remote sensing

MDPI

Article

LCMorph: Exploiting Frequency Cues and Morphological Perception for Low-Contrast Road Extraction in Remote Sensing Images

Xin Li [1,2,*], Shumin Yang [1,2], Fan Meng [3], Wenlong Li [1,2], Zongchi Yang [1,2] and Ruoyu Wei [1,2]

[1] Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China; z22070021@s.upc.edu.cn (S.Y.); z23070032@s.upc.edu.cn (W.L.); z23070091@s.upc.edu.cn (Z.Y.); z24070092@s.upc.edu.cn (R.W.)
[2] Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, China University of Petroleum (East China), Qingdao 266580, China
[3] Institute of Future Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China; meng@nuist.edu.cn
* Correspondence: lix@upc.edu.cn

Abstract: Road extraction in remote sensing images is crucial for urban planning, traffic navigation, and mapping. However, certain lighting conditions and compositional materials often cause roads to exhibit colors and textures similar to the background, leading to incomplete extraction. Additionally, the elongated and curved road morphology conflicts with the rectangular receptive field of traditional convolution. These challenges significantly affect the accuracy of road extraction in remote sensing images. To address these issues, we propose an end-to-end low-contrast road extraction network called LCMorph, which leverages both frequency cues and morphological perception. First, Frequency-Aware Modules (FAMs) are introduced in the encoder to extract frequency cues, effectively distinguishing low-contrast roads from the background. Subsequently, Morphological Perception Blocks (MPBlocks) are employed in the decoder to adaptively adjust the receptive field to the elongated and curved nature of roads. MPBlock relies on snake convolution, which mimics snakes' twisting behavior for accurate road extraction. Experiments demonstrate that our method achieves state-of-the-art performance in terms of F1 score and IoU on the self-constructed low-contrast road dataset (LC-Roads), as well as the public DeepGlobe and Massachusetts Roads datasets.

Keywords: remote sensing; road extraction; low-contrast roads; frequency cues; morphological perception

Academic Editor: Andrea Garzelli

Received: 5 October 2024
Revised: 20 December 2024
Accepted: 7 January 2025
Published: 13 January 2025

Citation: Li, X.; Yang, S.; Meng, F.; Li, W.; Yang, Z.; Wei, R. LCMorph: Exploiting Frequency Cues and Morphological Perception for Low-Contrast Road Extraction in Remote Sensing Images. *Remote Sens.* **2025**, *17*, 257. https://doi.org/10.3390/rs17020257

1. Introduction

The resolution of remote sensing images is continuously advancing, providing several advantages, such as extensive coverage, frequent updates, easy access, and detailed information. Therefore, remote sensing images have been widely used in extracting geographic information. As a fundamental source of geographic information, roads are essential for urban planning [1], traffic management [2], route navigation [3], mapping [4], and other applications. Consequently, road extraction from remote sensing images has become a widely researched topic. Based on various implementation techniques, we categorize the main road extraction methods into traditional methods and deep learning-based methods.

Traditional road extraction methods primarily rely on shallow road features, including radial, topological, texture [5–7], and morphological features [8,9]. For instance,

Yager et al. [10] utilized edge features in conjunction with a support vector machine to extract roads from remote sensing images. To address the challenge of identifying intersections in urban road extraction, Gamba et al. [11] employed road topology to verify intersection extraction results. Additionally, Shi et al. [12] focused on the morphological characteristics of roads, extracting reliable road segments and applying local linear kernel regression to determine smooth road centerlines. However, most of these shallow features are manually designed, which is time-consuming and requires expert knowledge. Furthermore, their extraction accuracy is often limited because shallow features are typically too simplistic to handle complex background environments effectively.

Deep learning methods have recently been widely adopted for road extraction due to their ability to automatically learn complex features from large-scale datasets [13–15]. These methods treat road extraction as a semantic segmentation task, classifying each pixel as either "road" or "background". As a result, many classical semantic segmentation models have been directly applied to road extraction, including UNet [16], SegNet [17], LinkNet [18], and DeepLabV3+ [19]. Subsequently, specific neural networks have been proposed for road extraction. These networks primarily adopt three key strategies:

(1) Expanding the receptive field and capturing long-range context. D-LinkNet [20] enhances LinkNet by integrating dilated convolution operations and skip connections, which expand the receptive field while preserving detailed information. SIINet [21] adopts a spatial information inference structure to learn both local visual features of roads and global information, alleviating the occlusion problem. To capture long-range context, NL-LinkNet [22] incorporates nonlocal operations into LinkNet's encoder. Moreover, the Transformer architecture excels at modeling long-range dependencies, leading to the development of Transformer-based road extraction methods. Luo et al. [23] proposed BDTNet, which employs a Bi-direction Transformer Module (BDTM) to capture contextual information of roads. To extract roads precisely, UMiTNet [24] was developed. It consists of four mix-Transformer blocks for global feature extraction and a Dilated Attention Module (DAM) for semantic feature fusion. Inspired by the sparse target pixels in remote sensing images, Chen et al. [25] proposed the Sparse Token Transformer (STT) to learn sparse feature representations. STT not only reduces computational complexity but also enhances extraction accuracy.

(2) Emphasizing the geometric attributes of roads. Roads exhibit distinctive geometric attributes, such as direction, connectivity, shape, and topology. Ding et al. [26] proposed the Direction-Aware Residual Network (DiResNet), incorporating direction supervision during training. Besides road extraction, CoANet [27] employs a connectivity attention module to explore the relationship between neighboring pixels. Consequently, road connectivity is well preserved. RSANet [28] is proposed to address the challenges of complex road shapes. It uses the Efficient Strip Transformer Module (ESTM) to model the long-distance dependencies required by long roads. And Road Edge Focal loss (REF loss) is introduced to alleviate sample imbalance caused by thin roads. Considering the topology of road networks, SDUNet [29] was designed to learn multi-level features and global prior information of road networks. From the perspective of constraints on model learning, Hu et al. [30] proposed PolyRoad. It uses a polyline matching cost and additional losses for improved road topology.

(3) Reducing the labeled data needed for training. Constructing large-scale labeled datasets is both costly and time-consuming. To address this issue, a semi-supervised network, SemiRoadExNet [31], was proposed to leverage pseudo-label information. Road extraction methods based on unsupervised learning do not rely on labeled datasets. To tackle the domain shift challenge, Zhang et al. [32] designed RoadDA, a two-stage unsupervised domain adaptation network for road extraction. Besides

these standard models, researchers have explored large models for road extraction. For example, Chen et al. [33] proposed RSPrompter to learn the generation of appropriate prompts for the Segment Anything Model (SAM) [34]. RSPrompter enables the SAM to produce semantically discernible segmentation results for remote sensing images. Moreover, Hetang et al. [35] improved the SAM by designing SAM-Road to extract road networks.

Although deep learning-based road extraction methods have achieved great success, they still face the following challenges.

C1: In certain scenes, roads often exhibit low contrast with their surroundings due to similar textures or colors. Consequently, existing methods face significant challenges in distinguishing these low-contrast roads from the background.

C2: In remote sensing images, roads often occupy a relatively small proportion and have elongated and curved morphology, particularly in rural areas. This further increases the difficulty of road extraction.

Existing road extraction methods typically rely on RGB cues within the spatial domain, which are ineffective in distinguishing low-contrast roads from the background (C1). Inspired by predator hunting systems, where frequency information often proves more advantageous than RGB features for distinguishing specific prey in natural environments [36], we leverage frequency cues to address challenge C1. Subsequent research [37,38] has shown that low-contrast objects and backgrounds are more distinguishable in the frequency domain. Therefore, we utilize frequency cues to enhance the localization of low-contrast roads, as illustrated in Highlight 1 of Figure 1.

