of HSIs and adversely affect the performance of downstream applications, including classification, target detection, reconstruction, and segmentation (4; 12). As a result, HSID has become an essential preprocessing step to recover meaningful information from corrupted data. This step is

particularly important because even small distortions in spectral signatures can propagate through subsequent processing stages, reducing the reliability of higher-level analysis.

Among the many applications of HSI, environmental monitoring has gained significant attention due to growing concerns regarding climate change and greenhouse gas emissions (13). Methane (CH_4) is recognized as one of the most influential greenhouse gases contributing to global warming (14). It is emitted from both natural sources, such as wetlands and oceans, and human-driven activities, including agriculture, fossil-fuel extraction, coal mining, and waste management (15; 16; 17). Although CH_4 remains in the atmosphere for a shorter duration than carbon dioxide, its global warming potential over a 20-year period is approximately 81 times greater than that of CO_2 (18; 19). Recent studies estimate that CH_4 is responsible for nearly 30% of current global warming, making its rapid mitigation one of the most effective strategies for slowing near-term climate change (20). In response to this urgency, around 160 countries have joined the Global Methane Pledge, committing to reduce CH_4 emissions by at least 30% from 2020 levels by 2030 (21).

Achieving these emission-reduction targets requires accurate monitoring systems capable of detecting CH_4 leaks and delineating plume boundaries over large geographical regions. In this context, HSI has emerged as a valuable remote sensing technology because its high spectral resolution enables the identification of CH_4 absorption features that are often invisible in conventional imagery (22; 23; 24). CH_4 plume segmentation from HSI plays an important role in environmental surveillance, leak localization, and emission quantification. However, this task remains difficult in practice due to weak, diffuse plume signals, complex, heterogeneous backgrounds, and high spectral redundancy arising from strong inter-band correlations (25; 26; 27).

Recent developments in DL have substantially improved the performance of both HSI denoising and CH_4 plume segmentation. Nevertheless, the effectiveness of these approaches is often constrained by the limited availability of annotated hyperspectral datasets (28; 29). Acquiring large-scale labeled HSI data is expensive, time-consuming, and frequently impractical, particularly for CH_4 plume monitoring (30; 31). Consequently, many models struggle to fully learn general spectral–spatial patterns, which affects their performance when conditions change across sensors, environments, or acquisition settings.

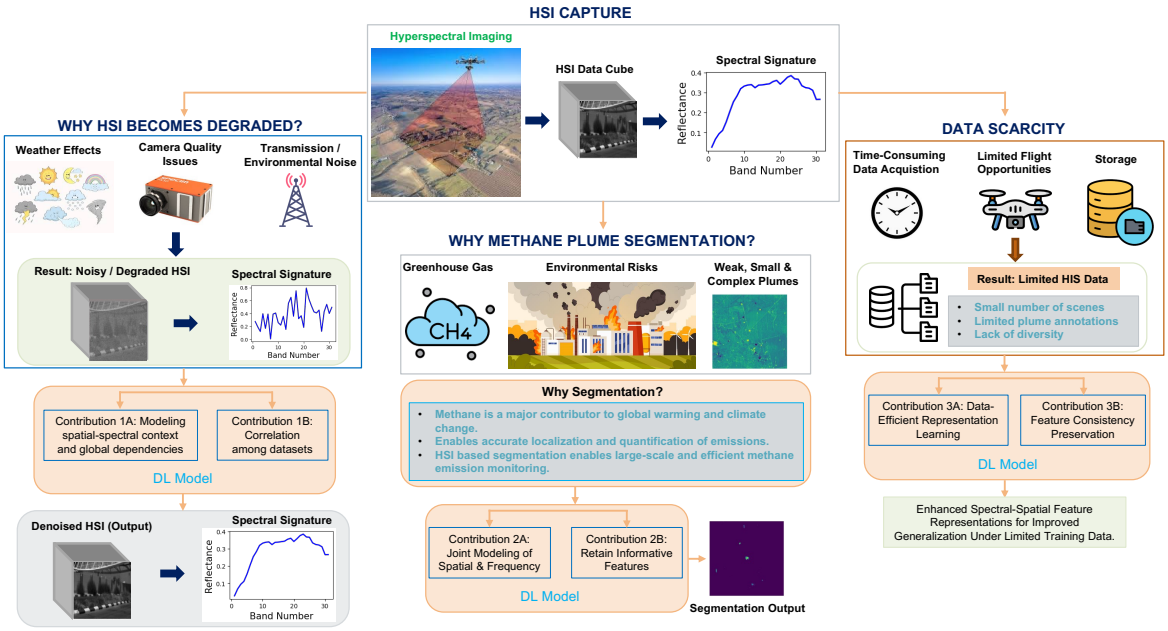


Figure 1.1: Overview of the key contributions of the thesis.

1.2 Thesis Contributions

This research addresses key challenges in HSID and CH_4 plume segmentation by designing DL frameworks specifically designed for hyperspectral data. The proposed studies explore SSR learning, long-range dependency modeling, frequency-aware feature extraction, and SCL to better utilize the rich information contained in HSIs, particularly in scenarios with limited annotated data. As illustrated in Figure 1.1, the research is organized into three major contributions, each addressing a different aspect of HSI analysis and CH_4 emission monitoring. The contributions are summarized as follows:

1.2.1 Contribution 1: Hybrid 3D CNN-Transformer and Spatial–Spectral Attention Framework for Hyperspectral Image Denoising

The first contribution addresses HSID, an essential preprocessing step for many HSI analysis applications. Although recent DL approaches have shown notable progress in reducing noise, preserving subtle spectral characteristics while effectively utilizing the rich spatial–spectral information in HSIs remains challenging. Moreover, hyperspectral data exhibit complex relationships across both spatial and spectral dimensions, making it important

to capture local structures as well as long-range dependencies (LRD) during denoising.

Research Gaps: Existing HSI denoising methods predominantly rely on either convolutional operations or attention-based mechanisms. CNN-based approaches effectively capture local spectral-spatial features but have limited capability in modeling long-range dependencies (32), whereas Vision Transformer (ViT) based methods capture global interactions through self-attention (SA) (33) but may not adequately preserve local details and typically focus on relationships within individual samples rather than across the dataset (34). In addition, many existing methods treat spatial and spectral information separately without explicitly modeling their joint interactions. These limitations motivate the development of a unified denoising framework that can simultaneously leverage local and global spatial-spectral details while preserving structural image integrity under diverse noise conditions.

Thesis Contributions: To address these challenges, we propose `IDEATION`, a hybrid framework that combines 3D CNN, a 3D UNet architecture, and a 3D Transformer for HSID. The architecture jointly exploits local spectral–spatial patterns and long-range contextual dependencies while preserving the structural integrity of HSI data through multi-scale encoder-decoder feature fusion. In addition, we introduce a novel **S**patial-**S**pectral **C**ubic Trans**former** (SSCformer), which models spatial, spectral, and joint spatial-spectral relationships using the proposed GFA mechanism. Unlike conventional SA, which captures dependencies within a single sample, GFA leverages correlations across the training dataset to learn representations consistent with the underlying noise characteristics of HSI data acquired by the same sensing device. By jointly modeling intra-image and inter-image relationships, the proposed framework facilitates richer spatial-spectral feature learning and improves HSID. The main contributions of this work are summarized as follows:

- We propose `IDEATION`, a hybrid framework for HSID that integrates 3D CNNs, a 3D U-Net with encoder-decoder skip connections, and a 3D Transformer to jointly exploit local spatial-spectral information and long-range contextual dependencies while preserving the structural integrity and fine details of HSI data.

- We introduce the **S**patial-**S**pectral **C**ubic Trans**former** (SSCformer), which sequentially models spatial, spectral, and joint spatial-spectral relationships. In addition,

learnable weighting parameters are incorporated to adaptively balance spatial and spectral information according to the characteristics of the input data.

- We design a GFA module that captures correlations not only within individual HSIs but also across the training dataset, enabling the model to learn more informative representations for HSID.

- We conduct extensive experiments on benchmark datasets under various noise settings, demonstrating that the proposed framework consistently achieves competitive performance in both quantitative evaluations and qualitative visual assessments compared with existing state-of-the-art methods.

Together, these contributions present a unified framework for HSID that exploits spatial–spectral information and models both local and long-range dependencies. By preserving structural details and spectral characteristics across diverse noise conditions, the proposed approach facilitates improved HSI restoration.

1.2.2 Contribution 2: Frequency-Aware Transformer and Enhanced Skip Connections for Hyperspectral Methane Plume Segmentation

The second contribution focuses on CH_4 plume segmentation from hyperspectral imagery, an important task for environmental monitoring, greenhouse gas emission assessment, and climate-change mitigation. While recent DL methods have improved plume detection performance, accurately delineating CH_4 plumes remains challenging due to weak spectral signatures, diffuse plume boundaries, complex background interference, and the high spectral redundancy inherent in hyperspectral data.

Research Gaps: Existing CH_4 plume segmentation methods primarly rely on CNN- or Transformer-based architecture. CNNs are effective at extracting local spatial features but have limited ability to capture long-range contextual dependencies, whereas Transformer-based approaches model global interactions through SA but may not adequately preserve

local structural details and boundary continuity that are essential for accurate plume delineation (35). Moreover, existing methods operate predominantly in the spatial domain and do not explicitly exploit complementary frequency-domain information, which can aid in distinguishing weak plume signatures from complex backgrounds (36). In addition, encoder–decoder frameworks commonly employ direct skip connections that transfer raw HSI features from the encoder to the decoder without accounting for spectral redundancy or noise arising from highly correlated bands (37). These limitations motivate the development of a framework that jointly leverages spatial and spatial-frequency representations while selectively propagating informative spectral features to improve CH_4 plume segmentation.

Thesis Contributions: To address these challenges, we propose `ALKANE`, an attention-based framework for hyperspectral CH_4 plume segmentation that combines spatial and spatial–frequency analysis within a unified encoder-decoder architecture. The proposed framework introduces a **S**patial **F**requency **T**ransformer (SFT), which replaces conventional SA with an **A**daptive **S**patial **F**requency **I**ntegration (ASFI) mechanism to jointly model spatial structures and complementary spatial–frequency information derived from magnitude and phase components. This design enables more effective characterization of weak and diffuse CH_4 plumes while maintaining boundary continuity. In addition, we incorporate an EGCA module into the skip connections to selectively propagate spectrally informative and spatially relevant features while suppressing redundant or noise-dominated channels. The main contributions of this work are summarized as follows:

- We propose `ALKANE`, an attention-based encoder-decoder framework for hyperspectral CH_4 plume segmentation that jointly exploits spatial and spatial-frequency representations to improve plume localization and boundary delineation.
- We introduce an SFT equipped with an ASFI mechanism that combines spatial features with magnitude- and phase-based frequency information, enabling complementary representation learning for CH_4 plume segmentation.
- We design an EGCA module for the skip connections that selectively emphasizes spectrally informative and spatially relevant features while suppressing redundant or noise-dominated channels prior to feature fusion.

- We perform extensive evaluations on multiple publicly available CH_4 plume datasets acquired from different hyperspectral sensors and sensing environments, demonstrating the robustness, generalization capability, and better segmentation performance of the proposed framework compared with existing state-of-the-art methods.

Together, these contributions establish a unified attention-driven framework that integrates spatial and spatial–frequency representations for hyperspectral CH_4 plume segmentation. By combining complementary feature cues and refining feature propagation during decoding, the proposed approach improves plume localization and boundary delineation across diverse hyperspectral datasets.

1.2.3 Contribution 3: Supervised Contrastive Learning and Inpainting for Hyperspectral Denoising and Methane Plume Segmentation

The third contribution focuses on SCL and inpainting-based representation learning for HSID and CH_4 plume segmentation. Although recent DL methods have shown promising results in both tasks, they performance is often limited by the scarcity of annotated HSI data and their reliance on task-specific optimization objectives, which may restrict the learning of meaningful spatial-spectral representations (38).

Research Gaps: Existing HSID methods primarily optimize reconstruction losses, while CH_4 plume segmentation approaches mainly rely on segmentation objectives. Consequently, the learned representations may not fully capture the intrinsic spatial-spectral characteristics of HSI data. Moreover, the limited availability of annotated HSIs further complicates training DL models (39). Although CL has shown promise for RL, its application to HSI analysis remains unexplored (40). In particular, existing approaches typically depend on both positive and negative sample pairs, whereas the use of positive-only SCL with structured augmentations has received little attention (41; 42; 43; 44; 45). Likewise, image inpainting has been extensively studied for recovering missing regions but has rarely been explored as an auxiliary task for generating informative positive views and enhancing RL in HSID and downstream CH_4 plume segmentation (46; 47; 48). These observations highlight the need for a unified framework that can leverage SCL and inpainting to strengthen SSR learning and

better utilize the available training data for both HSI denoising and CH_4 plume segmentation (49).

Thesis Contributions: To address these challenges, we propose AUGMENT, a unified framework that integrates positive-only SCL with HSI inpainting to enhance SSR learning under limited training data (LTD). The framework employs inpainting and geometric transformations to generate structured positive views, enabling the model to learn contextual and spectral relationships without relying on negative sample pairs while supporting improved denoising and CH_4 plume segmentation. The specific contributions are as follows:

- We propose AUGMENT, a unified framework that incorporates SCL to enhance SSR learning. The framework can be integrated with existing HSI denoising and CH_4 plume segmentation pipelines, enabling the extraction of meaningful features from noisy hyperspectral data for both image denoising and plume delineation tasks.

- We introduce an inpainting-based augmentation strategy within the SCL framework to generate structured variations of HSIs. By encouraging consistency across transformed views, the proposed strategy promotes the learning of contextual spatial-spectral relationships and improves feature representation under LTD conditions.

- We demonstrate that AUGMENT can be seamlessly incorporated into existing HSID and CH_4 plume segmentation methods, thereby improving LTD-based learning across multiple datasets.

Together, these contributions establish a unified RL framework that enhances spatial–spectral feature discrimination, improves robustness under LTD, and benefits both HSID and CH_4 plume segmentation.

1.3 Thesis Organization

This thesis is organized into six chapters, each addressing a distinct aspect of HSI analysis, including HSI denoising, CH_4 plume segmentation, and spatial-spectral representation learning using SCL and inpainting.

- Chapter 1 (the current chapter) presents the background and motivation for this research. It discusses the significance of hyperspectral imaging, the challenges associated with HSI degradation and CH_4 emission monitoring, and the need for advanced learning frameworks for HSI analysis.

- Chapter 2 provides a comprehensive review of the existing literature related to HSID, CH_4 plume segmentation, SCL, and inpainting-based representation learning. The chapter highlights the strengths and limitations of current approaches and identifies the research gaps addressed in this work.

- Chapter 3 presents `IDEATION`, a hybrid framework for HSI denoising that combines 3D CNNs, a 3D U-Net, and a 3D Transformer to jointly model local spatial-spectral characteristics and long-range contextual dependencies.

- Chapter 4 introduces `ALKANE`, an attention-driven framework for CH_4 plume segmentation from hyperspectral imagery. The proposed method integrates spatial and spatial-frequency representations through a frequency-aware Transformer and enhanced feature fusion mechanisms to improve plume localization and boundary delineation under diverse sensing conditions.

- Chapter 5 introduces `AUGMENT`, a unified framework that integrates SCL and HSI inpainting to generate informative training samples and enhance spatial-spectral representation learning. The proposed approach is designed to be incorporated into existing HSI denoising and CH_4 plume segmentation methods, particularly in scenarios with limited training data.

- Chapter 6 concludes the thesis by summarizing the major contributions and findings of the proposed frameworks. The chapter discusses their implications for HSI analysis and CH_4 emission monitoring, and outlines potential directions for future research.

Chapter 2

Literature Survey

This chapter reviews the existing literature relevant to HSI denoising, CH_4 plume segmentation, and SCL with inpainting-based representation learning. The review covers traditional, DL, and transformer-based methods for HSID, followed by recent developments in CH_4 plume detection and segmentation using HSI. Furthermore, studies on SCL and inpainting-based learning strategies are discussed to examine their potential for improving spatial-spectral feature representation under limited training data conditions. The objective of this review is to summarize the current state of the art, identify existing challenges and limitations, and establish the motivation for the frameworks proposed in the subsequent chapters.

2.1 Hyperspectral Image Denoising

2.1.1 Traditional Methods for Hyperspectral Image Denoising

Early HSI denoising approaches were predominantly based on model-driven optimization techniques that exploit handcrafted priors and the intrinsic characteristics of hyperspectral data. These methods generally formulate denoising as an inverse problem and solve it through iterative optimization procedures (47; 50; 51). Since HSIs exhibit strong spectral correlations and spatial redundancies, many traditional techniques utilize low-rank and sparse representations to recover clean images from noisy observations. Among the repre-

sentative methods, BM4D (52), an extension of BM3D (53), exploits nonlocal self-similarity by grouping similar spatial-spectral cubes and performing collaborative filtering in a transform domain. Similarly, LLRT (54) integrates a hyper-Laplacian prior with low-rank tensor modeling to preserve spectral structures while suppressing noise. NGMeet (55) further combines low-rank Tucker decomposition with matrix rank minimization to jointly exploit global spectral correlations and nonlocal spatial similarities. These approaches have demonstrated promising denoising capabilities, particularly under relatively simple noise conditions.

Despite their effectiveness, traditional methods often rely on carefully designed priors and multiple manually tuned hyperparameters. Their performance can degrade significantly in complex or mixed-noise scenarios commonly encountered in real-world HSI (56). Moreover, the iterative optimization procedures involved in these methods are computationally expensive, limiting their practicality for large-scale applications.

2.1.2 CNN Methods for Hyperspectral Image Denoising

The remarkable success of DL in computer vision (CV) has motivated the adoption of CNNs for HSID (57). Unlike traditional approaches, CNN-based methods learn data-driven mappings directly from noisy observations to clean images, eliminating the need for handcrafted priors and complex optimization procedures. Initial CNN-based denoising models were largely inspired by conventional 2D image restoration techniques. For example, GRNet (58) employs multiple sequential CNN modules to extract local and global spatial-spectral features for noise removal. T3SC (59) formulates hyperspectral denoising as a trainable sparse coding problem and achieves better performance than many traditional approaches. However, because HSIs inherently possess three-dimensional structures comprising two spatial dimensions and one spectral dimension, 3D convolutional operations are generally more suitable for modeling spectral-spatial relationships. In this direction, HSID-CNN (11) combines 2D and 3D convolutional branches to capture spatial and spectral information separately. Subsequently, QRNN3D (60) further improves denoising performance by employing 3D convolutional operations to jointly model spatial and spectral dependencies.

Although CNN-based methods have significantly advanced HSI denoising, their receptive

fields remain inherently local. Consequently, capturing long-range dependencies and global contextual information remains challenging, particularly when complex spectral structures extend across distant regions of the image (61).

2.1.3 Transformer Methods for Hyperspectral Image Denoising

To overcome the limited receptive field of CNNs, transformer architectures have recently attracted considerable attention in HSID. Originally introduced for natural language processing (NLP) (62), transformers employ SA mechanisms that enable the modeling of long-range relationships and global contextual dependencies. The success of the ViT (63) has inspired numerous transformer-based approaches in CV and RS (64). By partitioning images into patches and establishing relationships among them using SA, transformers can effectively capture global information that CNNs often struggle to model (65). However, directly applying conventional 2D transformers to HSIs may disrupt the inherent three-dimensional spectral-spatial structure, potentially weakening spectral consistency across bands.

To address this challenge, several transformer-based denoising frameworks have been developed specifically for HSIs. SSTD (66) explores both local and global spectral-spatial dependencies using transformer modules. SST (65) employs spatial and spectral attention mechanisms to exploit correlations across both domains for effective denoising. Similarly, (32) introduces a spectral enhancement strategy alongside SA modules to improve feature representation and noise suppression. Despite their strong ability to capture global dependencies, transformer-based approaches exhibit several limitations. SA mechanisms primarily focus on pairwise relationships within individual samples and may overlook broader inter-sample correlations present across the training dataset (34). Furthermore, many existing methods either emphasize spatial and spectral information separately or lack an effective mechanism to adaptively balance the contributions of both domains. In addition, transformer architectures may not adequately preserve fine-grained local details, which remain important for accurate HSI reconstruction (56).

These limitations motivate the development of a DL framework that jointly exploits local spectral-spatial representations and global contextual information while effectively

balancing their contributions for improved HSI denoising.

2.2 Methane Plume Segmentation

2.2.1 Traditional Methods for Methane Plume Segmentation

Early CH_4 plume detection approaches primarily relied on physics-based and statistical techniques that exploit the characteristic absorption signatures of CH_4 in hyperspectral imagery. Representative methods include matched filtering (67), Iterative Maximum A Posteriori Differential Optical Absorption Spectroscopy (IMAP-DOAS) (68), and ratio-based products (69). These techniques enhance methane-related spectral responses by comparing observed spectra against predefined CH_4 absorption patterns and background characteristics.

Although such methods have demonstrated effectiveness in controlled environments, their performance is often affected by spectral interference arising from surface materials, atmospheric constituents, and illumination variations. Consequently, they frequently produce false detections in complex scenes and exhibit limited adaptability across diverse environmental conditions. To improve reliability, manual inspection is commonly employed; however, this process becomes impractical when processing large-scale airborne or satellite datasets (31). Furthermore, methods such as IMAP-DOAS often rely on integrating airborne and ground-based measurements, which limits their scalability for operational, continuous CH_4 monitoring applications.

2.2.2 CNN-Based Methods for Methane Plume Segmentation

The limitations of traditional techniques have motivated the adoption of DL approaches for CH_4 plume segmentation. CNNs have emerged as a popular choice due to their ability to automatically learn discriminative spatial features directly from data. MethaNet (70) and CH4Net (71) employ CNN-based architectures trained on simulated plume datasets to identify CH_4 emission regions. Subsequently, U-Plume (72) and HyperSTARCOP (31) further improved segmentation performance by incorporating encoder-decoder architectures

and multi-scale feature extraction.

These methods benefit from the strong local feature learning capability of convolutional operations and achieve substantial improvements over traditional approaches. Nevertheless, CNNs inherently possess a limited receptive field, making it challenging to capture long-range contextual dependencies that are often important for accurately delineating plume boundaries under heterogeneous backgrounds. As a result, segmentation performance may degrade in scenarios involving weak plumes, cluttered environments, or spectrally complex scenes (25; 26; 27).

2.2.3 Transformer-Based Methods for Methane Plume Segmentation

To overcome the restricted contextual modeling capability of CNNs, transformer architectures have recently been introduced for CH_4 plume segmentation. By employing SA mechanisms, transformers can model long-range relationships and global interactions across an image, thereby improving the representation of extended plume structures. Gasformer (35) utilizes transformer-based feature extraction to capture global contextual information relevant to CH_4 plume detection. Similarly, MethaneMapper (73) incorporates spectral-aware transformer modules to better exploit the spectral characteristics of CH_4 signatures. Moreover, transformer-based segmentation frameworks developed in CV have demonstrated the effectiveness of attention mechanisms for dense prediction tasks (74; 75).

Despite these advances, most existing transformer-based approaches rely predominantly on spatial-domain SA. While effective for modeling global relationships, such mechanisms do not explicitly exploit frequency-domain information that can provide complementary cues regarding texture, structural consistency, and boundary characteristics. Consequently, the delineation of plume boundaries remains challenging, particularly when plume regions exhibit low contrast or are embedded within complex backgrounds (76).

2.2.4 Frequency-based Segmentation Methods

Recognizing the importance of frequency information, recent studies have explored the integration of frequency-domain representations into CH_4 plume segmentation frameworks.

Among these, FSDINet (77) introduces a dual-domain strategy that jointly extracts features from both spatial and frequency representations. By leveraging multi-scale analysis, the method demonstrates improved plume structures and enhanced robustness under challenging conditions. However, FSDINet is primarily built upon CNN architectures, which may limit its ability to model long-range dependencies and global interactions between spatial and frequency-domain features. As a result, maintaining boundary consistency and capturing complex plume patterns across large spatial extents remains challenging.

Additionally, channel-attention mechanisms have been extensively investigated to improve feature discrimination and information selection. For example, ECA-Net (78) introduces a lightweight attention mechanism that captures local inter-channel dependencies using adaptive one-dimensional convolution without significantly increasing computational complexity. Such attention strategies have shown considerable success in enhancing feature representation while maintaining efficiency. Despite the progress achieved by recent CNN- and transformer-based approaches, several challenges remain unresolved. Traditional methods struggle to generalize across diverse environmental conditions and often suffer from high false alarm rates. CNN-based models are effective at learning local spatial patterns but have limited capability to capture long-range contextual dependencies (79). Transformer-based methods address this limitation through SA mechanisms; however, existing approaches operate primarily in the spatial domain and do not explicitly exploit frequency-domain characteristics that can provide valuable information for plume localization and boundary delineation.

Furthermore, effectively integrating spatial and frequency-domain information while maintaining computational efficiency remains challenging. These limitations highlight the need for a unified framework that can jointly model local and global dependencies, leverage complementary spatial and frequency representations, and improve CH_4 plume segmentation performance across varying sensing conditions and scene complexities.

2.3 Supervised Contrastive Learning and Inpainting-based Hyperspectral Image Denoising and Segmentation

2.3.1 Hyperspectral Image Denoising

HSI denoising is a fundamental preprocessing step that aims to remove sensor- and environment-induced degradations while preserving the underlying spectral-spatial information. Existing denoising techniques can be broadly categorized into traditional model-based approaches and DL-based methods. Traditional denoising methods (47; 50; 80; 81) typically rely on handcrafted priors and iterative optimization strategies. These approaches exploit the intrinsic redundancy and correlation present across spatial and spectral dimensions of HSIs. For example, low-rank representation techniques (54; 82) model hyperspectral data as low-dimensional structures embedded in high-dimensional spaces, thereby facilitating noise suppression while preserving essential image content. Similarly, NGMeet (55) combines global spectral low-rank characteristics with spatial nonlocal self-similarity to improve restoration quality. Furthermore, (80) used Tucker decomposition to explicitly capture low-rank structures across spatial and spectral dimensions, while (83) proposed the DECSC framework that jointly exploits local spatial-spectral correlations, nonlocal self-similarities, and global spatial consistency. Despite their effectiveness, these methods often struggle under complex and mixed-noise conditions due to their reliance on predefined priors and manually tuned hyperparameters.

To address these limitations, DL-based approaches have gained significant attention for their ability to learn hierarchical feature representations directly from data. Early CNN-based methods demonstrated promising performance by exploiting spatial-spectral information through convolutional operations. HSID-CNN (11) employed residual convolutional networks to extract spectral-spatial features for denoising. Similarly, (58) utilized cascaded convolutional architectures to capture local and global correlations, while QRNN3D (60) incorporated quasi-recurrent neural networks to model long-range dependencies in hyperspectral data. T3SC (59) further combined trainable sparse coding with deep neural networks (DNN) to enhance denoising capability.

Recently, transformer architectures (62) have emerged as a powerful alternative due to their ability to model long-range dependencies through SA mechanisms. Inspired by the success of ViT (63), several transformer-based denoising methods have been developed for hyperspectral imagery. SST (65) employs attention mechanisms to jointly capture spatial and spectral similarities, whereas (32) integrates SA with spectral enhancement modules to learn discriminative hyperspectral representations. More recently, (84; 85), and (61) demonstrated that transformer-based architectures can effectively capture complex spectral-spatial interactions and improve denoising performance. Nevertheless, the effectiveness of these DL methods remains constrained by the limited availability of annotated hyperspectral datasets, which impedes the learning of robust, generalized representations.

2.3.2 Methane Plume Segmentation

CH_4 plume segmentation plays a crucial role in environmental monitoring and greenhouse gas emission assessment. Early CH_4 detection and segmentation methods relied primarily on physically motivated models and spectral matching techniques. Approaches such as matched filtering (67), IMAP-DOAS (68), and ratio-product-based methods (69) utilize prior knowledge of CH_4 absorption characteristics to enhance plume visibility. Although these techniques have demonstrated effectiveness in controlled scenarios, they often suffer from false detections caused by atmospheric variability, surface reflectance changes, and spectral interference from other gases. Moreover, their dependence on handcrafted processing pipelines limits their adaptability to diverse environmental conditions and large-scale monitoring applications.

The limitations of traditional approaches have motivated the adoption of DL techniques for CH_4 plume segmentation. CNN-based methods such as MethaNet (70), CH4Net (71), U-Plume (72), and HyperSTARCOP (31) leverage convolutional architectures to learn plume-related spatial patterns directly from data. These approaches have achieved improved segmentation performance by exploiting local spatial features and contextual information. However, the locality of convolution operations restricts their ability to capture long-range dependencies and global contextual relationships, which are often important for delineating

plume boundaries in complex scenes. To overcome these limitations, transformer-based segmentation methods have recently been explored. Gasformer (35) utilizes transformer encoders to model long-range contextual interactions, while MethaneMapper (73) incorporates spectral-aware transformer modules to improve plume representation learning (RL). Furthermore, recent advances in CV segmentation (74; 75) have demonstrated the effectiveness of attention mechanisms for capturing global context. However, most existing CH_4 plume segmentation approaches primarily focus on spatial-domain analysis and do not explicitly exploit complementary frequency-domain information that may assist in plume localization and boundary refinement.

Recognizing this limitation, FSDINet (77) introduced a frequency-aware framework that jointly analyzes spatial and frequency-domain representations using a CNN. By incorporating frequency information, the method improves plume segmentation performance and boundary delineation. Nevertheless, its reliance on convolutional operations limits its ability to model global dependencies and long-range interactions across spatial and spectral dimensions. Consequently, there remains a need for approaches that can effectively integrate local features, global contextual information, and complementary representations to achieve robust CH_4 plume segmentation under diverse sensing conditions.

