Road Segmentation from Satellite Imagery Based on an Improved SAM Model

Bingquan Yao[1] Haolin Liu[1,*] Wenjuan Mao[1] Jin Zhou[1]

[1]National Center for Quality Inspection and Testing of Surveying and Mapping Products,Beijing,China - (634468022@qq.com)

Keywords: Satellite Imagery, Road Extraction, Semantic Segmentation, Segment Anything Model, LoRA.

ABSTRACT:

Road network is an important infrastructure of urban spatial structure and traffic system. Its accurate acquisition is of great significance for urban traffic analysis, automatic driving map construction and disaster emergency response. With the wide acquisition of high-resolution remote sensing images, automatic extraction of road masks from remote sensing images has become an important research direction in the field of remote sensing image understanding. However, the existing deep learning methods still face the problems of obvious modal differences and insufficient modeling of road structure continuity in remote sensing scenes. To solve the above problems, this paper proposes a remote sensing image road segmentation model LR-SAM based on SAM (Segment Anything Model). In this model, the LoRA (Low Rank Adaptation) fine-tuning strategy is introduced to achieve efficient parameter updating, and the MS multi-scale feature interaction module is designed in the coding phase to enhance the expression ability of the linear structure and fine-grained information of the road. At the same time, the original prompt encoder is removed and a lightweight AD decoder is constructed to achieve multi-scale feature fusion. In the reasoning stage, TTA (Test Time Augmentation) strategy is introduced to improve the stability and segmentation accuracy of the model. Experimental results based on CHN6-CUG and SAT-MTB datasets show that the proposed method achieves 97.20% and 85.06% mIoU and 96.67% and 84.94% F1-score, respectively, which is significantly better than the mainstream road segmentation methods, and verifies the effectiveness of the proposed improvement points.

1. Introduction

1.1 Background and Related Work

Modern urban traffic network is the core of urban engineering, and the accurate description of its spatial topology is directly related to key scenarios such as traffic flow optimization and public safety emergency response (Mo et al., 2024). The road mask obtained by remote sensing image segmentation can not only support the real-time update of high-precision navigation map, but also provide spatial decision-making basis for the optimization of rescue path in flood disaster and the road condition perception of automatic driving system (Zhao et al., 2024). Road network extraction based on deep learning has become a research hotspot, and the end-to-end image segmentation technology has become the mainstream research direction because of its efficiency, which also provides an extremely reliable and efficient technical route for the development of remote sensing image road extraction research (Yuan et al., 2021).

In recent years, Transformer architecture has achieved great success in the field of natural language processing (Vaswani et al., 2017). Because of its powerful feature extraction ability and long-distance dependency modeling ability, it has been gradually introduced into the field of computer vision. The road extraction model based on transformer shows significant advantages in processing high-resolution satellite images and complex scenes. Yang took the lead in introducing transformer into the task of road segmentation in remote sensing images and proposed the hybrid architecture of TransRoadNet (Yang et al., 2022). In the encoder phase, CNN is used to extract local texture features, and the transformer module is used to establish the global association between pixels, which effectively solves the problem of discontinuous segmentation caused by objects similar to roads and occlusion in traditional methods. The RoadFormer proposed by Liu et al. is combined with spatial channel separable convolution for road extraction (Liu et al., 2023). The model first uses the Swin Transformer (Liu et al., 2021) multi-scale encoder to extract long-distance information, introduces the separable convolution of space and channel to obtain fine-grained and global features, and connects the expansion blocks after the spatial convolution to capture more road structures. Although the above improved architecture has made significant progress in the task of road extraction, most of the current research focuses on small-scale data sets or annotation data in specific fields, and there are problems such as insufficient modeling ability of road boundary continuity and high training cost, which restrict its application.

