sensors

MDPI

Review

A Survey of Deep Learning Road Extraction Algorithms Using High-Resolution Remote Sensing Images

Shaoyi Mo [1], Yufeng Shi [1,*], Qi Yuan [1] and Mingyue Li [2]

[1] College of Civil Engineering, Nanjing Forestry University, Nanjing 210047, China; moshaoyi@njfu.edu.cn (S.M.); yq@njfu.edu.cn (Q.Y.)
[2] School of Foreign Studies, Nanjing Forestry University, Nanjing 210047, China; mylee@njfu.edu.cn
* Correspondence: yfshi@njfu.edu.cn

Abstract: Roads are the fundamental elements of transportation, connecting cities and rural areas, as well as people's lives and work. They play a significant role in various areas such as map updates, economic development, tourism, and disaster management. The automatic extraction of road features from high-resolution remote sensing images has always been a hot and challenging topic in the field of remote sensing, and deep learning network models are widely used to extract roads from remote sensing images in recent years. In light of this, this paper systematically reviews and summarizes the deep-learning-based techniques for automatic road extraction from high-resolution remote sensing images. It reviews the application of deep learning network models in road extraction tasks and classifies these models into fully supervised learning, semi-supervised learning, and weakly supervised learning based on their use of labels. Finally, a summary and outlook of the current development of deep learning techniques in road extraction are provided.

Keywords: road extraction; high-resolution remote sensing images; deep learning; supervised learning; network model

Citation: Mo, S.; Shi, Y.; Yuan, Q.; Li, M. A Survey of Deep Learning Road Extraction Algorithms Using High-Resolution Remote Sensing Images. *Sensors* **2024**, *24*, 1708. https://doi.org/10.3390/s24051708

Academic Editor: Yun Zhang

Received: 15 January 2024
Revised: 26 February 2024
Accepted: 4 March 2024
Published: 6 March 2024

1. Introduction

There are various types of roads in remote sensing images, such as urban roads, suburban roads, mountain roads, expressways, overpasses, etc. As the resolution of remote sensing images continues to improve, high-resolution images contain more information about the texture, shape, structure, and neighborhood relationships of roads compared to low- and medium-resolution remote sensing images, enabling more accurate road information extraction [1]. Extracting road information from high-quality remote sensing images has always been challenging due to multiple factors. These include complex and cluttered backgrounds (such as buildings, vegetation, and various road types), diverse road shapes (which vary in width and length), and poor image perspectives (resulting from occlusions by clouds and fog, as well as lighting effects). Furthermore, as urban areas expand, the topological structure of roads becomes exceptionally complex, with numerous buildings obstructing large portions of road areas [2].

Road extraction is typically regarded as a semantic segmentation task, where road and non-road labels are assigned to all pixels in an image, achieving binary semantic segmentation. With the rapid advancement of deep learning, there has been widespread interest in its powerful data fitting and information processing capabilities. Previous reviews have focused on the progress of road extraction techniques in remote sensing images. They summarize both traditional and deep learning methods. For instance, Abdollahi et al. [3] summarized road extraction methods in remote sensing imagery as being based on deep learning techniques, such as DCNN [4], FCN [5], deconvolution [6], and GANs [7]. Lian et al. [8] further categorized extraction methods into heuristic and data-driven road extraction approaches. Heuristic methods predominantly employ semi-automatic or fully automatic

traditional techniques for road extraction, such as snake model-based contour extraction [9], geodesic path-based approaches [10], dynamic programming-based methods [11], and template matching [12]. Automated extraction methods include machine learning segmentation algorithms like SVM [13], K-Means [14], and Bayesian classifiers [15], edge analysis-based methods [16], and map-based techniques [17]. The data-driven module, based on [3], also adds a summary of graph-based methods [18]. Jia et al. [19] discussed the applications of active and passive remote sensing technologies in road extraction, including high-resolution, hyperspectral, synthetic aperture radar (SAR), and airborne laser scanning (ALS) technologies, and also provided a summary of the current state and future prospects of multi-source data fusion. Liu et al. [20] summarized previous data-driven methods as fully supervised learning methods and introduced weakly supervised and unsupervised learning methods. Currently, mainstream road extraction network models can be broadly categorized into fully supervised and semi-supervised (weakly supervised) extraction. The differentiation between these two learning methods primarily depends on whether the model requires substantial label data support during training. Fully supervised learning relies on a large number of pixel-level training labels for model training. This approach often achieves high-precision segmentation structures, but its generalization capability is relatively weak, resulting in limited segmentation performance in unknown scenarios. Moreover, obtaining pixel-level labels often requires a significant amount of manual annotation work, and these annotated data exhibit a high degree of subjectivity, potentially impacting the accuracy of road segmentation by the model. Semi-supervised (weak) learning relies on fewer training label data, which can be in the form of points, lines, and other weak labels for model training. While semi-supervised (weak) learning generally lags behind in segmentation performance compared to fully supervised learning, it offers certain advantages. This approach reduces the dependency on label data, thus alleviating the burden of manual annotation.

To address issues of insufficient labels and high annotation costs in road extraction tasks https://www.isprs.org/education/benchmarks/UrbanSemLab/ (accessed on 2 March 2024), this paper classifies network models based on the use of pixel-level labels, including fully supervised learning, semi-supervised learning, and weakly supervised learning. In this paper, “road extraction”, “deep learning”, and “remote sensing” were chosen as searching keywords. The Web of Science (WOS) and Google Scholar databases were used as literature search tools to primarily retrieve relevant literature from 2020 to 2023. We organized the publicly available datasets mentioned in the retrieved literature over 40 datasets (2013–2023). This compilation includes 22 publicly accessible road datasets, with images primarily sourced from Google Earth, OpenStreetMap (OSM), open APIs, drone imagery, and satellite imagery, covering urban, suburban, rural, and forested areas. Furthermore, we observed that multiple publicly available road datasets such as Massachusetts [21], ISPRS[1], CasNet [22], DeepGlobe [23], SpaceNet [24], Roadtracer [25], Ottawa [26], and CHN6-CUG [27] were utilized two or more times between 2020 and 2023, as depicted in Figure 1. In Figure 1, the leftmost column represents the number of times datasets were used during these four years, while the rightmost column indicates the number of times corresponding network models utilized the datasets. Additionally, we conducted research on pre-processing and post-processing work related to remote sensing images in the relevant literature. For instance, a real-time multi-temporal color data enhancement technique was introduced for improving Sentinel-1 multi-polarization and Sentinel-2 multi-spectral imagery datasets [28]. Image quality was enhanced through the application of the contrast-limited adaptive histogram equalization (CLAHE) algorithm to mitigate mountain shadow issues [29]. Post-processing tasks included road vectorization [30], road information, and label reconstruction [31], among others. Due to space constraints, this paper primarily focuses on the analysis and discussion of road feature extraction research based on fully supervised deep learning network models. The structure of this paper is as follows: Section 1 introduces and briefly elucidates the challenges and methods in the field of road extraction from remote sensing images. Section 2 delves

into road feature extraction using fully supervised deep learning network models while studying the strengths and limitations of these network models. Section 3 explores road feature extraction through semi-supervised (weak) deep learning. Section 4 presents a comprehensive review of road extraction methodologies, conducting a comparative analysis of diverse models in terms of their performance. Ultimately, we objectively discuss the limitations inherent in current supervised learning models. Section 5 put forwards future prospects of road extraction and challenges.

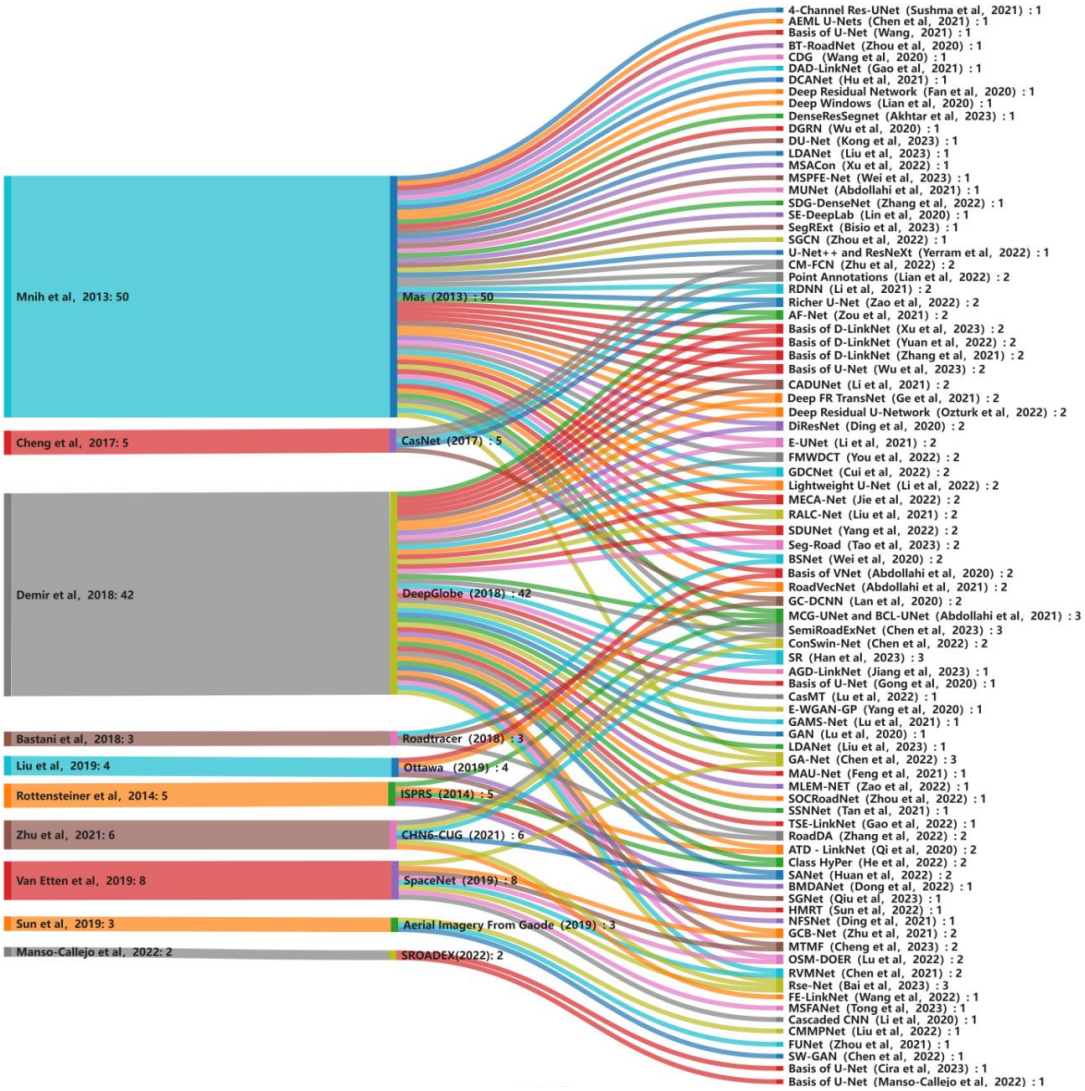

Figure 1. Public datasets used more than twice from 2020–2023.

2. Road Feature Extraction Based on Fully Supervised Deep Learning Network Models

Mnih [32] first introduced convolutional neural networks (CNNs) into road extraction tasks. Initially, in the field of deep learning for road extraction, many researchers used block-based CNN models to process roads within images. For example, finite state machine (FSM) and patch-based CNN (as shown in Figure 2) methods were employed [33] to track and extract roads separately. These patch-based CNN models performed excellently in aerial images with a spatial resolution of 1.2 m but struggled to achieve satisfactory results in higher-resolution (0.15 m) image extraction. To address this issue, Rezaee and Zhang [34] improved traditional patch-based CNN methods, enabling them to outperform support vector machine (SVM) methods in road extraction from high-resolution image datasets (0.15 m spatial resolution). However, patch-based CNN methods overly relied on the sliding window approach, which involved feature extraction through convolutional and pooling

layers, followed by backpropagation to fine-tune the final parameters. This resulted in relatively low extraction efficiency, which was insufficient for meeting the requirements of practical applications. Additionally, choosing an appropriate sliding window size was a challenging task. It was not until the emergence of fully convolutional neural networks (FCNs), that this problem was effectively solved. The FCN model was first introduced into the field of image segmentation [35], as shown in Figure 3, and it significantly improved segmentation efficiency. In contrast to traditional patch-based CNN models, an FCN is capable of pixel-level image classification, meaning it classifies each pixel into a category, with the output providing the category for each pixel. The FCN replaces fully connected layers with convolutional layers, achieving end-to-end semantic segmentation. This overcomes the inefficiency issue of patch-based CNN methods and allows for the extraction of target semantic information while preserving spatial information [1]. While the FCN enhanced the CNN by enabling pixel-to-pixel classification, it disregarded the relationships between pixels. Therefore, subsequent models introduced various attention mechanism modules to strengthen the relationships between pixels. Furthermore, the FCN's structure has offered novel insights into encoder–decoder network architectures.

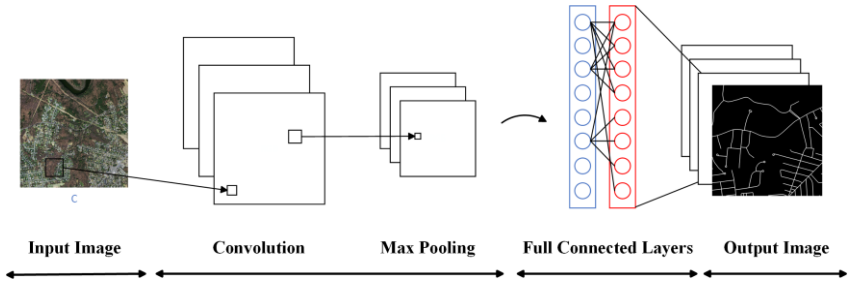

Figure 2. Patch-based CNN model.

