A Unified Frequency and Spatial Domain Network for Intelligent Remote Sensing Image Analysis

Lokeshwari P
Department of CSE
SRM IST
Ramapuram,Chennai,India
plokeshwari01@gmail.com

Shanmugapriya R
Department of CSE
SRM IST
Ramapuram,Chennai,India
shanmugapriya000000@gmail.com

Sreelekha AC
Department of CSE
SRM IST
Ramapuram,Chennai,India
sreelekhaac@gmail.com

Manju A Department
of CSE SRM IST
Ramapuram,Chennai,India
manjua1@srmist.edu.in

***Abstract*— The semantic segmentation of remote sensing images seeks to provide semantic labels for every pixel in remote sensing images to ensure accurate and fine-grained interpretations of objects on the ground. However, it still faces some challenges, such as interference from noise, limited inter- class difference, uneven scales of objects, and complicated correlations between spatial and frequency domains. The existing CNN networks have limitations in extracting long- range contextual information, and most of the existing transformer networks are mainly focused on feature modeling in the spatial domain, ignoring frequency domain features. Therefore, in this paper, a unified frequency and spatial domain network for intelligent remote sensing image analysis is presented to solve these problems. The key idea of this paper is to use a preprocessing-assisted frequency and spatial domain fusion strategy, in which frequency domain features are learned by using Fourier decomposition to emphasize high-frequency details, and spatial domain features are used to preserve global semantic structures. Moreover, a multi-scale context modeling mechanism is introduced to adapt to scale changes, and channel-spatial collaborative attention is used to optimize feature representations in different dimensions to improve the recognition accuracy of small targets. Significant experiments performed on benchmark remote sensing datasets show that the suggested method surpasses existing state-of-the-art approaches in terms of performance metrics. Ablation experiments also confirm the effectiveness of frequency-spatial fusion and multi-scale modeling in disambiguating inter-class confusion and improving semantic boundary accuracy.**

I. INTRODUCTION

Remote sensing image analysis has a great significance in a variety of applications, including environmental monitoring, urban planning, agricultural assessment, disaster management, and military surveillance. Due to the rapid advancement of satellite and aerial image acquisition technologies, remote sensing systems are capable of collecting a huge amount of high-resolution images. The effective extraction of information from images has become a critical challenge. Traditionally, remote sensing image analysis methods are based on various hand-crafted features, which are typically derived in the spatial domain as well as the frequency domain. In the spatial domain, texture, shape, and context are used, while in the frequency domain, various transforms like the Fourier transform and the wavelet transform are employed.

Although both the spatial and frequency domains are effective in providing important information, most of the traditional methods process images separately, which limits the effectiveness of the features as well as the performance of the image analysis systems. Recent advances in deep learning, particularly convolutional neural networks, have improved the performance of various remote sensing image processing tasks, including classification, detection, and segmentation. However, the standard CNN architectures tend to be mostly based on the spatial domain and may not be able to fully leverage the frequency domain characteristics, which are quite significant for the recognition of fine textures and multi-scale patterns.

In order to overcome these challenges, this paper presents a novel architecture called the Unified Frequency and Spatial Domain Network for intelligent analysis of images in the field of remote sensing. The proposed architecture has the capability to learn discriminative features from images in the spatial and frequency domains using a unified approach. The proposed network combines the strength of frequency domain transformations and spatial domain feature extraction methods, thus providing enhanced representation and analytical capability for images in the field of remote sensing.
The proposed network is expected to provide enhanced analytical capability for images in the field of remote sensing, and the experimental results prove its effectiveness in comparison to conventional networks that are based on spatial and frequency domain feature extraction methods. In addition, the proposed Unified Frequency and Spatial Domain Network is based on a novel and adaptive architecture for the analysis of images in the field of remote sensing, and it can be utilized for various applications such as image classification, object detection, semantic segmentation, and change detection. These features are very important in the identification of complex patterns in land cover, such as buildings, vegetation, water bodies, roads, and fields, among others. Instead of using Fourier transforms, the frequency branch uses a combination of Spatial Variance Features and wavelet transforms to learn structural intensity variations and patterns in the images. Specifically, the use of SVF is focused on the determination of the variation in pixel intensities in a given region, enabling the identification of texture irregularities, object boundaries, and surface heterogeneity.