2.3.3 Contrastive Learning

CL has emerged as an effective RL paradigm that aims to maximize the similarity between related samples while simultaneously increasing the separation between unrelated samples in the feature space (40). Through this objective, CL enables models to learn discriminative and robust representations without requiring extensive supervision. SCL extends this concept by incorporating class-label information during optimization (86). Instead of treating each sample independently, SCL encourages representations belonging to the same class to cluster together while separating samples from different classes. This additional supervision allows the learned features to exhibit stronger semantic consistency and improved discriminative power. Several CL frameworks, including SimCLR (87), MoCo (88), GenSCL (89), and SPTA (90), utilize both positive and negative sample pairs to learn meaningful repre-

sentations. In contrast, approaches such as BYOL (91) and ALPINE (45) demonstrate that effective representations can also be obtained using only positive sample relationships. Previous studies have shown that CL improves feature discrimination, robustness to noise, and generalization under limited-data conditions (49). These advantages have led to successful applications across various CV tasks, including classification, segmentation, and RL.

Despite these successes, the potential of CL for HSI analysis remains relatively underexplored. In particular, mostly HSI denoising methods rely on reconstruction objectives, while CH_4 plume segmentation approaches primarily focus on segmentation losses. As a result, the learned representations may not fully exploit the intrinsic spatial-spectral relationships present within HSI data.

2.3.4 Inpainting

Inpainting aims to reconstruct missing or corrupted regions by exploiting contextual information from the surrounding pixels. Traditional inpainting approaches primarily rely on diffusion-based and patch-matching techniques, where missing regions are filled using local spatial continuity or nonlocal similarities (46; 92). Although effective for natural images with small occlusions, these methods often struggle to preserve complex spatial structures and spectral consistency in hyperspectral imagery. Recent advances in DL have significantly improved inpainting performance by enabling models to learn semantic and contextual representations directly from data. CNN-based methods (93; 94) employ encoder-decoder architectures to infer missing content from neighboring regions. In the context of HSI, inpainting has been explored for recovering missing bands, dead pixels, and corrupted regions by leveraging spectral-spatial correlations across neighboring pixels and spectral channels (2). These studies demonstrate that inpainting can effectively exploit contextual information and improve RL.

Beyond image restoration, recent work has shown that masking and inpainting strategies can play an important role in self-supervised learning (SSL) by creating structured perturbations of the input data (95). By requiring the model to reconstruct the masked regions, the learned representations become more robust and semantically meaningful. Despite these

advantages, the integration of inpainting within SCL frameworks for HSID and CH_4 plume segmentation remains unexplored. Despite the success of CL in RL, its application to HSI analysis remains relatively limited. Furthermore, although masking and inpainting strategies have recently shown promise for learning robust visual representations, their integration with SCL for HSI analysis has received less attention. Consequently, the learned representations may not fully exploit the intrinsic spatial-spectral relationships present in HSI data, particularly under limited training data conditions.

Chapter 3

Hybrid 3D CNN–Transformer and Spatial–Spectral Attention Framework for Hyperspectral Image Denoising

3.1 Introduction

DL has significantly advanced the field of HSI analysis by enabling data-driven modeling of complex spatial–spectral relationships. In particular, HSID has witnessed substantial improvements through the use of DNN, which are capable of learning hierarchical feature representations for effective noise suppression (11; 58; 60). CNNs have been extensively explored for HSID due to their ability to capture local spatial–spectral structures by leveraging neighborhood information across both spatial and spectral dimensions (11). To further enhance reconstruction quality, several methods incorporate U-Net (96) based architectures, which are widely adopted for image restoration tasks. The encoder–decoder structure of U-Net, combined with skip connections, facilitates multi-scale feature aggregation while preserving fine-grained structural details that may otherwise be degraded during the encoding process (58).

Despite these advantages, CNN-based methods are inherently constrained by their local receptive fields, which restrict their ability to capture long-range dependencies present in

HSI data (97). To overcome this limitation, Transformer-based models have been introduced, leveraging SA mechanisms to model global contextual relationships across spatial and spectral dimensions (62; 63). Although these models enhance the representation of long-range interactions, conventional SA mechanisms operate on individual samples and do not explicitly capture correlations across different samples within a dataset (34). In HSID, such inter-sample correlations are significant, as data acquired under similar conditions often exhibit shared noise characteristics and statistical distributions. Ignoring these dependencies can reduce the effectiveness of the learned representations and impact denoising quality.

Furthermore, existing spatial–spectral approaches often treat spatial and spectral attention independently, without explicitly modeling their joint interactions (65). However, HSI data inherently exhibits strong coupling between spatial and spectral dimensions, requiring a unified representation that can simultaneously capture both aspects while adapting to data-specific characteristics. As a result, current methods face challenges in jointly preserving fine-grained local details, modeling global contextual information, and exploiting inter-sample correlations. To overcome these limitations, this chapter proposes a unified DL method that integrates complementary learning paradigms for HSID. The key contributions of the proposed methodology are summarized as follows:

- A hybrid HSID framework that integrates 3D CNNs, 3D Transformer-based encoders, and 3D U-Net architecture to jointly capture local spatial–spectral features and global contextual dependencies.

- A spatial–spectral Transformer block that sequentially models spatial, spectral, and joint dependencies for enhanced feature representation.

- A GFA mechanism that captures correlations not only within individual HSIs but also across samples in the dataset, enabling data-driven feature learning.

- A unified framework designed to handle complex and mixed noise conditions while supporting improved performance in downstream hyperspectral analysis tasks.

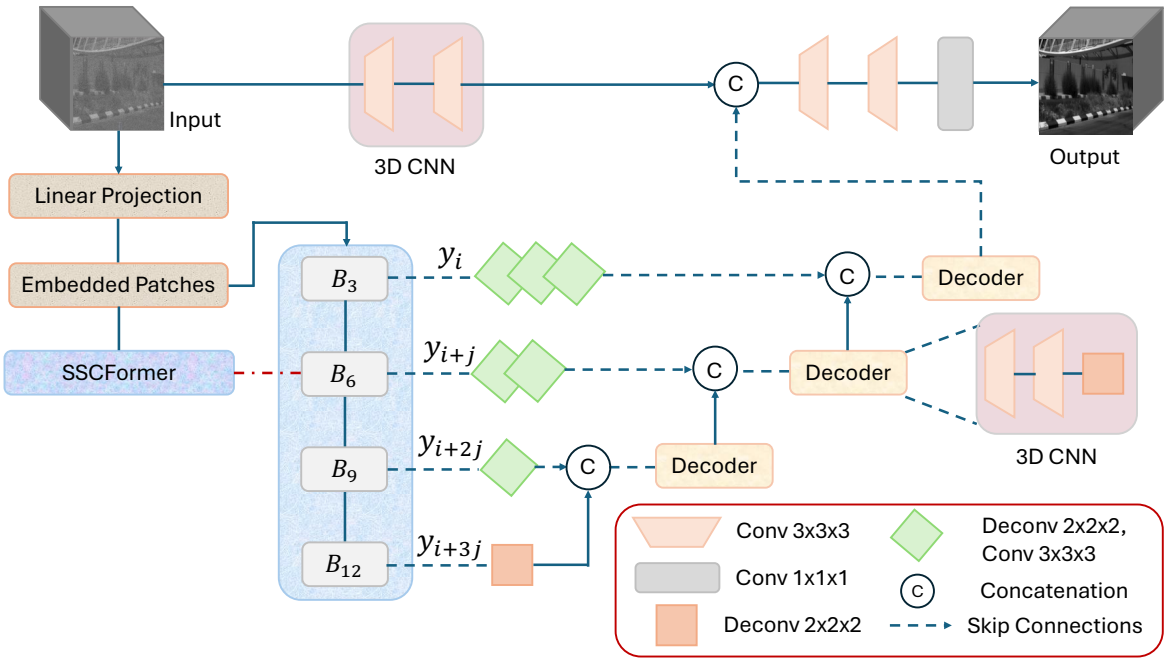


Figure 3.1: Overview of our proposed method, IDEATION.

3.2 Proposed Methodology

In this chapter, we present IDEATION (hybr**I**d 3**D** cnn–transform**E**r **A**nd spa**T**ial–spectral attent**I**on framework for hyperspectral image den**O**isi**N**g), a unified DL method designed for HSID. The design of the proposed method is motivated by the inherent properties of HSI data, where relevant information is distributed jointly across spatial and spectral dimensions, and noise often follows structured patterns across samples acquired under similar imaging conditions (11). Existing methods generally focus on either local spatial–spectral feature extraction or global contextual modeling, but do not effectively integrate these aspects within a single framework. In addition, inter-sample relationships, which are important in HSI data due to shared acquisition characteristics, are often not explicitly considered (82; 98). To address these limitations, IDEATION combines 3D CNN-based feature extraction, Transformer-based global representation learning, and a dataset-aware attention mechanism within a unified architecture.

In the standard ViT framework (63), the input is typically a 2D image denoted as $\mathbf{X} \in \mathbb{R}^{h \times w \times s}$, where h, w, and s correspond to the height, width, and number of channels, respectively. The input image is divided into N non-overlapping patches, where each patch is

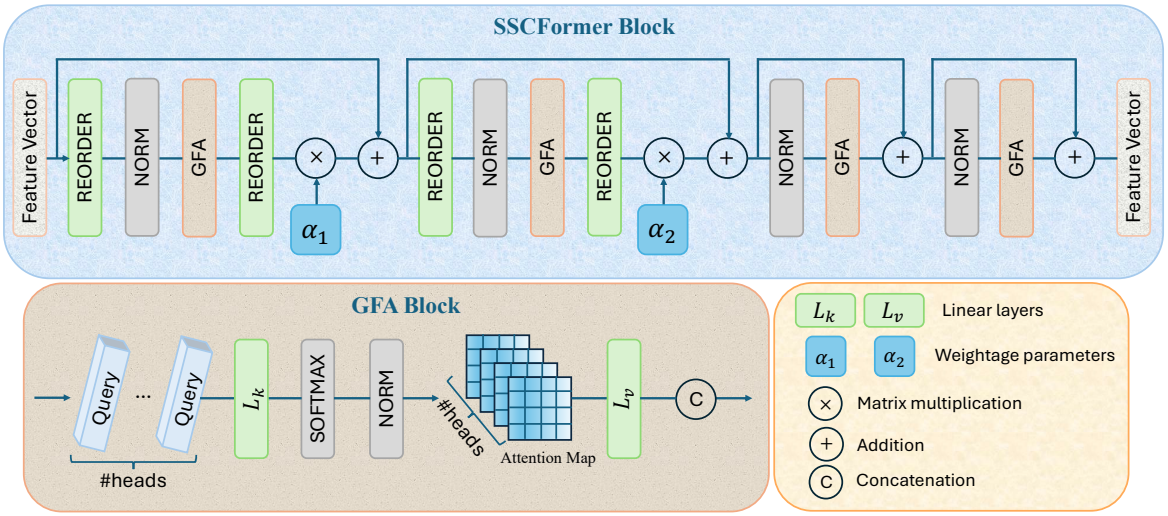


Figure 3.2: Illustration of the proposed SSCformer block incorporating the GFA module. The input feature representations are sequentially reordered, normalized, and processed through GFA modules, in which learnable weighting parameters (α_1 and α_2) govern feature aggregation. The GFA module employs multi-head attention to model spatial, spectral, and joint spatial-spectral relationships, thereby enhancing feature representation.

flattened into a vector representation $\mathbf{x}_k \in \mathbb{R}^{N\times(k^2\cdot s)}$. Here, each patch has a dimensionality of $k^2 \cdot s$, and the total number of patches is given by $\frac{h\cdot w}{k^2\cdot s}$.

In contrast, HSID involves volumetric data, where the input is represented as $\mathbf{X} \in \mathbb{R}^{h\times w\times s\times d}$. In this representation, d denotes the number of spectral bands, and for the considered HSIs, $s = 1$ since each spectral band contains a single channel. Extending the ViT formulation to this setting, the 3D input is partitioned into non-overlapping cubical patches, denoted as $\mathbf{x}_k \in \mathbb{R}^{N\times k^3}$, where each patch has dimensions $k \times k \times k$. The total number of such patches is computed as $\frac{h\cdot w\cdot d}{k^3}$. Each cubical patch is subsequently flattened and projected into a higher-dimensional embedding space of dimension D using a learnable linear transformation represented by $\mathbf{E} \in \mathbb{R}^{k^3\times D}$. To preserve positional information, a learnable positional embedding $\mathbf{e}_{pe} \in \mathbb{R}^{N\times D}$ is added to the projected features. The resulting embedded representation is given by:

$$\mathbf{z}_k^0 = \mathbf{x}_k\mathbf{E} + \mathbf{e}_{\text{pe}} \tag{3.1}$$

This stage completes the preprocessing of HSIs before they are passed to a sequence of SSCformer blocks. The SSCformer serves as the core component of the proposed framework,

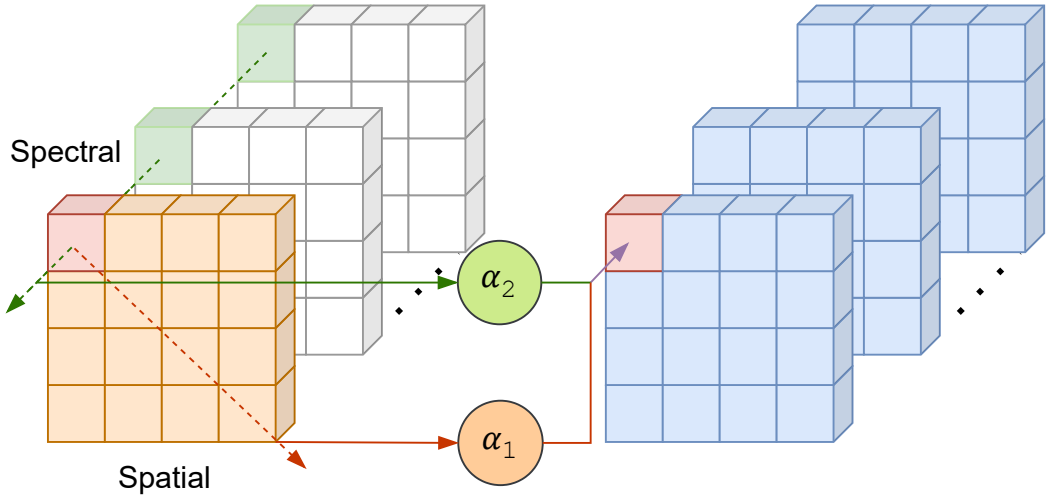


Figure 3.3: HSIs are divided into $16\times16\times16$ patches. The red patch is the focus, with green patches assessed for spectral correlation and orange patches for spatial correlation. Weights α_1 and α_2 combine spatial and spectral correlations respectively to enhance the red patch, which is then examined with blue patches for inter-domain correlations.

responsible for learning expressive feature representations from hyperspectral data (detailed in Section 3.2.1). It leverages multiple attention operations through the proposed GFA mechanism to capture meaningful correlations and generate informative feature embeddings. A detailed description of GFA is provided in Section 3.2.2. The refined features produced by the SSCformer are subsequently fed into a U-Net architecture for the final denoising task.

In the modified U-Net design, the encoder receives inputs from selected SSCformer blocks. Specifically, outputs are extracted from every j^{th} block starting from the i^{th} block, denoted as y_i, y_{i+j}, y_{i+2j}, and y_{i+3j}. In this work, the values $i = 3$ and $j = 3$ are adopted. These feature maps are reshaped to the size $\left(\frac{h}{k}\right) \times \left(\frac{w}{k}\right) \times \left(\frac{d}{k}\right) \times D$ prior to being processed by the encoder blocks. Each encoder block consists of $2 \times 2 \times 2$ deconvolution layers followed by a $3\times3\times3$ convolution layer. The output of the final SSCformer block, y_{12}, is also passed through a $2 \times 2 \times 2$ deconvolution layer to ensure compatibility in spatial dimensions with the encoder outputs. These feature maps are then combined within the decoder through concatenation, followed by upsampling using two $3\times3\times3$ convolutional layers and $2\times2\times2$ deconvolution layers. This procedure is repeated across multiple decoder stages (refer to Fig. 3.1).

In the final stage, the last decoder block consists of two $3 \times 3 \times 3$ convolution layers followed by a $1 \times 1 \times 1$ convolution layer. This block integrates the output of the preceding decoder stage along with features derived from the input image processed through additional $3 \times 3 \times 3$ convolution layers. The output of this final layer corresponds to the denoised HSI.

3.2.1 Spatial-Spectral Cubic Transformer Block

Following preprocessing, the embedded representations are processed through a stack of SSCformer blocks. Each block comprises multiple GFA modules followed by a Multilayer Perceptron (MLP) (99), and is designed to sequentially model spatial, spectral, and joint spatial–spectral dependencies. The operations within the ℓ^{th} block are defined as:

$$z_\ell^1 = \alpha_1 GFA\left(\text{Norm}\left(z_{(\ell-1,sp)}\right)\right) + z_{\ell-1} \tag{3.2}$$

$$z_\ell^2 = \alpha_2 GFA\left(\text{Norm}\left(z_{(\ell,st)}^1\right)\right) + z_\ell^1 \tag{3.3}$$

$$z_\ell^3 = GFA\left(\text{Norm}\left(z_\ell^2\right)\right) + z_\ell^2 \tag{3.4}$$

$$z_\ell = MLP\left(\text{Norm}\left(z_\ell^3\right)\right) + z_\ell^3 \tag{3.5}$$

where $\ell = 1, \ldots, L$.

In Eqn. 3.2, patches are reorganized into $\frac{d}{k}$ groups, each containing $\frac{h \cdot w}{k^2}$ patches sharing the same spatial location. This rearrangement is achieved using the REORDER block (refer to Fig. 3.2). The representation, $z_{(\ell-1,sp)}$ denotes spatially aligned patch groups. After layer normalization, GFA is applied to capture spatial dependencies within each group. The scaling factor, α_1, regulates the contribution of spatial attention, while the residual connection preserves information from the previous layer. In Eqn. 3.3, the grouping strategy is modified to capture spectral dependencies. Here, patches are divided into $\frac{h \cdot w}{k^2}$ groups, each containing $\frac{d}{k}$ spectrally aligned patches. The notation $z_{(\ell-1,st)}$ represents these spectrally grouped patches. GFA is then applied to model spectral correlations, with α_2 controlling its contribution. A residual connection ensures continuity of feature information. Subsequently, Eqn. 3.4 captures interactions across both spatial and spectral dimensions without explicit grouping. This allows the model to learn both inter- and intra-dimensional relationships

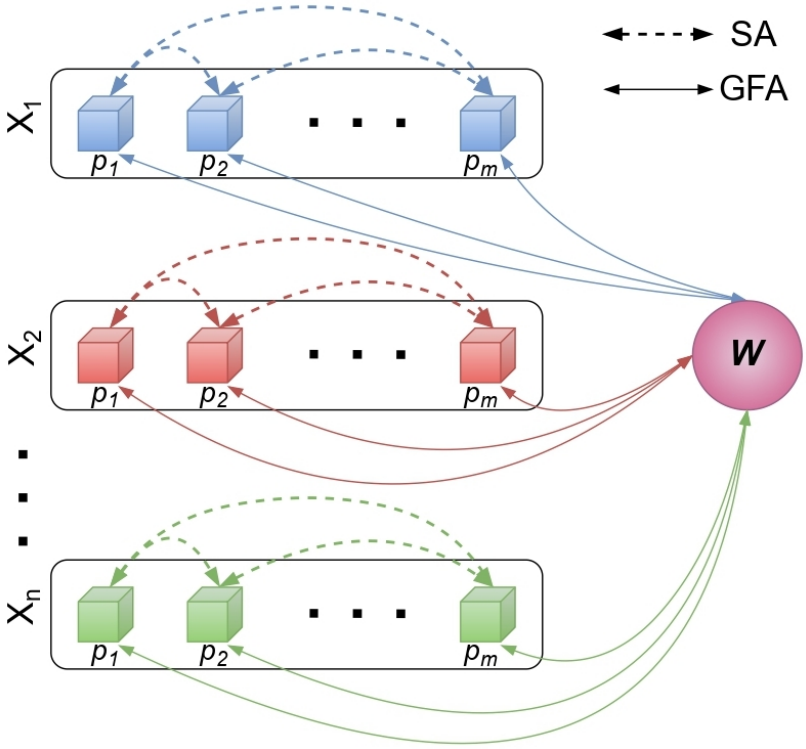


Figure 3.4: Visualization of SA vs GFA: In SA, attention is focused between patches within an image, X_i. In GFA, attention is focused between all patches and a common weight matrix, $\mathbf{W}$. This way, similarities among different image samples are captured by $\mathbf{W}$.

(refer to Fig. 3.3). The resulting feature representation can be expanded as:

$$z_\ell^3 = z_\ell^2 + GFA\left(z_\ell^2\right) \tag{3.6}$$

$$= z_\ell^1 + \alpha_2 GFA\left(z_{(\ell,st)}^1\right) + GFA\left(z_\ell^2\right) \tag{3.7}$$

$$= z_{\ell-1} + \alpha_1 GFA\left(z_{(\ell-1,sp)}\right) + \alpha_2 GFA\left(z_{(\ell,st)}^1\right) + GFA\left(z_\ell^2\right) \tag{3.8}$$

The parameters α_1 and α_2 control the relative contributions of spatial and spectral attention, respectively. The refined representation is then passed through the MLP block (Eqn. 3.5) to produce the final output of the SSCformer block.

3.2.2 Global Feature Attention

To introduce the proposed GFA mechanism, we first revisit the standard SA. In SA, the attention matrix, $\mathcal{A}$, is initially computed, followed by the calculation of the output patch

embedding, $\boldsymbol{Y}$. The process can be mathematically written as follows:

$$\mathcal{A} = SM(Q \cdot K^T / \sqrt{d_k}) \tag{3.9}$$

$$\boldsymbol{Y} = \mathcal{A} \cdot \boldsymbol{V} \tag{3.10}$$

where, Q, K, and V are obtained through linear projections of the input X, defined as $\boldsymbol{Q} = X \cdot \boldsymbol{W_Q}$, $\boldsymbol{K} = X \cdot \boldsymbol{W_K}$, and $\boldsymbol{V} = X \cdot \boldsymbol{W_V}$, where $\boldsymbol{W_Q}$, $\boldsymbol{W_K}$, and $\boldsymbol{W_V}$ are learnable weight matrices associated with query, key and value projections. The softmax function $SM(\cdot)$ ensures that the attention weights are normalized. This mechanism captures relationships within a single input sample. However, hyperspectral datasets often contain samples with similar characteristics (100), and conventional SA does not explicitly model relationships across different samples (34). To address this limitation, GFA introduces globally learned parameters that enable the modeling of inter-sample correlations (see Fig. 3.4). The GFA formulation is given by:

$$\mathcal{A}_i = \text{NORM}(\text{SM}(\boldsymbol{Q}_i \cdot \boldsymbol{W}_{\boldsymbol{K}_i})) \tag{3.11}$$

$$a_i = \mathcal{A}_i \cdot \boldsymbol{W}_{\boldsymbol{V}_i} \tag{3.12}$$

$$\boldsymbol{Y} = \text{Concat}(a_1, a_2, \ldots, a_n) \cdot \boldsymbol{W}_{out} \tag{3.13}$$

The input to the GFA module is denoted as $z_\ell^i \in \mathbb{R}^{N \times D}$, projected into a query representation $\boldsymbol{Q}_i$ using a learnable weight matrix $\boldsymbol{W}_{\boldsymbol{Q}_i} \in \mathbb{R}^{D \times D'}$, such that $\boldsymbol{Q}_i = z_\ell^i \boldsymbol{W}_{\boldsymbol{Q}_i}$. Here, $i \in \{1, \ldots, n\}$ denotes the index of the attention head, where n is the total number of heads in the GFA module. Subsequently, as shown in Eq. 3.11, the attention matrix, $\mathcal{A}_i$, is derived by applying the global learnable layer, $\mathcal{L}_k$, to $\boldsymbol{Q}_i$ (see Fig. 3.2). The weight matrix for $\mathcal{L}_k$ is represented as $\boldsymbol{W}_{\boldsymbol{K}_i} \in \mathbb{R}^{D' \times h}$, where h denotes the hidden dimension of $\mathcal{L}_k$. The resulting representation is normalized using the softmax function, followed by an L_1 normalization operation NORM$(\cdot)$ to ensure stable attention distribution. Kindly note that $\boldsymbol{W}_{\boldsymbol{K}_i}$ is not derived from the current input sample, unlike conventional SA. Instead, it is a globally learnable parameter that is optimized across the entire dataset. This enables the model to capture correlations between different HSI samples, rather than being restricted

to intra-sample relationships. Following this, as defined in Eq. 3.12, the intermediate representation a_i is obtained by applying another learnable transformation parameterized by $\boldsymbol{W}_{\boldsymbol{V}_i} \in \mathbb{R}^{h \times D'}$. This transformation refines the feature representation based on the attention weights encoded in $\mathcal{A}_i$. Finally, the outputs from all attention heads are concatenated and projected back to the original embedding space using $\boldsymbol{W}_{out} \in \mathbb{R}^{D' \times D}$, as shown in Eq. 3.13. This ensures consistency between input and output feature dimensions.

The subscripts $\boldsymbol{K}$ and $\boldsymbol{V}$ indicate that the corresponding weight matrices are analogous to key and value transformations in standard SA. However, in contrast to SA, the matrices $\boldsymbol{W}_{\boldsymbol{K}_i}$ and $\boldsymbol{W}_{\boldsymbol{V}_i}$ are independent of the input features and are learned across all samples in the dataset. As a result, the output $\boldsymbol{Y}$ captures inter-sample relationships by leveraging shared statistical properties across hyperspectral data. In summary, the GFA mechanism extends conventional attention by incorporating globally learned transformations, enabling the extraction of feature representations that encode both intra-sample and inter-sample correlations in a data-driven manner.

3.2.3 Training Loss Function

To guide the optimization of the proposed `IDEATION` framework, the Structural Similarity Index Measure (SSIM) (101) is adopted as the training objective. Unlike conventional pixel-wise loss functions (such as ℓ_1 or ℓ_2 norms), SSIM evaluates the similarity between two images by jointly accounting for luminance consistency, contrast preservation, and structural information.

Given a clean HSI $\mathbf{X}_{\text{clean}} \in \mathbb{R}^{h \times w \times d}$ and its reconstructed counterpart $\hat{\mathbf{X}}$, the SSIM between them is defined as:

$$SSIM(\mathbf{X}_{\text{clean}}, \hat{\mathbf{X}}) = \frac{(2\mu_X \mu_{\hat{X}} + c_1)(2\sigma_{X\hat{X}} + c_2)}{(\mu_X^2 + \mu_{\hat{X}}^2 + c_1)(\sigma_X^2 + \sigma_{\hat{X}}^2 + c_2)},$$

where μ_X and $\mu_{\hat{X}}$ denote the mean intensities of the clean and reconstructed images, respectively. The terms σ_X^2 and $\sigma_{\hat{X}}^2$ represent the corresponding variances, while $\sigma_{X\hat{X}}$ denotes the covariance between the two images. The constants c_1 and c_2 are small positive values introduced to avoid numerical instability. Since HSIs consist of multiple spectral bands,

SSIM is computed in a band-wise manner and averaged across all d bands to obtain the final similarity score. Accordingly, the training loss is defined as:

$$\mathcal{L}_{SSIM} = 1 - \frac{1}{d}\sum_{k=1}^{d} SSIM\big(\mathbf{X}_{\text{clean}}^{(k)}, \hat{\mathbf{X}}^{(k)}\big), \tag{3.14}$$

where $\mathbf{X}^{(k)}$ denotes the k^{th} spectral band. By optimizing this objective, the proposed framework promotes the preservation of structural characteristics and spatial–spectral consistency in the reconstructed hyperspectral data. This is particularly beneficial in scenarios involving complex or mixed noise, where maintaining structural information is critical for subsequent analysis tasks.

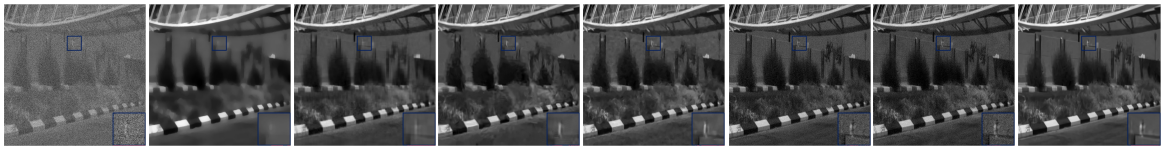
3.3 Experimental Settings

3.3.1 Dataset

The experiments in this study are conducted on the ICVL hyperspectral dataset (102), which contains 201 HSIs. Each HSI has a spatial resolution of 1392×1300 pixels and comprises 31 spectral bands spanning 400 nm to 700 nm. For model development and evaluation, the dataset is partitioned into three subsets: 100 images for training, 20 images for validation, and the remaining images for testing. Prior to feeding the data into the network, all images are normalized to the range $[0, 1]$ to ensure stable training. To facilitate efficient learning and to follow established practices in HSID (32), the images are divided into overlapping volumetric patches. Specifically, patches of size $64 \times 64 \times 31$ are extracted for training, while others $512 \times 512 \times 31$ are used during testing. This strategy is consistent with commonly adopted protocols in recent HSID works, such as (103). In addition to the ICVL dataset, the generalization capability of the proposed framework is further assessed using real-world HSI datasets, including the University of Pavia (PU) (104), Indian Pines (IP) (105), and Urban (106). These datasets provide diverse scene characteristics and noise distributions, enabling a comprehensive evaluation of the model's robustness.

Table 3.1: Denoising Results of various methods under different Gaussian Noise Levels on the ICVL Dataset

	σ = 10			σ = 30			σ = 50			σ = 70		
Methods ↓	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
Noisy	28.13	0.8792	18.72	18.59	0.5523	37.90	14.15	0.3476	49.01	11.23	0.2301	56.45
LLRT (54)	46.72	0.9983	1.60	41.12	0.9920	2.52	38.24	0.9830	3.47	36.23	0.9732	4.46
NGMeet (55)	47.90	0.9988	1.39	42.44	0.9816	2.06	39.69	0.9658	2.49	38.05	0.9531	2.83
HSID-CNN (11)	43.14	0.9918	2.12	40.30	0.9854	3.14	37.72	0.9746	4.27	34.95	0.9521	5.84
QRNN3D (60)	45.61	0.9977	1.80	42.18	0.9955	2.21	40.05	0.9929	2.63	38.09	0.9883	3.42
SST (65)	48.28	0.9990	1.30	43.32	0.9971	1.87	41.09	0.9950	2.19	39.55	0.9921	2.46
SERT (32)	47.72	0.9988	1.36	43.56	0.9969	1.77	41.33	0.9949	2.05	39.82	0.9929	2.30
IDEATION	**48.92**	**0.9992**	**1.24**	**45.50**	**0.9977**	**1.66**	**44.90**	**0.9971**	**1.88**	**43.88**	**0.9958**	**2.05**



Noisy LLRT NGMeet HSID QRNN3D SST SERT IDEATION

Figure 3.5: Visual denoising results for the ICVL dataset with Gaussian noise at a noise level of $\sigma = 50$.

