Deep Learning-Based Methods for Road Extraction From Remote Sensing Images

A vision, survey, and future directions

XUAN WANG, XIZHI JIN, ZHE DAI, YUXUAN WU, AND ABDELLAH CHEHRI

xxxxx

INTRODUCTION

With the continuous development of hardware technologies in remote sensing and aerospace industries, the spatial resolution of remote sensing images is increasing, and these high-resolution remote sensing images are widely used in the applications of remote sensing technology. How to rapidly and accurately extract the target features of interest from high-resolution images captured by remote sensing technology has become an essential area for research. Roads play a fundamental role in people's daily travel and transportation, serving as a crucial infrastructure that has supported human civilization in numerous ways. With the rapid development of urbanization, there is a growing need to update a significant amount of road information in terms of cartography, map updating, and sustainable urban planning. The traditional method of updating maps is inefficient and relies heavily on manual observation and long-term surveys. The results of road extraction can be utilized to help unmanned systems choose the best route for real-time path planning. This can be achieved by integrating the traffic management system and road data information [1], [2], [3].

Remote sensing images are of utmost importance in environmental monitoring and are indispensable for attaining sustainable development goals. As an illustrative example, satellite and unmanned aerial vehicle (UAV) imagery have been employed to monitor natural ecosystems. Climate change is widely recognized as having a detrimental effect, as it will necessitate increased investments in road networks and infrastructure repair and upgrades. Using remote sensing images in road planning can be beneficial in reducing the need for costly retrofitting and lowering upfront expenses.

On the other hand, roads are essential geographic indicators in remote sensing imagery. This is particularly evident in natural disasters, such as earthquakes, floods, hurricanes, tropical storms, and wildfires, resulting in damage or disruptions to road infrastructure. Therefore, it is crucial to identify

Digital Object Identifier 10.1109/MGRS.2024.3491014

the most appropriate rescue route promptly. In recent times, synchronized satellites and UAV technology advancements have greatly facilitated the rapid collection of road information in hazardous regions [4]. Road extraction has benefited from using high-resolution remote sensing images and UAV images. These images contain significant information regarding different features, including road material, road undulation, and road distribution. They offer valuable data and technical support for assessing road accessibility and monitoring structural health [5].

High-resolution images captured by remote sensing technology can be a valuable and cost-effective data source for extracting road information in hazardous or inaccessible regions. These images provide an objective and safe means of gathering data in such areas [6]. The use of road databases helps in creating and managing geospatial databases by incorporating large amounts of high-resolution imagery collected through remote sensing technology [7]. However, this task remains challenging, as roads blocked by trees, shadows, and buildings are difficult to extract. In the meantime, there are significant variations in different types of roads, with little difference between buildings and roads, resulting in insufficient road background in these images.

In the early stage of the research on road extraction from remote sensing images, researchers mainly focused on methods based on morphology [8] and machine learning [9], [10] to extract roads. Due to the lack of abundant computing power, these traditional manual methods can only handle grey-scale images, which are incapable of semantic segmentation, suffer from high costs, are time-consuming, and have large amounts of errors caused by human operation [11]. In addition, these traditional methods involving mathematical morphology and texture analysis have proven inefficient in dealing with multiscale roads [12], [13], [14].

However, with the advent of deep learning, the field of road extraction from remote sensing images has witnessed significant advancements. Models based on this technique have demonstrated remarkable results in feature extraction and segmentation of remote sensing images. As a result, the application of these models in road extraction has become a prominent area of research in the field of remote sensing.

Researchers have made considerable progress in remote sensing techniques and road extraction using DNNs [15], [16], [17], [18], [19]. These methods are often trained end-to-end by semantic segmentation, extracting information image features from annotated data. For example, Alshehhi et al. [20] proposed a single block-based convolutional neural network (CNN) to extract roads and buildings from remote sensing images. Then, Costea et al. [21] proposed a double-hop generative adversarial network (GAN) to detect roads. Zhou et al. [22] proposed a semantic segmentation neural network D-LinkNet based on LinkNet [23] by adding a dilation convolutional layer in the middle of the encoder and decoder. Due to the excellence of deep learning, the exploration of road extraction methods from remote sensing images is getting deeper. Figure 1 shows the key developments in the field of road extraction.

Previously, several reviews have analyzed deep learning techniques for extracting roads from remote sensing images [7], [13], [14], [24], [25]. However, only four methods were classified in [7]: piecewise-based CNNs, fully convolutional networks (FCNs), deconvolutional neural networks, and GANs.

On top of this, weakly supervised and unsupervised learning methods are introduced in [25], but new techniques, such as transformer-based networks and graph convolution-based networks (GCNs) are not introduced. This review adds to and improves on the aforementioned

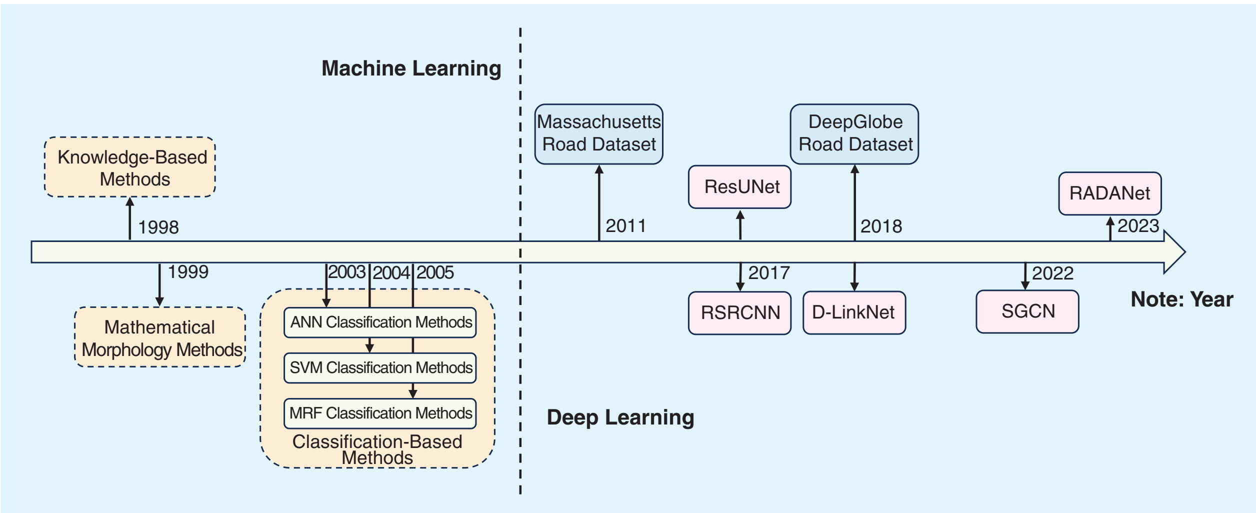

FIGURE 1. Key technologies in the development of road extraction. ANN: artificial neural network; MRF: Markov random field; RSRCNN: road structure refinement CNN; SVM: support vector machine.

research results, and the overview of this article is shown as Figure 2.

The main contributions of this article are as follows:

- The present article gives a complete review of the task of remote sensing imaging road extraction using deep learning. Particularly, these methods are mainly applied to optical images and very high-resolution (VHR) images. This includes problem specification, main datasets, and conventional evaluation methodologies, providing a thorough foundation for this review.
- This article presents a taxonomy of road extraction algorithms for remote sensing images based on the numerous available learning methodologies.
- This review also looks at how well typical centralized remote sensing image road extraction algorithms work on benchmark datasets and presents remote sensing image road extraction approaches proposed in recent years.
- The visual effects of the remote sensing corps' images using traditional road extraction methods are described.
- This review analyzes the problems and challenges faced by remote sensing image road extraction from multiple perspectives and makes valuable suggestions to clarify its future development.

The remainder of this review is structured as follows. The "Datasets and Evaluation Methods" section briefly discusses the datasets and evaluation metrics commonly used for remote sensing image road extraction based on deep learning. The "State-of-the-Art Methods for Road Extraction" section describes representative DNN architectures for road extraction tasks in detail. In the "Comparison and Discussion" section, several evaluation metrics are used to compare the performance of remote sensing image road extraction mentioned in the "State-of-the-Art Methods for Road Extraction" section. The "Current Challenges and Future Work" section discusses the current challenges of research for social responsibility and its possible development and the article concludes with a summary of research work in the "Conclusions" section.

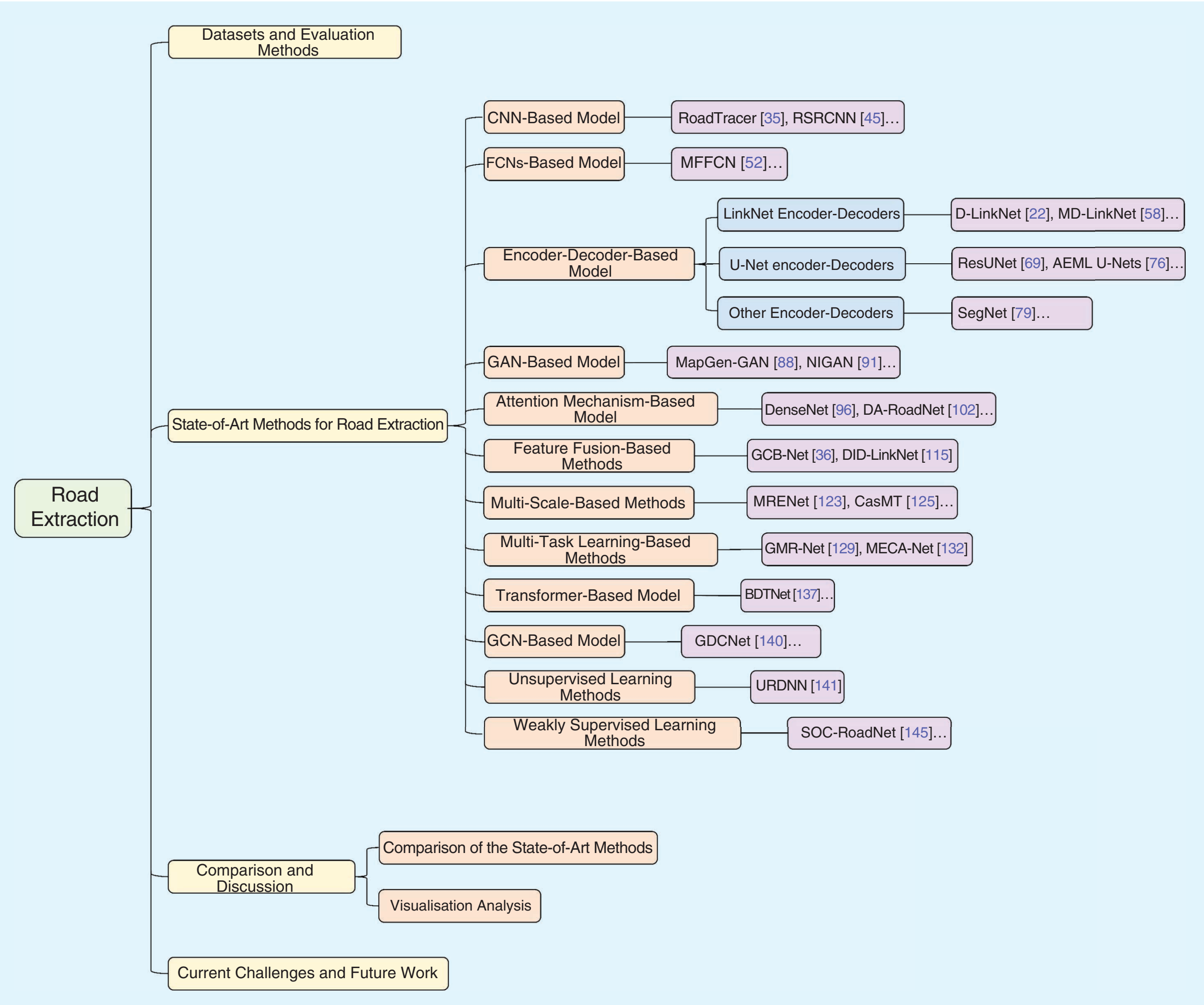

FIGURE 2. Hierarchically structured classification of road extraction in this article. AEML: Adaboost-based end-to-end multiple lightweight; MFFCN: multifeature FCN; RSRCNN: road structure refinement CNN; URDNN: unsupervised restricted deconvolutional neural network.

DATASETS AND EVALUATION METHODS

TYPICAL DATASETS

Datasets have played an indispensable role in the development of the field of road extraction from remote sensing images. On the one hand, datasets push researchers to address the possibility of road extraction in different scenarios. On the other hand, classical datasets have become a common basis for performance evaluation in road extraction. In the past decade, several classical datasets have been released that effectively contribute to the development of road extraction, as shown in Table 1.

1) *Massachusetts Roads Dataset [26]:* This dataset consists of 1,171 aerial images of Massachusetts in the United States. Each image is 1,500 × 1,500 pixels in size. The dataset was split into a training set of 1,108 images, a validation set of 14 images, and a test set of 49 images.
2) *DeepGlobe Roads Dataset [27]:* This dataset contains data with pixel-level annotations from Thailand, India, and Indonesia. The ground resolution of each image is 50 cm/pixel, and the pixel resolution is 1,024 × 1,024. As the labels of the test set are not given, researchers often divide the original training set into new training and test sets.
3) *SpaceNet Roads Dataset [31]:* The dataset encompasses four distinct areas: Las Vegas, Paris, Shanghai, and Khartoum. It comprises several labeled scenes, each of size 200 m × 200 m (650 px × 650 px), containing a total of 302,701 building footprints distributed across both urban and suburban settings. The dataset was divided into three subsets: 60% for training, 20% for testing, and 20% for validation. Each area is represented by a single image strip, ensuring consistent sun, satellite, and atmospheric conditions across the entire scene.
4) *CHN6-CUG Roads Dataset [36]:* This dataset is acquired from Google Earth. The images are selected from six cities with different levels of urbanization. Each image has a size of 512 × 512 pixels and a resolution of 50 cm/pixel. Finally, 4,511 labeled images are split into 3,608 images for training and 903 for testing.
5) *Gansu Mountain Road Dataset [39]:* The Gansu dataset is mainly derived from the mountainous areas of southern Gansu, China. The dataset consists of 255 images with an image size of 256 × 256, and the training, validation, and test sets consist of 204, 40, and 11 images, respectively.

EVALUATION METHODS

In this section, the extraction of road regions in remote sensing images, i.e., the pixel-level binary classification problem with positive samples of road pixels and negative samples of background pixels in semantic segmentation, is divided into four categories.

True positive (TP) indicates the number of correctly classified road pixels, false positive (FP) is the number of background pixels misclassified as road pixels, true negative (TN) indicates the number of correctly extracted background pixels, and false negative (FN) is the number of road pixels extracted as background pixels. Based on the aforementioned metric values, the most commonly used metrics to evaluate road extraction methods are intersection over union (IoU), precision, recall, F1 score, floating point operations (FLOPs), and the number of parameters (Params).

