Towards Satellite Image Road Graph Extraction: A Global-Scale Dataset and A Novel Method

Pan Yin[1,2*], Kaiyu Li[4*], Xiangyong Cao[1,2†], Jing Yao[6], Lei Liu[7], Xueru Bai[7], Feng Zhou[7], Deyu Meng[2,3,5]
[1]School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China
[2]Ministry of Education Key Laboratory of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an, 710049, China
[3]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, China
[4]School of Software Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
[5]Pengcheng Laboratory [6]Chinese Academy of Sciences [7]Xidian University

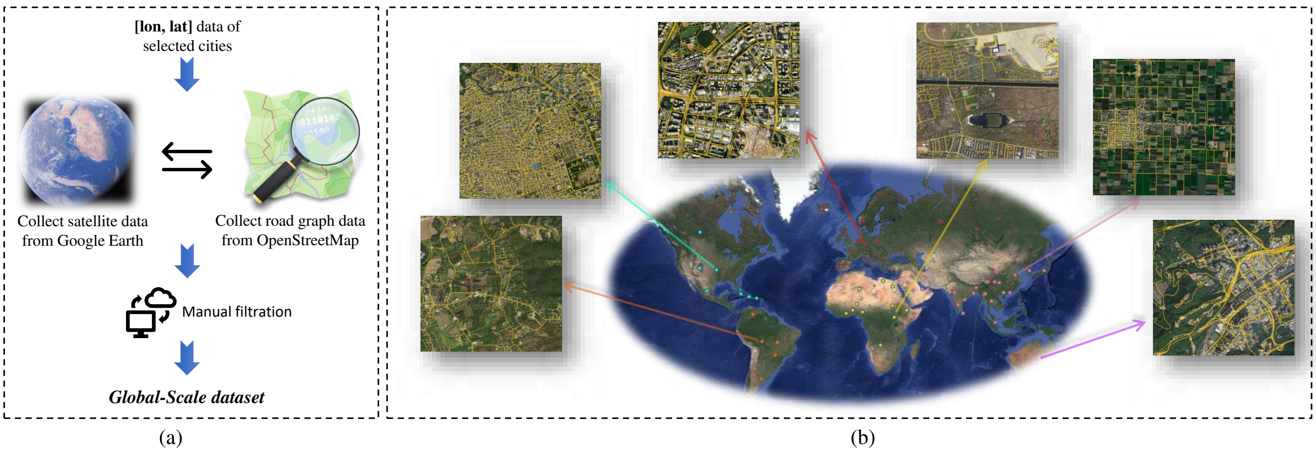

Figure 1. (a) The collection pipeline of our *Global-Scale* dataset; (b) The world map showing the region location of the collected images in the *Global-Scale* dataset.

Abstract

Recently, road graph extraction has garnered increasing attention due to its crucial role in autonomous driving, navigation, etc. However, accurately and efficiently extracting road graphs remains a persistent challenge, primarily due to the severe scarcity of labeled data. To address this limitation, we collect a global-scale satellite road graph extraction dataset, i.e. Global-Scale dataset. Specifically, the Global-Scale dataset is $\sim 20\times$ *larger than the largest existing public road extraction dataset and spans over 13,800* km^2 *globally. Additionally, we develop a novel road graph extraction model, i.e. SAM-Road++, which adopts a node-guided resampling method to alleviate the mismatch issue between training and inference in SAM-Road [17], a pioneering state-of-the-art road graph extraction model. Furthermore, we propose a simple yet effective "extended-line" strategy in SAM-Road++ to mitigate the occlusion issue on the road. Extensive experiments demonstrate the validity of the collected Global-Scale dataset and the proposed SAM-Road++ method, particularly highlighting its superior predictive power in unseen regions. The dataset and code are available at https://github.com/earth-insights/samroadplus.*

*Equal contribution
†Corresponding author (caoxiangyong@mail.xjtu.edu.cn)

1. Introduction

In recent years, daily travel has increasingly relied on navigation systems [34], particularly with the advent of autonomous driving technology [23], which has greatly enhanced convenience in everyday life [37, 53]. These advancements demand higher accuracy and real-time performance in extracting road graphs from satellite images [7, 44]. Existing approaches to end-to-end road graph extraction can be categorized into two main types: iterative methods [3, 49, 50] and global-based methods [2, 14, 15, 17, 48].

Iterative methods generate a road graph point by point from the edges of an image. While effective in road graph extraction, this step-by-step connection process can lead to error accumulation and significant computational burdens. In contrast, global-based methods can directly produce a complete road graph, addressing the limitations of iterative approaches. For instance, SAM-Road [17] performs global road graph prediction in two stages: road segmentation and relationship prediction between key points. However, these two stages are uncoupled during training, the first stage employs a conventional road segmentation model to capture the road mask, while the second stage uses node information from the ground truth to train a classifier for node connectivity. This leads to a mismatch during inference, as the classifier's inputs are based on key points selected from the predicted road mask rather than the training data.

To tackle this mismatch issue, we propose a novel "node sampling" strategy called node-guided resampling. Instead of directly using labeled nodes from the ground truth during classifier training, we resample nodes from the predicted road mask that correspond to the coordinates with the highest probability near the labeled nodes. This approach allows the classifier to leverage training experiences more effectively, resulting in greater consistency between the training and inference processes. Additionally, we recognize that occlusion presents a significant challenge in road graph extraction [51]. Since models can only extract information from a single overhead view in satellite images, accurately determining the connectivity of road nodes becomes difficult when occlusions (e.g., trees, shadows, etc.) obstruct visibility [4, 10]. To tackle this challenge, we aim to provide the classifier with more contextual information for identifying occluded scenes. Our strategy is based on a straightforward assumption: if road lines exist on either side of two neighboring nodes along a straight road, it is likely that a road also connects these two nodes. Motivated by this assumption, we introduce an "extended-line" strategy that utilizes the extended line information between two nodes as an additional criterion for determining road connectivity. This enables the model to better navigate occlusion issues in complex environments.