3.3.1.1 Noise Settings

In real-world scenarios, HSIs are frequently corrupted by various types of noise arising from sensor imperfections, environmental conditions, and acquisition processes. Common noise sources include Gaussian noise, stripe artifacts, dead lines or pixels, and impulse noise (82; 107). To evaluate the denoising performance of IDEATION, we simulate multiple noise conditions following widely used experimental protocols (60). First, additive independent and identically distributed (i.i.d.) Gaussian noise is introduced at levels ranging from 10 to 70, in increments of 20. This allows the assessment of model performance under varying noise intensities. Furthermore, to mimic more realistic and challenging degradation scenarios, five types of complex noise configurations are considered. These include non-i.i.d. Gaussian noise, a combination of Gaussian and stripe noise, Gaussian noise with deadline artifacts, Gaussian noise combined with impulse noise, and a mixture of multiple noise types. Such diverse settings ensure a comprehensive evaluation of the model under both simple and highly degraded conditions.

Table 3.2: Denoising performance of various methods across five complex noise scenarios on the ICVL Dataset

Noise Type →	Non-i.i.d Gaussian			Gaussian+Stripe			Gaussian+Deadline			Gaussian+Impulse			Gaussian+Mixture		
Methods ↓	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
Noisy	18.29	0.5116	46.20	17.51	0.4867	46.98	17.50	0.4770	47.55	14.93	0.3758	46.98	13.91	0.3396	51.53
BM4D (52)	36.18	0.9767	5.78	35.63	0.9730	6.26	33.77	0.9615	6.85	29.79	0.8613	21.59	28.01	0.8419	23.59
LLRT (54)	34.18	0.9618	4.88	34.27	0.9628	4.93	32.98	0.9559	5.29	28.85	0.8819	18.17	28.06	0.8697	19.37
NGMeet (55)	34.90	0.9745	5.37	34.88	0.9660	5.42	33.41	0.9665	6.55	27.02	0.7884	31.20	26.13	0.7796	31.89
HSID-CNN (11)	39.28	0.9819	3.80	38.09	0.9765	4.59	38.33	0.9783	3.99	36.21	0.9663	5.48	35.30	0.9588	6.29
QRNN3D (60)	42.18	0.9950	2.84	41.68	0.9943	2.97	41.69	0.9942	2.61	40.32	0.9914	4.31	39.08	0.9892	4.80
SST (65)	43.57	0.9970	2.05	42.97	0.9960	2.15	42.74	0.9961	2.21	41.66	0.9950	2.60	38.78	0.9911	2.99
SERT (32)	44.20	0.9971	1.69	43.68	0.9969	1.97	43.66	0.9969	1.99	42.67	0.9959	2.30	40.00	0.9937	2.84
IDEATION	**46.95**	**0.9974**	**1.57**	**45.12**	**0.9975**	**1.88**	**46.05**	**0.9978**	**1.93**	**44.78**	**0.9968**	**2.20**	**45.37**	**0.9949**	**2.72**

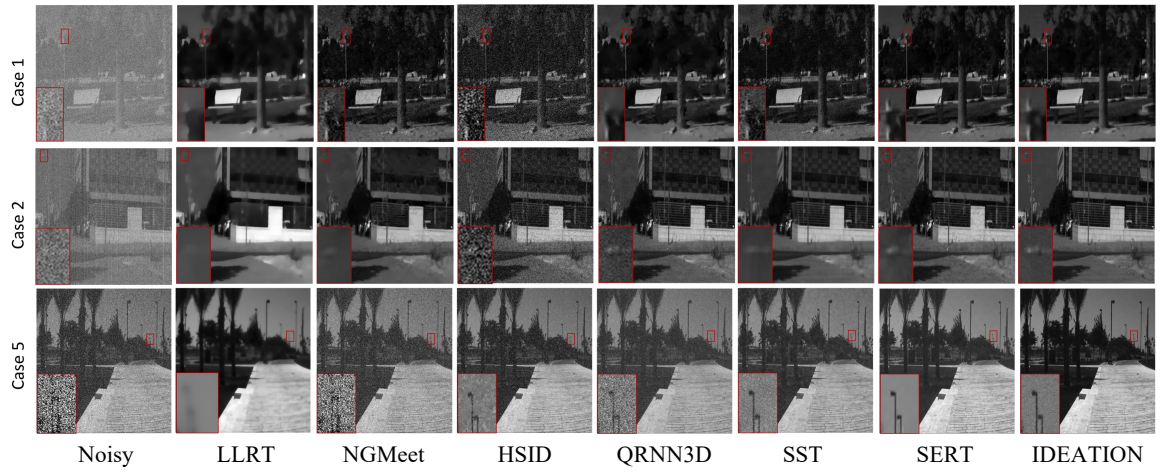


Figure 3.6: Visual comparison of the denoised images for different cases in the ICVL dataset

3.3.2 Implementation Details

The proposed method is implemented using the PyTorch DL library[1] along with the MONAI framework[2]. All experiments are conducted on a system equipped with an NVIDIA GeForce RTX 3050 GPU, an AMD Ryzen 5 5600X processor, and 16 GB of RAM. For optimization, the AdamW optimizer (108) is employed with an initial learning rate of 1×10^{-4}. The model is trained using a batch size of 4 over 500 epochs. The total training time is approximately 15 hours. The Transformer backbone follows the standard ViT configuration (ViT-B16) (63).

3.4 Experimental Results

3.4.1 Competing Methods

To evaluate the performance of the proposed `IDEATION` framework, we compare it with both traditional and DL-based HSID methods under a variety of noise conditions, including Gaussian and complex mixed noise scenarios. Conventional methods typically perform well when the noise distribution is known or follows predefined assumptions. However, DL methods offer greater adaptability by learning data-driven representations, enabling them to

[1] `http://pytorch.org/`
[2] `https://monai.io/`

handle diverse and complex noise characteristics more effectively.

In this work, we compare IDEATION with several SOTA methods representing different methodological categories. These include BM4D (52), a non-local filtering-based approach; LLRT (54), which leverages low-rank tensor modeling; and NGMeet (55), based on non-local orthogonal transformations. Additionally, we include comparisons with prominent DL-based methods, including HSID-CNN (11), QRNN3D (60), SST (65), and SERT (32).

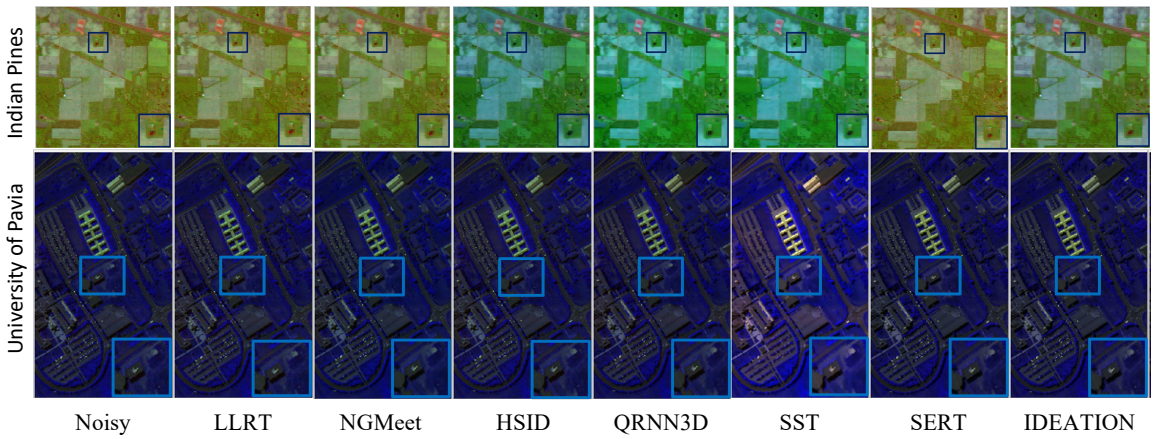


Figure 3.7: Real denoised results on IP and PU datasets for bands (3, 103, 203) and (5, 13, 95), respectively.

3.4.2 Simulated Experimental Results

3.4.2.1 Gaussian Noise

To assess the denoising capability of the proposed framework under controlled settings, zero-mean additive white Gaussian noise with varying standard deviations (σ) is added to the clean HSIs. The quantitative evaluation results, presented in Table 3.1, demonstrate that IDEATION consistently outperforms all competing methods across different noise levels. The results highlight the robustness and strong generalization capability of the proposed method, particularly under varying noise intensities. It is observed that while traditional and DL-based approaches tend to show comparable performance at higher noise levels, DL-based models generally exhibit superior performance under moderate and low noise conditions due to their ability to learn complex spatial–spectral representations.

For qualitative analysis, Fig. 3.5 presents visual comparisons of denoised HSIs at a representative noise level of $\sigma = 50$. The proposed IDEATION framework effectively suppresses noise while preserving essential structural details and spatial textures. Furthermore, to assess the performance of different denoising methods in the spectral domain, Fig. 3.9 illustrates the corresponding spectral profiles. The results demonstrate that IDEATION reconstructs spectral signatures that closely match the ground truth, thereby validating its capability to preserve both spatial structures and spectral fidelity.

3.4.2.2 Complex Noise

Table 3.2 reports the quantitative performance of the proposed IDEATION framework in comparison with other competing methods on the ICVL dataset under various complex noise conditions. For qualitative assessment, Fig. 3.6 presents visual comparisons across different noise scenarios, clearly demonstrating the superior capability of the proposed method in handling diverse and challenging noise degradations.

Conventional denoising techniques often exhibit limited performance in the presence of mixed and impulse noise, as such degradations tend to corrupt both spatial structures and spectral signatures. These methods rely on handcrafted priors, which are typically insufficient to model complex noise distributions without prior assumptions about the noise characteristics (65). In contrast, DL-based approaches are more effective in addressing such challenges, as they learn robust spatial-spectral representations directly from data. Motivated by this advantage, the proposed IDEATION framework leverages a data-driven learning strategy to effectively suppress complex noise while preserving essential image information.

Table 3.3: Classification performance on IP, PU, and Urban datasets before and after hyperspectral image denoising

Dataset	**Index**	**Noisy**	(54)	(55)	(11)	(60)	(65)	(32)	**IDEATION**
IP	OA	71.25	81.99	83.01	83.31	84.38	87.01	88.17	**90.89**
	Kappa	0.6696	0.7952	0.8130	0.8101	0.8228	0.8422	0.8628	**0.8960**
PU	OA	86.54	91.88	92.25	92.81	93.10	95.67	96.52	**98.70**
	Kappa	0.8068	0.9007	0.9113	0.9167	0.9218	0.9376	0.9588	**0.9770**
Urban	OA	86.11	91.27	92.33	92.76	92.99	94.17	94.91	**96.02**
	Kappa	0.8385	0.8847	0.8909	0.9088	0.9086	0.9189	0.9267	**0.9528**

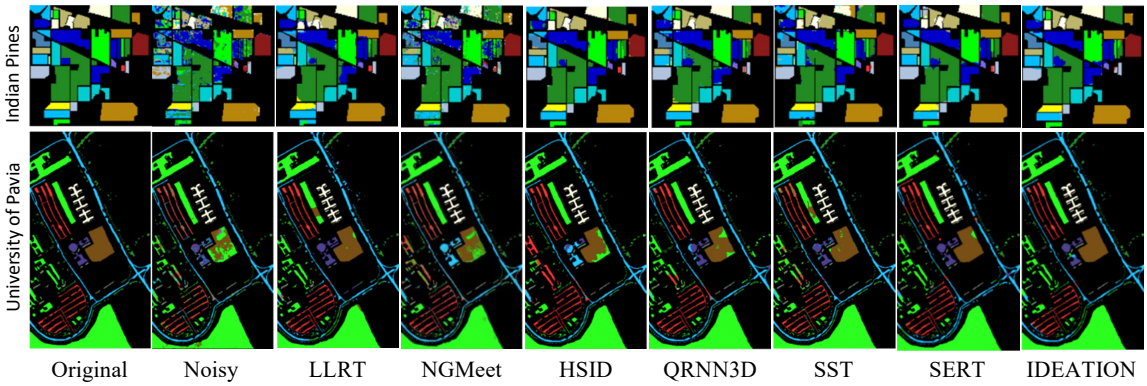


Figure 3.8: Classification outcomes for IP and PU datasets.

3.4.3 Real-Data Experimental Results

In this section, we evaluate the effectiveness of the proposed IDEATION framework on real-world noisy HSIs to demonstrate its practical applicability and generalization capability. Unlike simulated experiments, real-data analysis presents additional challenges due to unknown and complex noise distributions inherent in real acquisition systems. For this purpose, experiments are conducted on widely used benchmark datasets, including IP and PU. The IP dataset consists of images with a spatial resolution of 145×145 pixels and 220 spectral bands, whereas the PU dataset contains images with a spatial resolution of 610×340 pixels and 103 spectral bands. In addition, we further validate the proposed method on the Urban dataset (1) and the RealHSI dataset (2), which provide diverse real-world scenarios for comprehensive evaluation.

Qualitative denoising results for different spectral band combinations are presented in Fig. 3.7. It can be observed that the proposed IDEATION framework produces visually cleaner outputs with better preservation of spatial structures and finer details compared to existing approaches. The superiority of the proposed method is further evident in Fig. 3.10, where the reconstructed images exhibit enhanced sharpness and reduced noise artifacts.

To quantitatively assess the impact of denoising on downstream tasks, we evaluate HSI classification performance before and after denoising. A Support Vector Machine (SVM) classifier (4) is employed for this purpose, and the results are measured using Overall Accuracy (OA) and the Kappa coefficient (109). The classification maps for the IP and PU datasets are illustrated in Fig. 3.8, while the results for the Urban dataset are shown in Fig. 3.11.

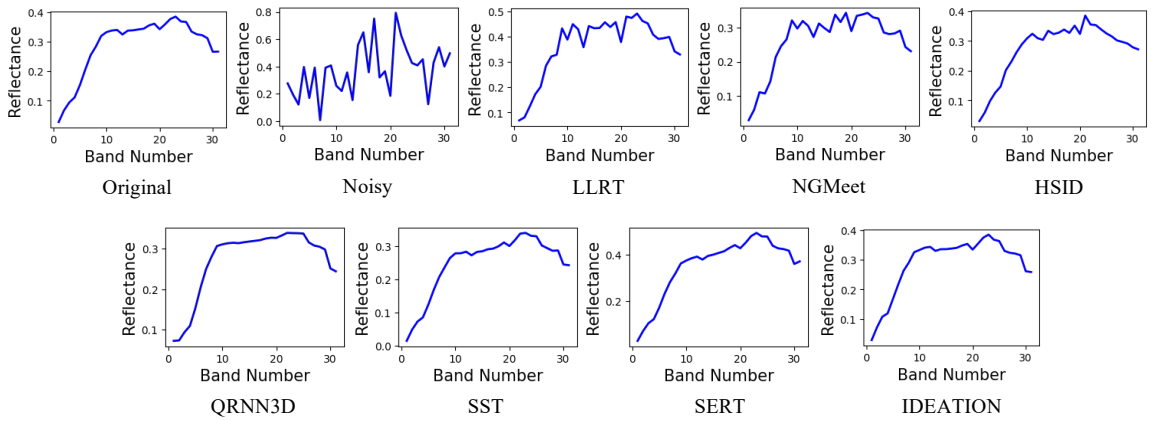


Figure 3.9: Spectral curves of the ICVL dataset at coordinates (110, 110) under Gaussian noise ($\sigma = 50$).

Corresponding quantitative comparisons are summarized in Table 3.3. The results clearly indicate that denoising using the proposed IDEATION framework significantly improves classification performance, highlighting its ability to recover meaningful spatial–spectral information. Furthermore, to analyze spectral fidelity, Fig. 3.9 presents spectral profile comparisons. The proposed method closely follows the ground truth signatures, demonstrating its effectiveness in preserving spectral consistency while suppressing noise. Overall, both qualitative and quantitative evaluations confirm that the proposed IDEATION framework outperforms existing methods on real-world datasets, achieving superior denoising performance and enhancing downstream hyperspectral analysis tasks.

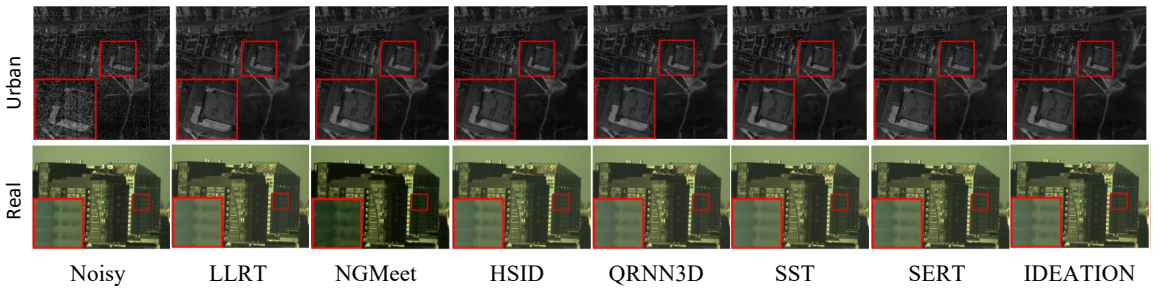


Figure 3.10: Visual outcomes for real-world noise reduction on the Urban (1) and Real (2) Datasets.

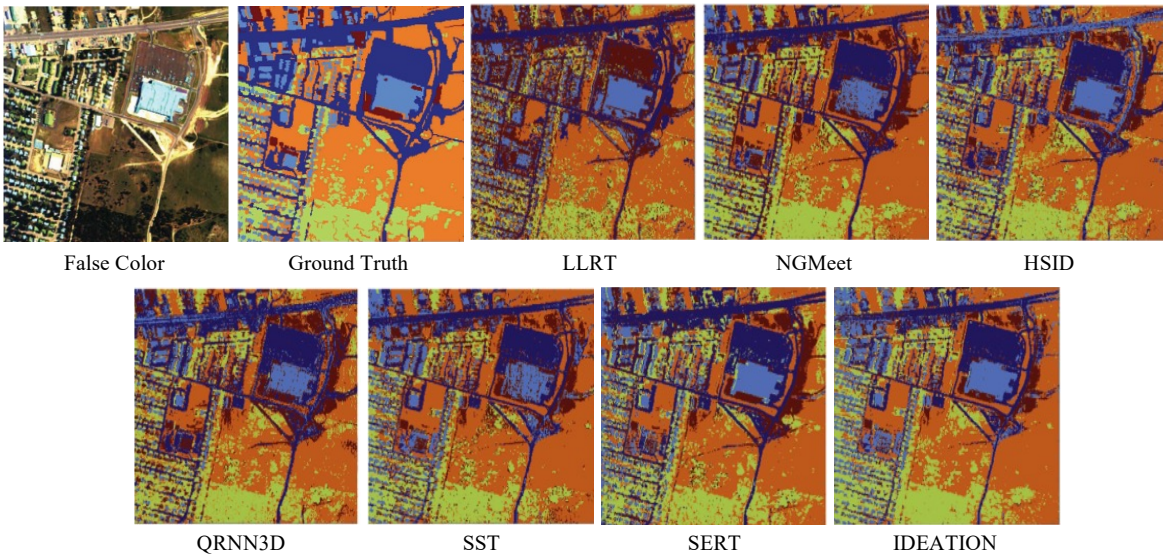


Figure 3.11: Classification outcomes for Urban (1) dataset.

3.4.4 Model Complexity

We analyze the computational efficiency of the proposed `IDEATION` framework in terms of floating-point operations (GFLOPs), model complexity (number of parameters), and execution time for both training and inference. The proposed method requires approximately 144.47 GFLOPs, significantly lower than recent SOTA approaches such as SERT, which reports 1018.9 GFLOPs. In terms of model complexity, the number of learnable parameters in `IDEATION` remains comparable to existing methods while achieving improved denoising performance, offering a favorable trade-off between accuracy and efficiency. Furthermore, the average inference time of the proposed framework is approximately 5.1 seconds per sample, which is reasonably close to SERT (3.2 seconds). The total training time for `IDEATION` is approximately 15 hours under the specified experimental setup.

3.4.5 Ablation Study

To comprehensively evaluate the effectiveness of the proposed framework, we perform ablation studies at both the architectural level and the SSCformer block level. All variants are trained and evaluated under identical experimental settings on the ICVL dataset to ensure a fair comparison. Table 3.4 presents the architectural ablation study, which investigates

Table 3.4: Ablation study of different architectural variants of the proposed IDEATION framework on the ICVL dataset.

Components			$\sigma = 30$		$\sigma = 50$		$\sigma = 70$	
CNN	**Transformer**	**U-Net**	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
2D	✗	✗	36.89	0.9688	35.42	0.9622	34.05	0.9522
3D	✗	✗	38.28	0.9731	37.23	0.9617	36.33	0.9551
3D	✗	3D	42.32	0.9709	40.25	0.9715	38.60	0.9662
✗	3D (SA)	3D	43.12	0.9804	42.28	0.9732	41.14	0.9701
✗	3D (GFA)	3D	43.68	0.9822	42.86	0.9761	41.73	0.9724
3D	2D (SA)	3D	41.79	0.9893	39.83	0.9847	37.96	0.9849
3D	2D (GFA)	3D	42.21	0.9901	40.35	0.9856	38.54	0.9855
3D	3D (SA)	✗	42.86	0.9721	42.14	0.9712	40.88	0.9637
3D	3D (SA)	3D	43.78	0.9821	42.96	0.9758	41.92	0.9742
3D	3D (GFA)	3D	45.50	0.9977	44.90	0.9971	43.88	0.9958

the contributions of the 3D CNN, 3D U-Net, and Transformer components. As expected, replacing the 2D CNN with a 3D CNN consistently improves denoising performance by jointly exploiting spatial and spectral information. Incorporating the 3D U-Net further enhances results by preserving structural details through encoder-decoder skip connections and enabling multi-scale feature fusion. Introducing transformer-based modeling leads to additional improvements by capturing long-range dependencies that are difficult to model using convolution alone. A comparison between Transformer variants employing conventional SA and the proposed GFA-based attention demonstrates that GFA consistently achieves superior performance across all noise levels. Furthermore, the best results are obtained when the 3D CNN, 3D U-Net, and the proposed 3D GFA-based Transformer are integrated into a unified framework, validating the complementary roles of volumetric convolution, multi-scale feature aggregation, and dataset-aware attention in HSID.

To further analyze the proposed SSCformer, Table 3.5 investigates the influence of the GFA module and the learnable scaling parameters α_1 and α_2. The GFA module enables the Transformer to capture correlations not only within an individual HSI but also across the training dataset, resulting in more informative feature representations than conventional SA. The results show that removing either α_1 or α_2 leads to a noticeable reduction in performance, indicating that both spatial-enhanced and spectral-enhanced features contribute to effective

Table 3.5: Ablation study of the proposed method with the SSCformer block on the ICVL dataset.

Components ↓			$\sigma = 30$		$\sigma = 50$		$\sigma = 70$	
GFA	α_1	α_2	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
✓	✗	✗	44.38	0.9834	43.22	0.9768	41.14	0.9782
✗	✓	✗	43.22	0.9887	42.85	0.9743	40.75	0.9719
✗	✗	✓	43.93	0.9815	43.11	0.9736	42.80	0.9718
✓	✓	✗	43.35	0.9833	42.98	0.9860	41.02	0.9831
✗	✓	✓	44.03	0.9807	43.81	0.9853	42.18	0.9769
✓	✗	✓	44.55	0.9819	43.57	0.9802	42.15	0.9843
✓	✓	✓	45.50	0.9977	44.90	0.9971	43.88	0.9958

representation learning. The degradation is more pronounced when α_2 is omitted, suggesting that spectral information plays a particularly important role in HSID. Moreover, using a single scaling parameter biases the feature representation toward a single domain, whereas jointly learning both parameters allows the network to adaptively balance spatial and spectral contributions according to the characteristics of the data. Overall, the highest denoising performance is achieved when GFA, α_1, and α_2 are all incorporated within the SSCformer, demonstrating the effectiveness of the proposed attention mechanism and adaptive feature weighting strategy.

3.5 Summary

HSID is a fundamental preprocessing step that directly impacts the performance of subsequent hyperspectral analysis tasks, including classification, segmentation, and target detection. In this chapter, we presented `IDEATION`, a hybrid DL framework developed to address the challenges associated with noise removal in HSIs by combining multiple complementary learning strategies. The proposed framework integrates 3D CNNs, Transformer-based architectures, and a spatial–spectral attention mechanism to capture diverse characteristics of hyperspectral data. By adopting a volumetric representation, the model jointly leverages spatial and spectral correlations, enabling comprehensive feature extraction. A core component of the framework is the SSCformer block, which sequentially models spatial, spectral,

and joint dependencies to enhance feature representation. In addition, the proposed GFA mechanism extends conventional SA by incorporating correlations across different hyperspectral samples, allowing the model to learn dataset-level representations. The integration of a modified 3D U-Net architecture further facilitates multi-scale feature fusion through skip connections, preserving fine structural information during the reconstruction process. The model is trained using an SSIM-based loss function, which emphasizes perceptual similarity and structural preservation rather than relying solely on pixel-wise differences. Experimental evaluations conducted on both simulated and real-world datasets demonstrate that the proposed `IDEATION` framework achieves improved performance compared to existing methods under various noise conditions. Furthermore, the denoised outputs contribute to better outcomes in downstream tasks such as HSI classification. Overall, this chapter presents `IDEATION` as a comprehensive framework for HSID, addressing key limitations of existing approaches and enabling improved spatial–spectral representation learning.

Chapter 4

Frequency-Aware Transformer and Enhanced Skip Connections for Hyperspectral Methane Plume Segmentation

4.1 Introduction

HSI has emerged as an important modality for environmental monitoring, particularly for the detection and spatial characterization of CH_4 plumes (15; 16; 17). Segmenting CH_4 plumes from HSI data is a crucial step in emission analysis, as it directly affects the estimation of plume extent and subsequent quantification processes (110; 111; 112). However, this task remains challenging due to strong inter-band correlation, background interference, and the inherently weak and diffuse nature of CH_4 spectral signatures (25; 26; 27). Recent DL methods have shown notable progress in HSI-based CH_4 plume segmentation by utilizing CNNs and encoder-decoder architectures (31; 71; 72; 77; 113; 114). These methods are effective at capturing local spatial structures; however, their reliance on limited receptive fields limits their ability to model long-range contextual dependencies (115). Transformer-based models mitigate this limitation through SA mechanisms, enabling global context

modeling. Nevertheless, existing transformer-based approaches operate primarily in the spatial domain and do not explicitly exploit frequency-domain characteristics, which can provide additional cues for representing subtle and diffuse plume patterns (35; 73; 116).

In HSI data, plume signatures often exhibit variations that are not easily separable in the spatial domain due to background complexity and spectral redundancy (117). Representing features in the frequency domain offers an alternative perspective, where structural and textural patterns can be analyzed in terms of spatial-frequency components (75). In particular, the magnitude component reflects the distribution of spatial-frequency energy, highlighting dominant structures, while the phase component preserves spatial alignment and geometric continuity, which are essential for maintaining plume boundaries (77; 118). By jointly considering magnitude and phase information, both plume morphology and boundary characteristics can be better represented, especially in scenarios involving weak or diffuse emissions (36). Furthermore, conventional skip connections in encoder-decoder architectures transfer features directly from the encoder to the decoder without selective refinement. In the context of HSI data, this direct propagation may introduce redundant or spectrally irrelevant information during feature fusion, which can affect the separation of plume and background regions (37; 79).

To address these challenges, we propose ALKANE (**A**ttention-based **L**earning and **K**nowledge **A**ggregation **N**etwork with **E**nhanced skip connections) for hyperspectral CH_4 plume segmentation. The proposed framework integrates frequency-domain representations with improved feature fusion strategies within a unified architecture. A frequency-aware Transformer is employed to jointly model spatial and spatial-frequency information, enabling better representation of weak and diffuse plume structures. In addition, an enhanced skip connection mechanism selectively propagates informative features while suppressing redundant components, thereby improving plume-background separation. The contributions of this chapter are summarized as follows:

- We introduce ALKANE, an attention-based framework for hyperspectral CH_4 plume segmentation that integrates spatial and spatial-frequency representations within a unified encoder-decoder architecture.

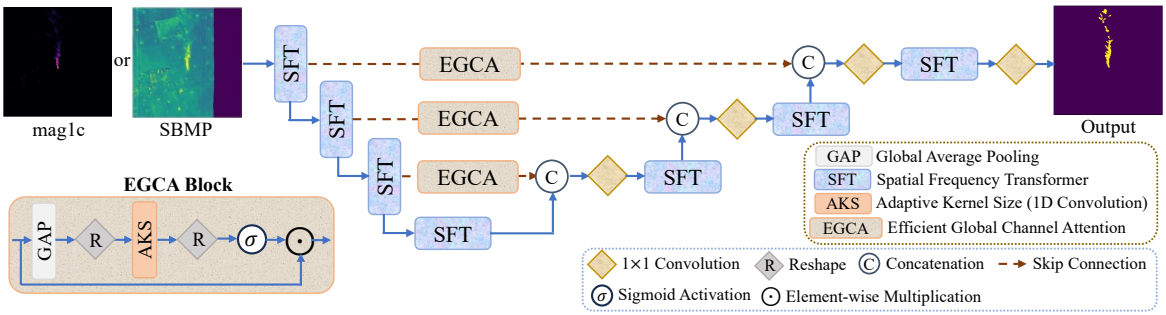


Figure 4.1: Overview of our proposed method ALKANE.

