International Journal of Geo-Information

MDPI

Article

Topology-Aware Road Extraction from Remote Sensing Images Using Deep Learning and Graph-Based Connectivity Refinement

Zixuan Teng [1], Zezhong Zheng [1,2], Xiangyang Sun [3,*] and Hao Xue [4]

[1] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China; 202522070212@std.uestc.edu.cn (Z.T.); zezhongzheng@uestc.edu.cn (Z.Z.)
[2] The Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313000, China
[3] SHU-SUCG Research Centre for Building Industrialization, SILC Business School, Shanghai University, Shanghai 201800, China
[4] School of Geographical & Earth Science, University of Glasgow, Glasgow G12 8QQ, UK; 3007050X@student.gla.ac.uk
* Correspondence: sun_x_y1104@shu.edu.cn

Abstract

Road networks are fundamental components of transportation infrastructure and play a crucial role in various geospatial applications. Although deep learning-based semantic segmentation models have achieved promising results in extracting roads from high-resolution remote sensing imagery, the resulting networks often suffer from topological fragmentation due to occlusions and shadows. To address this issue, we propose a topology-aware road extraction method that integrates deep learning-based segmentation with a graph-based connectivity refinement strategy. Specifically, a Pyramid Scene Parsing Network (PSPNet) is first employed to generate initial road probability maps. Subsequently, a connectivity-oriented post-processing pipeline is introduced, which incorporates a multi-source cost function strategy and a direction-aware Dijkstra search algorithm. By utilizing endpoint tangent vectors as inertial weights, the algorithm effectively reconstructs fragmented segments while ensuring geometric smoothness and topological consistency. Furthermore, a dynamic road width restoration strategy is applied to transform refined skeletons into physically consistent road entities. Experiments conducted on two publicly available datasets, CHN6-CUG and DeepGlobe, demonstrate the effectiveness of the proposed method. Quantitative results show that the refinement process significantly enhances road connectivity with a minimal trade-off in pixel-level accuracy. Specifically, the Conn metric increases by 0.1989 on the CHN6-CUG dataset and 0.3055 on the DeepGlobe dataset, while MIoU remains high with only marginal decreases of 1.07% and 0.45%, respectively. These findings indicate that the method effectively restores structural continuity, helping with reliable road network generation and subsequent integration into Geographic Information System (GIS)-based applications such as urban planning and autonomous navigation.

Keywords: road extraction; semantic segmentation; graph-based connectivity refinement

Check for updates

Academic Editors: Wolfgang Kainz and Maria Antonia Brovelli

Received: 17 March 2026
Revised: 6 May 2026
Accepted: 6 May 2026
Published: 9 May 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. 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.

1. Introduction

Road networks serve as essential components of geographic infrastructure, connecting cities, towns, and rural areas while supporting the movement of people and goods. They play a vital role in urban planning [1,2], transportation management [3,4], and emergency response [5,6], and thus are critical for socio-economic activities and regional development.

In the past few decades, remote sensing technologies have developed rapidly, and high-resolution imagery from satellites, aerial platforms, and Unmanned Aerial Vehicles (UAVs) have become increasingly available [7]. These data enable detailed observation of ground objects at unprecedented spatial and temporal scales, providing a valuable foundation for geospatial information acquisition such as disaster monitoring [8]. Similarly, as a task where fine-grained spatial details and timely updates are essential, road extraction clearly relies heavily on the support of remote sensing imagery.

Road extraction from remote sensing images is typically formulated as a semantic segmentation task, where each pixel in an image is assigned a road or non-road label [9]. According to the level of automation, existing approaches can be broadly categorized into heuristic methods and data-driven methods [10]. Heuristic methods typically rely on prior knowledge such as geometric, spectral, and contextual features of roads, thus limiting their adaptability to complex scenes and diverse imaging conditions [11]. Additionally, the lack of sufficient nonlinear expressivity limits how well deep semantic visual traits are identified [12]. In contrast, data-driven methods, particularly those based on deep learning, have become the dominant approaches for road extraction due to their ability to automatically learn hierarchical features and generalize across different scenes [13]. The Fully Convolutional Network (FCN) [14] introduced an end-to-end paradigm for semantic segmentation and has inspired a series of improved architectures in subsequent studies [15]. Building upon the FCN framework, various encoder–decoder architectures have been proposed, including UNet, SegNet, and DeepLabv3+, many of which have been successfully applied to road extraction tasks [16,17]. Among them, the widely used UNet model has inspired numerous variants aimed at enhancing segmentation performance. Typical strategies include customized loss functions [18], multi-scale contextual information fusion [19] and the incorporation of attention mechanisms [20], which have been shown to achieve higher accuracy in road extraction from high-resolution remote sensing imagery.

More recently, alternative network designs have been explored. For example, approaches based on the Generative Adversarial Network (GAN) have been introduced to road extraction tasks, where adversarial learning is used to encourage the predicted roads to better match the structural characteristics of real road networks [21]. In GAN-based frameworks, various architectural improvements have been explored, such as incorporating encoder–decoder structures into the generator or discriminator [22], integrating multi-scale feature fusion [23], and employing multiple discriminators [24] to enhance the quality and structural consistency of the generated road maps. Transformer-based architectures have also shown strong potential for road extraction, as self-attention mechanisms enable them to capture long-range dependencies and global contextual information more effectively. Numerous models that integrate the Convolutional Neural Network (CNN) and Transformer architectures have achieved promising performance in road extraction tasks [25,26]. In addition, transformer-based modules have also been widely incorporated into segmentation networks. For example, Hu et al. introduced MDTNet, where a multi-scale deformable transformer module is employed to capture richer feature representations [27].

Despite these advances, deep learning-based road extraction methods often focus primarily on pixel-level classification accuracy, while the structural properties of road networks receive relatively less attention. This challenge has motivated researchers to explore methods that explicitly consider the structural relationships among road segments. Graph-based approaches have thus attracted increasing attention, where the road network is represented as a graph structure to model the connectivity among road segments [28,29]. For example, RoadTracer formulates road extraction as an iterative graph construction process that incrementally traces the road network [30], while RoadCorrector converts segmentation results into road graphs and performs topology correction to improve road

connectivity [31]. Furthermore, some recent studies incorporate topology-aware constraints directly into deep neural networks to encourage structural consistency during training. Meanwhile, post-processing strategies have also been employed to refine segmentation outputs by leveraging structural cues such as skeleton representations or connectivity analysis. Compared with graph-based deep learning approaches, post-processing strategies are generally simpler to implement and can be easily integrated with existing segmentation models without modifying the network architecture. However, some existing post-processing methods rely on simple distance-based criteria to reconnect fragmented road segments, which often leads to unreliable or geometrically implausible connections when facing complex road patterns.

Based on these observations, this paper proposes a collaborative method that integrates deep learning-based semantic segmentation with graph-based post-processing. The method first employs a Pyramid Scene Parsing Network (PSPNet) to generate initial road probability maps, followed by a graph-based connectivity refinement method that systematically restores fragmented road segments through a multi-source cost function and a direction-aware path-finding algorithm. The primary contributions of this work are three-fold: (1) we design a multi-source cost-driven mechanism that integrates semantic probability and edge-aware geometric constraints to guide accurate topological reconstruction; (2) we develop a direction-aware Dijkstra search algorithm that incorporates endpoint tangent vectors as inertial weights to ensure the geometric smoothness of generated paths; and (3) we validate the method across diverse urban and rural datasets, demonstrating significant improvements in connectivity metrics and proving its practical potential for mapping and navigation.

2. Methodology

2.1. Baseline Semantic Segmentation Network

PSPNet, proposed by Zhao et al. in 2017 [32], is a semantic segmentation model that adopts a fully convolutional architecture. The overall structure of the network is illustrated in Figure 1. The core innovation of PSPNet lies in its Pyramid Pooling Module (PPM), highlighted by the box in Figure 1. The PPM applies four parallel adaptive average pooling operations with different grid sizes to the input feature map, allowing the model to capture contextual information at multiple spatial scales.