In 2023, the large image segmentation model SAM (Segment Anything Model) was born (Kirillov et al., 2023). As a general image segmentation model based on vision transformer, it achieves cross task generalization ability through pre-training on SA-1B large-scale data set, and can achieve better segmentation accuracy in the absence of labeled data. Later, many scholars applied it to the task of road extraction from remote sensing images. For example, Hetang proposed an improved version of SAM-road based on SAM model (Hetang et al., 2024). The model designed a graph neural network based on lightweight transformer for the task of road network topology prediction, which can use the embedded features of SAM images to estimate the probability of the existence of edges between nodes. Feng proposed a general model road SAM for road extraction based on SAM (Feng et al., 2024). In the training phase of the model, the explicit visual cue task specific input module is adopted, which can use the embedded features and high-frequency component information as the cue input model. At the same time, the frequency adapter fine-tuning mechanism is also designed. Through lightweight but efficient fine-tuning technology, the remote sensing knowledge

*Corresponding author

in specific fields is integrated into the road extraction model, which can significantly improve the segmentation performance and realize the efficient utilization of computing resources.

1.2 Problem Statement and Motivations

However, due to the following problems, the existing SAM or its improved model still performs poorly in the task of remote sensing image road segmentation: (1) The problem of modal offset is significant, and it is difficult to generalize to satellite remote sensing images: the pre-training data of SAM model mainly comes from natural images. There are modal differences between its semantic distribution and remote sensing images, which makes its feature extraction and matching ability unstable in the task of road extraction from remote sensing images, and the phenomenon of misjudgment and omission occurs frequently (Houlsby et al., 2019). (2) Lack of modeling ability for road boundary continuity: SAM mainly focuses on class-level segmentation rather than structural continuity. The road object in remote sensing image has the characteristics of strong connectivity and large width change, which puts forward the requirements of multi-scale perception and spatial context modeling for segmentation model. The original SAM structure was not structurally improved for road continuity and scale sensitivity, resulting in unstable segmentation performance of the edge connection region (Schick et al., 2020). (3) Complex calculation and redundant model: the original SAM structure design focused on semantic generalization ability, and did not do directional compression and acceleration processing for downstream tasks. It has a large number of parameters and takes a long time to reason, so it is not suitable for deployment in remote sensing image segmentation tasks requiring high time-consuming processing (Dai et al., 2023). (4) Simple full-scale continuous training is difficult to decouple task-related parameters and general feature expression: the traditional fine-tuning method cannot distinguish the boundary between "task specific layer" and "general visual layer", and it is difficult to achieve efficient modular adaptation. At the same time, it is easy to cover the basic feature capabilities obtained in the pre-training of the model, such as edge localization, shape perception and so on, leading to the decline of generalization (Parulekar et al., 2024).

2. Method

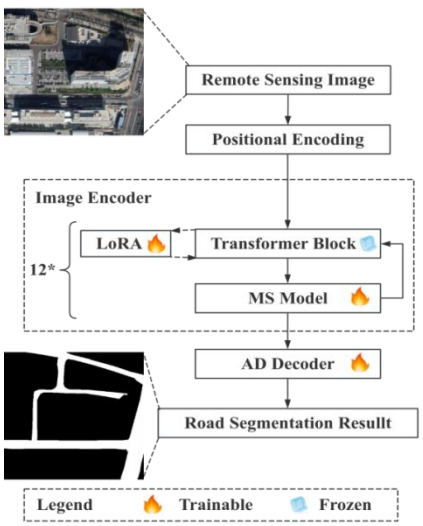

Figure 1. Technical route of this paper.

Aiming at the existing problems, based on the SAM framework, this paper focuses on structural innovation and training strategy improvement, and proposes the LR-SAM model. Firstly, LoRA structure and multi-scale fusion module are added based on the original module in the coding process to enhance the adaptability of the model to downstream tasks. Secondly, the original SAM prompt encoding embedded module is removed and a lightweight decoder is designed to replace the original decoder to fuse the multi-scale feature map for pre-test output. In the process of training, TTA strategy is introduced to enhance the data of training samples to achieve high-precision remote sensing image road segmentation (Perez and Wang, 2017). The overall technical process of the algorithm in this paper is shown in Figure 1 (Mikołajczyk and Grochowski, 2018).