- We design a frequency-aware Transformer that incorporates both magnitude and phase information, enabling joint modeling of spatial structures and spatial-frequency patterns for improved characterization of weak and diffuse plume regions.

- We propose an enhanced skip connection mechanism that selectively propagates spectrally relevant and spatially informative features while suppressing redundant components, thereby facilitating better feature fusion during decoding.

- We conduct extensive experiments on publicly available HSI datasets acquired from different sensors, demonstrating the effectiveness of the proposed framework in comparison with existing methods.

4.2 Proposed Methodology

This section presents ALKANE, an HSI-based CH_4 plume segmentation framework designed using an encoder–decoder architecture. The encoder extracts hierarchical feature representations via the proposed SFT, which integrates spatial context with spatial–frequency representations using the ASFI module. By jointly modeling magnitude and phase components, the framework facilitates the representation of weak and diffuse plume structures while maintaining boundary continuity.

To further refine feature propagation between the encoder and decoder, an EGCA mechanism is incorporated within the skip connections. This module selectively emphasizes spectrally informative and spatially relevant channels, while suppressing redundant and

Algorithm 1 ALKANE

Input: Feature tensor $\mathbf{f} \in \mathbb{R}^{C \times H \times W}$
Output: CH_4 plume segmentation map $\mathbf{Y}$
for each encoder stage **do**
    // SFT: Query–Key–Value Projection
    Generate query, key, and value representations (Eq. (4.1))
    // ASFI applied to Query and Key (Eq. (4.2)**)**
    for $\boldsymbol{\varphi} \in \{\boldsymbol{\varphi}_q, \boldsymbol{\varphi}_k\}$ **do**
        Transform to spatial-frequency domain (Eq. (4.6))
        Refine magnitude and phase (Eq. (4.7))
        Reconstruct spatial features (Eq. (4.8))
        Compute fusion weights (Eqs. (4.9)–(4.10))
        Apply feature modulation (Eqs. (4.11)–(4.12))
        Concatenate refined features (Eq. (4.13))
    end for
    // Multi-Head Attention
    Reshape and normalize tensors
    Compute attention map (Eq. (4.3))
    Aggregate features (Eqs. (4.4)–(4.5))
end for
// EGCA in Skip Connections
Compute channel descriptor (Eq. (4.14))
Compute attention weights (Eqs. (4.15)–(4.16))
Refine encoder features (Eq. (4.17))
for each decoder stage **do**
    Upsample decoder features
    Apply EGCA on encoder features
    Fuse encoder and decoder features
    Apply SFT
end for
Apply final convolution to obtain $\mathbf{Y}$

noise-dominated responses. By combining spatial, spatial–frequency, and channel-wise representations, ALKANE improves plume–background separation while maintaining structural consistency across varying environmental conditions. The overall architecture is illustrated in Fig.4.1, and the computational procedure is summarized in Algorithm 1.

4.2.1 Spatial Frequency Transformer (SFT)

Transformer-based architectures are capable of capturing long-range dependencies through SA mechanisms; however, they primarily operate in the spatial domain and do not explicitly exploit frequency-domain characteristics. Consequently, subtle structural variations and diffuse patterns distributed across multiple spatial scales remain difficult to represent, particularly under complex background conditions (36; 115). To overcome this limitation, the proposed SFT incorporates the ASFI module into the attention mechanism, enabling the joint modeling of spatial context and spatial–frequency information for CH_4 plume segmentation. The overall architecture of the SFT is shown in Fig. 4.2.

Let $\mathbf{f} \in \mathbb{R}^{C\times H\times W}$ denote the input feature tensor, where C represents the number of channels, and H and W denote spatial dimensions. The input is first projected into query ($\boldsymbol{\varphi}_q$), key ($\boldsymbol{\varphi}_k$), and value ($\boldsymbol{\varphi}_v$) representations using a 1×1 convolution $C_1(\cdot)$, followed by a 3×3 depthwise convolution $\mathcal{D}(\cdot)$ to incorporate local contextual information:

$$\boldsymbol{\varphi}_q,\ \boldsymbol{\varphi}_k,\ \boldsymbol{\varphi}_v = \mathrm{S}(\mathcal{D}(C_1(\mathbf{f}))) \tag{4.1}$$

where, $\mathrm{S}(\cdot)$ splits the projected features into three distinct components. To embed spatial-frequency characteristics into the attention computation, the query and key representations are processed through the ASFI module (refer to Section 4.2.2), yielding frequency-enhanced features:

$$\boldsymbol{\varphi}'_q = \mathrm{ASFI}(\boldsymbol{\varphi}_q), \qquad \boldsymbol{\varphi}'_k = \mathrm{ASFI}(\boldsymbol{\varphi}_k) \tag{4.2}$$

The value representation $\boldsymbol{\varphi}_v$ is preserved in the spatial domain to maintain content information during aggregation. For multi-head attention, the tensors $\boldsymbol{\varphi}'_q$, $\boldsymbol{\varphi}'_k$, and $\boldsymbol{\varphi}_v$ are reshaped into $\mathbb{R}^{h\times d\times HW}$, where h denotes the number of attention heads, $d = C/h$ represents the

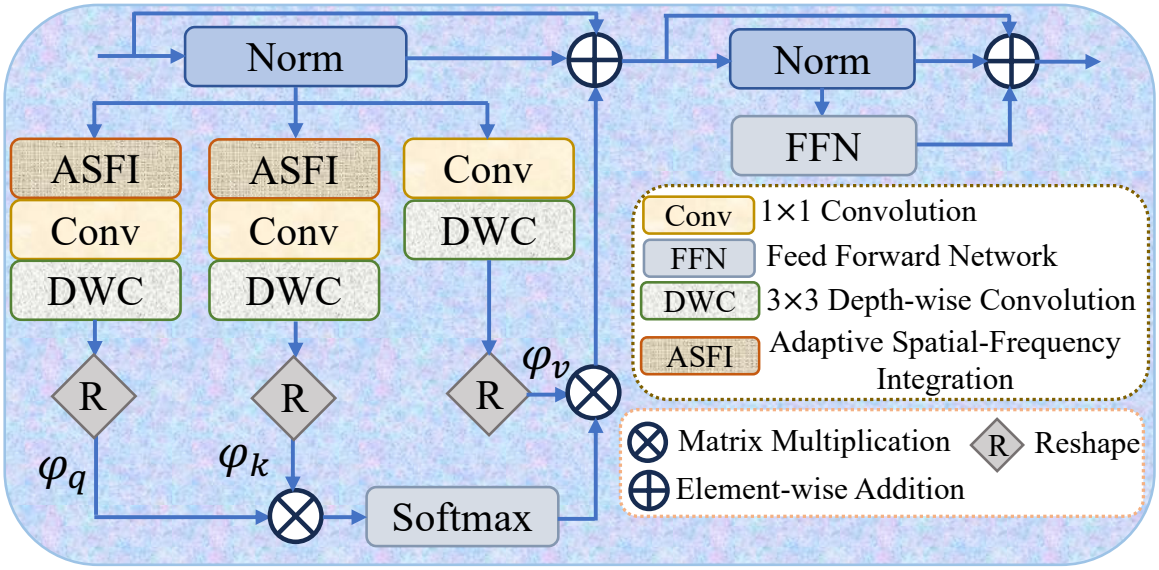


Figure 4.2: Architecture of the proposed SFT block.

channel dimension per head, and HW corresponds to the flattened spatial dimension. To stabilize the attention computation, ℓ_2 normalization is applied along the last dimension. The attention map is then computed as:

$$\mathbf{A} = \text{Softmax}\left(\frac{\varphi'_q \varphi'^{\top}_k}{\tau}\right) \tag{4.3}$$

where $\tau \in \mathbb{R}^{h\times 1\times 1}$ is a learnable temperature parameter controlling the sharpness of the attention distribution (63). The attention weights are applied to the value tensor to obtain aggregated features:

$$\mathbf{Z} = \mathbf{A} \cdot \varphi_v \tag{4.4}$$

Finally, the aggregated representation is reshaped back to $\mathbb{R}^{C\times H\times W}$ and passed through a 1×1 convolution to produce the output:

$$\mathbf{Y} = C_1(\mathbf{Z}) \tag{4.5}$$

By integrating ASFI within the attention mechanism, SFT extends spatial SA to incorporate spatial–frequency representations derived from magnitude and phase components. This enables improved representation of weak and diffuse CH_4 plume structures while maintaining

spatial coherence and boundary continuity under complex environmental conditions.

4.2.2 Adaptive Spatial-Frequency Integration (ASFI)

The detection of CH_4 plumes is particularly challenging due to their weak and diffuse nature, as well as the presence of atmospheric disturbances and complex background clutter (117). Features derived solely in the spatial domain may be insufficient to capture subtle structures distributed across multiple scales, whereas frequency-domain representations often lack precise spatial localization and boundary preservation (77). To mitigate these challenges, the proposed ASFI module jointly exploits spatial-domain and spatial-frequency representations, enabling enhanced plume characterization while maintaining boundary continuity (refer Fig. 4.3).

Let $\mathbf{f} \in \mathbb{R}^{C\times H\times W}$ denote an intermediate feature map. The ASFI module first applies a two-dimensional Fast Fourier Transform (FFT) along the spatial dimensions to obtain a complex representation in the frequency domain. From this representation, magnitude and phase components are extracted as:

$$\rho_{\mathbf{f}} = |\mathrm{FFT}(\mathbf{f})|\,, \quad \varphi_{\mathbf{f}} = \angle\mathrm{FFT}(\mathbf{f}) \tag{4.6}$$

Here, the magnitude component $\rho_{\mathbf{f}}$ reflects dominant structural and textural characteristics, while the phase component $\varphi_{\mathbf{f}}$ retains spatial alignment and geometric structure. These two components are processed separately to capture complementary spatial–frequency information. Specifically, each component is passed through a 1×1 convolution $C_1(\cdot)$ for channel-wise transformation, followed by a GELU activation $\delta(\cdot)$ and a 3×3 depthwise convolution $\mathcal{D}(\cdot)$ to incorporate local contextual information:

$$\varphi_d = \mathcal{D}(\delta(C_1(\varphi_{\mathbf{f}})))\,, \quad \rho_{\mathbf{f}}^d = \mathcal{D}(\delta(C_1(\rho_{\mathbf{f}}))) \tag{4.7}$$

The processed magnitude $\rho_{\mathbf{f}}^d$ and phase φ_d are then recombined and transformed back to the spatial domain using the inverse FFT, $\mathcal{F}^{-1}$, resulting in a frequency-enhanced spatial

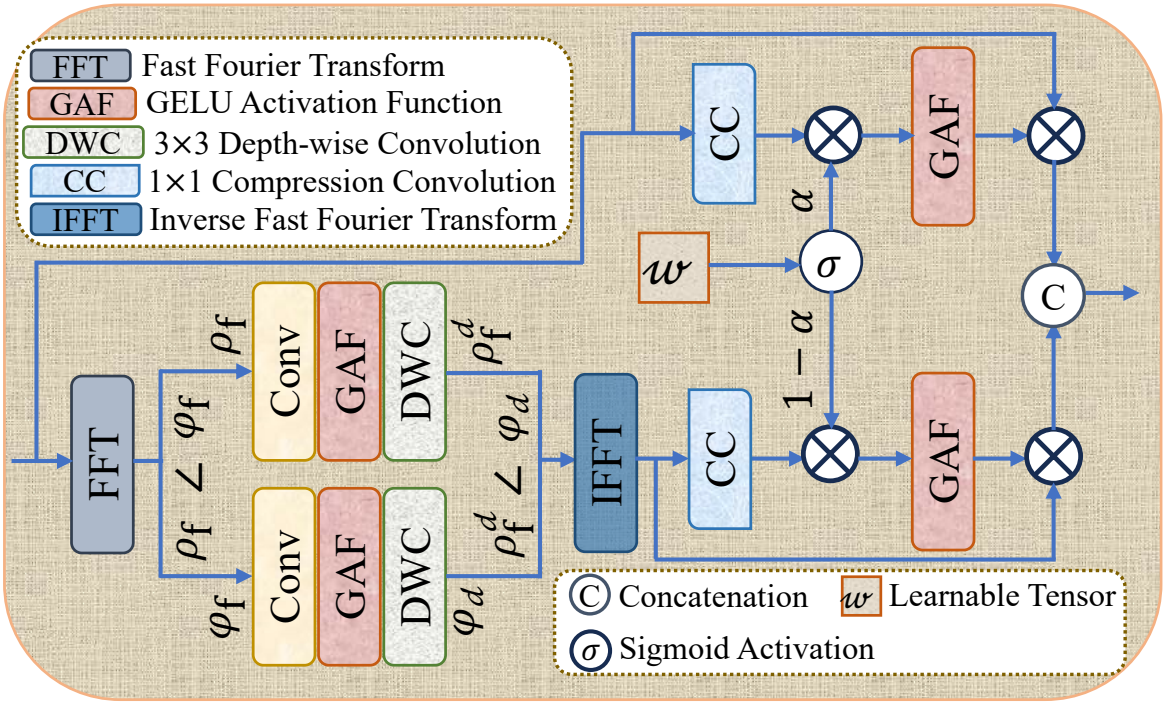


Figure 4.3: Architecture of the proposed ASFI block

representation $\mathbf{f}_d$:

$$\mathbf{f}_d = \mathcal{F}^{-1}(\varphi_d, \rho_{\mathbf{f}}^d) \tag{4.8}$$

To effectively combine spatial and spatial–frequency representations, a learnable scalar parameter $w \in \mathbb{R}$ is introduced to generate adaptive fusion weights:

$$\alpha_1 = \sigma(w) \tag{4.9}$$

$$\alpha_2 = 1 - \alpha_1 \tag{4.10}$$

where $\sigma(\cdot)$ denotes the sigmoid function, ensuring that the weights lie within $(0, 1)$ and sum to unity, thereby forming a convex combination. The parameter w is initialized such that $\alpha_1 \approx \alpha_2$, allowing both branches to contribute equally at the beginning of training, while enabling adaptive reweighting during optimization. Using these weights, the spatial and frequency-enhanced representations are modulated independently:

$$\mathbf{f}'_d = \delta(\alpha_1 \cdot C_1(\mathbf{f}_d)) \odot \mathbf{f}_d \tag{4.11}$$

$$\mathbf{f}' = \delta(\alpha_2 \cdot C_1(\mathbf{f})) \odot \mathbf{f} \tag{4.12}$$

where $\odot$ denotes element-wise multiplication. Finally, the refined features from both branches are combined through channel-wise concatenation:

$$\mathbf{f}'' = \text{Concat}(\mathbf{f}'_d, \mathbf{f}') \tag{4.13}$$

This formulation enables the integration of complementary spatial and spatial–frequency information, facilitating consistent delineation of CH_4 plume structures under complex environmental conditions.

4.2.3 Efficient Global Channel Attention (EGCA)

In encoder–decoder architectures, skip connections play a crucial role in transferring spatial information from the encoder to the decoder across different network depths (79). However, in the context of HSIs, directly forwarding encoder features can introduce redundancy and noise due to strong inter-band correlation and the limited spatial discriminability of several spectral channels (119). To address this limitation, the proposed EGCA module is integrated within the skip connections to adaptively control the flow of encoder features.

Let $\mathbf{f} \in \mathbb{R}^{C\times H\times W}$ denote the encoder feature tensor. The EGCA module first applies global average pooling (GAP) to extract a compact channel-wise descriptor:

$$\mathbf{g} = \text{GAP}(\mathbf{f}) \in \mathbb{R}^{C\times 1\times 1} \tag{4.14}$$

The resulting descriptor $\mathbf{g}$ is reshaped to $\mathbb{R}^{1\times C}$ and passed through a one-dimensional convolution to model local dependencies across channels:

$$\mathbf{s} = \sigma(\text{Conv1D}(\mathbf{g})) \tag{4.15}$$

where $\mathbf{s} \in (0, 1)^{1\times C}$ represents the learned channel-wise attention weights obtained after applying the sigmoid activation function. Following the efficient channel attention formulation

(78), the kernel size of the 1D convolution is adaptively determined as:

$$k = \text{odd}\big(\eta \log_2 C + \beta\big) \tag{4.16}$$

where $\eta = 0.5$ regulates the rate at which the receptive field grows with respect to the channel dimension, $\beta = 0.5$ defines the lower bound for the kernel size, and $\text{odd}(\cdot)$ ensures that the kernel size is an odd integer, maintaining symmetry in the convolution operation. This adaptive formulation enables the modeling of appropriate channel interactions without introducing unnecessary complexity. The attention weights $\mathbf{s}$ are then reshaped to $\mathbb{R}^{C\times1\times1}$ and applied to the original feature tensor through channel-wise multiplication:

$$\mathbf{Y} = \mathbf{f} \odot \mathbf{s} \tag{4.17}$$

where $\odot$ denotes element-wise multiplication. The resulting feature map $\mathbf{Y} \in \mathbb{R}^{C\times H\times W}$ emphasizes spectrally informative and spatially relevant channels while reducing the influence of redundant or noise-affected responses. By incorporating EGCA into the skip connections, the decoder receives refined feature representations, which supports improved plume–background distinction and more consistent boundary delineation in CH_4 plume segmentation.

4.3 Experimental Results

4.3.1 Experimental Settings

4.3.1.1 Datasets

We assess the proposed method on three commonly utilized HSI datasets, namely AVIRIS-NG, WorldView-3 (WV-3), and EMIT, which collectively represent variations in spatial and spectral resolutions. In accordance with the plume categorization strategy adopted in (31; 77), CH_4 plumes with estimated emission rates greater than 1000 kg/h are categorized as “strong”, whereas those with lower emission rates are classified as “weak”. The AVIRIS-NG dataset provides spectral measurements spanning 380-2510 nm with a

Table 4.1: Quantitative comparison on the AVIRIS-NG dataset for strong and weak methane plumes

Method	Strong Plumes					Weak Plumes				
	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
MethaNet (70)	70.04	69.79	62.13	46.89	0.5454	51.15	69.48	60.70	41.49	0.3180
U-Plume (72)	72.32	74.28	74.84	62.58	0.8976	54.39	69.11	64.12	47.36	0.2665
CH4Net (71)	74.67	75.67	70.83	58.10	0.4507	62.05	57.98	65.02	45.67	0.2017
HyperSTARCOP (31)	82.19	65.86	68.40	56.14	0.2194	73.07	43.22	59.35	37.01	0.1055
Gasformer (35)	79.17	79.89	77.45	65.31	0.5964	64.51	80.92	72.46	60.48	0.2233
MethaneMapper (73)	81.03	76.29	78.14	65.21	0.1461	73.02	82.35	76.81	64.47	0.1840
FSDINet (77)	87.75	80.13	77.05	65.56	0.1036	78.94	78.25	77.80	65.20	0.0771
ALKANE	**90.04**	**86.79**	**85.13**	**76.99**	**0.0013**	**80.13**	**83.02**	**83.41**	**71.18**	**0.0005**

spectral resolution of 5 nm and spatial dimensions of 512×512 pixels (120). Following the methodology in (31), additional training samples are generated using CH_4 plume annotations derived from the mag1c product (121), which are subsequently refined through manual correction. This process produces 167,825 training patches of size 128×128 with a 64-pixel overlap, along with 342 test images of size 512×512.

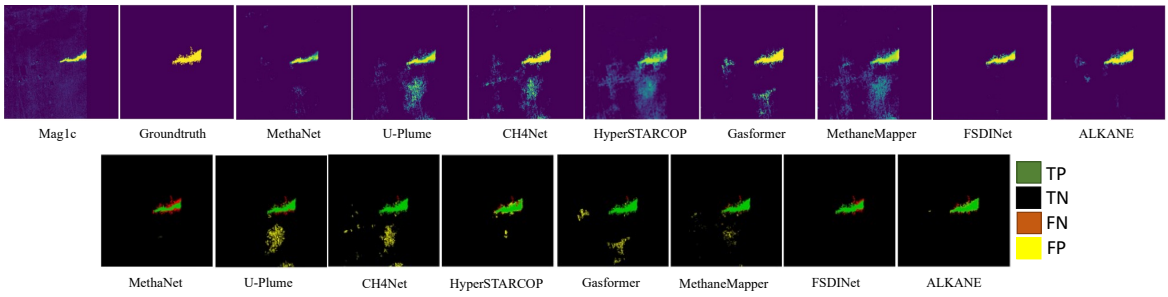


Figure 4.4: Visual comparison of methane plume segmentation results on strong plumes from the AVIRIS-NG dataset.

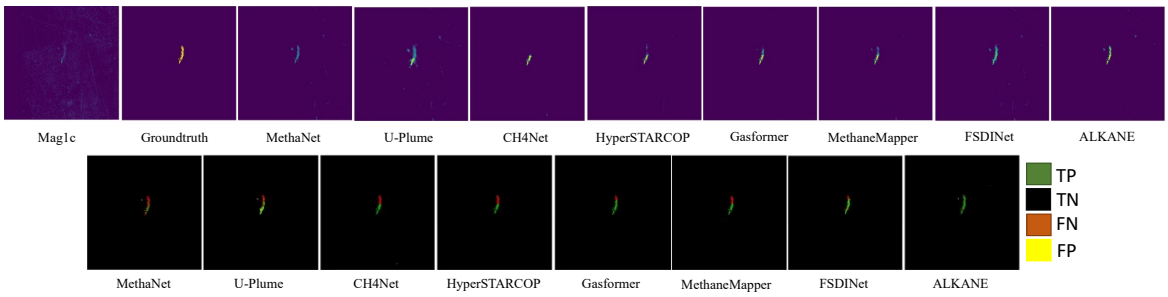


Figure 4.5: Visual comparison of methane plume segmentation results on weak plumes from the AVIRIS-NG dataset.

For the WV-3 dataset, hyperspectral data are synthetically generated by convolving

Table 4.2: Quantitative comparison on the WV-3 dataset for strong and weak methane plumes

Method	Strong Plumes					Weak Plumes				
	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
MethaNet (70)	9.75	75.20	16.43	9.39	14.7511	0.99	55.36	2.35	0.97	10.1400
U-Plume (72)	13.83	65.65	20.55	11.99	6.7680	1.53	64.69	3.46	1.51	6.2530
CH4Net (71)	16.80	50.11	21.65	13.14	4.4825	2.45	53.25	5.29	2.34	3.8310
MultiSTARCOP (31)	22.84	25.77	23.01	13.10	1.3340	2.82	29.29	8.38	2.58	1.2451
Gasformer (35)	23.14	21.32	19.81	10.70	1.0179	2.94	25.36	9.72	2.44	1.1030
MethaneMapper (73)	23.07	17.14	20.10	9.65	1.1523	3.87	22.65	12.46	3.24	0.8380
FSDINet (77)	31.83	23.09	27.81	13.69	0.8603	5.73	32.49	11.49	4.28	0.7377
ALKANE	**34.26**	**77.18**	**34.98**	**20.90**	**0.0160**	**8.16**	**68.12**	**15.78**	**7.04**	**0.0015**

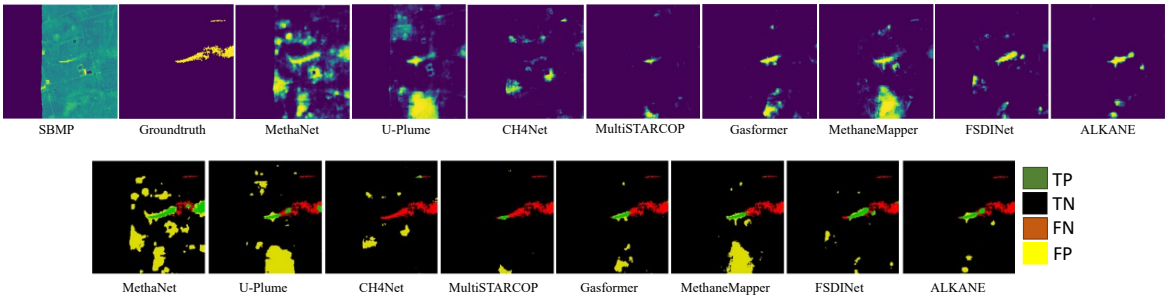


Figure 4.6: Visual comparison of methane plume segmentation results on strong plumes from the WV-3 dataset.

AVIRIS-NG measurements with the WV-3 spectral response function (122). The same data partitioning, patch extraction strategy, and preprocessing pipeline are retained. However, plume annotations are obtained from Single-Band Multipass (SBMP) ratio products (123) instead of mag1c, consistent with the practices followed in (31; 77). The EMIT dataset captures spectral reflectance in the 400–2500 nm range with a spectral sampling interval of 10 nm, consisting of 512×512 pixel scenes collected from 11 methane-emitting sites and 9 non-emitting sites. In line with existing CH_4 plume segmentation studies (31; 77), we adopt the commonly used Level-1 reflectance terminology for EMIT data without explicitly analyzing the official product definitions. This dataset is utilized solely for evaluating cross-sensor generalisation, without performing any fine-tuning.

4.3.1.2 Implementation Details

All experiments were performed using PyTorch 2.6.0 with CUDA 12.4 on a system equipped with an AMD Ryzen 5 5600X processor, 16 GB of RAM, and an NVIDIA

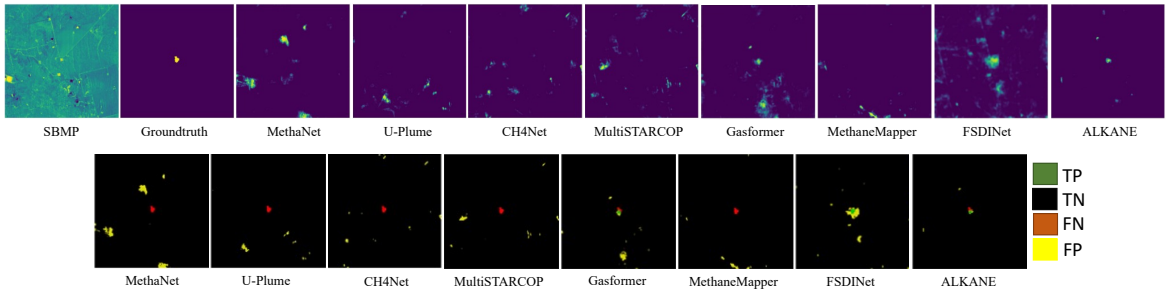


Figure 4.7: Visual comparison of methane plume segmentation results on weak plumes from the WV-3 dataset.

RTX 3050 GPU. The training process utilizes Binary Cross-Entropy (BCE) loss (124) to quantify pixel-wise discrepancies between the predicted probability map and the corresponding binary ground-truth mask. The model is trained with a batch size of 8 for 15 epochs using the AdamW optimizer. The initial learning rate is set to 1×10^{-4} and gradually reduced to 1×10^{-6} following a cosine annealing schedule. To accommodate variations in spatial and spectral characteristics across different sensors, we follow established practices and employ mag1c-based band inputs for AVIRIS-NG and SBMP-based band inputs for WV-3, consistent with prior CH_4 plume segmentation studies (31; 77). The preprocessing procedures for mag1c and SBMP band selection adhere to standard implementations reported in previous works and are publicly available[1]. During training, data augmentation is applied to improve generalization, including random rotations as well as horizontal and vertical flipping operations.

4.3.1.3 Evaluation Metrics

To quantitatively evaluate the segmentation performance, five widely adopted metrics are considered, namely Mean Precision (MP), Mean Recall (MR), Mean F1-score (MF1), Mean Intersection over Union (MIoU), and Mean False Positive Rate (MFPR) (117). These metrics provide complementary insights into detection quality, overlap accuracy, and error characteristics. For a comprehensive assessment, each metric is computed for individual classes and then averaged across all classes. The mathematical definitions of these metrics

[1] https://github.com/spaceml-org/STARCOP

Table 4.3: Quantitative comparison on the EMIT dataset for strong and weak methane plumes

Method	Strong Plumes					Weak Plumes				
	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
MethaNet (70)	50.25	26.59	30.08	23.90	0.0848	8.17	28.75	11.64	6.41	0.1189
U-Plume (72)	56.93	36.26	37.72	29.07	0.0785	20.07	6.29	9.04	3.28	0.0531
CH4Net (71)	60.42	35.48	33.31	23.93	0.1814	12.14	23.20	20.80	9.57	0.0309
HyperSTARCOP (31)	64.48	44.45	45.87	33.95	0.1313	21.46	15.88	18.78	9.23	0.0376
Gasformer (35)	67.08	58.77	52.82	38.70	0.4186	4.94	27.36	11.97	4.26	0.1434
MethaneMapper (73)	73.50	62.35	58.01	43.72	0.4511	7.81	21.92	14.81	5.57	0.1355
FSDINet (77)	74.26	72.39	60.72	47.53	0.1365	24.86	48.13	21.16	10.01	0.1133
ALKANE	**76.52**	**78.11**	**64.13**	**52.47**	**0.0098**	**27.01**	**51.64**	**28.33**	**13.17**	**0.0056**

are expressed as:

$$\text{MP} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FP_i} \tag{4.18}$$

$$\text{MR} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FN_i} \tag{4.19}$$

$$\text{MF1} = \frac{1}{C}\sum_{i=1}^{C}\frac{2 \cdot TP_i}{2 \cdot TP_i + FP_i + FN_i} \tag{4.20}$$

$$\text{MIoU} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{4.21}$$

$$\text{MFPR} = \frac{1}{C}\sum_{i=1}^{C}\frac{FP_i}{FP_i + TN_i} \tag{4.22}$$

In the above expressions, TP_i, FP_i, FN_i, and TN_i denote the number of true positives, false positives, false negatives, and true negatives corresponding to the i^{th} class, respectively, while C represents the total number of classes. Together, these metrics facilitate a balanced evaluation of segmentation outcomes by jointly considering detection capability, overlap accuracy, and the tendency of the model to produce false alarms.