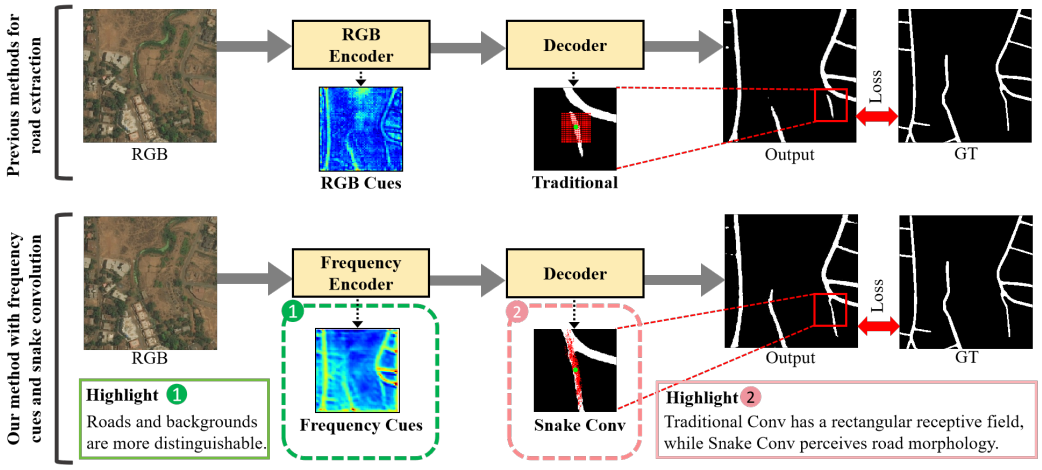

Figure 1. Motivation of our method. "Traditional" stands for traditional convolution, and "Snake Conv" stands for snake convolution. The major highlights of our method lie in frequency cues and snake convolution. Frequency cues effectively distinguish low-contrast roads from the background, while snake convolution accurately perceives road morphology.

To tackle the second challenge, C2, we introduce snake convolution [39] to enhance the morphological perception in road extraction as shown in Highlight 2 of Figure 1. Roads often present significant difficulties for accurate extraction due to their curved morphology and narrowness, which conflicts with the rectangular receptive field of traditional convolution. In contrast, snake convolution emulates the twisting motion of snakes, dynamically adjusting the receptive field to conform more closely to the elongated and curved road morphology. This mechanism allows snake convolution to achieve finer and more accurate road extraction, effectively addressing the geometric complexities of road morphology.

Above all, we propose an end-to-end network, LCMorph, for extracting low-contrast roads. LCMorph comprises frequency-enhanced localization and morphology-enhanced

extraction. During the frequency-enhanced localization stage, we introduce the Frequency-Aware Module (FAM) to extract frequency cues from RGB features. Frequency cues serve as critical points for distinguishing low-contrast roads. They help identify the rough positions of low-contrast roads, resulting in a coarse mask that suppresses background interference. Subsequently, the Morphological Perception Block (MPBlock) is proposed in the morphology-enhanced extraction stage to perceive road morphology. MPBlock is based on snake convolution, which enables accurate road extraction.

The remaining sections of this paper are organized as follows:

Section 2 describes the datasets and model architecture of LCMorph. Section 3 covers implementation details and results from comparison experiments. Section 4 provides a comprehensive discussion on ablation experiments and computational efficiency. Finally, Section 5 concludes this paper and discusses future work.

2. Materials and Methods

2.1. Datasets

2.1.1. DeepGlobe Dataset

The DeepGlobe dataset is derived from the DeepGlobe Road Extraction Challenge [40]. It covers 2220 square kilometers of land in Thailand, Indonesia, and India and contains urban and rural roads. Each image is 1024 × 1024 pixels with a resolution of 0.5 m/pixel. DeepGlobe provides pixel-level annotations categorized into road and background. There are 6226 images in DeepGlobe, and we divided them into training, validation, and test sets in a ratio of 8:1:1, resulting in 4980, 623, and 623 images, respectively. To improve the efficiency of model training, the original images were cropped into patches with an image size of 512 × 512. Example images from the DeepGlobe dataset are displayed in Figure 2.

Figure 2. Example images and ground truth from the DeepGlobe dataset, the Massachusetts Roads dataset, and our LC-Roads dataset. There are two classes in the ground truth: background (black) and road (white).

2.1.2. Massachusetts Roads Dataset

The Massachusetts Roads dataset, built by Mihn and Hinton [41], covers a wide variety of urban, suburban, and rural areas in the Massachusetts state. It includes 1108 training images, 14 validation images, and 49 test images. Each image is 1500 × 1500 pixels with a resolution of 1 m/pixel. After cropping the original images to 512 × 512 pixels, we obtained 9972 images for training, 126 images for validation, and 441 images for testing. Example images from the Massachusetts Roads dataset are shown in Figure 2.

2.1.3. LC-Roads Dataset

There is no existing dataset specifically for low-contrast roads, and the DeepGlobe dataset not only has a high resolution but also covers a variety of scenes. Therefore, we selected low-contrast road images from the DeepGlobe dataset and constructed a low-contrast road dataset, LC-Roads. The criteria for image selection were as follows: (a) low-contrast roads had to have textures or colors similar to the background, or the road boundaries had to be blurred; (b) low-contrast roads had to account for one-third or more of the total number of roads in each image; (c) low-contrast roads had to be labeled in the ground truth. The LC-Roads dataset has 2272 images with an image size of 512 × 512. The numbers of images in training, validation, and test sets were 1818, 227, and 227, respectively, according to a ratio of 8:1:1. Example images from our LC-Roads dataset are displayed in Figure 2.

2.2. Proposed Method

LCMorph consists of frequency-enhanced localization and morphology-enhanced extraction, as shown in Figure 3. The former leverages frequency cues to distinguish low-contrast roads from the background, while the latter continually recovers road details through morphological perception, ultimately achieving accurate road extraction.

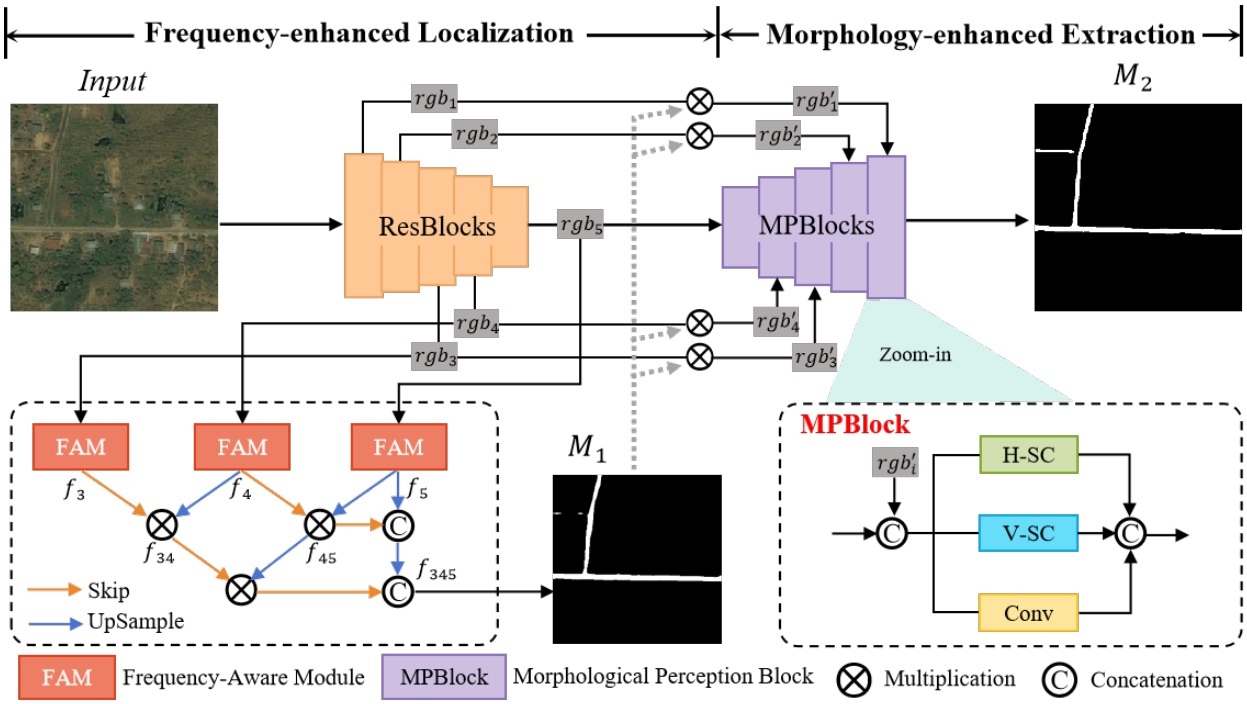

Figure 3. Overview of our LCMorph. It adopts an encoder–decoder architecture. In the frequency-enhanced localization stage, ResNet101 is employed as the encoder to extract RGB features rgb_i. Subsequently, Frequency-Aware Modules (FAMs) mine frequency cues from high-level RGB features. Frequency cues of different levels are fused in pairs to determine low-contrast roads' rough positions M_1. Under the guidance of M_1, RGB features are refined and then input into the decoder. In the morphology-enhanced extraction stage, the decoder comprises Morphological Perception Blocks (MPBlocks). Each MPBlock contains a horizontal snake convolution operation (H-SC), a vertical snake convolution operation (V-SC), and a traditional convolution operation (Conv). The decoder can perceive road morphology and ultimately produce accurate extraction results M_2.