2.1 LoRA-based Fine-tuning

LoRA, as an efficient fine-tuning technology, is suitable for improving training efficiency and reducing computing costs when data resources are limited (Lu and Weng, 2024). During fine-tuning, LoRA performs low rank decomposition on the weight matrix W of the
original model, which can be expressed as the equation 1:

$$W = W_0 + \Delta W = W_0 + A \cdot B \,, \tag{1}$$

where W_0 = original weight in the pre-training model
$\Delta W = A \cdot B$ = low rank adaptation matrix

Unlike the traditional fine-tuning method, which directly updates the parameters of W, LoRA only updates A and B. The rank of these two matrices is usually much lower than that of the original matrix. Therefore, LoRA reduces the computational cost of fine-tuning by reducing the number of parameters that need to be updated (Sun et al., 2023).

For the task of remote sensing image road segmentation, this paper adds a low rank matrix with rank of 8 to each transformer block in the encoder part of the original SAM model. In the training process, only these low rank parameters are trained and the original model parameters are frozen, so that the model can not only learn the knowledge of remote sensing image road segmentation at low computational cost, but also avoid the problems of excessive update and catastrophic forgetting that may occur in traditional fine-tuning.

2.2 MS Feature Interaction Module

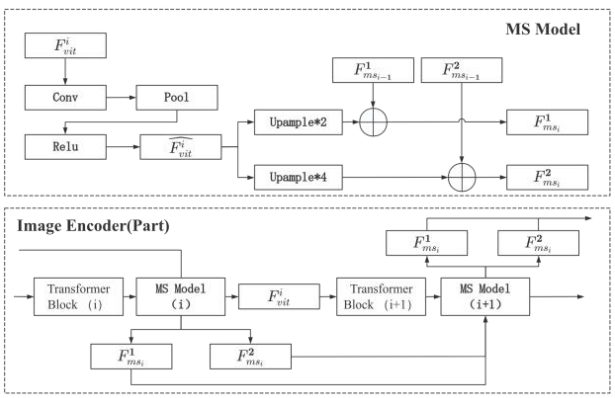

Figure 2. MS Feature Interaction Module.

In the task of road segmentation for satellite video images, because the segmentation object is the road, it has obvious linear features and is sensitive to edge details. However, the original SAM image is 16 times subsampled in the encoding process, which will lead to the loss of the position information and edge information of the segmented object in the feature map. In this paper, the MS multi-scale feature interaction module is introduced to retain the linear features and edge details in different scale feature maps (Yu et al., 2024). Its structure is shown in Figure 2.

Specifically, the original SAM image encoder contains 12 transformer blocks. After adding the MS multi-scale feature interaction module, the layer I MS module receives the output feature fivit from the layer I of the image encoder and the layered feature FKMS from the (i-1) layer. Two features fkmrm are output, and their resolutions are 8 times and 4 times of the original image size respectively. The above process can be formalized as equation 2:

$$\hat{F}^{k}_{ms_i} = \mathrm{deconv}_k\left(\mathrm{conv}_{1*1}\left(F^{j}_{vit}\right)\right), k = 1, 2, \quad (2)$$

Among them, deconv represents the deconvolution module, which performs 2 and 4 times of up sampling on fjvit to obtain high-resolution features of two different scales. In addition, the obtained high-resolution features need to be fused with the hierarchical features output from the previous layer of MS, as shown in the equation 3:

$$F^{k}_{ms_i} = \tilde{F}^{k}_{ms_i} + F^{k}_{ms_{i-1}}, k = 1, 2, \quad (3)$$

2.3 AD Lightweight Decoder

Due to the introduction of MS multi-scale feature interaction module, the image can be represented by three scales of 16 times subsampling, 8 times subsampling and 4 times subsampling. In order to make full use of the feature information contained in each scale feature map provided by the encoder and ensure the lightweight of the model, first remove the prompt encoder and mask decoder of the original SAM model, and then design a lightweight decoder AD to realize the output of road segmentation results of remote sensing images using multi-scale image coding feature map. Its architecture is shown in Figure 3.