4.3.2 Comparative Analysis

The proposed ALKANE framework is evaluated against several existing approaches, including CNN-based methods such as MethaNet (70), U-Plume (72), CH4Net (71), HyperSTARCOP, MultiSTARCOP (31), and FSDINet (77), as well as transformer-based methods

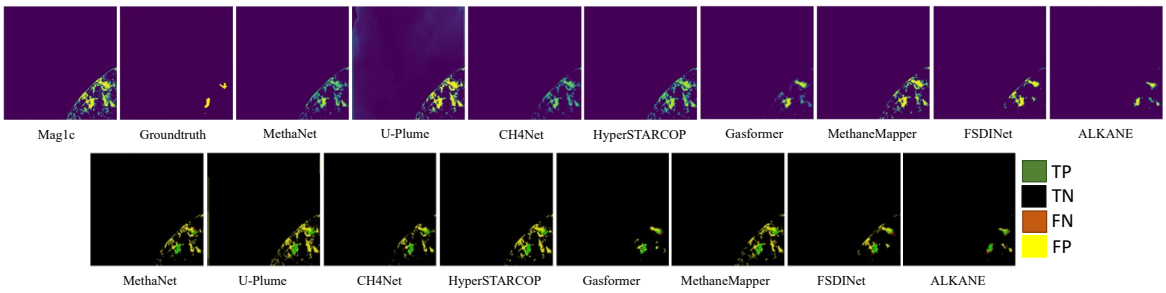


Figure 4.8: Visual comparison of methane plume segmentation results on strong plumes from the EMIT dataset.

including Gasformer (35) and MethaneMapper (73). The evaluation is conducted across three benchmark datasets. The comparison is carried out through quantitative analysis using standard segmentation metrics, complemented by qualitative visual inspection of segmentation outputs. In addition, cross-sensor generalization is examined by evaluating the trained models on the unseen EMIT dataset, providing insight into their ability to handle variations in spatial and spectral characteristics across different sensors.

4.3.2.1 Quantitative Comparison

As presented in Tables 4.1, 4.2, and 4.3, the proposed ALKANE framework demonstrates superior performance for both strong and weak plume categories across all evaluation metrics on both hyperspectral and multispectral datasets. To assess the consistency of the results, all experiments were repeated five times using different random seeds while maintaining identical training–testing splits, preprocessing steps, and optimization settings. On the AVIRIS-NG dataset, the strongest competing method, *FSDINet*, reports an average performance of MP = 87.78 ± 0.31, MR = 80.09 ± 0.28, MF1 = 77.03 ± 0.26, and MIoU = 65.54 ± 0.34 for strong plumes, with MFPR = 0.104 ± 0.003. For weak plumes, it achieves MP = 78.91 ± 0.29, MR = 78.22 ± 0.31, MF1 = 77.77 ± 0.27, and MIoU = 65.17 ± 0.33, along with MFPR = 0.077 ± 0.002.

In comparison, ALKANE yields higher mean values across all metrics, along with reduced variation over repeated trials. For strong plumes, it attains MP = 90.02 ± 0.23, MR = 86.70 ± 0.30, MF1 = 85.10 ± 0.30, and MIoU = 76.91 ± 0.34, while maintaining a notably low MFPR = 0.0014 ± 0.0002. For weak plumes, the method achieves MP = 80.12 ± 0.21, MR

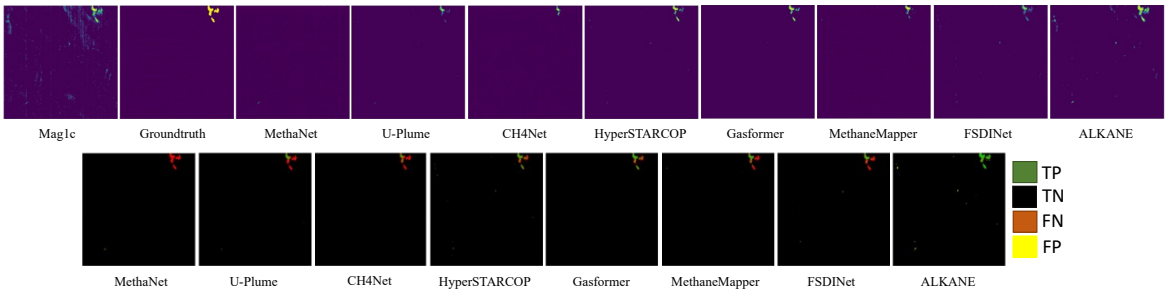


Figure 4.9: Visual comparison of methane plume segmentation results on weak plumes from the EMIT dataset.

= 82.96 ± 0.34, MF1 = 83.41 ± 0.34, and MIoU = 71.12 ± 0.39, with MFPR further reduced to 0.0005 ± 0.0001. Furthermore, a segmentation-to-emission analysis conducted on the AVIRIS-NG test set shows a Pearson correlation coefficient (125) of R^2 = 0.82 between segmentation-derived emission estimates and archival values, indicating the suitability of the proposed approach for supporting emission-rate estimation.

4.3.2.2 Qualitative Comparison

Visual comparisons presented in Figs. 4.4, 4.5, 4.6, and 4.7 indicate that `ALKANE` produces more complete and well-defined plume regions, with reduced false positives and false negatives relative to existing methods. Overall, the results suggest that the proposed framework can capture both global contextual relationships and fine-scale spatial structures, even under challenging conditions such as low-resolution multispectral data and weak plume scenarios.

4.3.2.3 Zero-Shot Generalization on EMIT

To examine cross-sensor generalization, the trained model is directly evaluated on the EMIT dataset without any fine-tuning. As shown in Table 4.3 and Figs. 4.8, 4.9, the proposed method outperforms all competing approaches for both strong and weak plume categories, demonstrating clear improvements over the strongest baseline, FSDINet. These results indicate that the proposed architecture maintains consistent plume representation even under unseen spectral and spatial configurations.

Table 4.4: Ablation study on EGCA kernel-size scaling parameters on the AVIRIS-NG dataset (Strong Plumes).

(η, β)	**MP** ↑	**MR** ↑	**MF1** ↑	**MIoU** ↑	**MFPR** ↓
(0.25, 0.5)	88.71	84.92	84.18	75.46	0.0019
(0.5, 0.0)	89.12	85.43	84.63	76.01	0.0016
(0.5, 0.5)	**90.04**	**86.79**	**85.13**	**76.99**	**0.0013**
(0.75, 0.5)	89.81	86.11	84.91	76.63	0.0015
(1.0, 0.5)	89.34	85.62	84.57	76.18	0.0018

4.3.3 Model Complexity

This section examines the computational characteristics of the proposed framework compared with existing methods, including model size (in millions of parameters, M), computational cost in GFLOPs (G), inference latency, throughput, and GPU memory usage. Lightweight architectures such as MethaNet (M = 1.05, G = 0.08) and CH4Net (M = 0.21, G = 2.87) require relatively low computational resources; however, their segmentation performance remains limited. Methods such as U-Plume (M = 1.94, G = 13.79) and HyperSTARCOP (M = 6.63, G = 0.85) introduce moderate increases in complexity, but their performance degrades under more challenging scenarios. Transformer-based approaches, including Gasformer (M = 3.65, G = 9.95) and MethaneMapper (M = 78.65, G = 86.50), provide improved segmentation outcomes, although this is accompanied by substantially higher computational and memory requirements, leading to increased inference time.

FSDINet (M = 8.40, G = 6.34) offers competitive segmentation performance; however, it incurs higher memory consumption and exhibits limitations in boundary delineation. In contrast, the proposed `ALKANE` framework achieves a model size of M = 3.80 and computational cost of G = 5.25, which are lower than those of MethaneMapper and FSDINet. In terms of runtime characteristics, `ALKANE` achieves an average inference latency of approximately 18.6 ms, corresponding to a throughput of 53.8 frames per second, while maintaining a peak GPU memory usage of approximately 1.4 GB during inference. The training duration is approximately 17 hours, compared to 22 hours for FSDINet under identical experimental settings. These observations indicate that the proposed framework maintains a balanced trade-off between computational cost and segmentation performance, making it suitable for practical CH_4 plume segmentation tasks.

Table 4.5: Ablation study of key components on AVIRIS-NG dataset (Strong Plumes).

Components		MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
ASFI	EGCA					
×	×	85.42	79.68	80.25	70.35	0.0031
✓	×	87.34	81.59	82.23	73.15	0.0022
×	✓	86.85	81.04	81.82	72.69	0.0025
✓	✓	**90.04**	**86.79**	**85.13**	**76.99**	**0.0013**

4.3.4 Ablation Study

The contribution of each proposed component is analyzed on the AVIRIS-NG dataset through a series of controlled ablation experiments, as summarized in Tables 4.4, 4.5, and 4.6. In this analysis, ASFI introduces spatial–frequency integration within the attention mechanism, while EGCA performs channel-wise feature modulation along the skip connections. As reported in Table 4.5, each component individually contributes to performance improvement. In particular, EGCA reduces the influence of redundant spectral responses, whereas ASFI enhances plume boundary delineation. When ASFI is removed, the model reduces to a conventional SA formulation, resulting in a noticeable decline in performance, which emphasizes the role of frequency-domain information. The complete `ALKANE` framework, incorporating both ASFI and EGCA, achieves the highest overall performance, indicating that these components provide complementary benefits.

Table 4.6: Impact of frequency-domain components in ASFI (Strong Plumes).

Frequency Components		MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
Magnitude	Phase					
×	×	86.85	81.04	81.82	72.69	0.0025
✓	×	88.15	83.22	83.01	74.60	0.0019
×	✓	87.32	82.07	82.45	73.58	0.0021
✓	✓	**90.04**	**86.79**	**85.13**	**76.99**	**0.0013**

To further analyze the design of ASFI, the magnitude and phase components are evaluated independently, as shown in Table 4.6. The use of either component alone improves performance compared to the spatial-only baseline, while their combined utilization yields the best results. This observation suggests that the magnitude component captures global structural characteristics of the plume, whereas the phase component preserves spatial alignment and boundary continuity. Additionally, the sensitivity of EGCA to the adaptive kernel-size

parameters in Eq. 4.16 is investigated. As presented in Table 4.4, moderate kernel sizes provide a balance between effective channel interaction and suppression of irrelevant responses, whereas larger kernels do not consistently improve performance. Based on this empirical evaluation, the default parameter setting ($\eta = 0.5,\ \beta = 0.5$) is adopted for all experiments.

4.4 Summary

This chapter presented ALKANE, an attention-based framework designed for HSI CH_4 plume segmentation. The proposed method addresses key challenges in HSI data, including inter-band redundancy, weak plume signatures, and background interference, by integrating spatial and spatial–frequency representations within a unified architecture. An SFT was introduced to extend conventional SA by incorporating frequency-domain cues into the feature interaction process. By embedding the ASFI module within the attention mechanism, the model jointly utilizes magnitude and phase information to capture both structural patterns and spatial continuity of CH_4 plumes. This design enables improved representation of diffuse plume regions and clearer delineation of plume boundaries.

To further refine feature propagation, an EGCA module was incorporated within skip connections. This module selectively emphasizes spectrally relevant channels while suppressing redundant responses, thereby improving feature fusion between encoder and decoder stages. The proposed framework was evaluated on multiple datasets, including AVIRIS-NG, WV-3, and EMIT, covering diverse sensing conditions. Experimental results demonstrated consistent improvements over existing CNN-based and transformer-based approaches across both strong and weak plume scenarios. Additional analyses, including ablation studies, confirmed the complementary roles of ASFI and EGCA, as well as the importance of jointly modeling magnitude and phase components. Cross-sensor evaluation further highlighted the model's ability to maintain performance under unseen spectral characteristics.

In summary, this chapter establishes the importance of integrating spatial and frequency-domain representations along with selective feature fusion strategies for improved CH_4 plume segmentation in hyperspectral imagery.

Chapter 5

Supervised Contrastive Learning and Inpainting for Hyperspectral Denoising and Methane Plume Segmentation

5.1 Introduction

HSI captures rich spectral information across numerous contiguous wavelength bands, enabling detailed characterization of materials and atmospheric constituents. This capability makes HSI particularly valuable for environmental monitoring tasks such as CH_4 plume detection and analysis. However, HSI data are often degraded during acquisition due to atmospheric interference, sensor noise, and photon-related distortions, which can obscure subtle spectral signatures and degrade downstream processing (5; 6; 126; 127; 128).

Improving the quality of HSI data through denoising is therefore an essential preprocessing step for CH_4 plume segmentation (129). Traditional denoising methods rely on handcrafted priors to model spatial and spectral correlations, but their performance is often limited when applied to complex or unseen conditions (32; 130; 131). DL based methods have shown improved capability in learning joint spatial–spectral representations, yet their performance is strongly influenced by the availability of large-scale annotated datasets (11; 58; 60). In practice, acquiring clean and well-annotated HSI data is challenging, leading

to limited training samples and reduced generalization in real-world scenarios (132; 133).

To address this limitation, CL has emerged as a representation learning strategy that leverages relationships between samples by bringing similar instances closer while pushing dissimilar ones apart in the feature space (40; 45). However, standard CL methods do not use label information, which limits their effectiveness for supervised tasks such as denoising and segmentation (87; 88). SCL extends this paradigm by incorporating class labels, enabling the model to learn more discriminative representations by grouping samples of the same class and separating different classes (86). Despite its success in HSI classification, its application to HSI denoising and CH_4 plume segmentation remains limited (38; 41; 42; 43; 44; 134).

In parallel, image inpainting provides a complementary perspective by reconstructing missing or corrupted regions using contextual information (46). While inpainting has been explored in HSI, it is typically treated as a standalone or auxiliary restoration task (47; 48; 135). Its potential to support RL, particularly in generating structured variations of data for contrastive objectives, has not been fully investigated. Motivated by these observations, this chapter introduces a unified framework that combines SCL and inpainting to improve HSID and CH_4 plume segmentation. The key idea is to use inpainting to generate meaningful variations of input data that preserve spatial–spectral consistency, which are then leveraged within an SCL framework to enhance feature representation. By jointly integrating these components, the proposed method encourages the model to better distinguish plume structures from background noise while maintaining spectral integrity.

To this end, we propose `AUGMENT` (supervised contr**A**stive learning and inpainting for hyperspectral denoising and methane pl**U**me se**GMENT**ation), a framework that incorporates contrastive representation learning with reconstruction-based objectives. This design enables simultaneous learning of denoising and segmentation tasks while addressing the challenge of limited training data in HSI. Our key contributions are as follows:

- We propose `AUGMENT`, a unified framework that integrates SCL and HSI inpainting to enrich training data and improve spatial-spectral representation learning for HSI analysis.

- We incorporate HSI inpainting within the SCL paradigm to generate structured vari-

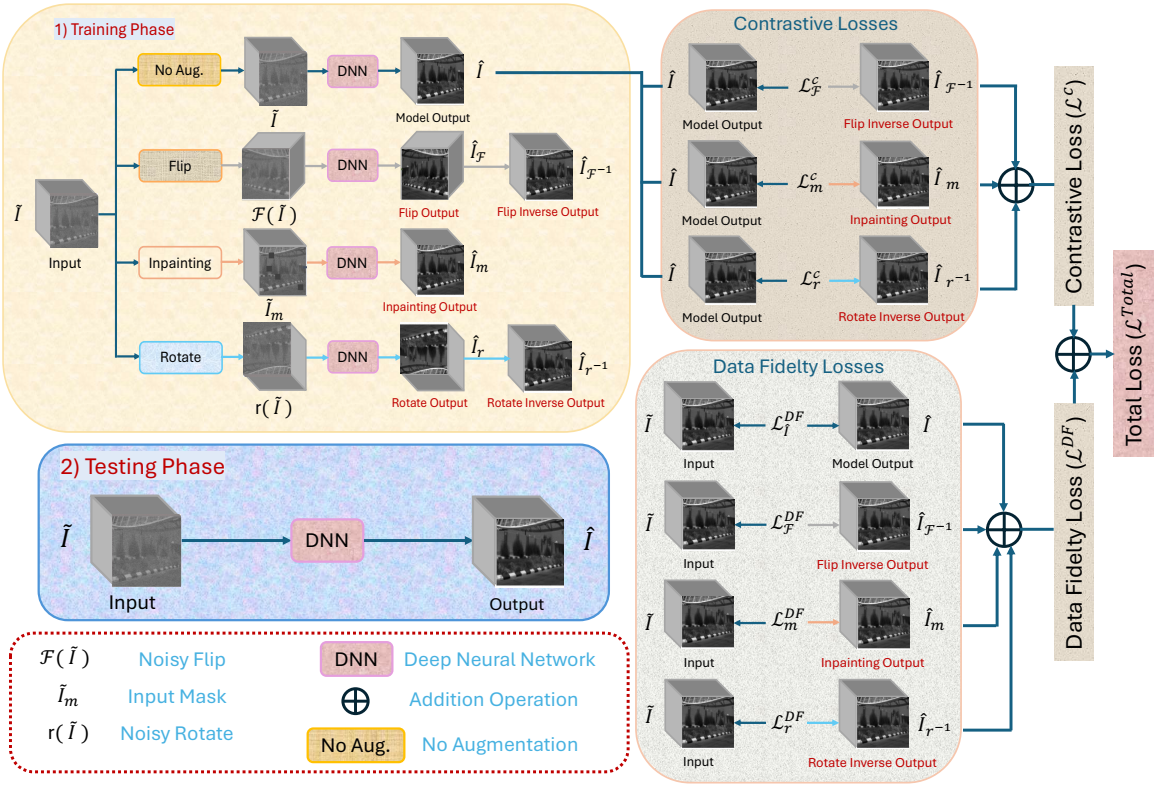


Figure 5.1: Overview of the proposed AUGMENT framework.

ations of input data. This integration enables the model to exploit contextual spatial-spectral relationships and improves RL performance under limited training data.

- We demonstrate that the proposed framework can be integrated into existing HSID pipelines and improves plume segmentation performance across diverse datasets and sensing conditions.

5.2 Proposed Methodology

HSI plays an important role in improving data quality for downstream tasks such as CH_4 plume segmentation. However, the presence of complex noise patterns and limited training samples often limits deep models' ability to learn meaningful spatial-spectral representations. In such cases, relying solely on reconstruction-based objectives may not sufficiently constrain the learning process. To address this, we introduce AUGMENT, which incorporates SCL along

with inpainting-based augmentation to encourage consistent and informative feature learning from noisy HSI (see Fig. 5.1).

Let $\tilde{I}$ denote a noisy HSI obtained by corrupting a clean image I with additive noise N:

$$\tilde{I} = I + N \tag{5.1}$$

where $\tilde{I}, I$, and $N \in \mathbb{R}^{H \times W \times B}$ share identical spatial and spectral dimensions. Here, H and W denote spatial dimensions, and B represents the number of spectral bands. The objective is to estimate a restored image $\hat{I}$ from $\tilde{I}$ such that the underlying spatial-spectral structure is preserved. This is achieved using a DNN, $D(\cdot)$:

$$\hat{I} = D(\tilde{I}) \tag{5.2}$$

The restored representation is further utilized for CH_4 plume segmentation, where discriminative spatial-spectral features are required to separate plume regions from the background. However, DNN models require large amounts of data to learn more effective features (39). Unfortunately, there is a scarcity of image sample in the HSI domain. To improve feature learning under limited data conditions, we incorporate SCL along with inpainting-driven augmentation. The overall objective combines a contrastive consistency term with a reconstruction-based data fidelity term.

5.2.1 Contrastive Loss

To construct multiple views of the same input, three transformations are applied to $\tilde{I}$: vertical flipping $\mathcal{F}(\cdot)$, rotation $r(\cdot)$, and masking $m(\cdot)$ (see Fig. 5.1). These generate augmented samples $\tilde{I}_{\mathcal{F}}$, $\tilde{I}_r$, and $\tilde{I}_m$. Although flipping and rotation are widely adopted data augmentation techniques, within the proposed `AUGMENT` framework they play a more specialized role for HSI. These transformations modify only the spatial arrangement of pixels while preserving the spectral characteristics of each pixel, thereby retaining the inherent spectral–spatial information present in HSIs. As a result, the SCL module is encouraged to learn representations that remain consistent under geometric variations, improving feature

robustness and generalization.

To further enrich the learning process, masking-based inpainting is incorporated as an additional augmentation strategy. By masking localized spectral–spatial regions, the network is required to reconstruct the missing information from the surrounding context. This process promotes a deeper understanding of spectral–spatial dependencies and strengthens the model's ability to capture meaningful contextual relationships. The combined use of flipping, rotation, and inpainting, therefore, provides diverse yet structurally consistent views of the same HSI sample, facilitating more effective CL while preserving the integrity of the original data. Consequently, `AUGMENT` learns more discriminative and robust spectral–spatial representations, leading to enhanced performance in both HSID and methane plume segmentation tasks. Formally, these transformations can be expressed as follows:

5.2.1.1 Flipping

Vertical flipping is applied to alter the spatial arrangement of the HSI while preserving its spectral information. This transformation reflects the image along the horizontal axis, effectively reversing the order of rows. The flipped image is defined as:

$$\tilde{I}_{\mathcal{F}} = \mathcal{F}(\tilde{I}) \tag{5.3}$$

For a pixel located at spatial position (h, w) in the b^{th} spectral band, the transformed intensity is given by:

$$\tilde{I}_{\mathcal{F}}[h, w, b] = \tilde{I}[H - h + 1, w, b] \tag{5.4}$$

where H denotes the image height. This transformation preserves the spectral signature of each pixel while modifying its spatial position, encouraging the model to learn features that are invariant to vertical shifts.

5.2.1.2 Rotation

Rotation is used to further diversify spatial configurations by rotating the HSI while keeping spectral responses unchanged. In this work, we consider a rotation of 180°, which

corresponds to reversing both spatial dimensions. The rotated image is expressed as:

$$\tilde{I}_r = r(\tilde{I}) \tag{5.5}$$

For a pixel at location (h, w) in the b^{th} band, the rotated intensity is computed as:

$$\tilde{I}_r[h, w, b] = \tilde{I}[2h_c - h, 2w_c - w, b] \tag{5.6}$$

where (h_c, w_c) represents the center of the image. This transformation preserves the underlying spatial-spectral structure while presenting it in a different orientation, enabling the network to learn consistent representations under geometric variations.

5.2.1.3 Masking (Inpainting)

To further enrich the diversity of positive samples used in SCL, `AUGMENT` incorporates a masking-based transformation to simulate missing or corrupted regions in HSIs. Unlike geometric augmentations, masking introduces localized spectral-spatial occlusions, forcing the network to recover the missing information from the surrounding context. This auxiliary inpainting task encourages the learning of more discriminative and robust spectral-spatial representations, which are beneficial for both HSID and CH_4 plume segmentation.

Specifically, for each spectral band (b), (k) square patches of size $(m \times m)$ are randomly selected, and the pixel values within these regions are replaced by the minimum intensity value of the corresponding spectral band. The masked HSI $(\tilde{I}_m)$ is defined as:

$$\tilde{I}_m[h, w, b] = \begin{cases} \min(\tilde{I}[:, :, b]) & \text{if } [h, w] \in P_b^i,\ \forall i \in [1, k] \\ \tilde{I}[h, w, b] & \text{otherwise} \end{cases} \tag{5.7}$$

where (P_b^i) denotes the set of pixel locations belonging to the (i^{th}) masked patch in the (b^{th}) spectral band, and $(\min(\tilde{I}[:, :, b]))$ represents the minimum intensity value within that band. By intentionally removing local information, the network is encouraged to infer the missing spectral and spatial content from neighbouring regions, thereby strengthening its ability to model contextual dependencies and inter-band correlations. Consequently,

the learned representations become more robust to spectral distortions, noise, and incomplete observations. The original noisy image and its transformed variants are subsequently processed by the denoising network, $D(\cdot)$ to obtain their corresponding restored outputs:

$$\hat{I}_{\mathcal{F}} = D(\tilde{I}_{\mathcal{F}}), \qquad \hat{I}_r = D(\tilde{I}_r), \qquad \hat{I}_m = D(\tilde{I}_m), \tag{5.8}$$

where $(\hat{I}_{\mathcal{F}})$, $(\hat{I}_r)$, and $(\hat{I}_m)$ denote the outputs corresponding to the flipped, rotated, and masked inputs, respectively. It is important to note that flipping and rotation are covariant transformations. A covariant transformation preserves the underlying semantic and structural content of an image while altering its spatial arrangement or orientation (136). Consequently, although the visual appearance changes, the intrinsic spectral information remains unchanged. To establish pixel-wise correspondence between the denoised outputs obtained from transformed inputs and those of the original input, the inverse transformations are applied to the corresponding outputs. The aligned outputs are expressed as $(\hat{I}_{\mathcal{F}^{-1}})$ and $(\hat{I}_{r^{-1}})$, respectively:

$$\hat{I}_{\mathcal{F}^{-1}} = \mathcal{F}^{-1}(\hat{I}_{\mathcal{F}}), \qquad \hat{I}_{r^{-1}} = r^{-1}(\hat{I}_r). \tag{5.9}$$

Unlike flipping and rotation, the masking operation does not preserve geometric correspondence and is therefore not covariant. As a result, no inverse transformation is required for the masked output. The main objective of `AUGMENT` is to ensure that the denoising network produces consistent representations irrespective of the transformations applied to the input. Ideally, a denoising model should generate equivalent restored outputs for the same spectral content regardless of spatial orientation or localized occlusions. However, in practice, DL models often exhibit epistemic uncertainty due to limited training data and incomplete feature learning (132; 133). Such uncertainty may cause inconsistent predictions when transformed versions of the same image are presented.

To alleviate this issue, `AUGMENT` incorporates SCL to explicitly encourage agreement among the denoised outputs obtained from different transformed views. Flipping and rotation promote orientation-invariant spectral-spatial representations, while masking encourages the network to recover missing contextual information and strengthen feature discrimination. This consistency objective is enforced through the contrastive loss (L_{CL}), defined as

$$L_{CL} = -\Big(\lambda_1 \cdot \text{sim}(\hat{I}, \hat{I}_{\mathcal{F}^{-1}}) + \lambda_2 \cdot \text{sim}(\hat{I}, \hat{I}_{r^{-1}}) + \lambda_3 \cdot \text{sim}(\hat{I}, \hat{I}_m)\Big) \tag{5.10}$$

where λ_1, λ_2, and λ_3 are weighting coefficients controlling the contribution of each transformed view, and $(\text{sim}(\cdot, \cdot))$ denotes the similarity measure between two denoised outputs. By maximizing agreement among multiple transformed representations, `AUGMENT` encourages the learning of transformation-invariant spectral-spatial features, reduces epistemic uncertainty, and improves the robustness of the denoising process. Kindly note that the weighting coefficients are set to $(\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3})$, ensuring equal contributions from flipping, rotation, and masking during contrastive optimization (see Sec. 5.4 for further discussion).

5.2.2 Data Fidelity Loss

While the L_{CL} enforces consistency between the outputs corresponding to the original and transformed inputs, it does not explicitly ensure that the restored image is close to the underlying clean hyperspectral signal. In the context of HSID, such a constraint is essential to preserve the true spatial-spectral characteristics, which directly influence downstream CH_4 plume segmentation performance.

To address this, we incorporate a data fidelity term that guides the network to reconstruct outputs that are aligned with the ground-truth clean image I_{gt}. This loss enforces pixel-wise and spectral consistency between the predicted outputs and the reference image. Formally, the data fidelity loss L_{DF} is defined as:

$$L_{DF} = -\Big(\gamma_1 \cdot \text{sim}(\hat{I}_{\mathcal{F}^{-1}}, I_{gt}) + \gamma_2 \cdot \text{sim}(\hat{I}_{r^{-1}}, I_{gt}) + \gamma_3 \cdot \text{sim}(\hat{I}_m, I_{gt}) + \gamma_4 \cdot \text{sim}(\hat{I}, I_{gt})\Big) \tag{5.11}$$

where $\hat{I}$ denotes the denoised output corresponding to the original input, while $\hat{I}_{\mathcal{F}^{-1}}$, $\hat{I}_{r^{-1}}$, and $\hat{I}_m$ represent the outputs corresponding to flipped, rotated, and masked inputs, respectively. The function $\text{sim}(\cdot, \cdot)$ measures similarity between two HSIs, capturing both spatial and spectral agreement. The coefficients $\gamma_1, \gamma_2, \gamma_3$, and γ_4 control the relative contribution of

each term. In this work, all coefficients are set equally as:

$$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \frac{1}{4}$$

as determined empirically (see Sec. 5.4). This formulation ensures that all augmented views contribute to learning a representation that remains consistent with the clean signal, thereby supporting both denoising quality and reliable plume delineation.

5.2.3 Total Loss

The overall training objective combines the contrastive consistency constraint and the data fidelity term. The total loss is defined as:

$$L_T = \alpha L_{CL} + L_{DF} \tag{5.12}$$

where α balances the contribution of the contrastive loss. In this work, $\alpha = 1$, based on empirical evaluation. The model parameters θ of the network $D(\cdot)$ are optimized by minimizing the total loss:

$$\hat{\theta} = \arg\min_{\theta} \; L_T = \arg\min_{\theta} \left(L_{CL} + L_{DF}\right) \tag{5.13}$$

By jointly optimizing these objectives, the proposed framework encourages consistent representations across transformations while maintaining fidelity to the clean hyperspectral signal, which is essential for subsequent CH_4 plume segmentation.

5.3 Experimental Results

This section presents the experimental evaluation of the proposed `AUGMENT` framework from two complementary perspectives. First, we investigate its effectiveness for HSID under both synthetic and real-world noise conditions. Subsequently, we examine how the denoised hyperspectral representations influence downstream CH_4 plume segmentation performance across multiple benchmark datasets.

5.3.1 Hyperspectral Image Denoising

5.3.1.1 Dataset

To evaluate the denoising capability of the proposed AUGMENT framework, we utilize the ICVL HSI dataset (102), which contains 201 images with 31 spectral bands and a spatial resolution of 1392×1300. The dataset is divided into 100 images for training, 20 for validation, and the remaining for testing. All images are normalized to the range $[0, 1]$ prior to training. Following standard practices in HSID (32; 79), the images are partitioned into spatial-spectral cubes. Specifically, patches of size $64 \times 64 \times 31$ are used during training, while larger patches of size $512 \times 512 \times 31$ are used for testing. To further examine generalization, we conduct experiments on real-world HSI datasets, including IP (105), and PU (137). These datasets provide diverse spectral characteristics and acquisition conditions, enabling evaluation under different real-world scenarios.