2.2.1. Frequency-Enhanced Localization

Frequency-enhanced localization involves extracting RGB features, mining frequency cues, and aggregating multi-scale features.

We employ ResNet101 [42] to extract different levels of RGB features rgb_i, where $i \in \{1, 2, 3, 4, 5\}$:

$$rgb_i = ResBlock(rgb_{i-1}). \tag{1}$$

Frequency cues are more appropriate for low-contrast road extraction than RGB cues. Unlike Discrete Fourier Transform (DFT) [43], Discrete Cosine Transform (DCT) [44], and Discrete Wavelet Transform (DWT) [45], octave convolution [46] can learn frequency cues in an end-to-end manner. It decomposes images into high-frequency and low-frequency components. Therefore, based on octave convolution, we propose the FAM to mine frequency cues for low-contrast roads. The specific process is shown in Figure 4.

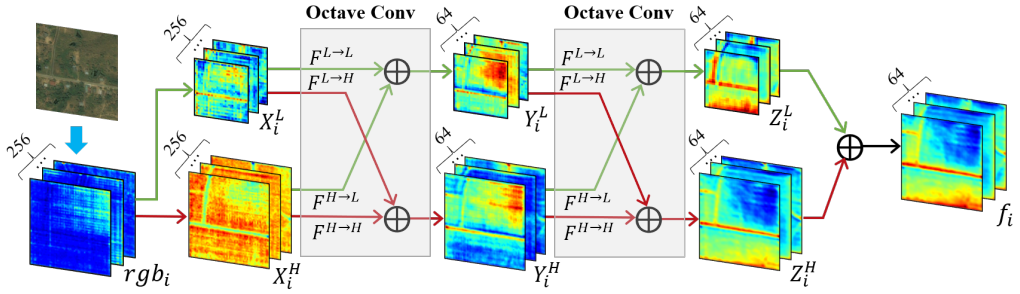

Figure 4. Illustration of Frequency-Aware Module (FAM). "Octave Conv" stands for octave convolution. Green and red paths represent the learning of low- and high-frequency information, respectively.

For high-level rgb_i, it is decomposed into X_i^L and X_i^H in Equation (2), with $i \in \{3, 4, 5\}$:

$$\begin{aligned} X_i^L &= F(Pool(rgb_i), W^{H\to L}), \\ X_i^H &= F(rgb_i, W^{H\to H}), \end{aligned} \tag{2}$$

where $F(X, W)$ indicates a convolution operation with parameters W and $Pool(\cdot)$ indicates an average pooling operation. Next, we obtain low-frequency components Y_i^L and high-frequency components Y_i^H by using the first octave convolution operation:

$$\begin{aligned} Y_i^L &= F(X_i^L, W^{L\to L}) \oplus F(Pool(X_i^H), W^{H\to L}), \\ Y_i^H &= F(X_i^H, W^{H\to H}) \oplus Up(F(X_i^L, W^{L\to H})), \\ &\Downarrow \\ Y_i^L, Y_i^H &= \overline{OctConv}(X_i^L, X_i^H), \end{aligned} \tag{3}$$

where $Up(\cdot)$ represents an up-sampling operation using nearest-neighbor interpolation and $\oplus$ denotes element-wise addition. And low-frequency components Z_i^L and high-frequency components Z_i^H are further obtained in the second octave convolution operation:

$$Z_i^L, Z_i^H = OctConv(Y_i^L, Y_i^H). \tag{4}$$

Considering that both high-frequency and low-frequency components play crucial roles in low-contrast road extraction, Z_i^L and Z_i^H are fused to form complete frequency cues in Equation (5):

$$f_i = Up(Z_i^L) \oplus Z_i^H, \tag{5}$$

where $Up(\cdot)$ is an up-sampling operation and $\oplus$ denotes element-wise addition.

To extract roads of different sizes, frequency cues of various levels are gradually fused, fully leveraging cross-layer semantic information, as shown in Figure 3. The specific fusion process is described in Equation (6):

$$\begin{aligned} f_{34} &= f_3 \otimes Up(f_4), \\ f_{45} &= f_4 \otimes Up(f_5), \\ f_{345} &= (f_{34} \otimes Up(f_{45})) \copyright Up(f_{45} \copyright Up(f_5)), \end{aligned} \tag{6}$$

where $\otimes$ denotes element-wise multiplication, $Up(\cdot)$ denotes an up-sampling operation, $\copyright$ denotes concatenation followed by a 3×3 convolution operation, and f_{345} represents the final frequency cues. A simple convolution operation is then used to obtain the coarse mask M_1 from f_{345}, which indicates the rough positions of low-contrast roads.

Ultimately, RGB features are refined under the guidance of M_1:

$$rgb_i^{'} = rgb_i + rgb_i \otimes \sigma(M_1), \tag{7}$$

where $\otimes$ denotes element-wise multiplication, $\sigma(\cdot)$ denotes the Sigmoid function, rgb_i denotes RGB features before refinement, and $rgb_i^{'}$ denotes the refined features.

2.2.2. Morphology-Enhanced Extraction

In the morphology-enhanced extraction stage, the decoder consists of MPBlocks. Each MPBlock contains a horizontal snake convolution operation (H-SC) [39], a vertical snake convolution operation (V-SC), and a traditional convolution operation.

Traditional convolution usually has a rectangular receptive field, which is incompatible with elongated and curved roads. However, snake convolution incorporates roads' snake-like morphology as prior knowledge to adjust the receptive field adaptively. The schematic illustration of snake convolution is shown in Figure 5.

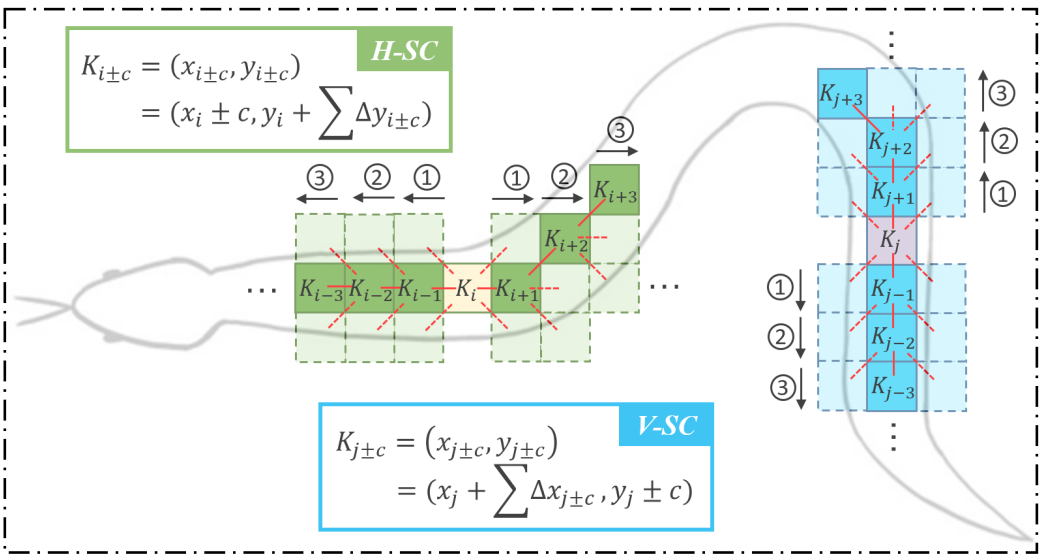

Figure 5. Illustration of snake convolution. "H-SC" (green squares) and "V-SC" (blue squares) denote horizontal and vertical snake convolution operations, respectively. The solid-line squares represent actual sampling points, while the dashed-line squares denote potential sampling points. The arrows illustrate the step-by-step process of determining sampling points, extending from the central sampling point toward both directions. The numbers on the arrows indicate the horizontal or vertical distance between the current sampling point and the central one.

