ISSN: 2424-8460(O)
2251-2608(P)

Original Research Article

A review of research progress in deep learning methods for road extraction from remote sensing images

Wei Jiang [1, 2], Shamsul Arrieya Bin Ariffin [3], Yanpeng Li [2, *]
[1] *City University Malaysia, Kuala Lumpur, 46100, Malaysia*
[2] *Jiangxi College of Applied Technology, Jiangxi, Gan zhou, 341000, China*
[3] *Sultan Idris Education University, Tanjung Malim Perak, 35900, Malaysia*

Abstract: Driven by the combination of high-resolution remote sensing and deep learning, road extraction research has advanced along three main lines (2023-2025): semi- and weakly supervised learning, CNN-Transformer hybrid architectures, and graph/vector modeling. Semi-supervision has proven effective in robustly improving road connectivity in annotation-scarce scenarios. Hybrid architectures that fuse local texture and global context are gradually replacing single CNN/Transformer architectures in feature representation. Direct prediction of road network graphs (nodes and edges) has significantly improved topological consistency and achieved breakthroughs in large-scale inference efficiency. This paper reviews representative methods and datasets/evaluation systems, and provides a horizontal comparison based on publicly available comparison results from the past two years. It summarizes current bottlenecks and trends: Learning paradigms focused on connectivity, cross-domain generalization, and zero-shot geographic transfer, along with more economical supervisory signals, will be the primary areas of focus.
Keywords: deep learning; remote sensing imagery; road extraction; semi-supervised learning; transformer

1. Introduction

Road extraction, a core task in automated mapping of basic geographic features, has long been plagued by challenges such as elongated objects, scale variations, occlusion, and shadows. Since 2023, researchers have used semi-supervision and weak supervision to alleviate the annotation costs and "long-tail sample"problem. Furthermore, structural design has evolved from CNNs to a hybrid CNN+Transformer approach to balance local boundaries and global geometry [1–3]. Simultaneously, the vectorization/graph learning approach has shifted from "pixel segmentation → post-processing vectorization"to end-to-end road network prediction, achieving leadership in topological metrics (TOPO/APLS) and significantly improving city-level large-scale inference speed [4,5]. These advances collectively point to the need for simultaneous advancements in supervision paradigms, structural design, and representation spaces to maintain a connected, complete, and usable road network output in complex urban scenarios. This article provides a comprehensive review, supporting data, and comparative methods to highlight the necessity and direction of future research.

2. Methodology (2023-2025)

2.1. Semi-/weak supervision: Less annotation for stronger generalization

SemiRoadExNet (ISPRS JPRS, 2023) introduces adversarial semi-supervision into road segmentation.

ISSN: 2424-8460(O)
2251-2608(P)

On DeepGlobe and other data , the IoU is improved by 0.96–5.3.8% compared with various semi-supervised segmentation methods. At low annotation rates (5%/10%/20%), the IoU is improved by 3.70%/12.2.5.%/10.93% respectively, demonstrating its effectiveness in sparsely labeled scenarios [1]. Furthermore, the semi-supervised strategy for road tasks was systematically verified in the 2024 IJGI research and can be used as a priority option for reducing costs and increasing efficiency in engineering practice [6] .

2.2. CNN-transformer hybrid: Balancing boundaries and global geometry

RoadCT proposed in 2024 combines CNN local representation with Transformer global dependency. It is reported that F1 and IoU are improved by about 1.1.–3.9%/1.6–6.0% relative to SOTA, showing the advantages of hybrid structure in slender targets and cross-scale receptive fields [7] . At the same time, SASwin Transformer combines spatial attention and SMLP modules with the Swin backbone, achieving IoU = 65.04% and 65.60% on Massachusetts and DeepGlobe respectively , which is 1.88%/1.84% higher than D- LinkNet (2024), and the statistical test results are significant, verifying the effectiveness of local-global collaboration [2] . In addition, the solution combining frequency domain features (FFT) and global spatial learning was also reported in 2024, pointing to the potential of cross-domain information fusion [9] .