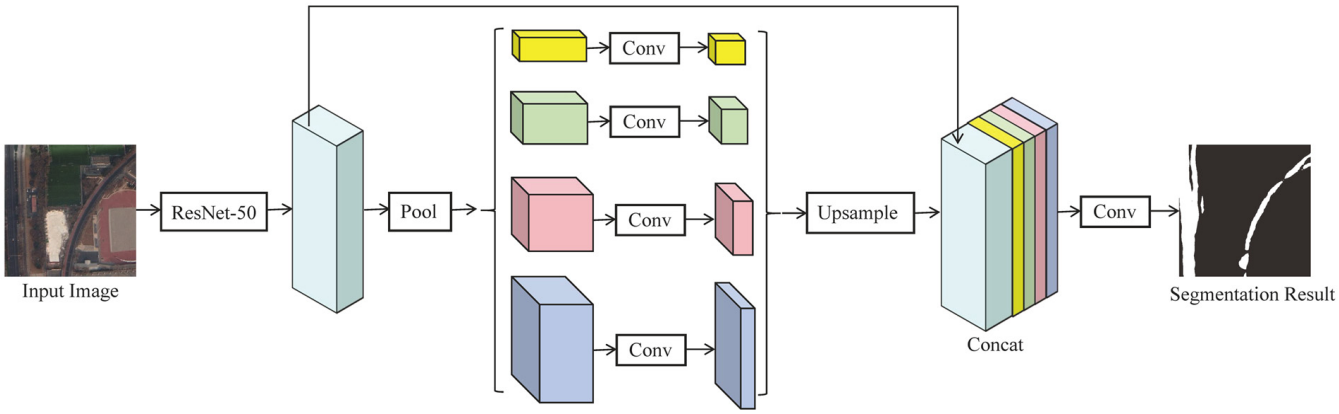


Figure 1. Structure of PSPNet.

Although more recent segmentation architectures based on Transformers or hybrid CNN–Transformer designs have been proposed, PSPNet remains a competitive choice due to its effective multi-scale context modeling and relatively simple architecture. Compared with many Transformer-based models that often require larger training datasets and

higher computational cost, PSPNet provides stable performance on high-resolution remote sensing imagery. Therefore, PSPNet is adopted in this paper to generate the initial road segmentation results for the subsequent topology refinement framework.

2.2. Graph-Based Connectivity Refinement Framework

While PSPNet is highly effective at capturing multi-scale contextual information, it mainly optimizes pixel-level classification and does not explicitly consider the topological continuity of road networks. As a result, the predicted road masks may contain fragmented structures, broken segments, or discontinuities, particularly in regions where roads are partially occluded by buildings, vegetation, or shadows. Such issues hinder the reconstruction of a coherent road network, which is essential for applications such as mapping and navigation. To address this problem, a graph-based connectivity refinement framework incorporating a multi-source cost function is proposed. By integrating semantic probability and edge-aware constraints, this framework refines initial segmentation results and reconstructs a road network with enhanced continuity and topological consistency. The overall workflow is illustrated in Figure 2.

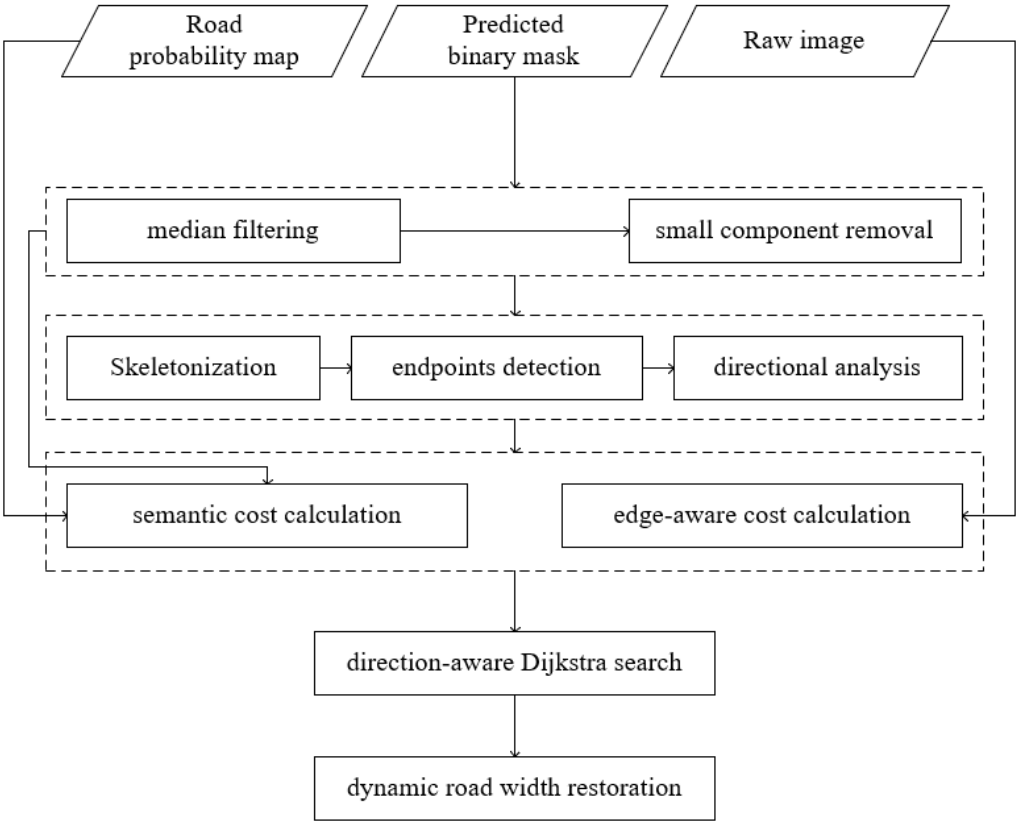


Figure 2. Workflow of the connectivity refinement framework.

(1) Noise removal and small component filtering

In the initial stage of the refinement process, the raw prediction masks generated by PSPNet are filtered to eliminate fragmentation and segmentation artifacts. A 3×3 median filtering is applied to the binary mask to smooth road boundaries and suppress isolated impulse noise while preserving the underlying structural integrity. Following this, small component removal is performed to filter out spurious detections. Specifically, a minimum area threshold is defined as 0.01% of the total image pixels. Connected components with an area smaller than this threshold are discarded. This step ensures that the subsequent skele-

tonization process focuses exclusively on primary road structures, effectively preventing the generation of false topological nodes caused by minor noise clusters.

(2) Topology extraction and directional analysis

Following initial filtering, the refined road mask is transformed into a one-pixel-wide centerline representation using morphological thinning. Based on this skeleton, the algorithm identifies endpoints, which are nodes with only a single neighbor in their 8-neighborhood. To ensure that newly generated paths follow the natural extension of the road network, a tangent direction vector is calculated for each identified endpoint by backtracking 12 pixels along the existing skeleton. This directional analysis allows the refinement process to maintain geometric continuity during the subsequent path-finding stage. Furthermore, to broaden the connectivity scope beyond simple endpoint-to-endpoint pairs, the algorithm establishes a comprehensive set of target candidates. This set includes other endpoints, boundary nodes, and all existing skeleton coordinates. This approach allows the refinement process to restore complex topological structures, such as T-junctions or roads extending to the image borders.

(3) Multi-source cost map construction

In this step, a comprehensive cost map is constructed to guide the path-finding process, ensuring that reconstructed connections are both geometrically plausible and semantically accurate. Unlike traditional methods that rely solely on Euclidean distance, the proposed framework employs a multi-source cost function. The total cost $C(i,j)$ for each pixel unit is computed as (1) to form a cost grid of the same dimensions as the input image.

$$C(i,j) = C_{sem}(i,j) + C_{edge}(i,j) \tag{1}$$

where $C(i,j)$ represents the total traversal cost at pixel coordinates (i,j), with i and j denoting the row and column indices respectively. The term $C_{sem}(i,j)$ refers to the semantic cost component derived from model predictions, while $C_{edge}(i,j)$ is the edge-aware penalty term used to block non-road obstacles.

The semantic cost $C_{sem}(i,j)$ forms the foundation of the map. This cost is calculated using (2):

$$C_{sem}(i,j) = \frac{\alpha}{P(i,j)^{\gamma} + \varepsilon} \tag{2}$$

where $P(i,j)$ denotes the road-class probability predicted by the deep learning model at pixel coordinates (i,j), and γ is the steepness exponent that controls the sensitivity of the cost relative to probability fluctuations. ε is a negligible smoothing constant introduced to prevent division by zero in non-road regions. To prioritize the utilization of existing segmentation results, α is set to 0.5 for pixels located within the initial prediction mask to lower the traversal cost, while a standard weight of 1.0 is maintained for pixels outside the mask. The inverse-power function is designed to provide low-resistance paths through high-probability regions while exponentially penalizing low-confidence areas. By adjusting γ for nonlinear mapping, the algorithm ensures that subtle gaps remain traversable, effectively guiding the repair of the road network from reliable segments.

The edge-aware cost $C_{edge}(i,j)$ is derived from the original input image by applying a Canny detector to the grayscale image, followed by a 3×3 dilation to expand the boundary influence. A constant penalty of 2.0 is assigned to these dilated edge regions, while other areas remain zero. The primary role of $C_{edge}(i,j)$ is to act as a geometric barrier. By imposing an extra traversal cost at physical boundaries, it prevents the path-finding algorithm from drifting into non-road obstacles, ensuring that the reconstructed segments remain aligned with the visual road margins.

