International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

AN INTELLIGENT FRAMEWORK FOR AERIAL IMAGE DETECTION USING U-NET AND YOLOV10

Bismah Tehreem Ahsen[1], Dr. Lalitha Saroja[2]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, bismahahsen@gmail.com
[2]Associate Professor, Department of CSE, Shadan Women's College of Engineering and Technology

ABSTRACT

We provide a unique method that combines a U-Net architecture with YOLOv10 to handle the difficulties of few-shot aerial picture semantic segmentation, where unseen-category items in query aerial images must be identified using only a few annotated support images. Typically, few-shot segmentation involves segmenting query images pixel-by-pixel using category prototypes that are extracted from support samples. However, objects in aerial photos frequently have irregular spatial distributions and arbitrary orientations, which results in notable differences in their properties. Low confidence scores and incorrect classification of same-category items with varying orientations are common outcomes of conventional approaches that do not take orientation changes into account. In order to get over these restrictions, our method combines YOLOv10 for quick and accurate object identification with U-Net for precise semantic segmentation, allowing for reliable item localisation even in challenging aerial scenarios. Because of YOLOv10's sophisticated detection capabilities, the system can recognise objects in a variety of sizes and orientations, producing trustworthy bounding boxes that direct the U-Net segmentation procedure. Combining these two architectures improves the network's ability to recognise rotated and dispersed aerial objects and reduces oscillation in confidence scores by reliably segmenting same-category objects independent of rotation or size. This U-Net + YOLOv10 framework successfully bridges the gap between effective object recognition and high-fidelity semantic segmentation in difficult aerial imagery by providing a scalable, rotation-robust few-shot aerial image segmentation solution.

1. INTRODUCTION

In many practical applications, such as urban planning, disaster management, environmental monitoring, and surveillance, aerial picture analysis is essential. However, because of the unpredictable orientations, different scales, and intricate spatial distributions of things like buildings, roads, and cars, semantic segmentation of aerial photos poses particular difficulties. In few-shot learning circumstances, where there are few annotated support samples available to help the segmentation of unrecognised categories, these difficulties are exacerbated. Due to their sensitivity to rotation and feature distortions, conventional approaches frequently struggle in these situations, resulting in misclassification and decreased accuracy. We suggest a unique framework that combines YOLOv10, a cutting-edge real-time object identification model, with U-Net, which is renowned for its strong pixel-level segmentation capabilities, to address these problems. The suggested U-Net + YOLOv10 architecture provides a scalable and rotation-robust solution for few-shot aerial picture semantic segmentation by fusing YOLOv10's capacity to localise objects across various scales and orientations with U-Net's fine-grained segmentation power. We suggest a unique U-Net + YOLOv10 architecture for few-shot aerial picture semantic segmentation in order to overcome these difficulties. Because of its encoder–decoder structure with skip connections, which maintain spatial information, U-Net is renowned for its robust pixel-level segmentation performance. Nevertheless, complicated object localisation and rotation variations in aerial imagery are beyond the capabilities of U-Net alone. We use YOLOv10, a cutting-edge object identification model, into the segmentation pipeline in order to get over this restriction. The framework can identify objects of interest at different scales and orientations thanks to YOLOv10's reliable and fast object localisation. In order to ensure precise border delineation while lowering false positives, the identified regions are subsequently fed into U-Net for fine-grained segmentation. The combination of YOLOv10 with U-Net improves robustness against rotation-induced feature distortions and sparse annotation settings in addition to increasing segmentation accuracy.

2. EXISTING SYSTEM OF PROJECT

Few-shot aerial picture semantic segmentation is a difficult problem that uses only a few annotated support images to accurately separate items from unseen categories in query photographs. Working with aerial photos, where items like buildings, roads, and cars might appear in unpredictable orientations and cause large feature variations, makes the process considerably more challenging. Conventional techniques, which match support and query images pixel by pixel, frequently overlook these orientation changes, resulting in inaccurate segmentations. Furthermore, mistaken

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

recognition is caused by the fluctuation of confidence scores in current rotation-insensitive algorithms, especially for rotated objects with lower confidence scores.