Beyond the algorithms, another bottleneck constraining the road graph extraction task is the limited availability of data. The most commonly used datasets in existing road graph extraction methods, such as City-Scale [14] and SpaceNet [42], are constrained by either the number of images or their locations, focusing primarily on urban roads [31]. This limited data volume and diversity lead to two main issues: 1) unfaithful evaluations of algorithms and 2) challenges in model generalization capabilities. To address these limitations, we collect a new road graph extraction dataset, ***Global-Scale***, which is nearly $\sim$ 20$\times$ larger than the largest existing public datasets in geographic coverage. *Global-Scale* encompasses all continents except Antarctica and includes roads from urban, rural, mountainous, and other complex environments. Additionally, to provide a more comprehensive benchmark for algorithm evaluation, we design both in-domain and out-of-domain testing sets within *Global-Scale*, aiming to account for the domain differences present in global road networks. The out-of-domain set consists of data from regions not included in the training set, enhancing the robustness of our evaluations. In summary, the main contributions of this work are threefold:

- We establish a novel road graph extraction model, namely SAM-Road++, by coupling the road segmentation and node connectivity prediction sub-networks as a whole. In SAM-Road++, a node-guided resampling strategy is ably introduced to address the mismatch problem between the training and inference phases for the first time.
- To tackle the issue of road occlusion in satellite images, we propose a novel extended-line strategy that utilizes the correlation between resampled key nodes to facilitate the identification and extraction of connected roads.
- We curate a new benchmark dataset for road extraction, *Global-Scale*, which contains the latest satellite images and faithful road graph maps with larger data volumes, broader coverage, and more diverse scenes, enabling a more comprehensive evaluation of road extraction tasks for the community.

2. Related Work

2.1. Existing Road Extraction Datasets

Existing road extraction datasets can be divided into two categories: segmentation-labeled datasets [9, 32, 33, 55] and graph-labeled datasets [14, 42].

Segmentation-labeled data are typical image-mask pairs. The Massachusetts [33] dataset covers a variety of scenes in urban and rural areas, with rich terrain and landform features. The DeepGlobe [9] dataset provides more than 10,000 satellite images in Thailand, Indonesia, and India. However, due to the lack of vector information, these datasets are not suitable for road graph extraction tasks.

Graph-labeled data provides vector graphs of the road network. SpaceNet [42] dataset is first presented in the SpaceNet Challenge. As listed in Tab. 1, it contains images cover mainly Las Vegas, Paris, and Shanghai. However, the coverage of SpaceNet is limited to urban roads, lacking descriptions of complex terrains like farmland and mountainous regions. On the other hand, the City-Scale [14] dataset contains only 180 satellite images of 20 cities in the United States, which still predominantly focuses on urban environments and does not adequately represent non-urban scenes, resulting in an overall insufficient data volume. Consequently, current graph-labeled datasets fall short of meet-

https://spacenet.ai/challenges/

Table 1. Summary of publicly available road extraction datasets. Grey rows indicate datasets that do not contain graph labels. U, R, and M denote urban, rural, and mountainous areas, respectively.

Dataset	Graph Label	Size	Train	Val	Test^{ID}	Test^{OOD}	GSD	Region	Region Type	Time
Massachusetts [33]	✗	$1,500^2$	1,108	14	49	✗	1.0	Massachusetts	U, R	2013
DeepGlobe [9]	✗	$1,024^2$	6,226	243	1,101	✗	0.5	Thailand, Indonesia, India	U, R	2013
SpaceNet [42]	✓	400^2	2,167	✗	567	✗	0.3	Paris, Las Vegas, Shanghai	U	2018
City-Scale [14]	✓	$2,048^2$	144	9	27	✗	1.0	20 city in the U.S.	U	2020
Global-Scale	✓	$2,048^2$	2,375	339	624	130	1.0	Global	U, R, M	2024

ing modern model requirements [19, 46] for large-scale and diverse data coverage.

2.2. Road Graph Extraction

The existing road graph extraction methods can be divided into two categories: segmentation-based method with post-processing and end-to-end graph-based method.

The segmentation-based method [1, 6, 11, 25, 28, 47, 55] leverage deep learning technology [12, 22] to obtain the segmentation mask of the road from images, and then extracts the road graph based on a series of complex post-processing methods [8, 24, 54]. For instance, Gao et al. [11] proposed the Multi-Feature Pyramid Network (MFPN), which uses the Feature Pyramid Network (FPN) [27] to capture multi-scale semantic features and weighted balanced loss function, improving road extraction accuracy. In contrast, Li et al. [25] developed a CNN-based [20, 26] framework that extracts road features from small Synthetic Aperture Radar (SAR) image patches [35], identifies candidate road regions and connectivity. Although segmentation-based methods can generate road graph, their performance in terms of topological connectivity is limited [45].