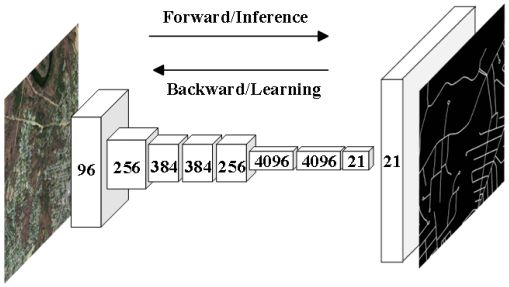

Figure 3. Fully Convolutional Neural Network (FCN) model.

2.1. Road Feature Extraction Based on Encoder–Decoder Structure

Following the FCN, network structures based on encoders and decoders have emerged and been widely applied. Their operation involves multiple downsampling of the original image by the encoder to obtain multi-level image feature information, followed by upsampling through the decoder to restore spatial information (Figure 4). Models based on this structure include SegNet [36], U-Net [37], PSPNet [38], LinkNet [39], DeepLab V3+ [40], and more. Among them, U-Net is one of the most classic networks with a symmetrical U-shaped encoder–decoder structure, initially applied in medical image segmentation tasks. This model employs an encoder–decoder structure for multi-scale feature fusion and pixel-level classification, while utilizing skip connections to acquire spatial information from the encoder and achieve feature fusion. The U-Net was extended by Chen et al. [41] to propose the Reconstruction Bias U-Net network. They added the ReLU function and a maxpooling layer and introduced decoding branches in the decoder to capture multiple semantic information from various upsampling processes. At present, there is a profusion of road extraction models based on encoder–decoder structures, encompassing models like LinkNet, D-LinkNet [42], U-Net and its variants VNet [43], U-Net++ [44], U^2-Net [45],

Dense-UNet [46], Res-UNet [47], MC-UNet [48], and others. While their structures exhibit slight variations, the primary distinctions lie in the encoder and decoder backbone models, intermediate layers, skip connection layers, and network model optimizations. In recent years, the rapid development of transfer learning has facilitated model training, especially when dealing with limited training data, significantly reducing training time and costs. Many scholars use network models pre-trained on ImageNet, such as VGG [49] and ResNet [50], as the backbone structure for their models. For instance, the pre-trained VGG16 from ImageNet was introduced by DeepLab V1 [51], along with the proposal of spatial convolution (dilated/atrous convolution) to increase the receptive field, addressing the issue of reduced resolution due to repeated pooling and downsampling. ResNet-50 was adopted as the backbone structure for PSPNet, which introduced spatial pyramid pooling (SPP) to gather contextual information from different regions, thereby enhancing its ability to obtain global information. DeepLab V2 [52] replaced the VGG16 backbone of DeepLab V1 with ResNet-101 and, inspired by SPP, introduced atrous spatial pyramid pooling (ASPP) to integrate multi-scale information. The emergence of SPP and ASPP resolved the issue of needing to resize images before they enter the neural network, especially for fixed-size inputs like 224 $\times$ 224 images. At present, some scholars introduce SPP and ASPP modules into models to enhance the extraction of road features from images through feature fusion. Lan et al. [53] and Gao et al. [54] have respectively proposed the GC-DCNN and Tes-LinkNet models based on the U-Net and LinkNet models. The former introduces the SPP module into the intermediate layers, while the latter uses the ASPP module. Huan et al. [55] introduced the SANet model pre-trained with ResNet-50 and introduced the ASPP module in the encoder. Inspired by dense convolution, Q. Wu et al. [56] introduced the dense and global spatial pyramid pooling module (DGSPP) into the decoder and encoder to enhance the network's perception and aggregation of contextual information. Wei and Zhang [57] integrated the multi-level strip pooling module (MSPM) into the skip connection layers to ensure road connectivity by aggregating long-range dependencies from different levels. LinkNet used ResNet-18 as the encoder backbone and improved segmentation efficiency by directly connecting the encoder and decoder. D-LinkNet employed the pre-trained ResNet-34 as the encoder backbone and introduced dilated convolutions in the intermediate layers. The design of D-LinkNet includes four progressively larger dilated convolution layers, forming a stacked pyramid pattern, also known as the D-Block, making the output of each layer the input to the next. This design expands the receptive field while maintaining image resolution, contributing to its championship in the DeepGlobe 2018 Road Extraction Challenge. However, there is a potential issue with the dilated convolutions in the intermediate layers of the D-LinkNet model, as it may lead to the loss of continuous information between neighboring pixels and introduce some unrelated contextual information, affecting road extraction's connectivity and integrity. Therefore, some scholars have enhanced the dilated convolutions in the intermediate layers of the D-LinkNet model. Gong et al. [58] replaced dilated convolutions with dense dilated convolutions, enabling multi-scale information fusion while expanding the receptive field. Wang et al. [59] restructured the D-Block into the DP-Block, inspired by the pyramid attention network [60]. They introduced global pooling and designed dense connections between convolutions to fully utilize global and dense information for enhancing road features. J. Zhang et al. [61], on the other hand, took inspiration from MobileNet V2 [62] and introduced bottleneck modules (bottleneck block) within the D-Block, forming D-Blockplus, thereby reducing network parameters and improving network performance.

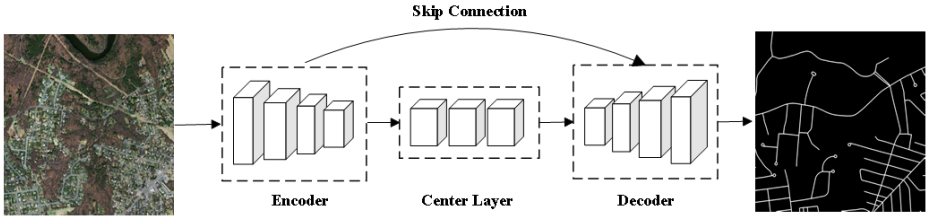

Figure 4. Network Models Based on Encoder–Decoder Structures.

2.2. Road Feature Extraction Based on Feature Fusion

Feature fusion refers to the combination and superimposition of features from different layers or branches using techniques such as weighting or concatenation. These features possess distinct characteristics. Low-level features have higher resolution, containing more positional and detailed information, but due to fewer convolutions, their semantic information is relatively less and may contain some level of noise. High-level features, on the other hand, contain richer semantic information but have lower resolution and a less effective ability to perceive detailed information. Feature fusion employs various strategies, such as feature concatenation, feature summation (including mean, pooling, weighted summation, like ASPP and SPP mentioned earlier), element-wise multiplication of feature elements, skip connections, deconvolution, attention mechanisms, and multi-scale feature fusion. These methods comprehensively utilize features of different levels and properties, making them a crucial component in network models.

2.2.1. Feature Fusion Based on Attention Mechanisms

The attention mechanism is a crucial module in deep learning networks and is considered as an additional neural network that can effectively integrate with neural networks [63]. In road feature extraction research, issues such as fragmented extraction results and poor connectivity often arise due to obstructions from buildings, trees, or background interference with similar textures. In such cases, by appropriately introducing attention modules, the model can focus more on information at road edges and intersections, leading to more connected and complete road extraction results.

In recent years, attention mechanisms have gained considerable traction in the domain of road extraction. Extensive research has delved into self-attention, channel attention [64], spatial attention [65,66], and hybrid attention mechanisms [67]. The integration of the multi-head attention mechanism from Transformer [68] into architectures like ConSwinNet [69] and Seg-Road [70] has effectively addressed the limitations of conventional CNNs, markedly enhancing the ability to perceive road texture intricacies and contextual information. Modules like the self-attention feature transfer module (SAFM) [71] have further facilitated comprehensive information integration within models, significantly bolstering the performance and robustness of road extraction tasks.

The foundational mechanisms of the channel attention module (CAM) and spatial attention module (SAM) play pivotal roles in road extraction. Networks such as Nested SE-DeepLab [72] and RALC-Net [1] have overcome challenges in road feature extraction by leveraging the squeeze-and-excitation (SE) and residual attention (RA) modules. Additionally, the incorporation of serial or parallel attention mechanisms like the convolutional block attention module (CBAM) [73] and ProCBAM [74] markedly improved the network's focus on road information, thereby elevating the performance of road extraction tasks. These innovative methods and varied applications of attention mechanisms comprehensively showcase effective strategies for enhancing model performance in road extraction tasks, enabling more efficient capture of road-related information. We have summarized the prevalent attention mechanism modules in current road extraction tasks in Table 1.

Table 1. Attention Mechanisms and Methods.

Model/Method	Attention Mechanism	Highlight(s)/Strength(s)
ConSwin-Net [69]	Multi-Head Self-Attention	Introduction of dual Swin Transformers and a residual block within the U-Net network structure, creating the ConSwin-Net, which mitigates CNN limitations in extracting global contextual features, thereby enhancing the model's perception of road texture details and global information
Seg-Road [70]	Self-Attention	Incorporation of a Transformer structure into the encoder combined with a convolutional neural network (CNN) decoder leads to improved connectivity in road segmentation and enhanced prediction result robustness
FSNet [71]	Self-Attention	Integration of the self-attention feature transfer module (SAFM) into the hidden layers of convolutional neural networks establishes relationships between each hidden layer and its contextual hidden layers. This facilitates the transfer of hidden layer feature information to the original feature map, resulting in improved road extraction performance
Nested SE-DeepLab network [72]	Channel Attention (SE)	Introduction of SE module into the encoder and decoder effectively merges and retains both shallow and deep information, addressing model imbalance issues in narrow road extraction
DSDNet [29]	Channel Attention (SE)	Integration of the SE module into the encoder of the D-LinkNet network assists ResNet in feature extraction for mountain roads
TSE-LinkNet [54]	Channel Attention (SE)	Combining the SE module with the ASPP module during downsampling enhances topological relationships between adjacent road pixels in images
BMDANet [75]	Modified Efficient Channel Attention (MECA)	Utilization of the improved MECA module enhances the continuity of road features based on the characteristics of RSI roads
MSACon [76]	Spatial Attention (SAM)	Construction of an MSACon dual-encoder network with a spatial attention-based fusion (SAF) mechanism improves road extraction by utilizing contextual relationships between roads and buildings

Table 1. *Cont.*

Model/Method	Attention Mechanism	Highlight(s)/Strength(s)
RALC-Net [1]	Spatial Attention (SAM)	Development of a dual-encoder RALC-Net network with a residual attention (RA) module integrates spatial contextual information to emphasize local semantics, aiding in the extraction of local road features
GCB-Net [28], CDG [77], CADUNet [78]	Global Attention (GA)	Focusing on highlighting high-level road features to improve segmentation results
CADUNet [78]	Core Attention (CA)	Ensuring the maximum transmission of road information between dense blocks and coordinating multi-scale road information acquisition through the global attention module
SANet [55]	Strip Attention (SAM)	Facilitating the fusion of lower-level and higher-level road features
FE-LinkNet [59]	Criss-Cross Attention (CCA)	Enhancing pixel-level representation capabilities by capturing long-range contextual information in horizontal and vertical directions
SegRExt-F [67]	Convolutional Block Attention Module (CBAM)	Improving network focus on images through concatenation of channel and spatial attention using CBAM
DU-Net [74]	Pro Convolutional Block Attention Module (ProCBAM)	Enhancing the integration of road information through ProCBAM with added SE module
SDG-LinkNet [61]	Position Attention Module (PAM) with D-Blockplus	Introducing the position attention module (PAM) and global information recovery module (GIRM) in parallel for global information acquisition
Meca-Net [66]	Long-Range Context-Aware Module (LCAM)	Designed to alleviate road occlusion issues by acquiring long-range context information through channel and spatial attention
GAN [79], MAU-Net [80], GAMSNet [81], CM-FCN [82]	Parallel Channel and Spatial Attention	Enhancing road information extraction and segmentation performance through the integration of parallel channel and spatial attention
MAU-Net [80]	Feature Fusion based on Attention Mechanism (FFBAM)	Introduced a feature fusion mechanism (FFBAM) for better fitting multi-scale road information
BMDANet [75]	Block Multi-Dimensional Attention (BMDA) Module	Introduced BMDA for feature extraction in blocks, integrating them through channel and spatial attention

Table 1. *Cont.*

Model/Method	Attention Mechanism	Highlight(s)/Strength(s)
CMAFE [83]	Cascaded Multi-Scale Attention Feature Enhancement (CMAFE)	Coarse feature extraction with dilated convolution pooling, followed by boundary enhancement in the lightweight U-Net network
Rse-Net [84]	Multi-Scale Convolutional Attention Module (CSAM)	Introduced Rse-Net with multi-scale convolutional attention module, focusing on boundary information and expanding the receptive field for more semantic information

2.2.2. Feature Fusion Based on Multi-Scale Images

The term "multi-scale" refers to images of different resolutions or different levels of image features (low-level features, high-level features). The purpose of feature fusion is to explore how to effectively utilize these multi-scale images to obtain more accurate road feature information [85].

The design of multi-scale feature fusion modules often draws inspiration from parallel or serial multi-branch network architectures, such as feature pyramid networks (FPNs) [86], Inception [87], and HRNet [88]. This section provides an overview of the multi-scale feature fusion modules and methods employed in road image segmentation tasks. Researchers have utilized supervised learning by combining edge information with image features to enhance road image segmentation networks. Various module designs have been proposed to address issues related to extracting road shapes and enhancing connectivity, such as the multi-scale context augmentation module [89], spatial context module [90], and feature review module [91]. Some modules are particularly adept at capturing elongated road shapes, while others focus on enhancing global features. Additional modules aim for multi-scale feature fusion. Solutions tailored for narrow, continuous, and expansive roads in high-resolution remote sensing images have also been proposed, incorporating multiple modules to optimize spatial feature preservation, shape enhancement, and multi-feature fusion. These innovative modules and methods collectively drive advancements in road extraction tasks, providing crucial technical support for more accurate identification of road shapes and improved segmentation outcomes. Due to space limitations, detailed method characteristics are summarized in Table 2.