II. RELATED WORKS

Mao and Lazaro.[1] proposed a framework for image segmentation using deep learning techniques in remote sensing images. In this framework, convolutional neural networks like U-Net, SegNet, and DeepLab were employed for extracting spatial and semantic features from satellite images. The main aim of this framework is to improve the classification of land cover and object boundaries in high- resolution images of the earth acquired using remote sensing technology. The performance of the proposed models was tested using benchmark images, and the performance of the models was evaluated using metrics like Pixel Accuracy and Intersection over Union (IoU).

Disadvantage: However, the proposed method needs large amounts of labeled data and high computational resources for model training, which makes the implementation challenging in certain environments where such data and resources are limited.

Zhang et al. [2] presented SAIP-Net, which is a network for remote sensing image segmentation by incorporating spectral information and spatial feature learning for remote sensing image segmentation. The network utilizes a spectral adaptive information propagation mechanism to effectively explore the relationship between spectral bands and improve feature representation for multi-spectral remote sensing images. The spectral information propagation mechanism effectively enhances the ability of the network to differentiate between classes with similar spatial features by propagating spectral information in different network layers. The performance of the network in remote sensing image segmentation was validated by conducting experiments using benchmark datasets for remote sensing images.

Disadvantage: However, the proposed SAIP-Net framework increases the complexity of the model due to the incorporation of the extra modules that handle the propagation of the spectral information, which increases the computational cost and the training time, making it difficult to implement the proposed framework in real-time systems.

Li et al. [3] presented a framework known as ISWSST (Index- Space Wave State Superposition Transform) for enhancing feature representation in remote sensing image segmentation tasks. In this framework, a new index space wave state superposition mechanism is employed for efficient modeling of spatial information in images. By transforming feature representations in a wave space, this framework can effectively model complex spatial dependencies in images. The results showed that this approach can enhance image segmentation accuracy by effectively differentiating similar land cover classes in remote sensing images.

Disadvantage: However, the use of the ISWSST framework results in additional computational costs, mainly because of the additional transformation and feature superposition operations, which might increase the processing time and require additional computational resources for training and deployment**.**

Mou et al. [4] presented RiFCN, for the segmentation of high-resolution images in the field of remote sensing, which is based on the recurrent fully convolutional network. In this approach, the recurrent connections and fully convolutional network were combined to progressively fuse multi-level features extracted from different layers of the network. The recurrent fusion mechanism is able to effectively learn low- level spatial features and high-level semantic information, which can be helpful in the segmentation of complex patterns in remote sensing images. The experimental results on benchmark datasets indicate that the proposed approach can obtain higher segmentation accuracy and improve the boundary detection compared with the traditional fully convolutional network approach.

Disadvantage: However, the feature fusion mechanism based on the repeating pattern has increased the complexity of the model's training process, which in turn has increased the computational cost of the model**.**

Chen et al. in [5], proposed a Wavelet Transform Feature Enhancement approach for remote sensing image segmentation, where wavelet transform techniques are employed for feature enhancement in remote sensing images. In this approach, feature extraction from satellite images is performed by decomposing them into multiple frequency components, which enables the model to effectively capture both texture details and spatial information in remote sensing images. By enhancing key features in remote sensing images, this approach effectively segments complex objects in remote sensing images. The results showed that feature enhancement using wavelet transform techniques improves the accuracy of remote sensing image segmentation.

Disadvantage: However, the process of using the wavelet transform requires additional steps in the computation process, which may result in increased complexity in the computation process and subsequently increase the time taken in the computation process.

Wang et al. [6]. proposed A frequency-aware adaptive filtering approach for the segmentation of images in the field of remote sensing. In this approach, the authors have tried to incorporate spatial as well as spectral features in the model for image segmentation. The frequency domain is utilized in this approach to improve the feature extraction capability of the model for the efficient segmentation of complex images in the field of remote sensing. Spatial features are combined with spectral features in the model for efficient image segmentation.

