



Article

A High-Resolution Remote Sensing Road Extraction Method Based on the Coupling of Global Spatial Features and Fourier Domain Features

Hui Yang ¹, Caili Zhou ², Xiaoyu Xing ³, Yongchuang Wu ⁴ and Yanlan Wu ^{4,5,6,7,*}¹ Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China; yanghui@ahu.edu.cn² School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China; x22301098@stu.ahu.edu.cn³ Hubei Institute of Land Surveying and Mapping, Wuhan 430034, China; nynuxxy@163.com⁴ School of Artificial Intelligence, Anhui University, Hefei 230601, China; wyc_ahu@stu.ahu.edu.cn⁵ Information Materials and Intelligent Sensing Laboratory of Anhui Province, Hefei 230601, China⁶ Anhui Engineering Research Center for Geographical Information Intelligent Technology, Hefei 230601, China⁷ Engineering Center for Geographic Information of Anhui Province, Hefei 230601, China

* Correspondence: wuyanlan@ahu.edu.cn

Abstract: Remote sensing road extraction based on deep learning is an important method for road extraction. However, in complex remote sensing images, different road information often exhibits varying frequency distributions and texture characteristics, and it is usually difficult to express the comprehensive characteristics of roads effectively from a single spatial domain perspective. To address the aforementioned issues, this article proposes a road extraction method that couples global spatial learning with Fourier frequency domain learning. This method first utilizes a transformer to capture global road features and then applies Fourier transform to separate and enhance high-frequency and low-frequency information. Finally, it integrates spatial and frequency domain features to express road characteristics comprehensively and overcome the effects of intra-class differences and occlusions. Experimental results on HF, MS, and DeepGlobe road datasets show that our method can more comprehensively express road features compared with other deep learning models (e.g., Unet, D-Linknet, DeepLab-v3, DCSwin, SGCN) and extract road boundaries more accurately and coherently. The IOU accuracy of the extracted results also achieved 72.54%, 55.35%, and 71.87%.

Keywords: road extraction; remote sensing; frequency domain learning; multi-perspective learning



Citation: Yang, H.; Zhou, C.; Xing, X.; Wu, Y.; Wu, Y. A High-Resolution Remote Sensing Road Extraction Method Based on the Coupling of Global Spatial Features and Fourier Domain Features. *Remote Sens.* **2024**, *16*, 3896. <https://doi.org/10.3390/rs16203896>

Academic Editors: Valerio Baiocchi, Alessandro Mei and Xianfeng Zhang

Received: 24 September 2024

Revised: 16 October 2024

Accepted: 18 October 2024

Published: 20 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Road extraction is an important topic in the field of remote sensing [1], which can provide essential data support for disaster emergency responses [2], urban planning [3], digital city construction [4], and autonomous driving [5]. However, a road presents narrow strip-shaped characteristics on a remote sensing image and is obstructed by shadows, trees, etc. [6], which makes remote sensing road extraction still subject to many challenges.

Early road extraction methods in remote sensing were primarily based on manually crafted features, which include threshold segmentation [7] and edge detection [8]. Threshold segmentation methods [9,10] are based on the grayscale or color features of remote sensing images, setting appropriate thresholds to distinguish roads from other objects and achieve road extraction. Edge detection methods [8,11] mainly utilize edge detection algorithms such as Sobel, Canny, Hough, etc., to extract road information. These methods can meet the requirements of remote sensing road extraction under single background conditions. However, all of these methods suffer from the problem of heavily relying on handcrafted features and expert experience.

Fully Convolutional Networks (FCNs) enable end-to-end pixel-level semantic segmentation while retaining spatial information [12] and have become a mainstream method for road extraction based on deep learning [13,14]. These include proposed road extraction methods based on classical semantic segmentation architectures such as UNet [15–17], DeepLab [18], and feature pyramid structures [19]. For example, to obtain multi-scale features and overcome the influence of occlusion, C-UNet [20] combines multi-scale dense atrous convolutions with UNet to obtain multi-scale road features and improve the accuracy of road segmentation. To tackle the issue of losing elongated road features in deeper layers, Res-UNet [21] introduces residual connections based on UNet to improve the capability of deep-level road feature extraction in complex scenes. Similarly, LDMM [22] integrates Dirichlet mixture models based on UNet to learn deeper road features for more accurate extraction results. To further enhance the ability of UNet to capture local road features, RDUN [23] uses residual dense blocks to aggregate local features and construct hierarchical features, improving road segmentation. Then, in remote sensing images, different feature channels often contain distinct semantic information. In order to enhance the utilization rate of DeepLab-based architecture channels, Nested SE-Deeplab [24] uses the SE module to apply weights to different feature channels and performs multi-scale upsampling to preserve and fuse shallow and deep information, improving road extraction results. Additionally, to address the insufficient fusion of shallow and deep features in the FPN structure, RoadCapsFPN [25] constructs a capsule feature pyramid network that extracts and integrates multiscale capsule features to restore high-resolution, semantically rich road features. DFPN [26] merges shallow and deep feature channels through deep fusion, enabling the learning of layered deep detail features. Both of these methods effectively address the issue of inadequate feature extraction from deep and shallow layers. However, in addition to the difference in multi-scale features, roads usually appear as meandering coherence, and the effective perception of their global features is also crucial for accurate road extraction. Although the above methods employ various strategies to improve FCN-based approaches for road extraction, the following key limitation remains: FCN architectures' relatively small receptive fields prevent them from effectively capturing global road information. Vision Transformer (ViT) [27], which uses multi-head attention to describe images, offers superior multi-scale and global information capture compared with FCN-based methods [28,29]. Transformer-based road extraction methods have subsequently been proposed [30,31]. For instance, RoadCT [32], GLNet [33], and DOSA [34] introduce local information description modules based on Transformer architectures, enabling the representation of both global and local detail features for improved road extraction accuracy. RoadTransNet [35] combines multi-head self-attention, cross-attention, and skip connections to fuse local and global context features, ensuring complete road extraction. The DRCNet [36] employs DenseNet-121 as its encoder, combined with Recurrent Criss-Cross Attention and Convolutional Block Attention Module, addressing the challenges posed by complex road geometries as well as vegetation and structural obstacles. These methods have improved the results of road extraction. Akhtarmanesh [37] adopted hard attention (data preprocessing) and soft attention (attention modules in the model) to enhance the model's focus on roads, thereby addressing the bias in the dataset. Jamali [38] integrated the advantages of HetConv, residual learning, the UNet architecture, and NAT, forming a novel deep semantic segmentation method that effectively addresses the challenge of accurately extracting road information from aerial images. Sundarapandi [39] combined quantum computing with dilated convolutions and introduced the Archimedes optimization algorithm to optimize network parameters. This integration allows the network to capture road features in remote sensing images more effectively, thus enhancing the accuracy and efficiency of road extraction. However, the aforementioned methods still suffer from issues such as inefficiency. To address this, Toni [40] proposed a deep learning architecture named AM-Unet, which improves the UNet architecture by refining the designs of its encoder, decoder, and skip connections. This method fuses low-level and high-level features and employs an attention mechanism

to weigh each channel, thereby enhancing computational efficiency and the accuracy of road information extraction.

However, in complex remote sensing images, different road information often exhibits varying frequency distributions and texture characteristics, and it is often difficult to express the comprehensive characteristics of roads effectively from a single spatial domain perspective [41]. The Frequency-to-Spectrum Mapping method [42–44] incorporates frequency domain features by transforming spatial domain images into the Fourier frequency domain, effectively separating high-frequency and low-frequency information. In complex environments, this method enhances edge information while reducing noise interference through frequency-to-spectrum mapping. Additionally, the Blind Block Reconstruction Network [45] employs a blind block reconstruction mechanism combined with a guard window, which minimizes the influence of center pixels on anomaly representation, allowing the model to focus more on learning background pixels. This approach not only improves adaptability to complex backgrounds but also enhances the extraction of road information of various sizes through multi-scale analysis, leading to significant detection performance in complex remote sensing imagery. Related studies [46,47] have shown that combining frequency domain and spatial domain feature learning significantly outperforms spatial domain feature learning with similar network structures.

Fourier transform can transform images from the spatial domain to the frequency domain [48]. Using Fourier transform to construct a frequency domain learning module, then utilizing frequency domain learning to achieve the separation of high-frequency and low-frequency information [49–51], can enhance the difference between target boundaries and backgrounds and reduce noise interference [52]. Therefore, this article proposes a road extraction method that couples global spatial learning with Fourier frequency domain learning. This method first utilizes a transformer to capture global road features. Then, it applies Fourier transform to separate and enhance high-frequency and low-frequency information. Finally, it integrates spatial and frequency domain features to express road characteristics comprehensively and overcome the effects of intra-class differences and occlusions. The main contributions of this work are as follows:

- (1) We propose a road extraction method that couples global spatial features with Fourier domain features to express road features from multiple perspectives and improve the separability of roads from other objects by integrating spatial domain features and frequency domain features.
- (2) We construct Fourier feature blocks to achieve the separation of high-frequency and low-frequency road information, enhance the distinction between roads and the background, and improve the feature expression of roads from a frequency domain perspective.
- (3) We explore the impact of the Fourier feature blocks on road extraction across different layers of the network.

The rest of this paper is organized as follows. Section 2 introduces the construction of our model and the implementation of the Fourier feature block. Section 3 provides a detailed explanation of the experimental results and analysis. Section 4 presents the ablation study results and analysis. Finally, Section 5 concludes this paper.

2. Methodology

2.1. Fourier Transform

In remote sensing, discrete signals on raster images can be analyzed and computed in-depth using harmonic analysis and spectral graph theory. Fourier transform, as a fundamental tool in signal processing, plays a crucial role by transforming signals from the time domain to the frequency domain, enabling us to analyze and process data from a different perspective. It decomposes the image into multiple frequency components, revealing the image's characteristics in the frequency domain [53]. Specifically, Fourier transform achieves the conversion of images from the spatial domain to the frequency domain, while the inverse Fourier transform is responsible for restoring frequency domain

information to the spatial domain. Together, they form the basic framework for frequency analysis in image processing, as shown in Equations (1) and (2).