(4) Graph-based topology refinement

Algorithm 1 outlines a systematic pipeline for restoring road network continuity by bridging fragmented segments. The process begins by prioritizing potential connections; a target set T is constructed to include endpoints, boundary nodes, and existing skeleton pixels, and candidate pairs are sorted in ascending order of Euclidean distance to ensure that nearby breakpoints are stabilized first. To ensure geometric smoothness and topological continuity, a direction-aware Dijkstra search is implemented (Algorithm 1, Lines 9–1), where the algorithm incorporates geometric inertia by calculating the cosine similarity $\cos\theta$ between each search step and the pre-defined orientation vector $\vec{v}_i$. Through the penalty operator ω, paths aligning with the road's heading receive a reward λ, while sharp deviations are heavily penalized, forcing the generated connections to maintain a smooth and collinear trajectory. Finally, each optimized path L_{opt} is validated against an average semantic probability threshold to filter out noise.

Algorithm 1: Graph-based Topology Refinement

Input: Skeleton map S, Endpoint set E, Orientation vectors V, Cost map C
Output: Refined road skeleton $S_{refined}$

Define target set $T = \{E \cup boundary\ nodes \cup S\}$;
Generate candidate pairs $P_{pairs} = \{(e_i, t_j)\}$ where $e_i \in E, t_j \in T$;
Sort P_{pairs} in ascending order of Euclidean distance;
for each $(e_i, t_j) \in P_{pairs}$ **do**
  if e_i *is already marked as connected* **then**
    continue;
  end
  Define a local search ROI centered around the midpoint of (e_i, t_j);
  Find path L_{opt} using Direction-aware Dijkstra on C:
  For each search step $\vec{v}_{step}$, calculate alignment $\cos\theta = \frac{\vec{v}_i \cdot \vec{v}_{step}}{\|\vec{v}_i\| \|\vec{v}_{step}\|}$;
  Update step cost: $Cost_{step} = Cost_{base} \times \omega(\cos\theta)$, where $\omega = \lambda$ (if $\cos\theta > 0.85$), 0.8 (if $\cos\theta > 0.6$), else 2.0;
  if *mean probability of pixels in* L_{opt} > *threshold* **then**
    Update skeleton $S \leftarrow S \cup L_{opt}$;
    Mark e_i as connected;
  end
end
return $S_{refined} \leftarrow S$;

(5) Dynamic road width restoration

To conclude the post-processing workflow, a dynamic road width recovery strategy is applied to transform the refined centerlines into physically consistent road regions. This adaptive approach utilizes a distance transform map derived from the initial binary mask to estimate the local road width at the connection anchors e_i and t_j. Specifically, the dilation radius is determined by taking the average of the distance transform values at the two anchor points and applying a scaling factor of 1.2 to facilitate a seamless transition. Finally, these paths are dilated using a disk-shaped structuring element and integrated into the global mask, restoring the natural geometry of the road network.

2.3. Evaluation Metrics

To comprehensively evaluate the performance of the proposed method, both conventional segmentation metrics and topology-aware connectivity metrics are adopted. The

conventional metrics measure pixel-level classification accuracy, while the connectivity-oriented metrics focus on the structural and morphological completeness of the predicted road networks.

Intersection over Union (IoU) and Mean IoU (MIoU): IoU evaluates the overlap between the predicted mask and the ground truth for each class. MIoU averages the IoU values over all classes to provide an overall segmentation score. The metrics are defined as (3) and (4) [33]:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{3}$$

$$\text{MIoU} = \frac{1}{N}\sum_{i=1}^{N} \text{IoU}_\text{i} \tag{4}$$

where TP, TN, FP and FN denote the numbers of true positive, true negative, false positive, and false negative pixels, respectively, and represents the number of classes.

Connectivity (Conn): To assess the impact of the post-processing step on the restoration of road connectivity, the Conn metric introduced by Wei et al. [34] was employed with specific modifications to enhance its numerical stability. This metric measures the local structural completeness and topological consistency of the extracted road networks. Specifically, both the predicted and ground truth road masks are skeletonized to generate one-pixel-wide centerlines, preserving the network topology. The total numbers of connected components in the ground truth and prediction are denoted as N_{gt} and N_{pred}, respectively. N_{gt}^{match} denotes the number of ground truth segments that are successfully overlapped by any predicted segment, and N_{pred}^{match} signifies the number of predicted segments that correctly overlap with any ground truth segment. This bidirectional matching mechanism ensures that the Conn value remains within the range of [0, 1]. The final connectivity score is calculated as (5):

$$\text{Conn} = \frac{N_{gt}^{match} + N_{pred}^{match}}{N_{gt} + N_{pred}} \tag{5}$$

Although Conn effectively measures the structural completeness of road networks, it relies on skeletonized representations and can be sensitive to small geometric deviations or noise in the predicted results. Moreover, the metric mainly evaluates whether road segments are connected, but does not explicitly assess whether the reconstructed connections are geometrically or topologically correct. As a result, incorrect connections may still yield relatively high connectivity scores. Therefore, several refinement steps are incorporated in the post-processing pipeline to mitigate these issues and obtain more stable and reliable evaluation results. In terms of evaluation, both MIoU and Conn are considered to provide a more comprehensive assessment of segmentation accuracy and network connectivity.

3. Experiments and Results

3.1. Datasets

Two publicly available road extraction datasets, CHN6-CUG and DeepGlobe, are employed in this paper to evaluate the effectiveness and generalization ability of the proposed method. Representative samples from the two datasets are illustrated in Figure 3.

The CHN6-CUG Road Dataset proposed by Zhu et al. [35] is a large-scale satellite dataset covering six representative Chinese cities. With a spatial resolution of 0.5 m/pixel, it contains dense and complex urban road networks, as illustrated in Figure 3a–d, providing a benchmark for road extraction in urban scenes. Since a portion of the original images contain little useful information (e.g., nearly blank scenes), the dataset was cleaned by removing unsuitable samples. The remaining images were randomly divided into

2488 training images, 311 validation images, and 312 test images, resulting in 3111 images in total.

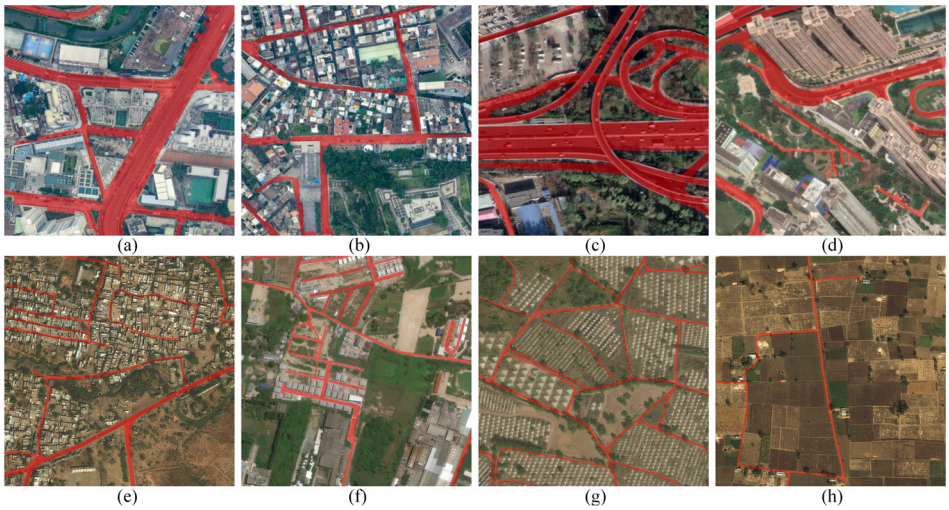


Figure 3. Example samples from the CHN6-CUG and DeepGlobe datasets with ground truth road annotations overlaid on the images. (**a**–**d**) CHN6-CUG dataset; (**e**–**h**) DeepGlobe dataset. The semi-transparent red regions indicate the ground truth road annotations overlaid on the original images.

The DeepGlobe Road Extraction Dataset is another widely used benchmark for road extraction from high-resolution satellite imagery [36]. Compared with CHN6-CUG, DeepGlobe contains images from diverse geographic regions with more heterogeneous landscapes, including urban and rural areas, as shown in Figure 3e–h. The road structures therefore exhibit larger variations in scale, continuity, and surrounding environments, posing additional challenges for road segmentation and connectivity preservation. The dataset consists of high-resolution satellite images with a spatial resolution of 0.5 m/pixel, and each image has a size of 1024 $\times$ 1024 pixels. In this paper, the 6226 annotated images were randomly re-partitioned into 4980 training images, 622 validation images, and 624 test images.