Disadvantage: However, the proposed approach involves extra processing in the frequency domain, which makes it less efficient in terms of computational complexity and the memory and processing requirements that it demands**.**

Yang et al. [7] proposed A novel deep learning framework named SFFNet, which stands for Spatial and Frequency Domain Fusion Network for image segmentation in the field of remote sensing images. The proposed framework is based on the fusion of spatial domain feature extraction and frequency domain representation learning. The proposed framework utilizes the separation of the high and low frequency features to learn the image features, which can effectively learn the image features and improve the image segmentation accuracy in the complex scene of remote sensing images. The proposed framework is also enhanced with the multiscale dual representation alignment filter to improve the image segmentation accuracy and the consistency of the image features. The proposed framework has achieved the state-of-the-art image segmentation accuracy compared with the other deep learning methods.

Disadvantage: However, the addition of the wavelet decomposition and the dual representation alignment increases the complexity of the model and the computational cost, which may result in slowing down the training process.

Gao et al. [8] proposed a paper "Adaptive Frequency Enhancement Network for Remote Sensing Semantic Segmentation" proposed the Adaptive Frequency Enhancement Network (AFENet), which is specifically designed for the task of semantic segmentation of images in the field of remote sensing, with a focus on the improvement of the interaction between spatial and frequency domain features. In addition, the authors of this paper proposed a selective feature fusion module for the integration of global contextual information and local detailed information for enhanced feature representation. The performance of the proposed network is evaluated using experimental results on benchmark datasets

Disadvantage: However, the use of adaptive frequency enhancement and fusion increases the complexity of network architecture, and this may require additional computational needs and possibly high hardware resources.

III. FREQUENCY SPATIAL DOMAIN APPROACH

In this project, we propose, a Unified Spatial– Frequency Domain Network enhanced with Spatial Variance Filtering (SVF) for intelligent remote sensing image segmentation. The primary objective of this work is to improve boundary localization, texture discrimination, and global structural understanding by integrating spatial and frequency representations within a single end-to-end deep learning framework. Unlike conventional spatial-only segmentation models, the proposed architecture emphasizes local intensity variations using SVF and combines them with frequency- domain structural cues to achieve richer feature representation. The model is specifically designed to handle high-resolution aerial and satellite imagery while maintaining robustness against noise, illumination changes, and heterogeneous land- cover patterns.

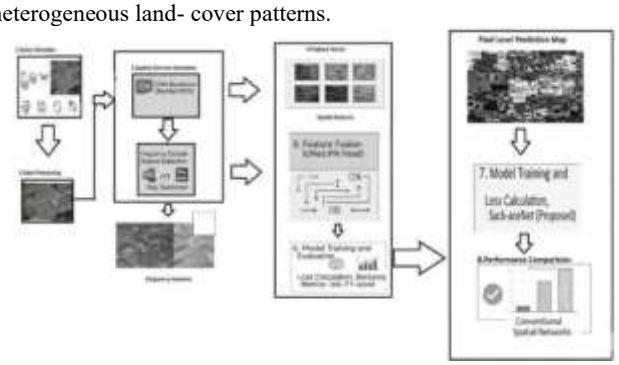

Fig 1. Architecture Diagram

A. Data Preparation and Preprocessing

The first module of the deep learning model is centered on data preparation and preprocessing to ensure that the images are standardized for use in deep learning. Benchmark datasets for remote sensing images in semantic segmentation are collected and prepared appropriately. The images are all resized to a fixed dimension to ensure uniformity in feature extraction from all images.

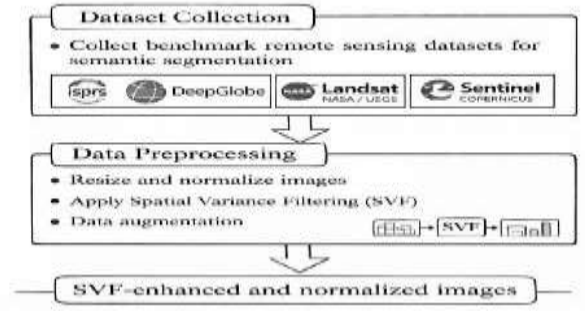

Fig 2. Data Preparation and Preprocessing

B. Dual-Domain Feature Extraction

The second module is in charge of extracting complementary information from both spatial and frequency domains. In the spatial domain branch, SVF-enhanced images undergo convolutional operations that extract hierarchical features from images, including edge, texture, and detailed structural feature information. In this module, we mainly emphasize learning local contextual information for precise image segmentation.