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy \quad (1)$$

$$f(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{j2\pi(ux+vy)} dx dy \quad (2)$$

FFT is an optimized version of the Discrete Fourier Transform (DFT) algorithm. When implemented correctly, FFT can significantly reduce computation time compared with slow DFT, thus increasing processing speed [54]. Therefore, this paper utilized FFT to transform signals or images into the frequency domain, allowing us to extract spectral features for road tasks and utilize these features to train deep learning models, as shown in Equations (3) and (4).

$$F(u, v) = \sum_{M_x=0}^{M_x-1} \sum_{M_y=0}^{M_y-1} f(x, y) e^{-j2\pi(\frac{ux}{M_x} + \frac{vy}{M_y})}, \quad (3)$$

$$u, v = 0, 1, 2, \dots, Mu - 1 \mid Mv - 1$$

$$f(x, y) = \frac{1}{M_u \cdot M_v} \sum_{M_u=0}^{M_u-1} \sum_{M_v=0}^{M_v-1} F(u, v) e^{-j2\pi(\frac{ux}{M_x} + \frac{vy}{M_y})}, \quad (4)$$

$$x, y = 0, 1, 2, \dots, Mx - 1 \mid My - 1$$

As shown in Figure 1, remote sensing images undergo Fourier transformation, transforming them from the spatial domain to the frequency domain. In the frequency domain, they can be further decomposed into low-frequency and high-frequency components, similar to natural images [55–57]. The low-frequency component represents regions in the image where brightness or grayscale changes slowly, corresponding to large flat areas in the image. It describes the main parts of the image. On the other hand, the high-frequency component corresponds to areas with rapid changes in the image, such as edges, contours, noise, and details. In other words, detailed texture information in the image primarily exists in the high-frequency component, while rich global information is stored in the low-frequency component.

Figure 2 illustrates the changes in the relationship between amplitude and frequency components in the Fourier domain by adjusting the filter radius ratio, along with the corresponding frequency component images. Studies have shown significant differences in low-frequency and high-frequency images obtained after applying different filter radius ratios. In particular, the application of a smaller filter radius ratio results in low-pass filtering, causing the image to become blurred while reducing internal differences within the same land cover type. Conversely, high-pass filtering enhances the high-frequency components of the image, thereby strengthening the delineation of boundaries between different land cover types. Furthermore, increasing the filter radius ratio can increase internal differences within land cover types, revealing richer texture details. Different land cover types exhibit varying sensitivities to frequency; thus, selecting an appropriate filter radius ratio is crucial for balancing boundary features and internal consistency of land cover types.

2.2. Network Architecture

Figure 3 illustrates the details of our model's architecture, which integrates elements such as patch partitioning, Aggregated Feature Integration (AFI) blocks, Swin Transformer blocks, Fourier feature blocks, and skip connections.

Encoding Stage: In the encoding stage, the input road image is first divided into small patches via patch partitioning, where each patch contains four adjacent pixels. This transforms the original size from $H \times W \times 3$ to $H/4 \times W/4 \times 48$. This ensures that both fine

and extensive road features are retained. These patches are then fed into a linear embedding layer and Swin Transformer modules to extract road features, helping capture long-range dependencies and local contextual information within road images. Subsequently, Fourier feature blocks perform frequency domain analysis. The Fourier feature blocks process significant frequency components related to road textures and contours, ensuring effective capture and emphasis of the smooth structural features and detailed textures of roads. Next, a downsampling operation is performed through a 3×3 layer with a stride of 2. The combination of Fourier feature blocks and Swin Transformer blocks in the downsampling process is repeated twice, with the latter two layers employing only Swin Transformer blocks to enhance global features and avoid excessive filtering of frequency components. Fourier feature blocks utilize a dynamic frequency domain filtering mechanism, combining multi-scale low-pass and high-pass filters specifically designed to eliminate non-road target frequency components in the Fourier spectrum while retaining and enhancing signals within the desired frequency range. When these features are inverse-transformed back to the image feature space, they integrate global and local road information, highlighting important semantic features of the roads.

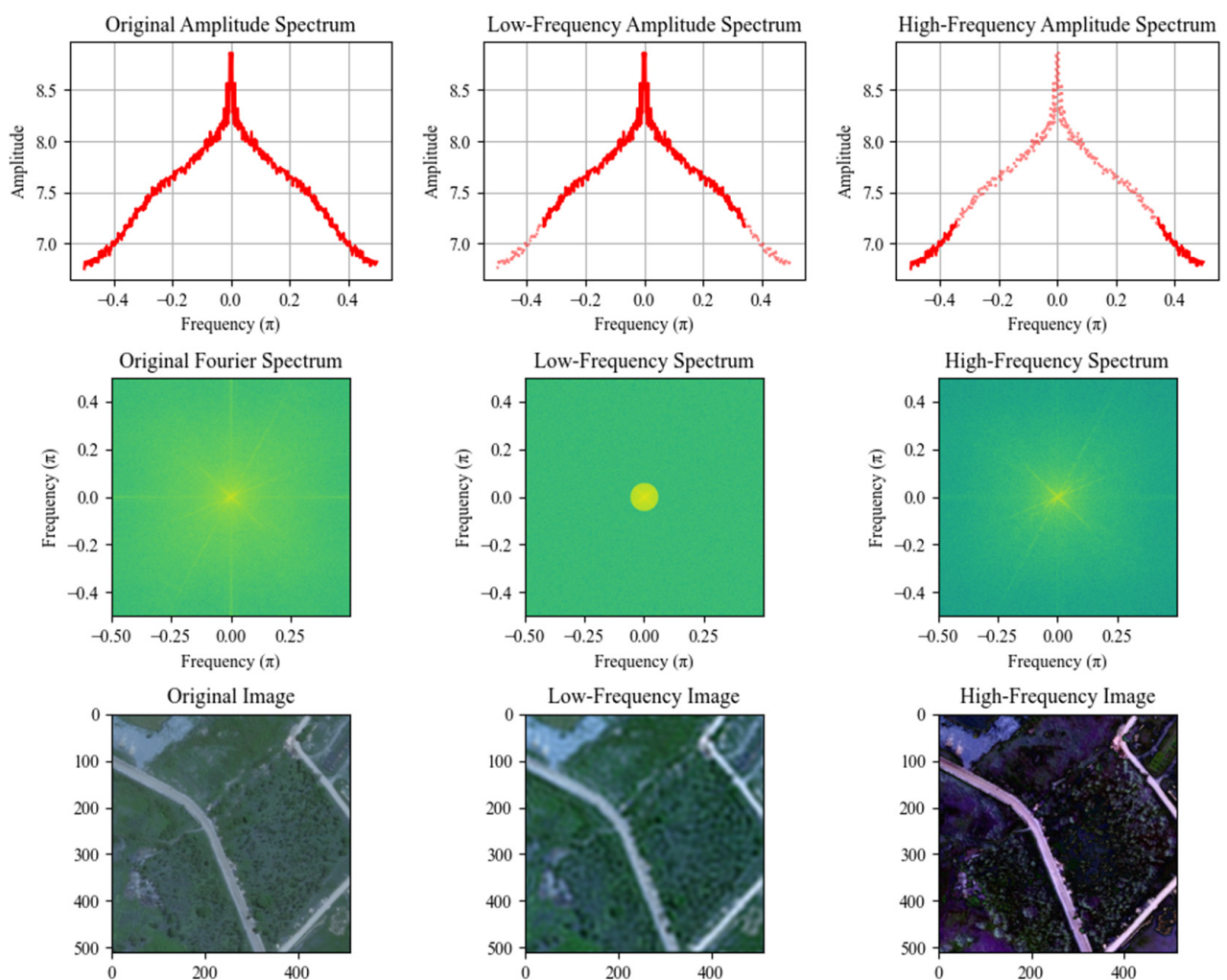


Figure 1. Decomposition of remote sensing images into low-frequency components and high-frequency components.

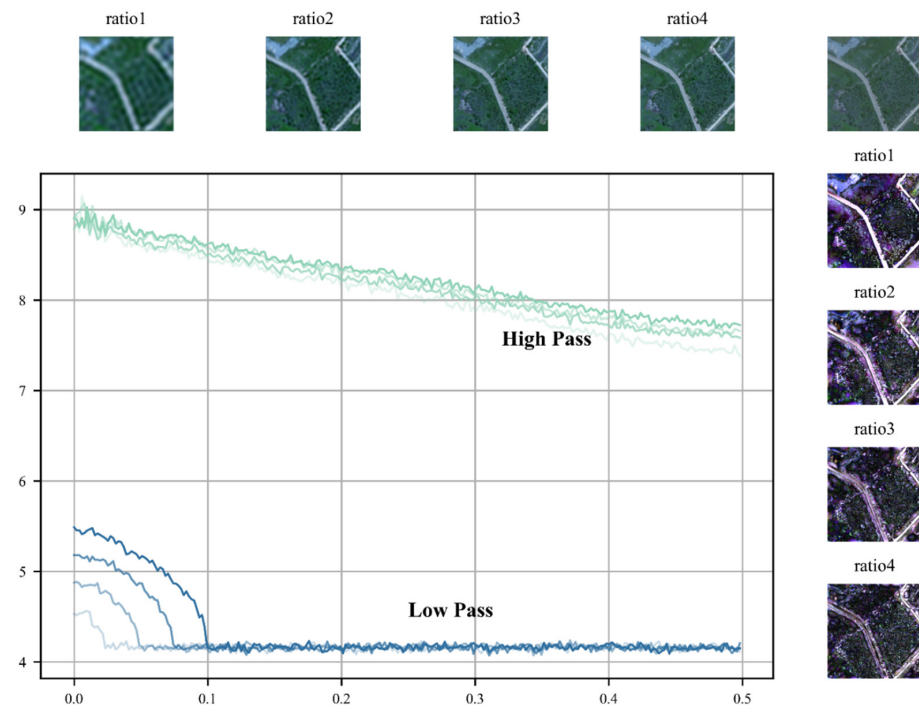


Figure 2. Frequency domain analysis of remote sensing images using different filter radius ratios.