3.2. Implementation Details

The experimental environment was built with Python and PyTorch 1.10.0 with CUDA 11.3 support. The proposed model was implemented using the MMSegmentation framework. Model training, inference, and post-processing were performed on NVIDIA RTX 4090D GPUs (NVIDIA, Santa Clara, CA, USA). The SGD optimizer (momentum = 0.9, weight decay = 1×10^{-4}) was used with an initial learning rate of 0.005, adjusted dynamically by a Warmup PolyLR schedule (1500 warm-up iterations, decay power = 0.9). Mixed-precision (FP16) training was applied to enhance memory efficiency. The cross-entropy loss served as the main objective, optionally combined with an auxiliary branch for deep feature supervision. Model performance was evaluated on the validation set using MIoU, and the best weights were saved automatically.

3.3. Comparative Study of Segmentation Models

To select an optimal deep learning baseline for the subsequent post-processing experiments, a series of comparative experiments were conducted on the CHN6-CUG and DeepGlobe datasets. Quantitative comparisons against several representative segmentation models are performed, including SegFormer [37], Swin Transformer [38], ConvNet [39], and BiSeNet V2 [40], which are widely used architectures covering both Transformer-based and CNN-

based segmentation models. The results are summarized in Tables 1 and 2, while representative segmentation results are presented in Figures 4 and 5 for qualitative comparison.

Table 1. Quantitative segmentation results of models on the CHN6-CUG dataset.

Models	Conn	MIoU (%)	Road IoU (%)
SegFormer	0.4335	71.15	48.36
Swin Transformer	0.4503	62.85	33.08
ConvNet	0.4777	66.55	39.93
BiSeNet V2	0.5416	72.86	51.42
PSPNet	**0.5806**	**78.78**	**62.16**

Note: Bold values indicate the best performance.

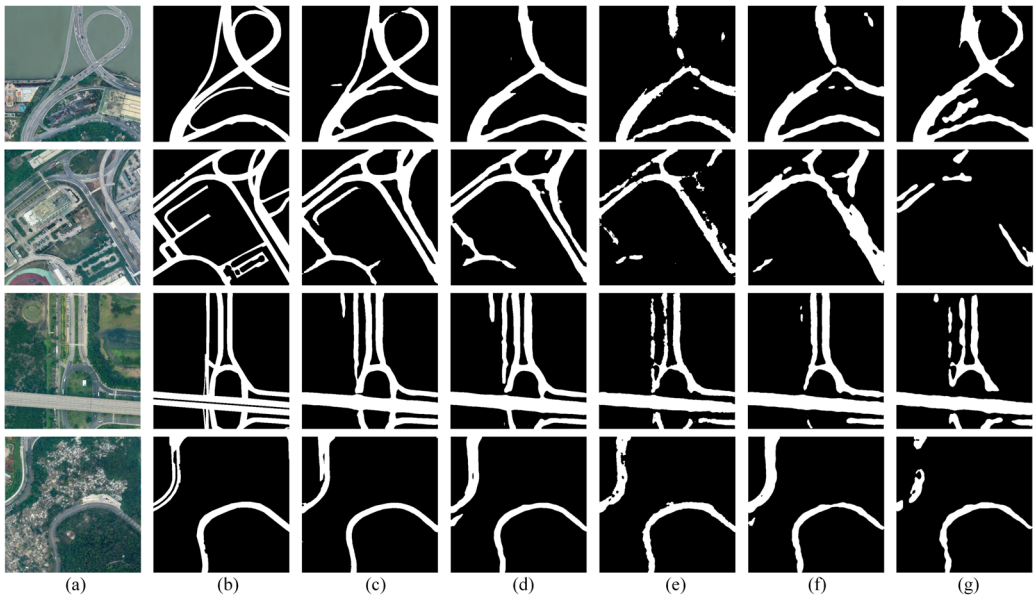


Figure 4. Visualized segmentation results of models on the CHN6-CUG dataset. (**a**) Image; (**b**) ground truth; (**c**) PSPNet; (**d**) BiSeNet V2; (**e**) SegFormer; (**f**) ConvNet; (**g**) Swin Transformer.

As shown in Table 1, PSPNet achieves the best performance among all of the compared models on the CHN6-CUG dataset, obtaining the highest scores on Road IoU (62.16%), MIoU (78.78%), and Conn (0.5806). These results indicate that PSPNet can accurately identify road regions while maintaining strong structural consistency in the extracted road networks. Among the comparison methods, BiSeNet V2 ranks second, with an MIoU of 72.86%, a Road IoU of 51.42%, and a Conn score of 0.5416, showing a significant performance gap compared to PSPNet across all metrics. Regarding the remaining three models, SegFormer exhibits the highest segmentation accuracy, performing closely to BiSeNet V2 and notably outperforming ConvNet and Swin Transformer. However, from the perspective of connectivity, ConvNet shows relatively better results compared to the other two models. This observation suggests that segmentation accuracy and structural connectivity do not always change consistently. Therefore, in road extraction tasks, segmentation accuracy alone is insufficient for practical applications, and the connectivity and integrity of the extracted road network should also be considered.

The qualitative results shown in Figure 4 further support the quantitative results, showing that the segmentation results produced by PSPNet exhibit the most complete and continuous road structures among all of the compared methods. Among the other models, BiSeNet V2 also achieves relatively good segmentation results overall. However, in the case of the ring-shaped multi-level highway interchange (first row in Figure 4), several road

segments are clearly missed by BiSeNet V2. The other three models exhibit more severe omissions and discontinuities in their predictions.

Table 2. Quantitative segmentation results of models on the DeepGlobe dataset.

Models	Conn	MIoU (%)	Road IoU (%)
SegFormer	0.4112	68.66	40.38
Swin Transformer	0.4553	70.90	44.50
ConvNet	0.4268	71.24	45.31
BiSeNet V2	0.4851	76.06	54.48
PSPNet	**0.5222**	**78.59**	**59.29**

Note: Bold values indicate the best performance.

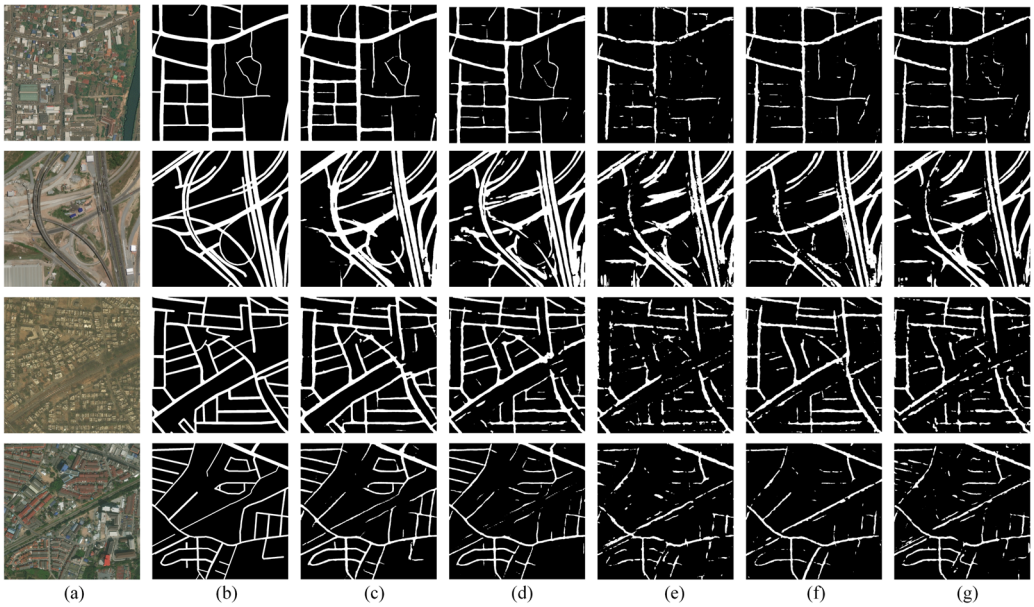


Figure 5. Visualized segmentation results of models on the DeepGlobe dataset. (**a**) Image; (**b**) ground truth; (**c**) PSPNet; (**d**) BiSeNet V2; (**e**) SegFormer; (**f**) ConvNet; (**g**) Swin Transformer.

On the DeepGlobe dataset, the overall performance of the models follows a trend similar to that observed on the CHN6-CUG dataset. As shown in Table 2, PSPNet still achieves the best performance, with a Road IoU of 59.29%, an MIoU of 78.59%, and a Conn score of 0.5222, outperforming the other compared models in terms of both segmentation accuracy and road connectivity. The consistent MIoU and Road IoU values observed across both datasets underscore the robust adaptability of PSPNet to different geographic scenarios. While BiSeNet V2 remains the second-best performing model overall on both datasets, its advantage in segmentation accuracy over the remaining baselines is significantly more pronounced on DeepGlobe than on CHN6-CUG. Notably, the overall Conn scores are slightly lower on DeepGlobe compared to CHN6-CUG for all models. One possible explanation is that the road networks in DeepGlobe images exhibit more complex spatial structures, including larger variations in road width and more complicated intersections. In addition, these images are more frequently affected by shadows, building occlusions, and background textures that resemble roads, which further increases the difficulty of maintaining continuous road structures.