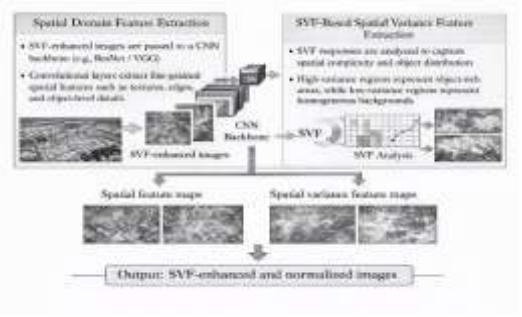

Fig 3. Dual-Domain Feature Extraction

In parallel with this, in the frequency domain branch, wavelet- based multi-scale analysis is conducted on SVF-enhanced images to extract global structural feature information from images, including object scale, spatial complexity, and global layout information in a scene. By learning both local spatial information and global frequency information from images in a joint fashion, a comprehensive feature representation is learned for precise land cover class discrimination from similar classes.

C. Feature Fusion and Segmentation Decoding

The third module combines the spatial and frequency features in order to produce the final segmentation results. The extracted spatial features are enhanced using the SVF and then fused with the frequency domain features in order to produce a unified feature map in the multi-domain representation.

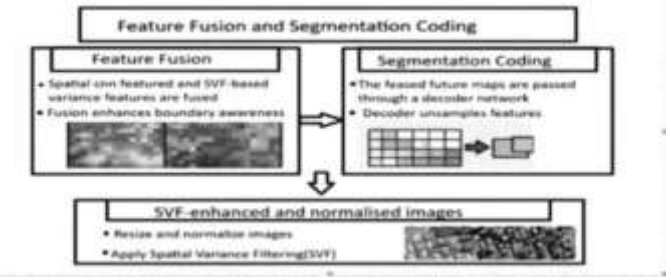

Fig 4. Feature Fusion Segmentation

The feature fusion is done through adaptive interactive operations such as the attention mechanism and feature concatenation. This allows the two features to interact and exchange information in a meaningful way. The fused feature map is then subjected to the segmentation decoding process in order to produce the pixel-level prediction map through the progressive sampling of the feature map.

D. Performance Stability and Robustness

Significant experimental results on various benchmark datasets, such as ISPRS Potsdam and Vaihingen datasets, have demonstrated that this methodology greatly reduces inter-class confusion. It has been observed by researchers that when frequency cues are added to the model, it is able to obtain a "global intuition" about the environment. This implies that it is unlikely to be misled by local irregularities such as shadows or temporarily placed vehicles, as it checks every detection in space against the spectral signature of the environment.

The pixel values are also normalized to ensure stability in training the model. An integral part of this module is the use of Spatial Variance Filtering (SVF), which increases local intensity variations, sharpens edges, and enhances textured areas, thereby removing any homogeneous backgrounds. Data augmentation techniques are also used in this module to increase the diversity of the dataset for use in training the model, thereby enhancing its ability to generalize well in different environmental conditions.

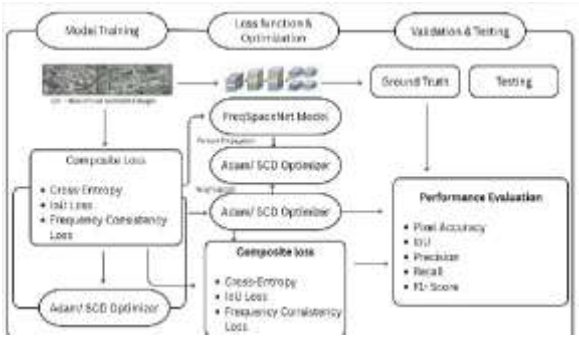

Fig 5. Model Training, Evaluation, and Performance Analysis

Moreover, the training process includes optimization using suitable loss functions and backpropagation for improving the learning ability of the network. In addition, the network learns to utilize the spatial and frequency domain features in an efficient manner, thus facilitating a deeper understanding of complex patterns in images related to remote sensing. The evaluation results provide further insight into the robustness and ability of the proposed architecture.