We present FRINet (Few-shot Rotation-Invariant Network), a unique deep learning model created especially for few-shot aerial picture semantic segmentation, to address these issues. The fundamental innovation of FRINet is its capacity to extract rotation-invariant characteristics, which guarantees accurate recognition and segmentation of objects belonging to the same category regardless of their orientation. This strategy overcomes the drawbacks of traditional techniques, which have trouble matching items in support and query images when those objects are arbitrarily rotated.

3. RELATED WORK

In order to put our suggested methodology in context, this section reviews recent semantic segmentation research, with a special emphasis on few-shot semantic segmentation methods and strategies designed for aerial photography.

A. Segmentation via Semantics

In computer vision, semantic segmentation has been studied in great detail. Real-time semantic segmentation on high-resolution images is pioneered by the FCN [4]. A symmetric expanding path for accurate localisation and a contracting path for context capture are introduced by U-Net [33]. A number of initiatives have been put out in recent years to enhance semantic segmentation model performance. For example, Chen et al. [34] introduced DeepLabv3+, a cutting-edge semantic segmentation network that uses a decoder module and atrous spatial pyramid pooling (ASPP) for improved performance. In the meantime, Zhao et al. CCNet [36] is another noteworthy study in this area that presents an encoder-decoder architecture with a unique crisscross attention module to improve the segmentation accuracy of the feature maps. As shown by Zhao et al. [37], GCNs [27] have also gained popularity for collecting pixel dependencies and producing superior feature representations. In order to increase the accuracy of semantic segmentation models, these studies emphasise the significance of integrating contextual information, attention mechanisms, and feature coupling. However, the substantial annotation needs of supervised segmentation techniques provide difficulties because of costly and constrained datasets as well as possible annotation errors.

B. Semantic Segmentation in Few Shots

Early techniques in the field of visual scenes presented a number of approaches, including as semi-supervised learning [10], weakly supervised learning [38], and unsupervised learning [39], with the goal of overcoming the problem of inadequate annotated sample data.Few-shot learning has received a lot of attention lately, especially when it comes to its extension to semantic segmentation [40]. The issue of few-shot semantic segmentation has been addressed in a number of recent studies. For example, a few-shot segmentation method based on an FCN that can learn to build classifier weights for each class was first proposed by Shaban et al. [41]. The goal of prototype learning frameworks like PL [42] and PANet [15] is to learn consistent embedding prototypes and adjust to new requirements. Other research, like CANet [43], which integrates an iterative optimisation module on the merged query and support feature to iteratively refine results, concentrate on mining hierarchical information to optimise the parsing result for few-shot segmentation.

4. PROPOSED SYSTEM

We present a unique solution by merging a U-Net architecture with YOLOv10 to handle the difficulties of few-shot aerial picture semantic segmentation, where unseen-category items in query aerial photographs must be parsed with only a few annotated support images. In few-shot segmentation, query images are typically segmented by pixel-wise matching using category prototypes that are taken from support samples. However, the unpredictable distribution and orientations of aerial objects in these photos frequently lead to notable feature fluctuations, rendering traditional methods—which do not take orientation changes into account—ineffective. Significant feature distortions are caused by the rotation sensitivity of aerial photos, which results in poor confidence ratings and incorrect classification of objects of the same category with different orientations.

We suggest an improved approach to get around these restrictions by fusing YOLOv10 for precise object identification and localisation with U-Net for reliable semantic segmentation. This design, which we call YOLO-U-Net, makes use of YOLOv10's quick and accurate detection capabilities to recognise objects in a variety of scales and orientations, giving the U-Net segmentation process trustworthy advice. The network can efficiently extract orientation-varying but category-consistent features from support images by aligning the detected object regions with U-Net's segmentation algorithm.

6. MODULES

1. Data Preprocessing Module:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Aerial picture datasets are gathered and meticulously prepared for evaluation and training in this module. Preprocessing techniques like normalisation, scaling, and augmentation are used because aerial photographs frequently experience problems like scale fluctuations, illumination changes, and rotation. To guarantee that the model learns orientation-invariant characteristics, transformations like rotation, flipping, cropping, and scaling are especially crucial. This module makes sure that the input data is reliable, varied, and able to successfully handle few-shot learning tasks.