Table 2. Multi-Scale Feature Fusion Module and Methods.

Model/Method	Multi-Scale Feature Fusion Module	Highlight(s)/Strength(s)
Geographic Feature-Enhanced Network [92]	Joint Shared Learning and Feature Fusion	Enhancing road extraction connectivity through joint learning of pixel-level, edge-level, and region-level road features, followed by feature fusion
DA-CapsUNet [89]	Multi-Scale Context Augmentation (CTA)	Enlarging the receptive field and integration of context information from different scales
BT-RoadNet [90]	Coarse Map Predicting Module (CMPM) and Spatial Context Module	Serially connected spatial context module effectively captures elongated road shapes

Table 2. *Cont.*

Model/Method	Multi-Scale Feature Fusion Module	Highlight(s)/Strength(s)
Deep FR TransNet [91]	Feature Review (FR)	Evaluating road features of varying scales, with an emphasis on contour characteristics, to improve road profile information
DCANet [93]	Discriminative Context-Aware Feature (DCF) Module	Aligning feature maps across scales to extract high-frequency information, with a refine decoder (RD) for spatial information retention and feature representation
AF-Net [94]	All-Scale Feature Fusion (AF) Module	Recursive integration of features from two pathways, leveraging scale features with varying spatial and semantic information, to provide accurate spatial and semantic information for road extraction
NFSNet [71]	Global Feature Refinement (GFR) Module	Improved semantic information of feature maps for more detailed segmentation outputs
ConSwin-Net [69]	Feature-Enhanced Connection (FC) and Shape-Augmented Connection (SC)	Enhanced and separate transmission of structural and textural features to the decoder, improving overall model performance
MLEM-NET [95]	Multi-Scale Line Enhancement Module (MLEM)	Utilizing the Hough transform (HT) to enhance local and global linear features in remote sensing images
SDUNet [96]	Densely Connected Encoder Block and Spatial Intensifier (DULR) Module	Constructing spatial relationships between features at different positions and introducing skip connection layers to preserve the topological structure
Meca-Net [66]	Multi-Scale Feature Encoding Module (MFEM)	Utilizing convolution kernels of different scale sizes and aggregating multi-scale features through a parallel strategy for recognizing elongated roads
MSPFE-Net [57]	Feature Enhancement Module (FEM) with Stripe Pooling	Extracting and merging features from various levels to accomplish multi-scale feature fusion
LDANet [97]	Feature Expansion Module and Deep Feature Association Module	Expanding and merging features to address challenges posed by narrow and complex rural roads, improving feature associations, and promoting multi-feature fusion
MTMF [98]	Canny Operator and HRNet	Improving road image segmentation through the fusion of edge information and image features

2.2.3. Feature Fusion Based on Multi-Modal Fusion

Solely relying on optical remote sensing imagery to provide learning information for network models does not guarantee excellent learning outcomes. This is due to spectral similarities between buildings and roads and the potential for occlusions caused by tall buildings and trees. These factors can lead to inaccurate identification and acquisition of

road feature information by the model, ultimately affecting road extraction results. Additionally, sensor imaging and lighting conditions can also adversely affect the recognition and acquisition of road feature information. Recognizing this challenge, researchers have explored multi-modal data, including multi-spectral (hyperspectral) data, synthetic aperture radar (SAR) [99], light detection and ranging (LiDAR), unmanned aerial vehicle (UAV) data, GPS trajectory data, and multi-temporal data. The penetrative and oblique observation properties of synthetic aperture radar (SAR) have been ingeniously leveraged by J. Zhang et al. [61] to address issues arising from shadows and occlusions caused by vegetation and buildings in optical remote sensing, providing network models with more detailed road information. On the other hand, dual-temporal optical remote sensing imagery has been employed [100] to detect and update road databases. Sensors with high revisit times, such as Sentinel-1 and Sentinel-2, have been utilized by Ayala et al. [28] to enhance datasets with multi-temporal multi-spectral and SAR data through color data augmentation.

Multi-modal fusion involves feature integration between different data sources, particularly for cross-source fusion between GPS trajectory data and remote sensing imagery. Similarly, we have provided a more intuitive tabular summary of methods related to multi-modal feature fusion in Table 3.

Table 3. Multi-modal Fusion Module and Methods.

Model/Method	Module Name	Fusion Data Sources	Highlight(s)/Strength(s)
DeepDualMapper [101]	Gated Fusion Module (GFM)	GPS trajectory data and remote sensing imagery	GFM was designed to control and integrate information from both modalities in a complementary perception manner
MTMSAF [102]	Adaptive Fusion Module (AFM)	GPS trajectory data and remote sensing imagery	AFM was utilized to integrate road features from trajectory data and remote sensing imagery
CMMPNet [103]	Dual Enhancement Module (DEM)	Cross-source fusion of images and trajectory data	DEM was introduced to enhance and complement features from both images and trajectory data bidirectionally, applicable for LiDAR and remote sensing imagery data
MSFANet [104]	Cross-source Feature Fusion Module (CFFM)	Traditional remote sensing imagery and hyperspectral imagery	Hyperspectral and remote sensing imagery are combined to alleviate discontinuous outputs and using CFFM to correct and fuse spectral features at different scales, reducing noise and redundancy

Attention mechanisms themselves are models with advantages such as fewer parameters, faster processing speed, and good performance. Compared to CNNs, attention mechanisms have lower model complexity, fewer parameters, and lower computational requirements. Furthermore, attention mechanisms address the issue of non-parallel computation in RNNs [105], as they do not rely on the results of the previous step, enabling efficient parallel computation. Hence, they have become an important component of feature fusion in network models. However, it is worth noting that the introduction of attention mechanisms may lead to model overfitting. If a network model is already complex, incorporating attention mechanisms can increase the number of model parameters, potentially causing overfitting issues. Additionally, fusing different features together may introduce noise and other challenges. Attention itself is a type of feature, so when integrating it with other features, careful consideration is needed to assess whether it might negatively impact the network model's performance. For multi-modal data, while it provides richer semantic information to networks, there may be differences in semantics among different modalities.

Therefore, addressing noise reduction and semantic differences while fusing these features is an issue to be focused on in the future.

2.3. Road Feature Extraction Based on GANs

In 2014, generative adversarial networks (GANs) were introduced by Goodfellow et al. [106] operating on an unsupervised learning approach, consisting of a generator G and a discriminator D. The task of the generator is to generate data closely resembling real images, attempting to "deceive" the discriminator. The discriminator's role is to determine whether the data generated by the generator is correct and provide feedback to enhance the generator's ability to "fabricate". This process forms a cycle, continuing until neither can deceive the other. Essentially, it is a zero-sum game, also known as the Bash game. However, because the generator does not require training labels, data can be generated too freely, including images, text, or even sound from noise, which is not ideal for image recognition tasks. To address this issue, the introduction of some conditions to both the generator and discriminator was proposed. In the context of image recognition tasks, conditions could be introduced to the discriminator to make it generate only images. In the same year, conditional generative adversarial networks (CGANs) [107] were introduced (Figure 5). CGANs are generative adversarial network models with constraint conditions. Incorporating variables y into both the generator and discriminator, these variables guide the data generation process by the generator. The variables y can be labels or even images, marking a shift of GANs from unsupervised learning towards supervised learning.

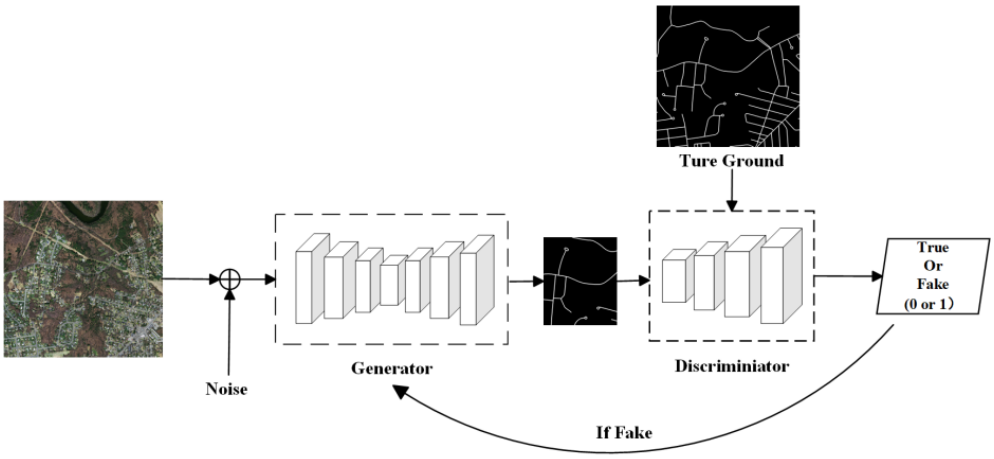

Figure 5. Network Model Based on Conditional Generative Adversarial.

In 2017, the Pix2pix [108] model was introduced, which is based on the structure of conditional generative adversarial networks (CGAN) for image-to-image transformations, also referred to as domain adaptation. In this approach, the generator of the model utilizes a U-Net network, while the discriminator is designed using the PatchGAN architecture. Many researchers continue to reference this model in current road extraction tasks. For instance, Yang and Wang [109] followed the structure of Pix2pix and introduced the WGAN-GP network for rural road extraction. They used both U-Net and BiSeNet as generators, employing an ensemble strategy to combine their inference outputs for better road vector generation. The discriminator in their model used PatchGAN. Cira et al. [110,111] applied the Pix2pix model to post-process road extraction. They improved the integrity of road surface area extraction by contaminating labels and reconstructing them. In addition, Abdollahi et al. [7] proposed a deep learning approach using conditional generative adversarial networks (CGANs) for road segmentation in high-resolution aerial imagery. They utilized an enhanced U-Net model (MUNet) as a generator to segment images and obtain high-resolution segmented maps of road networks. NIGAN [112], comprising two CGAN networks, was used for scene selection in mountainous road scenarios. This was caried out to pre-select areas that contain mountainous road scenes, thereby reducing the workload

in subsequent segmentation and road extraction tasks. The generator in their model is based on an encoder–decoder structure, utilizing ResNet-34 as the backbone. Middle layers incorporate dilated convolutions, which are helpful for extracting small objects like roads and expanding the receptive field while enhancing global information.

Conditional generative adversarial networks (CGANs) have played a crucial role in road extraction tasks. They are not only used for road segmentation but also for preprocessing road extraction, enriching road information in images, and reducing the workload for subsequent segmentation networks. Additionally, in post-processing, employing adversarial training techniques to enhance segmentation results has reduced issues related to fragmentation while improving road connectivity.

2.4. Road Feature Extraction Based on Cumulative Integration of Multiple Models

In road extraction tasks, ensemble strategies have been increasingly adopted by researchers to combine multiple models serially or in parallel. Integrated models with strong generalization capabilities, high robustness, and exceptional segmentation performance have been highly sought after in research endeavors. Parallel strategies (Figure 6) are most commonly used. For example, Senthilnath et al. [113] employed three relatively mature network models, FCN-32, Pix2Pix, and CycleGAN [114], for transfer learning. Both Pix2Pix and CycleGAN are commonly used in domain transfer tasks. The key difference is that Pix2Pix requires training data to be in pairs, which is challenging to find in the natural world. The emergence of CycleGAN effectively solves this problem. They proposed the Deep TEC integrated classifier, which utilizes a parallel strategy to integrate the results of road segmentation from three models. This approach achieved outstanding integration performance in extracting urban road networks from drones. Cira et al. [115] combined improved CNN, VGG, ResNet-50, and Inception-ResNet [116] models in parallel and fused extraction results using an averaging structure. This strategy aims to leverage the strengths of each model while minimizing their weaknesses, ultimately resulting in a classifier with reduced classification error. Chen et al. [117] employed ResNet-50 models with three distinct convolution kernel sizes for road extraction, integrating the results to form a ResNet-50 training block enriched with high-level information. Li et al. [118] reorganized the layers of U-Net and duplicated a single submodel N times, creating an ensemble model E consisting of N parallel submodels. Following optimization and prediction, they ultimately established an E-UNet model with 14 layers. Abdollahi et al. [119] adopted a parallel approach by linking two improved U-Net models, BCL-UNet (ConvLSTM [120] + U-Net) and MCG-UNet (BConvLSTM + SE + dense convolutions [121]). They introduced dense convolutions and compression activation modules in the upsampling layers of the standard U-Net. They employed bidirectional convolutional long short-term memory (BConvLSTM) for skip connections, enabling the generation of high-resolution segmentation maps even in challenging backgrounds while preserving edge information. The graph-based dual convolutional network (GDCNet) [122] integrates graph convolutional networks (GCNs) and CNNs. Employing a ResNet-50 backbone that included encoder and decoder convolutional neural networks, researchers applied a parallel approach for road extraction, effectively addressing concerns associated with poor connectivity and discontinuities. This was achieved by generating complementary spatial–spectral features at both superpixel and pixel levels and efficiently propagating these features between graph nodes and image pixels using a graph decoder. Sun et al. [123] employed a parallel network model consisting of dual branches for road and building extraction. One branch is the multi-resolution semantic extraction branch, composed of three parallel ResNet networks, used to extract semantic features of roads and buildings at different resolutions. The other branch is the Transformer semantic extraction branch, which utilizes a ResNet-18 backbone and features a Transformer-based encoder–decoder. This parallel strategy successfully addresses the current limitation of semantic segmentation networks in terms of receptive field by fusing the output results of the two branches.