2.3. Vectorization/Graph learning: Directly predicting road network graphs

SAM-Road (CVPR 2024 Workshops) uses the basic model SAM for geometric-topological two-stage road network prediction: first, it generates high-quality road/intersection probability masks to extract vertices, and then uses a lightweight Transformer-GNN to predict edge connectivity. This method achieves or exceeds the existing SOTA in TOPO and APLS indicators on City-scale and SpaceNet , and is 40× faster than RNGDet ++ on the city-level test set , demonstrating the engineering advantages of parallel sliding window reasoning [4] . In addition, the 2024 DSVNet improves topological consistency on high-resolution images through vertex supervision and deformable attention, adding evidence to the "direct graph modeling"route [5] .

3. Comparison and analysis of public results in the past two years

To facilitate horizontal comparison, we selected two commonly used pixel-level benchmarks from the past two years using the IoU metric, and compiled the results from the 2024 SASwin public **Tables 2** and **1** (using the same comparison criteria as in the table itself). As shown in **Table 1** , hybrid/Transformer models generally outperform pure CNNs on both Massachusetts and DeepGlobe , with a higher relative advantage (SASwin: 65.60%) on DeepGlobe , where occlusion and scale variations are more pronounced.

Table 1. IoU (%) of representative methods on public benchmarks in the past two years.

method	Massachusetts	DeepGlobe	illustrate
D- LinkNet	63.1.6	63.76	CNN residual-hole connection baseline
RoadExNet	64.09	61.49	Structural consistency prior enhancement
RemainNet	64.91	64.1.5	Mask modeling to alleviate occlusion
SASwin	65.04	65.60	Swin+Spatial Attention+SMLP

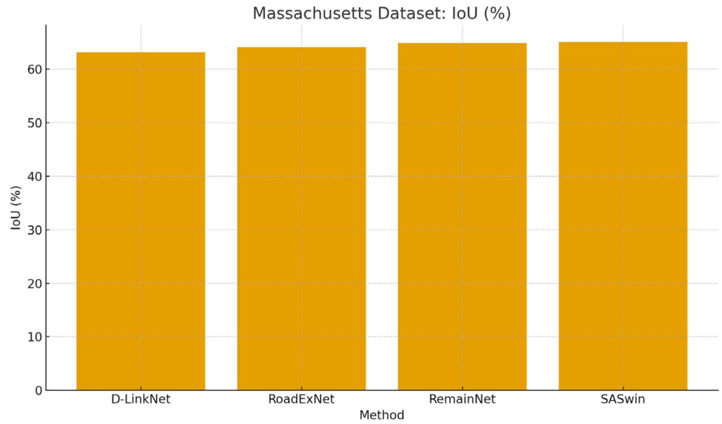

Figure 1. Massachusetts benchmark IoU comparison (%).

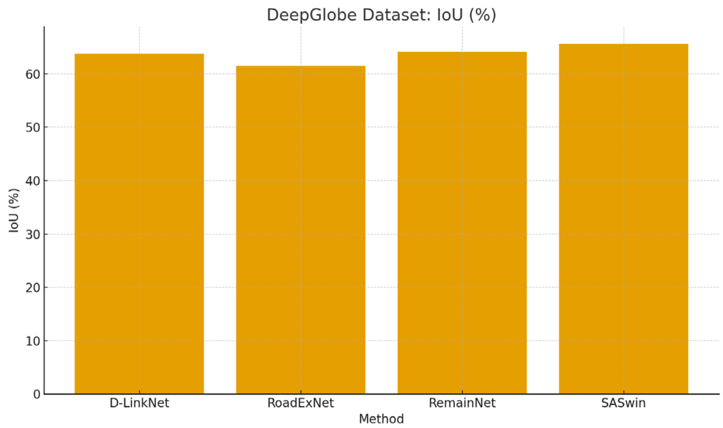

Figure 2. DeepGlobe benchmark IoU comparison (%).

Note: The data used in **Figures 1** and **2** are both quoted from SASwin's public comparison table and are consistent in caliber; only the common strong baselines and representative new methods from 2023 to 2024 are shown for easy intuitive comparison.