In the case of a horizontal snake convolution operation unfolded from a 3×3 traditional convolution operation, the coordinates of each sampling point $K_{i\pm c}$ are $(x_{i\pm c}, y_{i\pm c})$,

where $K_i = (x_i, y_i)$ is the central sampling point and $c = \{0, 1, 2, 3, 4\}$ denotes the horizontal distance from K_i. In snake convolution, determining each sampling point's coordinates is progressive. Starting from K_i, the coordinates of the sampling point away from K_i depend on the previous sampling point and an offset $\Delta \in [-1, 1]$: compared with K_{i-1} and K_{i+1}, Δ is added to K_{i-2} and K_{i+2}, respectively. Eventually, the coordinates of sampling points in a horizontal snake convolution operation are calculated as Equation (8):

$$K_{i\pm c} = \begin{cases} (x_{i+c}, y_{i+c}) = (x_i + c, y_i + \sum_i^{i+c} \Delta y) \\ (x_{i-c}, y_{i-c}) = (x_i - c, y_i + \sum_{i-c}^{i} \Delta y). \end{cases} \tag{8}$$

Similarly, the coordinates of a vertical snake convolution operation can be deduced as in Equation (9):

$$K_{j\pm c} = \begin{cases} (x_{j+c}, y_{j+c}) = (x_j + \sum_j^{j+c} \Delta x, y_j + c) \\ (x_{j-c}, y_{j-c}) = (x_j + \sum_{j-c}^{j} \Delta x, y_j - c). \end{cases} \tag{9}$$

2.2.3. Loss Function

Road extraction can be regarded as a binary classification task at the pixel level. Therefore, binary cross-entropy loss (BCE Loss) is chosen as a term of the loss function. As shown in Equation (10), $p_i \in [0, 1]$ denotes the probability that the model predicts a pixel as road. y_i is the label of this pixel and takes the value of 0 or 1 to represent background or road, respectively. N is the total number of pixels in an image.

$$L^{BCE} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]. \tag{10}$$

However, roads cover a small proportion of remote sensing images, suffering from sample imbalance during training. Dice Loss can alleviate the issue of foreground's small proportion and perform well in binary classification tasks. Therefore, Dice Loss is selected as the second term of the loss function. In Equation (11), p_i, y_i, and N have the same meanings as in Equation (10), and ε is a tiny positive number used to avoid the case where the denominator is zero.

$$L^{Dice} = 1 - \frac{2 \sum_{i=1}^{N} y_i p_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \varepsilon}. \tag{11}$$

As shown in Figure 3, LCMorph has two outputs: a coarse mask M_1 and a final mask M_2. To enhance the learning ability, both outputs need to be supervised. The overall loss of LCMorph is

$$L = \sum_{i=1}^{2} (L_i^{BCE} + L_i^{Dice}). \tag{12}$$

3. Results

3.1. Implementation Details and Evaluation Metrics

The training epochs were 100 for the DeepGlobe dataset, 120 for the Massachusetts Roads dataset, and 80 for the LC-Roads dataset. The batch size was set to 8. We used the Adam optimizer to optimize our network. And the PolyLR policy was adopted to gradually reduce the learning rate, as shown in Equation (13):

$$lr = init\_lr(1 - \frac{iter}{max\_iter})^{power}, \tag{13}$$

where lr denotes the learning rate, $init\_lr = 0.01$ denotes the initial learning rate, $iter$ denotes the number of batches engaged in training, max_iter denotes the total number of batches, and $power = 3$ is a hyper-parameter. For all methods and variants, we used the same dataset division and hyper-parameter settings to ensure fairness. Additionally, we expanded the datasets with data augmentation techniques such as random rotation, horizontal flipping, and Gaussian blurring to enhance the model's generalization.

To evaluate the road extraction performance of our method, we used four popular metrics: precision (P), recall (R), F1 score (F1), and intersection over union (IoU). Since road extraction results are susceptible to sample imbalance, F1 score and IoU are more comprehensive and objective for assessing model performance [47].

3.2. Comparison Experiments on LC-Roads Dataset

To validate LCMorph's superiority in the low-contrast road extraction task, some advanced methods were selected:

(1) UNet [16]. U-Net is a widely used semantic segmentation model featuring a symmetric encoder–decoder architecture with skip connections, enabling precise segmentation and context capture.
(2) SegNet [17]. SegNet utilizes an encoder–decoder architecture, where the decoder employs pooling indices from the encoder for up-sampling, which preserves spatial details.
(3) LinkNet [18]. LinkNet is an efficient semantic segmentation model designed for real-time applications. It combines an encoder–decoder architecture with residual connections to maintain high accuracy with fewer parameters.
(4) DeepLabV3+ [19]. DeepLabV3+ employs an encoder–decoder architecture with atrous convolution. The encoder captures multi-scale contextual information, and the decoder is simple yet effective.
(5) PSPNet [48]. To capture global contextual information, PSPNet introduces pyramid pooling modules, enhancing the model's ability to understand various object scales.
(6) D-LinkNet [20]. D-LinkNet is a classical road extraction model. Based on LinkNet, D-LinkNet contains dilated convolution layers to expand the receptive field. It won first place in the CVPR DeepGlobe 2018 Road Extraction Challenge.
(7) SIINet [21]. SIINet enhances road extraction by facilitating multidirectional message passing between pixels. It effectively captures both local and global spatial information.
(8) CoANet [27]. CoANet is a road extraction model which integrates strip convolution operations with a connectivity attention module. It addresses occlusions and achieves good results.
(9) NL-LinkNet [22]. NL-LinkNet is the first road extraction model to use nonlocal operations. The nonlocal block enables the model to capture long-range dependencies and distant information.
(10) SDUNet [29]. SDUNet is a spatially enhanced and densely connected UNet. It aggregates multi-level features and preserves road structure.
(11) DSCNet [39]. DSCNet is a tubular structure segmentation model applied to both vessel segmentation and road extraction. Snake convolution is proposed by DSCNet.
(12) RoadExNet. RoadExNet is the generator of SemiRoadExNet [31]. For fair comparison, we trained RoadExNet in a fully supervised manner.
(13) OARENet [49]. OARENet is a road extraction model designed to address dense occlusions. It proposes an occlusion-aware decoder, achieving excellent performance in complex scenes.

As shown in Table 1, LCMorph achieves comprehensive state-of-the-art (SOTA) performance compared with classical semantic segmentation models. Among the road extraction models, LCMorph achieves the highest recall, F1 score, and IoU, with precision ranking third. CoANet, DSCNet, and RoadExNet employ specific modules for road shapes, leading to high precision. However, DSCNet's recall is significantly lower than that of LCMorph, indicating its limited ability to distinguish low-contrast roads from the background. Although OARENet can address the occlusion problem, it lacks a model structure capable of handling low-contrast scenes. Therefore, it performs poorly on the LC-Roads dataset. In comparison with CoANet, which ranks second, LCMorph achieves a 3% improvement in recall, a 1.1% improvement in F1 score, and a 1.45% improvement in IoU. The largest boost in recall is mainly due to LCMorph's effectiveness in extracting low-contrast roads often missed by other methods.

Table 1. Quantitative comparison of our LCMorph with some advanced methods on LC-Roads dataset, in which **bold** denotes the best and underline denotes the second best.