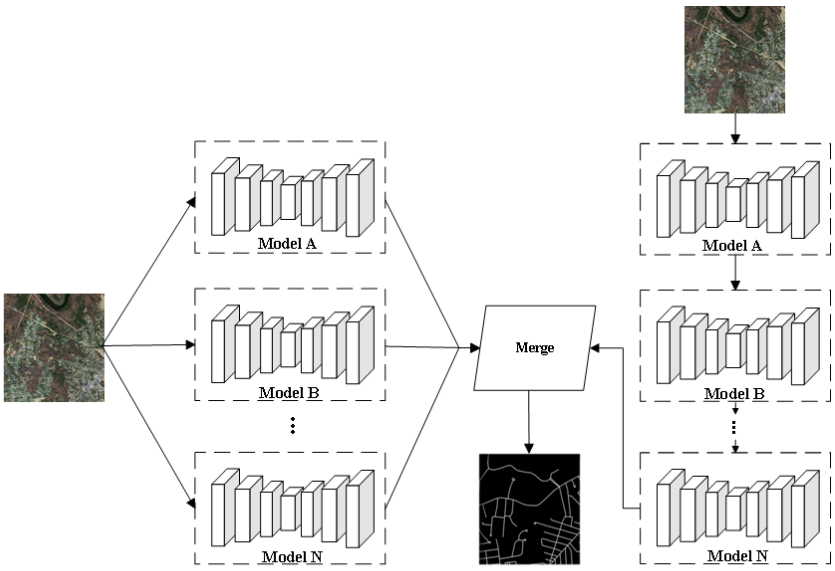

Figure 6. Network Model Based on Cumulative Integration of Multiple Models.

Certainly, a serial strategy employing multiple models for road extraction is also utilized by some researchers. For instance, a direction-aware residual network, DiResNet [124]. DiResNet comprises a ResNet segmentation network (DiResSeg) based on the decoding layers with structural supervision and a refinement network (DiResRef) based on U-Net. The former is dedicated to enhancing the learning of road topology, while the latter further refines the road segmentation results. Z. Chen et al. [125] drew inspiration from the AdaBoost classification algorithm and combined multiple lightweight U-Net models by connecting them in a serial manner, forming AdaBoost-like end-to-end multiple lightweight U-Nets (AEML U-Nets). Under this serial strategy, the output of the previous network serves as the input for the next one. To ensure the training quality of each U-Net, the researchers designed a multi-objective optimization strategy for joint training of all U-Nets. Finally, the output results of each U-Net are fused to obtain the ultimate road extraction result.

With the continuous development of deep learning, models are gradually evolving towards greater depth and width. However, it is important to note that increasing depth and width does not always lead to improved model performance and can potentially result in issues like overfitting. In this section, we summarize how scholars leverage the unique characteristics of different models and employ ensemble strategies to integrate these models. These characteristics include having fewer model parameters, fast recognition speed, strong generalization, and expertise in extracting road features in various scenarios. By combining multiple models, whether they are simple or mature, researchers have achieved better road feature extraction results than with a single model. Nonetheless, it is essential to be aware that multiple independent models do not always outperform a deeper and larger single model. This is because these models are trained independently, and their training outcomes may vary. In parallel extraction, individual models may perform poorly, becoming bottlenecks for overall performance. In serial extraction, if the same model is used for serial processing, it may lead to a series of problems. For instance, determining strategies to ensure consistent training results for each model and whether an excessive number of models effectively deepens the model's depth, potentially leading to gradually declining performance. These issues are worthy of in-depth consideration and exploration.

2.5. Road Feature Extraction Based on Multiple Tasks

The focus of most current road extraction tasks is primarily on extracting road surfaces. However, roads encompass various elements, including road centerlines, road edges, road nodes, and more, all of which are equally important. Consequently, the challenge of

achieving multi-task road extraction persists. Many researchers are exploring network models for accomplishing multi-task road extraction in remote sensing images, surpassing the scope of surface extraction alone (Figure 7).

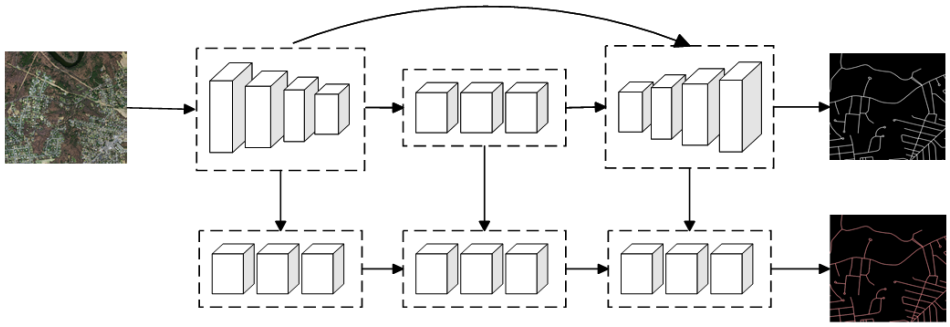

Figure 7. Network Models Based on Multiple tasks.

In the road surface and centerline extraction tasks, the D-LinkNet model was employed [126]. Initially, the imagery was coarsely segmented for road extraction. Subsequently, the boosting segmentation network (BSNet) based on the ResNet-34 network architecture was used to enhance the connectivity and accuracy of the coarse segmentation results. Road intersections simultaneously generated starting points by employing multi-start point tracking. Finally, an iterative search strategy embedded with convolutional neural networks (CNNs) was used to track a continuous and complete road network. Refined extraction of road surfaces and centerlines was achieved by integrating segmentation, tracking results, semantic information, and topological data. A dual-task end-to-end convolutional neural network (MRENet) [127] with a dual-branch structure was developed. These two branches facilitated feature sharing, with the main branch responsible for road surface extraction, and the other branch utilizing features extracted from the main branch as conditions for centerline extraction. This information exchange and parameter sharing approach helped mitigate potential issues arising from insufficient centerline samples. To address the problem of poor connectivity in road extraction often caused by complex backgrounds, Lu et al. [128] identified interconnections between different extraction tasks. For example, the road surface segmentation results influenced the final position of centerlines and edges, and the integrity of road edges was closely related to road surface connectivity. Therefore, they proposed a cascaded multi-task (CasMT) road extraction framework to simultaneously extract road surfaces, centerlines, and edges. This framework fully leveraged the interrelationships between these tasks, promoting interconnectivity within the road network.

To improve the connectivity of road surfaces, additional information about roads, such as road nodes and intersections, is also extracted by many scholars in multi-task extraction. D. Chen et al. [129], while using network models to extract road surfaces, also extract information about road nodes. This node information provides supervision for road surfaces, contributing to their continuous improvement in connectivity. X. Chen et al. [130] constructed a node inference branch within the network, modeling road nodes together with road surfaces, thereby enhancing the topological structure of roads and reducing surface fragmentation. Roads and intersections are two crucial elements in road network generation. Li et al. [102] using trajectory data and remote sensing images, and not only extracted road surfaces but also recovered intersection information from road area features, simultaneously performing road surface and intersection extraction tasks. Additionally, some researchers apply multi-tasking to segmentation and change detection. M. Zhou et al. [100] proposed a neural network with dual-task road change detection, called dual-task dominant Transformer-based neural network (DT-RoadCDNet). This network takes input from two-phase remote sensing images and can perform both segmentation and change identification tasks, resulting in two road surface segmentation images before and after changes and one road change image.

Roads are not only composed of road surfaces but also include elements such as road centerlines, road edges, and road nodes. The emergence of multi-task road extraction has the potential to enhance road information, facilitating better road pipeline planning. However, in current road extraction tasks, research focused on road centerlines as the primary extraction task is relatively scarce, with most relying on labeled data provided by OpenStreetMap (OSM). Road centerlines are not only vital components of roads but can also serve as weak labels for subsequent tasks based on weak supervision learning. Additionally, road edges and road nodes are equally crucial. Edges determine the integrity and continuity of road surfaces, while linear elements consist of nodes. Nodes can be used as additional information for predicting and inferring road surface breakpoints and completing linear elements, thus improving road connectivity. They can also serve as road backbones, facilitating subsequent road vectorization processing. Road networks evolve and change each year, and electronic maps require timely updates of road networks. Traditional methods often require substantial human and material resources for field surveys. Road change detection tasks rely on neural networks and remote sensing images, automating the extraction of road changes from images, reducing the need for manual intervention. However, due to limitations in data sources and labels, change detection tasks still face issues of missed detections and false alarms, necessitating further improvement in data source quality, label quality, and network model quality.

2.6. Road Feature Extraction Based on Network Optimization

The various strategies employed by research scholars in optimizing the training of network models are research hotspots, and the primary focus is loss functions. Loss functions play an indispensable role in the training of network models, as they measure the difference between the model's predictions and the ground truth. Model performance is typically evaluated by calculating the loss value, where lower loss signifies better model performance, indicating that the model's predictions are closer to the ground truth.

We find that the dice coefficient loss, binary cross entropy loss, and cross entropy loss are the most commonly used loss functions. Since road extraction tasks are typically binary semantic segmentation tasks, binary cross entropy loss is more common than cross entropy loss. Additionally, in model training, the dice coefficient loss is used to measure the similarity between predicted results and labels, while binary cross entropy loss is employed to assess the distance between predicted results and actual labels. For instance, Lin et al. [72] introduced both of these loss functions into their proposed SE-DeepLab network and compared their effectiveness in model training. They found that the dice loss was better suited for their model, significantly enhancing its performance during training and prediction. Similarly, Lan et al. [53] also argued that the dice coefficient loss is more suitable for road segmentation tasks because it conducts global assessment, whereas binary cross entropy loss is pixel-wise. When extreme imbalance exists between foreground and background, binary cross entropy loss may not effectively address this issue. However, the dice coefficient loss is sensitive to noise and may overlook boundary information, leading to poorer road edge segmentation. To address this concern, Zao and Shi [131] proposed an edge-focused loss, which guides the network to pay more attention to road edge regions. Additionally, they introduced an enhancement factor that assigns higher loss contributions to pixels closer to the edges, thereby improving road boundary segmentation.

Different types of loss functions are combined, which is a training strategy used by the D-LinkNet. The loss functions were integrated by using various combinations of strategies [58,79,132] to fully exploit their respective advantages in road extraction. For example, Abdollahi et al. [133] introduced the VNet network model for road extraction and proposed a new dual-loss function called cross entropy and dice loss (CEDL). This loss function combines cross entropy (CE) and dice loss (DL) because cross entropy considers local information while dice loss focuses more on global information. Introducing the CEDL loss function into VNet can reduce the impact of class imbalance issues, thus improving road extraction results. Since high-resolution remote sensing images typically include

complex backgrounds such as occlusion, shadows, and similar textures in the surrounding terrain, many roads are difficult to identify successfully, leading to a relatively high rate of omissions. To address this challenge, Lu et al. [128] introduced the hard example mining (HEM) loss function. This loss function, by jointly using dice and binary cross entropy loss functions, pays more attention to hard samples, enhancing road recognition and further improving road completeness.

To address the issue of sample imbalance, the focal loss function has been employed by some researchers [28,89,134]. Additionally Wei and Zhang [57] combined focal loss with the dice function. The focal loss function [135] differs from traditional cross entropy functions by focusing on resolving sample imbalances and confounding pixel categories. Abdollahi et al. [136] introduced a loss function called median frequency balancing focal loss weighted (MFB_FL) based on the focal loss function to deal with highly imbalanced datasets, where positive samples are scarce. The introduction of MFB_FL eases the burden on simple samples, allowing more time to be spent learning difficult samples, thereby improving road extraction and road vectorization results. The issue has also been addressed by some researchers through modifications to the loss function. Yang and Wang [109] added a spatial penalty term to the loss function to address the typical class imbalance issue in road extraction. Additionally, the softmax cross entropy loss (SCE), Jaccard, and Lovasz softmax (LZS) loss functions have been applied in binary road extraction tasks. J. Zhang et al. [61] combined Jaccard and cross entropy losses in the training of the SDG-LinkNet model to avoid the problem of single cross entropy easily falling into local optima. Furthermore, Sushma et al. [137] simultaneously used LZS and boundary loss functions during model training, with results showing their superiority over the mean squared error (MSE) loss.

With relatively limited research on loss functions in road extraction tasks, an attention loss function called GapLoss was proposed by Yuan and Xu [138]. This function can be combined with any segmentation network. Firstly, a binary prediction mask is obtained using a deep learning network. Secondly, a vector skeleton is extracted from the prediction mask. Thirdly, for each pixel, eight adjacent pixels with the same value are calculated, and if the value is 1, the pixel is identified as an endpoint. Fourthly, based on the number of endpoints within a buffer range, the corresponding weight is assigned to each pixel in the predicted image. Finally, the weighted average of the cross entropy of all pixels in the batch is used as the final loss function value. GapLoss was introduced into four relatively basic network models (PSPNet, U-Net++, SegNet, and MUNet), and the training results outperformed the use of the three loss functions: dice, binary cross entropy, and focal. This suggests that GapLoss not only improves the connectivity of predicted roads but also enhances the accuracy of road predictions. Xu et al. [139], based on the D-LinkNet, compared twelve well-known loss functions, categorizing them into region-based (such as dice, Jaccard, and focal), distribution-based (such as binary cross entropy), and composite-based (such as a combination of dice and binary cross entropy). They found that different loss functions performed significantly differently under different models. Region-based loss functions generally outperformed distribution-based ones, while the performances of region-based and composite-based loss functions were comparable. This indicates that the choice of the most suitable loss function should be based on the model's design.