Table 2. Methodological comparison of representative works (2023-2025).

category	Representative method (year)	Key Idea	Key data/indicators
Semi-supervised	SemiRoadExNet (2023)	Adversarial semi-supervision improves connectivity and robustness	DeepGlobe /CHN6-CUG; IoU improvement of 0.96–5.3.8%
Hybrid structure	RoadCT (2024)	CNN+Transformer dual decoding/ relational fusion	Two public sets; F1/IoU +1–6%
Hybrid structure	SASwin (2024)	Swin+Spatial Attention+SMLP	Mass./DeepGlobe IoU =65.04/65.60
Graph/Vectorization	SAM-Road (2024)	Mask Geometry + Transformer-GNN Topology	City-scale/ SpaceNet ; TOPO/APLS SOTA; **40×** rate
Graph/Vectorization	DSVNet (2024)	Vertex Deep Supervision + Deformable Attention	High-resolution remote sensing; topological consistency enhancement
Frequency domain fusion	Yang et al. (2024)	Global space + Fourier frequency domain joint	Remote Sensing 16(20):3896

4. Conclusion

Evidence from the past three years indicates that semi- and weak-supervision effectively alleviates the annotation bottleneck, and that CNN-Transformer hybrids have become the mainstream architecture for high-res-

ISSN: 2424-8460(O)
2251-2608(P)

olution remote sensing road extraction. Furthermore, vectorization and graph learning are evolving from a "post-processing step"to an end-to-end objective, offering significant advantages in topological consistency and large-scale reasoning. Future directions worth noting include: (i) end-to-end training with topology as the objective function; (ii) cross-domain self-supervision and integration of geographic priors; (iii) multimodal (SAR/terrain/intersection POI) integration and frequency-domain-spatiotemporal fusion; and (iv) efficient parameter adaptation to underlying visual models.

Funding

This work was supported in part by the Research and Innovation Center for Natural Resources Ecological Balance (Grant No. 2302880001) ; in part by the Research Project on Joint Assimilation Inversion Model of Ozone (O_3) and Fine Particulate Matter (PM2.5.) (Grant No. 3203) ; and in part by the Science and Technology Research Project of Jiangxi Provincial Department of Education (Grant No. GJJ214913).

References

[1] Chen H, Li Z, Wu J, Xiong W, Du C. SemiRoadExNet: A semi-supervised network for road extraction from remote sensing imagery via adversarial learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 198: 169–183.

[2] Zhu X, Huang X, Cao W, et al. Road Extraction from Remote Sensing Imagery with Spatial Attention Based on Swin Transformer[J]. Remote Sensing, 2024, 16(7): 1183.

[3] Mo S, Makungu D, Nhamo L. A Survey of Deep Learning Road Extraction Algorithms for High-Resolution Remote Sensing Images[J]. Sensors, 2024, 24(1): 109.

[4] Hetang C, Xue H, Le C, et al. Segment Anything Model for Road Network Graph Extraction[C]//CVPR 2024 Workshops. 2024: 2556–2566.

[5] Zhao Y, Chen Z, Zhao Z, et al. A Deeply Supervised Vertex Network for Road Network Graph Extraction in High-Resolution Images[J]. International Journal of Applied Earth Observation and Geoinformation, 2024, 133.

[6] Yu H, Zhou J, Li Q, et al. Improved Road Extraction Models through Semi-Supervised Learning[J]. ISPRS International Journal of Geo-Information, 2024, 13(10): 347.

[7] Liu W, Gao S, Zhang C, Yang B. RoadCT: A Hybrid CNN-Transformer Network for Road Extraction From Satellite Imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 2501805.

[8] Li Z, Wang Y, Huang S, et al. RemainNet: Explore Road Extraction from Remote Sensing Images Based on Mask Image Modeling[J]. Remote Sensing, 2023, 15(17): 4215.

[9] Yang H, Zhang Y, Zhou Y, et al. A High-Resolution Remote Sensing Road Extraction Method Based on Global Spatial Learning and Fourier Frequency Domain Learning[J]. Remote Sensing, 2024, 16(20): 3896.

[10] Bai X, Guo L, Huo H, et al. Rse-Net: Road-shape Enhanced Neural Network for Road Extraction in High-Resolution Remote Sensing Image[J]. International Journal of Remote Sensing, 2023, 44(??): 1–22.

[11] Zhang J, Hu X, Wei Y, Zhang L. Road Topology Extraction From Satellite Imagery by Joint Learning of Nodes and Their Connectivity[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5602613.

[12] Luo Z, Zhou K, Tan Y, et al. AD-RoadNet: An Auxiliary-Decoding Road Extraction Network Improving Connectivity While Preserving Multiscale Road Details[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 8049–8062.