1) *IoU:* IoU is a common metric for semantic segmentation and target detection, which indicates the ratio of the overlapping area between ground truth and predicted area to the total area, quantifying the fit between the extracted results and the true label. The larger the value

TABLE 1. COMMON DATASETS.

DATASETS	YEAR	IMAGE SIZE	TRAIN SET	VAL SET	TEST SET	TOTAL SET	SPATIAL RESOLUTION/M
Massachusetts [26]	2013	1,500 × 1,500	1,108	14	49	1,171	1
DeepGlobe [27]	2018	1,024 × 1,024	6,226	243	1,101	8,570	0.5
Gaofen-2 [28]	2018	512 × 512	36,000	–	4,000	40,000	0.8
WHU [29]	2020	–	–	–	–	–	0.8–2
LRSNY [30]	2021	1,000 × 1,000	716	220	432	1,368	0.5
SpaceNet [31]	2018	650 × 650	35,392	220	432	1,368	0.5
Abu Dhabi [20]	2017	41,411 × 31,894	150	30	30	8,570	0.5
Conghua [32]	2019	6,000 × 6,000	37.6	–	9.4	47	0.2
CNDS [33]	2017	600 × 600	180	14	30	224	1.2
RNBD [34]	2019	–	14 Regions	1 Region	6 Regions	21 Regions	0.2
RTDS [35]	2018	4,096 × 4,096	–	–	–	300	–
CHN6-CUG [36]	2021	512 × 512	3,608	–	903	4,511	0.5
ShaoShan [37]	2020	1,589 × 1,131	29	–	20	49	0.5
GE-road [38]	2020	800 × 800	12,000	1,000	7,000	20,000	0.3–0.6
Gansu [39]	2021	256 × 256	204	40	11	255	2
RDCME [40]	2022	256 × 256	–	–	–	775	0.6

of IoU, the more overlapping areas of the result. The IoU is defined as follows:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \tag{1}$$

2) *Precision:* Precision indicates how many positive samples are predicted to be positive. It represents the proportion of correct matches in the road extraction results

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{2}$$

3) *Recall:* Recall indicates the number of correctly predicted positive samples in the sample. It indicates the proportion of matched pixels in ground truth road conditions

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{3}$$

4) *F1 score:* F1 score is a broad measure of the accuracy of a dichotomous model. It represents the measure of the harmonic mean of Recall and Precision. The F1 score is defined as follows:

$$\mathrm{F1\ score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{4}$$

The ranges of precision, recall, and F1 score are 0–1, and the higher the value, the better the performance of the two-class model.

5) *FLOPs:* The number of floating point calculations which can be used to measure the complexity of the model.
6) *Params:* The total number of parameters to be trained in the network model.

STATE-OF-ART METHODS FOR ROAD EXTRACTION

CNN-BASED MODEL

CNN is capable of extracting routes from remote sensing images. CNN classifiers are initially trained on small blocks, then specific sliding blocks are used, and ultimately, all labeled blocks are combined to generate a map of the entire image. In this section, we explore the extraction of roads from remote sensing images using CNN models, as shown in Figure 3.

In a groundbreaking contribution, Minh and Hinton [41] pioneered the integration of CNN models into remote sensing image road extraction. The road extraction performance experienced notable improvement by incorporating unsupervised learning methods for initializing the feature detectors and enhancing the local spatial coherence of the output labels. On this basis, restricted Boltzmann machine [26] and finite state machine [42] were combined with DNN, respectively, to demonstrate the feasibility of the CNN model in remote sensing image road extraction.

Saito et al. [43] combined a CNN and channelized inhibitory softmax to propose a new semantic segmentation model for extracting buildings and roads from aerial images. This model takes the initial pixels as input and predicts and outputs images of buildings and roads through feature extraction and classification.

Li et al. [44] used CNNs to capture local information on roads, combined with a smoothing algorithm based online integral convolution to extract the centerline of roads.

Wei et al. [45] proposed a road structure refinement CNN (RSRCNN) based on the VGG [46] model. In addition, a loss function based on the minimum Euclidean distance was proposed to effectively improve the aerial image road extraction by embedding the geometric structure information of the road.

Gao et al. [47] proposed a modified deep residual CNN (RDRCNN) based on an optical satellite image for road extraction. The RDRCNN model combines the symmetric structure of U-Net with residual connected units and integrates the occluded misclassified regions using a blind voting method.

Bastani et al. [35] offered a new method, RoadTracer, to construct road network graphs. RoadTracer effectively mitigates the effect of noise on CNN models by iterating through the CNN framework to decide on different roads to build the network.

Abdollahi et al. [48] presented an approach based on a CNN to extract remote sensing image roads by combining principal component analysis and object-based image analysis. Principal component analysis was used to replace irrelevant parameters to improve classification, and object-based image analysis used spectral and spatial information to improve intelligent decision making.

Although researchers have tried to continuously introduce various machine-learning methods in CNN models, in practical applications, the results of image extraction still have disadvantages, such as a lot of noise and low accuracy.

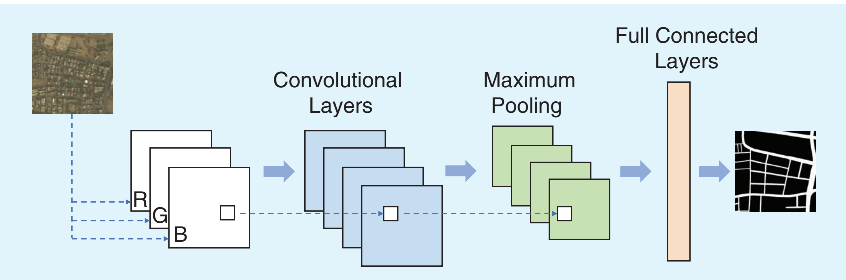

FIGURE 3. CNN method for road extraction.

FCNs-BASED MODEL

Compared to CNN models that accept only images with fixed sizes, FCN models use an interpolation layer after the final convolution layer to upsample the feature maps and recover images of similar size to the input, as shown in Figure 4. This section describes a study on the extraction of roads from remote sensing images using a model of FCNs.

Varia et al. [49] combined an FCN and conditional GAN (CGAN) to extract a new deep learning technique for road extraction from UAV remote sensing images. FCN-32 can capture road features well and segment them so that the completeness of road extraction is good.

Zhang et al. [50] proposed an FCN-based aerial image road extraction method. In addition, an integrated method based on spatial consistency is proposed, which incorporates the advantages of different models and avoids the determination of weights to cope with the imbalance of road and background areas.

Henry et al. [51] proposed a model FCN-8 based on evaluating FCNs for synthetic aperture radar image road extraction.

Zhang et al. [52] provided a road extraction method for remote sensing images of mountainous areas based on a multifeature full convolutional network (MFFCN). To improve the model training speed, MFFCN removes six layers of convolution based on the FCN model, effectively avoiding repeated storage and computation. Researchers have proposed some UAV road extraction methods based on fuzzy networks.

For example, UFCN [53] uses jump connections between the convolutional and deconvolutional stacks to preserve local information, significantly improving the model's performance.

Pan et al. [54] introduced a novel approach that utilizes an FCN to automatically extract roads from VHR remote sensing images. The road centerlines used for the labels in model training and validation are derived automatically from OpenStreetMap.

Although the aforementioned FCN-based methods have been applied to remote sensing image road extraction, with the deepening of the network layers, the input information is constantly diluted and road details are lost. Therefore, the FCN-based methods have the disadvantages of inefficiency, shadow effects, and susceptibility to occlusion.

ENCODER–DECODER-BASED MODEL

The primary concept underlying the encoder–decoder model involves extracting feature information from an image through convolutional pooling operations in the encoder. Subsequently, the image information is systematically recovered by employing an upsampling transposition convolutional structure in the decoder. Among these models, LinkNet [23] and U-Net [55] are the two typical image semantic segmentation codec models that are most applied in remote sensing image road extraction. In this section, we describe the study of road extraction from remote sensing images using the encoder–decoder structure.

LINKNET ENCODER–DECODERS

Since Zhou et al. [22] proposed the D-LinkNet model (Figure 5) and won the Conference on Computer Vision and Pattern Recognition DeepGlobe 2018 road extraction challenge, several researchers have made many improvements based on it.

For example, some researchers have used a variant of ResNet instead of ResNet34 in the D-LinkNet model encoder to construct the D-LinkNet50 [56] and the D-LinkNetPlus [57], respectively, where the D-LinkNetPlus network has fewer parameters and better extraction accuracy.

Gu et al. [58] proposed a multidilation (MD)-Link NeSet network for road extraction from remote sensing images of mining areas by UAV. A ResNeSt coding structure is used to extract the feature maps of the images, and an MD module is designed to enrich the network features through the fusion of multichannel feature maps. To improve the extraction effect of open pit mine roads, a road loss function is introduced in MD-LinkNeSt to obtain roads with better continuity.

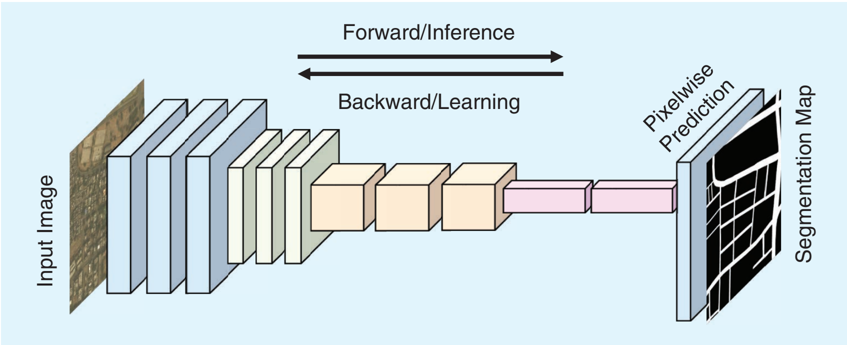

FIGURE 4. FCN method for road extraction.

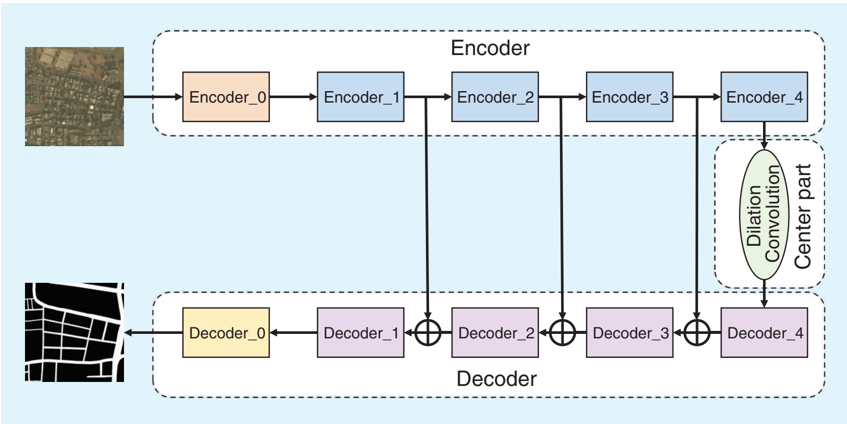

FIGURE 5. The structural framework of D-LinkNet [22].

In their study, Xu et al. [59] proposed an improved road extraction network, known as *DSDNet*. This network is built upon the foundation of D-LinkNet. The system used the limited adaptive histogram equalization algorithm to address the adverse effects of clouds and mountain shadows on the roads. The subsequent step involves the extraction of roads using the advanced DSDNet network. Postprocessing techniques are used in conjunction with terrain constraints to effectively mitigate the problem of false detection and enhance the overall quality of the image.

Zhou et al. [60] proposed a fusion network (FuNet) that fuses remote sensing images and location data using D-LinkNet as the backbone network. To enhance the learning capability of the network, a generalized iterative reinforcement model is designed to self-correct and optimize the model. In addition, the data preprocessing method using histogram equalization is also used to improve the image contrast.

Wang et al. [61] proposed an efficient nonlocal LinkNet (NL-LinkNet) to extract roads in VHR satellite images with fewer parameters and faster training convergence time. The nonlocal operation can better capture remote information on all features. Tran et al. [62] introduced the pyramid pooling (PP) [63] module into the LinkNet model and designed an improved DNN PP-LinkNet.

Wang et al. [64] replace the D-Block of the central structure of the D-LinkNet model with the DenseRe-Block and propose an improved Re-DlinkNet. Taking advantage of the easy construction of receptive fields through the attentional mechanisms in the transfomer blocks, Miao et al. [65] spread the stacks of transformer blocks across the layers of LinkNet and proposed a novel network called *TransLinkNet*.

Gu [66] proposed a multimodal road extraction model (MDNet) adapted to additional infrared channel remote sensing images by adding a unidirectional D-LinkNet branch identical to the backbone based on the D-LinkNet network with infrared images as input.

Wang et al. [67] use a cascading approach to redesign the D-Block into a DP-Block and then use a self-attentive mechanism in the upsampling phase to propose a novel feature-enhanced D-LinkNet (FE-LinkNet) to solve the problem of poor connectivity. Yang et al. [68] designed a new residual dense U-Net (RDUN) by combining the advantages of residual learning, dense nets, and U-Net.

U-NET ENCODER–DECODERS

U-Net [55] was first proposed by medical image segmentation applications and has achieved remarkable results in several semantic segmentation fields due to its superior performance. U-Net uses an encoder shrinkage path and decoder expansion path to capture the contextual information of the image using downsampling and recover the image position information using the upsampling operation to recover the image details gradually. Combining the advantages of residual networks and U-Net, Zhang et al. [69] constructed a deep residual U-Net (ResUNet) model with residual units as the basic neural units.

Xin et al. [32] proposed a DenseUNet network consisting of densely connected cells and skip connections and solved the problem of tree and building shadow occlusion to some extent.

Buslaev et al. [70] proposed an automatic extraction method based on the U-Net net family of FCNs. To improve the accuracy and recall of road extraction, Singh and Dash [71] designed a two-step deep CNN based on U-Net.

In their study, Sun et al. [72] proposed a U-Net model that utilizes multiple-output superposition for the purpose of road extraction from satellite images. A postprocessing method is employed to address the issue of unbalanced training data classes. This method utilizes a hybrid loss function, incorporating a shortest-path search with a decreasing threshold.

Constantin et al. [73] proposed a DNN based on U-Net and introduced atrous spatial pyramidal pools (ASPPs) to deepen the network in the lower levels of U-Net. A loss function based on binary cross-entropy and Jaccard distance is designed to avoid false detection and improve binary classification accuracy.

Hou et al. [74] proposed a remote sensing image road extraction method based on complementary U-Net (C-UNet). First, the standard U-Net is used for coarse extraction of roads from remote sensing images; then, multiscale dense expansion convolution (MD-UNet) is introduced to extract finer roads. Finally, the two extraction results are fused to obtain the final segmentation results.