The graph-based methods are gradually emerging to get a better graph of the road topology in an end-to-end fashion. It can be further divided into iteration-based method [3, 49, 50] and global-based method [14, 15, 17]. The iteration-based methods build the complete road structure by predicting road nodes step by step. The early RoadTracer [3] model started from the initial node and gradually built a road graph with fixed angles and step sizes. RNGDet [49] and RNGDet++ [50] combine CNN and Transformer [41, 43, 52] to extract features and iteratively predict vertices, which significantly improves the ability to capture the global structure of the road. However, due to point-by-point iteration, this type of method is time-consuming, and errors tend to accumulate as points are iterated. In contrast, the global-based method [14, 15, 17] can directly generate a complete road graph with significantly improved efficiency. For example, Sat2Graph [14] trains the model through graph coding tensors, directly predicts the graph tensor coding of roads, and generates vector graphs through post-processing. Notably, SAM-Road [17] first extracts segmentation masks, and then uses key point location information (from segmentation masks) to directly predict global connectivity. While SAM-Road reduces the post-processing steps, its dependence on node labels during training leads to a mismatch between the training and inference phases.

3. *Global-Scale* Dataset

A major challenge facing the road graph extraction task is the lack of comprehensive datasets and benchmarks [30, 39]. Table 1 shows the survey results of currently available road extraction datasets, revealing several key issues. First, the existing graph-labeled datasets are small in scale and mainly focus on urban areas. In contrast, segmentation-labeled datasets are not only larger in scale but also cover multiple types such as cities and farmland. Second, current road extraction datasets are limited to the city scale, lacking a global scale road extraction dataset. Finally, although the training set of SpaceNet is large in scale, the size of each image is small (400×400), resulting in a limited coverage area, and roads are usually elongated and connectivity [56], such a size may only include some shorter roads, resulting in limited contextual information. Although City-Scale provides a larger image size (2,048 $\times$ 2,048), which solves the coverage problem, it has a small number of images, especially the test set has only 27 images, which makes the model evaluation susceptible to significant impact from the performance of a single image.

Our *Global-Scale* dataset is a comprehensive road graph resource covering all continents except Antarctica, designed to address the gaps in existing graph-labeled datasets. Figure 1 illustrates the collection methodology of the *Global-Scale* dataset. Specifically, we manually selected the longitude and latitude of various types of roads, including urban, rural, and mountainous, using Google Earth [29, 36]. We then gathered satellite images from the Google Static Map API [40] based on the selected latitude and longitude information, as well as from the commonly used open-source OpenStreetMap [13] database to obtain corresponding road graph data as ground truth. Each image has a spatial resolution of 1m/pixel, following the standards set by [14]. However, the annotation completeness of OpenStreetMap can be affected by variations in population size and economic disparities across regions [16]. To ensure the annotation com-

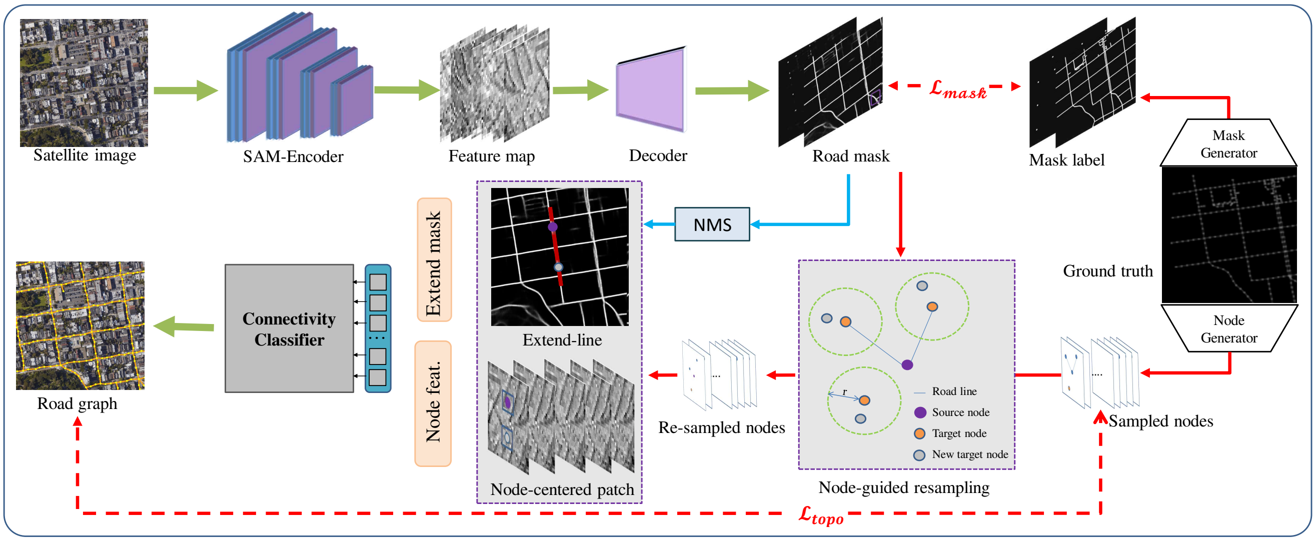

Figure 2. Overview of our proposed SAM-Road++. The red line indicates ***training only*** and the blue line indicates ***inference only***. The satellite image is fed into SAM [21] Encoder and Decoder to get the feature map and road mask. During training, the proposed node-guided resampling uses the ground truth and road mask to get the re-sampled nodes. In the inference stage, the nodes are obtained using only road masks through Non-Maximum Suppression (NMS). Finally, the connectivity classifier determines whether a road exists between nodes based on the node features and the extended line between the nodes. Both loss functions $\mathcal{L}_{mask}$ and $\mathcal{L}_{topo}$ are binary cross-entropy loss, where $\mathcal{L}_{mask}$ and $\mathcal{L}_{topo}$ are used to supervise the road segmentation and the topological connectivity of the road, respectively.

pleteness of the *Global-Scale* dataset, we conducted a second screening of the annotations after the initial data collection to confirm the integrity of our road graph labels.