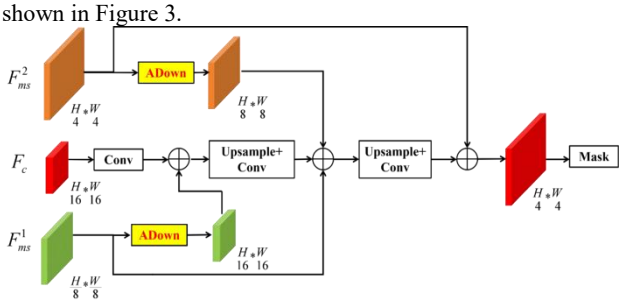

Figure 3. AD Lightweight Decoder.

The resolution of MS multi-scale feature interaction module is 8 times and 4 times of down sampling feature map, respectively. For high-resolution features with different scales, a two-stage injection method is used. Firstly, the key features obtained by adown operation are injected into the 16 times subsampled feature map, and then two complete high-resolution hierarchical features are injected into the upsampled feature map. Through this two-stage injection, the enhanced feature representation is obtained, which enables the decoding process to achieve more refined segmentation results.

Adown combines the smoothness of average pooling with the saliency detection ability of maximum pooling to reduce the amount of calculation while retaining key features (Gao et al., 2023). It enhances the expression ability through multi-scale feature fusion, reduces the loss of target details in each resolution image, and makes the model perform better in complex background (Wang et al., 2023). Its overall structure is shown in Figure 4.

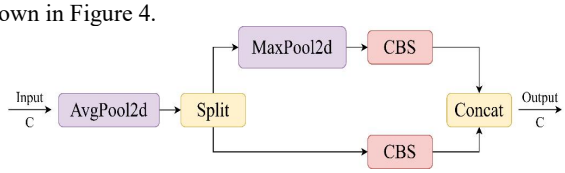

Figure 4. Adown.

Firstly, the module implements avgpool2d downsampling operation on the C-channel input characteristic map to obtain the low resolution characteristic substrate. Then, the feature is decomposed into two path branches through the channel split strategy. The main path maintains the current resolution for feature preservation, and the auxiliary path uses the maximum pool to strengthen the local discriminant feature. The nonlinear transformation of the two features is realized by the CBS (conv batch norm Silu) module, and its mathematical expression is shown in the equation 4. Finally, the original channel dimensions are restored to form multi-scale fusion output through channel dimension splicing and reorganization, which effectively enhances the module's ability to represent the characteristics of occluded targets and low resolution images.

$$CBS_{out} = SiLU(BatchNorm(Conv2d(X))), \quad (4)$$

where X = original weight in the pre-training model

3. Experiment and Result Analysis

3.1 Dataset

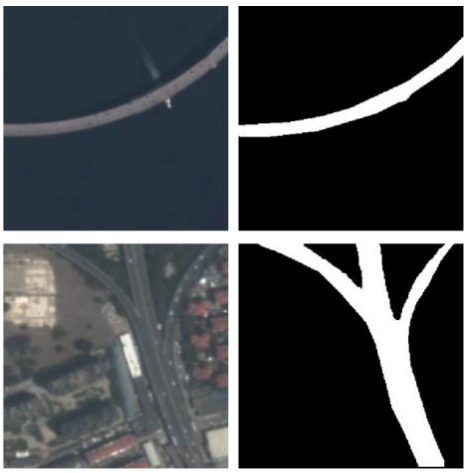

Figure 5. Road Segmentation Label Based on SAT-MTB Dataset.

This paper selects the open remote sensing image road data set CHN6-CUG released by China University of Geosciences and the satellite video image SAT-MTB data set released by the space application engineering and technology center of Chinese Academy of Sciences. The CHN6-CUG dataset covers high-

resolution remote sensing images of six cities in China (Zhu et al., 2021). The base map is from Google Earth with a resolution of 50 cm/pixel. A total of 4511 512 × 512 images are included and divided into training set and test set according to 3608:903. SAT-MTB data set is collected by Jilin-1 03 satellite in many countries and regions around the world (Li et al., 2023). The original data contains 249 full resolution satellite video images with a video frame rate of 10 fps, and the road labels are made by the author of this paper, as shown in Figure 5. The introduction of satellite video image data sets can further verify the generalization ability of this paper in dealing with the road extraction task of different modal remote sensing images.