In addition to the utilization of loss functions for optimizing model training, the traditional batch normalization (BN) layer has been replaced with filter response normalization (FRN) in the upsampling layer by some researchers [27,140]. With the introduction of this layer, the model decreases its dependence on random batches, thereby benefiting model optimization and enhancing training efficiency.

This section primarily introduces the fundamentals of network optimization in road extraction tasks, with an emphasis on the utilization of loss functions. Additionally, it briefly mentions adjustments made between different layers of the model to enhance the model's training capabilities. Concerning the application of loss functions, binary cross entropy, dice loss, and their combinations represent the most commonly employed loss functions in model training. However, due to variations inherent in different models, the performance

of various loss functions may exhibit differences. Furthermore, it is worth noting that there is relatively limited in-depth research on loss functions in the road extraction field. Although dice loss and binary cross-entropy–dice combinations are presently regarded as more suitable loss functions, the question of whether these loss functions can consistently perform well in new models that are deeper, wider, and larger warrants consideration. Therefore, one of the future research directions involves the design of loss functions with strong generalization capabilities aimed at improving performance on diverse models.

3. Road Feature Extraction Based on Semi-Supervised (Weak) Deep Learning Network Models

Semi-supervised learning falls within the domain of weakly supervised learning, combining elements of both unsupervised and supervised learning. It consists of a supervised learning part and an unsupervised learning part. Zhou [141] subdivided weakly supervised learning into three categories: (1) incomplete supervision refers to the situation where only a portion of the training data are labeled, and the rest are unlabeled. (2) Inexact supervision refers to the provision of coarse-grained label information in the training data, which is more common in tasks such as object detection and instance segmentation but less prevalent in road extraction tasks, where road extraction is typically a binary semantic segmentation problem. (3) Inaccurate supervision means that the labels in the training data may contain errors or inaccuracies, which are inevitable in road datasets because road labeling typically involves manual annotation. The author proposes corresponding solutions for these three types of supervision. For incomplete supervision problems, active learning or semi-supervised learning methods are used. Additionally, multi-instance learning can be applied to address inexact supervision problems. For inaccurate supervision problems, learning with label noise strategies is employed, introducing noise to the labels for model training. In summary, both semi-supervised learning and weakly supervised learning rely on a small amount of labeled data and a large amount of unlabeled data for training models and improving performance. In the field of road extraction, researchers have used various methods to address the issue of limited labeled data. This section will explore this issue from the perspectives of weakly supervised learning and semi-supervised learning.

3.1. Road Feature Extraction Based on Weakly Supervised Learning

In weakly supervised road extraction tasks, the challenge of acquiring pixel-level labeled data at a high cost and difficulty is encountered by researchers. Therefore, the exploration of alternatives such as weak label data, such as point or line annotations, has become a focus. These data are comparatively easier to obtain and more abundant than pixel-level labels, making them the preferred choice for researchers. For instance, a method known as "deep windows" [142] effectively utilizes point annotation data in road centerline extraction tasks. A block-based road center point estimation model was initially designed, inspired by the stacked hourglass networks applied in the field of human pose estimation [143]. This model was then trained using point annotations (indicating the center points of roads in training blocks) to predict road center points within local blocks. Subsequently, the direction of the road was estimated using the Fourier spectrum analysis algorithm. Guided by the CNN model, road center points within blocks were iteratively tracked and connected along the road's direction, completing the road centerline extraction. Building upon this method, Lian and Huang [144] further developed a point-based weakly supervised road segmentation method for road surface extraction. Point annotation data were initially utilized to detect road seed points and background points in remote sensing images. These points were then used to train a support vector machine classifier (SVC) for classifying each pixel in the image as road or non-road. Simultaneously, a multi-scale and multi-direction Gabor filter was introduced to estimate the road potential of each pixel based on the preliminary classification results, taking into consideration the local geometric and directional features of the road. Finally, an active contour model algorithm based on

local binary fitting energy (LBF-Snake) was introduced to extract road contours from non-uniform road potential maps and optimize road regions through simple post-processing.

The weakly supervised road surface extraction method "ScRoadExtractor" was proposed [145]. This method utilizes road centerlines as line drawing label data and combines remote sensing images with a road label propagation algorithm to generate pseudo-labels. Holistically nested edge detection (HED) was employed for edge detection within the imagery boundary. Additionally, a network model with a dual-semantic branch (DBNet) was designed for training. The model's primary branch is based on an encoder–decoder structure, with ResNet-34 serving as the encoder backbone. The intermediate layer incorporates atrous spatial pyramid pooling (ASPP). The decoder includes road surface segmentation and road boundary detection branches, which utilize segmentation and boundary loss functions to assess the similarity between the segmentation results and pseudo-labels and the edge segmentation results and edge detection. This enables the network to iteratively optimize and improve road extraction. M. Zhou et al. [146] observed that in the presence of background occlusion and spectral confusion in remote sensing images, road edges tend to appear blurry. Using single-pixel-width line drawing labels alone to approximate the position of road centerlines does not offer sufficient supervision for road boundary learning. Consequently, this results in decreased accuracy in road surface segmentation when employing line drawing supervision methods. They also considered the label propagation algorithm to be overly complex and, as a result, opted not to use it. Instead, they introduced a weakly supervised road segmentation network, SOC-RoadNet, based on structural and directional consistency. SOC-RoadNet utilizes line drawing labels as weak supervision for road surface extraction from remote sensing images. SOC-RoadNet features a dual-branch architecture, encompassing a road segmentation branch and a road direction prediction branch. The road segmentation branch directly learns road surface features from the line drawing labels, while the direction prediction branch predicts continuous road directions to enhance road connectivity. Rather than regularizing road boundaries using unreliable edge maps, SOC-RoadNet improves the accuracy of road boundaries by introducing a structural consistency loss function. These methods illustrate how to judiciously leverage point and line annotations to enhance road extraction performance and accuracy within a weakly supervised learning framework.

3.2. Road Feature Extraction Based on Semi-Supervised Learning

When applying semi-supervised learning to road extraction tasks, three main aspects are typically addressed. The first involves consistency regularization, often entailing two branches, each dealing with samples subject to different perturbations. Through loss functions, the predictions of these two branches are encouraged to remain consistent. This means that some form of perturbation (e.g., flipping, rotating, cropping, and mirroring) is applied to unlabeled sample data, and the model's predictions should exhibit minimal changes. The second aspect pertains to adversarial training, wherein adversarial strategies are applied to unlabeled data to align the outputs of unlabeled data as closely as possible with the distribution of real data. Finally, pseudo-labeling is the third aspect, involving an initial model training using labeled data. Subsequently, the trained model is utilized to make predictions for unlabeled data, high-confidence samples (above a pre-defined threshold) are selected, and their predicted results are used as pseudo-labels. These pseudo-labeled data are integrated into the labeled dataset, and the model undergoes further training on this expanded labeled dataset through an iterative process aimed at ongoing model optimization. In general, these methods are aimed at addressing challenges such as limited label availability and high annotation costs.

(1) Based on the consistency regularization

When applying semi-supervised learning to road extraction tasks, the three approaches mentioned above have been utilized by researchers. For instance, the introduction of the idea of consistency regularization into road extraction was presented [147]. A semi-supervised semantic segmentation method for fine-grained road scene understanding was

designed. Four perturbation strategies were employed, encompassing random grayscale, random blur, random color jitter (brightness, contrast, saturation, etc.), and random Gaussian noise. A dual-branch structure was implemented, with one branch perturbing unlabeled data and the other branch preserving the original image. The combination of labeled and unlabeled samples in a U-Net model, with a balanced strategy of supervised and unsupervised losses, enabled the efficient extraction of road scene information, including vehicles, road lines, crosswalks, ground markings, and lane widths. This approach not only improved the classification accuracy of semantic segmentation networks but also mitigated the negative impact of limited labeled data on network performance. In another study [148], which focused on consistency regularization in semi-supervised learning, perturbation schemes were reviewed, and prominent data-level perturbation schemes, CutMix and ClassMix (a development from CutMix), as well as model-level perturbation representatives, mean teacher (MT) and cross pseudo-supervision (CPS), were identified. Inspired by these four perturbation methods, an end-to-end semi-supervised semantic segmentation framework named "ClassHyPer" was proposed. This framework is based on the ClassMix structure and simultaneously incorporates MT and CPS perturbations to form a mixed perturbation strategy. The images subjected to these mixed perturbations were then processed through a classic FCN with VGG16 as the backbone structure. By employing various loss functions to calculate sample correlations, ClassHyper exhibited strong performance on five different urban and road datasets, demonstrating its potential in enhancing model performance when confronted with limited labeled data.

(2) Based on the consistency regularization and pseudo-labels

The concept of consistency regularization and pseudo-labeling was introduced into semi-supervised road extraction tasks by You et al. [149], who proposed a novel semi-supervised remote sensing road extraction method called "FMWDCT". This method comprises two key components: dual-network cross training (DCT) and foreground pasting (FP). The objective of dual-network cross training is to address common challenges in remote sensing image segmentation tasks, such as limited training data and high annotation costs. Foreground pasting involves the integration of foreground pixels from labeled images into unlabeled images, generating mixed input images. This strategy aims to tackle the issue of imbalanced positive and negative training samples in road extraction tasks. In FMWDCT, each network includes both an initial network and an enhancement network. Mixed pseudo-labels are generated by combining high-confidence predictions from the enhancement network and labeled masks. Subsequently, these mixed pseudo-labels are employed to guide cross training in another adversarial base network and to facilitate smoothing updates in the corresponding enhancement network. This approach contributes to the enhancement of road extraction in situations involving limited labeled data while harnessing the potential of unlabeled data and pseudo-labeling.

(3) Based on adversarial training and pseudo-labels

The semi-supervised road extraction problem was addressed [150] through the utilization of adversarial training and pseudo-labeling. They introduced an innovative semi-supervised road extraction network known as "SemiRoadExNet", which is designed based on generative adversarial networks (GANs) and comprises a generator and two discriminators. The generator follows an encoder–decoder structure, utilizing ResNet-34 as the encoder backbone, and introduces channel attention and spatial attention in a serial strategy. Additionally, multiple dilated convolutions with skip connections are incorporated in the middle layers. Two discriminators, based on the U-Net architecture, are employed for different tasks. The working principle of SemiRoadExNet is as follows: first, labeled and unlabeled images are input into the generator network for road extraction. The generator's output includes road segmentation results and their corresponding entropy maps. The entropy map represents the confidence level for each pixel's prediction of road or non-road. Next, two discriminators are utilized to enforce the consistency of feature distributions between the road prediction maps and entropy maps of labeled and unlabeled data. Through

adversarial training, the generator is continuously regularized, exploring latent information within unlabeled data and enhancing the model's generalization capability. This method aims to maximize the utilization of potential information in low-confidence pixels in pseudo-labels, further enhancing semi-supervised road extraction models, reducing reliance on labeled data, and improving network performance.

3.3. Road Feature Extraction Based on Semi-Weakly Supervised Learning

A novel approach [151] combines the strengths of semi-supervised and weakly supervised learning, resulting in a method known as semi-weakly supervised learning. In this context, adversarial training from semi-supervised learning and the utilization of weak labels (such as road centerlines) from weakly supervised learning were leveraged to propose a remote sensing image road extraction model named "SW-GAN". SW-GAN comprises two generators and one discriminator. These generators include a fully supervised generator based on the D-LinkNet model and a weakly supervised generator based on the Res-UNet model, which incorporates learnable pyramid dilated modules into the middle and skip connection layers to expand the receptive field. The training dataset includes both fully supervised and weakly supervised datasets. During the training process, the fully supervised generator uses both the fully supervised and weakly supervised datasets, while the weakly supervised generator utilizes only the weakly supervised dataset. The output of the weakly supervised generator is employed as a feature to augment the fully supervised generator. To ensure consistency between the fully supervised and weakly supervised generators on the weakly supervised dataset, a consistency loss function is designed to encourage both generators to produce results that are as similar as possible. The discriminator employs an FCN model, aiming to distinguish whether the generated road network is a pixel-level manually annotated road network or fully supervised synthesized road network. SW-GAN effectively utilizes a limited amount of fully supervised data and a substantial amount of weakly supervised data for road network extraction in remote sensing images, combining the advantages of semi-supervised and weakly supervised learning and achieving outstanding road extraction results.

4. Discussions

This paper starts from the perspective of supervised learning in deep learning, emphasizing the technical intricacies involved in road extraction from remote sensing images, and categorizes supervised learning into four methods based on the use of pixel-level label data. The advantages and disadvantages of the four learning methods are listed in Table 4.

For a more comprehensive evaluation of model performances, we primarily assess the accuracy of the models based on five key metrics, namely intersection over union (IoU), overall accuracy (OA), Precision, Recall, and F1. IoU indicates the overlap between the predicted and ground truth road areas in road extraction tasks. OA denotes the accuracy, signifying the ratio of correctly predicted pixels to the total pixels. Precision reflects the proportion of accurately predicted road pixels by the model, while Recall measures the number of roads identified by the model. F1 is the harmonic mean of Precision and Recall. Simultaneously, we have outlined the performance of several models on the road dataset of Massachusetts, as depicted in Table 5.

LDANet [97] demonstrates exceptional performance in terms of Recall, Precision, and F1-Score, showcasing its ability to accurately identify road pixels while effectively reducing false positives. Furthermore, LDANet boasts an impressively low parameter count of only 0.2M, positioning itself as an outstanding lightweight model, thereby highlighting a promising direction for future research and adoption. Seg-Road-I, DU-Net, CM-FCN, and others exhibit commendable performance across multiple metrics, showcasing elevated levels of Recall, Precision, and F1-Score. Similar to LDANet, they serve as representatives of high-performance models in this domain.