Finally, the *Global-Scale* dataset contains 3,468 satellite images, each with a size of 2048 $\times$ 2048, providing extensive spatial coverage and a variety of road types. The dataset is divided into training (2,375 images), validation (339 images), and test sets (624 images). Notably, to evaluate the model's inference capability in unseen cities, we additionally collected 130 images from Hong Kong, Shenzhen, and Lucerne, which are not included in the training set, serving as an out-of-domain test set. This design ensures the generalization ability of the model in different geographical environments can be evaluated. In summary, the *Global-Scale* dataset not only provides abundant training resources but also establishes a robust benchmark for future road graph extraction research, allowing researchers to more comprehensively evaluate and compare the performance of different road graph extraction models.

4. Methodology

4.1. Preliminary

SAM-Road [17] is currently the latest global-based end-to-end road graph extraction method and the first to bring the foundation model [18, 38] into the field of road graph extraction. Assume that the road graph is $G = (V, E)$, where V is the set of road nodes of the road graph (*i.e.* vertices) and E is the set of road lines (*i.e.* edges). The training phase of SAM-Road consists of two stages. In the first stage, an RGB satellite image is fed into the SAM encoder to extract image feature embeddings. The decoder then predicts per-pixel probabilities for the existence of road lines and nodes in the form of masks. In the second stage, SAM-Road randomly samples nodes (position information) and corresponding edges (road information) from the ground truth. Using these node pairs, SAM-Road employs the image embeddings of the node pairs obtained in the previous stage as input to train a road classifier to get the probability of a road existing between two nodes.

During the inference phase, ground truth is not available, therefore, SAM-Road selects road nodes from the masks predicted in the first stage as inputs for the classifier of the second stage. As for the selection of nodes, SAM-Road employs a Non-Maximum Suppression (NMS) strategy [17], which prioritizes nodes with higher prediction probabilities while filtering out surrounding nodes.

However, a significant issue arises from the fact that the classifier trained on ground truth nodes does not perform effectively on the nodes obtained from segmentation masks during the inference process, which indicates a mismatch between training and inference. Consequently, ensuring that the classifier exhibits consistent behavior during both training and inference is critical for road graph extraction. To address this challenge, we propose a novel "node sampling" strategy.

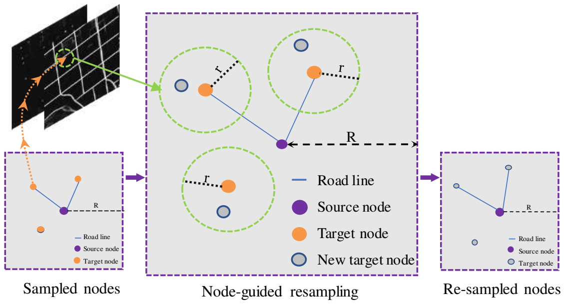

Figure 3. Illustrating the process of node-guided resampling. The sampled nodes are obtained by sampling from the ground truth, and R represents the maximum distance threshold between the source node and target node during the sampling process. Then for each target node, our node-guided resampling strategy will find the maximum probability point of the mask around the target node and save it as the new target node. r represents the maximum distance threshold between the target node and new target node.

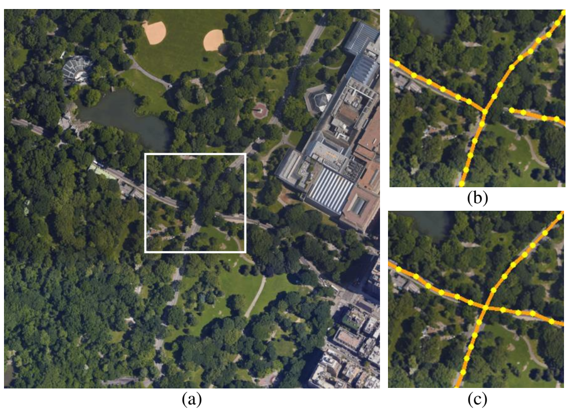

Figure 4. Illustration of occlusion challenge in road nodes connectivity from satellite images. (a) is the raw satellite image, (b) is the prediction of SAM-Road [17], and (c) is the prediction with the "extended-line" strategy.

4.2. Node-guided resampling

Our goal is to align the nodes provided to the classifier during training more closely with those used in inference. However, if we directly adopt the inference process for node selection—specifically using the NMS strategy—the connectivity information necessary to supervise the learning of the classifier becomes unavailable. To address this issue, we propose a compromise strategy called node-guided resampling, as illustrated in Figure 3.

First, we sample N nodes (referred to as source nodes) from the ground truth through a node generator. For each source node, we then identify and save all target nodes within a distance R, along with the connectivity information between these target nodes and the source node. This process yields the sampled nodes (both source and target). Additionally, to ensure diversity among the sampled nodes, we leverage the inherent degree attributes of the graph for source node selection, specifically favoring nodes with rarer degree attributes, which are easier to sample.

For the sampled nodes, to align more closely with the inference process, we leave the source nodes unchanged and use a radius of r to select a new target node with the highest probability, centered around each source node. We then save the coordinate information of this new target node in place of the old one. This results in the creation of re-sampled nodes. The re-sampled nodes not only retain the source nodes of the ground truth and the connectivity between the nodes, but the position information of the new target nodes also matches the node selection strategy in the inference. At the same time, by using the maximum probability of the predicted road mask to pick the new target nodes, it takes full advantage of the experience gained in the previous stage. This strategy compensates for the shortcomings in the training process and effectively improves the consistency of the model during training and inference.