IV. RESEARCH AND ANALYSIS

Analysis of the research and performance of the FreqSpaceNet and similar unified frameworks indicates the capability to address "spectral confusion" issues wherein different objects, such as green roofs and grass, have the exact same appearance in the spatial domain but have unique "signatures" in the frequency domain. Based on the basic framework of the FreqSpaceNet architecture, the methodology is now focused on a highly integrated approach wherein the frequency and spatial domains are not just parallel paths but rather a unified system with the purpose of overcoming the most difficult challenges in the field of remote sensing, including inter-class similarity and the accurate delineation of small objects from complex backgrounds.

A. Frequency Band Self-Attention (FreqBand)

The methodology also introduces the FreqBand Self Attention mechanism as a different approach compared to the conventional spatial attention method, utilizing information from various frequency bands. The model processes frequency bands ranging from low-frequency structural information to high-frequency texture information, assigning different weights to this information based on different geographic regions.

For example, mid-frequencies are used to emphasize texture information such as forests, while unnecessary information such as "salt and pepper" noise, which might appear in the classification process, is reduced based on high-frequency information. In this context, the normalized input image I of size M x N is converted into the frequency domain, where a frequency shift operation is performed to shift the zero- frequency component to a more suitable position. The absolute spectrum of this image, denoted as

$M(u,v) = |F(u,v)|$, is then improved through a logarithmic scaling operation, where the result is denoted as

$M\_log(u,v) = \log(1 + M(u,v))$. After this, a high-frequency noise filtering operation might be performed, followed by a normalization operation to the range [0,1] to generate the final frequency feature map F, which might be used in the segmentation process.B. Multiscale Contextual Modeling

FreqSpaceNet employs a Multiscale Global Class Aware module for effectively dealing with large scale variations in objects, from smaller ones such as cars to larger ones such as airport terminals. In this module, dilated convolution operations help in increasing the receptive field without reducing spatial resolution. This enables class consistency so that larger objects are not segmented into smaller ones for incorrect classification while still being able to classify smaller objects correctly. In spatial frequency fusion, the algorithm takes two feature maps: spatial feature map S and frequency feature map F as inputs and generates a feature representation SVF. First, both spatial feature map S and frequency feature map F need to be ensured to be of the same dimension, followed by feature fusion using either a weighted fusion approach where

SVF = alpha*S + beta*F,

with alpha + beta = 1, or a concatenation fusion approach where SVF = Concatenate(S, F). Finally, feature fusion is normalized to produce SVF feature map for further processing in the segmentation model.

C. Synergistic Boundary Awareness

One of the significant aspects of the research conducted for the FreqSpaceNet approach is its focus on boundary sharpening. In regular networks, boundaries are often "blurred" because they are based on spatial gradients, which can be quite weak in satellite images with low contrast.

The approach of FreqSpaceNet is to utilize the phase information of the frequency domain to "guide" the spatial branch of the network. The phase information in the Fourier Transform is naturally related to the positions of edges and shapes in an image; thus, the network is able to make predictions with pixel accuracy even when the contrast between a building and the soil is very low.

D. Performance Stability and Robustness

Significant experimental results on various benchmark datasets, such as ISPRS Potsdam and Vaihingen datasets, have demonstrated that this methodology greatly reduces inter-class confusion. It has been observed by researchers that when frequency cues are added to the model, it is able to obtain a "global intuition" about the environment. This implies that it is unlikely to be misled by local irregularities such as shadows or temporarily placed vehicles, as it checks every detection in space against the spectral signature of the environment. This leads to stability in segmentation results, which are consistent regardless of sensor type and atmospheric conditions.