This observation is further supported by the qualitative results presented in Figure 5. Although PSPNet still produces the best predictions among all compared models on the DeepGlobe dataset and can recover most of the main road structures, the predicted results show more noticeable local gaps, missing branches, and incomplete connections compared with those on CHN6-CUG, especially in scenes with dense road networks,

complex branches, or thin road segments. These issues break the continuity of the road skeleton and consequently lead to a reduction in the Conn metric. Similar phenomena can also be observed in the predictions of the other models, where missed detections and fragmented road structures appear more frequently.

Overall, the experimental results on both datasets demonstrate that PSPNet shows strong robustness across different road extraction datasets and consistently achieves superior performance on multiple evaluation metrics. Therefore, PSPNet is selected to generate the initial road segmentation results in this paper. However, when dealing with road scenes that have complex spatial distributions and diverse topological structures, relying solely on segmentation networks is still insufficient to fully guarantee the structural completeness of road networks. These findings further highlight the necessity of introducing a graph-based connectivity refinement post-processing strategy, which aims to repair broken road segments and enhance the connectivity of road networks based on the initial segmentation results.

3.4. Performance Evaluation of the Proposed Connectivity Refinement Framework

After the segmentation results were obtained from the PSPNet model, a connectivity-preserving post-processing procedure was applied to repair discontinuous or fragmented road segments. The post-processing pipeline consisted of several sequential steps, as described in Section 2.2. In the implementation of our proposed post-processing pipeline, several key parameters are involved across various stages, including noise filtering, cost map construction, path searching, and path validation. Specifically, parameters related to noise removal, such as the median filter size and the minimum area threshold for component filtering, as well as parameters for Canny edge detection, are set to default or empirical values based on the existing literature and visual inspections. Given their relatively straightforward roles in establishing a clean baseline and reasonable edge constraints without drastically altering the core topology optimization, we skip formal parameter sensitivity experiments for these settings, assuming stability within a reasonable range.

Instead, our analysis focuses on two pivotal hyperparameters that directly govern the effectiveness of our directional graph-based refinement: the power γ to which the semantic probability $P(i, j)$ is raised in the semantic cost calculation, and the coefficient λ used for the cosine-based dynamic directional compensation in the alignment constraint. Specifically, γ determines how heavily the cost function penalizes lower-probability regions; a higher power value creates a steeper cost landscape, forcing the path searching algorithm (Algorithm 1) to strictly follow high-confidence areas, potentially leading to more fragmented paths if predictions are noisy, while a lower value may yield smoother but less semantically accurate connections. Complementarily, λ controls the degree to which the collinearity constraint is enforced during path refinement. This parameter directly dictates how much the cost is adjusted to favor paths that align with the established skeleton orientation, thereby mitigating sharp turns and enhancing geometric continuity. Given that these two hyperparameters inherently define the trade-offs between model fidelity and topological regularity—the core objective of our connectivity-preserving strategy—performing sensitivity experiments is essential to identify the optimal parameter configuration that robustly reconnects fragmented roads without sacrificing their structural integrity. Consequently, 30 representative images were selected from each dataset based on their typical road scenes and reliable initial segmentation performance. On these samples, six candidate parameter configurations were sequentially applied to the post-processing pipeline. The resulting MIoU and Conn values are summarized in Table 3.

Table 3. Results of sensitivity experiments on key parameters in the connectivity refinement framework.

γ	λ	MIoU (CHN6-CUG)	Conn (CHN6-CUG)	MIoU (DeepGlobe)	Conn (DeepGlobe)
1.0	0.2	83.99	0.8568	77.52	0.8509
1.0	0.5	84.06	0.8568	77.63	0.8516
1.5	0.2	83.87	0.8476	77.53	0.8516
1.5	0.5	84.07	0.8513	77.65	0.8516
2.0	0.2	83.90	0.8513	77.66	0.8516
2.0	0.5	83.99	0.8596	77.64	0.8516

Note: The unit of MIoU in the table is percentage (%).

Overall, the results in Table 3 demonstrate that the proposed post-processing framework maintains highly stable performance across varying parameter settings on both datasets, with fluctuations in both MIoU and Conn values remaining within an exceptionally narrow margin. This stability underscores that the connectivity restoration is primarily driven by the inherent structural logic of the pipeline—namely the synergy between topology extraction and directional graph search—rather than being overly dependent on delicate hyperparameter tuning in our approach. The fact that the Conn metric remains identical or near-identical across the majority of parameter combinations on the DeepGlobe dataset further supports this conclusion, indicating that the framework is robust enough to identify the correct topological paths even under different penalty intensities.

Based on these experimental results, $\gamma = 1$ and $\lambda = 0.5$ are identified as the optimal configuration for the CHN6-CUG dataset, striking a balance between segmentation accuracy and topological connectivity while maximizing its potential for road network repair. Regarding the DeepGlobe dataset, $\gamma = 2$ and $\lambda = 0.2$ are selected. All subsequent quantitative and qualitative post-processing results presented for both datasets are based on these parameter settings.

Based on the segmentation results generated by PSPNet, the connectivity refinement method was applied on both the CHN6-CUG and DeepGlobe datasets. The resulting MIoU, Conn, and the average processing time per image are reported in Table 4. To further demonstrate the advantages of this post-processing framework over standalone deep learning models, the performance metrics for BiSeNet V2, which achieved the second-best results in Section 3.3, are also included for comparison.

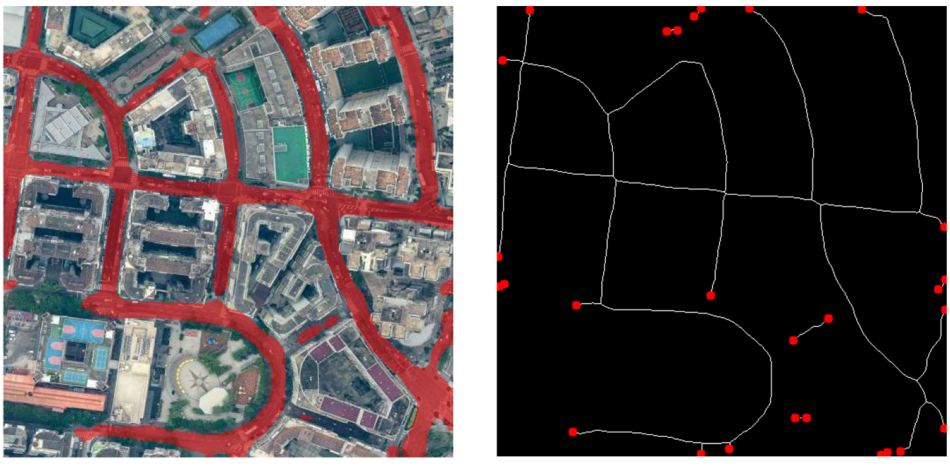
Table 4. Quantitative results and processing efficiency of the proposed connectivity refinement framework on the CHN6-CUG and DeepGlobe datasets.

Dataset	Method	MIoU (%)	Conn	Average Time (s/Image)
CHN6-CUG	BiSeNet V2	72.86	0.5416	3.80
	PSPNet	78.78	0.5806	
	PSPNet + post-processing	77.71	0.7795	
DeepGlobe	BiSeNet V2	76.06	0.4851	10.82
	PSPNet	78.59	0.5222	
	PSPNet + post-processing	78.14	0.8277	

As shown in Table 4, the connectivity refinement significantly improves the structural continuity of the extracted road networks on both datasets. On the CHN6-CUG dataset, the Conn metric increases from 0.5806 to 0.7795 (+0.1989) after the post-processing procedure, indicating that many fragmented road segments in the original PSPNet predictions are successfully reconnected. Meanwhile, the decline in MIoU is marginal, from 78.78% to 77.71% (−1.07%), which demonstrates that segmentation accuracy is largely preserved.