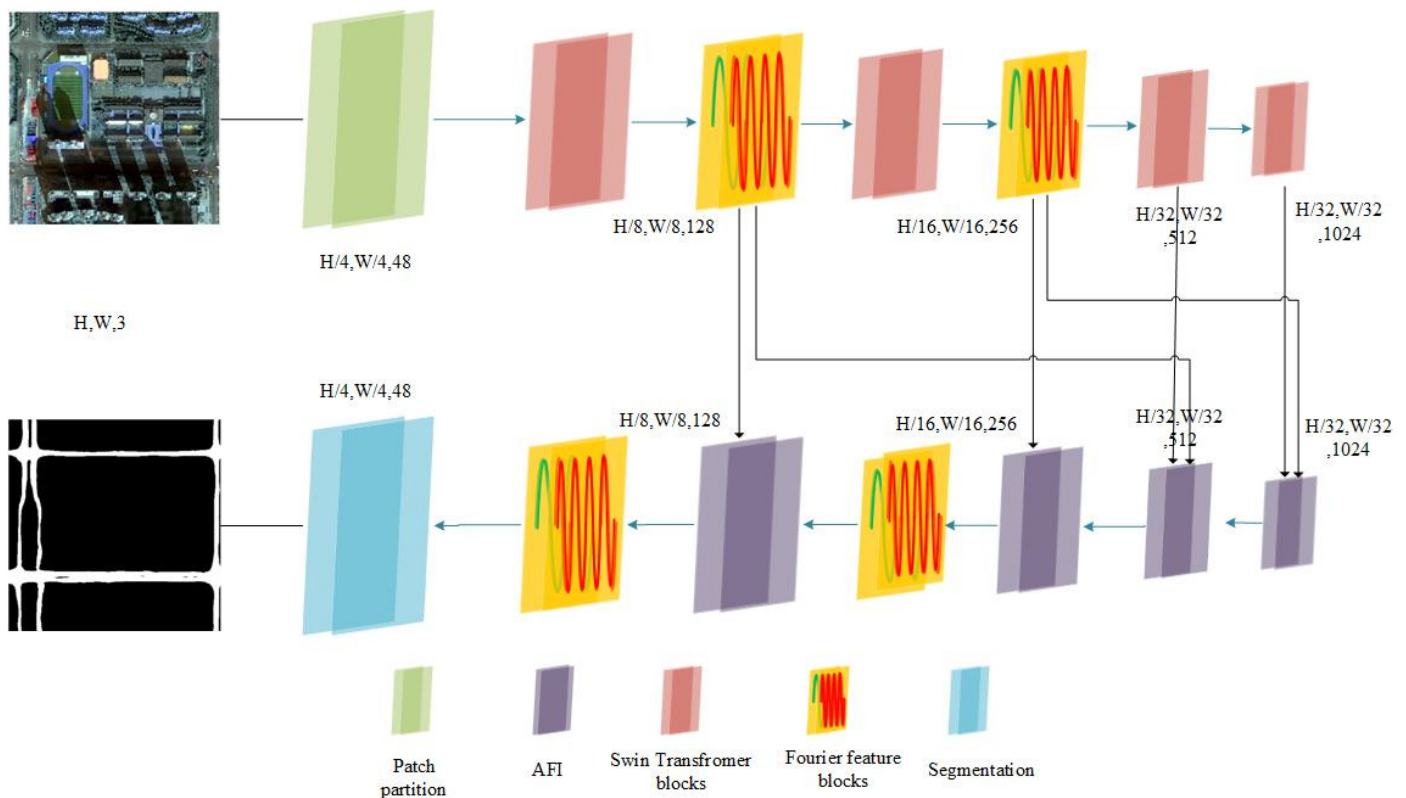


Figure 3. The structural composition of our model. Swin Transformer blocks provide spatial domain features, and Fourier feature blocks provide frequency domain features.

Decoding Stage: In the decoding stage, the lowest two layers' features are not processed by Fourier feature blocks. The decoding layers create AFI modules to integrate multi-level, multi-semantic road contextual information and further learn road frequency domain features in the Fourier space. Each decoder sub-network starts with upsampling,

using 2×2 transposed convolutions with a stride of 2, gradually restoring the image to its original size. Skip connections between the Fourier feature blocks of specific encoding layers allow for the fusion of low-level and high-level semantic road information, mitigating gradient vanishing issues and effectively capturing road features. Finally, segmentation is performed to complete the road image segmentation task.

This core methodology is meticulously designed for the extraction and analysis of road features. Every element of our model's architecture considers the specific characteristics of roads. Patch partitioning ensures the retention of fine- and large-scale road features. The Swin Transformer, with its robust global feature learning capability, is adept at capturing road-specific dependencies and contextual information. As each layer of the Swin Transformer progressively learns global road features, it reduces intra-class differences within roads. Fourier feature blocks focus on processing frequency components related to road textures and contours, filtering out frequency components of these intra-class differences, which is crucial for detailed road analysis. The dynamic filtering mechanism within the Fourier feature blocks effectively isolates and emphasizes frequency ranges representing road features, enhancing the overall accuracy and reliability of road image segmentation and analysis tasks.

2.3. Fourier Feature Blocks

We propose a Fourier feature blocks module (as shown in Figure 4) designed to effectively extract and process the frequency domain features of road information in images, thereby enhancing image processing and analysis performance. The design of this module is based on handling road features through the following distinct processing groups: frequency domain feature processing and spatial information feature processing.

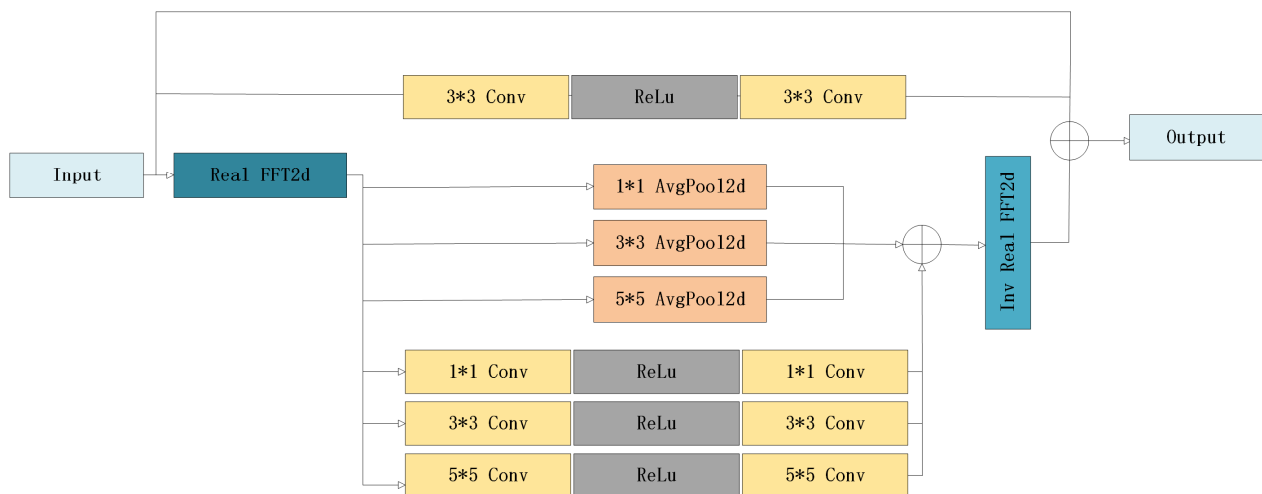


Figure 4. The structural composition of Fourier feature blocks. The upper part is the spatial information feature processing group; the middle part is a low-pass filter; and the lower part is the high-pass filter.

Frequency Domain Feature Processing Group: In this group, the input image information is initially transformed into the frequency domain using FFT. The transformed data are then divided into six sets as follows: smooth structural features are mapped to low-frequency signals, while detailed features such as contours and textures are converted into high-frequency signals. These frequency domain signals are processed using specially designed filters aimed at extracting or enhancing signals within specific frequency ranges.

After FFT, three sets of signals are passed through convolutional layers equipped with kernels of various sizes. These convolutional layers act as high-pass filters [58], specifically designed to remove unnecessary low-frequency information from the image while retaining the high-frequency components crucial for depicting global road boundaries. The strategy

of using convolutional kernels of different sizes simulates high-pass filters with varying cutoff frequencies, allowing for more precise retention of the high-frequency information needed from the image.

Conversely, the remaining three sets of signals are processed through low-pass filters with average pooling layers, where pooling layers of different window sizes simulate low-pass filters with varying cutoff frequencies. Average pooling helps retain the low-frequency components of the image, which is vital for preserving the internal structural features of different road types, suppressing high-frequency noise, and reducing intra-class variations in roads. After this series of low-pass and high-pass filtering, the processed results from each channel group are integrated to achieve cross-channel fusion of the frequency domain signals. These processed frequency domain signals are then reconstructed back to the spatial domain via inverse FFT, ensuring that the output feature maps maintain the same dimensions and shape as the original input image.

Spatial Information Feature Processing Group: To address the potential issue of local detail loss during frequency domain feature extraction, the spatial information feature processing group is designed. This group employs a strategy where the input signal and the result of convolution on the input signal are summed as a residual term. Finally, the spatial domain feature information is combined with the information obtained from the inverse Fourier transform. This strategy not only helps retain the global features of the image but also enhances the capture of local details, thus improving the internal consistency and smoothness of road features while significantly enhancing boundary clarity.

Through this designed module, we demonstrate how to combine global and local feature learning effectively when processing image features, ultimately enhancing the model's overall performance in feature extraction.