Method	Source	P (%)	R (%)	F1 (%)	IoU (%)
Semantic Segmentation Models					
UNet	MICCAI'15	70.89	73.90	72.36	56.70
PSPNet	CVPR'17	62.96	77.60	69.52	53.27
LinkNet	VCIP'17	71.16	75.71	73.37	57.94
SegNet	TPAMI'17	72.32	76.19	74.20	58.99
DeepLabV3+	ECCV'18	66.63	77.87	71.81	56.02
Road Extraction Models					
D-LinkNet	CVPRW'18	71.87	79.65	75.56	60.72
SIINet	ISPRS'19	71.16	72.72	71.93	56.23
CoANet	TIP'21	72.76	78.99	75.74	60.95
NL-LinkNet	GRSL'22	71.48	75.63	73.50	58.10
SDUNet	PR'22	72.03	73.04	72.53	56.95
DSCNet	ICCV'23	71.09	75.65	73.30	57.91
RoadExNet	ISPRS'23	**72.82**	75.84	74.33	59.14
OARENet	TGRS'24	72.14	72.51	72.21	56.51
LCMorph	Ours	72.32	**81.99**	**76.85**	**62.40**

In addition to the quantitative results, the qualitative results of each method were compared on the LC-Roads test set. Due to limited space, Figure 6 only displays the six best methods, including LCMorph. The red boxes highlight differences among these methods, and the yellow boxes indicate roads ignored by the ground truth. In the first row of Figure 6, roads have similar colors to the background, resulting in low-contrast roads below the image. In the second row of Figure 6, roads and their surroundings are composed of the same material, making both color and texture similar, which complicates road extraction. In these low-contrast scenes, roads extracted by LCMorph are both complete and continuous, particularly for roads with blurred boundaries in the red boxes. In the third row of Figure 6, the road in the red box is tightly wrapped by dense buildings, making it difficult to extract. Additionally, its morphology is irregular, unlike common straight roads. However, LCMorph still achieves complete and accurate extraction results compared with other methods. Furthermore, the road in the yellow box is neglected in the ground truth but is successfully extracted by all deep learning methods. This suggests that neural networks can learn the intrinsic characteristics of roads rather than simply fitting to the ground truth. In the last row of Figure 6, road morphology in the red box is more complicated, and tree occlusion further increases the extraction difficulty. LCMorph's extraction result has the most adequate road details, and its road morphology is the closest to the ground

truth. Table 1 and Figure 6 demonstrate the advantages of LCMorph in low-contrast road extraction tasks from both quantitative and qualitative perspectives.

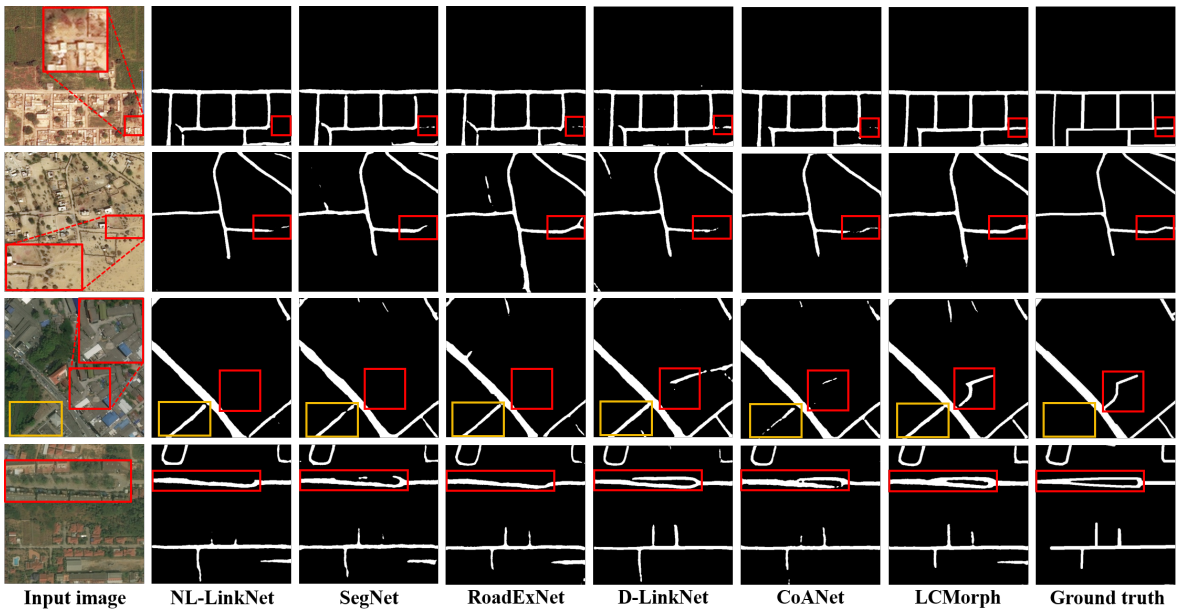

Figure 6. Qualitative comparison of our LCMorph with some advanced methods on LC-Roads dataset. Red boxes highlight differences among methods. Yellow boxes indicate roads ignored by ground truth.

3.3. Comparison Experiments on Public Datasets

Compared with LC-Roads, the DeepGlobe dataset and Massachusetts Roads dataset contain more diverse road scenes, particularly salient roads. Therefore, we also conducted comparison experiments on these two public datasets to verify LCMorph's versatility. When faced with diverse road scenes, LCMorph attains satisfactory performance on the DeepGlobe dataset and Massachusetts Roads dataset. According to Table 2, it achieves the highest F1 score and IoU on these two datasets.

On the DeepGlobe dataset, LCMorph ranks second in precision and has competitive recall. Compared with OARENet (SOTA precision), LCMorph improves recall by 4.32%, and it improves precision by 4.49% compared with CoANet (SOTA recall).

On the Massachusetts Roads dataset, LCMorph exhibits the second-highest recall that is 5.62% superior to RoadExNet (SOTA precision). And LCMorph's precision surpasses D-LinkNet's by 2.36%. These results demonstrate that LCMorph can better balance precision and recall, resulting in the best F1 score and IoU.

Figure 7 displays LCMorph's road extraction results in four representative scenes from these two datasets. LCMorph accurately extracts low-contrast roads in Figure 7a, particularly the curved road in the green box. For land cover such as ridges in the field that resemble roads (yellow boxes), LCMorph avoids misidentifying them as roads. Although easy to distinguish, the roads in Figure 7b are occluded by trees, buildings, and shadows. LCMorph completely extracts partially occluded roads (green boxes). As for fully occluded roads (red boxes), LCMorph's performance needs improvement. Severe occlusion is also a challenge for other road extraction methods. In Figure 7c, roads in the upper-left green box exhibit low contrast, while those in the lower-left green box display significant curvature. All these roads are accurately extracted with the FAM and snake convolution. Roads in Figure 7d have complex morphology, including curved roads, forked roads, merged roads, and adjacent roads, as shown in the green box. With the powerful morphological perception of snake convolution, LCMorph extracts roads with complex morphology in

a fine-grained manner. The qualitative results in Figure 7 provide strong evidence of LCMorph's versatility.

Table 2. Quantitative comparison of our LCMorph with some advanced methods on DeepGlobe dataset and Massachusetts Roads dataset, in which **bold** denotes the best and underline denotes the second best.

Method	Source	DeepGlobe Dataset				Massachusetts Roads Dataset			
		P (%)	R (%)	F1 (%)	IoU (%)	P (%)	R (%)	F1 (%)	IoU (%)
Semantic Segmentation Models									
UNet	MICCAI'15	73.20	70.10	71.62	59.06	77.29	72.13	73.19	59.46
PSPNet	CVPR'17	75.35	78.71	76.99	62.95	76.53	69.47	72.83	57.27
LinkNet	VCIP'17	71.33	79.81	75.33	61.34	79.18	73.71	75.52	61.74
SegNet	TPAMI'17	79.84	75.92	77.83	63.79	72.79	77.41	74.26	60.11
DeepLabV3+	ECCV'18	78.20	76.24	75.69	62.33	75.47	77.97	76.70	62.25
Road Extraction Models									
D-LinkNet	CVPRW'18	73.50	81.38	77.24	63.36	74.57	**78.85**	75.58	61.75
SIINet	ISPRS'19	75.42	83.15	79.09	64.35	73.47	70.89	72.16	56.85
CoANet	TIP'21	74.02	**85.31**	79.27	65.65	75.15	77.89	76.48	61.94
NL-LinkNet	GRSL'22	74.99	77.50	76.23	62.65	79.14	74.17	76.57	62.19
SDUNet	PR'22	78.40	80.43	79.40	65.91	77.56	74.57	75.23	61.34
DSCNet	ICCV'23	77.03	75.91	76.47	62.76	75.83	77.47	76.64	62.22
RoadExNet	ISPRS'23	77.76	77.14	77.45	63.51	**82.46**	72.89	77.38	63.10
OARENet	TGRS'24	**79.88**	76.70	78.26	64.04	77.79	75.23	76.49	61.96
LCMorph	Ours	78.51	81.02	**79.74**	**66.30**	76.93	78.51	**77.71**	**63.55**