Notably, this refined MIoU still maintains a significant advantage over the standalone performance of BiSeNet V2 (72.86%). A similar trend is observed on the DeepGlobe dataset, where the Conn score rises sharply from 0.5222 to 0.8277 (+0.3055). The MIoU decreases only moderately from 78.59% to 78.14% (−0.45%), remaining substantially higher than that of BiSeNet V2 (76.06%). These results suggest that the proposed framework effectively prioritizes topological integrity with minimal cost to pixel-level accuracy, consistently outperforming secondary segmentation models. In addition, the average processing time of the post-processing pipeline is 3.80 s per image on CHN6-CUG and 10.82 s per image on DeepGlobe, indicating that the proposed method introduces an acceptable computational overhead while remaining practical for road extraction workflows. The increased processing time observed on the DeepGlobe dataset is primarily attributed to its larger image dimensions and the higher density of narrow, elongated road segments. These characteristics result in a more complex skeleton with a significantly higher number of candidate paths, thereby requiring more iterations during the graph-based refinement process. However, this increased computational effort is justified by the superior connectivity gains achieved; the framework effectively navigates this complexity to resolve more extensive fractures, leading to a more pronounced improvement in the Conn metric compared to other datasets.

Figure 6 illustrates the initial stage of the connectivity refinement framework, focusing on the identification of topological deficiencies within the raw model output. As observed, although the model captures the primary road areas, distinct discontinuities occur in regions characterized by tree shading or complex building shadows. These gaps result in a fragmented road network that fails to satisfy the requirements for practical navigation or routing applications. To address these discontinuities, the binary mask is first reduced to a one-pixel-wide skeleton, represented by the white lines in Figure 6b. Based on this representation, the framework automatically detects all dead-ends, which are marked as endpoints. These endpoints serve as the logical anchors for the subsequent connectivity restoration, precisely identifying the locations where the road network structure is interrupted.



(a) (b)

Figure 6. Initial road segmentation and topological feature extraction. (**a**) Initial overlay. (**b**) Skeleton and endpoints. In (**a**), the initial segmentation results are visualized as a semi-transparent red overlay on the original imagery. In (**b**), the detected endpoints are indicated by the red circles.

Figure 7 reveals the decision-making core of the framework, where a multi-source cost map is constructed to facilitate intelligent path searching. In Figure 7a, a comprehensive cost landscape is visualized to represent the difficulty of traversing different pixels. Areas with high road probability and low edge gradients form low-cost valleys (indicated in deep

blue), whereas non-road pixels and complex building structures are assigned significantly much higher traversal costs (indicated by warm colors). This cost map ensures that the subsequent path-searching process remains within the most plausible road regions. It is worth noting that the cost disparity between these two types of regions is quite substantial; this is primarily due to the mathematical form of the semantic cost function, which imposes exponential penalties on low-probability areas. However, the critical decisions where the connections are formed typically occur within these low-cost valleys where semantic cost differences are marginal. To provide finer guidance in these regions, our framework incorporates an edge cost to further constrain the paths along geometric boundaries. As observed in Figure 7b, the framework utilizes a Dijkstra-based searching algorithm to calculate optimal paths between disconnected endpoints. The visualization demonstrates that the generated paths (white lines) successfully bridge gaps by navigating through the low-cost valleys of the map. By incorporating directional constraints derived from the initial skeleton, the searching process ensures that the newly created connections maintain geometric smoothness and structural consistency with the existing road segments.

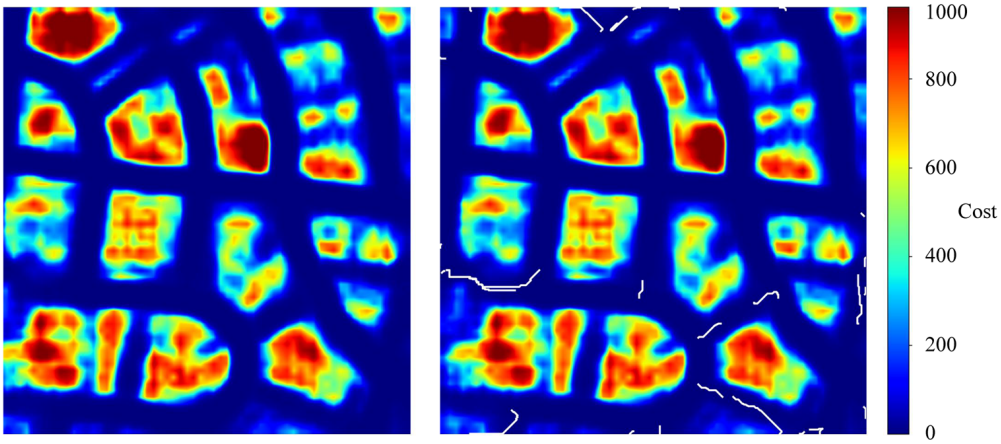


Figure 7. Cost map construction and direction-aware path searching. (**a**) Multi-source cost map. (**b**) Path searching.

The final results of the post-processing framework are displayed in Figure 8. Figure 8a presents the refined skeleton after the path-searching process. Compared to Figure 6b, the previously isolated endpoints are now connected, forming a continuous and closed topological network. Then, by restoring the road width to the refined skeleton, the refined road network is reconstructed. To provide a clear and intuitive demonstration of the post-processing framework's effectiveness, several key regions are highlighted in Figure 8b for comparison with the ground truth in Figure 8c. Specifically, yellow circles indicate successfully reconnected road segments where the algorithm correctly bridged existing gaps. Blue circles represent remaining discontinuities that were not successfully reconnected, while red circles point to incorrect connections that do not align with the actual road topology. Overall, the visualization demonstrates that the proposed algorithm achieves a substantial improvement in connectivity. A majority of the fragmented road sections have been effectively restored, while the occurrences of false or missed connections remain relatively low. These results confirm the robustness of the framework in repairing complex topological errors within extracted road networks.

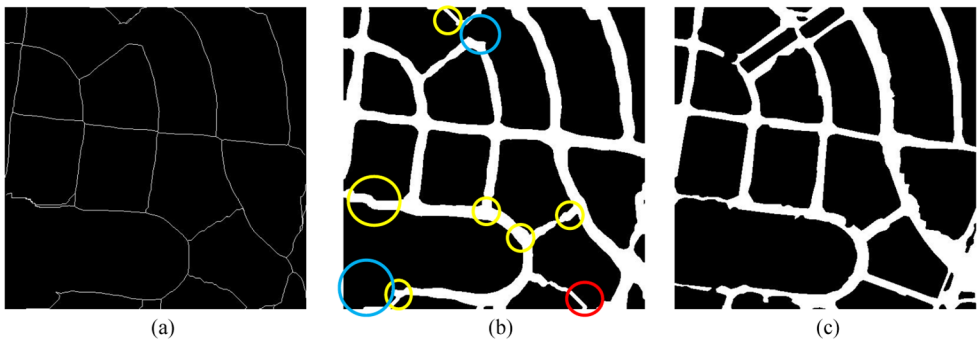


Figure 8. Topological refinement results and comparison with ground truth. (**a**) Refined skeleton. (**b**) Final result. (**c**) Ground truth. In (**b**), yellow circles indicate successfully reconnected road segments; blue circles represent remaining discontinuities, while red circles point to incorrect connections.

Figure 9 presents six representative examples selected from the CHN6-CUG dataset and the DeepGlobe dataset. By comparing the original images, predictions, refinement results and ground truth labels, Figure 9 illustrates both the effectiveness and the limitations of the proposed post-processing method. As shown in Figure 9b,f, the initial segmentation results produced by PSPNet contain several discontinuities or missing road segments, particularly along narrow roads or in areas where roads are partially occluded by surrounding objects. After applying the proposed connectivity refinement procedure, as shown in Figure 9c,g, several previously disconnected road segments are successfully reconnected, leading to an improvement in the overall continuity of the road network.

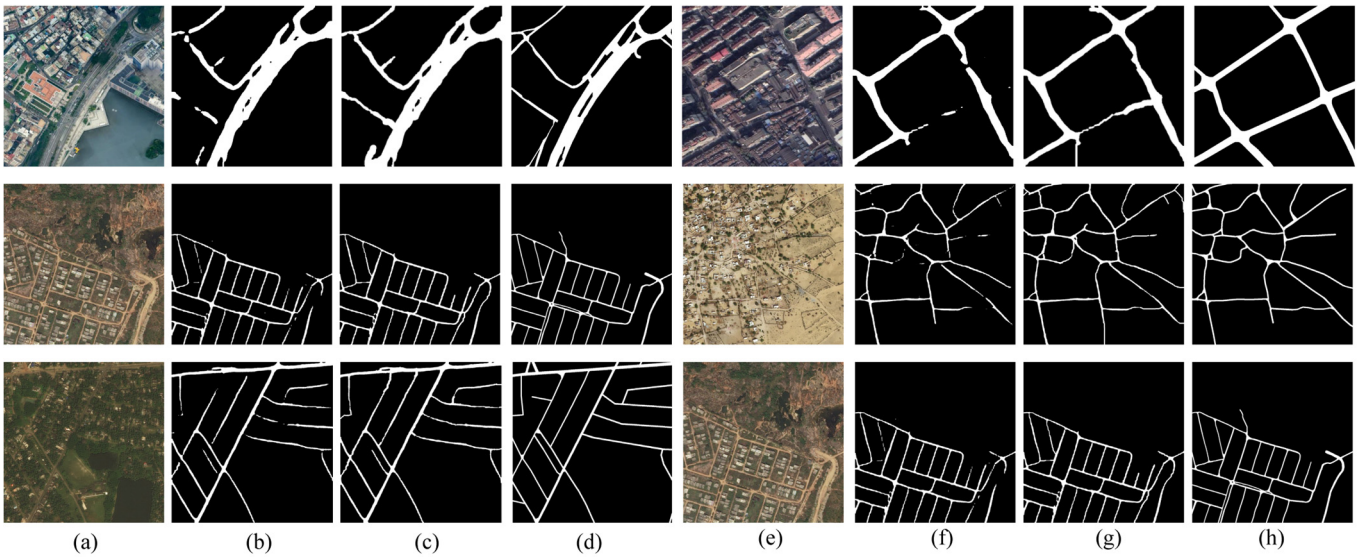


Figure 9. Examples demonstrating the effectiveness of the proposed connectivity refinement framework on two datasets. (**a**,**e**) image; (**b**,**f**) prediction; (**c**,**g**) refinement result; (**d**,**h**) ground truth.