3. Experiments and Evaluation

3.1. Setting Loss Functions and Evaluation Metrics

In semantic segmentation tasks, Dice loss is commonly applied to enhance segmentation performance and effectively mitigate the influence of imbalanced classes. This loss function calculates the intersection ratio of predicted labels and ground truth labels, multiplied by 2 and divided by the sum of their element counts, ensuring a range between [0, 1].

$$DiceLoss = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

In our study, we employed multiple evaluation metrics, including mean Intersection over Union (mIoU), recall, precision, and F1 score, to assess model performance. OA represents the overall accuracy, that is, the proportion of the number of correct classifications of all samples to the total number of samples. The F1 score represents the harmonic mean of precision and recall, providing a comprehensive evaluation of both metrics. Meanwhile, mIoU reflects the degree of overlap between model predictions and actual road boundaries. To compare our proposed model with other popular models comprehensively, we conducted thorough calculations of these metrics in the experiments. The specific formulas for calculating these metrics are as follows:

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FP + FN} \quad (9)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (10)$$

True Positives (TPs) and True Negatives (TNs) represent the number of pixels correctly predicted as positive class and negative class, respectively. False Positives and False Negatives represent the number of non-road pixels incorrectly classified as positive class and road pixels incorrectly classified as negative class, respectively.

3.2. Experimental Configuration

In this study, PyTorch 1.13.1 was utilized as the deep learning framework, and development was conducted using the Python 3.7 programming language within the JetBrains PyCharm 2022 development platform. The experimental setup included a computer equipped with an Intel (R) Core (TM) i7-9700 CPU and an NVIDIA 1080Ti GPU. Adam was employed as the optimizer, with each dataset input being an image of size 512×512 , a batch size of 4, and a total of 50 training epochs. The initial learning rate was set to 1×10^{-3} , and the weight decay rate was 2.5×10^{-4} . Notably, during the model training process, the learning rate was automatically reduced as the number of epochs increased to accelerate the model's convergence.

3.3. Experimental Dataset

3.3.1. HF Dataset

We created a dataset called Hefei (HF) using GF-2 remote sensing satellite images from the Hefei area of China. The data resolution of the GF-2 satellite is 1 m. We appropriately cropped and grouped the data, and ultimately obtained 12,628 samples with a size of 512×512 pixels, as shown in Figure 5. Among them, we selected 108 images containing various land cover categories as the test dataset. The remaining samples were roughly divided into a training dataset and a validation dataset in a 4:1 ratio. The validation dataset contained 2504 images, while the training dataset contained 10,016 images.

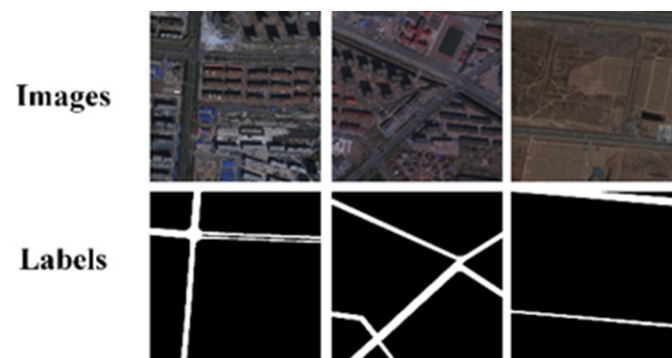


Figure 5. Example from the HF dataset.

3.3.2. Massachusetts Roads Dataset

The Massachusetts Roads dataset [59] covers an area of over 2600 km² in the state of Massachusetts, USA, including a variety of regions such as urban areas, towns, rural areas, and mountainous regions. The spatial resolution of the images is approximately 1 m. This dataset consists of 1171 pairs of 1500×1500 -pixel RGB aerial images, as shown in Figure 6, encompassing various road scene areas including urban, suburban, and rural environments. To prevent data loss and enhance the dataset, we applied flipping and rotation transformations to the original data. After cropping, a total of 11,710 images with a size of 512×512 pixels were generated. Following foreground filtering, 660 images without road foregrounds, along with their corresponding labels, were removed. Finally, a randomized selection process yielded 8684 images for the training set, 2170 for the validation set, and 196 for the test set.



Figure 6. Example from the Massachusetts dataset.

3.3.3. DeepGlobe Dataset

The DeepGlobe road dataset [60] covers various scenes in Thailand, India, and Indonesia, comprising 6226 pairs of 1024×1024 -pixel RGB satellite images and labels, with a resolution of 0.5 m per image. These images were collected by Digital Globe satellites. During data preprocessing, the samples were cropped into 24,904 images of 512×512 pixels, as shown in Figure 7. The test set included 308 images, while the training and validation sets were split at a ratio of 4:1, containing 19,676 and 4920 images, respectively.



Figure 7. Example from the DeepGlobe dataset.

3.4. Comparative Test of Various Road Datasets

To validate the effectiveness of our proposed method, we conducted both qualitative and quantitative performance comparisons of our model against U-Net [61], D-LinkNet [62], DeepLab-v3 [63], DCSwin [64], and the separable graph convolutional network (SGCN) [65] on the HF dataset, Massachusetts road dataset, and DeepGlobe road dataset. Specifically, DeepLab-v3 is a convolutional neural network-based semantic segmentation model that utilizes dilated convolution to expand the receptive field. D-LinkNet enhances road extraction capabilities in complex environments through a pretrained encoder and dilated convolution. U-Net is a fundamental and highly effective image segmentation network known for its strong segmentation capabilities. DCSwin employs the Swin Transformer as its backbone network and incorporates DCFAM to restore image resolution, producing precise segmentation maps. Meanwhile, SGCN is a hierarchical separable graph convolutional network specifically designed for road extraction from high-resolution remote sensing images in complex environments.

3.4.1. Experiment on HF Dataset

As shown in Figure 8, we compared the road extraction performance of our proposed model with other comparative models on the HF dataset. Overall, the extraction performance of our model is superior to other models, with fewer instances of fragmentation, discontinuity, and omission. While other models can extract the main contours of roads, they exhibit issues in detail. In Figure 8a, our model successfully extracts a complete and regular contour of a small road segment, while most of the other comparative models perform poorly, extracting only incomplete and fragmented road segments. In Figure 8c, because of occlusion by building shadows, the extraction performance of Unet is poor, and D-LinkNet, DeepLab-v3, and DCSwin also perform poorly in the presence of shadows, with only SGCN and our model able to extract roads completely. However, compared with our model, the overall performance of SGCN is significantly inferior. In Figure 8d, because of interference from building shadows, only our model can extract roads completely, while the other models exhibit noticeable discontinuities. In Figure 8f–i, our model can extract small roads more completely and smoothly, overcoming the influence of shadows on road extraction. As shown in the analysis of fragmented and omitted road extraction, our model can effectively capture local details, thereby improving the internal consistency and smoothness of road features and significantly enhancing the clarity of boundaries. The learning strategy in the Fourier domain effectively integrates global and local features, coordinates the internal consistency of road patches, and enhances road boundary information, thereby improving the issues of fragmentation and omission in extraction.

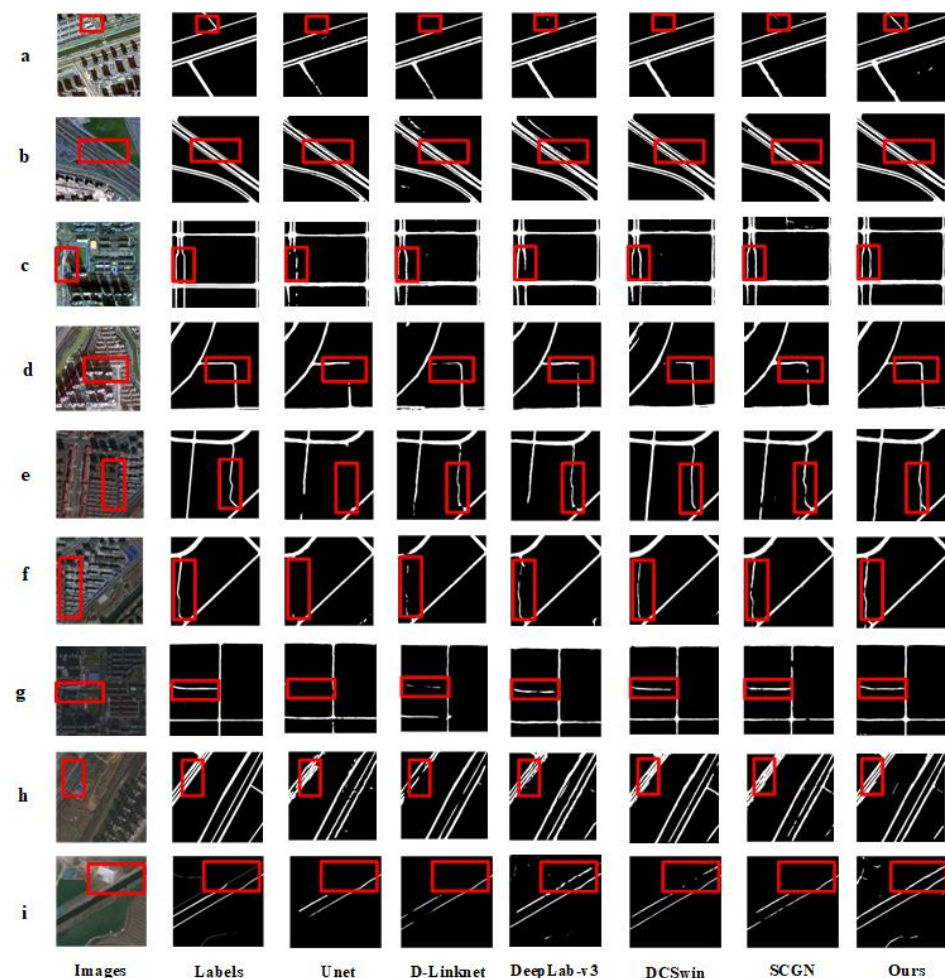


Figure 8. Comparative experimental results on the HF dataset (a–i). In each image, the red boxes highlight road details and the differences between the various models.

Table 1 quantitatively demonstrates the overall evaluation metrics on the test dataset. Compared with the other networks, our model achieves improvements in IOU of 8.57%, 5.98%, 4.69%, 2.33%, and 2.76%, respectively. The recall improvements are 6.51%, 3.66%, 7.20%, 2.97%, 1.50%, and 2.43%, respectively. The F1 score improvements are 6.63%, 4.39%, 3.66%, 1.95%, and 2.09%, respectively. The maximum improvements in average precision and OA are 4.65% and 1.20%, respectively. From both qualitative and quantitative perspectives, although DeepLab-v3, D-LinkNet, U-Net, DCSwin, and SGCN can extract the basic contours of roads, these models do not perform well for roads with significant shadow and noise interference. In contrast, our model produces smoother and more complete road extraction results. Additionally, compared with the other methods, our model also demonstrates improvements in multiple quantitative evaluation metrics, thereby validating its feasibility in road extraction tasks.