2. YOLOv10 Detection Module:

This module is in charge of quickly and precisely identifying objects in aerial photos. YOLOv10 is used to create bounding boxes that offer localisation hints and to identify objects of interest. YOLOv10's ability to handle objects at different scales and orientations, in contrast to older approaches, is crucial for aerial photos where objects may seem rotated or scattered. This module's output provides a solid basis for directing the segmentation process, guaranteeing that only pertinent areas are further processed.

3. U-Net Segmentation Module:

After obtaining the bounding boxes from YOLOv10, this module uses U-Net to segment at the pixel level. Because the U-Net architecture can collect both high-level and low-level picture information, it is very useful for accurate boundary identification. U-Net can concentrate on areas of interest and produce precise segmentation masks that capture fine-grained object structures by using bounding box information as a guide. In spite of complicated spatial distributions, this improves the system's overall accuracy in identifying objects of the same category.

4. Few-Shot Prototype Learning Module:

Enabling the system to function in few-shot learning scenarios is the main goal of this module. The segmentation of query photos is guided by category prototypes that are extracted from a limited collection of annotated support images. With little training data, the model may identify previously undiscovered categories thanks to these prototypes that capture the essential characteristics of each object class. Because labelled datasets are frequently scarce in aerial images, our method guarantees adaptation to new item classes.

5. Integration and Evaluation Module:

The results of YOLOv10 and U-Net are combined into a single framework in the final module. The system achieves robust, rotation-invariant segmentation performance by combining semantic segmentation with object detection. In order to evaluate system performance, this module also supports evaluation using common measures including accuracy, confidence stability, and mean Intersection over Union (mIoU). To show how well the suggested method handles rotation, scale fluctuations, and few-shot situations, comparative experiments with baseline methods are carried out.

6. Performance Evaluation and Monitoring:

Using common metrics like accuracy, precision, recall, and F1-score, the assessment module focuses on evaluating the fire detection system's efficacy. It guarantees that the system is operating as anticipated and offers suggestions for enhancements. In order to ensure that the system continues to be effective and scalable in production settings, this module also monitors the system's capacity to handle massive amounts of data in real-time.

6. TECHNIQUES USED IN PROJECT

6.1 Few Shot Rotation Invariant Transform Network (FRINET)

Despite its benefits, FRINet has certain drawbacks. Compared to traditional architectures, the model's computational complexity is increased by the addition of rotation-invariant methods, which could slow down inference. Large-scale labelled datasets are not as necessary thanks to its few-shot capability, but performance can still drastically deteriorate when there are very few or noisy support samples. Furthermore, the model's reliance on rotation-invariant transformations may occasionally result in the loss of fine-grained orientation features that are crucial for particular tasks, like identifying runway orientation or vehicle heading. Additionally, training FRINet necessitates precise hyperparameter adjustment to strike a compromise between rotation invariance and generalisation, which may make deployment more difficult in practical settings.

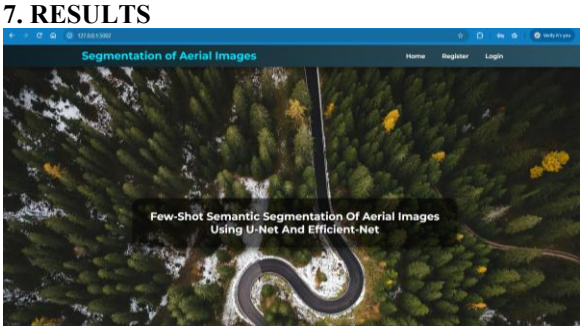
6.2 ALGORITHM USED:

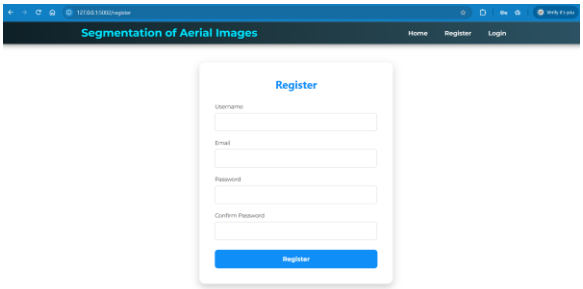
Even when things are photographed from various perspectives, it can aid in the precise detection and segmentation of buildings, roads, and other infrastructure in urban planning. After natural disasters like earthquakes or floods, when objects may surface at different angles, YOLO-U-Net can be used in disaster response to identify and evaluate damage to buildings or infrastructure. The model's adaptability to various scales and rotations is especially helpful in environmental monitoring when following changes in water bodies, forests, or land cover over time. YOLO-U-Net is a potent tool for security and defence applications since it may assist in the identification and tracking of cars, planes, or other moving objects across various angles and scales in aerial surveillance.

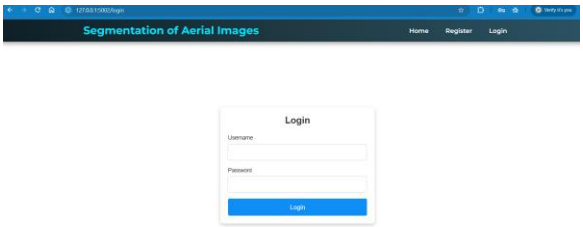
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

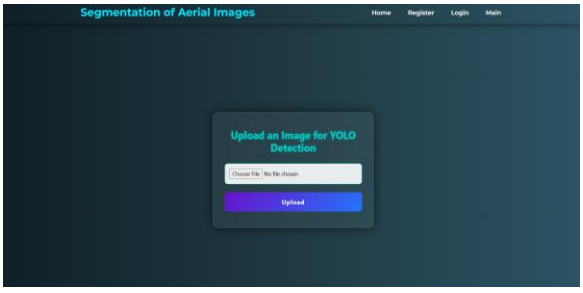
By using YOLOv10's rotation- and scale-aware detection to direct U-Net's segmentation, this model improves traditional few-shot segmentation. This is crucial for aerial images, where objects like buildings and roads frequently appear in arbitrary orientations, resulting in notable feature variations.