3.2 Experimental Setup

The experimental environment of this paper is Windows 10, CPU (Intel (R) core (TM) i7-11700 @ 2.50 GHz), and memory 16 GB. The algorithm implementation is based on PyCharm compilation tool, using Python 3.9 language, PyTorch 2.1.2 with CUDA 12.3 architecture. For the algorithm level, the LR-SAM model training parameters proposed in this paper are set as follows: the number of training epochs is 40, the initial learning rate is 0.0001, and the optimizer uses SGD.

Secondly, the TTA (test time augmentation) strategy is introduced in the process of model reasoning. Its core idea is to introduce data enhancement in the model reasoning stage. By generating multiple enhanced versions of the same input image, such as horizontal flip, vertical flip, multi-directional rotation, etc., the input model is predicted separately, and then all the prediction results are inversely converted back to the original space and fused, so as to obtain a more stable and accurate segmentation output. This method does not need additional training, and only adds a small amount of calculation in the reasoning stage, which can effectively reduce the uncertainty of a single prediction and improve the adaptability of the model to perspective changes, noise and complex scenes.

3.3 Evaluation Index

In order to comprehensively measure the effect of the proposed model on the task of road segmentation in satellite images, mIoU and F1 score parameters are used as accuracy evaluation indexes in this paper. IoU is obtained by calculating the ratio of the intersection and union of the predicted results and the actual results. mIoU is the average IoU of all kinds of ground objects, and F1 score is the harmonic average of the accuracy rate and recall rate.

3.4 Comparative Experiment

DATASAT	MODEL	mIoU	F1-Score
CHN6-CUG	UNet++	91.15	85.29
	PSPNet	89.05	81.59
	DeepLabV3	87.51	80.64
	LR-SAM	**97.20**	**96.67**
SAT-MTB	UNet++	72.46	79.08
	PSPNet	75.72	80.36
	DeepLabV3	68.95	73.39
	LR-SAM	**85.06**	**84.94**

Table 1. Quantitative results of comparative experiments with different models

In order to verify the advancement of the LR-SAM algorithm proposed in this paper in the task of road segmentation from satellite images, the CHN6-CUG dataset and SAT-MTB dataset were compared with a variety of mainstream road segmentation methods (UNet++ (Zhou et al., 2020), PSPNet (Zhao et al., 2017), DeepLabV3 (Yang et al., 2022)), and the comparison results are shown in Table 1. It can be seen that the mIoU and F1 scores can reach 97.20%, 96.67% and 85.06%, 84.94% respectively when using the model proposed in this paper for road segmentation on the open satellite image road data set, and the accuracy is better than other models.

The visual comparison results of the experiment are shown in Figure 6. It can be seen from the figure that the detection results of UNet++ are missing to a certain extent, which is manifested in the road connection or the place related to the background, especially in the part of the sea highway in Scene 2, which fails to distinguish the road from the sea; PSPNet is relatively stable in dealing with urban roads and marine roads, but when the road surface color changes, it can not solve the intra class differences, resulting in missed detection; In the detection results of DeepLabV3, a large number of urban buildings were identified as roads, and there were also problems such as the inability to extract marine roads, which performed worst in the listed models; Based on SAM visual model, this paper uses LoRA continuation training strategy and improved model structure to distinguish most of the inter class differences and fit most of the intra class differences in image segmentation tasks, and obtains high-precision road segmentation results.