4.3. "Extended-line" strategy

Both in the inference and training process, once the locations of the nodes are determined, the road classifier needs to make road prediction between the source node and target node. However, since the model can only extract information from a single overhead view in satellite images, the determination of connectivity between road nodes is susceptible to occlusion, as shown in Figure 4. To tackle this issue, we propose the "extended-line" strategy.

For a pair of re-sampled nodes required for connectivity discrimination, we use their coordinates in the image to extract their node-centered patch in feature map obtained from SAM-Encoder, called source and target features. Considering the extensibility of the road and the influence of factors such as tree shadows on road judgment, we believe that the information between the two nodes and the information on their extensions can also effectively assist the model in predicting the existence of the road. Therefore, we uniformly sampled the mask values at both ends of the respective extensions n times by using the road masks previously generated by the model. In addition, we also uniformly sample m times on the line between these two nodes.

4.4. Inference

In the inference process, we input the satellite images into the SAM encoder and decoder and obtain the predicted segment masks, including road mask and key point mask. First, we use NMS to deal with the road mask in three steps: 1) we set all pixel values in the mask less than the threshold t_1 to 0, and 2) find the coordinates $(x_i,\, y_i)$ of the largest probability in the mask and save it to the set V_1, then set all pixel values around $(x_i,\, y_i)$ within the threshold radius r_1

Table 2. Comparison of SAM-Road++ with other methods on currently publicly available datasets, The best results are highlighted in bold font, and the second result is underlined. * denotes pre-training on the *Global-Scale* dataset. For all the metrics, larger values indicate better performance.

	City-Scale				SpaceNet			
	F1	Precision	Recall	APLS	F1	Precision	Recall	APLS
Sat2Graph [14]	76.26	80.70	72.28	63.14	80.97	85.93	**76.55**	64.43
RNGDet [49]	76.87	85.97	69.87	65.75	81.13	90.91	73.25	65.61
RNGDet++ [50]	78.44	85.65	72.58	67.76	**82.51**	91.34	<u>75.24</u>	67.73
SAM-Road [17]	77.23	**90.47**	67.69	<u>68.37</u>	80.52	93.03	70.97	71.64
Ours	<u>80.01</u>	88.39	<u>73.39</u>	68.34	81.57	<u>93.68</u>	72.23	<u>73.44</u>
Ours*	**80.66**	<u>89.08</u>	**74.07**	**69.55**	<u>82.07</u>	**93.97**	72.84	**74.35**

to 0. Next, 3) repeat the second step until all the non-zero pixel points have been traversed. For the key point mask, we use the threshold l_2 and the threshold radius r_2 to repeat the above three steps to obtain the set V_2. Finally, we merge the obtained location sets V_1 and V_2 together to obtain the final road nodes information, which can be used for the subsequent connectivity prediction. It is worth mentioning that the use of thresholds l_1, l_2 and radius r_1, r_2 ensures that we get the node with the largest probability within a certain range, which also corresponds well to the node-guided resampling that we adopt in the training phase.

In the connectivity discrimination stage, each node is considered as a source node, and the nodes in its surrounding range as target nodes. For each pair of source and target nodes, we use the "extended-line" strategy to extract information input to the connectivity classifier to predict the probability of whether a road exists between the pair of nodes. Since each node will be a source node and the same roads may be predicted multiple times, we average all the obtained road scores to get the final road graph.

5. Experiments

5.1. Datasets

In this paper, we provide a comprehensive evaluation of three datasets, City-Scale [14], SpaceNet [42] and our *Global-Scale*. Table 1 shows the basic information of these three datasets, which includes important details such as dataset size and spatial resolution. Additionally, to ensure the fairness of experiments, for SpaceNet, we refer to the pre-processing method of previous works [14, 17]. Specifically, we adjust the spatial resolution of satellite images in the SpaceNet dataset to 1m/pixel.

5.2. Metrics

To evaluate the performance of models, we use two metrics: TOPO [5] and Average Path Length Similarity (APLS) [42]. The TOPO compares the similarity of reachable subgraphs of the same vertices in the predicted graph and ground truth in terms of precision, recall and F1. The APLS metric, on the other hand, focuses on the accuracy of the shortest path between two locations in the road graph. This metric is very sensitive to path quality and is well suited to assessing the accuracy of road connectivity.

5.3. Implementation

For both the City-Scale and *Global-Scale* datasets, we extracted 512 $\times$ 512 pixel image patches from the satellite images with the batch size set to 16, and set the number of sampled source nodes N for each patch to 512. For SpaceNet, the patch size is set to 256 × 256, the batch size to 64, and the number of source nodes N per patch is set to 128. In the sampling phase, R is set to 16 pixels, and set r to 8 pixels in the sampling phase for all datasets. In the implementation of "extended-line", since the resolution of all the datasets is 1m/pixel, we set the length on the ends of the respective extended lines to 8 pixels and then set the width of the line to 3 pixels to simulate a road, as well as setting the number of samples n to 15 and m to 20. In the inference phase, we set radius r_1 to 16 and radius r_2 to 8. The binary classification thresholds t_1, t_2, and t_3 are the thresholds that yield the highest F_1 score on the validation set.

We use the Adam optimizer with a basic learning rate of 0.001. For SpaceNet and City-Scale datasets, we train SAM-Road++ until the validation metrics stabilize, following [17]. For *Global-Scale*, SAM-Road++ is trained for 150 epochs. All experiments are conducted on one 4090 GPU.