Meanwhile, some instances of false or incorrect connections are observed, typically occurring when the original prediction already contained inaccuracies that misled the algorithm into misinterpreting the road structure. In addition, real-world road networks often possess irregular geometries, exhibiting varying widths, curved shapes, and complex intersections. Such characteristics present challenges for a purely topology-based refinement strategy to perfectly reconstruct the road network in every situation. Consequently, while the proposed post-processing method effectively improves connectivity in many cases, its performance remains dependent, to some extent, on the quality of the initial segmentation results.

4. Discussions

4.1. Analysis of Segmentation Models

Although several recently proposed segmentation architectures, including SegFormer, Swin Transformer, ConvNet, and BiSeNet V2, have demonstrated strong performance in various semantic segmentation benchmarks, the experimental results in this study show that PSPNet achieves comparatively better performance for road extraction on both the CHN6-CUG and DeepGlobe datasets. This phenomenon can be attributed to several task-related factors. Similar observations have also been reported in recent road extraction studies, where CNN-based architectures or their improved variants still outperform Transformer-based segmentation models in certain scenarios [9,41].

First, PSPNet is well suited for extracting elongated structures such as roads. Its PPM aggregates contextual information at multiple spatial scales, which helps capture long and continuous road patterns. Although Transformer-based models can also model long-range dependencies through self-attention, the datasets used in this paper consist of cropped image tiles, which limits the available spatial context and may reduce the advantage of global attention. Second, many Transformer-based segmentation models are designed for complex multi-class semantic segmentation tasks. In contrast, this study focuses on a binary road extraction problem, where all other land-cover types are treated as background. Under such conditions, convolution-based architectures with strong structural inductive biases may still perform competitively. Third, the scale of the training data may also affect model performance. Transformer-based models typically benefit from very large training datasets, whereas the datasets used in this study are relatively limited in size. As a result, the advantages of these architectures may not be fully realized.

4.2. Analysis of the Graph-Based Connectivity Refinement Framework

The experimental results validate that the proposed graph-based refinement framework significantly mitigates the fragmentation issues inherent in pixel-level segmentation. Unlike conventional morphological operations, the core strength of our approach lies in the integration of multi-source cost modeling and orientation-aware path-finding.

Despite these improvements, several limitations persist in complex topological scenarios. One primary issue is proximity-induced misconnection, particularly in regions with dense, parallel road structures where the cost function may lack sufficient discriminative power. Additionally, the framework faces a challenge in distinguishing between accidental fractures and intentional dead-ends (e.g., cul-de-sacs). Furthermore, the effectiveness of the post-processing strategy, particularly the road width restoration, remains highly dependent on the quality of the initial segmentation. If the initial binary mask is uneven or significantly narrowed due to severe occlusions, the distance transform-based width estimation may produce inconsistent results, leading to geometric distortions in the recovered segments. Finally, achieving robust generalizability across diverse scenes with varying spatial resolutions remains a challenge.

To address these limitations, several promising directions for future research are identified. First, higher-order topological constraints such as curvature continuity could be integrated into the graph construction to better distinguish between intersecting and non-intersecting segments. Second, to reduce the dependency on initial segmentation quality, instead of relying on purely heuristic cost functions and post hoc width estimation, incorporating learnable connectivity modules is a promising path. By training the model to predict link probabilities and geometric parameters directly from both local image patches and global context, the framework could potentially compensate for initial segmentation errors and achieve more stable, physically consistent road reconstruction.

4.3. Application Potential in GIS Tasks

From a geospatial perspective, the value of connectivity preservation lies not only in improving the visual completeness of extracted roads, but also in enhancing their functional usability in downstream spatial analyses. In many Geographic Information System (GIS) workflows, road data are not used as raster masks but as network structures that support topology-aware operations, such as connectivity queries, shortest-path analysis, and transportation network modeling. To further explore the potential of the proposed connectivity refinement method in supporting GIS tasks, a simple simulated routing experiment is conducted. As illustrated in Figure 10, starting and destination points are selected within the road skeleton network to analyze the routing performance before and after the refinement process. The results highlight improvement in path reachability. In Figure 10a, gaps from shadows or occlusions often cause routing algorithms to fail when a rupture exists between coordinates. Furthermore, the refinement optimizes path quality. Even when a redundant connection exists in the raw output, the fragmented network often forces circuitous routing. In Figure 10b, our framework enables the planning of shorter, more geographically logical paths. These findings confirm that the proposed restoration not only resolves road discontinuity but also enhances the reliability of extracted networks for real-world applications such as logistics and traffic simulation.

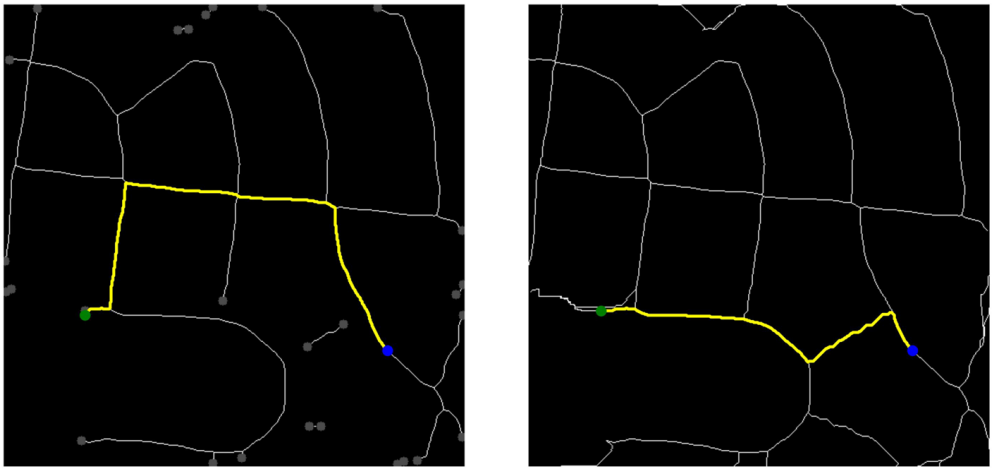


Figure 10. Comparison of routing performance between (**a**) the initial skeleton and (**b**) the refined skeleton. The green and blue dots represent the starting and ending points, respectively, and the yellow lines indicate the generated routes.

5. Conclusions

In this paper, we present a collaborative method that integrates deep learning-based semantic segmentation with graph-based post-processing. Our approach effectively bridges the gap between pixel-level semantic segmentation and topological consistency through the design of a multi-source cost function and an direction-aware Dijkstra search algorithm. Experimental results across urban and rural datasets demonstrate that the proposed method significantly enhances the structural continuity of road networks, as evidenced by substantial improvements in the Conn and MIoU metrics. Specifically, on the CHN6-CUG dataset, the Conn metric increased by 0.1989 with a marginal MIoU decrease of only 1.07%; similarly, on the DeepGlobe dataset, the Conn improved by 0.3055 while the MIoU dropped by a negligible 0.45%. These results indicate that our refinement strategy achieves a boost in topological connectivity with a minimal trade-off in pixel-level accuracy. Overall, the proposed method produces physically consistent road entities ready for GIS and navigation applications. Future work will focus on improving the system's adaptability through learn-

able connectivity modules and multi-modal data integration, ensuring robust performance in large-scale and complex geospatial environments.