Table 1. Accuracy evaluation for the comparative experiments on the HF dataset.

Model	IOU	Recall	F1	Precision	OA
U-Net	63.97%	70.48%	75.59%	84.30%	96.83%
D-Linknet	66.56%	73.33%	77.83%	84.94%	97.11%
DeepLab-v3	67.85%	74.02%	78.56%	85.17%	97.35%
DCSwin	70.21%	75.49%	80.27%	87.87%	97.63%
SGCN	69.78%	74.56%	80.13%	84.94%	97.54%
ours	72.54%	76.99%	82.22%	89.59%	98.03%

3.4.2. Experiment on Massachusetts Roads Dataset

The comparison of extraction results between our model and the other models on the Massachusetts road dataset is shown in Figure 9. Overall, our model is more effective compared with the other models, where the results of the other comparative models appear to be inferior. Although they all extract the main contours of the roads, they lack consideration for the consistency of features along road edges and internally, resulting in road fragmentation and breakage. For example, in Figure 9a, a section of the road is obscured by tree shadows in the bottom right corner, leading to fragmented and broken road extraction by U-Net, D-LinkNet, DeepLab-v3, DCSwin, and SGCN. Only in our model, after filtering and considering feature consistency, can the obscured and shadowed roads be smoothed and made more consistent. In Figure 9c,d,f, the red box highlights a road segment with inconsistent brightness, indicating non-uniform lighting conditions and surface conditions influenced by adjacent trees, which significantly interferes with road extraction by various models. Models such as U-Net and D-LinkNet, compared here, can only extract this internally different road well after our model's smoothing and internal consistency learning strategies, reducing intra-class differences. In summary, on the Massachusetts road dataset, our model outperforms the compared models in road extraction. These models mainly learn road information from the spatial domain unilaterally, extracting from a relatively single perspective. In contrast, our proposed model combines spatial and frequency domains, providing multiple perspectives to extract road features, resulting in better road extraction performance and reducing fragmentation and omission in road extraction.

Table 2 quantitatively demonstrates the overall evaluation metrics on the test dataset. Relative to the other comparative networks, our model achieves IOU improvements of 10.42%, 10.21%, 6.01%, 2.81%, and 1.89%, respectively. The recall rates improve by 10.86%, 5.21%, 2.69%, 3.08%, and 4.83%, respectively. The F1 scores improve by 9.75%, 6.24%, 6.33%, 3.46%, and 1.85%, respectively. From both qualitative and quantitative perspectives, although DeepLab-v3, D-LinkNet, U-Net, DCSwin, and SGCN can also extract the basic contours of roads, they perform poorly when shadows or inconsistent internal characteristics of roads are present. In contrast, our model produces smoother and more complete road extraction results.

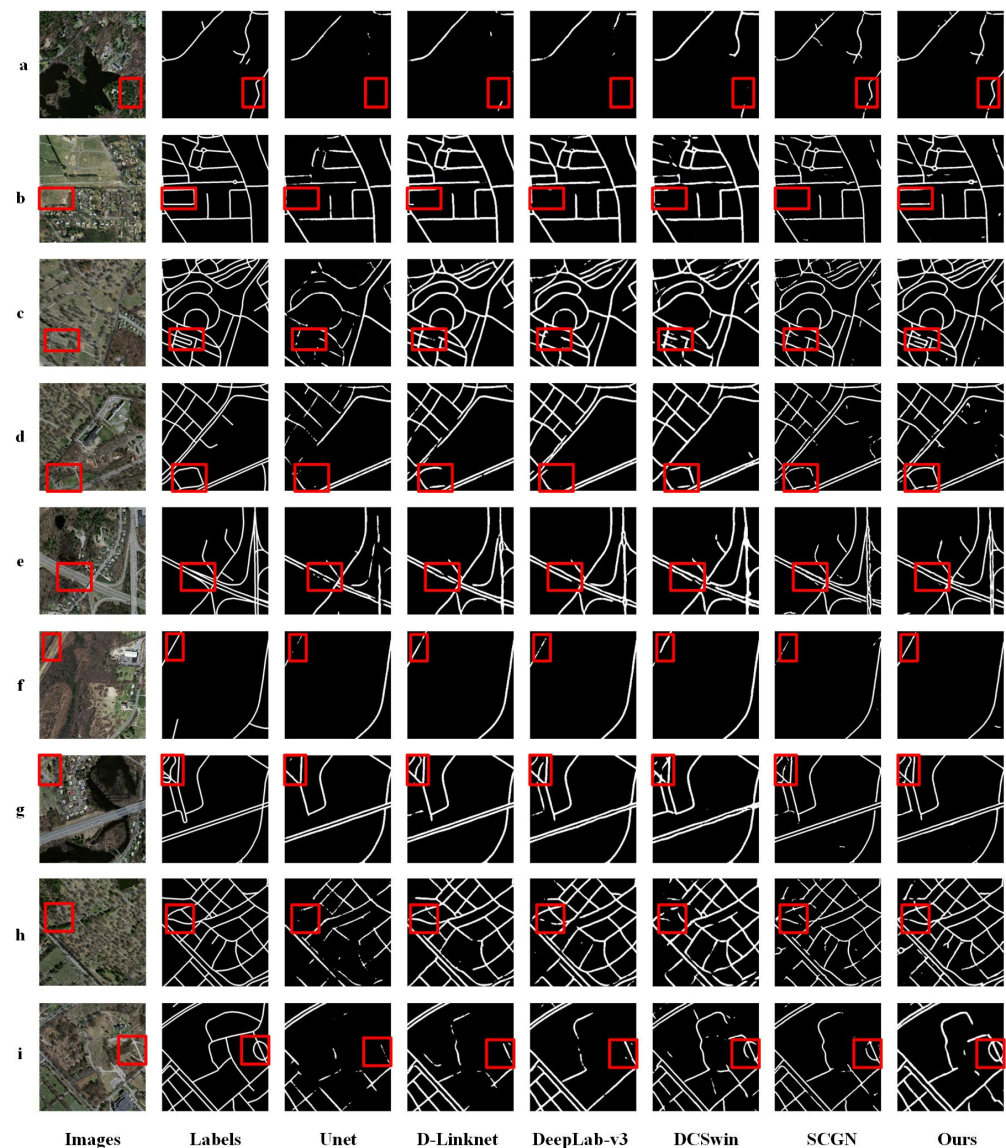


Figure 9. Comparative experimental results on the Massachusetts roads dataset (a–i). In each image, the red boxes highlight road details and the differences between various models.

Table 2. Accuracy evaluation for the comparative experiments on the Massachusetts roads dataset.

Model	IOU	Recall	F1	Precision	OA
U-Net	44.93%	55.77%	58.29%	65.76%	97.01%
D-Linknet	45.14%	61.42%	58.80%	60.13%	96.80%
DeepLabv3	48.34%	63.94%	61.71%	63.56%	97.08%
DCSwin	52.54%	63.55%	64.58%	69.71%	97.53%
SGCN	53.46%	61.80%	66.19%	73.77%	97.68%
ours	55.35%	66.63%	68.04%	71.83%	97.74%

3.4.3. Experiment on DeepGlobe Dataset

In general, the performance of our model on the DeepGlobe dataset is also significantly superior to the models we compared it against, as shown in Figure 10. In Figure 10b,c,f,i, some roads are covered by vegetation, making road extraction more challenging. As a result, the compared models struggle to extract roads covered by vegetation completely, leading to fragmented and broken road extraction. However, after Fourier transformation, our model considers global feature information in the frequency domain. Therefore, when

shadows and vegetation occlusion are present, it can extract road information from another feature perspective. In Figure 10g, because of various influences, the reflectance of the same road varies, with some sections appearing dark and others appearing bright. This leads to fragmented and broken road extraction by U-Net, D-LinkNet, DeepLab-v3, DCSwin, and SGCN. In contrast, our model, which considers the consistency of features along road edges and internally, can handle intra-class differences well and extract roads more excellently compared with the other models. Overall, on the DeepGlobe road dataset, our model can transform from the spatial domain to the frequency domain and then combine spatial and frequency domains to learn road information, providing more extraction perspectives. This results in better road extraction performance, reducing fragmentation and omissions in road extraction.