Table 4. Comparison of 4 learning methods.

Learning Type	Labeled Data Usage	Extraction Accuracy	Generalization Ability	Prospects for Future Research	Disadvantages
Fully Supervised Learning	Large amount of high-quality labeled data	High accuracy	Relatively poor	Excellent results with sufficient labeled data, limited generalization	Requires substantial human effort and cost to label data. It may overfit to labeled data and lack adaptability to unseen scenarios
Semi-Supervised Learning	Small amount of labeled data + unlabeled data	Lower than fully supervised	Better than fully supervised	Potential improvements through utilizing both labeled and unlabeled data	Complexity in designing algorithms that effectively leverage both labeled and unlabeled data, risk of error propagation from weak labels
Weakly Supervised Learning	Large amount of weakly labeled data	Lower than fully supervised	Strong generalization	Promising due to ease of obtaining weak labels and better generalization	Difficulty in ensuring accuracy due to the noise or ambiguity present in weak labels, potential inconsistency in labeling quality
Semi-Weakly Supervised Learning	Combination of small amount of labeled data + large amount of weakly labeled data	Moderate accuracy	Strong generalization	Opportunity to harness the benefits of both labeled and weakly labeled data	Balancing accuracy from labeled data with generalization from weak labels, potential challenges in harmonizing the different types of labeled data

Table 5. The Performance Comparison of Models on the Massachusetts Dataset.

Method	Recall	Precision	F1-Score	OA	IoU	mIoU	Parameters (M)
SegRExt-A [67]	68.29	76.95	-	97.53	56.82	-	-
SegRExt-F [67]	63.84	74.88	-	96.62	52.85	-	-
MSPFE-Net [57]	75.50	73.11	74.29	-	59.09	-	-
LDANet [97]	97.07	97.55	97.31	-	68.34	-	0.20
SemiRoadExNet [150]	-	-	70.23	-	54.66	-	-
Seg-Road-I [70]	92.86	87.34	90.02	-	68.38	83.89	28.67
DU-Net [74]	96.96	97.48	96.72	-	-	67.05	-
SR [31]	77.50	80.41	78.93	-	-	65.30	-
MECA-Net [66]	78.19	80.63	79.39	-	65.82	-	-
GA-Net [130]	76.89	84.10	80.33	-	67.13	-	
SDG-DenseNet [61]	77.67	81.86	79.63	-	66.47	-	265.00
SDUNet [96]	75.70	81.20	78.40	-	74.10	-	80.24
MUNet [138]	-	-	67.40	97.20	-	74.00	-
U-Net++ + Resnext [152]	95.10	94.30	94.70	-		-	
Deep residual U-Net [153]	80.00	84.00	81.00	-	72.00	-	-
CM-FCN [82]	77.87	79.45	78.65	97.98	67.55	-	56.45
CRAE-Net [83]	79.35	80.04	79.52	-	66.27	-	49.18
SGCN [154]	73.91	84.82	78.99	-	81.65	65.28	42.73
Richer U-Net [131]	-	-	-	-	58.63	-	-
GDCNet [122]	71.21	84.43	-	-	62.94	-	-
ConSwin [69]	79.17	81.11	80.13	98.15	66.84	-	-
RALC-Net [1]	-	-	74.70	-		59.61	
RoadVecNet [136]	-	-	92.51	-	86.31	-	-

Table 5. *Cont.*

Method	Recall	Precision	F1-Score	OA	IoU	mIoU	Parameters (M)
MCG-UNet [119]	86.59	91.18	88.74	-	79.92	-	
AEML U-Nets [125]	76.33	81.06	78.62	-	64.77	-	-
RVgg19 [155]	91.02	84.98	87.90	-	-	-	-
CADUNet [78]	76.55	79.45	77.89	98.00	64.12	-	-
AF-Net [94]	-	-	-	-	67.25	-	-
E-UNet [118]	81.30	80.71	80.45	97.59	68.56	-	-
DCANet [93]	79.54	80.20	79.84	98.09	66.45	82.23	11.1
Deep FR TransNet [91]	78.13	83.72	-	97.48		62.86	-
Prop-GAN [7]	92.92	91.54	92.20	-		87.43	-
DGRN [56]	71.97	-	76.59	-	62.48	-	-
CNN-Based [126]	85.88	78.47	-	-	78.65	-	-
Nested SE-Deeplab [72]	-	85.80	85.70	96.70	73.87	-	-
DiResNet [124]	79.41	80.38	79.70	98.13	-	-	-
CDG [77]	71.80	81.41	76.10	-	61.90	-	-
VNet+CEDL [133]	-	-	91.18	-	83.82	-	-

ConSwin, DCANet, and DiResNet all have overall accuracy (OA) exceeding 98%. This high OA indicates that these models exhibit a very high level of accuracy in correctly classifying road and non-road pixels within the dataset they were evaluated on.

Prop-GAN, DCANet, and Seg-Road-I exhibit high mIoU, with Prop-GAN achieving the highest mIoU among these models. This signifies their robustness and precision in road extraction tasks, indicating their capability to accurately identify and extract road information.

In conclusion, we have provided a more detailed summary of the limitations and challenges associated with current models in the context of road extraction. The following points encapsulate our findings:

(1) Model Complexity vs. Inference Speed

Complex models generally confer superior accuracy, however, at the potential expense of increased computational overhead and a higher number of parameters during the inference phase. Looking forward, achieving a nuanced equilibrium between model complexity and predictive speed is imperative, particularly in the context of real-time applications for road extraction.

(2) Generalization vs. Specialization

When confronted with unfamiliar road data, models demonstrating excessive specialization may encounter challenges, while those characterized by an overly generalized nature may fail to comprehensively capture the nuanced complexities within specific road domains. Achieving a judicious balance is crucial for optimizing performance across diverse road scenarios.

(3) Interpretability vs. Model Performance

Simplified models are often prized for their interpretability, yet they may fall short of matching the performance of their more intricate counterparts. While road extraction may superficially appear as a straightforward binary classification task, certain deep neural networks—especially sophisticated architectures like the Transformer—are frequently characterized as "black-box" models. This characterization poses challenges in deciphering their decision making processes and assessing their suitability for deployment in binary classification tasks. Furthermore, we underscore the notion that employing overly complex models for ostensibly simple tasks might be construed as an instance of "overengineering". Therefore, meticulous consideration is warranted in the selection of models, navigating the delicate balance between interpretability and performance.

5. Prospects

Despite significant progress in the field of road extraction from remote sensing images in recent years, there are still some issues that require further research and development, summarized as follows:

(1) Obtaining High-Quality Labeled Sample Data

This can be addressed by employing semi-supervised and weakly supervised learning methods, combining limited labeled sample data with a large amount of unlabeled data. Although these methods may not achieve the same level of accuracy in road extraction as full supervision, they provide new approaches to addressing this challenge. Furthermore, we have observed that there is a relatively limited availability of open road datasets in complex mountainous terrains when organizing the dataset. Therefore, there is a need to further expand data resources in this regard.

(2) Differences in Spectral Information Due to Factors Such as Sensors and Solar Angles

Additionally, when dealing with challenges like road occlusion and complex background information, relatively simple neural networks can be employed to separate road and non-road areas in advance, thereby enhancing the robustness of the model in subsequent recognition tasks. However, it is worth noting that research in areas such as image denoising and super-high-resolution reconstruction remains relatively limited in the field of data enhancement.

(3) Utilizing Multi-Modal Data

Currently, the application of multi-modal data in road extraction research is relatively limited. Multi-spectral (hyperspectral) data provide us with rich spectral information, while SAR data compensate for the limitations of optical images when dealing with issues like vegetation occlusion. However, LiDAR data are distinctive, typically in the form of three-dimensional point cloud data, and there are significant differences in spatial representation compared to two-dimensional road data. Therefore, further research is needed in the area of data fusion. Scholars in this field have conducted relatively limited research, leaving room for further exploration in the future. With the continuous expansion of crowdsourced data and the advantages of GNSS and other trajectory data, which do not contain additional environmental information and have minimal interference, they have played a significant role when combined with optical images. This combination provides us with complementary information and effectively mitigates issues such as the loss of road intersection information and incomplete connections. In the future, crowdsourced datasets from platforms like Google, Amap, Didi, Baidu, and others will further support and assist road extraction.

(4) Optimization of Fully Supervised Learning Models

From generative adversarial networks (GANs) to conditional generative adversarial networks (CGANs), and from unsupervised learning to supervised learning, these advancements all emphasize the advantages of supervised learning in road feature extraction to achieve more ideal road extraction results. Models based on the encoder–decoder structure are still a popular research direction in the current deep learning field. Introducing attention mechanism modules in different structures, achieving multi-scale feature fusion, considering the introduction of Transformer, GCNs, and deep convolutional separation structures, and even introducing corresponding loss functions based on the model's characteristics during the training process all contribute to improving the model's road feature extraction performance in images. As models move towards greater depth and width, an increase in model size may lead to an excess of parameters, thereby raising training costs. Therefore, seeking lighter, more efficient, and more highly generalizable models becomes an important direction for future research.

(5) Optimization of Semi-Supervised (Weak) Learning Models

With the emergence of semi-supervised (weak) learning, we have successfully overcome the challenges of high costs and the difficulty of obtaining labels by using a small amount of labeled data and a large amount of weakly labeled annotation data. We have employed various methods and strategies for model training, achieving training results approximating those of fully supervised learning. However, despite the significant progress made in semi-supervised and weakly supervised learning, there is still a substantial gap in accuracy when it comes to road extraction compared to fully supervised learning. Additionally, there is relatively limited research on models based on semi-weakly supervised learning. Therefore, future research directions should explore how to fully integrate the respective strengths of semi-supervised and weakly supervised learning to compensate for their shortcomings and build more powerful semi-weakly supervised models.

(6) Road Extraction Post-Processing

Road segmentation is not the end of road extraction. After road segmentation, there is still significant room for the post-processing of road extraction. This is because the quality of the model's extraction cannot be solely measured by high or low accuracy. Further observation is required to assess whether the connectivity of roads in the image is intact or if there are issues like fragmentation. Relevant post-processing methods can be used to repair damaged roads and improve the connectivity of poorly connected intersections. Additionally, attention should be given to specific tasks such as vectorization of roads, estimation of road areas, and registration of road features with aerial imagery. These tasks are of great significance to fields such as geographic information systems (GISs), urban road networks, and electronic map updates. Conditional generative adversarial networks (CGANs) can be applied not only to road extraction tasks but also provide new avenues for road extraction post-processing. By utilizing the differences between the generator and discriminator backbone models and additional conditions like adding noise and artifacts, they offer extensive opportunities for the future development of post-processing in this field.

Author Contributions: Conceptualization, Y.S.; investigation, S.M. and Q.Y.; writing—original draft preparation, S.M.; writing—review and editing, Y.S. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20201387 and QingLan Project of Jiangsu Province (QL2021), China.

Data Availability Statement: Data will be made available on request.

Acknowledgments: The authors would like to thank the contributions of the editor and reviewers.

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

References

1. Liu, Z.; Wang, M.; Wang, F.; Ji, X. A Residual Attention and Local Context-Aware Network for Road Extraction from High-Resolution Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 4958. [CrossRef]
2. Li, P.; He, X.; Qiao, M.; Miao, D.; Cheng, X.; Song, D.; Chen, M.; Li, J.; Zhou, T.; Guo, X.; et al. Exploring Multiple Crowdsourced Data to Learn Deep Convolutional Neural Networks for Road Extraction. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102544. [CrossRef]
3. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-of-the-Art Review. *Remote Sens.* **2020**, *12*, 1444. [CrossRef]
4. Wei, Y.; Wang, Z.; Xu, M. Road Structure Refined CNN for Road Extraction in Aerial Image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 709–713. [CrossRef]
5. Abdollahi, A.; Pradhan, B.; Shukla, N. Extraction of Road Features from UAV Images Using a Novel Level Set Segmentation Approach. *Int. J. Urban Sci.* **2019**, *23*, 391–405. [CrossRef]
6. Xin, J.; Zhang, X.; Zhang, Z.; Fang, W. Road Extraction of High-Resolution Remote Sensing Images Derived from DenseUNet. *Remote Sens.* **2019**, *11*, 2499. [CrossRef]
7. 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]