5.4. Comparative Results

We evaluate the SAM-Road++ model on the City-Scale and SpaceNet datasets and quantitatively compared it with other methods, as shown in Table 2. We compare four typical methods, including two iteration-based methods (RNGDet, RNGDet++) and two global-based methods (Sat2Graph, SAM-Road). The experimental results show that our method outperforms existing SOTA methods in F1 metrics for City-Scale and SpaceNet datasets, demonstrating the advantages in overall topology extraction. However, on the City-Scale dataset, although our recall is higher than other

Figure 5. The visualized road network graph predictions of SAM-Road++ and two baseline methods. Better zoom-in and view in color. Overall, the prediction accuracy of SAM-Road++ is higher. In the crossroads regions (a and b), SAM-Road++ successfully predicted the complexity of multiple roundabouts and overpasses, and in predicting the tree-shaded region c, SAM-Road++'s prediction result is also more complete compared to the other two baselines.

Table 3. Comparison with different methods on *global-scale*. Our method significantly outperforms the other methods on the APLS metrics on both in-domain and out-of-domain test sets, and although we do not achieve sota on the precision metric, it performs far better on the more comprehensive F1 metric on both in-domain and out-of-domain test sets.

	Global-Scale (In-Domain)				*Global-Scale* (Out-of-Domain)			
	F1	Precision	Recall	APLS	F1	Precision	Recall	APLS
SegFormer [47]	42.49	78.23	29.17	24.65	37.27	72.91	25.03	21.36
Sat2Graph [14]	35.53	90.15	22.13	26.77	30.64	**84.73**	19.75	22.49
RNGDet [49]	52.59	79.89	40.72	49.43	42.62	68.79	32.60	36.33
RNGDet++ [50]	55.04	79.02	45.23	52.72	47.34	70.22	35.71	38.08
SAM-Road [17]	59.80	**91.93**	45.64	59.08	46.64	84.54	33.81	40.51
Ours	**62.33**	88.95	**49.27**	**62.19**	**48.34**	82.21	**36.04**	**43.17**

methods, the precision is slightly lower than SAM-Road. This is mainly due to the fact that in the inference, a series of node pairs obtained through NMS do not exactly match the ground truth. The node-guided resampling strategy makes the node pairs used in the training phase closer to the inference. As a result, our model is able to identify roads between the node pairs that may deviate from the ground truth, albeit with reduced score.

For the APLS metrics, the performance on the SpaceNet is better than other methods, which indicates that our proposed "extended-line" strategy helps the model to infer road lengths closer to the ground truth. Meanwhile, the APLS of our method on the City-Scale dataset is comparable to that of SAM-Road. We note that the City-Scale test set contains so few images (only 27 images) that the per-image prediction has a significant impact on the final APLS metric, which results in the APLS metric being unreliable. In addition, to further validate the usefulness of our pro-

Figure 6. Visualisation of the different methods on the out-of-domain test set. In the rural road region with farmland as context ((a) and (b)), our method successfully avoids road disconnection. In the road region (c) with the shadow of a tall building, our proposed "extended-line" strategy accurately predicts the connectivity of the shaded portions.

posed *Global-Scale* dataset, we pre-trained SAM-Road++ on *Global-Scale* and fine-tuned it on the City-Scale and SpaceNet datasets. The results show that the introduction of the *Global-Scale* dataset improves the model's performance in terms of APLS and TOPO metrics on both datasets.

5.5. Benchmarking on *Global-Scale*

Figure 5 and Figure 6 show the qualitative results of SAM-Road++ on the in-domain and out-of-domain test sets of the global-scale dataset, respectively. As can be seen from the circled areas in regions (a) and (b) in Figure 5, when facing complex intersections such as overpasses and multiple roundabouts intersections, the key nodes predicted by the models are not very accurately localized on the road, but owing to the node-guided resampling strategy used in the training phase, SAM-Road++ is able to accurately determine the connectivity of road nodes in this complex scene during the inference phase. The areas circled in Figure 5(c) and Figure 6(c) are obscured by the shadows of buildings and vegetation on the side of the road, which is an inherent challenge to the road graph extraction task. Previous methods cannot discriminate under such conditions, and the "extended line" strategy in SAM-Road++ effectively avoids road breaks based on the extended properties of the road.

For further comparison, we trained the six models mentioned in the previous section on the *Global-Scale* dataset, which contains more comprehensive data, and validated them on both the in-domain test set and out-of-domain test set and the results are shown in Table 3. On the in-domain test set, our method significantly outperforms previous methods in both APLS and TOPO metrics, which indicates that our method outperforms other methods in a wide range of scenarios. Although the overall performance on the out-of-domain test set is lower than that on the in-domain test set, our method still significantly outperforms the other methods, suggesting that our method is more robust to unseen regional data than the other methods, and is more suitable for real-world applications. Besides, the methods [14, 47] perform poorly on the dataset, probably because their single post-processing way leads to a poor generalization on *Global-Scale* with different road types.

Finally, by comparing the results in Table 2 and Table 3, it can be seen that all five models do not perform as well on *Global-Scale* as they did on City-Scale and SpaceNet, which further demonstrates that *Global-Scale* is a more challenging dataset to better evaluate the generalization ability of models in complex scenarios.

Table 4. Ablation results for "extended line" strategy and node-guided resampling on the SpaceNet dataset.