5.3.1.2 Noise settings

Real-world HSIs are commonly affected by multiple forms of degradation, including dead pixels, Gaussian noise, impulse noise, and stripe noise (82; 107). To comprehensively evaluate the denoising capability of the proposed AUGMENT framework, we generate noisy hyperspectral observations using widely adopted synthetic noise models (60).

Specifically, two categories of noise settings are considered. The first category consists of additive i.i.d. Gaussian noise with noise levels ranging from 10 to 70 in increments of 20. The second category includes more challenging mixed-noise scenarios designed to better approximate real-world hyperspectral degradations. These settings include: (1) non-i.i.d. Gaussian noise, (2) mixed noise corruption, (3) Gaussian noise combined with deadline noise, (4) Gaussian noise combined with impulse noise, and (5) Gaussian noise combined with stripe noise. These diverse degradation settings enable detailed analysis of the proposed framework under both controlled and complex noise conditions that are frequently encountered in practical hyperspectral imaging systems.

Table 5.1: Quantitative evaluation of denoising methods on the ICVL dataset in the presence of Gaussian noise.

		σ = 10			σ = 30			σ = 50			σ = 70		
Methods		**PSNR↑**	**SSIM↑**	**SAM↓**	**PSNR↑**	**SSIM↑**	**SAM↓**	**PSNR↑**	**SSIM↑**	**SAM↓**	**PSNR↑**	**SSIM↑**	**SAM↓**
Noisy		28.13	0.8792	18.72	18.59	0.5523	37.90	14.15	0.3476	49.01	11.23	0.2301	56.45
HSID-CNN (11)	w/o	43.14	0.9918	2.12	40.30	0.9854	3.14	37.72	0.9746	4.27	34.95	0.9521	5.84
	AUGMENT	**43.44**	**0.9921**	**2.08**	**40.78**	**0.9860**	**3.10**	**37.98**	**0.9753**	**4.20**	**35.23**	**0.9529**	**5.77**
QRNN3D (60)	w/o	45.61	0.9977	1.80	42.18	0.9955	2.21	40.05	0.9929	2.63	38.09	0.9883	3.42
	AUGMENT	**45.98**	**0.9978**	**1.76**	**42.71**	**0.9956**	**2.18**	**40.33**	**0.9930**	**2.58**	**38.36**	**0.9885**	**3.37**
T3SC (59)	w/o	45.81	0.9979	2.02	42.44	0.9957	2.44	40.39	0.9933	2.85	38.80	0.9904	3.26
	AUGMENT	**46.16**	**0.9980**	**1.99**	**42.94**	**0.9959**	**2.42**	**40.71**	**0.9935**	**2.81**	**39.10**	**0.9908**	**3.22**
GRNet (58)	w/o	45.25	0.9976	1.83	42.09	0.9957	2.18	40.25	0.9936	2.42	38.95	0.9914	2.63
	AUGMENT	**45.57**	**0.9978**	**1.76**	**42.58**	**0.9958**	**2.14**	**40.49**	**0.9937**	**2.40**	**39.17**	**0.9918**	**2.60**
SST (65)	w/o	48.28	0.9990	1.30	43.32	0.9971	1.87	41.09	0.9950	2.19	39.55	0.9921	2.46
	AUGMENT	**48.66**	**0.9991**	**1.28**	**43.88**	**0.9972**	**1.84**	**41.44**	**0.9955**	**2.13**	**39.83**	**0.9927**	**2.41**
SERT (32)	w/o	47.72	0.9988	1.36	43.56	0.9969	1.77	41.33	0.9949	2.05	39.82	0.9929	2.30
	AUGMENT	**48.07**	**0.9989**	**1.31**	**44.13**	**0.9971**	**1.75**	**41.72**	**0.9952**	**1.98**	**40.19**	**0.9933**	**2.25**
IDEATION	w/o	49.08	0.9992	1.22	45.63	0.9977	1.63	44.94	0.9972	1.85	43.96	0.9958	2.02
	AUGMENT	**49.33**	**0.9993**	**1.19**	**45.81**	**0.9979**	**1.62**	**45.21**	**0.9973**	**1.81**	**44.17**	**0.9960**	**1.97**

Note: "w/o" denotes the baseline denoising model trained without the proposed AUGMENT framework and using only conventional augmentation strategies.

5.3.1.3 Implementation Details

All experiments are implemented using PyTorch and conducted on a system equipped with an AMD Ryzen 5 5600X six-core CPU, 16 GB RAM, and an NVIDIA GeForce RTX 3050 GPU. The proposed AUGMENT framework is trained using the AdamW optimizer (108) with an initial learning rate of 1×10^{-4}. During training, a batch size of 4 is employed, and the network is optimized for 500 epochs. Unless otherwise stated, the same training configuration is used across all experiments to maintain consistency during performance evaluation.

5.3.1.4 Quantitative Metrics

Following established HSID protocols (79), we evaluate the performance of the proposed AUGMENT framework using three widely adopted quantitative metrics, namely Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) (138), and Spectral Angle Mapper (SAM) (139). Among these, PSNR and SSIM are utilized to measure spatial reconstruction quality, while SAM evaluates spectral consistency between the restored and reference HSIs. Higher PSNR and SSIM values indicate improved reconstruction quality, whereas lower SAM values correspond to better preservation of spectral information.

1. Peak Signal-to-Noise Ratio (PSNR): PSNR measures the reconstruction fidelity between the denoised image $\hat{I}$ and the ground-truth image I_{gt}. It is computed as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{MAX^2}{\text{MSE}} \right) \tag{5.14}$$

where MAX denotes the maximum possible pixel intensity value and MSE represents the mean squared error between $\hat{I}$ and I_{gt}, defined as:

$$\text{MSE} = \frac{1}{HWB} \sum_{h=1}^{H} \sum_{w=1}^{W} \sum_{b=1}^{B} \left(\hat{I}(h, w, b) - I_{gt}(h, w, b) \right)^2 \tag{5.15}$$

2. Structural Similarity Index Measure (SSIM): SSIM evaluates structural similarity between the denoised and reference images by jointly considering luminance, contrast, and

structural information. It is defined as:

$$\text{SSIM}(\hat{I}, I_{gt}) = \frac{(2\mu_{\hat{I}}\mu_{I_{gt}} + c_1)(2\sigma_{\hat{I}I_{gt}} + c_2)}{(\mu_{\hat{I}}^2 + \mu_{I_{gt}}^2 + c_1)(\sigma_{\hat{I}}^2 + \sigma_{I_{gt}}^2 + c_2)} \tag{5.16}$$

where μ, σ^2, and $\sigma_{\hat{I}I_{gt}}$ denote the mean, variance, and covariance, respectively, while c_1 and c_2 are small constants used for numerical stability.

3. Spectral Angle Mapper (SAM): SAM measures spectral similarity by computing the angular difference between the spectral signatures of the denoised and reference hyperspectral pixels. It is defined as:

$$\text{SAM} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\cos^{-1}\left(\frac{\hat{I}_{h,w}^T I_{gt,h,w}}{\|\hat{I}_{h,w}\|_2 \, \|I_{gt,h,w}\|_2}\right) \tag{5.17}$$

where $\hat{I}_{h,w}$ and $I_{gt,h,w}$ represent the spectral vectors corresponding to pixel location (h, w) in the denoised and reference HSIs, respectively.

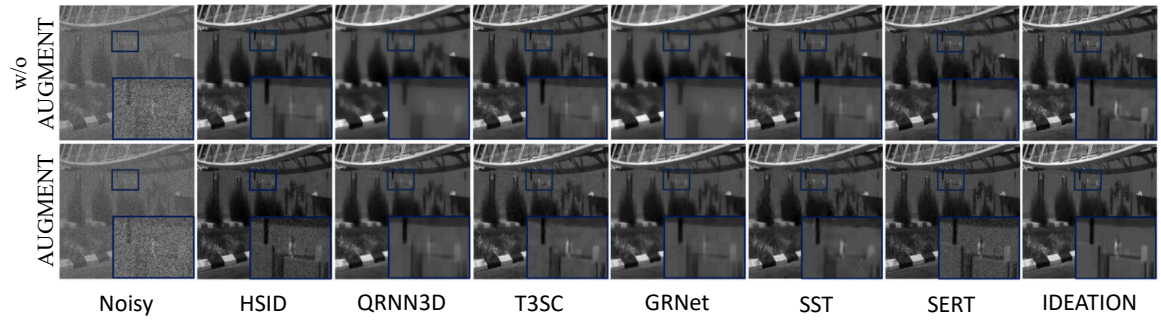


Figure 5.2: Visual comparison of denoising results on the ICVL dataset under Gaussian noise with $\sigma = 50$.

5.3.1.5 Experimental results of simulated data

Gaussian Noise: To evaluate the denoising capability of the proposed `AUGMENT` framework, zero-mean additive white Gaussian noise with different variance levels (σ) was introduced into the HSIs. The quantitative comparisons are summarized in Table 5.1. In the table, the term "w/o" denotes the original backbone networks trained using only standard augmentation strategies, without incorporating the proposed `AUGMENT` framework. The obtained results

Table 5.2: Quantitative evaluation of denoising methods on the ICVL dataset under complex noise.

Noise Type →		Non-i.i.d Gaussian			Gaussian+Stripe			Gaussian+Deadline			Gaussian+Impulse			Gaussian+Mixture		
Methods ↓		PSNR ↑	SSIM ↑	SAM ↓	PSNR ↑	SSIM ↑	SAM ↓	PSNR ↑	SSIM ↑	SAM ↓	PSNR ↑	SSIM ↑	SAM ↓	PSNR ↑	SSIM ↑	SAM ↓
Noisy		18.29	0.5116	46.20	17.51	0.4867	46.98	17.50	0.4770	47.55	14.93	0.3758	46.98	13.91	0.3396	51.53
HSID-CNN (11)	w/o	39.28	0.9819	3.80	38.09	0.9765	4.59	38.33	0.9783	3.99	36.21	0.9663	5.48	35.30	0.9588	6.29
	AUGMENT	**39.90**	**0.9821**	**3.76**	**38.33**	**0.9768**	**4.54**	**38.54**	**0.9785**	**3.94**	**36.55**	**0.9667**	**5.44**	**35.79**	**0.9592**	**6.24**
QRNN3D (60)	w/o	42.18	0.9950	2.84	41.68	0.9943	2.97	41.69	0.9942	2.61	40.32	0.9914	4.31	39.08	0.9892	4.80
	AUGMENT	**42.83**	**0.9953**	**2.81**	**41.95**	**0.9945**	**2.94**	**41.93**	**0.9944**	**2.58**	**40.68**	**0.9916**	**4.26**	**39.60**	**0.9895**	**4.76**
T3SC (59)	w/o	41.95	0.9922	4.18	41.32	0.9937	3.27	39.59	0.9924	4.86	37.85	0.9843	6.53	35.53	0.9767	8.12
	AUGMENT	**42.33**	**0.9927**	**4.14**	**41.65**	**0.9940**	**3.21**	**39.81**	**0.9927**	**4.81**	**38.17**	**0.9847**	**6.49**	**36.01**	**0.9771**	**8.09**
GRNet (58)	w/o	35.19	0.9780	5.19	34.85	0.9772	5.41	33.78	0.9744	5.42	32.78	0.9606	8.26	30.91	0.9617	8.26
	AUGMENT	**35.53**	**0.9783**	**5.16**	**35.15**	**0.9776**	**5.38**	**34.03**	**0.9748**	**5.39**	**33.13**	**0.9610**	**8.21**	**31.40**	**0.9621**	**8.21**
SST (65)	w/o	43.57	0.9970	2.05	42.97	0.9960	2.15	42.74	0.9961	2.21	41.66	0.9950	2.60	38.78	0.9911	2.99
	AUGMENT	**44.25**	**0.9972**	**2.01**	**43.26**	**0.9961**	**2.11**	**43.01**	**0.9963**	**2.18**	**42.06**	**0.9951**	**2.57**	**39.38**	**0.9914**	**2.97**
SERT (32)	w/o	44.20	0.9971	1.69	43.68	0.9969	1.97	43.66	0.9969	1.99	42.67	0.9959	2.30	40.00	0.9937	2.84
	AUGMENT	**44.91**	**0.9972**	**1.66**	**44.03**	**0.9972**	**1.95**	**43.96**	**0.9971**	**1.97**	**43.13**	**0.9961**	**2.27**	**40.72**	**0.9940**	**2.80**
IDEATION	w/o	47.08	0.9974	1.55	45.22	0.9975	1.85	46.23	0.9978	1.91	44.90	0.9969	2.17	45.50	0.9949	2.69
	AUGMENT	**47.53**	**0.9976**	**1.52**	**45.48**	**0.9976**	**1.83**	**46.54**	**0.9979**	**1.82**	**45.21**	**0.9970**	**2.11**	**45.96**	**0.9951**	**2.61**

Note: The term "w/o" denotes the baseline denoising models trained without the proposed AUGMENT framework, while "AUGMENT" indicates the corresponding models integrated with the proposed supervised contrastive learning and inpainting strategy.

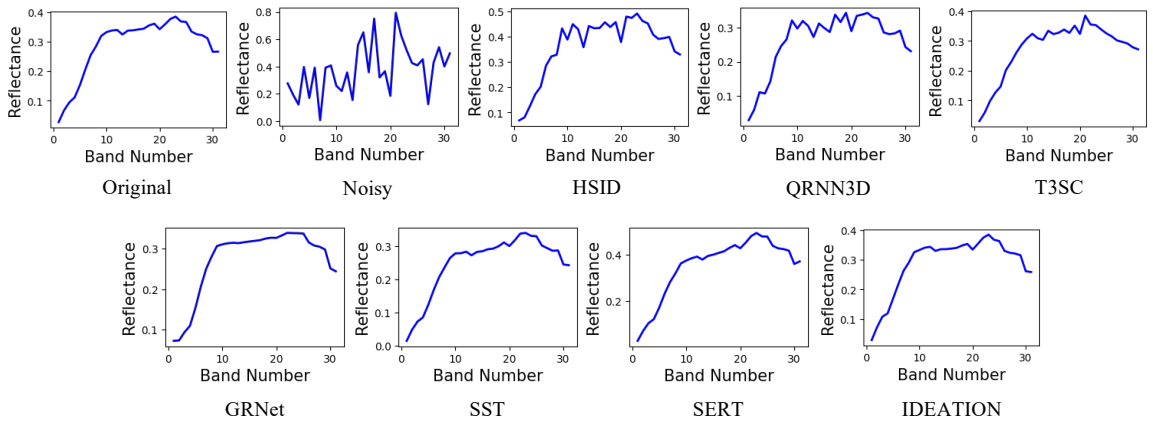


Figure 5.3: Comparison of reconstructed spectral signatures obtained by AUGMENT-integrated baseline methods on the ICVL dataset under Gaussian noise ($\sigma = 50$) at pixel location $(120, 120)$.

Table 5.3: SVM classification accuracy before and after HSI denoising.

Dataset	**Metric**	Noisy	HSID-CNN	QRNN3D	T3SC	GRNet	SST	SERT	IDEATION
IP	OA (w/o)	71.25	83.31	84.38	85.62	78.66	87.01	88.17	90.89
	OA (**AUGMENT**)	**71.25**	**84.03**	**85.12**	**86.34**	**79.46**	**87.95**	**89.04**	**91.56**
	Kappa (w/o)	0.6696	0.8101	0.8228	0.8366	0.7560	0.8422	0.8628	0.8960
	Kappa (**AUGMENT**)	**0.6696**	**0.8209**	**0.8337**	**0.8458**	**0.7684**	**0.8531**	**0.8734**	**0.9028**
PU	OA (w/o)	86.54	92.81	93.10	94.84	90.45	95.67	96.52	98.70
	OA (**AUGMENT**)	**86.54**	**93.47**	**93.81**	**95.50**	**91.02**	**96.28**	**97.29**	**99.24**
	Kappa (w/o)	0.8068	0.9167	0.9218	0.9330	0.8891	0.9376	0.9588	0.9770
	Kappa (**AUGMENT**)	**0.8068**	**0.9222**	**0.9273**	**0.9388**	**0.8954**	**0.9439**	**0.9652**	**0.9833**

Note: "w/o" denotes the baseline denoising models without the proposed AUGMENT framework, while "AUGMENT" denotes the corresponding models integrated with the proposed supervised contrastive learning and inpainting strategy.

indicate that integrating AUGMENT consistently improves the denoising performance of all baseline methods across different noise levels. The improvements are evident in higher PSNR and SSIM values and lower SAM values, indicating enhanced reconstruction quality in both the spatial and spectral domains. Furthermore, the proposed framework is compatible with various CNN- and Transformer-based denoising architectures, underscoring its applicability as a generic auxiliary learning strategy for HSID.

The qualitative comparisons presented in Figs. 5.2 and 5.3 further support the quantitative observations. Specifically, Fig. 5.2 illustrates the reconstructed visual outputs under Gaussian noise, where the denoised images generated using AUGMENT preserve finer spatial structures and reduce distortion artifacts more effectively than the baseline approaches. In addition, Fig. 5.3 presents the spectral profile comparisons, which are important for evaluat-

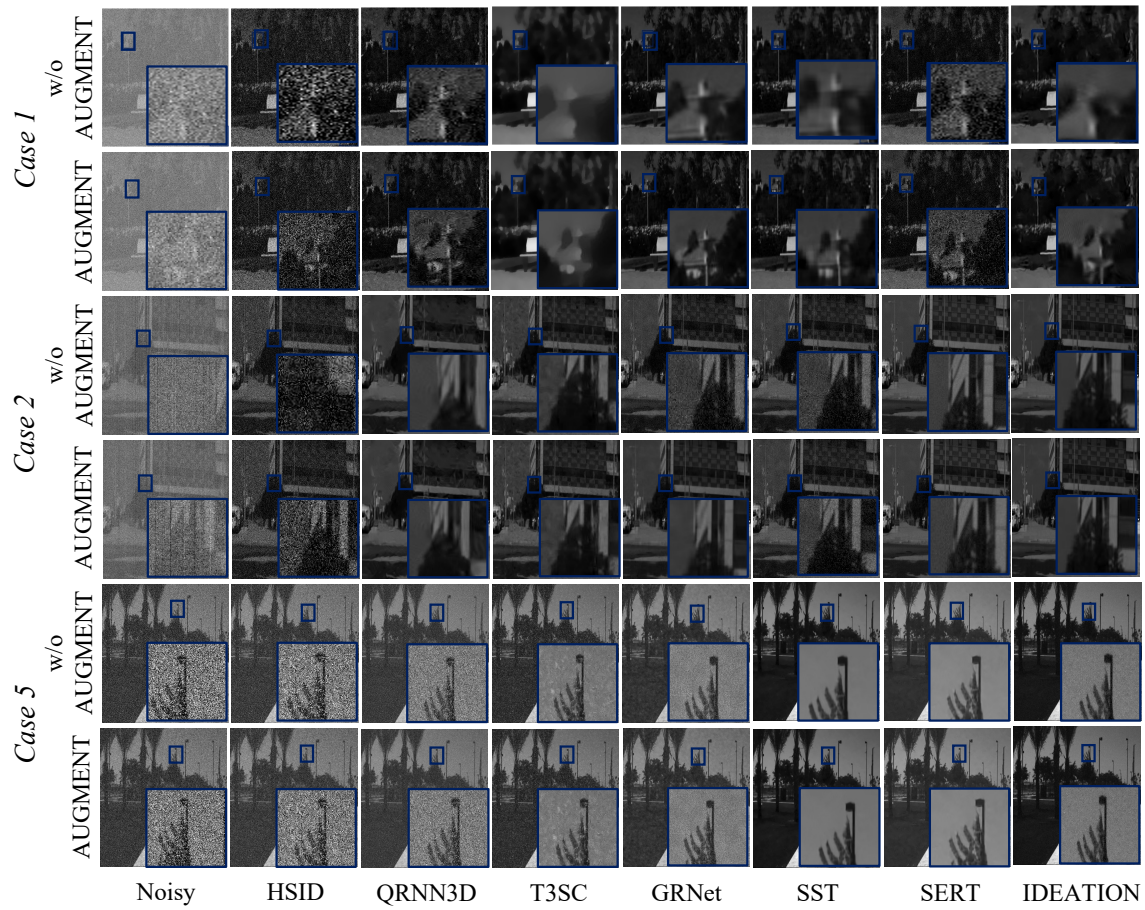


Figure 5.4: Visual comparison of denoising results on the ICVL dataset under different complex noise.

ing spectral consistency after denoising. The spectral curves produced by AUGMENT remain closer to the ground-truth signatures, demonstrating improved preservation of spectral characteristics that are essential for subsequent hyperspectral analysis tasks.

The residual error maps shown in Fig. 5.5 provide additional insight into reconstruction quality by highlighting the remaining errors after denoising. Compared with the baseline methods, the residual maps corresponding to AUGMENT contain fewer structural distortions and lower residual noise patterns. Furthermore, Fig. 5.6 presents false-color visualization results for bands 4, 11, and 19 of the ICVL dataset under Gaussian noise ($\sigma = 50$). The visual comparisons demonstrate that the proposed framework more effectively preserves spatial structures and spectral consistency while suppressing noise degradation across spectral bands.

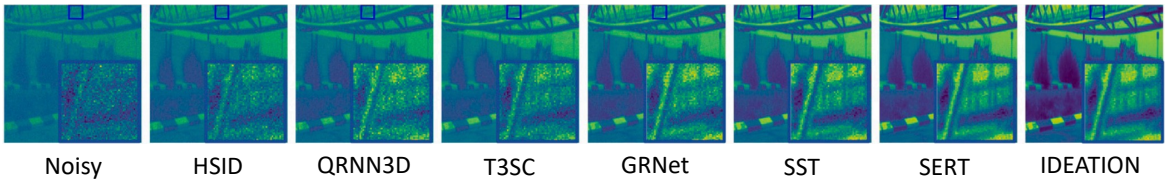


Figure 5.5: Residual error maps of baseline and AUGMENT-enhanced denoising methods on the ICVL dataset under Gaussian noise with $\sigma = 50$.

Table 5.4: Sensitivity analysis of AUGMENT integrated with IDEATION on the ICVL dataset under Gaussian noise ($\sigma = 50$). Bold values indicate the parameter configuration used in the main experiments.

Parameters	**Metrics**	0.1	0.2	0.25	0.3	**0.33 / 0.25**	0.35	0.4	0.5
λ ($\gamma = \frac{1}{4}$)	MPSNR (dB)	44.50	44.65	44.72	44.76	**44.83**	44.81	44.75	44.60
	MSSIM	0.9965	0.9967	0.9968	0.9969	**0.9970**	0.9969	0.9968	0.9966
γ ($\lambda = \frac{1}{3}$)	MPSNR (dB)	44.55	44.66	textbf44.83	44.74	44.70	44.77	44.68	44.59
	MSSIM	0.9966	0.9967	**0.9970**	0.9969	0.9968	0.9969	0.9967	0.9965

Note: The parameters λ and γ were varied independently while fixing the other parameter.
The highlighted values correspond to the final parameter setting adopted in AUGMENT.

Complex Noises: The denoising performance of the proposed AUGMENT framework under complex noise conditions is reported in Table 5.2. As in the previous experiment, the notation "w/o" denotes the baseline backbone networks trained without the proposed framework, using only conventional augmentation strategies. Visual comparisons under different complex noise conditions are presented in Fig. 5.4. The results demonstrate that AUGMENT improves the reconstruction performance of all baseline methods under various complex degradation scenarios, including non-i.i.d. Gaussian noise, Gaussian noise combined with stripe noise, deadline noise, impulse noise, and mixed noise conditions. The improvements are reflected through better spatial reconstruction and improved preservation of spectral information across different corruption patterns.

In particular, traditional handcrafted methods often struggle in the presence of impulse and mixed noise because such degradations severely disturb the spatial and spectral correlations of hyperspectral data. Since these methods rely heavily on predefined assumptions and manually designed priors, their performance degrades when the noise distribution becomes highly irregular or spatially variant. In contrast, DL-based methods learn feature representations directly from data and are therefore better suited for modeling complex degra-

dation patterns. By incorporating SCL and inpainting-based augmentation, the proposed AUGMENT framework further enhances feature discrimination and reconstruction consistency. The masking-based inpainting strategy encourages the network to recover missing spectral-spatial structures, while SCL promotes similarity among correlated representations. Consequently, the framework produces denoised outputs with improved spatial clarity and spectral consistency under challenging noise conditions.



Figure 5.6: False-color visualization of denoising results produced by baseline and AUGMENT-enhanced methods on the ICVL dataset under Gaussian noise ($\sigma = 50$), using bands 4, 11, and 19 for RGB composition.

5.3.1.6 Experimental results of real data

To further examine the applicability of the proposed AUGMENT framework under practical scenarios, experiments were conducted on real-world HSI datasets, namely IP and PU. The PU dataset contains HSIs of spatial size 610×340 with 103 spectral bands, whereas the IP dataset consists of 145×145 pixels with 220 spectral channels. Unlike synthetic experiments, where individual noise types are artificially introduced in a controlled manner, real-world HSIs generally contain multiple degradation sources simultaneously, including Gaussian noise, impulse noise, stripe artifacts, and sensor-related distortions (140; 141). Consequently, these datasets provide a more practical benchmark for evaluating the denoising capability of HSI restoration methods under mixed-noise environments.

The qualitative denoising results are illustrated in Fig. 5.7, where the reconstructed outputs for the noisiest spectral bands are presented. Specifically, bands 2, 101, and 205 are considered for the IP dataset, while bands 3, 11, and 98 are selected for the PU dataset. The visual comparisons demonstrate that the proposed AUGMENT framework suppresses noise artifacts while preserving important spectral-spatial structures within the HSI data.

To further analyze the influence of denoising on downstream hyperspectral applications, classification experiments were conducted before and after denoising. A SVM classifier

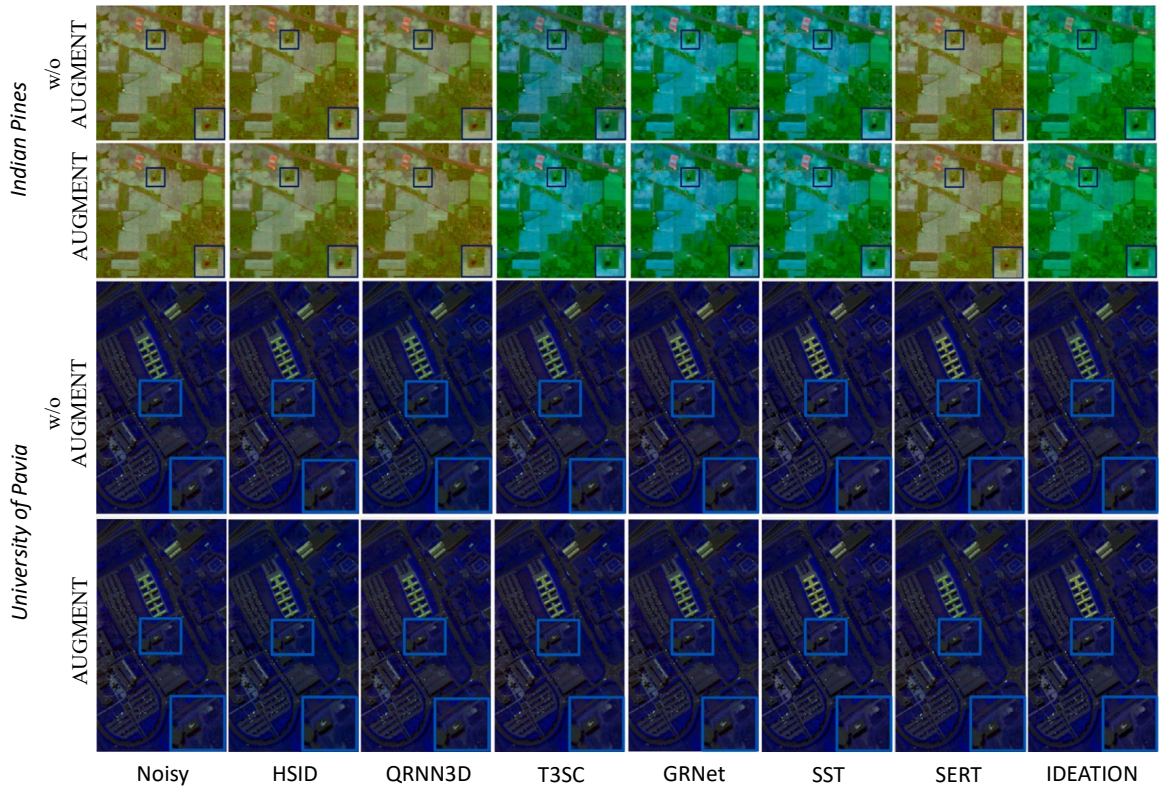


Figure 5.7: Visual comparison of denoising results on the real-world IP and PU datasets. The displayed bands correspond to (2, 101, 205) for IP and (3, 11, 98) for PU.

(142) was employed to maintain consistency across all experiments. The classification performance was evaluated using OA and the Kappa coefficient. The classification maps for the IP and PU datasets are shown in Fig. 5.8, while the quantitative comparisons are summarized in Table 5.3. The obtained results indicate that integrating AUGMENT with existing denoising frameworks improves classification performance across both datasets. The improvements in OA and Kappa values suggest that the proposed denoising strategy preserves meaningful spectral-spatial information that is beneficial for subsequent hyperspectral interpretation tasks.