Moreover, Abdollahi et al. [75] constructed a multilevel context-gated U-Net (MCG-U-Net) and a bidirectional ConvLSTM U-Net model (BCL-U-Net). In addition, a boundary-aware loss function was designed to allow the network to focus on hard semantic segmentation regions.

Chen et al. [76] proposed an Adaboost-based end-to-end multiple lightweight U-Nets (AEML U-Nets) approach for remote sensing image extraction. This model uses multiple lightweight U-Nets to enhance the segmentation capability of the model and combines them with the Adaboost strategy. Finally, numerous lightweight U-Nets are trained using a multiobjective learning strategy.

In recent years, Yang et al. [77] have designed a spatially enhanced and densely connected U-Net (SDUNet) that can better preserve the geometric structure of the road. SDUNet uses densely connected blocks to extract multilevel local features in the encoding phase and constructs a partial CNN, also called a *DULR module*, to reduce information loss during encoding.

OTHER ENCODER–DECODERS

Panboonyuen et al. [78] proposed a SegNet-based [79] network model for segmenting road targets in aerial images. Landscape metric thresholding is used to eliminate too many further detected roads and improve accuracy.

Inspired by the U-Net family of architectures, Doshi [80] designed a new residual first-start skip net. As in U-Net, a jump connection linking the same size layer is added between the encoder and decoder. To solve the problem of time-consuming and excessive noise caused by the symmetric structure of U-Net, Ding and Bruzzone [81] designed the direction-aware residual network (DiResNet).

A structurally supervised asymmetric residual segmentation network is designed in the model to enhance the learning of road topology. Then, directionally supervised learning of directed linear features is used to alleviate line segmentation and connection deficiency in road extraction.

Wang et al. [82] proposed a remote sensing road extraction method based on directional conditional random fields and an integrated codec network with inner convolution. By adding an inner convolutional network to the highest hidden layer of the codec network, the ability of the network to embed linear features and learn road-specific semantic information is enhanced, thus improving the learning of road topology and linear features. In addition, a postprocessing method with directed conditional random fields (CRF) is designed to improve the accuracy and connectivity of road extraction.

Ge et al. [83] presented a novel deep feature-review transmit network (TransNet) based on contour learning. Two feature-review modules are added to the encoder to enhance the learning of road contours, and a bridge is introduced in the encoder and decoder to recover the lost features.

The use of encoder–decoder-based methods has emerged as a prominent basis for various techniques, including attention mechanisms, multiscale approaches, and feature fusion. Numerous empirical investigations have consistently demonstrated that the LinkNet architecture, with its residual structure, exhibits a notably superior extraction performance compared to the U-Net architecture. The LinkNet structure, in particular, undergoes two primary enhancements: 1) The use of enhanced feature fusion modules, as opposed to skip links, enables the preservation of an increased amount of feature information. (2) For best results, consider including a module that improves perception while preserving the precise nuances of road information between the encoder and decoder. While this new method opens up a vast range of possibilities, the primary goal is to develop a model that is very efficient and performs well.

GAN-BASED MODEL

A GAN consists of a generator and a discriminator. The generator learns sample data features, and the discriminator identifies whether the samples are real or generated. This section describes a study of road extraction from remote sensing images using GAN-based models (Figure 6).

In recent years, several researchers have worked on the GAN model and used it for road extraction. Varia et al. [49] applied CGANs to road extraction in UAV images and demonstrated the feasibility of GAN in road extraction.

Shi et al. [84] proposed an end-to-end CNN based on generative adversarial training using Segnet as a generator. Cira et al. [85] proposed a pix2pix-based CGAN that introduces postprocessing methods to semantic segmentation of pavement regions. Then, a smoothing-based optimization method is used to extract road nodes and graph structure to improve performance.

Zhang et al. [86] combined deep convolutional GAN (DCGAN) and CGAN to propose an improved GAN for aerial image road extraction. The method uses a special FCN instead of an inverse convolution layer. While inheriting the FCN-4s model, the pooling layer is removed during downsampling, and low-level features are added to the sampling mapping. Good performance can still be achieved by training with a small amount of data.

Shamsolmoali et al. [87] proposed an adversarial spatial pyramid network with structured domain adaptation. To address the discrepancy between true and false images, structured domain adaption is applied with road extraction, and the model is optimized by using a joint reconstruction loss function.

Song et al. [88] proposed a model MapGen-GAN based on adversarial learning by using an improved BRB-Unet–based model as a generator, combined with the integration of roundness and geometric consistency constraints to reduce the semantic distortion of translation and to obtain richer information.

Abdollahi et al. [89] fused improved U-Net (MUNet) and GAN for extracting roads and improving segmentation map accuracy, respectively. Then, simple processing

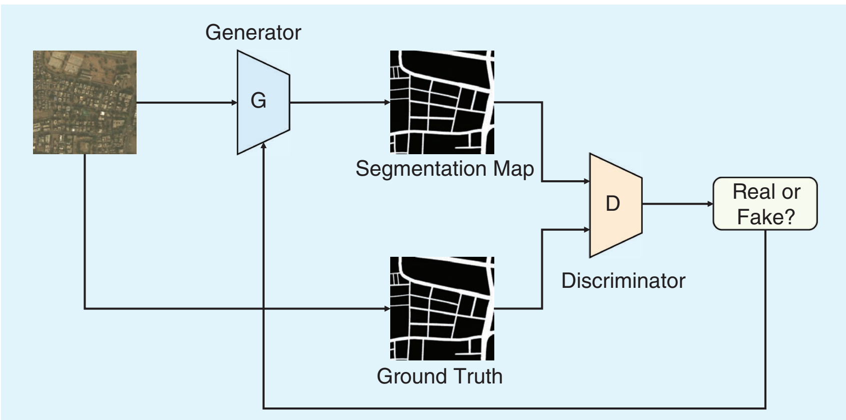

FIGURE 6. The structural framework of GAN [49].

combined with edge-preserving filtering significantly improves road network segmentation.

Hu et al. [90] proposed the weakly supervised GAN (WSGAN) for extracting roads from remote sensing images for specific scenes. First, the road network features are enhanced using the Wasserstein GAN with residual network blocks and gradient penalized loss. Then, the binary images are postprocessed to remove the salt-and-pepper noise. Experiments show that a more complete road can still be extracted even if buildings and a number of shadows cover it.

Chen et al. [91] proposed the NIGAN for extracting roads from images captured in complicated mountainous regions. The framework designs a new remote sensing road scene neighborhood probability enhancement method to classify road scenes. It introduces an improved dilation convolution model in the CGAN to subdivide roads from classified road scenes.

The GAN model has the ability to generate data continuously and distinguish between true and false data, which can effectively compensate for the scarcity of road feature information. Especially, it has better ability when facing the situation of less road information in a single image. However, when processing large-scale image data, the GAN model can be prone to gradient disappearance and mode collapse, leading to road information loss.

ATTENTION MECHANISM-BASED MODEL

The main details of the attention mechanism are designed to help ignore irrelevant information and focus on important details. The attention mechanism is usually inserted as a component in the backbone network to obtain attention weights through forward propagation and backward feedback.

For example, Hu et al. [92] proposed a novel "squeeze and excite" (SE) module, and subsequently, Lin et al. [93] proposed a nested SE-Deeplab model by integrating the SE module into the DeepLab v3 [94] network. It can apply weights to different feature roads and perform multiscale upsampling to retain more information.

With the continuous advancement of attention mechanisms and previous research, the attention mechanism has been widely used to extract roads from remote sensing images.

Xu et al. [95] proposed a deep learning framework for road extraction (GL-Dense-U-Net) by adding global and local attention models to DenseNet [96], drawing on the symmetric structure of the U-Net model.

Based on the DenseNet, the coordinate-dense-global (CDG) [97] was proposed by adding the global attention module (GAM) and the coordinated convolution (Coord-Conv) module. The GAM introduces high-level information using global averaging pooling, which can effectively avoid the loss of global contextual information (Figure 7).

Moreover, Li et al. [98] proposed a new model, CA-DUNet, based on the DenseUNet framework by embedding two global attention and core attention modules. The GAM is used to obtain contextual information on the road. In addition, the core attention module is used to ensure the network's maximum transmission of road information is densely fast, thus providing the integrity of the extracted road.

Qi et al. [99] proposed an attention-based mechanism for the ATD-LinkNet model, which is used for road extraction in remote sensing images. The ATD-LinkNet model incorporates both the benefits of a global attention mechanism and a multiscale model. This combination results in an alternative AT module that comprehensively captures contextual semantic information.

Li et al. [100] proposed an end-to-end deep learning model that effectively improves road extraction by introducing a direction-aware module to maintain the road topology and adaptively correct the features. Liu et al. [101] proposed a new residual attention local context-aware network (RALC-Net). A residual attention module has been developed to enhance the extraction of road local information. This module leverages residual connectivity to effectively integrate spatial contextual information with attention mechanisms, thereby emphasizing the significance of local semantics.

Wan et al. [102] proposed a dual attention road extraction neural network (DA-RoadNet) to solve the problem of discontinuous and incomplete road extraction under complex conditions. The dual attention mechanism (DAM) enhances the representation of road area features and makes road extraction more effective by using a hybrid function of the dice loss function and binary cross-entropy loss to avoid occlusion by shadows from trees and buildings.

Yang et al. [103] proposed a road extraction model RCFSNet based on a coordinated DAM (CDAM). CDAM is mainly used to enhance the representation of road features and its full-stage feature fusion. Xu et al. [104] proposed a spatial attention-based network model, MSACON. It uses the results of building structure

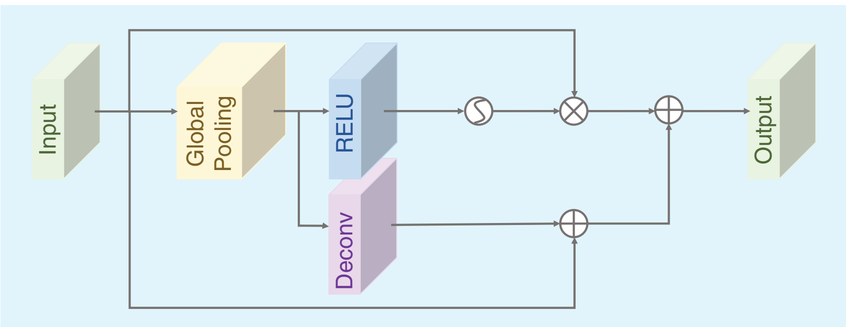

FIGURE 7. The structural framework of CDG [97].

extraction to compute standard deviations of the mean as fuzzy auxiliary information to infer potential roads in remote sensing images, especially covered and obscured roads.

In addition, a spatial attention-based fusion mechanism is integrated into the decoder to mine the contextual information of buildings and roads. Chen et al. [105] proposed an improved codec model DBRANet. A two-branch network module was proposed in the encoder module, using a residual network in one branch and a refined asymmetric block in the other branch. A new attention module-regional attention network module is used in the decoder module to cope with the characteristics of road twists in remote sensing images.

Wang et al. [106] introduce the dilation convolutional attention module in the middle of the encoder and decoder. In contrast, the convolutional block attention module is embedded in the parallel dilation convolution and pool branches to obtain more attention-aware features.

Huan et al. [107] proposed a new image segmentation network (SANet) to solve the problem of poor continuity in remote sensing image segmentation. The network is designed with a stripe attention module, which is used to extract road context information and road localization, and a channel attention fusion module, which is used to fuse high-bottom features.

To extract abundant features from a small number of samples, Gong et al. [108] made some modifications to U-Net by using a dense dilation convolution module to extract road features and a channel attention module to achieve feature fusion. Xiao et al. [109] proposed a RATT-UNet deep learning network for extracting mine roads based on the U-Net framework with residual connectivity and attention mechanism. The RATT model can achieve better extraction results with fewer parameters.

Ren et al. [38] provided a new dual-attention capsule U-Net model (DA-CapsUNet) by combining the advantages of capsule representation and attention mechanism. The atrous convolution-based channel feature attention module and the spatial feature attention module are used in the encoder and decoder to highlight salient features.

Shao et al. [2] proposed an end-to-end road extraction network with an embedded attention mechanism by adding a channel attention mechanism and a spatial attention mechanism based on the U-Net framework. The residual dilation convolution module is introduced in the network framework to extract network information at different scales. In contrast, the residual dense connectivity block is introduced to enhance the information flow transfer and feature reuse of feature maps at different levels.

Mei et al. [110] proposed a connectivity attention network (CoANet). To mitigate the effect of occlusion in the road area, a connected attention module combining graphical information is designed, which can better preserve the connectivity of the road. In addition, a strip convolution module has been developed to better conform to the road shape. Dai et al. [111] proposed a road-augmented deformable attention network (RADANet). RADANet embeds a deformable attention module at each layer of the encoder and encoder to capture multiscale road semantic information and inserts a road augmentation module after the fourth encoder level to capture the semantic shape information of the road.

Khan et al. [112] proposed an encoder–decoder network, DSMSA-Net, integrating two attention units. The first scale attention unit uses a multiscale strategy to extract multiscale information from the feature graph. The second spatial attention unit uses two different pooling operations to capture contextual information.

The attention block has the properties of accurate localization, expanding the perceptual field and effectively extracting contextual information. In the process of road extraction, it can effectively ignore irrelevant information. As more attention modules are proposed, the road extraction model based on the attention mechanism has more possibilities.

FEATURE FUSION-BASED METHODS

The basic idea of feature fusion is to add or stitch the extracted feature maps at the pixel level. In the feature extraction stage, the semantic information of the feature map is enriched by fusing multiscale feature information. The globally valid information is utilized in the feature utilization phase by fusing features at different levels. The following describes a study of road extraction from remote sensing images using a feature fusion-based model.

To address the problem of category imbalance in high-resolution aerial images, Lan et al. [113] proposed a global context-based dilation convolutional divine neural network (GC-DCNN). The algorithm is based on three remaining dilation blocks used to construct encoders to capture more additional features, and a pyramid pooling module was used to fuse multiscale global contextual features to enhance the feature representation.

Zhu et al. [36] proposed a novel road extraction network, GCB-Net, which incorporates a global context-aware module in the codec and uses a filter response normalization layer to eliminate batch dependency and further improve the robustness of the model (Figure 8).

Li et al. [114] proposed a convolutional network (Y-Net) shaped like the letter Y for segmenting and extracting multiscale roads from remote sensing images. The Y-Net consists of a two-armed feature extraction module and a fusion module. The feature extraction module extracts semantic and detailed features using two subnets.

Yan et al. [115] constructed an improved structure DID-LinkNet by upgrading the D-Block to DID-Block using dense connectivity and iterative fusion. Three fusion nodes were added to the D-Block to iteratively fuse the lateral outputs and aggregate the learned representations from each layer, effectively improving the connectivity of the road.

Since the image resolution decreases with increasing network depth and some important location information is weakened, Tan et al. [116] proposed a new end-to-end extraction model and designed a new decoder to capture road edge and shape information more accurately.