8. 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]
9. Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: Active Contour Models. *Int. J. Comput. Vis.* **1988**, *1*, 321–331. [CrossRef]
10. Miao, Z.; Wang, B.; Shi, W.; Zhang, H. A Semi-Automatic Method for Road Centerline Extraction from Vhr Images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1856–1860. [CrossRef]
11. Gruen, A.; Li, H. Road Extractionfrom Aerialand Satellite Images by Dynamic Programming. *ISPRS J. Photogrammetty Remote Sens.* **1995**, *50*, 11–20. [CrossRef]
12. Park, S.-R.; Kim, T. Semi-Automatic Road Extraction Algorithm from IKONOS Images Using Template Matching. In Proceedings of the 22nd Asian Conference on Remote Sensing, Singapore, 5–9 November 2001.
13. Yager, N.; Sowmya, A. Support Vector Machines for Road Extraction from Remotely Sensed Images. In *Computer Analysis of Images and Patterns*; Petkov, N., Westenberg, M.A., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2756, pp. 285–292. ISBN 978-3-540-40730-0.
14. Zhang, J.; Chen, L.; Zhuo, L.; Geng, W.; Wang, C. Multiple Saliency Features Based Automatic Road Extraction from High-resolution Multispectral Satellite Images. *Chin. J. Electron.* **2018**, *27*, 133–139. [CrossRef]
15. Yousefi, B.; Mirhassani, S.M.; Marvi, H. Classification of Remote Sensing Images from Urban Areas Using Laplacian Image and Bayesian Theory. In Proceedings of the International Symposium on Optomechatronic Technologies, Lausanne, Switzerland, 8–10 October 2007; Kofman, J., Lopez De Meneses, Y., Kaneko, S., Perez, C.A., Coquin, D., Eds.; SPIE: Bellingham, WA, USA, 2007; p. 67180F.
16. Karaman, E.; Çinar, U.; Gedik, E.; Yardımcı, Y.; Halıcı, U. A New Algorithm for Automatic Road Network Extraction in Multispectral Satellite Images. In Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil, 7–9 May 2012.
17. Manandhar, P.; Marpu, P.R.; Aung, Z. Segmentation Based Traversing-Agent Approach for Road Width Extraction from Satellite Images Using Volunteered Geographic Information. *Appl. Comput. Inform.* **2018**, *17*, 131–152. [CrossRef]
18. Tan, Y.-Q.; Gao, S.-H.; Li, X.-Y.; Cheng, M.-M.; Ren, B. VecRoad: Point-Based Iterative Graph Exploration for Road Graphs Extraction. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 8907–8915.
19. Jia, J.; Sun, H.; Jiang, C.; Karila, K.; Karjalainen, M.; Ahokas, E.; Khoramshahi, E.; Hu, P.; Chen, C.; Xue, T.; et al. Review on Active and Passive Remote Sensing Techniques for Road Extraction. *Remote Sens.* **2021**, *13*, 4235. [CrossRef]
20. Liu, P.; Wang, Q.; Yang, G.; Li, L.; Zhang, H. Survey of Road Extraction Methods in Remote Sensing Images Based on Deep Learning. *PFG—J. Photogramm. Remote Sens. Geoinf. Sci.* **2022**, *90*, 135–159. [CrossRef]
21. Mnih, V. Machine Learning for Aerial Image Labeling. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada, 2013.
22. Cheng, G.; Wang, Y.; Xu, S.; Wang, H.; Xiang, S.; Pan, C. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3322–3337. [CrossRef]
23. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 172–17209.
24. Van Etten, A.; Lindenbaum, D.; Bacastow, T.M. SpaceNet: A Remote Sensing Dataset and Challenge Series. *arXiv* **2019**, arXiv:1807.01232.
25. 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 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 4720–4728.
26. Liu, Y.; Yao, J.; Lu, X.; Xia, M.; Wang, X.; Liu, Y. RoadNet: Learning to Comprehensively Analyze Road Networks in Complex Urban Scenes from High-Resolution Remotely Sensed Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2043–2056. [CrossRef]
27. 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]
28. Ayala, C.; Aranda, C.; Galar, M. *Multi-Temporal Data Augmentation for High Frequency Satellite Imagery: A Case Study in Sentinel-1 and Sentinel-2 Building and Road Segmentation*; Jiang, J., Shaker, A., Zhang, H., Eds.; ISPRS: Nice, France, 2022; Volume 43, pp. 25–32.
29. Xu, Z.; Shen, Z.; Li, Y.; Xia, L.; Wang, H.; Li, S.; Jiao, S.; Lei, Y. Road Extraction in Mountainous Regions from High-Resolution Images Based on DSDNet and Terrain Optimization. *Remote Sens.* **2021**, *13*, 90. [CrossRef]
30. Zhang, T.; Dai, J.; Li, Y.; Zhang, Y. Vector Data Partition Correction Method Supported by Deep Learning. *Int. J. Remote Sens.* **2022**, *43*, 5603–5635. [CrossRef]
31. Han, L.; Hou, L.; Zheng, X.; Ding, Z.; Yang, H.; Zheng, K. Segmentation Is Not the End of Road Extraction: An All-Visible Denoising Autoencoder for Connected and Smooth Road Reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4403818. [CrossRef]
32. Mnih; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al. Learning to Detect Roads in High-Resolution Aerial Images. In *Computer Vision—ECCV 2010*; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6316, pp. 210–223. ISBN 978-3-642-15566-6.
33. Wang, J.; Song, J.; Chen, M.; Yang, Z. Road Network Extraction: A Neural-Dynamic Framework Based on Deep Learning and a Finite State Machine. *Int. J. Remote Sens.* **2015**, *36*, 3144–3169. [CrossRef]

34. Rezaee, M.; Zhang, Y. Road Detection Using Deep Neural Network in High Spatial Resolution Images. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, United Arab Emirates, 6–8 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–4.
35. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *arXiv* **2015**, arXiv:1411.4038.
36. Badrinarayanan, V.; Handa, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling. *arXiv* **2015**, arXiv:1505.07293.
37. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
38. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. *arXiv* **2017**, arXiv:1612.01105.
39. 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.
40. 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.
41. Chen, Z.; Wang, C.; Li, J.; Xie, N.; Han, Y.; Du, J. Reconstruction Bias U-Net for Road Extraction from Optical Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2284–2294. [CrossRef]
42. 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 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 192–1924.
43. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *arXiv* **2016**, arXiv:1606.04797.
44. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *arXiv* **2018**, arXiv:1807.10165.
45. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection. *Pattern Recognit.* **2020**, *106*, 107404. [CrossRef]
46. Cao, Y.; Liu, S.; Peng, Y.; Li, J. DenseUNet: Densely Connected UNet for Electron Microscopy Image Segmentation. *IET Image Process.* **2020**, *14*, 2682–2689. [CrossRef]
47. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
48. Chen, D.; Hu, F.; Mathiopoulos, P.T.; Zhang, Z.; Peethambaran, J. MC-UNet: Martian Crater Segmentation at Semantic and Instance Levels Using U-Net-Based Convolutional Neural Network. *Remote Sens.* **2023**, *15*, 266. [CrossRef]
49. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
50. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
51. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2016**, arXiv:1412.7062.
52. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *arXiv* **2017**, arXiv:1606.00915. [CrossRef]
53. Lan, M.; Zhang, Y.; Zhang, L.; Du, B. Global Context Based Automatic Road Segmentation via Dilated Convolutional Neural Network. *Inf. Sci.* **2020**, *535*, 156–171. [CrossRef]
54. Gao, C.; Gu, L.; Ren, R.; Jiang, M. *Deep Learning Combined with Topology and Channel Features for Road Extraction from Remote Sensing Images*; Butler, J., Xiong, X., Gu, X., Eds.; SPIE: Bellingham, WA, USA, 2022; Volume 12232.
55. Huan, H.; Sheng, Y.; Zhang, Y.; Liu, Y. Strip Attention Networks for Road Extraction. *Remote Sens.* **2022**, *14*, 4516. [CrossRef]
56. Wu, Q.; Luo, F.; Wu, P.; Wang, B.; Yang, H.; Wu, Y. Automatic Road Extraction from High-Resolution Remote Sensing Images Using a Method Based on Densely Connected Spatial Feature-Enhanced Pyramid. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 3–17. [CrossRef]
57. Wei, Z.; Zhang, Z. Remote Sensing Image Road Extraction Network Based on MSPFE-Net. *Electronics* **2023**, *12*, 1713. [CrossRef]
58. Gong, Z.; Xu, L.; Tian, Z.; Bao, J.; Ming, D. Road Network Extraction and Vectorization of Remote Sensing Images Based on Deep Learning. In Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; Xu, B., Mou, K., Eds.; IEEE: Piscataway, NJ, USA, 2020; pp. 303–307.
59. Wang, Q.; Bai, H.; He, C.; Cheng, J. FE-LinkNet: Enhanced D-LinkNet with Attention and Dense Connection for Road Extraction in High-Resolution Remote Sensing Images. In Proceedings of the IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 3043–3046.
60. Li, H.; Xiong, P.; An, J.; Wang, L. Pyramid Attention Network for Semantic Segmentation. *arXiv* **2018**, arXiv:1805.10180.
61. Zhang, J.; Li, Y.; Si, Y.; Peng, B.; Xiao, F.; Luo, S.; He, L. A Low-Grade Road Extraction Method Using SDG-DenseNet Based on the Fusion of Optical and SAR Images at Decision Level. *Remote Sens.* **2022**, *14*, 2870. [CrossRef]
62. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 4510–4520.

63. Chen, D.; Li, X.; Hu, F.; Mathiopoulos, P.T.; Di, S.; Sui, M.; Peethambaran, J. EDPNet: An Encoding–Decoding Network with Pyramidal Representation for Semantic Image Segmentation. *Sensors* **2023**, *23*, 3205. [CrossRef]
64. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *arXiv* **2019**, arXiv:1709.01507.
65. Gao, L.; Wang, J.; Wang, Q.; Shi, W.; Zheng, J.; Gan, H.; Lv, Z.; Qiao, H. Road Extraction Using a Dual Attention Dilated-LinkNet Based on Satellite Images and Floating Vehicle Trajectory Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10428–10438. [CrossRef]
66. Jie, Y.; He, H.; Xing, K.; Yue, A.; Tan, W.; Yue, C.; Jiang, C.; Chen, X. MECA-Net: A Multiscale Feature Encoding and Long-Range Context-Aware Network for Road Extraction from Remote Sensing Images. *Remote Sens.* **2022**, *14*, 5342. [CrossRef]
67. Bisio, I.; Garibotto, C.; Haleem, H.; Lavagetto, F.; Sciarrone, A. Traffic Analysis through Deep-Learning-Based Image Segmentation from UAV Streaming. *IEEE Internet Things J.* **2023**, *10*, 6059–6073. [CrossRef]
68. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
69. Chen, T.; Jiang, D.; Li, R. Swin Transformers Make Strong Contextual Encoders for VHR Image Road Extraction. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 3019–3022.
70. Tao, J.; Chen, Z.; Sun, Z.; Guo, H.; Leng, B.; Yu, Z.; Wang, Y.; He, Z.; Lei, X.; Yang, J. Seg-Road: A Segmentation Network for Road Extraction Based on Transformer and CNN with Connectivity Structures. *Remote Sens.* **2023**, *15*, 1602. [CrossRef]
71. Ding, C.; Weng, L.; Xia, M.; Lin, H. Non-Local Feature Search Network for Building and Road Segmentation of Remote Sensing Image. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 245. [CrossRef]
72. Lin, Y.; Xu, D.; Wang, N.; Shi, Z.; Chen, Q. Road Extraction from Very-High-Resolution Remote Sensing Images via a Nested SE-Deeplab Model. *Remote Sens.* **2020**, *12*, 2985. [CrossRef]
73. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521.
74. Kong, J.; Zhang, Y. DU-Net-Cloud: A Smart Cloud-Edge Application with an Attention Mechanism and U-Net for Remote Sensing Images and Processing. *J. Cloud Comput.-Adv. Syst. Appl.* **2023**, *12*, 1–14. [CrossRef]
75. Dong, S.; Chen, Z. Block Multi-Dimensional Attention for Road Segmentation in Remote Sensing Imagery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3137551. [CrossRef]
76. Xu, Y.; Chen, H.; Du, C.; Li, J. MSACon: Mining Spatial Attention-Based Contextual Information for Road Extraction. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3073923. [CrossRef]
77. Wang, S.; Yang, H.; Wu, Q.; Zheng, Z.; Wu, Y.; Li, J. An Improved Method for Road Extraction from High-Resolution Remote-Sensing Images That Enhances Boundary Information. *Sensors* **2020**, *20*, 2064. [CrossRef]
78. Li, J.; Liu, Y.; Zhang, Y.; Zhang, Y. Cascaded Attention denseUNet (CADUNet) for Road Extraction from Very-High-Resolution Images. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 329. [CrossRef]
79. Lu, X.; Zhong, Y.; Zheng, Z. A Novel Global-Aware Deep Network for Road Detection of Very High Resolution Remote Sensing Imagery. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 2579–2582.
80. Feng, D.; Shen, X.; Xie, Y.; Liu, Y.; Wang, J. Efficient Occluded Road Extraction from High-Resolution Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 4974. [CrossRef]
81. Lu, X.; Zhong, Y.; Zheng, Z.; Zhang, L. GAMSNet: Globally Aware Road Detection Network with Multi-Scale Residual Learning. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 340–352. [CrossRef]
82. Zhu, Y.; Long, L.; Wang, J.; Yan, J.; Wang, X. Road Segmentation from High-Fidelity Remote Sensing Images Using a Context Information Capture Network. *Cogn. Comput.* **2022**, *14*, 780–793. [CrossRef]
83. Li, S.; Liao, C.; Ding, Y.; Hu, H.; Jia, Y.; Chen, M.; Xu, B.; Ge, X.; Liu, T.; Wu, D. Cascaded Residual Attention Enhanced Road Extraction from Remote Sensing Images. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 9. [CrossRef]
84. Bai, X.; Guo, L.; Huo, H.; Zhang, J.; Zhang, Y.; Li, Z.-L. Rse-Net: Road-Shape Enhanced Neural Network for Road Extraction in High Resolution Remote Sensing Image. *Int. J. Remote Sens.* **2023**, *44*, 1–22. [CrossRef]
85. He, L.; Zhu, T.; Lv, M. Retracted: An Early Warning Intelligent Algorithm System for Forest Resource Management and Monitoring. *Comput. Intell. Neurosci.* **2023**, *2023*, 9853814. [CrossRef]
86. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. *arXiv* **2017**, arXiv:1612.03144.
87. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2014**, arXiv:1409.4842.
88. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. *arXiv* **2019**, arXiv:1902.09212.
89. Ren, Y.; Yu, Y.; Guan, H. DA-CapsUNet: A Dual-Attention Capsule U-Net for Road Extraction from Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 2866. [CrossRef]
90. Zhou, M.; Sui, H.; Chen, S.; Wang, J.; Chen, X. BT-RoadNet: A Boundary and Topologically-Aware Neural Network for Road Extraction from High-Resolution Remote Sensing Imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 288–306. [CrossRef]
91. Ge, Z.; Zhao, Y.; Wang, J.; Wang, D.; Si, Q. Deep Feature-Review Transmit Network of Contour-Enhanced Road Extraction from Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 3001805. [CrossRef]