Author Contributions: Conceptualization, Zixuan Teng and Zezhong Zheng; methodology, Zixuan Teng; software, Zixuan Teng; validation, Zixuan Teng, Zezhong Zheng and Hao Xue; formal analysis, Zixuan Teng; investigation, Zixuan Teng; resources, Xiangyang Sun; data curation, Zixuan Teng; writing—original draft preparation, Zixuan Teng; writing—review and editing, Zixuan Teng, Zezhong Zheng, Xiangyang Sun and Hao Xue; visualization, Zixuan Teng; supervision, Zezhong Zheng and Xiangyang Sun; project administration, Zezhong Zheng and Xiangyang Sun. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by Huzhou Municipal Science and Technology Bureau (Grant No. 2024GZ54) and in part by the Natural Science Foundation of Sichuan Province (Grant No. 2026NSFSC0219).

Data Availability Statement: The CHN6-CUG Road Dataset and the DeepGlobe Road Extraction Dataset used in this paper are both publicly accessible. The CHN6-CUG Road Dataset was developed by Qiqi Zhu and colleagues at the China University of Geosciences and can be obtained from the official project webpage at: https://grzy.cug.edu.cn/zhuqiqi/zh_CN/yjgk/32368/content/1734.htm (accessed on 1 October 2025). The DeepGlobe Road Extraction Dataset can be accessed through the DeepGlobe challenge website: https://deepglobe.org/challenge.html (accessed on 10 January 2026).

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

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
PSPNet	Pyramid Scene Parsing Network
MIoU	Mean Intersection over Union
IoU	Intersection over Union
UAV	Unmanned Aerial Vehicle
FCN	Fully Convolutional Network
GAN	Generative Adversarial Network
PPM	Pyramid Pooling Module
GIS	Geographic Information System

References

1. Qi, H.; Shi, J.; Chen, J.; Chi, C.; Shan, H. Research on the Complete Design, Construction and Management of Urban Road in Dalian City under the Concept of "People-Oriented Traffic". In Proceedings of the 2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT), Online, 15–17 May 2020; pp. 457–460.
2. Qian, D.; Wang, Y.; Zhang, X.; Zhao, D. Rationality evaluation of urban road network plan based on the ew-topsis method. In Proceedings of the 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Beihai, China, 16–17 January 2021; pp. 840–844.
3. Ait Ouallane, A.; Bahnasse, A.; Bakali, A.; Talea, M. Overview of road traffic management solutions based on IoT and AI. *Procedia Comput. Sci.* **2022**, *198*, 518–523. [CrossRef]
4. Bisio, I.; Garibotto, C.; Haleem, H.; Lavagetto, F.; Sciarrone, A. A systematic review of drone based road traffic monitoring system. *IEEE Access* **2022**, *10*, 101537–101555. [CrossRef]
5. Liu, K.; Zhai, C.; Dong, Y.; Meng, X. Post-earthquake functionality assessment of urban road network considering emergency response. *J. Earthq. Eng.* **2023**, *27*, 2406–2431. [CrossRef]
6. Yunus, S.; Abdulkarim, I.A. Road traffic crashes and emergency response optimization: A geo-spatial analysis using closest facility and location-allocation methods. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1535–1555. [CrossRef]
7. Dritsas, E.; Trigka, M. Remote sensing and geospatial analysis in the big data era: A survey. *Remote Sens.* **2025**, *17*, 550. [CrossRef]
8. Xie, Y.; Zhan, N.; Zhu, J.; Xu, B.; Chen, H.; Mao, W.; Luo, X.; Hu, Y. Landslide extraction from aerial imagery considering context association characteristics. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103950. [CrossRef]

9. Yang, Z.; Zhang, W.; Li, Q.; Ni, W.; Wu, J.; Wang, Q. C^2net: Road extraction via context perception and cross spatial-scale feature interaction. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5647011. [CrossRef]
10. 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]
11. Wang, W.; Yang, N.; Zhang, Y.; Wang, F.; Cao, T.; Eklund, P. A review of road extraction from remote sensing images. *J. Traffic Transp. Eng. (Engl. Ed.)* **2016**, *3*, 271–282. [CrossRef]
12. Xie, Y.; Liu, S.; Chen, H.; Cao, S.; Zhang, H.; Feng, D.; Wan, Q.; Zhu, J.; Zhu, Q. Localization, balance, and affinity: A stronger multifaceted collaborative salient object detector in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *63*, 4700117. [CrossRef]
13. Liu, P.; Gao, X.; Shi, C.; Lu, Y.; Bai, L.; Fan, Y.; Xing, Y.; Qian, Y. CGCNet: Road extraction from remote sensing image with compact global context-aware. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5638312. [CrossRef]
14. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
15. Pan, D.; Zhang, M.; Zhang, B. A generic FCN-based approach for the road-network extraction from VHR remote sensing images–using openstreetmap as benchmarks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2662–2673. [CrossRef]
16. 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.* **2023**, *17*, 1126–1136. [CrossRef]
17. Zhang, D.; Yang, Y.; Qu, F.; Liu, Y. Road extraction from remote sensing images based on improved Deeplabv3+ network. In Proceedings of the 2024 4th International Conference on Computer Science and Blockchain (CCSB), Shenzhen, China, 6–8 September 2024; pp. 446–449.
18. Hu, J.; Li, Q.; Wang, Q. Dualstrip-net: A strip-based unified framework for weakly-and semi-supervised road segmentation from satellite images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5617514. [CrossRef]
19. Shao, S.; Xiao, L.; Lin, L.; Ren, C.; Tian, J. Road extraction convolutional neural network with embedded attention mechanism for remote sensing imagery. *Remote Sens.* **2022**, *14*, 2061. [CrossRef]
20. Qu, S.; Liu, G.; Zhang, X.; Liu, Y. Heterogeneous dual-decoder network for road extraction in remote sensing images. *Sci. Rep.* **2025**, *15*, 31619. [CrossRef]
21. 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]
22. Abdollahi, A.; Pradhan, B.; Sharma, G.; Maulud, K.N.A.; Alamri, A. Improving road semantic segmentation using generative adversarial network. *IEEE Access* **2021**, *9*, 64381–64392. [CrossRef]
23. Lin, S.; Yao, X.; Liu, X.; Wang, S.; Chen, H.-M.; Ding, L.; Zhang, J.; Chen, G.; Mei, Q. MS-AGAN: Road extraction via multi-scale information fusion and asymmetric generative adversarial networks from high-resolution remote sensing images under complex backgrounds. *Remote Sens.* **2023**, *15*, 3367. [CrossRef]
24. 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]
25. Wang, R.; Cai, M.; Xia, Z.; Zhou, Z. Remote sensing image road segmentation method integrating CNN-Transformer and UNet. *IEEE Access* **2023**, *11*, 144446–144455. [CrossRef]
26. 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]
27. Hu, P.-C.; Chen, S.-B.; Huang, L.-L.; Wang, G.-Z.; Tang, J.; Luo, B. Road extraction by multiscale deformable transformer from remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 2503905. [CrossRef]
28. He, S.; Bastani, F.; Jagwani, S.; Alizadeh, M.; Balakrishnan, H.; Chawla, S.; Elshrif, M.M.; Madden, S.; Sadeghi, M.A. Sat2graph: Road graph extraction through graph-tensor encoding. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 51–67.
29. Zao, Y.; Zou, Z.; Shi, Z. Road graph extraction via transformer and topological representation. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 2502205. [CrossRef]
30. Bastani, F.; He, S.; Abbar, S.; Alizadeh, M.; Balakrishnan, H.; Chawla, S.; Madden, S.; DeWitt, D. Roadtracer: Automatic extraction of road networks from aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4720–4728.
31. Li, J.; He, J.; Li, W.; Chen, J.; Yu, J. RoadCorrector: A structure-aware road extraction method for road connectivity and topology correction. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5616018. [CrossRef]
32. 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.
33. Wang, J.; Chen, T.; Zheng, L.; Tie, J.; Zhang, Y.; Chen, P.; Luo, Z.; Song, Q. A multi-scale remote sensing semantic segmentation model with boundary enhancement based on UNetFormer. *Sci. Rep.* **2025**, *15*, 14737. [CrossRef] [PubMed]

34. Wei, Y.; Zhang, K.; Ji, S. Simultaneous road surface and centerline extraction from large-scale remote sensing images using CNN-based segmentation and tracing. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8919–8931. [CrossRef]
35. Zhu, Q.; Zhang, Y.; Wang, L.; Zhong, Y.; Guan, Q.; Lu, X.; Zhang, L.; Li, D. A global context-aware and batch-independent network for road extraction from VHR satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 353–365. [CrossRef]
36. 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, 18–22 June 2018; pp. 172–181.
37. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
38. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
39. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
40. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [CrossRef]
41. Zong, J.; Sun, Y.; Wang, R.; Xu, D.; Yang, X.; Zhao, X. PWFNet: Pyramidal Wavelet–Frequency Attention Network for Road Extraction. *Remote Sens.* **2025**, *17*, 2895. [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.