Zou et al. [117] proposed a full-scale feature fusion network, AF-Net, in which a full-scale feature fusion module (AF module) was designed to implement feature fusion at each scale and embed a convolutional attention block attention model in the model to improve the quality of road extraction.

Wu et al. [118] constructed a dense global spatial pyramid pool (DGSPP) module and a fuzzy neural network-based road extraction model DGRN. The model builds on the residual network (ResNet) and introduces the DGSPP module to reduce spatial feature dropout and improve the reusability of road features. DGSPP uses dense connections instead of parallel connections in traditional ASPP to obtain richer multiscale features.

In previous studies, researchers often used skip links to achieve simple feature fusion. Nowadays, more and more feature fusion modules are proposed and used in the decoding process, which can efficiently extract the road background information and fuse the full-stage features of the road.

MULTITASK LEARNING-BASED METHODS

Due to the relevance of information, such as roads, road centerlines, and road edges in remotely sensed images, researchers often use multitask learning techniques to extract this information from remotely sensed images for scene classification. Multitask learning can train models with better robustness and generalization ability by jointly learning multiple related tasks and using different tasks to provide additional useful information between them.

This section investigates a multitask-based road extraction method from remote sensing images. Liu et al. [34] proposed a novel multitask-based CNN model, RoadNet, for the simultaneous detection of roads, road centerlines, and road edges in VHR images. RoadNet uses a cascaded network for multiscale and multilevel feature extraction to cope with complex scenes. In addition, a cropping and bilinear mixing method is designed to segment VHR images to improve training performance.

Yang et al. [119] proposed a deep learning model RCNN-UNet based on recursive CNNs. A multitask learning framework was used to predict the road and road centerline using two predictors, respectively. Combined with the U-Net framework, an RCNN modular unit is designed to significantly preserve the rich underlying spatial features and effectively mitigate the effects of occlusion, noise, and complex backgrounds.

Lu et al. [120] proposed a multitask automatic road extraction framework (MSMT-RE) for simultaneously detecting roads and road centerlines. The framework combines multiscale feature fusion and adaptive loss function based on U-Net to enhance the prediction of obscured roads.

To solve the problem of road fragmentation and incorrect segmentation due to traditional semantic segmentation methods, Yi et al. [121] proposed an end-to-end model, EUNetMTL, based on semantic cut and directional learning. In the encoding phase, an EfficientNet-B4–based encoder network structure is used to improve feature extraction accuracy and expand the receptive field by adding dilated convolution. EUNetMTL also has a directional learning branching module to ensure the topology of the road extraction results.

Wei et al. [122] proposed a multilevel framework neural network to extract road regions and road centerlines simultaneously. First, the framework is optimized using a boosting strategy for segmented images. Then, a novel road centerline tracking method is designed to track the automatically generated starting points, which are iteratively searched by an embedded CNN to obtain the complete road network.

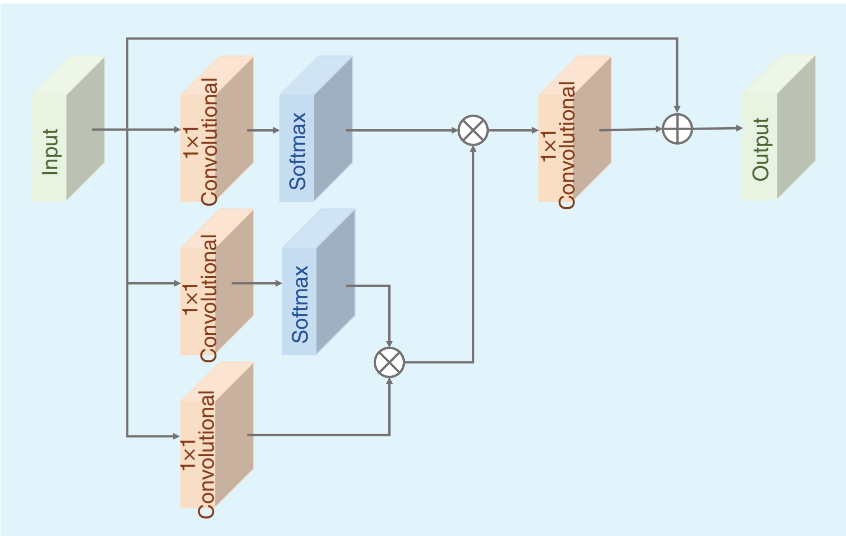

FIGURE 8. The structural framework of GCB-Net [36].

Shao et al. [123] proposed a multitask road correlation extraction network (MRENet) for extracting road regions and road centerlines from VHR images. To capture more feature information, MRENet uses Resblock and PSPooling modules to integrate feature information at different scales and road and centerlines. It also uses a weighted binary cross-entropy function to solve the problem of background imbalance.

Wei et al. [124] proposed a multitask deep learning framework to perform road detection and centerline extraction simultaneously. An ordinal regression-based road centerline method reduces the false-negative part of the road centerline extraction results. To consider road topological features, a new road-topology loss function is designed to improve the connectivity and integrity of the road network.

To improve road connectivity, Lu et al. [125] proposed a new cascaded multitasking (CasMT) road extraction framework that leverages the symbiotic relationship between road segmentation, centerline extraction, and edge detection. In addition, a context-aware module is used to achieve topology-aware learning, and a hard sample mining loss reduction fuzzy neural network is employed to improve road recognition rate and completeness further.

The multitask learning usually includes pavement segmentation, road edge detection, and road centerline extraction. The crossover of information between subtasks may be beneficial to both parties. The cascade between subtasks can effectively compensate for the problem of insufficient samples of a single task, which can improve the generalization effect and solve the shadows and occlusions of the road region. However, due to the presence of background noise and the unbalance of the ratio between the road and the background, it will come to a large number of extra computations.

MULTISCALE-BASED METHODS

The multiscale-based methods extract information at different scales by constructing multiple neural networks, and these abundant features help to improve the accuracy of the road extraction task from remote sensing images. This section describes a study on road extraction from remote sensing images using a multiscale approach.

Recently, Li et al. [126] proposed a hybrid convolutional network fusing multiple subnetworks. The hybrid convolutional network constructed three subnetworks containing FCN, modified U-Net, and VGG to extract coarse-, medium-, and fine-grained road features, respectively. Then, these multiple granularities were fused using shallow convolutional subnetworks to obtain better extraction results.

To solve the problems of burr and time-consuming in the existing road centerline extraction, Liu et al. [127] proposed a road centerline extraction method based on a multiscale Gabor filter and multidirectional nonmaximum suppression. First, CNN automatically learns structural features from the raw data, and a CNN-based pixel classifier is used to classify the aerial images. In addition, shape feature filters, multidirectional morphological filters, and void filling are used to extract accurate road centerlines.

Li et al. [128] proposed a multiscale skip-connected and asymmetric convolution-based U-Net (MACU-Net). Among them, multiscale skip connections can combine and improve the multilevel semantic features generated by U-Net. Moreover, it enhances the representation capability of standard convolutional layers by using asymmetric convolutional blocks.

Multiscale convolution can efficiently acquire layer-level features in different dimensions to improve road centerlines' completeness and recognition accuracy. To achieve high-accuracy segmentation of multiscale roads in complex scenes, Zhang et al. [129] proposed a global attention multipath dilation convolutional selective pass refinement network (GMR-Net).

Different regions' contextual information is aggregated using the multipath dilation convolution method to achieve multiscale road feature extraction. In addition, a gated refinement unit combining global and local information is proposed to refine the details progressively.

Wulamu et al. [130] combined ASPP with U-Net to improve the ability to obtain background information at multiple scales, He et al. [131], to better cope with complex road scenarios, integrated ASPP with an encoder–decoder network to improve road extraction performance.

In addition, a new structural similarity (SSIM) loss function is designed to improve the quality of road extraction. To cope with the problem that the shadows of trees and buildings obscure roads, Jie et al. [132] provided a new multiscale based road extraction network (MECA-Net). First, MECA-Net uses a multiscale feature module to improve the model's ability to recognize slender roads. Second, a long-range context-aware module (LCAM) consisting of a channel attention module and a strip pooling module is designed to obtain sufficient remote context information (Figure 9).

The multiscale-based methods obtain multiscale feature maps by performing feature extraction on input images of different scales. The expansion of the receptive field can effectively enhance the extraction of slender and occluded roads. Nowadays, researchers tend to use a combination of multiscale feature learning and feature fusion, which can effectively improve the accuracy and stability of road extraction.

TRANSFORMER-BASED MODEL

In recent years, Transformer's field of excellence in natural language processing has driven its application in computer vision tasks, such as image classification and image segmentation.

Yang et al. [133] introduced Swin transformer [135] to remote sensing image road extraction to better capture the context. In addition, a foreground contextual information supplement module algorithm is proposed to enhance the inference of the obscured roads.

Ge et al. [135] developed a new network Swin transformer U-net using a Swin converter to replace multiheaded self-attentive (MSA) with a window-based self-attentive (W-MSA) module or shifted window-based self-attentive (SW-MSA) module for better extraction of global and local information. In addition, an improved loss function based on binary cross-entropy is proposed by combining binary cross-entropy with dice loss to improve road extraction accuracy.

More specifically, Chen et al. [136] designed a new two-branch coding module, CoSwin, in a U-shaped network framework, using the global context modeling capability of Swin transformer and the local extraction capability of ResNet.

Luo et al. [137] proposed a bidirectional transform network (BDTNet) to enhance remote sensing images' global and local extraction. This method constructs a bidirectional transform module (BDTM) that extracts local and global features using the convolution operator and SA mechanism. A feature refinement module is introduced to fuse with the features extracted by BDTM to enhance the visualization capability.

Zhang et al. [40] proposed a light Rodaformer model to extract mountain roads in complex environments. A converter and a self-noticing module are used to extract road edge information accurately. Then, a postprocessing module based on road topological features is used to remove incorrectly predicted roads.

The existing methods mostly combine transformer blocks and CNNs. They complement each other and enhance the capability of feature extraction. However, using transformers inevitably causes problems, such as insufficient GPU memory and low computational efficiency.

GCN-BASED MODEL

The concept of GCN [138] was extended to graph-structured data by constructing adjacency matrices. GCNs were initially applied in knowledge graphs, and in recent years, they have been used to feature extraction of natural images. In this area, Cui et al. [139] extracted a new road extraction method based on joint superpixel segmentation and GCN.

GCN uses simple linear iterative clustering segmented superpixels as nodes, based on which the features of nodes are updated using the features of their neighbors to retain more spatial detail information.

Zhou et al. [39] proposed a separable GCN. Two GCNs are used to capture global contextual information in spatial features and channels to enhance the representation of road features.

To address the problem of incomplete and discontinuous road extraction, Cui et al. [140] combined a CGN and a CNN to construct a graph-based dual convolutional network (GDCNet) capable of feature learning for regions at different scales to enhance the extraction of tiny or covered roads.

The GCN methods help extract tiny, tortuous, and covered roads. However, GCN's receptive field of view depends on the input size. A large receptive field leads to too much noise, while a small receptive field leads to a lack of information. Therefore, it becomes important to choose a suitable input size.

UNSUPERVISED LEARNING METHODS

Although the remote sensing image road extraction method with supervised learning has produced some results, manually labeled datasets are often not achieved with favorable results after iterative training in practical applications.

Therefore, the best approach is introducing an unsupervised application method that minimizes the differences between the original and reconstructed images by tailoring the labels to the model.

For example, Tao et al. [141] proposed a framework called *unsupervised restricted deconvolutional neural network* (*URDNN*) by using a stacked convolutional autoencoder for unsupervised feature extraction training and alternating training between supervised and unsupervised approaches to constrain feature extraction and learning to obtain more generalized and abstract features from unlabeled data.

In a subsequent study, Tao et al. [142] introduced a novel approach in the form of a dual-stream deep-learning neural network. This study used an unsupervised hierarchical feature extraction technique to preserve the low-level

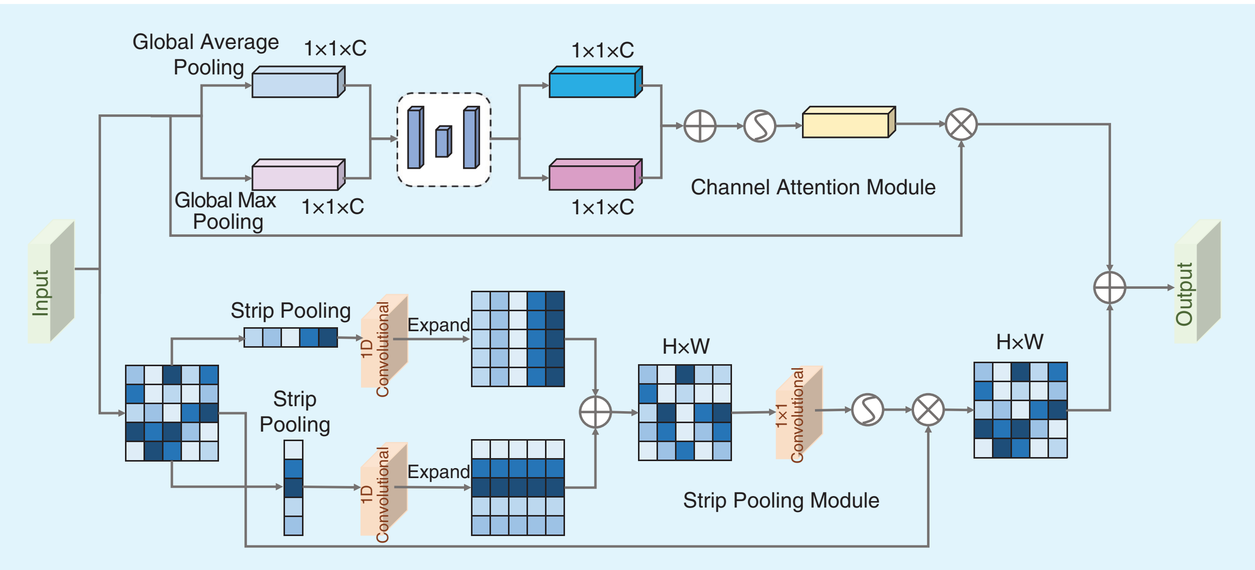

FIGURE 9. The structural framework of LCAM [133].

information within a supervised residual network's framework, specifically designed to extract features from multispectral images.

The efficacy of unsupervised learning methods in extracting information is subject to certain limitations compared to supervised learning systems. The researchers proposed utilizing a hybrid approach, incorporating both unsupervised and supervised learning methodologies, to incrementally improve the accuracy of road extraction.

WEAKLY SUPERVISED LEARNING METHODS

In practical applications, unsupervised learning methods exhibit limited performance. Researchers have proposed enhancing supervised learning techniques by integrating manually supervised information into the model. This approach effectively minimizes the requirement for manual labeling and subsequently reduces the workload associated with it. In the initial phase, researchers mainly combined CNNs with weakly supervised techniques for extracting road images from remote sensing images. Unfortunately, the results were unsatisfactory.