92. Li, X.; Wang, Y.; Zhang, L.; Liu, S.; Mei, J.; Li, Y. Topology-Enhanced Urban Road Extraction via a Geographic Feature-Enhanced Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8819–8830. [CrossRef]
93. Hu, L.; Niu, C.; Ren, S.; Dong, M.; Zheng, C.; Zhang, W.; Liang, J. Discriminative Context-Aware Network for Target Extraction in Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *15*, 700–715. [CrossRef]
94. Zou, S.; Xiong, F.; Luo, H.; Lu, J.; Qian, Y. AF-Net: All-Scale Feature Fusion Network for Road Extraction from Remote Sensing Images. In Proceedings of the 2021 Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 29 November–1 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 66–73.
95. Zao, Y.; Chen, H.; Liu, L.; Shi, Z. Enhance Essential Features for Road Extraction from Remote Sensing Images. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 3023–3026.
96. Yang, M.; Yuan, Y.; Liu, G. SDUNet: Road Extraction via Spatial Enhanced and Densely Connected UNet. *Pattern Recognit.* **2022**, *126*, 108549. [CrossRef]
97. Liu, B.; Ding, J.; Zou, J.; Wang, J.; Huang, S. LDANet: A Lightweight Dynamic Addition Network for Rural Road Extraction from Remote Sensing Images. *Remote Sens.* **2023**, *15*, 1829. [CrossRef]
98. Cheng, B.; Tian, M.; Jiang, S.; Liu, W.; Pang, Y. Multi-Task Learning and Multimodal Fusion for Road Segmentation. *IEEE Access* **2023**, *11*, 18947–18959. [CrossRef]
99. Yan, J.; Chen, Y.; Zheng, J.; Guo, L.; Zheng, S.; Zhang, R. Multi-Source Time Series Remote Sensing Feature Selection and Urban Forest Extraction Based on Improved Artificial Bee Colony. *Remote Sens.* **2022**, *14*, 4859. [CrossRef]
100. Zhou, M.; Sui, H.; Chen, S.; Chen, X.; Wang, W.; Wang, J.; Liu, J. UGRoadUpd: An Unchanged-Guided Historical Road Database Updating Framework Based on Bi-Temporal Remote Sensing Images. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21465–21477. [CrossRef]
101. Wu, H.; Zhang, H.; Zhang, X.; Sun, W.; Zheng, B.; Jiang, Y. DeepDualMapper: A Gated Fusion Network for Automatic Map Extraction Using Aerial Images and Trajectories. Machine Learning for Aerial Image Labeling. *arXiv* **2020**, arXiv:2002.06832.
102. Li, Y.; Xiang, L.; Zhang, C.; Jiao, F.; Wu, C. A Guided Deep Learning Approach for Joint Road Extraction and Intersection Detection from Rs Images and Taxi Trajectories. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8008–8018. [CrossRef]
103. Liu, L.; Yang, Z.; Li, G.; Wang, K.; Chen, T.; Lin, L. Aerial Images Meet Crowdsourced Trajectories: A New Approach to Robust Road Extraction. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 3308–3322. [CrossRef]
104. 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]
105. Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. *arXiv* **2015**, arXiv:1409.2329.
106. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661. [CrossRef]
107. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
108. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 5967–5976.
109. Yang, C.; Wang, Z. An Ensemble Wasserstein Generative Adversarial Network Method for Road Extraction from High Resolution Remote Sensing Images in Rural Areas. *IEEE Access* **2020**, *8*, 174317–174324. [CrossRef]
110. Cira, C.-I.; Kada, M.; Manso-Callejo, M.; Alcarria, R.; Bordel Sanchez, B. Improving Road Surface Area Extraction via Semantic Segmentation with Conditional Generative Learning for Deep Inpainting Operations. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 43. [CrossRef]
111. Cira, C.-I.; Manso-Callejo, M.; Alcarria, R.; Fernandez Pareja, T.; Bordel Sanchez, B.; Serradilla, F. Generative Learning for Postprocessing Semantic Segmentation Predictions: A Lightweight Conditional Generative Adversarial Network Based on Pix2pix to Improve the Extraction of Road Surface Areas. *Land* **2021**, *10*, 79. [CrossRef]
112. Chen, W.; Zhou, G.; Liu, Z.; Li, X.; Zheng, X.; Wang, L. NIGAN: A Framework for Mountain Road Extraction Integrating Remote Sensing Road-Scene Neighborhood Probability Enhancements and Improved Conditional Generative Adversarial Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3188908. [CrossRef]
113. Senthilnath, J.; Varia, N.; Dokania, A.; Anand, G.; Benediktsson, J.A. Deep TEC: Deep Transfer Learning with Ensemble Classifier for Road Extraction from UAV Imagery. *Remote Sens.* **2020**, *12*, 245. [CrossRef]
114. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *arXiv* **2020**, arXiv:1703.10593.
115. Cira, C.-I.; Alcarria, R.; Manso-Callejo, M.-A.; Serradilla, F. A Framework Based on Nesting of Convolutional Neural Networks to Classify Secondary Roads in High Resolution Aerial Orthoimages. *Remote Sens.* **2020**, *12*, 765. [CrossRef]
116. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Learning. *arXiv* **2016**, arXiv:1602.07261.
117. Chen, Z.; Fan, W.; Zhong, B.; Li, J.; Du, J.; Wang, C. Corse-to-Fine Road Extraction Based on Local Dirichlet Mixture Models and Multiscale-High-Order Deep Learning. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4283–4293. [CrossRef]
118. Li, J.; Meng, Y.; Dorjee, D.; Wei, X.; Zhang, Z.; Zhang, W. Automatic Road Extraction from Remote Sensing Imagery Using Ensemble Learning and Postprocessing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10535–10547. [CrossRef]
119. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Multi-Object Segmentation in Complex Urban Scenes from High-Resolution Remote Sensing Data. *Remote Sens.* **2021**, *13*, 3710. [CrossRef]

120. Song, H.; Wang, W.; Zhao, S.; Shen, J.; Lam, K.-M. Pyramid Dilated Deeper ConvLSTM for Video Salient Object Detection. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 11215, pp. 744–760. ISBN 978-3-030-01251-9.
121. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2018**, arXiv:1608.06993.
122. Cui, F.; Shi, Y.; Feng, R.; Wang, L.; Zeng, T. A Graph-Based Dual Convolutional Network for Automatic Road Extraction from High Resolution Remote Sensing Images. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 3015–3018.
123. Sun, Z.; Zhou, W.; Ding, C.; Xia, M. Multi-Resolution Transformer Network for Building and Road Segmentation of Remote Sensing Image. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 165. [CrossRef]
124. 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]
125. Chen, Z.; Wang, C.; Li, J.; Fan, W.; Du, J.; Zhong, B. Adaboost-like End-to-End Multiple Lightweight U-Nets for Road Extraction from Optical Remote Sensing Images. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102341. [CrossRef]
126. 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]
127. Shao, Z.; Zhou, Z.; Huang, X.; Zhang, Y. MRENet: Simultaneous Extraction of Road Surface and Road Centerline in Complex Urban Scenes from Very High-Resolution Images. *Remote Sens.* **2021**, *13*, 239. [CrossRef]
128. Lu, X.; Zhong, Y.; Zheng, Z.; Chen, D.; Su, Y.; Ma, A.; Zhang, L. Cascaded Multi-Task Road Extraction Network for Road Surface, Centerline, and Edge Extraction. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3165817. [CrossRef]
129. Chen, D.; Zhong, Y.; Zheng, Z.; Ma, A.; Lu, X. Urban Road Mapping Based on an End-to-End Road Vectorization Mapping Network Framework. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 345–365. [CrossRef]
130. Chen, X.; Sun, Q.; Guo, W.; Qiu, C.; Yu, A. GA-Net: A Geometry Prior Assisted Neural Network for Road Extraction. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103004. [CrossRef]
131. Zao, Y.; Shi, Z. Richer U-Net: Learning More Details for Road Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3081774. [CrossRef]
132. Fan, J.; Yang, Z. Deep Residual Network Based Road Detection Algorithm for Remote Sensing Images. In Proceedings of the 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin, China, 25–27 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1723–1726.
133. Abdollahi, A.; Pradhan, B.; Alamri, A. VNet: An End-to-End Fully Convolutional Neural Network for Road Extraction from High-Resolution Remote Sensing Data. *IEEE Access* **2020**, *8*, 179424–179436. [CrossRef]
134. Akhtar, N.; Mandloi, M. DenseResSegnet: A Dense Residual Segnet for Road Detection Using Remote Sensing Images. In Proceedings of the 2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), Hyderabad, India, 27–29 January 2023; Volume 1, pp. 1–4.
135. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2018**, arXiv:1708.02002.
136. Abdollahi, A.; Pradhan, B.; Alamri, A. RoadVecNet: A New Approach for Simultaneous Road Network Segmentation and Vectorization from Aerial and Google Earth Imagery in a Complex Urban Set-Up. *Giscience Remote Sens.* **2021**, *58*, 1151–1174. [CrossRef]
137. Sushma, B.; Fatimah, B.; Raj, P. Road Segmentation in Aerial Imagery by Deep Neural Networks with 4-Channel Inputs. In Proceedings of the 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 25–27 March 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 340–344.
138. Yuan, W.; Xu, W. GapLoss: A Loss Function for Semantic Segmentation of Roads in Remote Sensing Images. *Remote Sens.* **2022**, *14*, 2422. [CrossRef]
139. Xu, H.; He, H.; Zhang, Y.; Ma, L.; Li, J. A Comparative Study of Loss Functions for Road Segmentation in Remotely Sensed Road Datasets. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103159. [CrossRef]
140. Zhang, Y.; Zhu, Q.; Zhong, Y.; Guan, Q.; Zhang, L.; Li, D. A Modified D-LinkNet with Transfer Learning for Road Extraction from High-Resolution Remote Sensing. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1817–1820.
141. Zhou, Z.-H. A Brief Introduction to Weakly Supervised Learning. *Natl. Sci. Rev.* **2018**, *5*, 44–53. [CrossRef]
142. Lian, R.; Huang, L. DeepWindow: Sliding Window Based on Deep Learning for Road Extraction from Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1905–1916. [CrossRef]
143. Newell, A.; Yang, K.; Deng, J. Stacked Hourglass Networks for Human Pose Estimation. *arXiv* **2016**, arXiv:1603.06937.
144. Lian, R.; Huang, L. Weakly Supervised Road Segmentation in High-Resolution Remote Sensing Images Using Point Annotations. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3059088. [CrossRef]
145. Wei, Y.; Ji, S. Scribble-Based Weakly Supervised Deep Learning for Road Surface Extraction from Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 3061213. [CrossRef]
146. Zhou, M.; Sui, H.; Chen, S.; Liu, J.; Shi, W.; Chen, X. Large-Scale Road Extraction from High-Resolution Remote Sensing Images Based on a Weakly-Supervised Structural and Orientational Consistency Constraint Network. *ISPRS J. Photogramm. Remote Sens.* **2022**, *193*, 234–251. [CrossRef]

147. Xiao, R.; Wang, Y.; Tao, C. Fine-Grained Road Scene Understanding from Aerial Images Based on Semisupervised Semantic Segmentation Networks. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3059708. [CrossRef]
148. He, Y.; Wang, J.; Liao, C.; Shan, B.; Zhou, X. ClassHyPer: Classmix-Based Hybrid Perturbations for Deep Semi-Supervised Semantic Segmentation of Remote Sensing Imagery. *Remote Sens.* **2022**, *14*, 879. [CrossRef]
149. You, Z.-H.; Wang, J.-X.; Chen, S.-B.; Tang, J.; Luo, B. FMWDCT: Foreground Mixup into Weighted Dual-Network Cross Training for Semisupervised Remote Sensing Road Extraction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5570–5579. [CrossRef]
150. 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]
151. Chen, H.; Peng, S.; Du, C.; Li, J.; Wu, S. SW-GAN: Road Extraction from Remote Sensing Imagery Using Semi-Weakly Supervised Adversarial Learning. *Remote Sens.* **2022**, *14*, 4145. [CrossRef]
152. Yerram, V.; Takeshita, H.; Iwahori, Y.; Hayashi, Y.; Bhuyan, M.K.; Fukui, S.; Kijsirikul, B.; Wang, A. Extraction and Calculation of Roadway Area from Satellite Images Using Improved Deep Learning Model and Post-Processing. *J. Imaging* **2022**, *8*, 124. [CrossRef] [PubMed]
153. Ozturk, O.; Isik, M.S.; Sariturk, B.; Seker, D.Z. Generation of Istanbul Road Data Set Using Google Map API for Deep Learning-Based Segmentation. *Int. J. Remote Sens.* **2022**, *43*, 2793–2812. [CrossRef]
154. Zhou, G.; Chen, W.; Gui, Q.; Li, X.; Wang, L. Split Depth-Wise Separable Graph-Convolution Network for Road Extraction in Complex Environments from High-Resolution Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3128033. [CrossRef]
155. Li, P.; He, X.; Qiao, M.; Cheng, X.; Li, Z.; Luo, H.; Song, D.; Li, D.; Hu, S.; Li, R.; et al. Robust Deep Neural Networks for Road Extraction from Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6182–6197. [CrossRef]