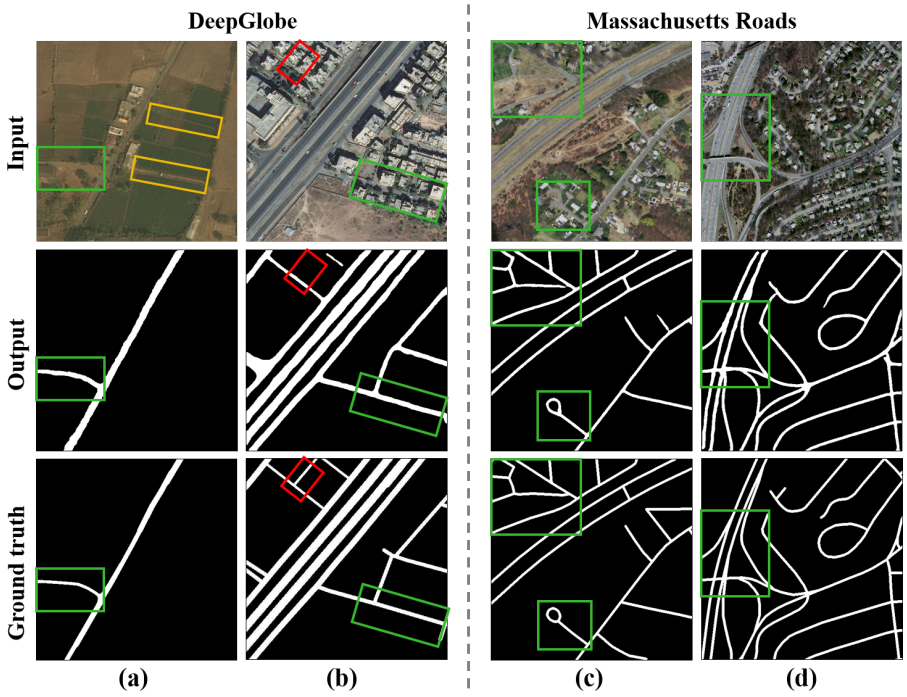

Figure 7. Qualitative results of our LCMorph on DeepGlobe dataset and Massachusetts Roads dataset. Green boxes denote challenging roads successfully extracted by LCMorph. Red boxes represent challenging roads that LCMorph fails to extract. Yellow boxes denote land cover similar to roads. (**a**) Low-contrast roads in DeepGlobe dataset. (**b**) Salient roads in DeepGlobe dataset. (**c**) Low-contrast roads in Massachusetts Roads dataset. (**d**) Salient roads in Massachusetts Roads dataset.

4. Discussion

4.1. Effect of Different Encoders

Roads in remote sensing images exhibit multi-scale variations. As a result, the encoder needs a powerful multi-scale feature extraction capability. With its hierarchical feature extraction mechanism, ResNet effectively captures information at different scales, from local details to semantic context. Therefore, we selected ResNet as the encoder. As ResNet has multiple variants, in order to determine the most suitable encoder, we explored the effect of different ResNet variants, as shown in Table 3. LCMorph-light adopts ResNet50 as its encoder, while LCMorph-heavy utilizes ResNet152. Apart from the encoder, LCMorph-light and LCMorph-heavy both have the same structure as LCMorph.

Table 3. Effect of different encoders on model performance on LC-Roads dataset.

Method	Encoder	P (%)	R (%)	F1 (%)	IoU (%)	Params (M)	FLOPs (G)
LCMorph-light	ResNet50	72.64	77.26	74.88	59.85	52.90	216.55
LCMorph	ResNet101	72.32	81.99	76.85	62.40	71.90	294.75
LCMorph-heavy	ResNet152	73.85	80.27	76.93	62.73	87.54	358.62

Despite its lighter model architecture, LCMorph-light proves significantly less effective in road extraction than LCMorph. Compared with ResNet101, ResNet152 is deeper. In Table 3, LCMorph-heavy achieves slightly higher F1 score and IoU than LCMorph, but its parameters increase by 21.75%, and its FLOPs increase by 21.67%. These results suggest that ResNet101 offers a better trade-off between model effectiveness and efficiency. Therefore, ResNet101 was selected as the encoder.

4.2. Effectiveness of Each Module in LCMorph

To verify the effectiveness of modules in LCMorph, we performed ablation experiments on the LC-Roads dataset. The following experiments use the same parameter settings. The experimental results are shown in Table 4, where "✓" denotes that the corresponding module is used. "FAM" stands for Frequency-Aware Module, and "MPBlock" stands for Morphological Perception Block. Additionally, method 4 is LCMorph, and method 1 is the baseline which replaces the FAM with a Receptive Field Block (RFB) [50] and snake convolution with traditional convolution.

Table 4. Ablation study for proposed modules on LC-Roads dataset, in which **bold** denotes the best.

No.	FAM	MPBlock	P (%)	R (%)	F1 (%)	IoU (%)
1			69.82	81.66	75.28	60.35
2	✓		69.77	**82.36**	75.54	60.70
3		✓	72.18	81.31	76.47	61.90
4	✓	✓	**72.32**	81.99	**76.85**	**62.40**

Method 2 improves recall by 0.7%, F1 score by 0.26%, and IoU by 0.35% in the presence of the FAM. Similarly, method 4 shows an increase compared with method 3. This demonstrates that the FAM has a significant positive effect. And the biggest increase in recall suggests that frequency cues are good at discovering low-contrast roads which are ignored by RGB cues. Incorporating snake convolution into the network also improves the road extraction results. Compared with the baseline, method 3 increases precision by 2.36%, F1 score by 1.46%, and IoU by 1.55%. Method 4 also substantially improves these three metrics compared with method 2. Snake convolution pays more attention to the elongated road region and extracts roads more finely, leading to the greatest boost in precision. Under the joint effect of the FAM and MPBlock, LCMorph achieves the highest precision, F1 score,

and IoU. We also provide a variant of LCMorph with higher recall, LCMorph w/o MPBlock (method 2), which can handle more challenging low-contrast situations.

Figure 8 displays the qualitative results of the FAM and snake convolution on the LC-Roads dataset. The first two rows of Figure 8 depict rural scenes, while the last row is a remote sensing image taken over a town. Figure 8b visualizes RGB cues from the RFB, and Figure 8c visualizes frequency cues from the FAM. Compared with the RGB cues, low-contrast roads are more noticeable in frequency cues, which indicates that frequency cues can effectively distinguish these roads from the background. Thus, frequency cues are more suitable for extracting low-contrast roads. Additionally, some road segments (green boxes) are selected in the ground truth, and sampling points of traditional convolution and snake convolution are drawn in Figure 8e. For elongated roads, traditional convolution's receptive field is obviously deviated, losing focus on the road region. In contrast, snake convolution's receptive field closely fits the road and presents elongated morphology. This demonstrates that snake convolution can effectively perceive the road morphology and focus on the road region.

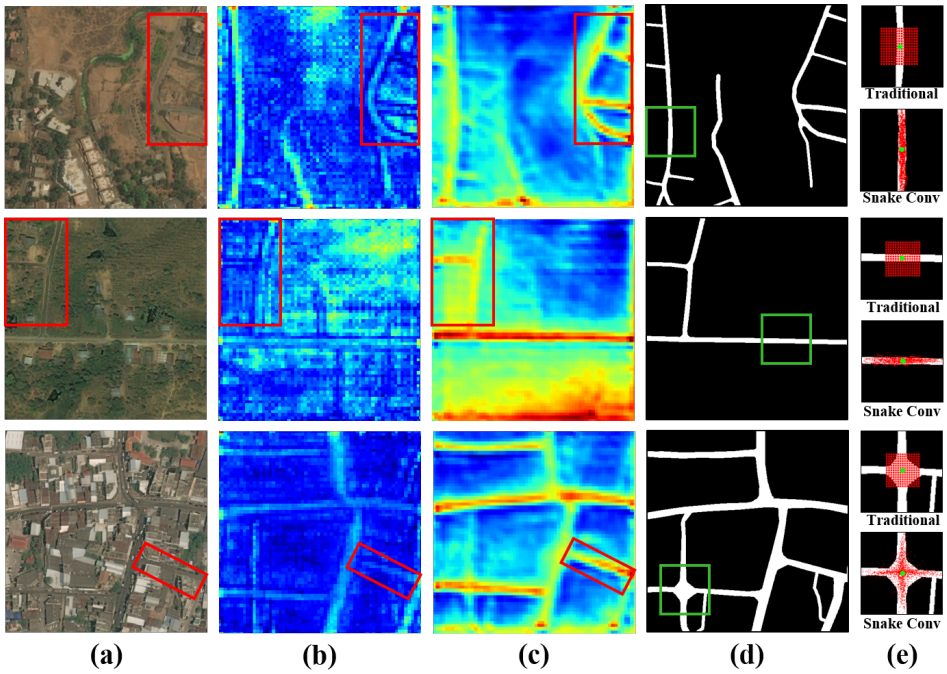

Figure 8. Qualitative results of proposed modules on LC-Roads dataset. Red boxes indicate low-contrast roads. Green boxes denote areas where receptive field is visualized. "Traditional" stands for traditional convolution, and "Snake Conv" stands for snake convolution. (**a**) Input image. (**b**) RGB cues without FAM. (**c**) Frequency cues with FAM. (**d**) Ground truth. (**e**) Receptive field.