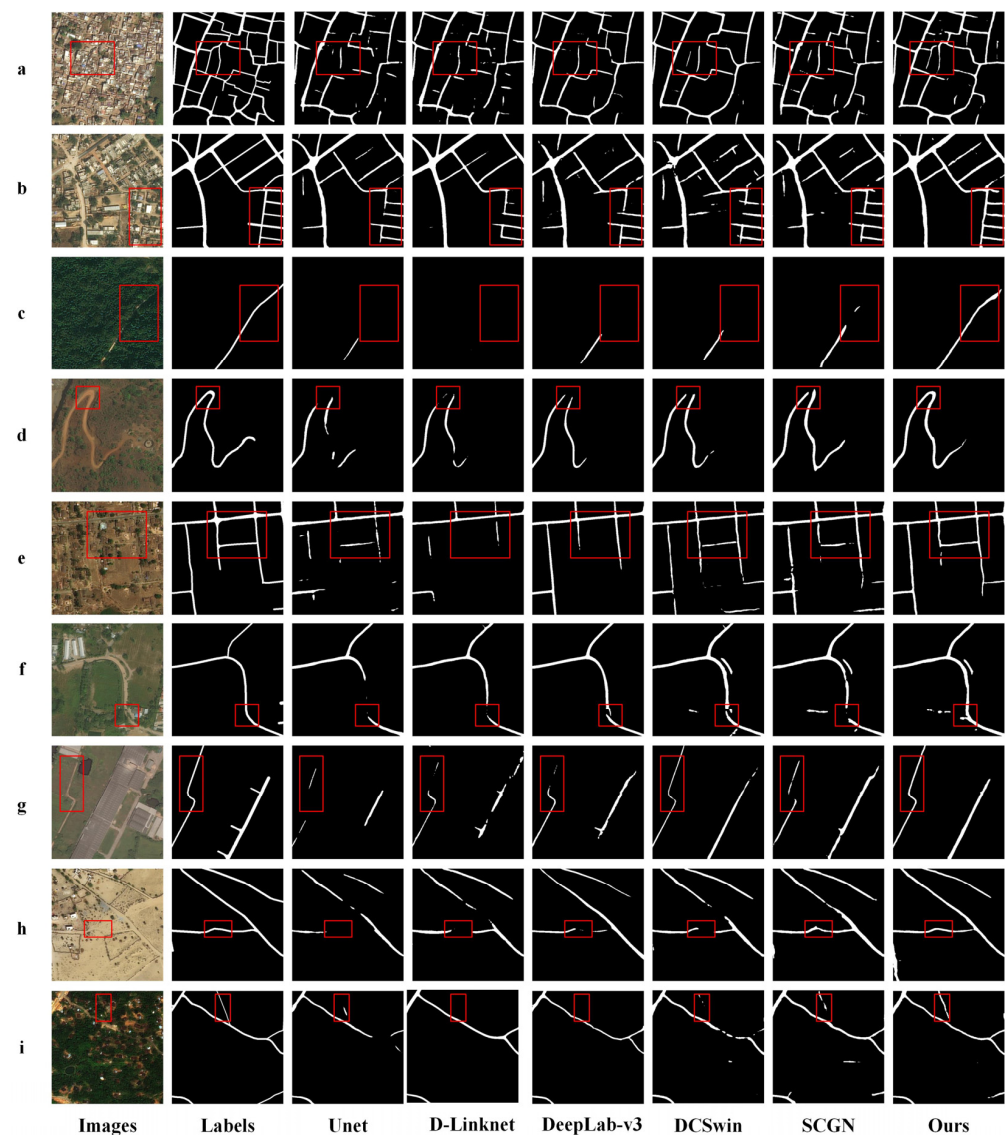


Figure 10. Comparative experimental results on the DeepGlobe roads dataset (a–i). In each image, the red boxes highlight road details and the differences between various models.

Table 3 compares the performance of the different models on the image segmentation task. Based on the provided metrics, our model demonstrates superior performance in IOU, recall, F1 score, AP, and OA. Specifically, our model achieves an IOU of 71.87%, with a maximum improvement of 4.96%, showing higher accuracy compared with the other models. Meanwhile, in terms of recall, our model also achieves 77.49%, slightly higher than other models. Regarding the F1 score, a comprehensive evaluation metric,

our model leads with a score of 82.67%, further proving its excellent balance between precision and recall. In terms of AP, our model reaches 89.95%, demonstrating higher average segmentation accuracy compared with the other models such as U-Net, D-LinkNet, DeepLab-v3, DCSwin, and SGCN. Finally, in terms of OA, a metric measuring the overall performance of the model, our model ranks first with a score of 98.58%, surpassing all the other models. In summary, our model demonstrates comprehensive and superior performance in the image segmentation task, achieving leading results in both individual metrics and overall performance.

Table 3. Accuracy evaluation for the comparative experiments on the DeepGlobe dataset.

Model	IOU	Recall	F1	Precision	OA
U-Net	60.79%	73.05%	75.96%	79.54%	97.83%
D-Linknet	64.36%	73.42%	77.65%	85.47%	97.88%
DeepLab-v3	65.51%	75.23%	79.30%	85.36%	97.96%
DCSwin	67.83%	75.85%	80.43%	88.17%	98.35%
SGCN	66.01%	74.89%	79.82%	84.75%	98.10%
ours	71.87%	77.49%	82.67%	89.95%	98.58%

4. Discussion

To further investigate the effectiveness of Fourier frequency filters, this paper conducted the following ablation experiments on the HF dataset: one where our model had its Fourier feature blocks removed, and another where Fourier blocks were added to every layer during downsampling and upsampling. In this context, our model with Fourier feature blocks in every layer is referred to as Full Fourier.

Figure 11 compares the prediction results of our model, Full Fourier, No Fourier, and road labels. Overall, both our model and Full Fourier perform better in extracting road features than No Fourier. However, Full Fourier is slightly less accurate than our model. A possible reason for this is that during the deepest feature sampling, certain road frequency components might be erroneously filtered out in the frequency domain, leading to less effective road extraction in some areas compared with our model. Our model, by applying Fourier feature blocks only in the first two layers of the network, balances the need to learn global and multi-perspective frequency domain features while avoiding excessive filtering of road frequency components in the deeper layers, resulting in a higher degree of road extraction completeness. For example, in Figure 11a,b,g, shadows from buildings, bridges, and other structures obscure parts of the roads. Our model, leveraging global road features, restores the road portions obscured by shadows, thereby enhancing the connectivity of road extraction. In contrast, without the Fourier modules, accurately recovering and extracting these shadowed areas is challenging. In Figure 11c,d, changes in road reflectivity due to factors such as road moisture and large objects may cause some road segments to appear darker while others are brighter in the images. In such cases, road extraction without the Fourier modules may result in fragmented and discontinuous sections. Our model, through Fourier processing, comprehensively considers both the edge details and internal consistency of the roads, thereby achieving more complete internal extraction and smoother edges and significantly improving the accuracy of road extraction.

These experiments clearly demonstrate the crucial role and effectiveness of Fourier frequency domain filtering blocks in enhancing the accuracy of road image segmentation.

Then, we quantitatively evaluated the effectiveness of Fourier feature blocks and their impact on different layers in our model. On the HF dataset, adding Fourier feature blocks only to the first two layers of the network resulted in the following improvements: IOU increased by 2.25%, the recall rate increased by 1.03%, the F1 score increased by 2.43%, the overall accuracy increased by 2.06%, and OA increased by 0.75% in Table 4. However, adding Fourier blocks to every layer of our model resulted in a decrease in accuracy compared with our model with Fourier feature blocks only in the first two layers. This indicates that incorporating Fourier modules provides a novel perspective for learning

road features through frequency domain learning, enhancing the network's perception of global road structures and improving road boundary features and internal consistency, thereby obtaining richer semantic information. Nonetheless, excessive filtering of frequency domain components does not yield better results; filtering is most effective when applied to shallow layers.

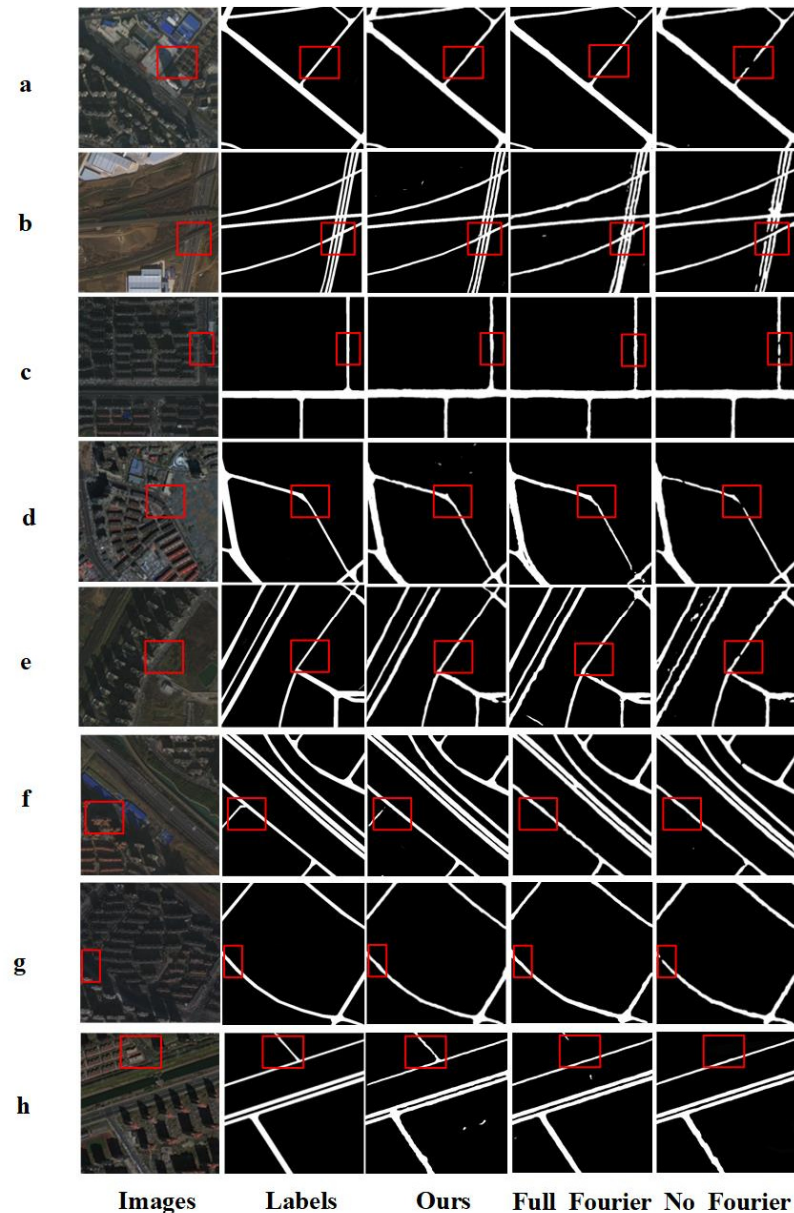


Figure 11. Results of ablation experiments on the HF dataset (a–h). In each image, the red boxes highlight road details and the differences observed in the ablation study.

Table 4. Evaluation of ablation experiment accuracy on the HF dataset.