There are two main groups of deep, weakly supervised remote sensing image road extraction methods. The first category is based on graffiti tag learning. Wei and Ji [143] proposed a graffiti-based weakly supervised road extraction method, ScRoadExtractor. Yuan et al. [144] suggested a graffiti-based weakly supervised remote sensing road extraction network, WR2E, and a road trilateration generation module to extend the semantic information of graffiti annotations.

Zhou et al. [145] proposed a weakly supervised road segmentation network (SOC-RoadNet) based on structure and directional consistency to learn road surface features directly from graffiti tags. Then, the road direction features are aggregated using the road refinement module to optimize the road network.

The second category is learning methods inspired by human pose. Lian and Huang [146] presented a point labeling-based road network extraction method. The method only needs to point out the centroid of each road segment in the training block, effectively reducing the labeling cost.

Likewise, inspired by the cascading hourglass model for human joint detection, Lian and Huang [147] proposed a weakly supervised road segmentation framework based on multipoint annotation.

In recent years, weakly supervised learning based on small samples has become a popular research trend in road extraction of remote sensing images. Researchers combine a small number of samples with GAN to avoid a large number of pixel-level training labeling samples. In particular, selecting the appropriate input image size and activation function is necessary when using GAN.

COMPARISON AND DISCUSSION

COMPARISON OF THE STATE-OF-ART METHODS

To describe the performance of the remote sensing image road extraction models mentioned in the "State-of-the-Art Methods for Road Extraction" section more intuitively, 20 of the representative models in each of Massachusetts [26] and DeepGlobe [27] datasets are presented in Tables 2 and 3, respectively. This includes the image input size and the results obtained on these models for metrics, such as IoU, Precision, Recall, and F1 score.

To visualize our experimental results on remote sensing image datasets, we selected several classical road extraction models and presented the visualization results to visually and comprehensively illustrate their performance comparisons. In particular, we retrained these models and tested them based on the DeepGlobe [27] and Gansu [39] datasets. The environment and framework used were Python v3.8, Pytorch 1.8.0, and CUDA 11.7. All experiments are implemented on 2 NVIDIA 3090 with 24 GB of memory.

By comparing these algorithms, the following conclusions can be drawn:

- *The Massachusetts road dataset* provides original 1,500 × 1,500 pixel experimental images that are too large to be fed into the network and prone to training instability, so dominant models tend to use 512 × 512 images as input. ICN-DCRF [82] based on encoder–decoder structure

TABLE 2. MASSACHUSETTS ROADS DATASET.

METHOD	INPUT SIZE	IoU	PRECISION	RECALL	F1 SCORE
Deep FR TransNet [83]	1,024 × 1,024	62.86	83.72	78.1	–
RCFSNet [103]	1,024 × 1,024	65.52	82.54	73.99	78.03
Two-Step CNN [71]	512 × 512	–	87.9	89.3	88.6
ICN-DCRF [82]	512 × 512	–	87.1	82.2	84.6
U-Net [132]	512 × 512	65.63	81	77.57	79.25
LinkNet [132]	512 × 512	65.43	80.77	77.51	79.1
SDUNet [77]	512 × 512	74.1	81.2	75.7	78.4
DBRANet [105]	512 × 512	65.72	81.21	77.48	78.48
GMR-Net [129]	512 × 512	–	83.91	87.6	85.7
DDU-Net [106]	512 × 512	–	82.54	73.99	78.03
MECA-Net [132]	512 × 512	65.82	80.63	78.19	79.33
CoSwin [136]	512 × 512	66.4	81.7	78	79.81
BDTNet [137]	512 × 512	66.04	82.99	76.37	79.55
DiResNet [81]	320 × 320	67	80.12	80.29	80.06
RDRCNN [47]	256 × 256	67.10	85.35	75.75	80.31
AEML U-Nets [76]	256 × 256	64.77	81.06	76.33	78.62
CADUNet [98]	256 × 256	64.12	79.45	76.55	77.89
DA-RoadNet [102]	256 × 256	64.19	79.16	77.25	78.19
GDCNet [140]	256 × 256	62.94	84.43	71.21	–
SGCN [39]	224 × 224	65.28	84.82	73.91	78.99

and GMR-Net [129] based on multiscale have achieved better results. The symmetric encoder–decoder can effectively retain the ability of spatial details and is suited for processing large-scale images. Based on this, the introduction of postprocessing methods and gated units to gradually refine the road details can significantly improve the quality of road extraction.

- *The mainstream model for the DeepGlobe road dataset* usually uses the original 1,024 × 1,024 image input, and the NL-LinkNet [61], RCFSNet [103], and TransRoadNet [133] models are all improved models based on D-LinkNet [22]. In the encoding stage, the pretrained ResNet34 [148] network weights on the ImageNet dataset are used for image feature extraction, which can accelerate the convergence of the model. On the one hand, the introduced diverse attention mechanism to enrich the contextual content of the image in the feature utilization stage can effectively improve road connectivity and completeness. On the other hand, although the models extracted in recent years have improved regarding road extraction performance, the number of parameters is getting larger, and the training time is getting longer.
- *In both DeepGlobe and Gansu datasets,* the extraction ability of the U-Net–based model is much less than that of the LinkNet-based model. U-Net–based models have the disadvantages of misclassification and poor generalization ability. Differences between LinkNet-based models existed mainly on roads that were influenced by other factors, especially roadside buildings and trees. On the major roads, the extracted results were largely similar.
- *Transformer methods:* CoSwin [136] and BDTNet [137] models both use the Swin transformer for extracting road information in remote sensing images. The Swin transformer structure demonstrates an extraordinary ability to identify and understand the global context. Earlier work has shown that the transformer structure may result in the loss of crucial local features and a decrease in road segmentation accuracy. As a result, pursuing a model that combines the transformer will emerge as the primary focus of future research.
- *GCN methods:* SGCN [39], and GDCNet [140] models are both graph convolution-based road extraction algorithms for remote sensing images. Compared with other models, GCN can expand the receptive domain and capture global contextual confidence, which helps to extract covered, winding, and fine roads, especially for small targets and complex roads. Therefore, the remote sensing image road extraction models used for graph convolution have tended to perform well in terms of Precision metrics. When the GCN module is employed, selecting a smaller image size as input is especially necessary.
- *In some of the experimental results,* there is often overextraction. And we believe that there are reasons for this problem. On the one hand, it is the extraction error of the model itself. On the other hand, the dataset is not fine enough, resulting in the correct roads being extracted not being shown in the labeled images. In order to make the model cope with the extraction of complex situations, we need more and more accurate datasets of roads from high-resolution remote sensing images.

TABLE 3. DEEPGLOBE DATASET.

METHOD	INPUT SIZE	IoU	PRECISION	RECALL	F1 SCORE
U-Net [103]	1,024 × 1,024	63.62	73.22	83.57	76.76
D-LinkNet [133]	1,024 × 1,024	67.58	79.01	83.07	79.55
NL-LinkNet [103]	1,024 × 1,024	68.58	80.92	81.77	80.02
MACU-Net [133]	1,024 × 1,024	66.57	79.67	80.55	78.94
RCFSNet [103]	1,024 × 1,024	69.34	78.98	85.46	81.01
TransRoadNet [133]	1,024 × 1,024	70.06	80.92	83.91	81.34
Deep FR TransNet [83]	1,024 × 1,024	72.44	87.3	81.15	—
RENA [2]	1,024 × 1,024	63.1	78.4	77	76.4
MECA-Net [132]	512 × 512	65.15	78.39	79.41	78.9
DBRANet [105]	512 × 512	66.16	79.88	79.56	78.02
BDTNet [137]	512 × 512	67.09	84.18	76.77	80.3
GMR-Net [129]	512 × 512	—	87.97	88.86	87.38
GCB-Net [36]	512 × 512	70.08	—	—	81.54
EUNetMTL [121]	512 × 512	69.52	—	83.58	82.02
SDUNet [77]	512 × 512	66.8	78.4	74.2	79.4
DiResNet [81]	320 × 320	66.80	78.76	81.46	79.09
GDCNet [140]	256 × 256	54.39	76.5	67.46	—
FCN-8s [102]	256 × 256	47.12	65.06	63.08	64.05
GL-Dense-U-Net [102]	256 × 256	51.65	75.34	62.16	68.12
DA-RoadNet [102]	256 × 256	55.7	73.65	69.55	71.54

VISUALIZATION ANALYSIS

The DeepGlobe [27] dataset covers a variety of scenarios. Since the test set does not give labeled images, we divided the original training set of 6,226 pairs of images into 5,500 pairs of images for training and 726 pairs for testing. The visualization method on the Gansu [39] dataset uses a 256 × 256 image input for all models.

Figures 10, 11, and 12 show a comparison of different road extraction methods used to extract roads from remote sensing images for the DeepGlobe [27] dataset. Figure 10 shows that the RCFSNet [103] model using the coordinated dual attention mechanism has better extraction results for roads obscured by trees and buildings, especially for fine roads.

Figure 11 shows the U-Net [55] and RCFSNet [103] models do not extract asphalt roads as well as the D-LinkNet [22] and NL-LinkNet [61] models using LinkNet [23] as the framework. Moreover, the U-Net [55] and RCFSNet [103] models are prone to misclassify the roads. The cause of this misclassification is not all a problem with the model itself; the lack of fine-grained road labeling in certain datasets is also a contributing factor. This result can be seen more obviously in Figure 12.

The extraction results of D-LinkNet [22] and NL-LinkNet [61] are similar, and the road continuity and completeness are better, but the training time is less in NL-LinkNet [61].

Figures 13 and 14 show a comparison of different road extraction methods used to extract roads from remote sensing images for the Gansu [39] dataset. U-Net [55] performed poorly on this dataset.

The extraction results of LinkNet [23], D-LinkNet [22], and NL-LinkNet [61] are similar. Although the general outline of the road can be recognized, the recognition accuracy is not enough, and this result is especially on the MACU-Net [129] model, where slender roads are recognized as too coarse. The SGCN [39] based on graph convolution has the best recognition accuracy and recognition results, especially for tiny roads.

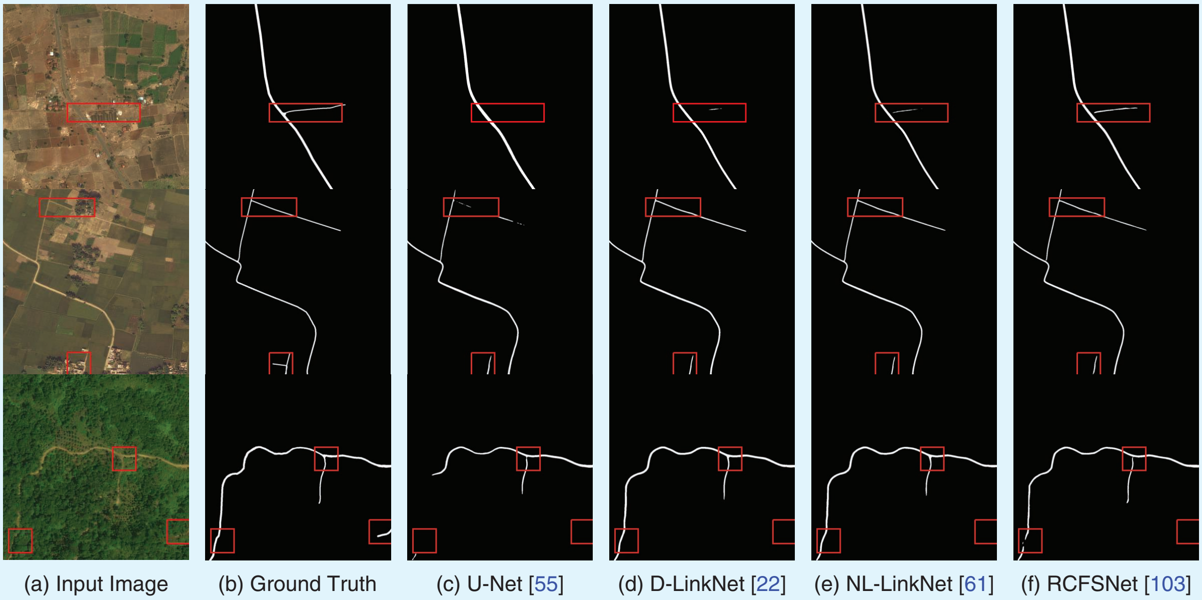

FIGURE 10. (a)–(f) Comparison of visual results of different road extraction methods on the DeepGlobe [27] dataset.

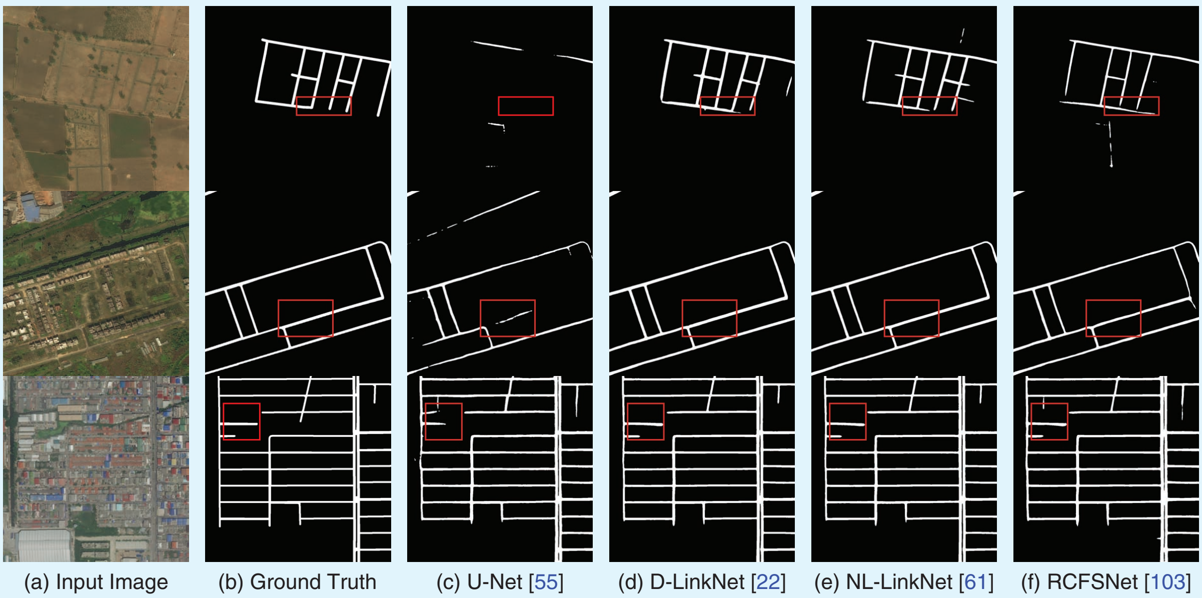

FIGURE 11. (a)–(f) Comparison of visual results of different road extraction methods on the DeepGlobe [27] dataset.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