4.3. Effect of Different Frequency Components

The effectiveness of the FAM has been verified in the above. However, the FAM decomposes RGB features into low-frequency components and high-frequency components, necessitating further investigation into how to use these frequency features for optimal road extraction results. We made different modifications to the second octave convolution operation in the FAM and primarily explored three approaches: using only low-frequency components, using only high-frequency components, and fusing both low-frequency and high-frequency components. The quantitative results are shown in Table 5, and the visualization of different frequency components is displayed in Figure 9.

Table 5. Quantitative results of different frequency components on LC-Roads dataset, in which **bold** denotes the best.

Approach	P (%)	R (%)	F1 (%)	IoU (%)
Low frequency	68.15	72.02	70.03	53.81
High frequency	**78.28**	72.18	75.11	60.14
Low and high frequency	76.95	**76.25**	**76.85**	**62.40**

The metrics of high-frequency components significantly outperform those of low-frequency components. This suggests that high-frequency components are more robust and discriminative for low-contrast roads, as supported by Figure 9b,c. This finding aligns with the human visual system, which typically uses high-frequency information to identify targets in uncertain regions. Further analysis of Figure 9 shows that high-frequency components focus on the details and textures of roads, including edges and lines, while low-frequency components focus on the overall layouts and structures of images, including smooth parts of road regions and background regions. Although high-frequency components are more prominent, low-frequency components also contain key cues needed to extract roads. Therefore, the optimal approach is to fuse high-frequency and low-frequency components, as validated in the last row of Table 5.

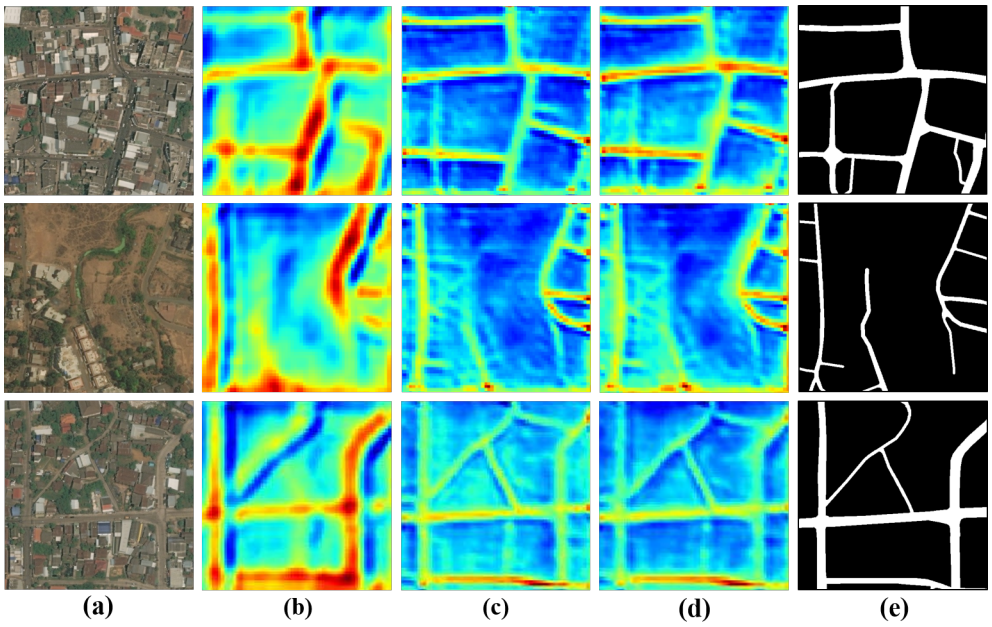

Figure 9. Visualization of different frequency components. (**a**) Input image. (**b**) Low-frequency components. (**c**) High-frequency components. (**d**) Fusion of low- and high-frequency components. (**e**) Ground truth.

4.4. Computational Efficiency

Additionally, we explored the computational efficiency of each method in terms of the number of parameters (Params) and Floating-Point Operations (FLOPs). As shown in Table 6, LCMorph achieves the highest IoU across all three datasets. Compared with methods such as UNet, SegNet, DeepLabV3+, CoANet, and OARENet, although LCMorph shows decreased computational efficiency, its IoU improves substantially. Moreover, LCMorph outperforms PSPNet and SDUNet in both computational efficiency and IoU. Overall, LCMorph's computational efficiency is moderate and could be further optimized through techniques such as pruning in the future.

Table 6. The computational efficiency of different methods. IoU_{DG} represents the IoU on the DeepGlobe dataset, IoU_{Mass} denotes the IoU on the Massachusetts Roads dataset, and IoU_{LC} indicates the IoU on the LC-Roads dataset.

Method	Source	Params (M)	FLOPs (G)	IoU_{DG} (%)	IoU_{Mass} (%)	IoU_{LC} (%)
Semantic Segmentation Models						
UNet	MICCAI'15	26.36	223.88	59.06	59.46	56.70
PSPNet	CVPR'17	86.06	327.02	62.95	57.27	53.27
LinkNet	VCIP'17	11.53	12.09	61.34	61.74	57.94
SegNet	TPAMI'17	29.48	170.45	63.79	60.11	58.99
DeepLabV3+	ECCV'18	54.70	83.24	62.33	62.25	56.02
Road Extraction Models						
D-LinkNet	CVPRW'18	31.10	33.60	63.36	61.75	60.72
SIINet	ISPRS'19	7.36	36.10	64.35	56.85	56.23
CoANet	TIP'21	59.15	277.58	65.65	61.94	60.95
NL-LinkNet	GRSL'22	21.82	32.07	62.65	62.19	58.10
SDUNet	PR'22	80.24	353.26	65.91	61.34	56.95
DSCNet	ICCV'23	4.52	40.38	62.76	62.22	57.91
RoadExNet	ISPRS'23	31.13	33.84	63.51	63.10	59.14
OARENet	TGRS'24	71.30	99.90	64.04	61.96	56.51
LCMorph	Ours	71.90	294.75	66.30	63.55	62.40

5. Conclusions

High inter-class similarity and complex road morphology pose significant challenges for road extraction in low-contrast scenes. To overcome these challenges, we propose LCMorph, the first end-to-end network designed for low-contrast road extraction. Additionally, we constructed a specialized dataset, LC-Roads, for low-contrast roads, with the aim of facilitating future research in this field. In summary, the contributions of this paper are as follows:

1. The Frequency-Aware Module (FAM) is introduced to enhance the distinction between low-contrast roads and the background. With its help, LCMorph effectively identifies overlooked low-contrast roads.
2. To handle elongated and curved roads, we propose the Morphological Perception Blocks (MPBlocks). These blocks adaptively adjust the receptive field to the road morphology, achieving accurate road extraction.
3. LCMorph achieves state-of-the-art performance in terms of F1 score and IoU on the LC-Roads, DeepGlobe, and Massachusetts Roads datasets. And the effectiveness of the FAM and MPBlock is validated through adequate ablation experiments.

Although low-contrast roads and backgrounds have similar colors or textures in RGB images, they exhibit different spectra when the materials are different. Therefore, spectral information may be helpful in distinguishing low-contrast roads. In the future, we plan to introduce spectral information into LCMorph to further improve the extraction of low-contrast roads.

Author Contributions: Conceptualization, X.L. and S.Y.; methodology, X.L. and S.Y.; software, S.Y. and W.L.; validation, F.M. and Z.Y.; formal analysis, X.L. and F.M.; investigation, S.Y., W.L., and Z.Y.; resources, X.L.; data curation, S.Y.; writing—original draft preparation, S.Y.; writing—review and editing, X.L. and F.M.; visualization, R.W.; supervision, X.L. and F.M.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Key Research and Development Project of China under grant 2021YFA1000103, in part by the Natural Science Foundation of Shandong Province of China under grant ZR2024MF048, and in part by the Key Laboratory of Marine Hazard Forecasting, Ministry of Natural Resources, under grant LOMF2202.