"extended-line"	node-guided resampling	APLS	F1
✗	✗	71.64	80.52
✗	✓	71.90	81.77
✓	✗	73.22	80.89
✓	✓	73.44	81.57

5.6. Ablation Studies

In this section, we conduct ablation studies to verify the rationality of the design of SAM-Road++, including the node-guided resampling and "extended-line" strategy, as shown in Table 4. First, we completely remove the node-guided resampling strategy from the training and observe a significant decrease in the model's performance on the F1 metric. This suggests that the node-guided resampling indeed simulates the inference process in training, enabling the model to predict better topologies. Next, we eliminate the "extended-line" strategy and find that the APLS decreases, suggesting that our strategy effectively helps the model predict road lengths that are closer to the ground truth.

6. Conclusion

In this paper, we present a large-scale dataset, *Global-Scale*, and a novel method, SAM-Road++, for road graph extraction. The *Global-Scale* encompasses six continents and has been meticulously curated to include a diverse array of scenes, such as urban and rural areas. SAM-Road++ effectively addresses the mismatch between training and inference in global-based methods while mitigating the occlusion challenges inherent in road graph extraction tasks. Extensive experiments demonstrate that *Global-Scale* serves as a more comprehensive and challenging benchmark. In addition, SAM-Road++ achieves superior performance on both existing public datasets and *Global-Scale*, without incurring significant inference costs. Looking ahead, we plan to further expand the *Global-Scale* dataset and endeavor to innovate the paradigm of the road extraction task.

Acknowledgment. This work was supported by the National Natural Science Foundation of China under Grant 62192781, Grant 62272375, Grant 62272374, Grant 62250009, Grant 62137002, Grant 62276208, and Grant 62425113, in part by Natural Science Foundation of Shaanxi Province under Grant 2024JC-JCQN62, in part by Natural Science Basic Research Program of Shaanxi Province under Grant 2024JC-JCQN-02, in part by Project of China Knowledge Center for Engineering Science and Technology, in part by Project of Chinese academy of engineering "The Online and Offline Mixed Educational Service System for "The Belt and Road" Training in MOOC China," in part by the Major Key Project of PCL under Grant PCL2024A06, and in part by Tianyuan Fund for Mathematics of the National Natural Science Foundation of China under Grant 12426105.

References

[1] Abolfazl Abdollahi, Biswajeet Pradhan, and Abdullah Alamri. Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *Ieee Access*, 8:179424–179436, 2020. 3

[2] Gaetan Bahl, Mehdi Bahri, and Florent Lafarge. Road extraction from overhead images with graph neural networks. *arXiv preprint arXiv:2112.05215*, 2021. 1

[3] Favyen Bastani, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Sam Madden, and David DeWitt. Roadtracer: Automatic extraction of road networks from aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4720–4728, 2018. 1, 3

[4] Anil Batra, Suriya Singh, Guan Pang, Saikat Basu, CV Jawahar, and Manohar Paluri. Improved road connectivity by joint learning of orientation and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10385–10393, 2019. 2

[5] James Biagioni and Jakob Eriksson. Inferring road maps from global positioning system traces: Survey and comparative evaluation. *Transportation research record*, 2291(1): 61–71, 2012. 6

[6] Hao Chen, Zhenghong Li, Jiangjiang Wu, Wei Xiong, and Chun Du. Semiroadexnet: A semi-supervised network for road extraction from remote sensing imagery via adversarial learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 198:169–183, 2023. 3

[7] Ziyi Chen, Liai Deng, Yuhua Luo, Dilong Li, José Marcato Junior, Wesley Nunes Gonçalves, Abdul Awal Md Nurunnabi, Jonathan Li, Cheng Wang, and Deren Li. Road extraction in remote sensing data: A survey. *International journal of applied earth observation and geoinformation*, 112: 102833, 2022. 1

[8] Guangliang Cheng, Ying Wang, Shibiao Xu, Hongzhen Wang, Shiming Xiang, and Chunhong Pan. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6):3322–3337, 2017. 3

[9] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 172–181, 2018. 2, 3

[10] Vikas Dhiman, Quoc-Huy Tran, Jason J Corso, and Manmohan Chandraker. A continuous occlusion model for road scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4331–4339, 2016. 2

[11] Xun Gao, Xian Sun, Yi Zhang, Menglong Yan, Guangluan Xu, Hao Sun, Jiao Jiao, and Kun Fu. An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network. *IEEE Access*, 6:39401–39414, 2018. 3

[12] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016. 3

[13] Mordechai Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18, 2008. 3

[14] Songtao He, Favyen Bastani, Satvat Jagwani, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Mohamed M Elshrif, Samuel Madden, and Mohammad Amin Sadeghi. Sat2graph: Road graph extraction through graph-tensor encoding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, pages 51–67. Springer, 2020. 1, 2, 3, 6, 7, 8

[15] Yang He, Ravi Garg, and Amber Roy Chowdhury. Td-road: top-down road network extraction with holistic graph construction. In *European Conference on Computer Vision*, pages 562–577. Springer, 2022. 1, 3

[16] Benjamin Herfort, Sven Lautenbach, João Porto de Albuquerque, Jennings Anderson, and Alexander Zipf. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in openstreetmap. *Nature Communications*, 14(1):3985, 2023. 3

[17] Congrui Hetang, Haoru Xue, Cindy Le, Tianwei Yue, Wenping Wang, and Yihui He. Segment anything model for road network graph extraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2556–2566, 2024. 1, 2, 3, 4, 5, 6, 7

[18] Danfeng Hong, Bing Zhang, Xuyang Li, Yuxuan Li, Chenyu Li, Jing Yao, Naoto Yokoya, Hao Li, Pedram Ghamisi, Xiuping Jia, et al. Spectralgpt: Spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 4

[19] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020. 3

[20] Teja Kattenborn, Jens Leitloff, Felix Schiefer, and Stefan Hinz. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 173:24–49, 2021. 3