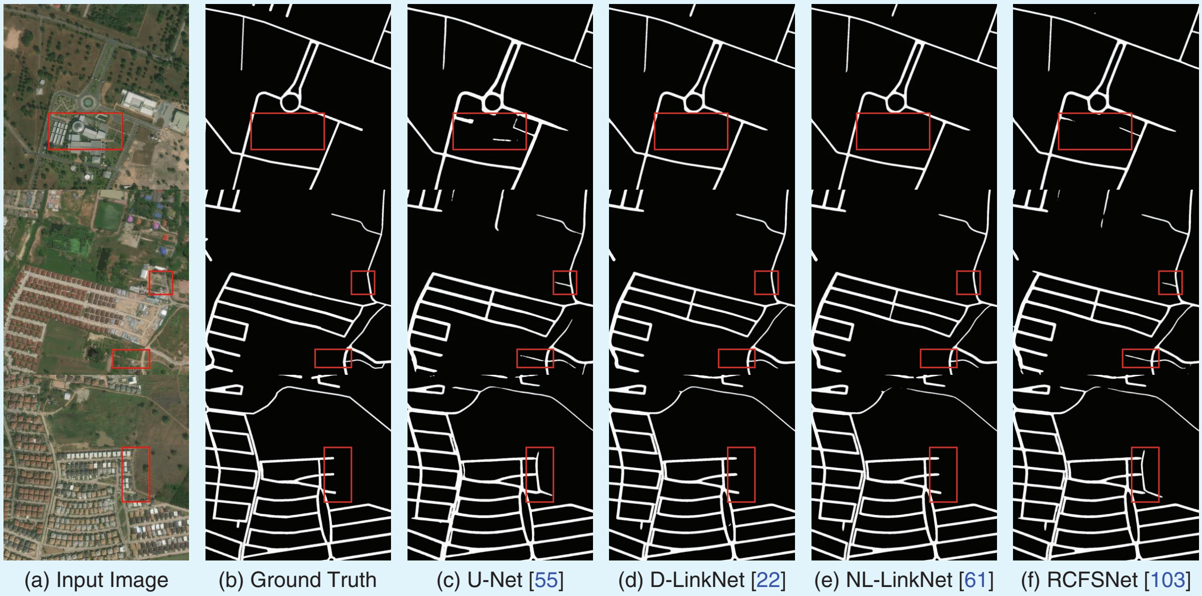

FIGURE 12. (a)–(f) Comparison of visual results of different road extraction methods on the DeepGlobe [27] dataset.

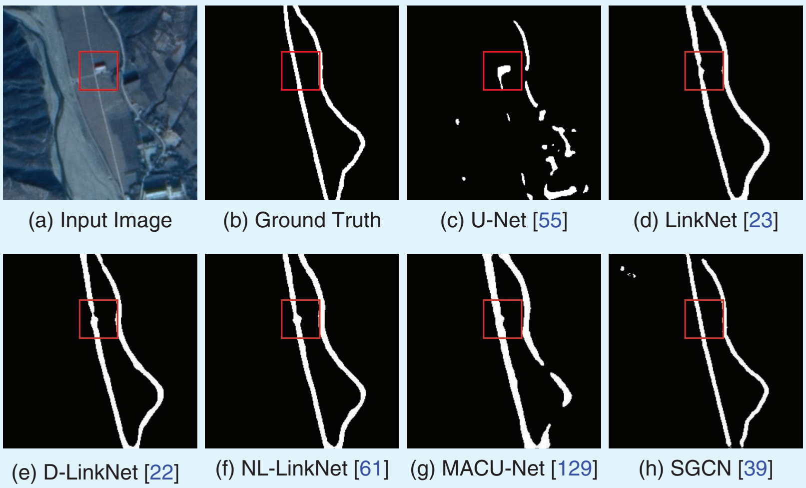

FIGURE 13. (a)–(h) Comparison of visual results of different road extraction methods on the Gansu [39] dataset.

CURRENT CHALLENGES AND FUTURE WORK

The "State-of-the-Art Methods for Road Extraction" and "Comparison and Discussion" sections showcased models that have demonstrated exceptional performance in extracting roads from remotely sensed images. The application results of these models on remote sensing images show their significant contribution to the advancement of road extraction techniques for remote sensing images. However, high-resolution remote sensing images still present numerous unresolved issues and challenges. Roads exhibit several characteristics. First, they can span long distances, often extending for many miles. Second, their boundaries can be indistinct and not clearly defined. Also, roads can have irregularities, which may not be perfectly smooth or uniform. It is important to note that these characteristics can vary significantly depending on the type of road, such as tarmac, dirt roads, asphalt roads, gravel roads, and earthen roads. However, it is important to note that road information is subject to various interferences that can impact its accuracy and reliability. These interferences include the presence of buildings, which can obstruct the line of sight and affect data collection. Additionally, vegetation shading can further complicate the process by creating shadows and reducing visibility. Furthermore, the movement of pedestrians and traffic flow can introduce dynamic elements that may interfere with collecting and interpreting road information. These interferences pose significant challenges in accurately capturing and analyzing road data.

In this section, we discuss these issues and present some popular and promising future research directions. Remote sensing image road extraction can contribute to map creation and update research by breaking through the limitations of technology level and environment.

We believe these methods help foster the development of cutting-edge research on remote sensing image road extraction and "pave the way" (without playing on words) for future advancements in this important field. In the following, we will outline potential exciting future directions for network design and application.

- *Producing high-resolution road remote sensing datasets:* Compared with image classification datasets, the datasets for image segmentation tasks are much smaller, especially in remote sensing image segmentation datasets. The current datasets used by the fully supervised algorithm are all labeled by humans. The number of finely labeled images generated is also small, and the next research is devoted to producing high-quality remote sensing image road datasets. In addition, designing a better model to label image datasets automatically may become another research hotspot.
- *Transformer structure:* The visual transformer structure has recently shown promising results in various computer vision tasks. Currently, only a few remote sensing image road extraction models utilize transformer technology. These models have shown that the transformer structure effectively captures contextual semantic information. In the future, a transformer structure suitable for remote sensing image road extraction tasks will be designed based on the unique features of remote sensing images, which differ from natural images.
- *Explore semisupervised and unsupervised research:* In supervised learning models, networks that have been trained well often struggle to achieve good performance when the dataset size is limited. Semantic segmentation of images is a more challenging task compared to target detection and image classification. As a result, it requires a larger number of labeled images and higher accuracy requirements. However, the process of manually labeling these images is time-consuming and requires a significant amount of effort. Semisupervised and unsupervised learning methods leverage a smaller number of labeled datasets and a larger number of unlabeled datasets, offering advantages in dataset scaling. It is important to thoroughly investigate and analyze the impact of semisupervised and unsupervised learning on semantic segmentation networks, specifically focusing on ensuring high model extraction accuracy.
- *Lighter weight and more efficient network structure:* Existing models for extracting roads from remote sensing images have shown improved accuracy, but they tend to be slower and require more time in comparison. It is expected that there will be a significant increase in the demand for real-time road extraction from remote sensing images in the coming years. When investigating lightweight road extraction, it is essential to prioritize precision. The forthcoming models will likely showcase the capacity to extract roads with exceptional speed and accuracy. Hence, the future of research in this particular field is anticipated to be marked by significant endeavors.

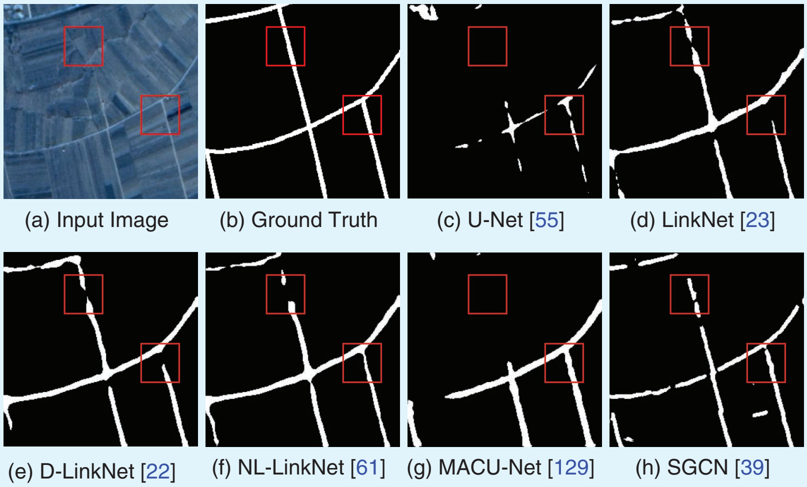

FIGURE 14. (a)–(h) Comparison of visual results of different road extraction methods on the Gansu [39] dataset.

CONCLUSIONS

This study provides an in-depth review of methods for extracting roads from remote sensing images using deep learning techniques. The analysis encompasses widely used datasets and evaluation criteria, along with a thorough comparison of various conventional methodologies. Additionally, a visual representation of the results is provided, further enhancing the comprehensiveness of this review.

While there have been recent advancements in remote sensing image road extraction methods, particularly those that use deep learning, there are still substantial obstacles that need to be addressed. Some challenges in this field include the limited availability of high-precision datasets and the difficulty of extracting obscured roads.

Our analysis has identified two primary areas that should be the main focus of research and development priorities. First, the development of methodologies that enable the design of lightweight and efficient models could revolutionize the field. Second, exploring the potential benefits that can be achieved by combining unsupervised learning algorithms could open up new avenues of research and innovation.

The present review is expected to provide valuable insights for researchers seeking to enhance their comprehension of using remote sensing image road extraction. Consequently, this is likely to contribute to the ongoing expansion of this field.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of Shandong Province (ZR2022QF037).

AUTHOR INFORMATION

Xuan Wang (xuanwang91@ytu.edu.cn) was as born in Weihai, Shandong, China in 1991. She received her B.S. and Ph.D. degrees from Traffic Information Engineering and Control, Chang'an University, China, in 2013 and 2018, respectively. She is currently an associate professor at the School of Computer Science and Control Engineering in Yantai University, Yantai 264005, China. Her research interests include intelligent traffic control, artificial Intelligence and computer vision. She is a Senior Member of IEEE.

Xizhi Jin (jinxizhi@s.ytu.edu.cn) was born in Zibo, Shandong, China in 2001. He received his B.S. degree from Yantai University, China, in 2023. He is currently pursuing a B.S. degree in the Guangdong Laboratory of Artificial Intelligence and Digital Economy, Shenzhen 518107, China. His research interests include deep learning and remote sensing image processing.

Zhe Dai (zhedai@chd.edu.cn) is a lecturer at the College of Transportation Engineering, Chang'an University, Xi'an 710064, China. Currently, he is leading five research projects, including the China Postdoctoral Science Foundation and the Central University Fund, among others. He has published more than 10 academic articles, and has been granted five invention patents. His research interests include intelligent transportation systems, traffic information perception, and deep learning.

Yuxuan Wu (wuyuxuan@chd.edu.cn) was born in Nanchang, Jiangxi, China in 2000. He received his B.S. degree in transportation engineering from Chang'an University, China, in 2022. He is currently pursuing a Ph.D. degree at Chang'an University, Xi'an 710064, China. His research interests include intelligent transportation systems, traffic information perception, and accident prevention in transportation systems.

Abdellah Chehri (chehri@rmc.ca) is an associate professor at the Royal Military College of Canada, Kingston, ON K7K 7B4, Canada. Before joining the Royal Military College, he was an associate professor at the University of Quebec Chicoutimi (UQAC). He was an affiliate professor at the University of Quebec in Outaouais, UQAC, and an adjunct professor at the University of Ottawa. He has served as guest/associate editor for several well-reputed journals. He is a member of the IEEE Communication Society, IEEE Vehicular Technology Society, and IEEE Photonics Society. He is a Senior Member of IEEE.

REFERENCES

[1] B. Feizizadeh, Z. Ronagh, S. Pourmoradian, H. Abedi Gheshlaghi, T. Lakes, and T. Blaschke, "An efficient GIS-based approach for sustainability assessment of urban drinking water consumption patterns: A study in Tabriz city, Iran," *Sustain. Cities Soc.*, vol. 64, Jan. 2021, Art no. 102584, doi: 10.1016/j.scs.2020.102584.

[2] S. Shao, L. Xiao, L. Lin, C. Ren, and J. Tian, "Road extraction convolutional neural network with embedded attention mechanism for remote sensing imagery," *Remote Sens.*, vol. 14, no. 9, 2022, Art no. 2061, doi: 10.3390/rs14092061.

[3] Z. Chen et al., "Road extraction in remote sensing data: A survey," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 112, Aug. 2022, Art no. 102833, doi: 10.1016/j.jag.2022.102833.

[4] J. Senthilnath, N. Varia, A. Dokania, G. Anand, and J. A. Benediktsson, "Deep TEC: Deep transfer learning with ensemble classifier for road extraction from UAV imagery," *Remote Sens.*, vol. 12, no. 2, 2020, Art no. 245, doi: 10.3390/rs12020245.

[5] D. Reagan, A. Sabato, and C. Niezrecki, "Feasibility of using digital image correlation for unmanned aerial vehicle structural health monitoring of bridges," *Struct. Health Monit.*, vol. 17, no. 5, pp. 1056–1072, 2018, doi: 10.1177/1475921717735326.

[6] X. Sun et al., "FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 184, pp. 116–130, Feb. 2022, doi: 10.1016/j.isprsjprs.2021.12.004.

[7] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review," *Remote Sens.*, vol. 12, no. 9, 2020, Art no. 1444, doi: 10.3390/rs12091444.

[8] S. Valero, J. Chanussot, J. A. Benediktsson, H. Talbot, and B. Waske, "Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images," *Pattern Recognit. Lett.*, vol. 31, no. 10, pp. 1120–1127, 2010, doi: 10.1016/j.patrec.2009.12.018.

[9] P. Deepan, S. Abinaya, G. Haritha, and V. Iswarya, "Road recognition from remote sensing imagery using machine learning," *Int. Res. J. Eng. Technol.*, vol. 5, no. 3, pp. 3677–3683, 2018.

[10] J. Zhang, L. Chen, C. Wang, L. Zhuo, Q. Tian, and X. Liang, "Road recognition from remote sensing imagery using incremental learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 2993–3005, Nov. 2017, doi: 10.1109/TITS.2017.2665658.

[11] B. Liu, H. Wu, Y. Wang, and W. Liu, "Main road extraction from ZY-3 grayscale imagery based on directional mathematical morphology and VGI prior knowledge in urban areas," *PLoS One*, vol. 10, no. 9, 2015, Art no. e0138071, doi: 10.1371/journal.pone.0138071.

[12] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," *J. Traffic Transp. Eng. (English Edition)*, vol. 3, no. 3, pp. 271–282, 2016, doi: 10.1016/j.jtte.2016.05.005.