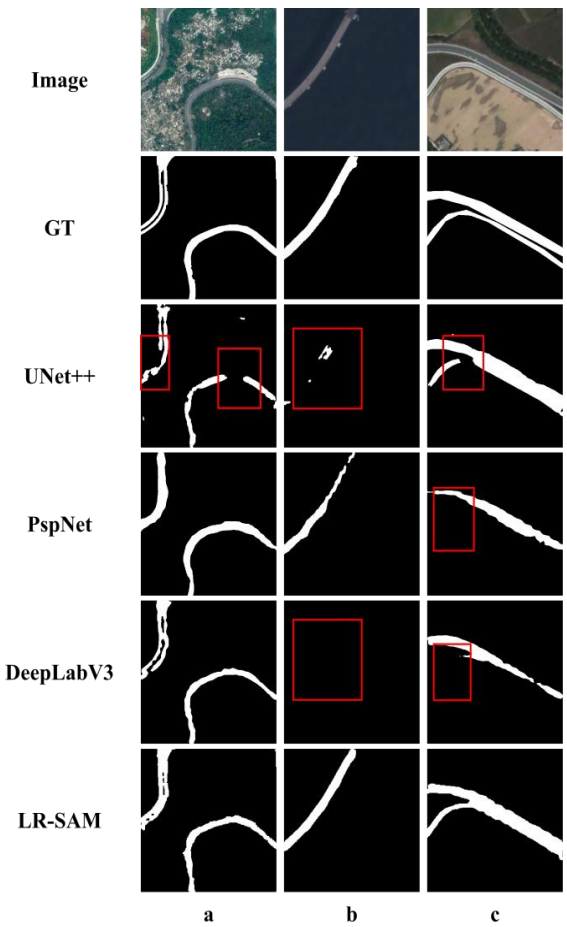

Figure 6. Visualization results of comparative experiments of different models.

3.5 Ablation Experiment

In order to verify the effectiveness of the various improved schemes proposed in this paper in terms of accuracy indicators, the LR-SAM proposed in this paper was ablated with the original SAM and the SAM continuation method with LoRA strategy. The results are shown in Table 2. It can be seen that the strategy of direct continuous training of SAM is difficult to learn effective improvement under the restriction of fewer cycles and smaller data set size; After adding the LoRA efficient continuation training strategy, it avoids the common problems of large model continuation training such as catastrophic forgetting on the premise of ensuring the computational efficiency through low rank decomposition, and can improve its adaptability to specific downstream tasks; Further, after the MS feature interaction module and AD lightweight decoder are added to the model (due to the limitation of the number of branches on the model input, the two model structures cannot be ablated and compared separately), the model can focus on the key features on multiple scales, which further improves the performance of the model compared with the continuous training image encoder strategy.

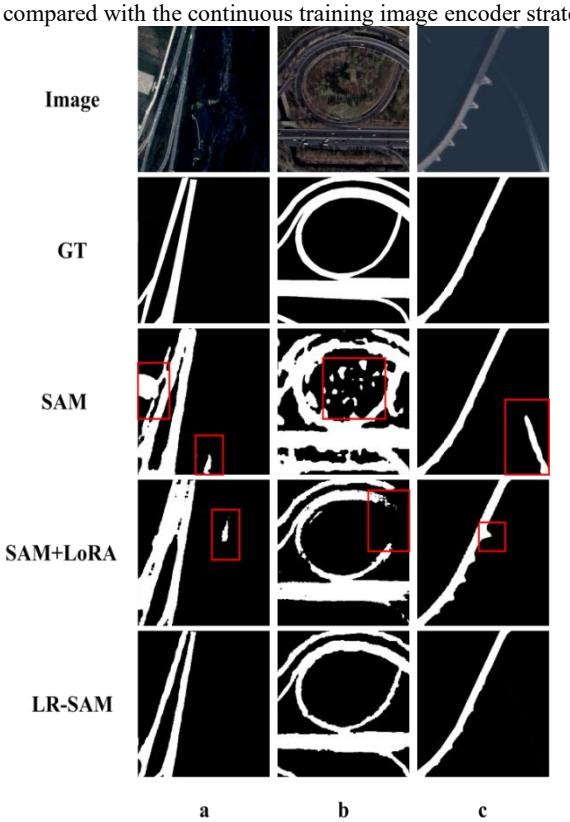

Figure 7. Visualization results of ablation experiments with different models.