5.3.1.7 Computational Complexity

The computational characteristics of the proposed AUGMENT framework were also analyzed by comparing the baseline backbone with and without the proposed training strategy. The incorporation of AUGMENT increased the total training duration from approximately

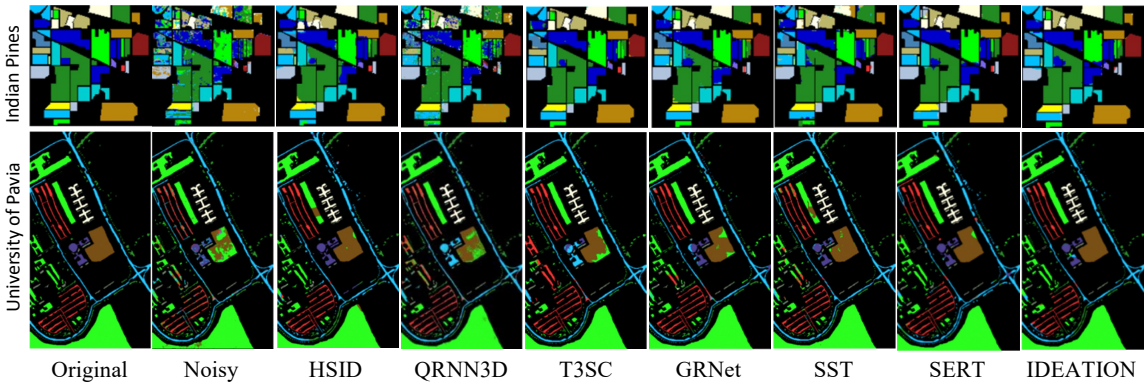


Figure 5.8: Comparison of SVM-based classification outcomes on the PU and IP datasets, highlighting the impact of AUGMENT on downstream HSI analysis.

Table 5.5: Ablation analysis of supervised contrastive learning within the proposed AUGMENT framework under Gaussian noise on the ICVL dataset.

Settings	$\sigma = 30$		$\sigma = 50$		$\sigma = 70$	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baseline	44.84	0.9973	44.39	0.9968	43.46	0.9956
Baseline + Augmentation	45.12	0.9973	44.59	0.9969	43.59	0.9956
Baseline + AUGMENT	**45.46**	**0.9976**	**44.83**	**0.9970**	**43.86**	**0.9958**

Note: The table highlights the contribution of SCL and augmentation strategies within the proposed AUGMENT framework. The best-performing configuration is highlighted in bold.

15 hours to 17 hours due to the additional SCL and inpainting operations performed during optimization. However, the inference stage remained unaffected since the proposed framework modifies only the training procedure and does not introduce additional operations during testing. The average inference time remained nearly identical at approximately 5.8 seconds for both configurations. Similarly, the computational complexity, measured in GFLOPs, increased marginally from 141.5 to 143.3. These observations indicate that the proposed framework introduces only limited additional computational overhead while improving denoising performance.

5.3.1.8 Sensitivity Analysis

A sensitivity analysis was conducted to study the influence of the weighting parameters λ and γ used in the proposed objective function. The experimental results are summarized

in Table 5.4. During the analysis, λ and γ were varied independently while fixing the other parameter. For the parameter λ, the values were varied from 0.1 to 0.5 while maintaining $\gamma = 1/4$. Similarly, for γ, the same range of values was explored while fixing $\lambda = 1/3$. The quantitative results indicate that performance variations remain relatively small across different parameter settings for both MPSNR and MSSIM. The highest reconstruction performance was obtained when $\lambda = 1/3$ and $\gamma = 1/4$, and these values were therefore selected for all experiments presented in this chapter. The observations from the sensitivity analysis suggest that the proposed AUGMENT framework maintains consistent behavior across a reasonably broad parameter range, indicating stable optimization characteristics during training.

5.3.2 Hyperspectral Methane Plume Segmentation

5.3.2.1 Datasets

To evaluate the effectiveness of the proposed AUGMENT framework for hyperspectral methane plume segmentation, experiments are conducted on three widely adopted remote sensing datasets, namely AVIRIS-NG, WV-3, and EMIT. These datasets offer substantial diversity in spectral, spatial, and sensor characteristics, enabling a comprehensive evaluation of segmentation performance and cross-sensor generalization. Following the plume categorization strategy adopted in recent methane plume detection studies (31; 77), methane plumes with estimated emission rates greater than 1000 kg/h are categorized as *strong plumes*, whereas plumes with lower emission rates are categorized as *weak plumes*. The AVIRIS-NG dataset (120) contains HSI acquired over the spectral range of 380-2510 nm with a spectral resolution of approximately 5 nm. Each scene has a spatial dimension of 512×512 pixels. Following the methodology described in (31), additional training annotations are generated using methane plume masks obtained from the mag1c product (121), which are subsequently refined through manual correction to improve annotation quality. The resulting dataset contains 167,825 training patches of size 128×128 extracted with an overlap of 64 pixels, together with 342 testing scenes of size 512×512.

For the WV-3 dataset, multispectral observations are synthetically generated by con-

volving AVIRIS-NG hyperspectral measurements with the spectral response function of the WorldView-3 sensor (122). The same preprocessing pipeline, data partitioning strategy, and patch extraction procedure used for AVIRIS-NG are retained to ensure consistent evaluation. In contrast to AVIRIS-NG, plume annotations are generated using SBMP ratio products (123), following the protocol adopted in (31; 77). The EMIT dataset consists of hyperspectral scenes collected over methane-emitting and non-emitting regions using the EMIT imaging spectrometer. The dataset spans the spectral range of 400-2500 nm with a spectral sampling interval of approximately 10 nm. Each scene has a spatial dimension of 512×512 pixels and includes observations from 11 methane-emitting sites and 9 non-emitting sites. Consistent with previous methane plume segmentation studies (31; 77), the commonly adopted Level-1 reflectance representation is utilized for evaluation. The EMIT dataset is used exclusively to assess zero-shot cross-sensor generalization without additional fine-tuning.

5.3.2.2 Implementation Details

All experiments are implemented using PyTorch 2.6.0 with CUDA 12.4 on a workstation equipped with an AMD Ryzen 5 5600X processor, 16 GB RAM, and an NVIDIA RTX 3050 GPU. The methane plume segmentation framework is optimized with the AdamW optimizer and a batch size of 8 over 15 training epochs. The initial learning rate is set to 1×10^{-4} and gradually decayed to 1×10^{-6} using a cosine annealing scheduling strategy. To supervise the segmentation process, Binary Cross-Entropy (BCE) loss (124) is employed to quantify the discrepancy between the predicted probability map and the corresponding binary ground-truth mask.

To accommodate spectral variations among different sensors, we follow the preprocessing strategies commonly adopted in methane plume segmentation literature (31; 77). Specifically, mag1c-based band representations are utilized for AVIRIS-NG, whereas SBMP-based band representations are employed for WV-3. The preprocessing implementations for both mag1c and SBMP representations follow publicly available standard protocols.[1]

In this chapter, the proposed `AUGMENT` framework is integrated into the methane plume

[1] `https://github.com/spaceml-org/STARCOP`

Table 5.6: Quantitative comparison of methane plume segmentation performance on the AVIRIS-NG dataset for strong and weak plumes.

Methods		**Strong Plumes**					**Weak Plumes**				
		MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
MethaNet (70)	w/o	70.04	69.79	62.13	46.89	0.5454	51.15	69.48	60.70	41.49	0.3180
	AUGMENT	**71.26**	**71.08**	**63.84**	**48.27**	**0.5211**	**53.02**	**71.15**	**62.48**	**43.06**	**0.2945**
U-Plume (72)	w/o	72.32	74.28	74.84	62.58	0.8976	54.39	69.11	64.12	47.36	0.2665
	AUGMENT	**73.84**	**75.92**	**76.13**	**64.01**	**0.8512**	**56.18**	**71.44**	**65.95**	**49.03**	**0.2417**
CH4Net (71)	w/o	74.67	75.67	70.83	58.10	0.4507	62.05	57.98	65.02	45.67	0.2017
	AUGMENT	**76.01**	**77.20**	**72.34**	**59.88**	**0.4175**	**63.71**	**60.13**	**66.48**	**47.12**	**0.1863**
HyperSTARCOP (31)	w/o	82.19	65.86	68.40	56.14	0.2194	73.07	43.22	59.35	37.01	0.1055
	AUGMENT	**83.45**	**67.58**	**70.06**	**57.82**	**0.1942**	**74.88**	**45.61**	**61.28**	**39.05**	**0.0931**
Gasformer (35)	w/o	79.17	79.89	77.45	65.31	0.5964	64.51	80.92	72.46	60.48	0.2233
	AUGMENT	**81.02**	**81.63**	**79.18**	**67.05**	**0.5480**	**66.29**	**82.74**	**74.11**	**62.14**	**0.2019**
MethaneMapper (73)	w/o	81.03	76.29	78.14	65.21	0.1461	73.02	82.35	76.81	64.47	0.1840
	AUGMENT	**82.57**	**78.04**	**79.86**	**67.18**	**0.1296**	**74.90**	**84.12**	**78.54**	**66.28**	**0.1615**
FSDINet (77)	w/o	87.75	80.13	77.05	65.56	0.1036	78.94	78.25	77.80	65.20	0.0771
	AUGMENT	**89.18**	**82.64**	**80.42**	**69.01**	**0.0618**	**80.76**	**80.41**	**80.15**	**68.04**	**0.0426**
ALKANE	w/o	90.04	86.79	85.13	76.99	0.0013	80.13	83.02	83.41	71.18	0.0005
	AUGMENT	**91.37**	**88.45**	**87.06**	**79.12**	**0.0008**	**82.04**	**84.96**	**85.72**	**73.84**	**0.0003**

Note: "w/o" indicates the original baseline models trained without the proposed `AUGMENT` framework. "`AUGMENT`" denotes the corresponding models trained using the proposed augmentation-driven learning strategy.

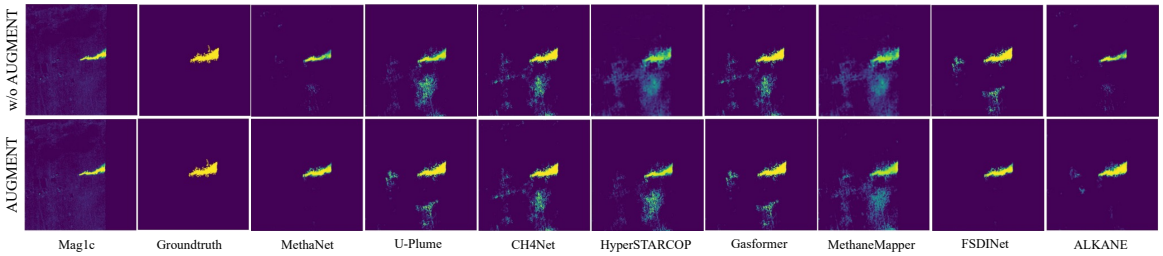


Figure 5.9: Visual comparison of methane plume segmentation results on strong plumes from the AVIRIS-NG dataset.

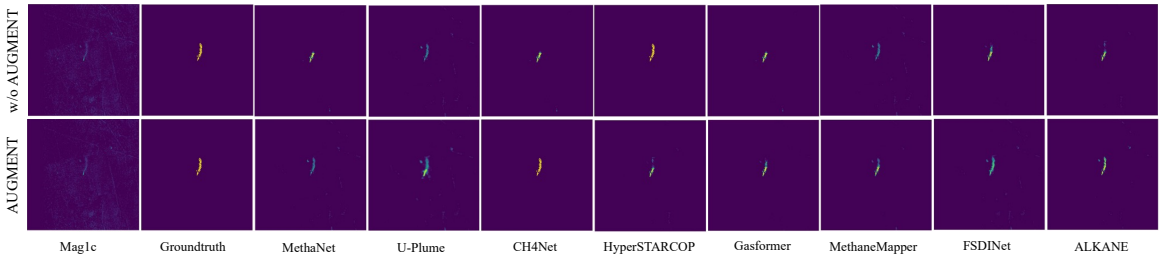


Figure 5.10: Visual comparison of methane plume segmentation results on weak plumes from the AVIRIS-NG dataset.

segmentation pipeline to investigate whether the augmentation-driven representation learning strategy can improve plume localization and segmentation robustness under varying spectral and spatial conditions. Similar to the HSI denoising experiments, `AUGMENT` is incorporated as a plug-and-play training strategy without modifying the inference-stage architecture.

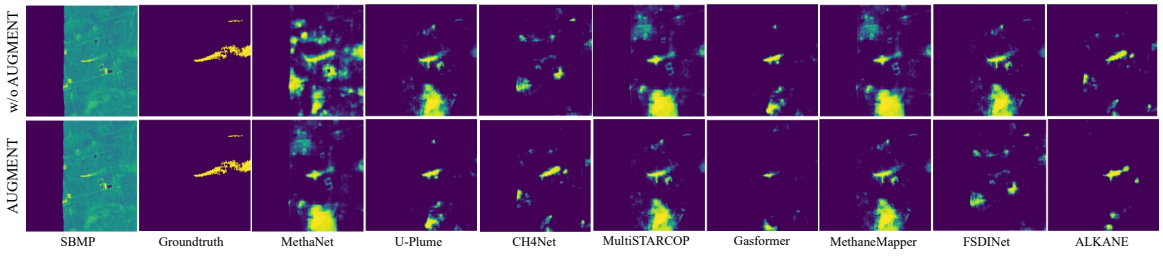


Figure 5.11: Visual comparison of methane plume segmentation results on strong plumes from the WV-3 dataset.

5.3.2.3 Evaluation Metrics

To quantitatively evaluate methane plume segmentation performance, five widely adopted evaluation metrics are considered, namely Mean Precision (MP), Mean Recall (MR), Mean F1-score (MF1), Mean Intersection over Union (MIoU), and Mean False Positive Rate (MFPR) (117). These metrics collectively evaluate detection accuracy, segmentation overlap quality, and false alarm characteristics. The mathematical definitions of these metrics are expressed as follows:

$$\text{MP} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FP_i} \tag{5.18}$$

$$\text{MR} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FN_i} \tag{5.19}$$

$$\text{MF1} = \frac{1}{C}\sum_{i=1}^{C}\frac{2TP_i}{2TP_i + FP_i + FN_i} \tag{5.20}$$

$$\text{MIoU} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{5.21}$$

$$\text{MFPR} = \frac{1}{C}\sum_{i=1}^{C}\frac{FP_i}{FP_i + TN_i} \tag{5.22}$$

where TP_i, FP_i, FN_i, and TN_i denote the true positives, false positives, false negatives, and true negatives associated with the i^{th} class, respectively, and C represents the total number of classes.

5.3.3 Comparative Analysis

To assess the effectiveness of the proposed AUGMENT framework, experiments are conducted using several existing methane plume segmentation approaches, including CNN-based methods such as MethaNet (70), U-Plume (72), CH4Net (71), HyperSTARCOP, MultiSTARCOP (31), and FSDINet (77), as well as transformer-based methods including Gasformer (35) and MethaneMapper (73).

For each baseline segmentation framework, experiments are performed both with and without the proposed AUGMENT strategy to evaluate its contribution toward improving

Table 5.7: Quantitative comparison of methane plume segmentation performance on the WV-3 dataset for strong and weak plumes.

Methods		**Strong Plumes**					**Weak Plumes**				
		MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
MethaNet (70)	w/o	9.75	75.20	16.43	9.39	14.7511	0.99	55.36	2.35	0.97	10.1400
	AUGMENT	**11.08**	**76.94**	**18.25**	**10.62**	**13.2045**	**1.42**	**57.18**	**3.14**	**1.28**	**9.2214**
U-Plume (72)	w/o	13.83	65.65	20.55	11.99	6.7680	1.53	64.69	3.46	1.51	6.2530
	AUGMENT	**15.26**	**67.48**	**22.41**	**13.20**	**5.9801**	**1.96**	**66.82**	**4.18**	**1.88**	**5.7146**
CH4Net (71)	w/o	16.80	50.11	21.65	13.14	4.4825	2.45	53.25	5.29	2.34	3.8310
	AUGMENT	**18.07**	**52.43**	**23.18**	**14.35**	**3.9072**	**2.98**	**55.14**	**6.08**	**2.86**	**3.2648**
MultiSTARCOP (31)	w/o	22.84	25.77	23.01	13.10	1.3340	2.82	29.29	8.38	2.58	1.2451
	AUGMENT	**24.31**	**27.62**	**24.88**	**14.76**	**1.0406**	**3.34**	**31.56**	**9.41**	**3.12**	**0.9823**
Gasformer (35)	w/o	23.14	21.32	19.81	10.70	1.0179	2.94	25.36	9.72	2.44	1.1030
	AUGMENT	**24.92**	**23.85**	**21.74**	**12.06**	**0.7864**	**3.56**	**27.91**	**10.88**	**2.96**	**0.8545**
MethaneMapper (73)	w/o	23.07	17.14	20.10	9.65	1.1523	3.87	22.65	12.46	3.24	0.8380
	AUGMENT	**24.85**	**19.73**	**22.18**	**10.98**	**0.9037**	**4.56**	**24.94**	**13.92**	**3.88**	**0.6244**
FSDINet (77)	w/o	31.83	23.09	27.81	13.69	0.8603	5.73	32.49	11.49	4.28	0.7377
	AUGMENT	**33.62**	**25.81**	**30.02**	**15.74**	**0.4126**	**6.91**	**35.14**	**13.84**	**5.37**	**0.3915**
ALKANE	w/o	34.26	77.18	34.98	20.90	0.0160	8.16	68.12	15.78	7.04	0.0015
	AUGMENT	**36.81**	**79.46**	**37.92**	**23.14**	**0.0009**	**9.84**	**70.73**	**18.06**	**8.91**	**0.0004**

Note: "w/o" indicates the original baseline models trained without the proposed `AUGMENT` framework. "`AUGMENT`" denotes the corresponding models trained using the proposed augmentation-driven learning strategy.

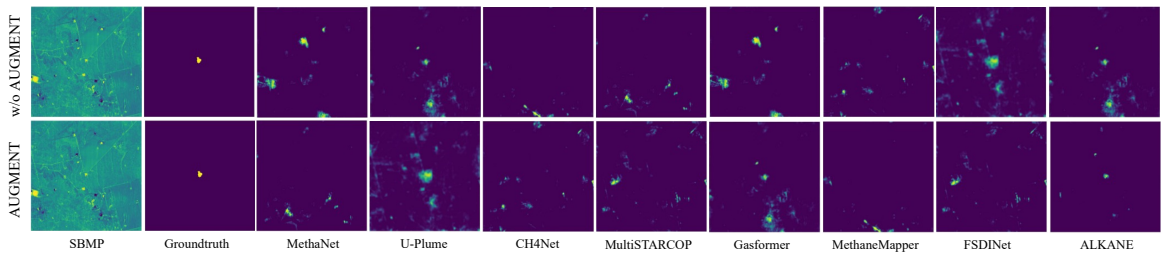


Figure 5.12: Visual comparison of methane plume segmentation results on weak plumes from the WV-3 dataset.

methane plume segmentation performance. As in the HSI denoising experiments, the notation "w/o" denotes the original baseline models trained only with conventional augmentation strategies, without incorporating the proposed AUGMENT framework. The comparison is conducted through quantitative analysis using standard segmentation metrics, complemented by qualitative visual inspection of segmentation outputs. Furthermore, cross-sensor generalization capability is examined by evaluating the trained models directly on the unseen EMIT dataset without additional fine-tuning.

5.3.3.1 Quantitative Comparison

Tables 5.6, 5.7, and 5.8 present the quantitative comparison results obtained on the AVIRIS-NG, WV-3, and EMIT datasets, respectively. The experimental results demonstrate that integrating the proposed AUGMENT framework consistently improves methane plume segmentation performance across both strong and weak plume categories. To ensure statistical consistency, all experiments are repeated five times using different random seeds while preserving identical preprocessing steps, data partitions, and optimization settings. The results indicate that incorporating AUGMENT improves Mean Precision, Mean Recall, Mean F1-score, and Mean Intersection over Union, while reducing the Mean False Positive Rate across different baseline architectures.

In particular, the proposed augmentation-driven representation learning strategy improves segmentation networks' ability to capture plume boundaries and preserve spatial continuity, especially under challenging weak-plume conditions, where spectral responses are relatively subtle. Moreover, the framework demonstrates enhanced robustness against spectral variability and background interference, leading to improved generalization across

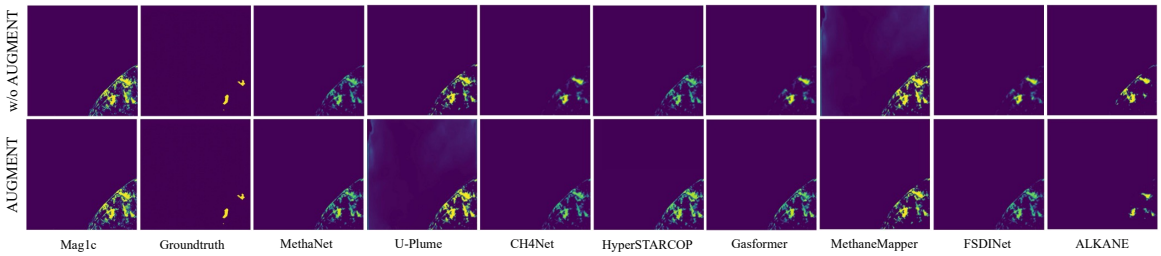


Figure 5.13: Visual comparison of methane plume segmentation results on strong plumes from the EMIT dataset.

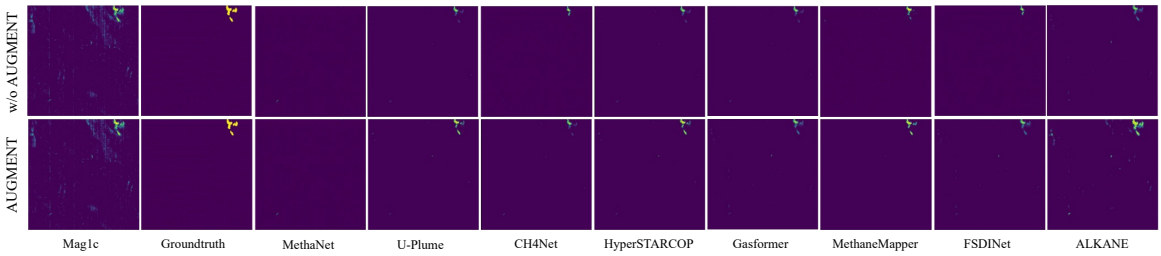


Figure 5.14: Visual comparison of methane plume segmentation results on weak plumes from the EMIT dataset.

different sensors and acquisition conditions. A segmentation-to-emission analysis conducted on the AVIRIS-NG test set further demonstrates the practical applicability of the proposed framework. The Pearson correlation coefficient between segmentation-derived emission estimates and archival emission measurements indicates that the proposed augmentation strategy effectively supports downstream methane emission quantification tasks.

5.3.3.2 Qualitative Comparison

Qualitative comparisons illustrated in Figs. 5.13, 5.14, 5.11, and 5.12 demonstrate that the integration of `AUGMENT` produces more complete and spatially coherent plume regions compared with baseline models trained without the proposed strategy. The segmentation outputs generated using `AUGMENT` exhibit fewer false positives and false negatives while preserving finer plume boundaries and structural continuity.

The visual results further indicate that the proposed framework effectively captures both global contextual dependencies and fine-grained spatial structures, enabling robust methane plume segmentation under challenging conditions such as weak plume scenarios and low-resolution multispectral imagery.

Table 5.8: Quantitative comparison of methane plume segmentation performance on the EMIT dataset for strong and weak plumes.

Methods		**Strong Plumes**					**Weak Plumes**				
		MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓	MP ↑	MR ↑	MF1 ↑	MIoU ↑	MFPR ↓
MethaNet (70)	w/o	50.25	26.59	30.08	23.90	0.0848	8.17	28.75	11.64	6.41	0.1189
	AUGMENT	**52.03**	**28.44**	**32.11**	**25.62**	**0.0735**	**9.52**	**30.61**	**13.08**	**7.34**	**0.1021**
U-Plume (72)	w/o	56.93	36.26	37.72	29.07	0.0785	20.07	6.29	9.04	3.28	0.0531
	AUGMENT	**58.76**	**38.19**	**39.84**	**31.12**	**0.0667**	**21.85**	**7.88**	**10.94**	**4.12**	**0.0418**
CH4Net (71)	w/o	60.42	35.48	33.31	23.93	0.1814	12.14	23.20	20.80	9.57	0.0309
	AUGMENT	**62.31**	**37.42**	**35.16**	**25.84**	**0.1620**	**13.88**	**25.37**	**22.61**	**10.92**	**0.0241**
HyperSTARCOP (31)	w/o	64.48	44.45	45.87	33.95	0.1313	21.46	15.88	18.78	9.23	0.0376
	AUGMENT	**66.22**	**46.38**	**47.91**	**35.84**	**0.1186**	**23.28**	**17.94**	**20.63**	**10.88**	**0.0294**
Gasformer (35)	w/o	67.08	58.77	52.82	38.70	0.4186	4.94	27.36	11.97	4.26	0.1434
	AUGMENT	**69.41**	**60.82**	**55.18**	**40.93**	**0.3721**	**6.43**	**29.74**	**13.81**	**5.22**	**0.1197**
MethaneMapper (73)	w/o	73.50	62.35	58.01	43.72	0.4511	7.81	21.92	14.81	5.57	0.1355
	AUGMENT	**75.62**	**64.71**	**60.84**	**46.03**	**0.3914**	**9.54**	**24.13**	**16.95**	**6.81**	**0.1148**
FSDINet (77)	w/o	74.26	72.39	60.72	47.53	0.1365	24.86	48.13	21.16	10.01	0.1133
	AUGMENT	**75.94**	**74.63**	**62.98**	**49.86**	**0.0917**	**26.74**	**50.58**	**24.03**	**11.84**	**0.0746**
ALKANE	w/o	76.52	78.11	64.13	52.47	0.0098	27.01	51.64	28.33	13.17	0.0056
	AUGMENT	**78.84**	**80.26**	**67.41**	**55.92**	**0.0032**	**29.46**	**54.28**	**31.74**	**15.82**	**0.0018**

Note: "w/o" indicates the original baseline models trained without the proposed AUGMENT framework. "AUGMENT" denotes the corresponding models trained using the proposed augmentation-driven learning strategy.

Table 5.9: Ablation analysis of different AUGMENT components under Gaussian noise on the ICVL dataset.

Components			σ = 30		σ = 50		σ = 70	
Flip	**Rotation**	**Inpainting**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
×	×	×	44.84	0.9973	44.39	0.9968	43.46	0.9956
✓	×	×	44.98	0.9974	44.45	0.9968	43.49	0.9957
×	✓	×	44.91	0.9973	44.41	0.9968	43.45	0.9956
×	×	✓	45.12	0.9975	44.57	0.9969	43.60	0.9957
✓	✓	×	45.03	0.9974	44.49	0.9968	43.55	0.9957
✓	×	✓	45.22	0.9976	44.62	0.9969	43.75	0.9958
×	✓	✓	45.18	0.9975	44.59	0.9968	43.67	0.9958
✓	✓	✓	**45.46**	**0.9976**	**44.83**	**0.9970**	**43.86**	**0.9958**

Note: The table presents the contribution of flipping, rotation, and inpainting modules within the proposed AUGMENT framework. The best-performing configuration is highlighted in bold.

5.3.3.3 Zero-Shot Generalization on EMIT

To evaluate the cross-sensor generalization capability of the proposed framework, the trained models are directly tested on the EMIT dataset without performing any additional fine-tuning. As shown in Table 5.8 and Figs. 5.13 and 5.14, integrating AUGMENT consistently improves segmentation performance for both strong and weak methane plume categories.

Compared with the corresponding baseline models trained without AUGMENT, the augmented frameworks achieve higher segmentation accuracy while maintaining lower false positive rates. These observations demonstrate that the proposed augmentation-driven learning strategy enables the segmentation networks to preserve robust plume representations even under unseen spectral and spatial distributions, thereby improving cross-sensor generalization capability.

5.3.4 Model Complexity

The computational complexity of the methane plume segmentation frameworks is analyzed in terms of model parameters, computational cost, inference latency, throughput, GPU memory consumption, and training duration. Since the proposed AUGMENT framework is incorporated only during training, the inference-stage architecture remains unchanged. Consequently, the average inference latency, throughput, and GPU memory consumption remain

nearly identical to those of the original baseline models. Although incorporating AUGMENT slightly increases overall training time due to additional augmentation and representation-learning operations, the computational overhead remains relatively small compared with the resulting improvements in segmentation performance. These observations demonstrate that the proposed AUGMENT framework provides consistent performance gains while introducing minimal additional computational complexity, making it suitable for practical hyperspectral methane plume segmentation applications.

5.4 Ablation Studies

To analyze the contributions of different components within the proposed AUGMENT framework, an ablation study was conducted with the IDEATION backbone under varying Gaussian noise levels. The proposed framework generates augmented positive samples using three transformation strategies, namely flipping, rotation, and masking-based inpainting. The quantitative results corresponding to different component combinations are summarized in Table 5.9. The first row of the table represents the configuration in which all transformation components are disabled. In this setting, the backbone network is trained without the proposed augmentation strategies, while the remaining optimization pipeline is preserved. The obtained results indicate comparatively lower denoising performance when none of the proposed transformations are utilized. When each component is introduced individually, improvements can be observed in both PSNR and SSIM values. Furthermore, combining multiple components leads to additional performance gains, while the highest reconstruction quality is achieved when all three components are incorporated simultaneously within the AUGMENT framework.

Among the individual transformations, masking-based inpainting contributes noticeably to the overall improvement. This behavior can be attributed to the auxiliary reconstruction objective introduced by masked regions, which encourages the network to recover missing spectral-spatial information during training. Similarly, geometric augmentations such as flipping and rotation assist the model in learning invariant spectral-spatial representations under different orientations. The combination of these transformations therefore enables the

network to learn more discriminative representations for hyperspectral reconstruction. The proposed framework employs three key loss functions, L_{CL}, L_{DF}, and L_T. The contrastive loss (L_{CL}) enforces representation consistency between the original HSI and its transformed counterparts generated through flipping, rotation, and masking operations. By maximizing the similarity between these augmented representations, the framework encourages invariant feature learning across different spatial transformations and masked conditions.