Data Availability Statement: The DeepGlobe dataset and Massachusetts Roads dataset are publicly available. Our LC-Roads dataset and codes can be accessedat https://github.com/VictorYang097/Low-contrast-roads (accessed on 5 July 2024).

Acknowledgments: We thank the authors of the DeepGlobe dataset and Massachusetts Roads dataset for providing the remote sensing road images used in our experiments.

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

References

1. Son, T.H.; Weedon, Z.; Yigitcanlar, T.; Sanchez, T.; Corchado, J.M.; Mehmood, R. Algorithmic urban planning for smart and sustainable development: Systematic review of the literature. *Sustain. Cities Soc.* **2023**, *94*, 104562. [CrossRef]
2. Żochowska, R.; Pamuła, T. Impact of Traffic Flow Rate on the Accuracy of Short-Term Prediction of Origin-Destination Matrix in Urban Transportation Networks. *Remote Sens.* **2024**, *16*, 1202. [CrossRef]
3. Chen, P.; Wu, J.; Li, N. A personalized navigation route recommendation strategy based on differential perceptron tracking user's driving preference. *Comput. Intell. Neurosci.* **2023**, *2023*, 8978398. [CrossRef] [PubMed]
4. Stewart, C.; Lazzarini, M.; Luna, A.; Albani, S. Deep learning with open data for desert road mapping. *Remote Sens.* **2020**, *12*, 2274. [CrossRef]
5. Lian, R.; Wang, W.; Mustafa, N.; Huang, L. Road extraction methods in high-resolution remote sensing images: A comprehensive review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5489–5507. [CrossRef]
6. Stoica, R.; Descombes, X.; Zerubia, J. A Gibbs point process for road extraction from remotely sensed images. *Int. J. Comput. Vis.* **2004**, *57*, 121–136. [CrossRef]
7. Sghaier, M.O.; Lepage, R. Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 1946–1958. [CrossRef]
8. Mohammadzadeh, A.; Tavakoli, A.; Valadan Zoej, M.J. Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened IKONOS images. *Photogramm. Rec.* **2006**, *21*, 44–60. [CrossRef]
9. Maurya, R.; Gupta, P.; Shukla, A.S. Road extraction using k-means clustering and morphological operations. In Proceedings of the 2011 International Conference on Image Information Processing, Shimla, India, 3–5 November 2011; pp. 1–6.
10. Yager, N.; Sowmya, A. Support vector machines for road extraction from remotely sensed images. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Groningen, The Netherlands, 25–27 August 2003; pp. 285–292.
11. Gamba, P.; Dell'Acqua, F.; Lisini, G. Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 387–391. [CrossRef]
12. Shi, W.; Miao, Z.; Debayle, J. An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 3359–3372. [CrossRef]
13. Nie, J.; Wang, Z.; Liang, X.; Yang, C.; Zheng, C.; Wei, Z. Semantic Category Balance-Aware Involved Anti-Interference Network for Remote Sensing Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4409712. [CrossRef]
14. Li, X.; Wang, Z.; Chen, C.; Tao, C.; Qiu, Y.; Liu, J.; Sun, B. SemID: Blind Image Inpainting with Semantic Inconsistency Detection. *Tsinghua Sci. Technol.* **2024**, *29*, 1053–1068. [CrossRef]
15. Liu, R.; Wu, J.; Lu, W.; Miao, Q.; Zhang, H.; Liu, X.; Lu, Z.; Li, L. A Review of Deep Learning-Based Methods for Road Extraction from High-Resolution Remote Sensing Images. *Remote Sens.* **2024**, *16*, 2056. [CrossRef]
16. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; pp. 234–241.
17. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
18. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
19. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.

20. 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 Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 182–186.
21. Tao, C.; Qi, J.; Li, Y.; Wang, H.; Li, H. Spatial information inference net: Road extraction using road-specific contextual information. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 155–166. [CrossRef]
22. 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]
23. Luo, L.; Wang, J.X.; Chen, S.B.; Tang, J.; Luo, B. BDTNet: Road extraction by bi-direction transformer from remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2505605. [CrossRef]
24. Deng, F.; Luo, W.; Ni, Y.; Wang, X.; Wang, Y.; Zhang, G. UMiT-Net: A U-shaped mix-transformer network for extracting precise roads using remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5801513. [CrossRef]
25. Chen, K.; Zou, Z.; Shi, Z. Building extraction from remote sensing images with sparse token transformers. *Remote Sens.* **2021**, *13*, 4441. [CrossRef]
26. Ding, L.; Bruzzone, L. DiResNet: Direction-aware residual network for road extraction in VHR remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 10243–10254. [CrossRef]
27. Mei, J.; Li, R.J.; Gao, W.; Cheng, M.M. CoANet: Connectivity attention network for road extraction from satellite imagery. *IEEE Trans. Image Process.* **2021**, *30*, 8540–8552. [CrossRef]
28. Wang, C.; Xu, R.; Xu, S.; Meng, W.; Wang, R.; Zhang, J.; Zhang, X. Toward accurate and efficient road extraction by leveraging the characteristics of road shapes. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4404616. [CrossRef]
29. Yang, M.; Yuan, Y.; Liu, G. SDUNet: Road extraction via spatial enhanced and densely connected UNet. *Pattern Recognit.* **2022**, *126*, 108549. [CrossRef]
30. Hu, Y.; Wang, Z.; Huang, Z.; Liu, Y. PolyRoad: Polyline Transformer for Topological Road-Boundary Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 5602112. [CrossRef]
31. Chen, H.; Li, Z.; Wu, J.; Xiong, W.; Du, C. SemiRoadExNet: A semi-supervised network for road extraction from remote sensing imagery via adversarial learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *198*, 169–183. [CrossRef]
32. Zhang, L.; Lan, M.; Zhang, J.; Tao, D. Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5609413. [CrossRef]
33. Chen, K.; Liu, C.; Chen, H.; Zhang, H.; Li, W.; Zou, Z.; Shi, Z. RSPrompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4701117. [CrossRef]
34. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 4015–4026.
35. Hetang, C.; Xue, H.; Le, C.; Yue, T.; Wang, W.; He, Y. Segment Anything Model for Road Network Graph Extraction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 2556–2566.
36. Zhong, Y.; Li, B.; Tang, L.; Kuang, S.; Wu, S.; Ding, S. Detecting camouflaged object in frequency domain. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4504–4513.
37. Cong, R.; Sun, M.; Zhang, S.; Zhou, X.; Zhang, W.; Zhao, Y. Frequency perception network for camouflaged object detection. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 1179–1189.
38. Xie, C.; Xia, C.; Yu, T.; Li, J. Frequency representation integration for camouflaged object detection. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 1789–1797.
39. Qi, Y.; He, Y.; Qi, X.; Zhang, Y.; Yang, G. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 6070–6079.
40. 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 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 172–181.
41. Mnih, V. *Machine Learning for Aerial Image Labeling*; University of Toronto: Toronto, ON, Canada, 2013.
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
43. Winograd, S. On computing the discrete Fourier transform. *Math. Comput.* **1978**, *32*, 175–199. [CrossRef]
44. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete cosine transform. *IEEE Trans. Comput.* **1974**, *100*, 90–93. [CrossRef]
45. Shensa, M.J. The discrete wavelet transform: Wedding the a trous and Mallat algorithms. *IEEE Trans. Signal Process.* **1992**, *40*, 2464–2482. [CrossRef]

46. Chen, Y.; Fan, H.; Xu, B.; Yan, Z.; Kalantidis, Y.; Rohrbach, M.; Yan, S.; Feng, J. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3435–3444.
47. Tong, Z.; Li, Y.; Zhang, J.; He, L.; Gong, Y. MSFANet: Multiscale fusion attention network for road segmentation of multispectral remote sensing data. *Remote Sens.* **2023**, *15*, 1978. [CrossRef]
48. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
49. Yang, R.; Zhong, Y.; Liu, Y.; Lu, X.; Zhang, L. Occlusion-aware road extraction network for high-resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5619316. [CrossRef]
50. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.