The visual comparison results of the experiment are shown in Figure 7. The three example scenarios are urban roads, elevated hubs and offshore bridges. As can be seen from the figure, the direct continuation training of the original SAM can not get a high performance in the small-scale downstream task data set due to the fact that the model itself is designed as a general model, and there are many false detections in terms of the detection results, such as the false identification of the tracks of offshore ships as roads, etc; After freezing the original image encoder and adding LoRA, the adaptability of the model to downstream tasks is improved, which is mainly reflected in the obvious decline of false detection, but there are still cases where the piers of offshore bridges are mistakenly identified as roads. At the same time, due to the continued training of the decoder from the original parameters, a certain degree of missing detection has occurred, such as the road fracture in Scene 3; The LR-SAM model proposed in this paper adds multi-scale feature fusion strategy and lightweight decoder, which is more suitable for the task of road segmentation in remote sensing images, and can achieve high-precision road extraction.

DATASAT	LoRA	MS+AD	mIoU	F1-Score
CHN6-CUG			69.17	71.01
	√		84.05	83.79
	√	√	**97.20**	**96.67**
SAT-MTB			58.19	71.72
	√		73.73	83.86
	√	√	**85.06**	**84.94**

Table 2. Quantitative results of ablation experiments with different models

4. Conclusions

Aiming at the problems of obvious modal differences, insufficient modeling of road structure continuity and high computational complexity in the task of road segmentation from remote sensing images, this paper proposes an improved model LR-SAM for road extraction from remote sensing images based on the SAM framework. By introducing LoRA efficient fine-tuning strategy, this method can achieve high-efficient adaptation to the task of remote sensing road extraction under the condition of freezing most of the parameters of the original model; At the same time, MS multi-scale feature interaction module is added to alleviate the loss of spatial details caused by the original coding structure in the process of sampling at high magnification; In addition, by removing the original prompt encoder and designing a lightweight AD decoder, the effective fusion of multi-scale features and the output of results are achieved.

The experimental results on CHN6-CUG and SAT-MTB remote sensing image data sets show that the LR-SAM model proposed in this paper achieves 97.20% mIoU and 96.67% F1 score in the road extraction task, and the overall performance is significantly better than the mainstream image segmentation models such as UNet++, PSPNet and DeepLabV3. At the same time, the ablation experiment further verifies the effectiveness of LoRA fine-tuning strategy, MS multi-scale feature interaction module and AD lightweight decoder in improving the performance of the model. Experimental results show that the proposed method can effectively improve the accuracy of road segmentation based on remote sensing images.

In general, LR-SAM model shows good accuracy in the task of road segmentation in remote sensing images, and provides a feasible and efficient technical scheme for automatic extraction of road network in remote sensing images. Future work can further combine road topology constraints and graph structure modeling methods to improve the expression ability of road connectivity and explore the generalization ability of the model in larger scale remote sensing data and complex urban environment.

References

Dai, H., Ma, C., Yan, Z., Diao, W., Chen, Y., Sun, J., 2023. SAMAug: point prompt augmentation for Segment Anything Model. *arXiv preprint*, arXiv:2307.01187. doi.org/10.48550/arXiv.2307.01187

Feng, W., Guan, F., Sun, C., Xu, W., 2024. Road-SAM: Adapting the Segment Anything Model to road extraction from large very-high-resolution optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 21, 1-5. doi.org/10.1109/LGRS.2024.3430900

Gao, X., Sun, Y., Xiao, Y., Gu, Y., Chai, S., Chen, B., 2022. Adaptive random down-sampling data augmentation and area attention pooling for low resolution face recognition. *Expert Systems with Applications*, 209, 118275. doi.org/10.1016/j.eswa.2022.118275

Hetang, C., Xue, H., Le, C., Yue, T., Wang, W., He, Y., 2024. Segment anything model for road network graph extraction. *In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2556-2566. doi.org/10.1109/CVPR52733.2024.00246

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S., 2019. Parameter-efficient transfer learning for NLP. *In: Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, Long Beach, CA, USA, 2790-2799. doi.org/10.48550/arXiv.1902.00751