On the other hand, the data fidelity loss (L_{DF}) constrains the reconstructed outputs to remain close to the clean ground-truth HSI. This objective ensures preservation of spectral-spatial information during the denoising process. The weighting coefficients associated with the contrastive loss were selected as $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3}$, such that equal importance is assigned to the three transformation branches while maintaining balanced optimization behavior. Similarly, for the data fidelity loss, the coefficients were chosen as $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \frac{1}{4}$ to equally balance the reconstruction constraints corresponding to the original and transformed outputs. The final optimization objective is defined as $L_T = \alpha L_{CL} + L_{DF}$, where $\alpha = 1$. Assigning equal importance to the contrastive and reconstruction objectives enables balanced optimization between feature representation learning and hyperspectral reconstruction quality. During experimentation, it was observed that increasing α beyond 1 caused the optimization process to emphasize representation similarity excessively, which negatively influenced reconstruction consistency. Conversely, using smaller values of α reduced the influence of supervised contrastive learning, causing the overall behavior to become similar to the baseline backbone models.

To further investigate the contribution of SCL, an additional ablation experiment was performed and the results are reported in Table 5.5. Three different configurations were considered: (B) the original baseline model without augmentation or contrastive learning, (B + A) the baseline combined with augmentation strategies only, and (B + SCL) the complete `AUGMENT` framework incorporating SCL. The results demonstrate that introducing augmentation strategies alone improves the denoising performance compared with the baseline configuration. However, incorporating SCL further improves the reconstruction quality across all evaluated noise levels. These observations indicate that the performance improvements are not solely caused by augmentation operations, but are additionally strengthened

through representation learning enabled by the supervised contrastive objective.

5.5 Summary

This chapter introduced AUGMENT, a unified augmentation-driven learning framework designed to improve HSID and methane plume segmentation through SCL. The proposed framework aims to learn more discriminative spatial-spectral representations by enforcing consistency among multiple augmented views of the same hyperspectral sample, thereby enhancing feature robustness and generalization. To achieve this objective, AUGMENT integrates several augmentation strategies, including flipping, rotation, and inpainting, within the SCL paradigm. In particular, the incorporation of HSI inpainting enables the generation of structured and semantically meaningful variations of the input data, allowing the network to better exploit contextual spatial-spectral relationships. This design is especially beneficial under limited training data conditions, where learning robust representations remains a challenging task.

Extensive experiments were conducted on both synthetic and real-world hyperspectral datasets for denoising applications. Quantitative and qualitative results demonstrated that integrating AUGMENT consistently improves the performance of a wide range of state-of-the-art denoising methods under Gaussian and complex noise scenarios. Additional analyses, including classification-based evaluation, sensitivity analysis, and ablation studies, further confirmed the effectiveness of the proposed framework. The generality of AUGMENT was further validated through methane plume segmentation experiments on the AVIRIS-NG, WV-3, and EMIT datasets. By integrating the proposed framework into existing segmentation architectures, consistent improvements were observed across multiple evaluation metrics, including precision, recall, F1-score, and intersection-over-union. Moreover, the framework enhanced segmentation quality for both strong and weak methane plumes while improving cross-sensor generalization performance under diverse sensing conditions.

Overall, the results presented in this chapter demonstrate that AUGMENT functions as a versatile plug-and-play framework that can be seamlessly incorporated into existing HSI analysis pipelines. By combining SCL with augmentation-driven representation learning and

hyperspectral inpainting, the proposed framework effectively improves denoising accuracy, strengthens methane plume segmentation performance, and promotes robust spatial-spectral feature learning across a variety of hyperspectral sensing applications.

Chapter 6

Conclusion and Future Work

This work has focused on advancing HSI analysis by developing DL frameworks for HSI denoising, CH_4 plume segmentation, and spatial-spectral representation learning. With the increasing availability of hyperspectral remote sensing data and the growing need for accurate environmental monitoring, there is a strong demand for methods that can effectively handle noise, exploit spectral-spatial information, and reliably detect methane emissions. However, challenges such as complex noise distributions, high inter-band redundancy, weak plume signatures, and limited training data continue to restrict the performance of existing approaches.

The contributions of this work span three major research directions, each addressing a different challenge in HSI analysis:

- **Hyperspectral Image Denoising:** We proposed `IDEATION`, a hybrid denoising framework that integrates 3D CNNs, 3D Transformer-based feature extraction, and 3D U-Net architecture. The framework jointly captures local spatial-spectral characteristics and global contextual dependencies, enabling effective restoration of hyperspectral images under complex and mixed noise conditions.

- **Methane Plume Segmentation:** This work introduced `ALKANE`, a frequency-aware Transformer framework for methane plume segmentation from hyperspectral imagery. By jointly exploiting spatial, spectral, and spatial-frequency representations, the framework improves plume localization, boundary delineation, and detection performance

for both strong and weak methane plumes across different sensing environments.

- **Supervised Contrastive Learning and Inpainting-based Representation Learning:** We proposed `AUGMENT`, a unified framework that integrates supervised contrastive learning and HSI inpainting with hyperspectral image denoising and methane plume segmentation. The framework enhances spectral-spatial representation learning through structured augmentations and contextual reconstruction, leading to improved denoising and downstream plume segmentation performance, particularly with limited training data.

Each proposed framework was extensively evaluated across multiple publicly available hyperspectral datasets acquired under varying sensing conditions. The experimental results demonstrate that it is possible to develop accurate, robust, and generalizable models by effectively leveraging complementary spatial, spectral, contextual, and frequency-domain information. One of the key observations throughout this work is that incorporating complementary learning paradigms consistently improves performance. While IDEATION benefits from combining CNN and Transformer-based feature extraction, ALKANE demonstrates the advantages of integrating spatial and frequency-domain representations for plume segmentation. Similarly, AUGMENT highlights the effectiveness of SCL and inpainting-based augmentation for learning discriminative spectral-spatial representations. These findings emphasize the importance of robust representation learning for HSI analysis tasks.

Despite the promising results, several limitations remain. First, the proposed frameworks have primarily been evaluated on benchmark datasets, and their performance on large-scale operational deployments remains to be investigated. Second, the computational complexity associated with deep Transformer-based architectures may pose challenges for real-time applications and resource-constrained platforms. Third, accurately identifying extremely weak methane plumes under highly complex environmental backgrounds remains a difficult problem requiring further research. This work opens several promising directions for future investigation:

- **Self-Supervised and Foundation Model Learning:** Future research may explore self-supervised learning and foundation-model paradigms to reduce dependence on

labeled hyperspectral datasets and improve cross-sensor generalization.

- **Joint Optimization of Denoising and Segmentation:** Developing end-to-end frameworks that simultaneously perform hyperspectral image denoising and methane plume segmentation may further enhance downstream performance.
- **Multimodal Remote Sensing Analysis:** Integrating hyperspectral imagery with complementary modalities such as thermal, multispectral, or LiDAR data could provide richer scene understanding and improve methane monitoring capabilities.
- **Efficient Deployment:** Future studies may investigate lightweight architectures, model compression techniques, and hardware-aware optimization strategies to enable deployment on edge devices and real-time monitoring systems.
- **Advanced Representation Learning:** Further exploration of contrastive learning, masked modeling, generative learning, and frequency-domain feature learning may lead to more robust and transferable spectral-spatial representations.

In conclusion, this work contributes toward advancing HSI analysis by addressing three interconnected challenges: HSI denoising, methane plume segmentation, and representation learning. Through the development of `IDEATION`, `ALKANE`, and `AUGMENT`, this work demonstrates how spatial-spectral learning, frequency-domain analysis, and SCL can be effectively combined to improve the quality and reliability of HSI understanding. The methodologies presented in this work provide a foundation for future research to develop accurate, robust, and scalable solutions for remote sensing and environmental monitoring applications.

Bibliography

[1] L. S. Kalman and E. M. Bassett III, "Classification and material identification in an urban environment using hydice hyperspectral data," in *Imaging Spectrometry III*, vol. 3118. SPIE, 1997, pp. 57–68.

[2] T. Zhang, Y. Fu, and C. Li, "Hyperspectral image denoising with realistic data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2248–2257.

[3] A. Bhargava, A. Sachdeva, K. Sharma, M. H. Alsharif, P. Uthansakul, and M. Uthansakul, "Hyperspectral imaging and its applications: A review," *Heliyon*, vol. 10, no. 12, 2024.

[4] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[5] Z. P. Lee, W. J. Rhea, R. Arnone, and W. Goode, "Absorption coefficients of marine waters: Expanding multiband information to hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 118–124, 2005.

[6] S. Bhadra, V. Sagan, S. Sarkar, M. Braud, T. C. Mockler, and A. L. Eveland, "Prosail-net: A transfer learning-based dual stream neural network to estimate leaf chlorophyll and leaf angle of crops from uav hyperspectral images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 210, pp. 1–24, 2024.

[7] J. Transon, R. d'Andrimont, A. Maugnard, and P. Defourny, "Survey of hyperspectral

earth observation applications from space in the sentinel-2 context," *Remote Sensing*, vol. 10, no. 2, p. 157, 2018.

[8] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[9] B. Fei, "Hyperspectral imaging in medical applications," in *Data handling in science and technology*. Elsevier, 2019, vol. 32, pp. 523–565.

[10] K. Bi, S. Xiao, S. Gao, C. Zhang, N. Huang, and Z. Niu, "Estimating vertical chlorophyll concentrations in maize in different health states using hyperspectral lidar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 8125–8133, 2020.

[11] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1205–1218, 2018.

[12] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8032–8041.

[13] M. A. Calin, A. C. Calin, and D. N. Nicolae, "Application of airborne and spaceborne hyperspectral imaging techniques for atmospheric research: Past, present, and future," *Applied Spectroscopy Reviews*, vol. 56, no. 4, pp. 289–323, 2021.

[14] M. Saunois, A. Martinez, B. Poulter, Z. Zhang, P. A. Raymond, P. Regnier, J. G. Canadell, R. B. Jackson, P. K. Patra, P. Bousquet *et al.*, "Global methane budget 2000–2020," *Earth System Science Data*, vol. 17, no. 5, pp. 1873–1958, 2025.

[15] Q. Sun, T. Liu, J. Xu, H. Li, and M. Huang, "Rapid recognition and concentration prediction of gas mixtures based on smlp," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–9, 2024.

[16] M. F. Gard, "A status report on environmental monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 4, pp. 782–785, 2002.

[17] I. Karakurt, G. Aydin, and K. Aydiner, "Sources and mitigation of methane emissions by sectors: A critical review," *Renewable energy*, vol. 39, no. 1, pp. 40–48, 2012.

[18] Q. Wang, Y. Sun, X. Pan, M. Xing, and Y. Zhao, "Optical gas imaging for accurate leakage source measurement based on optical flow analysis," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[19] M. Sobanaa, R. Prathiviraj, J. Selvin, and M. Prathaban, "A comprehensive review on methane's dual role: effects in climate change and potential as a carbon–neutral energy source," *Environmental Science and Pollution Research*, vol. 31, no. 7, pp. 10 379–10 394, 2024.

[20] M. Harmsen, D. P. van Vuuren, B. L. Bodirsky, J. Chateau, O. Durand-Lasserve, L. Drouet, O. Fricko, D. E. Gernaat, T. Hanaoka *et al.*, "The role of methane in future climate strategies: mitigation potentials and climate impacts," *Climatic Change*, vol. 163, pp. 1409–1425, 2020.

[21] E. Nisbet, R. Fisher, D. Lowry, J. France, G. Allen, S. Bakkaloglu, T. Broderick, M. Cain, M. Coleman, J. Fernandez *et al.*, "Methane mitigation: methods to reduce emissions, on the path to the paris agreement," *Reviews of Geophysics*, vol. 58, no. 1, 2020.

[22] C. Ma, M. Wan, J. Wu, X. Kong, A. Shao, F. Wang, Q. Chen, and G. Gu, "Light self-gaussian-attention vision transformer for hyperspectral image classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.

[23] X. Zhang, Z. Feng, S. Yang, Y. Wei, L. Song, L. Yang, and L. Jiao, "Semantic correntropy representation for hyperspectral image classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023.

[24] Y. Mi, B. Tu, Y. Chen, Z. Cao, and A. Plaza, "Hyperspectral anomaly detection via anchor generation," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[25] C. Shi, P. Chen, L. Fang, M. Zhao, X. Hei, and Q. Miao, "Multil-scale spatial-frequency domain cross transformer for hyperspectral image classification," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[26] X. Cheng, M. Zhang, S. Lin, Y. Li, and H. Wang, "Deep self-representation learning framework for hyperspectral anomaly detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–16, 2023.

[27] H. Ji, Z. Zuo, and Q.-L. Han, "A divisive hierarchical clustering approach to hyperspectral band selection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[28] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, 2021.

[29] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *Journal of imaging*, vol. 5, no. 5, p. 52, 2019.

[30] F. Ullah, I. Ullah, K. Khan, S. Khan, and F. Amin, "Advances in deep neural network-based hyperspectral image classification and feature learning with limited samples: a survey: F. ullah et al." *Applied Intelligence*, vol. 55, no. 6, p. 370, 2025.

[31] V. Ržička, G. Mateo-Garcia, L. Gómez-Chova, A. Vaughan, L. Guanter, and A. Markham, "Semantic segmentation of methane plumes with hyperspectral machine learning models," *Scientific Reports*, vol. 13, no. 1, p. 19999, 2023.

[32] M. Li, J. Liu, Y. Fu, Y. Zhang, and D. Dou, "Spectral Enhanced Rectangle Transformer for hyperspectral image denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5805–5814.

[33] T. Wang, J. Lan, Z. Han, Z. Hu, Y. Huang, Y. Deng, H. Zhang, J. Wang, M. Chen, H. Jiang *et al.*, "O-Net: a novel framework with deep fusion of CNN and Transformer

for simultaneous segmentation and classification," *Frontiers in Neuroscience*, vol. 16, p. 876065, 2022.

[34] M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu, "Beyond self-attention: External attention using two linear layers for visual tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5436–5447, 2022.

[35] T. T. Sarker, M. G. Embaby, K. R. Ahmed, and A. AbuGhazaleh, "Gasformer: A transformer-based architecture for segmenting methane emissions from livestock in optical gas imaging," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5489–5497.

[36] C. Liu, S. Jia, H. Wu, D. Zeng, F. Cheng, and S. Zhang, "A spatial-frequency domain associated image-optimization method for illumination-robust image matching," *Sensors*, vol. 20, no. 22, p. 6489, 2020.

[37] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, "Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 190, pp. 196–214, 2022.

[38] L. Huang, Y. Chen, X. He, and P. Ghamisi, "Supervised contrastive learning-based classification for hyperspectral image," *Remote Sensing*, vol. 14, no. 21, p. 5530, 2022.

[39] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.

[40] L. Stival, R. da Silva Torres, and H. Pedrini, "Semantically-aware contrastive learning for multispectral remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 223, pp. 173–187, 2025.

[41] S. Hou, H. Shi, X. Cao, X. Zhang, and L. Jiao, "Hyperspectral imagery classification based on contrastive learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[42] S. Xia, X. Zhang, H. Meng, J. Fan, and L. Jiao, "Two-stream networks for contrastive learning in hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 1903–1920, 2023.

[43] Y. Ning, J. Peng, Q. Liu, Y. Huang, W. Sun, and Q. Du, "Contrastive learning based on category matching for domain adaptation in hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[44] T. Lu, Y. Hu, W. Fu, K. Ding, B. Bai, and L. Fang, "Scl-net: An end-to-end supervised contrastive learning network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[45] L. Birla, S. Shukla, A. K. Gupta, and P. Gupta, "ALPINE: Improving remote heart rate estimation using contrastive learning," in *IEEE Winter Conference on Applications of Computer Vision*, 2023, pp. 5029–5038.

[46] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Conference on Computer Graphics and Interactive Techniques*, 2000, pp. 417–424.

[47] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 730–742, 2018.

[48] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. Figueiredo, "Block-gaussian-mixture priors for hyperspectral denoising and inpainting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2478–2486, 2020.

[49] H. Xu, X. Zhang, H. Li, L. Xie, W. Dai, H. Xiong, and Q. Tian, "Seed the views: Hierarchical semantic alignment for contrastive representation learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3753–3767, 2022.

[50] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5366–5380, 2017.

[51] B. Du, Z. Huang, N. Wang, Y. Zhang, and X. Jia, "Joint weighted nuclear norm and total variation regularization for hyperspectral image denoising," *International Journal of Remote Sensing*, vol. 39, no. 2, pp. 334–355, 2018.

[52] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119–133, 2012.

[53] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[54] Y. Chang, L. Yan, and S. Zhong, "Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4260–4268.

[55] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6868–6877.

[56] H. Chen, G. Yang, and H. Zhang, "Hider: A hyperspectral image denoising transformer with spatial-spectral constraints for hybrid noise removal," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 7, pp. 8797–8811, 2024.

[57] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis *et al.*, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

[58] X. Cao, X. Fu, C. Xu, and D. Meng, "Deep spatial-spectral global reasoning network for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[59] T. Bodrito, A. Zouaoui, J. Chanussot, and J. Mairal, "A trainable spectral-spatial sparse coding model for hyperspectral image restoration," *Advances in Neural Information Processing Systems*, vol. 34, pp. 5430–5442, 2021.

[60] K. Wei, Y. Fu, and H. Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 363–375, 2020.

[61] H. Zeng, K. Feng, X. Zhao, J. Cao, S. Huang, H. Luong, and W. Philips, "Degradation-noise-aware deep unfolding transformer for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[62] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[63] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.

[64] A. K. Gupta, R. Kumar, L. Birla, and P. Gupta, "RADIANT: Better rPPG estimation using signal embeddings and Transformer," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4976–4986.

[65] M. Li, Y. Fu, and Y. Zhang, "Spatial-Spectral Transformer for hyperspectral image denoising," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, 2023, pp. 1368–1376.

[66] M. Wang, W. He, and H. Zhang, "A spatial–spectral transformer network with total variation loss for hyperspectral image denoising," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[67] D. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. Green, S. Kratwurst, T. Krings *et al.*, "Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane," *Atmospheric Measurement Techniques*, vol. 8, no. 10, pp. 4383–4397, 2015.

[68] A. K. Ayasse, A. K. Thorpe, D. A. Roberts, C. C. Funk, P. E. Dennison, C. Frankenberg, A. Steffke, and A. D. Aubrey, "Evaluating the effects of surface properties on methane retrievals using a synthetic airborne visible/infrared imaging spectrometer next generation (aviris-ng) image," *Remote sensing of environment*, vol. 215, pp. 386–397, 2018.

[69] A. K. Thorpe, D. A. Roberts, E. S. Bradley, C. C. Funk, P. E. Dennison, and I. Leifer, "High resolution mapping of methane emissions from marine and terrestrial sources using a cluster-tuned matched filter technique and imaging spectrometry," *Remote Sensing of Environment*, vol. 134, pp. 305–318, 2013.

[70] S. Jongaramrungruang, A. K. Thorpe, G. Matheou, and C. Frankenberg, "Methanet–an ai-driven approach to quantifying methane point-source emission from high-resolution 2-d plume imagery," *Remote Sensing of Environment*, vol. 269, p. 112809, 2022.

[71] A. Vaughan, G. Mateo-García, L. Gómez-Chova, V. Růžička, L. Guanter, and I. Irakulis-Loitxate, "Ch4net: a deep learning model for monitoring methane super-emitters with sentinel-2 imagery," *EGUsphere*, vol. 2023, pp. 1–17, 2023.

[72] J. H. Bruno, D. Jervis, D. J. Varon, and D. J. Jacob, "U-plume: automated algorithm for plume detection and source quantification by satellite point-source imagers," *Atmospheric Measurement Techniques*, vol. 17, no. 9, pp. 2625–2636, 2024.

[73] S. Kumar, I. Arevalo, A. Iftekhar, and B. Manjunath, "Methanemapper: Spectral absorption aware hyperspectral transformer for methane detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 609–17 618.

[74] D. Niu, X. Wang, X. Han, L. Lian, R. Herzig, and T. Darrell, "Unsupervised universal image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 22 744–22 754.

[75] L. Chen, Y. Fu, L. Gu, D. Zheng, and J. Dai, "Spatial frequency modulation for seman-

tic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[76] V. V. Batchu, M. Conserva, A. Wilson, A. M. Michalak, V. Gulshan, P. G. Brodrick, A. K. Thorpe, and C. V. Arsdale, "Global monitoring of methane point sources using deep learning on hyperspectral radiance measurements from emit," *arXiv preprint arXiv:2604.10094*, 2026.

[77] Y. Liu, H. Shi, S. Wu, H. Ye, X. Wang, Y. Han, and W. Xiong, "Frequency and spatial domain injection network for methane plumes semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[78] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 534–11 542.

[79] A. Dixit, A. K. Gupta, P. Gupta, S. Srivastava, and A. Garg, "UNFOLD: 3D U-Net, 3D CNN and 3D Transformer based Hyperspectral Image Denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–10, 2023.

[80] C. Cheng, D. Sun, Y. Yang, Z. Guo, and J. Peng, "Hyperspectral image denoising via low-rank tucker decomposition with subspace implicit neural representation," *Remote Sensing*, vol. 17, no. 16, p. 2867, 2025.

[81] Q. Zhang, Q. Yuan, M. Song, H. Yu, and L. Zhang, "Cooperated spectral low-rankness prior and deep spatial prior for hsi unsupervised denoising," *IEEE Transactions on Image Processing*, vol. 31, pp. 6356–6368, 2022.

[82] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2013.

[83] J. Ye, J. Wang, F. Xiong, J. Chen, and Y. Qian, "Deep equilibrium convolutional sparse coding for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[84] Q. Zhang, Y. Dong, Y. Zheng, H. Yu, M. Song, L. Zhang, and Q. Yuan, "Three-dimension spatial-spectral attention transformer for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[85] M. Shao, X. Tan, K. Shang, T. Liu, and X. Cao, "A hybrid model of state space model and attention for hyperspectral image denoising," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[86] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in neural information processing systems*, vol. 33, pp. 18 661–18 673, 2020.

[87] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1597–1607.

[88] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738.

[89] J. Kim, H. Lee, J. Chang, and S. M. Park, "Generalized supervised contrastive learning," *arXiv preprint arXiv:2206.00384*, 2022.

[90] T. Lee and S. Yoo, "Augmenting few-shot learning with supervised contrastive learning," *IEEE Access*, vol. 9, pp. 61 466–61 474, 2021.

[91] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 271–21 284, 2020.

[92] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[93] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2536–2544.

[94] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5505–5514.

[95] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009.

[96] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[97] H. Zeng, J. Cao, K. Feng, S. Huang, H. Zhang, H. Luong, and W. Philips, "Degradation-Noise-Aware Deep Unfolding Transformer for hyperspectral image denoising," *arXiv preprint arXiv:2305.04047*, 2023.

[98] X. Fu, Y. Guo, M. Xu, and S. Jia, "Hyperspectral image denoising via robust subspace estimation and group sparsity constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–10, 2023.

[99] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS transactions on circuits and systems*, vol. 8, no. 7, pp. 579–588, 2009.

[100] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Hyperspectral image classification using spectral-spatial lstms," *Neurocomputing*, vol. 328, pp. 39–47, 2019.

[101] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[102] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proceedings of the European Conference on Computer Vision*, 2016, pp. 19–34.

[103] L. Pang, W. Gu, and X. Cao, "TRQ3DNet: A 3D quasi-recurrent and transformer based network for hyperspectral image denoising," *Remote Sensing*, vol. 14, no. 18, p. 4598, 2022.

[104] P. Gamba, "A collection of data for urban area characterization," in *IEEE International Geoscience and Remote Sensing Symposium*, vol. 1, 2004, p. 72.

[105] D. A. Landgrebe, *Signal theory methods in multispectral remote sensing*. John Wiley & Sons, 2003, vol. 24.

[106] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Proceedings of the European Conference on Computer Vision*, 2010, pp. 210–223.

[107] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-iid noise structure," *IEEE Transactions on Cybernetics*, vol. 48, no. 3, pp. 1054–1066, 2017.

[108] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.

[109] W. D. Thompson and S. D. Walter, "A reappraisal of the kappa coefficient," *Journal of clinical epidemiology*, vol. 41, no. 10, pp. 949–958, 1988.

[110] B. J. Schuit, J. D. Maasakkers, P. Bijl, G. Mahapatra, A.-W. Van den Berg, S. Pandey *et al.*, "Automated detection and monitoring of methane super-emitters using satellite data," *Atmospheric Chemistry and Physics Discussions*, vol. 2023, pp. 1–47, 2023.

[111] Z. Jiang, J. Wu, H. Yu, L. Lan, D. Pan, and W. Gui, "End-to-end methane concentration measurement model with deep residual shrinkage network for direct absorption spectroscopy," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[112] B. Peng, H. Liu, L. Zheng, M. Zhu, N. Liao, and B. Du, "Study on cross-instrumentation gas concentration detection based on domain adaptation," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[113] S. Kumar, C. Torres, O. Ulutan, A. Ayasse, D. Roberts, and B. Manjunath, "Deep remote sensing methods for methane detection in overhead hyperspectral imagery," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1776–1785.

[114] H. Shanableh, M. B. A. Gibril, A. Mansour, A. Dixit, R. Al-Ruzouq *et al.*, "A comparative analysis of deep learning methods for ghaf tree detection and segmentation from uav-based images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 805–811, 2025.

[115] A. Dixit, N. Hosamani, P. Gupta, and A. Garg, "VISIONARY: Novel spatial-spectral attention mechanism for hyperspectral image denoising," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2025, pp. 2736–2745.

[116] B. Rouet-Leduc and C. Hulbert, "Automatic detection of methane emissions in multispectral satellite imagery using a vision transformer," *Nature Communications*, vol. 15, no. 1, p. 3801, 2024.

[117] C. Chen, M. Fan, Z. Wang, M. Liang, J. Tao, and L. Chen, "Mpsunet: a deep learning-based segmentation framework for methane plume detection with space-based hyperspectral and multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[118] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 2005.

[119] J. Sun, H. Zhang, X. Ma, R. Wang, H. Sima, and J. Wang, "Spectral–spatial adaptive weighted fusion and residual dense network for hyperspectral image classification," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 28, no. 1, pp. 21–33, 2025.

[120] D. H. Cusworth, R. M. Duren, A. K. Thorpe, W. Olson-Duvall, J. Heckler, J. W. Chapman, M. L. Eastwood, M. C. Helmlinger, R. O. Green, G. P. Asner *et al.*, "Intermittency of large methane emitters in the permian basin," *Environmental Science & Technology Letters*, vol. 8, no. 7, pp. 567–573, 2021.

[121] M. D. Foote, P. E. Dennison, A. K. Thorpe, D. R. Thompson, S. Jongaramrungruang, C. Frankenberg, and S. C. Joshi, "Fast and accurate retrieval of methane concentration from imaging spectrometer data using sparsity prior," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 6480–6492, 2020.

[122] G. Thuillier, M. Hersé, D. Labs, T. Foujols, W. Peetermans, D. Gillotay, P. Simon, and H. Mandel, "The solar spectral irradiance from 200 to 2400 nm as measured by the solspec spectrometer from the atlas and eureca missions," *Solar Physics*, vol. 214, pp. 1–22, 2003.

[123] D. J. Varon, D. Jervis, J. McKeever, I. Spence, D. Gains, and D. J. Jacob, "High-frequency monitoring of anomalous methane point sources with multispectral sentinel-2 satellite observations," *Atmospheric Measurement Techniques Discussions*, vol. 2020, pp. 1–21, 2020.

[124] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.

[125] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.

[126] M. F. Guerri, C. Distante, P. Spagnolo, F. Bougourzi, and A. Taleb-Ahmed, "Deep learning techniques for hyperspectral image analysis in agriculture: A review," *ISPRS Open Journal of Photogrammetry and Remote Sensing*, p. 100062, 2024.

[127] P. Dou, C. Huang, W. Han, J. Hou, Y. Zhang, and J. Gu, "Remote sensing image classification using an ensemble framework without multiple classifiers," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 208, pp. 190–209, 2024.

[128] Q. Zhang, Q. Yuan, J. Li, F. Sun, and L. Zhang, "Deep spatio-spectral bayesian posterior for hyperspectral image non-iid noise removal," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 125–137, 2020.

[129] Q. Xu, X. Gu, P. Li, and X. Gu, "Multi-task deep learning for quantifying methane emissions from 2-d plume imagery with low signal-to-noise ratio," *International Journal of Remote Sensing*, vol. 46, no. 2, pp. 704–727, 2025.

[130] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Transactions on Geoscience and remote sensing*, vol. 49, no. 3, pp. 973–980, 2010.

[131] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 178–188, 2015.

[132] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Machine learning*, vol. 110, no. 3, pp. 457–506, 2021.

[133] S. Shukla, L. Birla, A. K. Gupta, and P. Gupta, "Trustworthy medical image segmentation with improved performance for in-distribution samples," *Neural Networks*, vol. 166, pp. 127–136, 2023.

[134] Z. Li, K. Bi, Y. Wang, Z. Fang, and J. Zhang, "Supervised contrastive learning for open-set hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[135] R. Wong, Z. Zhang, Y. Wang, F. Chen, and D. Zeng, "HSI-IPNet: Hyperspectral imagery inpainting by deep learning with adaptive spectral extraction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4369–4380, 2020.

[136] D. Seo, J. Ho, and B. C. Vemuri, "Covariant image representation with applications

to classification problems in medical imaging," *International journal of computer vision*, vol. 116, pp. 190–209, 2016.

[137] P. Gamba, "A collection of data for urban area characterization," in *IEEE International Geoscience and Remote Sensing Symposium*, vol. 1. IEEE, 2004.

[138] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[139] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. Shapiro, P. Barloon, and A. F. Goetz, "The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data," *Remote sensing of environment*, vol. 44, no. 2-3, pp. 145–163, 1993.

[140] Y. Wu, W. Xu, and L. Zheng, "Hyperspectral image mixed noise removal via double factor total variation nonlocal low-rank tensor regularization," *Remote Sensing*, vol. 16, no. 10, p. 1686, 2024.

[141] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227–1243, 2017.

[142] K. Tan, E. Li, Q. Du, and P. Du, "An efficient semi-supervised classification approach for hyperspectral imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 97, pp. 36–45, 2014.