Model	IOU	Recall	F1	Precision	OA
No Fourier	70.29%	75.96%	79.85%	87.52%	97.46%
Full Fourier	72.05%	76.11%	81.55%	86.71%	97.91%
Ours	72.54%	76.99%	82.22%	89.59%	98.03%

Finally, to further demonstrate the effectiveness of Fourier feature blocks, the feature maps of our model at different network structure layers were visualized, along with the

corresponding levels of our model without Fourier blocks, as shown in Figure 12. (Because of the extensive downsampling in the deeper layers of the network, the feature maps of the last two downsampling layers and the first two upsampling layers are not displayed.) It can be clearly observed that during the downsampling and upsampling process, under the influence of surrounding background noise, the road features in the feature maps without Fourier transformation are significantly less distinct, with cluttered spatial information that fails to distinguish road features effectively. In contrast, the road features extracted by the Fourier feature blocks are more prominent, with smoother road surfaces and more textured road edges.

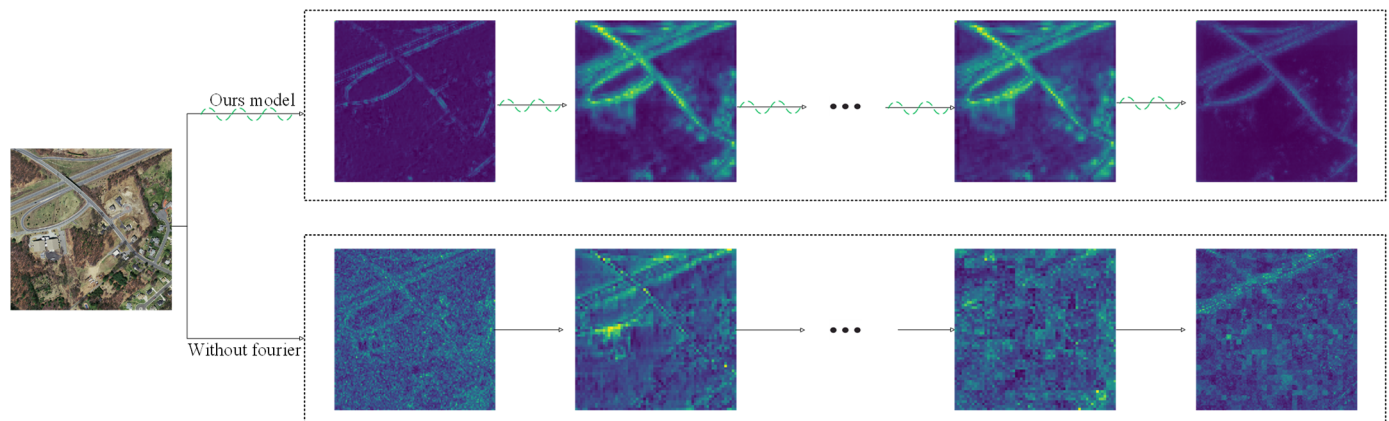


Figure 12. Feature maps for different structure layers of the network. The upper part is the feature map output by our model; the lower part is the feature map output after removing Fourier feature blocks.

5. Conclusions

In order to enhance the ability of deep learning models to express road features in complex remote sensing images, this article proposes a road extraction method that couples global spatial learning with Fourier frequency domain learning. This method first utilizes FFT technology combined with adaptive dynamic filters to transform global road feature maps into the frequency spectrum domain. In the frequency domain, the network learns road features in a novel manner, specifically through frequency domain information learning that captures critical low-frequency information and essential high-frequency details. By employing low-pass and high-pass filters with varying kernel sizes, the method effectively separates low-frequency components from global road information and high-frequency components from road edge information, allowing for dynamic adjustment of these features. By enhancing or attenuating frequency components, the method improves the high-frequency information of road edges while reducing the intra-class variation in low-frequency information within the road. After performing an inverse Fourier transform back to the spatial domain and combining it with spatial domain features, the dual features from the frequency and spatial domains are obtained, enhancing global road information and reducing intra-class differences, thereby improving model performance. Experimental results on the HF, MS, and DeepGlobe road datasets show that our model achieves higher boundary accuracy compared with other deep learning models (such as Unet, D-Linknet, DeepLab-v3, DCSwin, and SGCN), where IOU reaches 72.54%, 55.35%, and 71.87%, respectively. These results not only validate the effectiveness of our proposed global spatial and Fourier frequency domain coupled learning strategy but also demonstrate the significant advantages of this approach in addressing complex road extraction tasks. In particular, the model effectively reduces fragmentation, especially when roads are affected by shadow interference or internal variations. Additionally, the ablation experiments further confirm the critical role of the frequency domain learning module in enhancing the model's performance, particularly in strengthening road edge information and reducing intra-class variance. The impact of the Fourier feature blocks on road extraction across different layers

of the network is also discussed. This study showcases the potential of incorporating frequency domain information to improve road feature representation and indicates that this method holds promising prospects for future application in other remote sensing target extraction tasks. Although the proposed method has shown strong performance across multiple datasets, there is still room for improvement to enhance the model's generalization capability. Future research will focus on exploring more types of feature fusion strategies, particularly beyond the spatial and frequency domains, to further minimize information loss and improve the model's generalization ability and robustness.

Author Contributions: Conceptualization H.Y.; data curation. Y.W. (Yongchuang Wu) and C.Z.; funding acquisition, H.Y. and X.X.; investigation, C.Z., Y.W. (Yongchuang Wu) and H.Y.; methodology, C.Z.; resources, H.Y.; validation, X.X., Y.W. (Yanlan Wu) and Y.W. (Yongchuang Wu); visualization, Y.W. (Yanlan Wu); writing—original draft, C.Z.; writing—review and editing, C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant No. 42101381) (Corresponding author: Yanlan Wu).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available because of permissions issues.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-Of-The-Art Review. *Remote Sens.* **2020**, *12*, 1444. [\[CrossRef\]](#)
2. Sussi, E.; Husni, R.; Yusuf, A.; Harto, D.B.; Suwardhi, D.; Siburian, A. Utilization of Improved Annotations from Object-Based Image Analysis as Training Data for DeepLab V3+ Model: A Focus on Road Extraction in Very High-Resolution Orthophotos. *IEEE Access* **2024**, *12*, 67910–67923. [\[CrossRef\]](#)
3. Montenegro, A.L.; Rey-Gozalo, G.; Arenas, J.P.; Suárez, E. Streets Classification Models by Urban Features for Road Traffic Noise Estimation. *Sci. Total Environ.* **2024**, *932*, 173005. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Tao, Y.; Tian, L.; Wang, C.; Dai, W.; Xu, Y. A fine construction method of urban road DEM considering road morphological characteristics. *Sci. Rep.* **2022**, *12*, 14958. [\[CrossRef\]](#)
5. Xu, Y.; Chen, H.; Du, C.; Li, J. MSACon: Mining spatial attention-based contextual information for road extraction. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5604317. [\[CrossRef\]](#)
6. Wang, Y.; Seo, J.; Jeon, T. NL-LinkNet: Toward lighter but more accurate road extraction with nonlocal operations. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3000105. [\[CrossRef\]](#)
7. Lin, Y.; Saripalli, S. Road Detection and Tracking from Aerial Desert Imagery. *J. Intell. Robot. Syst.* **2012**, *65*, 345–359. [\[CrossRef\]](#)
8. Liu, W.; Zhang, L.; Li, L.; Chen, Y. Dictionary Learning-Based Hough Transform for Road Detection in Multispectral Image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2330–2334. [\[CrossRef\]](#)
9. Yang, J.; He, Y.; Caspersen, J. Region merging using local spectral angle thresholds: A more accurate method for hybrid segmentation of remote sensing images. *Remote Sens. Environ.* **2017**, *190*, 137–148. [\[CrossRef\]](#)
10. Courtrai, L.; Lefèvre, S. Morphological path filtering at the region scale for efficient and robust road network extraction from satellite imagery. *Pattern Recognit. Lett.* **2016**, *83*, 195–204. [\[CrossRef\]](#)
11. Yeom, J.; Kim, Y. A Regular Grid-Based Hough Transform for the Extraction of Urban Features Using High-Resolution Satellite Images. *Remote Sens. Lett.* **2015**, *6*, 409–417. [\[CrossRef\]](#)
12. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015, Boston, MA, USA, 7–12 June 2015; IEEE: New York, NY, USA, 2015; Volume 2015, pp. 3431–3440. [\[CrossRef\]](#)
13. Kestur, R.; Farooq, S.; Abdal, R.; Mehraj, E.; Narasipura, O.; Mudigere, M. UFCN: A fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle. *J. Appl. Remote Sens.* **2018**, *12*, 016020. [\[CrossRef\]](#)
14. Zhang, Y.; Xia, G.; Wang, J.; Lha, D. A Multiple Feature Fully Convolutional Network for Road Extraction From High-Resolution Remote Sensing Image Over Mountainous Areas. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1600–1604. [\[CrossRef\]](#)
15. Yang, X.; Li, X.; Ye, Y.; Lau, R.Y.K.; Zhang, X.; Huang, X. Road Detection and Centerline Extraction via Deep Recurrent Convolutional Neural Network U-Net. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7209–7220. [\[CrossRef\]](#)
16. Çalışkan, E.; Sevim, Y. Forest Road Extraction from Orthophoto Images by Convolutional Neural Networks. *Geocarto Int.* **2022**, *37*, 11671–11685. [\[CrossRef\]](#)
17. Eerapu, K.K.; Ashwath, B.; Lal, S.; Dell'Acqua, F.; Narasimha Dhan, A.V. Dense Refinement Residual Network for Road Extraction From Aerial Imagery Data. *IEEE Access* **2019**, *7*, 151764–151782. [\[CrossRef\]](#)