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything. *In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, 4015-4026. doi.org/10.1109/ICCV51070.2023.00371

Li, S., Zhou, Z., Zhao, M., Yang, J., Guo, W., Lv, Y., Kou, L., Wang, H., Gu, Y., 2023. A multi-task benchmark dataset for satellite video: Object detection, tracking, and segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-21. doi.org/10.1109/TGRS.2023.3278075

Liu, X., Wang, Z., Wan, J., Zhang, J., Xi, Y., Liu, R., Miao, Q., 2023. RoadFormer: Road extraction using a Swin transformer combined with a spatial and channel separable convolution. *Remote Sensing*, 15(4), 1049. doi.org/10.3390/rs15041049

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. *In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 10012-10022. doi.org/10.1109/ICCV48922.2021.00986

Lu, X., Weng, Q., 2024. Multi-LoRA fine-tuned segment anything model for urban man-made object extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-19. doi.org/10.1109/TGRS.2024.3435745

Mikołajczyk, A., Grochowski, M., 2018. Data augmentation for improving deep learning in image classification problem. *In: 2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Świnoujście, Poland, 117-122.

Mo, S., Shi, Y., Yuan, Q., Li, M., 2024. A survey of deep learning road extraction algorithms using high-resolution remote sensing images. *Sensors*, 24(5), 1708. doi.org/10.3390/s24051708

Parulekar, B., Singh, N., Ramiya, A. M., 2024. Evaluation of segment anything model (SAM) for automated labelling in machine learning classification of UAV geospatial data. *Earth Science Informatics*, 17(5), 4407-4418. doi.org/10.1007/s12145-024-01402-9

Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint*, arXiv:1712.04621. doi.org/10.48550/arXiv.1712.04621

Schick, T., Schütze, H., 2020. Exploiting cloze questions for few-shot text classification and natural language inference. *arXiv preprint*, arXiv:2001.07676. doi.org/10.48550/arXiv.2001.07676

Sun, X., Ji, Y., Ma, B., Li, H., 2023. A comparative study between full-parameter and LoRA-based fine-tuning on Chinese instruction data for instruction following large language model. *arXiv preprint*, arXiv:2304.08109. doi.org/10.48550/arXiv.2304.08109

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. *In: Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Long Beach, CA, USA, 5998-6008. doi.org/10.5555/3295222.3295349

Wang, G., Chen, Y., An, P., Hong, H., Hu, J., Huang, T., 2023. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors*, 23(16), 7190. doi.org/10.3390/s23167190

Yang, E., Zhou, W., Qian, X., Yu, L., 2022. MGCNet: Multilevel gated collaborative network for RGB-D semantic segmentation of indoor scene. *IEEE Signal Processing Letters*, 29, 2567-2571. doi.org/10.1109/LSP.2022.3224085

Yang, Z., Zhou, D., Yang, Y., Zhang, J., Chen, Z., 2022. TransRoadNet: A novel road extraction method for remote sensing images via combining high-level semantic feature and context. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5. doi.org/10.1109/LGRS.2022.3189642

Yu, Y., Xu, C., Wang, K., 2024. TS-SAM: fine-tuning segment-anything model for downstream tasks. *In: 2024 IEEE International Conference on Multimedia and Expo (ICME)*, Niagara Falls, ON, Canada. doi.org/10.48550/arXiv.2408.01835

Yuan, X., Shi, J., Gu, L., 2021. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, 169, 114417. doi.org/10.1016/j.eswa.2020.114417

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. *In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2879-2887. doi.org/10.1109/CVPR.2017.660

Zhao, Y. G., 2024. Application of remote sensing image technology in map compilation. *Intelligent Building & Smart City*, (2), 49-51.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J., 2020. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(10), 3068-3078. doi.org/10.1109/TMI.2020.3001005

Zhu, Q. Q., Zhang, Y. N., Wang, L. Z., Zhong, Y. F., Guan, Q. F., Lu, X. Y., Zhang, L. P., Li, D. R., 2021. 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. doi.org/10.1016/j.isprsjprs.2021.02.008