[13] R. Lian, W. Wang, N. Mustafa, and L. Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5489–5507, 2020, doi: 10.1109/JSTARS.2020.3023549.

[14] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, May 2021, Art no. 114417, doi: 10.1016/j.eswa.2020.114417.

[15] L. Mou and X. X. Zhu, "RiFCN: Recurrent network in fully convolutional network for semantic segmentation of high resolution remote sensing images," 2018, *arXiv:1805.02091.*

[16] R. Dong, X. Pan, and F. Li, "DenseU-Net-based semantic segmentation of small objects in urban remote sensing images,"

IEEE Access, vol. 7, pp. 65,347–65,356, 2019, doi: 10.1109/ACCESS.2019.2917952.

[17] M.-T. Pham, L. Courtrai, C. Friguet, S. Lefèvre, and A. Baussard, "Yolo-fine: One-stage detector of small objects under various backgrounds in remote sensing images," *Remote Sens.*, vol. 12, no. 15, 2020, Art no. 2501, doi: 10.3390/rs12152501.

[18] H. Wang, X. Chen, T. Zhang, Z. Xu, and J. Li, "CCTNet: Coupled CNN and transformer network for crop segmentation of remote sensing images," *Remote Sens.*, vol. 14, no. 9, 2022, Art no. 1956, doi: 10.3390/rs14091956.

[19] X. Zhou et al., "Edge-guided recurrent positioning network for salient object detection in optical remote sensing images," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 539–552, Jan. 2023, doi: 10.1109/TCYB.2022.3163152.

[20] R. Alshehhi, P. Reddy Marpu, W. L. Woon, and M. D. Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 139–149, Aug. 2017, doi: 10.1016/j.isprsjprs.2017.05.002.

[21] D. Costea, A. Marcu, E. Slusanschi, and M. Leordeanu, "Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 2100–2109, doi: 10.1109/ICCVW.2017.246.

[22] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 192–1924, doi: 10.1109/CVPRW.2018.00034.

[23] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Piscataway, NJ, USA: IEEE, 2017, pp. 1–4, doi: 10.1109/VCIP.2017.8305148.

[24] Y. Zhang, J. He, X. Kan, G. Xia, L. Zhu, and T. Ge, "Summary of road extraction methods for remote sensing images," *Comput. Eng. Appl.*, vol. 54, no. 13, pp. 1–10, 2018.

[25] P. Liu, Q. Wang, G. Yang, L. Li, and H. Zhang, "Survey of road extraction methods in remote sensing images based on deep learning," *PFG–J. Photogramm. Remote Sensing Geoinf. Sci.*, vol. 90, no. 2, pp. 135–159, 2022, doi: 10.1007/s41064-022-00194-z.

[26] V. Mnih, *Machine Learning for Aerial Image Labeling.* Toronto, ON, Canada: Univ. of Toronto, 2013.

[27] I. Demir et al., "DeepGlobe 2018: A challenge to parse the earth through satellite images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 172–17209, doi: 10.1109/CVPRW.2018.00031.

[28] C. Yang and Z. Wang, "An ensemble Wasserstein generative adversarial network method for road extraction from high resolution remote sensing images in rural areas," *IEEE Access*, vol. 8, pp. 174,317–174,324, 2020, doi: 10.1109/ACCESS.2020.3026084.

[29] M. Zhou, H. Sui, S. Chen, J. Wang, and X. Chen, "BT-RoadNet: A boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 288–306, Oct. 2020, doi: 10.1016/j.isprsjprs.2020.08.019.

[30] Z. Chen, C. Wang, J. Li, N. Xie, Y. Han, and J. Du, "Reconstruction bias U-Net for road extraction from optical remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2284–2294, 2021, doi: 10.1109/JSTARS.2021.3053603.

[31] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "SpaceNet: A remote sensing dataset and challenge series," 2018, *arXiv: 1807.01232*.

[32] J. Xin, X. Zhang, Z. Zhang, and W. Fang, "Road extraction of high-resolution remote sensing images derived from DenseU-Net," *Remote Sens.*, vol. 11, no. 21, 2019, Art no. 2499, doi: 10.3390/rs11212499.

[33] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, Jun. 2017, doi: 10.1109/TGRS.2017.2669341.

[34] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, and Y. Liu, "RoadNet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2043–2056, Apr. 2019, doi: 10.1109/TGRS.2018.2870871.

[35] F. Bastani et al., "RoadTracer: Automatic extraction of road networks from aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4720–4728, doi: 10.1109/CVPR.2018.00496.

[36] Q. Zhu et al., "A global context-aware and batch-independent network for road extraction from VHR satellite imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 353–365, May 2021, doi: 10.1016/j.isprsjprs.2021.03.016.

[37] Z. Chen, W. Fan, B. Zhong, J. Li, J. Du, and C. Wang, "Corse-to-fine road extraction based on local Dirichlet mixture models and multiscale-high-order deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4283–4293, Oct. 2020, doi: 10.1109/TITS.2019.2939536.

[38] Y. Ren, Y. Yu, and H. Guan, "DA-CapsUNet: A dual-attention capsule U-Net for road extraction from remote sensing imagery," *Remote Sens.*, vol. 12, no. 18, 2020, Art no. 2866, doi: 10.3390/rs12182866.

[39] G. Zhou, W. Chen, Q. Gui, X. Li, and L. Wang, "Split depthwise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021, doi: 10.1109/TGRS.2021.3128033.

[40] X. Zhang et al., "Complex mountain road extraction in high-resolution remote sensing images via a Light Roadformer and a new benchmark," *Remote Sens.*, vol. 14, no. 19, 2022, Art no. 4729, doi: 10.3390/rs14194729.

[41] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Proc. 11th Eur. Conf. Comput. Vis.: Part VI*, Heraklion, Crete, Greece: Springer, Sep. 5–11, 2010, pp. 210–223.

[42] J. Wang, J. Song, M. Chen, and Z. Yang, "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," *Int. J. Remote Sens.*, vol. 36, no. 12, pp. 3144–3169, 2015, doi: 10.1080/01431161.2015.1054049.

[43] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with convolutional neural networks,"

J. Electron. Imaging, vol. 28, no. 10, pp. 1–9, 2016, doi: 10.2352/ISSN.2470-1173.2016.10.ROBVIS-392.

[44] P. Li et al., "Road network extraction via deep learning and line integral convolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE, 2016, pp. 1599–1602, doi: 10.1109/IGARSS.2016.7729408.

[45] Y. Wei, Z. Wang, and M. Xu, "Road structure refined CNN for road extraction in aerial image," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 709–713, May 2017, doi: 10.1109/LGRS.2017.2672734.

[46] J. Yu, X. Ye, and Q. Tu, "Traffic sign detection and recognition in multiimages using a fusion model with YOLO and VGG network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16,632–16,642, Sep. 2022, doi: 10.1109/TITS.2022.3170354.

[47] L. Gao, W. Song, J. Dai, and Y. Chen, "Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network," *Remote Sens.*, vol. 11, no. 5, 2019, Art no. 552, doi: 10.3390/rs11050552.

[48] A. Abdollahi, B. Pradhan, and N. Shukla, "Road extraction from high-resolution orthophoto images using convolutional neural network," *J. Indian Soc. Remote Sens.*, vol. 49, no. 3, pp. 569–583, 2021, doi: 10.1007/s12524-020-01228-y.

[49] N. Varia, A. Dokania, and J. Senthilnath, "DeepExt: A convolution neural network for road extraction using RGB images captured by UAV," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Piscataway, NJ, USA: IEEE, 2018, pp. 1890–1895, doi: 10.1109/SSCI.2018.8628717.

[50] X. Zhang, W. Ma, C. Li, J. Wu, X. Tang, and L. Jiao, "Fully convolutional network-based ensemble method for road extraction from aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1777–1781, Oct. 2020, doi: 10.1109/LGRS.2019.2953523.

[51] C. Henry, S. Majid Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018, doi: 10.1109/LGRS.2018.2864342.

[52] Y. Zhang, G. Xia, J. Wang, and D. Lha, "A multiple feature fully convolutional network for road extraction from high-resolution remote sensing image over mountainous areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1600–1604, Oct. 2019, doi: 10.1109/LGRS.2019.2905350.

[53] R. Kestur, S. Farooq, R. Abdal, E. Mehraj, O. Narasipura, and M. Mudigere, "UFCN: A fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle," *J. Appl. Remote Sens.*, vol. 12, no. 1, 2018, Art no. 016020, doi: 10.1117/1.JRS.12.016020.

[54] D. Pan, M. Zhang, and B. Zhang, "A generic FCN-based approach for the road-network extraction from VHR remote sensing images–using OpenStreetMap as benchmarks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2662–2673, 2021, doi: 10.1109/JSTARS.2021.3058347.

[55] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assisted Interven. (MICCAI)*, Munich, Germany: Springer, October 5–9, 2015, pp. 234–241.

[56] S. Li and X. Liu, "Multi-type road extraction and analysis of high-resolution images with D-LinkNet50," in *Proc. 3rd Int. Conf. Geol. Mapping Remote Sens. (ICGMRS)*, Piscataway, NJ, USA: IEEE, 2022, pp. 244–248, doi: 10.1109/ICGMRS55602.2022.9849390.

[57] Y. Li, B. Peng, L. He, K. Fan, Z. Li, and L. Tong, "Road extraction from unmanned aerial vehicle remote sensing images based on improved neural networks," *Sensors*, vol. 19, no. 19, 2019, Art no. 4115, doi: 10.3390/s19194115.

[58] Q. Gu, B. Xue, S. Ruan, and X. Li, "A road extraction method for intelligent dispatching based on MD-LinkNeSt network in open-pit mine," *Int. J. Min. Reclam. Environ.*, vol. 35, no. 9, pp. 656–669, 2021, doi: 10.1080/17480930.2021.1949800.

[59] Z. Xu et al., "Road extraction in mountainous regions from high-resolution images based on DSDNet and terrain optimization," *Remote Sens.*, vol. 13, no. 1, 2020, Art no. 90, doi: 10.3390/rs13010090.

[60] K. Zhou, Y. Xie, Z. Gao, F. Miao, and L. Zhang, "FuNet: A novel road extraction network with fusion of location data and remote sensing imagery," *ISPRS Int. J. Geo-Inform.*, vol. 10, no. 1, 2021, Art no. 39, doi: 10.3390/ijgi10010039.

[61] Y. Wang, J. Seo, and T. Jeon, "NL-LinkNet: Toward lighter but more accurate road extraction with nonlocal operations," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3050477.

[62] A. Tran, A. Zonoozi, J. Varadarajan, and H. Kruppa, "PP-LinkNet: Improving semantic segmentation of high resolution satellite imagery with multi-stage training," in *Proc. 2nd Workshop Struct. Understanding Multimedia heritAge Contents*, 2020, pp. 57–64, doi: 10.1145/3423323.3423407.

[63] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[64] Y. Wang et al., "Re-DLinkNet: Based on DLinkNet and ReNet for road extraction from high resolution satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE, 2021, pp. 4664–4667, doi: 10.1109/IGARSS47720.2021.9553728.

[65] C. Miao, C. Liu, Z. Zhang, and Q. Tian, "TransLinkNet: LinkNet with transformer for road extraction," in *Proc. SPIE 12173 Int. Conf. Optics Mach. Vis. (ICOMV)*, 2022, pp. 138–143, doi: 10.1117/12.2634524.

[66] Y. Gu, "MDNet: A multi-modal dual branch road extraction network using infrared information," in *Proc. 3rd Int. Conf. Geol. Mapping Remote Sens. (ICGMRS)*, Piscataway, NJ, USA: IEEE, 2022, pp. 626–631, doi: 10.1109/ICGMRS55602.2022.9849226.

[67] Q. Wang, H. Bai, C. He, and J. Cheng, "FE-LinkNet: Enhanced D-LinkNet with attention and dense connection for road extraction in high-resolution remote sensing images," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, Piscataway, NJ, USA: IEEE, 2022, pp. 3043–3046, doi: 10.1109/IGARSS46834.2022.9883026.

[68] X. Yang et al., "Road detection via deep residual dense U-Net," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Piscataway, NJ, USA: IEEE, 2019, pp. 1–7, doi: 10.1109/IJCNN.2019.8851728.

[69] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018, doi: 10.1109/LGRS.2018.2802944.

[70] A. Buslaev, S. Seferbekov, V. Iglovikov, and A. Shvets, "Fully convolutional network for automatic road extraction from satellite

imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 197–1973, doi: 10.1109/CVPRW.2018.00035.

[71] P. Singh and R. Dash, "A two-step deep convolution neural network for road extraction from aerial images," in *Proc. 6th Int. Conf. Signal Process. Integrated Netw. (SPIN)*, Piscataway, NJ, USA: IEEE, 2019, pp. 660–664, doi: 10.1109/SPIN.2019.8711639.

[72] T. Sun, Z. Chen, W. Yang, and Y. Wang, "Stacked U-Nets with multi-output for road extraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 187–1874, doi: 10.1109/CVPRW.2018.00033.

[73] A. Constantin, J.-J. Ding, and Y.-C. Lee, "Accurate road detection from satellite images using modified U-Net," in *Proc. IEEE Asia Pacific Conf. Circuits Syst. (APCCAS)*, Piscataway, NJ, USA: IEEE, 2018, pp. 423–426, doi: 10.1109/APCCAS.2018.8605652.

[74] Y. Hou, Z. Liu, T. Zhang, and Y. Li, "C-UNet: Complement UNet for remote sensing road extraction," *Sensors*, vol. 21, no. 6, 2021, Art no. 2153, doi: 10.3390/s21062153.

[75] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Multi-object segmentation in complex urban scenes from high-resolution remote sensing data," *Remote Sens.*, vol. 13, no. 18, 2021, Art no. 3710, doi: 10.3390/rs13183710.

[76] Z. Chen, C. Wang, J. Li, W. Fan, J. Du, and B. Zhong, "Adaboost-like end-to-end multiple lightweight U-Nets for road extraction from optical remote sensing images," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 100, 2021, Art no. 102341, doi: 10.1016/j.jag.2021.102341.

[77] M. Yang, Y. Yuan, and G. Liu, "SDUNet: Road extraction via spatial enhanced and densely connected UNet," *Pattern Recognit.*, vol. 126, Jun. 2022, Art no. 108549, doi: 10.1016/j.patcog.2022.108549.

[78] T. Panboonyuen, P. Vateekul, K. Jitkajornwanich, and S. Lawawirojwong, "An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery," in *Proc. 13th Int. Conf. Comput. Inf. Technol. (IC2IT)*, Munich, Germany: Springer, 2018, pp. 191–201.

[79] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[80] J. Doshi, "Residual inception skip network for binary segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 216–219.

[81] L. Ding and L. Bruzzone, "DiResNet: Direction-aware residual network for road extraction in VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10243–10254, Dec. 2021, doi: 10.1109/TGRS.2020.3034011.