18. Das, S.; Fime, A.A.; Siddique, N.; Hashem, M.M.A. Estimation of Road Boundary for Intelligent Vehicles Based on DeepLabV3+ Architecture. *IEEE Access* **2021**, *9*, 121060–121075. [\[CrossRef\]](#)
19. Wang, R.; Cai, M.; Xia, Z. A Lightweight High-Resolution RS Image Road Extraction Method Combining Multi-Scale and Attention Mechanism. *IEEE Access* **2023**, *11*, 108956–108966. [\[CrossRef\]](#)
20. Hou, Y.; Liu, Z.; Zhang, T.; Li, Y. C-UNet: Complement UNet for Remote Sensing Road Extraction. *Sensors* **2021**, *21*, 2153. [\[CrossRef\]](#)
21. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [\[CrossRef\]](#)
22. Chen, Z.; Fan, W.; Zhong, B.; Li, J.; Du, J.; Wang, C. Coarse-to-Fine Road Extraction Based on Local Dirichlet Mixture Models and Multiscale-High-Order Deep Learning. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4283–4293. [\[CrossRef\]](#)
23. Yang, X.; Li, X.; Ye, Y.; Zhang, X.; Zhang, H.; Huang, X.; Zhang, B. Road detection via deep residual dense U-Net. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–7. [\[CrossRef\]](#)
24. Lin, Y.; Xu, D.; Wang, N.; Shi, Z.; Chen, Q. Road Extraction from Very-High-Resolution Remote Sensing Images via a Nested SE-Deeplab Model. *Remote Sens.* **2020**, *12*, 2985. [\[CrossRef\]](#)
25. Guan, H.; Yu, Y.; Li, D.; Wang, H. RoadCapsFPN: Capsule Feature Pyramid Network for road extraction from VHR optical remote sensing imagery. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11041–11051. [\[CrossRef\]](#)
26. Chen, S.; Zhang, Z.; Zhong, R.; Zhang, L.; Ma, H.; Liu, L. A dense feature pyramid network-based deep learning model for road marking instance segmentation using MLS point clouds. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 784–800. [\[CrossRef\]](#)
27. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
28. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on Vision Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 87–110. [\[CrossRef\]](#)
29. Islam, K. Recent advances in Vision Transformer: A survey and outlook of recent work. *arXiv* **2022**, arXiv:2203.01536. [\[CrossRef\]](#)
30. Zhu, X.; Huang, X.; Cao, W.; Yang, X.; Zhou, Y.; Wang, S. Road extraction from remote sensing imagery with spatial attention based on Swin Transformer. *Remote Sens.* **2024**, *16*, 1183. [\[CrossRef\]](#)
31. Han, Y.; Liu, Q.; Liu, H.; Hu, X.; Wang, B. PT-RE: Prompt-based multi-modal transformer for road network extraction from remote sensing images. *IEEE Sens. J.* **2024**. [\[CrossRef\]](#)
32. Liu, W.; Gao, S.; Zhang, C.; Yang, B. RoadCT: A hybrid CNN-transformer network for road extraction from satellite imagery. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 2501805. [\[CrossRef\]](#)
33. Wang, X.; Cai, Y.; He, K.; Wang, S.; Liu, Y.; Dong, Y. Global-local information fusion network for road extraction: Bridging the gap in accurate road segmentation in China. *Remote Sens.* **2023**, *15*, 4686. [\[CrossRef\]](#)
34. Kumar, K.M.; Velayudham, A. CCT-DOSA: A Hybrid Architecture for Road Network Extraction From Satellite Images in the Era of IoT. *Evol. Syst.* **2024**, *15*, 1939–1955. [\[CrossRef\]](#)
35. Kumar, K.M. RoadTransNet: Advancing remote sensing road extraction through multi-scale features and contextual information. *Signal Image Video Process.* **2024**, *18*, 2403–2412. [\[CrossRef\]](#)
36. Wei, D.; Li, P.; Xie, H.; Xu, Y. DRCNet: Road Extraction From Remote Sensing Images Using DenseNet With Recurrent Criss-Cross Attention and Convolutional Block Attention Module. *IEEE Access* **2023**, *11*, 126879–126891. [\[CrossRef\]](#)
37. Akhtarmanesh, A.; Abbasi-Moghadam, D.; Sharifi, A.; Yadkouri, M.H.; Tariq, A.; Lu, L. Road Extraction from Satellite Images Using Attention-Assisted UNet. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1126–1136. [\[CrossRef\]](#)
38. Jamali, A.; Roy, S.K.; Li, J.; Ghamisi, P. Neighborhood Attention Makes the Encoder of ResUNet Stronger for Accurate Road Extraction. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6003005. [\[CrossRef\]](#)
39. Sundarapandi, A.M.S.; Alotaibi, Y.; Thanarajan, T.; Rajendran, S. Archimedes Optimisation Algorithm Quantum Dilated Convolutional Neural Network for Road Extraction in Remote Sensing Images. *Heliyon* **2024**, *10*, e26589. [\[CrossRef\]](#)
40. Toni, Y.; Meena, U.; Mishra, V.K.; Garg, R.D.; Sharma, K.P. AM-UNet: Road Network Extraction from High-Resolution Aerial Imagery Using Attention-Based Convolutional Neural Network. *J. Indian Soc. Remote Sens.* **2024**. [\[CrossRef\]](#)
41. Mehmood, M.; Shahzad, A.; Zafar, B.; Shabbir, A.; Ali, N. Remote sensing image classification: A comprehensive review and applications. *Math. Probl. Eng.* **2022**, *2022*, 5880959. [\[CrossRef\]](#)
42. Wang, D.; Gao, L.; Qu, Y.; Sun, X.; Liao, W. Frequency-to-Spectrum Mapping GAN for Semisupervised Hyperspectral Anomaly Detection. *CAAI Trans. Intell. Technol.* **2023**, *8*, 1258–1273. [\[CrossRef\]](#)
43. Gao, L.; Wang, D.; Zhuang, L.; Sun, X.; Huang, M.; Plaza, A. BS3LNet: A New Blind-Spot Self-Supervised Learning Network for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5504218. [\[CrossRef\]](#)
44. Wang, J.; Guo, S.; Hua, Z.; Huang, R.; Hu, J.; Gong, M. CL-CaGAN: Capsule Differential Adversarial Continual Learning for Cross-Domain Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5517315. [\[CrossRef\]](#)
45. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Huang, M.; Plaza, A. BockNet: Blind-Block Reconstruction Network with a Guard Window for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5531916. [\[CrossRef\]](#)
46. Wang, Z.; Zhao, Y.; Chen, J. Multi-Scale Fast Fourier Transform Based Attention Network for Remote-Sensing Image Super-Resolution. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2023**, *16*, 2728–2740. [\[CrossRef\]](#)
47. Song, B.; Min, S.; Yang, H.; Wu, Y.; Wang, B. A Fourier Frequency Domain Convolutional Neural Network for Remote Sensing Crop Classification Considering Global Consistency and Edge Specificity. *Remote Sens.* **2023**, *15*, 4788. [\[CrossRef\]](#)

48. Yu, L.; Xie, J.; Zheng, X. The Relationship Between Graph Fourier Transform (GFT) and Discrete Cosine Transform (DCT) for 1D Signal and 2D Image. *SIViP* **2023**, *17*, 445–451. [\[CrossRef\]](#)
49. Wang, S.; Cheng, H.; Ying, L.; Xiao, T.; Ke, Z.; Zheng, H.; Liang, D. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magn. Reson. Imag.* **2020**, *68*, 136–147. [\[CrossRef\]](#)
50. Xi, J.; Ersoy, O.K.; Cong, M.; Zhao, C.; Qu, W.; Wu, T. Wide and Deep Fourier Neural Network for Hyperspectral Remote Sensing Image Classification. *Remote Sens.* **2022**, *14*, 2931. [\[CrossRef\]](#)
51. Yao, Z.; Fan, G.; Fan, J.; Gan, M.; Chen, C.L.P. Spatial–Frequency Dual-Domain Feature Fusion Network for Low-Light Remote Sensing Image Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4706516. [\[CrossRef\]](#)
52. Yu, B.; Yang, A.; Chen, F.; Wang, N.; Wang, L. SNNFD, spiking neural segmentation network in frequency domain using high spatial resolution images for building extraction. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102930. [\[CrossRef\]](#)
53. Ricaud, B.; Borgnat, P.; Tremblay, N.; Gonçalves, P.; Vanderghelynst, P. Fourier could be a data scientist: From graph Fourier transform to signal processing on graphs. *C. R. Phys.* **2019**, *20*, 474–488. [\[CrossRef\]](#)
54. Hu, Y.; Lu, L.; Li, C. Memory-accelerated parallel method for multidimensional fast fourier implementation on GPU. *J. Supercomput.* **2022**, *78*, 18189–18208. [\[CrossRef\]](#)
55. Singh, A.; Chougule, A.; Narang, P.; Chamola, V.; Yu, F.R. Low-Light Image Enhancement for UAVs with Multi-Feature Fusion Deep Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3513305. [\[CrossRef\]](#)
56. Chen, H.; Yokoya, N.; Chini, M. Fourier domain structural relationship analysis for unsupervised multimodal change detection. *ISPRS J. Photogramm. Remote Sens.* **2023**, *198*, 99–114. [\[CrossRef\]](#)
57. Zhu, P.; Zhang, X.; Han, X.; Cheng, X.; Gu, J.; Chen, P.; Jiao, L. Cross-Domain Classification Based on Frequency Component Adaptation for Remote Sensing Images. *Remote Sens.* **2024**, *16*, 2134. [\[CrossRef\]](#)
58. Wang, W.; Wang, J.; Chen, C.; Jiao, J.; Cai, Y.; Song, S.; Li, J. Fremae: Fourier transform meets masked autoencoders for medical image segmentation. *arXiv* **2023**, arXiv:2304.10864. [\[CrossRef\]](#)
59. Mnih, V. Machine Learning for Aerial Image Labeling. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada, 2013.
60. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; IEEE: New York, NY, USA, 2018; pp. 172–17209. [\[CrossRef\]](#)
61. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
62. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High-Resolution Satellite Imagery Road Extraction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; IEEE: New York, NY, USA, 2018; pp. 192–1924. [\[CrossRef\]](#)
63. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* **2018**, arXiv:1802.02611. [\[CrossRef\]](#)
64. Wang, L.; Li, R.; Duan, C.; Zhang, C.; Meng, X.; Fang, S. A Novel Transformer-Based Semantic Segmentation Scheme for Fine-Resolution Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6506105. [\[CrossRef\]](#)
65. Zhou, G.; Chen, W.; Gui, Q.; Li, X.; Wang, L. Split Depth-Wise Separable Graph-Convolution Network for Road Extraction in Complex Environments from High-Resolution Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5614115. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.