IV. CONCLUSION

In conclusion, the proposed Unified Frequency and Spatial Domain approach, operating in SVF mode, provides a comprehensive paradigm for advanced remote sensing image segmentation. While conventional deep learning models are based on the application of frequency-domain feature extraction, the SVF approach provides a comprehensive platform where both spatial structural features as well as frequency-domain features are incorporated into a single, unified platform. This provides the model with comprehensive

ability to incorporate fine local details as well as global dependencies associated with spectral frequency patterns. The system architecture, as presented in the context of the SVF approach, provides a comprehensive overview of ability of the model to incorporate both spatial as well as frequency- domain transformations, which are otherwise carried out independently of each other in conventional models. In SVF mode, the model provides accurate object-level discrimination as well as precise edge delineation, while the frequency-domain features provide robustness to the model against various types of noise, illumination, as well as scale inconsistencies.

Furthermore, the addition of the composite loss function—encompassing the cross-entropy loss function, IoU loss function, and frequency consistency loss function— guarantees the model's optimization in the pixel domain as well as the spectral domain. This results in the model's training in SVF mode being more stable and less susceptible to overfitting and complex urban and environmental characteristics.

The overall impacts and benefits of the SVF-based architecture are not limited to the model's performance in the segmentation task. Indeed, as Earth observation systems continue to produce ultra-high-resolution images with multiple spectral and temporal bands, the spatial-only model would be limited in its capacity to fully exploit the frequency information embedded within the datasets. The SVF model overcomes this limitation in the following manner. Indeed, the SVF model would allow the model to simultaneously perceive the spatial layout as well as analyze the frequency distribution in the spectral domain. This would be highly beneficial for the development of automated systems in the following domains: urban mapping, environmental monitoring, agricultural assessment, and disaster response systems. The modular design of FreqSpaceNet enables it to be combined with more advanced backbone networks or multi-scale feature extraction techniques, making it more suitable for potential advancements in EO technology in the near future. As more high-resolution, multi-spectral, and multi-temporal data sets emerge in the EO domain, the ability to effectively learn both local structural information and global contextual information is critical. SVF mode enables this by learning these two aspects in unison rather than in separate entities.

The proposed approach can be seen as contributing not only to the improvement in segmentation results but also in generalization, robustness, and scalability for a more intelligent remote sensing framework in the near future by effectively balancing spatial resolution with frequency domain awareness through the Unified Frequency-Spatial Domain framework driven by SVF.

REFERENCES

[1]. Mao & Lazaro (2025) – Deep Learning Models for Remote Sensing Segmentation. Mao and Lazaro proposed a deep learning–based framework using CNN architectures such as U-Net, SegNet, and DeepLab to extract spatial and semantic features from satellite imagery, improving land- cover classification

[2].Zhang et al. (2025) – SAIP-Net: Spectral Adaptive Information Propagation. Zhang and colleagues proposed SAIP-Net, which integrates spectral information with spatial feature learning to enhance feature representation and improve segmentation accuracy in multi-spectral remote sensing images.

[3]. Wang et al. (2025) – Frequency-Aware Adaptive Filtering for Segmentation. Wang and colleagues proposed a frequency- aware adaptive filtering approach that combines spatial and spectral features to enhance texture representation and improve segmentation performance in complex remote sensing scenes.

[4]. Li et al. (2024) – ISWSST: Index-Space Wave State Superposition Transform. Li and colleagues introduced ISWSST, which utilizes wave-based feature representation to capture multi-scale spatial patterns and improve segmentation accuracy in remote sensing images

[5]. Chen et al. (2023) – Wavelet Transform Feature Enhancement. Chen and colleagues proposed a wavelet transform–based feature enhancement method that decomposes images into multiple frequency components to improve texture representation and segmentation performance.

[6]. Mou et al. (2018) – RiFCN: Recurrent Fully Convolutional Network. Mou and Zhu proposed RiFCN, which integrates recurrent connections with fully convolutional networks to fuse multi-level spatial features for improved semantic segmentation of high- resolution remote sensing images..

[7]. Yang et al. (2024) – SFFNet: Spatial and Frequency Domain Fusion Network.Yang and colleagues proposed SFFNet, which combines spatial feature extraction with frequency-domain representation using wavelet decomposition to improve segmentation accuracy and feature consistency.

[8]. Gao et al. (2025) – Adaptive Frequency Enhancement Network (AFENet). Gao and colleagues proposed AFENet, which enhances interaction between spatial and frequency domain features using adaptive frequency enhancement and feature fusion modules to improve remote sensing image segmentation performance.