[82] S. Wang, X. Mu, D. Yang, H. He, and P. Zhao, "Road extraction from remote sensing images using the inner convolution integrated encoder-decoder network and directional conditional random fields," *Remote Sens.*, vol. 13, no. 3, 2021, Art no. 465, doi: 10.3390/rs13030465.

[83] Z. Ge, Y. Zhao, J. Wang, D. Wang, and Q. Si, "Deep feature-review transmit network of contour-enhanced road extraction from remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021, doi: 10.1109/LGRS.2021.3061764.

[84] Q. Shi, X. Liu, and X. Li, "Road detection from remote sensing images by generative adversarial networks," *IEEE Access*, vol. 6, pp. 25,486–25,494, 2017, doi: 10.1109/ACCESS.2017.2773142.

[85] C.-I. Cira, M.-Á. Manso-Callejo, R. Alcarria, T. F. Pareja, B. B. Sanchez, and F. Serradilla, "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*, vol. 10, no. 1, 2021, Art no. 79, doi: 10.3390/land10010079.

[86] X. Zhang, X. Han, C. Li, X. Tang, H. Zhou, and L. Jiao, "Aerial image road extraction based on an improved generative adversarial network," *Remote Sens.*, vol. 11, no. 8, 2019, Art no. 930, doi: 10.3390/rs11080930.

[87] P. Shamsolmoali, M. Zareapoor, H. Zhou, R. Wang, and J. Yang, "Road segmentation for remote sensing images using adversarial spatial pyramid networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4673–4688, Jun. 2021, doi: 10.1109/TGRS.2020.3016086.

[88] J. Song, J. Li, H. Chen, and J. Wu, "MapGen-GAN: A fast translator for remote sensing image to map via unsupervised adversarial learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2341–2357, 2021, doi: 10.1109/JSTARS.2021.3049905.

[89] A. Abdollahi, B. Pradhan, G. Sharma, K. N. A. Maulud, and A. Alamri, "Improving road semantic segmentation using generative adversarial network," *IEEE Access*, vol. 9, pp. 64,381–64,392, 2021, doi: 10.1109/ACCESS.2021.3075951.

[90] A. Hu, S. Chen, L. Wu, Z. Xie, Q. Qiu, and Y. Xu, "WSGAN: An improved generative adversarial network for remote sensing image road network extraction by weakly supervised processing," *Remote Sens.*, vol. 13, no. 13, 2021, Art no. 2506, doi: 10.3390/rs13132506.

[91] W. Chen, G. Zhou, Z. Liu, X. Li, X. Zheng, and L. Wang, "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.*, vol. 60, pp. 1–15, 2022, doi: 10.1109/TGRS.2022.3188908.

[92] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[93] Y. Lin, D. Xu, N. Wang, Z. Shi, and Q. Chen, "Road extraction from very-high-resolution remote sensing images via a nested SE-Deeplab model," *Remote Sens.*, vol. 12, no. 18, 2020, Art no. 2985, doi: 10.3390/rs12182985.

[94] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[95] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning," *Remote Sens.*, vol. 10, no. 9, 2018, Art no. 1461, doi: 10.3390/rs10091461.

[96] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, *arXiv:1404.1869*.

[97] R. Liu et al., "An intriguing failing of convolutional neural networks and the CoordConv solution," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 9628–9639.

[98] J. Li, Y. Liu, Y. Zhang, and Y. Zhang, "Cascaded attention DenseUNet (CADUNet) for road extraction from very-high-resolution images," *ISPRS Int. J. Geo-Inform.*, vol. 10, no. 5, 2021, Art no. 329, doi: 10.3390/ijgi10050329.

[99] X. Qi, K. Li, P. Liu, X. Zhou, and M. Sun, "Deep attention and multi-scale networks for accurate remote sensing image segmentation," *IEEE Access*, vol. 8, pp. 146,627–146,639, 2020, doi: 10.1109/ACCESS.2020.3015587.

[100] X. Li, Y. Wang, L. Zhang, S. Liu, J. Mei, and Y. Li, "Topology-enhanced urban road extraction via a geographic feature-enhanced network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8819–8830, Dec. 2020, doi: 10.1109/TGRS.2020.2991006.

[101] Z. Liu, M. Wang, F. Wang, and X. Ji, "A residual attention and local context-aware network for road extraction from high-resolution remote sensing imagery," *Remote Sens.*, vol. 13, no. 24, 2021, Art no. 4958, doi: 10.3390/rs13244958.

[102] J. Wan, Z. Xie, Y. Xu, S. Chen, and Q. Qiu, "DA-RoadNet: A dual-attention network for road extraction from high resolution satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6302–6315, 2021, doi: 10.1109/JSTARS.2021.3083055.

[103] Z. Yang, D. Zhou, Y. Yang, J. Zhang, and Z. Chen, "Road extraction from satellite imagery by road context and full-stage feature," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2022.3228967.

[104] Y. Xu, H. Chen, C. Du, and J. Li, "MSACon: Mining spatial attention-based contextual information for road extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2021.3073923.

[105] S.-B. Chen, Y.-X. Ji, J. Tang, B. Luo, W.-Q. Wang, and K. Lv, "DBRANet: Road extraction by dual-branch encoder and regional attention decoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3074524.

[106] Y. Wang et al., "DDU-Net: Dual-decoder-U-Net for road extraction using high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022, doi: 10.1109/TGRS.2022.3197546.

[107] H. Huan, Y. Sheng, Y. Zhang, and Y. Liu, "Strip attention networks for road extraction," *Remote Sens.*, vol. 14, no. 18, 2022, Art no. 4516, doi: 10.3390/rs14184516.

[108] Z. Gong, L. Xu, Z. Tian, J. Bao, and D. Ming, "Road network extraction and vectorization of remote sensing images based on deep learning," in *Proc. IEEE 5th Inf. Technol. Mechatron. Eng. Conf. (ITOEC)*, 2020, pp. 303–307, doi: 10.1109/ITOEC49072.2020.9141903.

[109] D. Xiao, L. Yin, and Y. Fu, "Open-Pit mine road extraction from high-resolution remote sensing images using RATT-UNet," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021, doi: 10.1109/LGRS.2021.3065148.

[110] J. Mei, R.-J. Li, W. Gao, and M.-M. Cheng, "CoANet: Connectivity attention network for road extraction from satellite imagery," *IEEE Trans. Image Process.*, vol. 30, pp. 8540–8552, 2021, doi: 10.1109/TIP.2021.3117076.

[111] L. Dai, G. Zhang, and R. Zhang, "RADANet: Road augmented deformable attention network for road extraction from complex high-resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023, doi: 10.1109/TGRS.2023.3237561.

[112] S. D. Khan, L. Alarabi, and S. Basalamah, "DSMSA-Net: Deep spatial and multi-scale attention network for road extraction in high spatial resolution satellite images," *Arab. J. Sci. Eng.*, vol. 48, no. 2, pp. 1907–1920, 2023, doi: 10.1007/s13369-022-07082-z.

[113] M. Lan, Y. Zhang, L. Zhang, and B. Du, "Global context based automatic road segmentation via dilated convolutional neural network," *Inf. Sci.*, vol. 535, pp. 156–171, Oct. 2020, doi: 10.1016/j.ins.2020.05.062.

[114] Y. Li, L. Xu, J. Rao, L. Guo, Z. Yan, and S. Jin, "A Y-Net deep learning method for road segmentation using high-resolution visible remote sensing images," *Remote Sens. Lett.*, vol. 10, no. 4, pp. 381–390, 2019, doi: 10.1080/2150704X.2018.1557791.

[115] H. Yan, C. Zhang, J. Yang, M. Wu, and J. Chen, "Did-Linknet: Polishing D-block with dense connection and iterative fusion for road extraction," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE, 2021, pp. 2186–2189, doi: 10.1109/IGARSS47720.2021.9554534.

[116] X. Tan, Z. Xiao, Q. Wan, and W. Shao, "Scale sensitive neural network for road segmentation in high-resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 533–537, Mar. 2021, doi: 10.1109/LGRS.2020.2976551.

[117] S. Zou, F. Xiong, H. Luo, J. Lu, and Y. Qian, "AF-Net: All-scale feature fusion network for road extraction from remote sensing images," in *Proc. Digital Image Comput. Techn. Appl. (DICTA)*, Piscataway, NJ, USA: IEEE, pp. 1–8, 2021, doi: 10.1109/DICTA52665.2021.9647235.

[118] Q. Wu, F. Luo, P. Wu, B. Wang, H. Yang, and Y. Wu, "Automatic road extraction from high-resolution remote sensing images using a method based on densely connected spatial feature-enhanced pyramid," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3–17, 2020, doi: 10.1109/JSTARS.2020.3042816.

[119] X. Yang, X. Li, Y. Ye, R. Y. Lau, X. Zhang, and X. Huang, "Road detection and centerline extraction via deep recurrent convolutional neural network U-Net," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7209–7220, Sep. 2019, doi: 10.1109/TGRS.2019.2912301.

[120] X. Lu et al., "Multi-scale and multi-task deep learning framework for automatic road extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9362–9377, Nov. 2019, doi: 10.1109/TGRS.2019.2926397.

[121] F. Yi, R. Te, Y. Zhao, and G. Xu, "EUNetMTL: Multitask joint learning for road extraction from high-resolution remote sensing images," *Remote Sens. Lett.*, vol. 13, no. 3, pp. 258–268, 2022, doi: 10.1080/2150704X.2021.2019344.

[122] Y. Wei, K. Zhang, and S. Ji, "Simultaneous road surface and centerline extraction from large-scale remote sensing images using CNN-based segmentation and tracing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8919–8931, Dec. 2020, doi: 10.1109/TGRS.2020.2991733.

[123] Z. Shao, Z. Zhou, X. Huang, and Y. Zhang, "MRENet: Simultaneous extraction of road surface and road centerline in complex urban scenes from very high-resolution images," *Remote Sens.*, vol. 13, no. 2, 2021, Art no. 239, doi: 10.3390/rs13020239.

[124] X. Wei, X. Lv, and K. Zhang, "Road extraction in SAR images using ordinal regression and road-topology loss," *Remote Sens.*, vol. 13, no. 11, 2021, Art no. 2080, doi: 10.3390/rs13112080.

[125] X. Lu et al., "Cascaded multi-task road extraction network for road surface, centerline, and edge extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2022.3165817.

[126] Y. Li, L. Guo, J. Rao, L. Xu, and S. Jin, "Road segmentation based on hybrid convolutional network for high-resolution visible remote sensing image," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 613–617, Apr. 2019, doi: 10.1109/LGRS.2018.2878771.

[127] R. Liu et al., "Multiscale road centerlines extraction from high-resolution aerial imagery," *Neurocomputing*, vol. 329, pp. 384–396, Feb. 2019, doi: 10.1016/j.neucom.2018.10.036.

[128] R. Li, C. Duan, and S. Zheng, "MACU-Net semantic segmentation from high-resolution remote sensing images," 2020, *arXiv:2007.13083*.

[129] Z. Zhang, X. Sun, and Y. Liu, "GMR-Net: Road-extraction network based on fusion of local and global information," *Remote Sens.*, vol. 14, no. 21, 2022, Art no. 5476, doi: 10.3390/rs14215476.

[130] A. Wulamu, Z. Shi, D. Zhang, and Z. He, "Multiscale road extraction in remote sensing images," *Comput. Intell. Neurosci.*, vol. 2019, Jul. 2019, Art no. 2373798, doi: 10.1155/2019/2373798.

[131] H. He, D. Yang, S. Wang, S. Wang, and Y. Li, "Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss," *Remote Sens.*, vol. 11, no. 9, 2019, Art no. 1015, doi: 10.3390/rs11091015.

[132] Y. Jie et al., "MECA-Net: A multiscale feature encoding and long-range context-aware network for road extraction from remote sensing images," *Remote Sens.*, vol. 14, no. 21, 2022, Art no. 5342, doi: 10.3390/rs14215342.

[133] Z. Yang, D. Zhou, Y. Yang, J. Zhang, and Z. Chen, "TransRoadNet: A novel road extraction method for remote sensing images via combining high-level semantic feature and context," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3171973.

[134] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9992–10022, doi: 10.1109/ICCV48922.2021.00986.

[135] C. Ge, Y. Nie, F. Kong, and X. Xu, "Improving road extraction for autonomous driving using Swin transformer UNet," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Piscataway, NJ, USA: IEEE, 2022, pp. 1216–1221, doi: 10.1109/ITSC55140.2022.9922395.

[136] T. Chen, D. Jiang, and R. Li, "Swin transformers make strong contextual encoders for VHR image road extraction," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE, 2022, pp. 3019–3022, doi: 10.1109/IGARSS46834.2022.9883628.

[137] L. Luo, J.-X. Wang, S.-B. Chen, J. Tang, and B. Luo, "BDTNet: Road extraction by bi-direction transformer from remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3183828.

[138] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907.*

[139] F. Cui, R. Feng, L. Wang, and L. Wei, "Joint superpixel segmentation and graph convolutional network road extraction for high-resolution remote sensing imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE, 2021, pp. 2178–2181, doi: 10.1109/IGARSS47720.2021.9554635.

[140] F. Cui, Y. Shi, R. Feng, L. Wang, and T. Zeng, "A graph-based dual convolutional network for automatic road extraction from high resolution remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE, 2022, pp. 3015–3018, doi: 10.1109/IGARSS46834.2022.9883088.

[141] Y. Tao, M. Xu, F. Zhang, B. Du, and L. Zhang, "Unsupervised-restricted deconvolutional neural network for very high resolution remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6805–6823, Dec. 2017, doi: 10.1109/TGRS.2017.2734697.

[142] Y. Tao, M. Xu, Y. Zhong, and Y. Cheng, "GAN-assisted two-stream neural network for high-resolution remote sensing image classification," *Remote Sens.*, vol. 9, no. 12, 2017, Art no. 1328, doi: 10.3390/rs9121328.

[143] Y. Wei and S. Ji, "Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2021, doi: 10.1109/TGRS.2021.3061213.

[144] G. Yuan, J. Li, X. Liu, and Z. Yang, "Weakly supervised road network extraction for remote sensing image based scribble annotation and adversarial learning," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7184–7199, 2022, doi: 10.1016/j.jksuci.2022.05.020.

[145] M. Zhou, H. Sui, S. Chen, J. Liu, W. Shi, and X. Chen, "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.*, vol. 193, pp. 234–251, Nov. 2022, doi: 10.1016/j.isprsjprs.2022.09.005.

[146] R. Lian and L. Huang, "DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1905–1916, 2020, doi: 10.1109/JSTARS.2020.2983788.

[147] R. Lian and L. Huang, "Weakly supervised road segmentation in high-resolution remote sensing images using point annotations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021, doi: 10.1109/TGRS.2021.3059088.

[148] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

GRS