[21] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 4

[22] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 3

[23] Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, et al. Towards fully autonomous driving: Systems and algorithms. In *2011 IEEE intelligent vehicles symposium (IV)*, pages 163–168. IEEE, 2011. 1

[24] Qi Li, Yue Wang, Yilun Wang, and Hang Zhao. Hdmapnet: An online hd map construction and evaluation framework. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 4628–4634. IEEE, 2022. 3

[25] Yue Li, Rong Zhang, and Yunfei Wu. Road network extraction in high-resolution sar images based cnn features. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1664–1667. IEEE, 2017. 3

[26] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12):6999–7019, 2021. 3

[27] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 3

[28] Yeneng Lin, Dongyun Xu, Nan Wang, Zhou Shi, and Qiuxiao Chen. Road extraction from very-high-resolution remote sensing images via a nested se-deeplab model. *Remote sensing*, 12(18):2985, 2020. 3

[29] Richard J Lisle. Google earth: a new geological resource. *Geology today*, 22(1):29–32, 2006. 3

[30] Ruyi Liu, Junhong Wu, Wenyi Lu, Qiguang Miao, Huan Zhang, Xiangzeng Liu, Zixiang Lu, and Long Li. A review of deep learning-based methods for road extraction from high-resolution remote sensing images. *Remote Sensing*, 16 (12):2056, 2024. 3

[31] Fanny Malin, Ilkka Norros, and Satu Innamaa. Accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention*, 122:181–188, 2019. 2

[32] Gellert Mattyus, Shenlong Wang, Sanja Fidler, and Raquel Urtasun. Enhancing road maps by parsing aerial images around the world. In *Proceedings of the IEEE international conference on computer vision*, pages 1689–1697, 2015. 2

[33] Volodymyr Mnih. *Machine learning for aerial image labeling*. University of Toronto (Canada), 2013. 2, 3

[34] Sherif AS Mohamed, Mohammad-Hashem Haghbayan, Tomi Westerlund, Jukka Heikkonen, Hannu Tenhunen, and Juha Plosila. A survey on odometry for autonomous navigation systems. *IEEE access*, 7:97466–97486, 2019. 1

[35] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013. 3

[36] Onisimo Mutanga and Lalit Kumar. Google earth engine applications, 2019. 3

[37] Ilja Nastjuk, Bernd Herrenkind, Mauricio Marrone, Alfred Benedikt Brendel, and Lutz M Kolbe. What drives the acceptance of autonomous driving? an investigation of acceptance factors from an end-user's perspective. *Technological Forecasting and Social Change*, 161:120319, 2020. 1

[38] Lucas Prado Osco, Qiusheng Wu, Eduardo Lopes de Lemos, Wesley Nunes Gonçalves, Ana Paula Marques Ramos, Jonathan Li, and José Marcato Junior. The segment anything model (sam) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation*, 124:103540, 2023. 4

[39] Zehang Sun, George Bebis, and Ronald Miller. On-road vehicle detection: A review. *IEEE transactions on pattern analysis and machine intelligence*, 28(5):694–711, 2006. 3

[40] Gabriel Svennerberg. *Beginning google maps API 3*. Apress, 2010. 3

[41] Michail Tarasiou, Erik Chavez, and Stefanos Zafeiriou. Vits for sits: Vision transformers for satellite image time series. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10418–10428, 2023. 3

[42] Adam Van Etten, Dave Lindenbaum, and Todd M Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018. 2, 3, 6

[43] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 3

[44] Peijuan Wang, Bulent Bayram, and Elif Sertel. A comprehensive review on deep learning based remote sensing image super-resolution methods. *Earth-Science Reviews*, 232: 104110, 2022. 1

[45] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 3

[46] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022. 3

[47] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34: 12077–12090, 2021. 3, 7, 8

[48] Jiakun Xu, Bowen Xu, Gui-Song Xia, Liang Dong, and Nan Xue. Patched line segment learning for vector road mapping. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6288–6296, 2024. 1

[49] Zhenhua Xu, Yuxuan Liu, Lu Gan, Yuxiang Sun, Xinyu Wu, Ming Liu, and Lujia Wang. Rngdet: Road network graph detection by transformer in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2022. 1, 3, 6, 7

[50] Zhenhua Xu, Yuxuan Liu, Yuxiang Sun, Ming Liu, and Lujia Wang. Rngdet++: Road network graph detection by transformer with instance segmentation and multi-scale features

enhancement. *IEEE Robotics and Automation Letters*, 8(5): 2991–2998, 2023. 1, 3, 6, 7

[51] Ruoyu Yang, Yanfei Zhong, Yinhe Liu, Xiaoyan Lu, and Liangpei Zhang. Occlusion-aware road extraction network for high-resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 2

[52] Li Yuan, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Zi-Hang Jiang, Francis EH Tay, Jiashi Feng, and Shuicheng Yan. Tokens-to-token vit: Training vision transformers from scratch on imagenet. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 558–567, 2021. 3

[53] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443–58469, 2020. 1

[54] Tongjie Y Zhang and Ching Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984. 3

[55] Qiqi Zhu, Yanan Zhang, Lizeng Wang, Yanfei Zhong, Qingfeng Guan, Xiaoyan Lu, Liangpei Zhang, and Deren Li. A global context-aware and batch-independent network for road extraction from vhr satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:353–365, 2021. 2, 3

[56] Aleš Žnidarič, Vikram Pakrashi, Eugene O'Brien, and Alan O'Connor. A review of road structure data in six european countries. *Proceedings of the Institution of Civil Engineers-Urban design and planning*, 164(4):225–232, 2011. 3