7. RESULTS



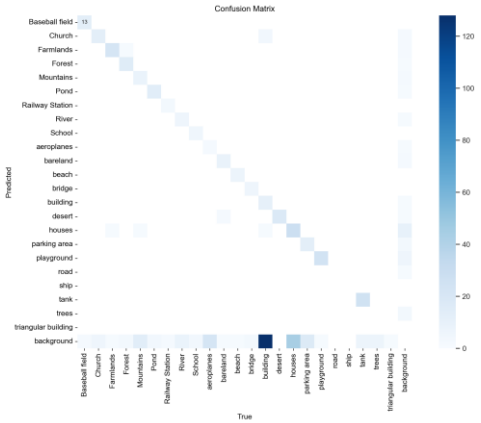


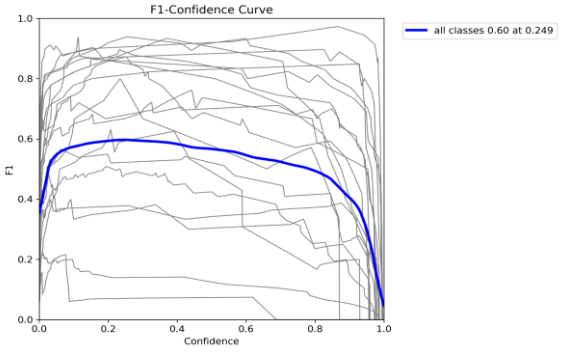




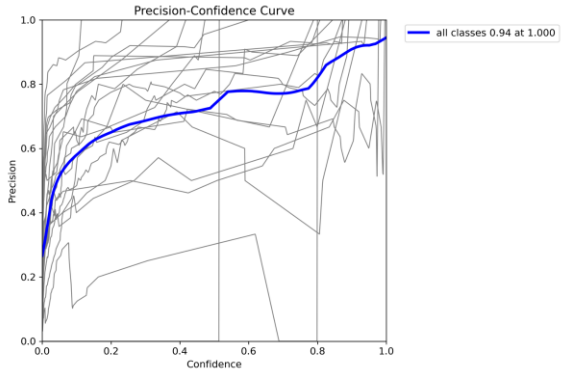


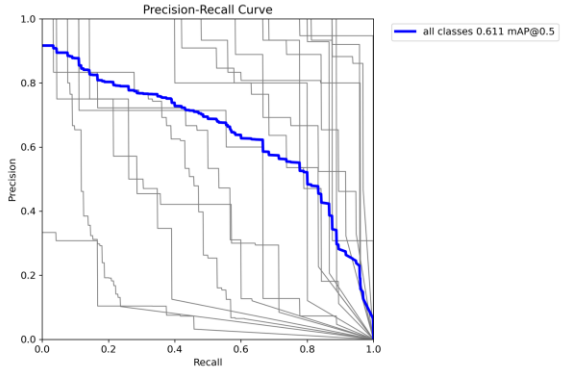


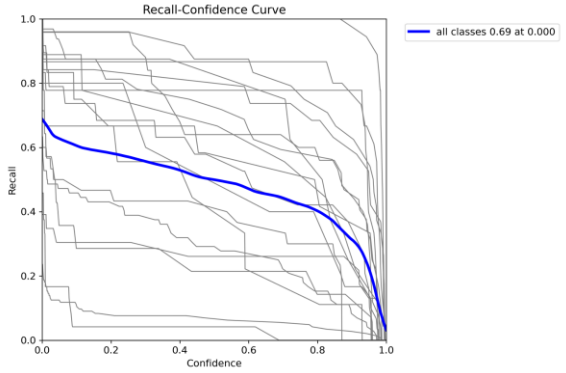


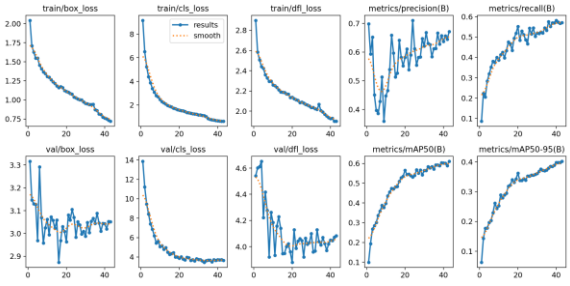


International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper









8. CONCLUSION

The suggested method successfully overcomes the drawbacks of traditional approaches that struggle with rotation sensitivity and feature variations by fusing YOLOv10's capacity for quick and accurate object detection across a variety of orientations and scales with U-Net's strength in producing precise pixel-level segmentation. In addition to offering a scalable and workable solution for real-world applications including urban planning, disaster response, environmental monitoring, and aerial surveillance, this hybrid framework guarantees reliable recognition and segmentation of unseen-category items in aerial imagery. The suggested method shows promise for bridging the gap between effective object detection and high-fidelity semantic segmentation. clearing the path for future aerial picture analysis that is more sophisticated, dependable, and rotation-invariant.

9. FUTURE ENHANCEMENT

The suggested U-Net and YOLOv10 integrated framework for few-shot aerial picture segmentation can be improved in the future to handle new problems in practical applications. The use of transformer-based designs, which can enhance the identification of intricate spatial patterns and capture long-range interdependence in aerial data, is one promising avenue. Furthermore, the model may be better able to generalise across various geographic locations and imaging settings without the need for sizable annotated datasets if self-supervised learning and domain adaption techniques are incorporated. Additionally, by maximising computing efficiency and utilising lightweight model compression techniques, the system might be expanded for real-time processing on edge devices. Additionally, the robustness of segmentation results can be improved through multi-modal data fusion with LiDAR, hyperspectral photography, or SAR data. especially in situations involving environmental monitoring and disaster management. These improvements would increase the framework's scalability, adaptability, and suitability for a greater variety of remote sensing and aerial surveillance applications.

REFERENCES

1] A. Kae, K. Sohn, H. Lee, and E. Learned-Miller, “Augmenting CRFs with Boltzmann machine shape priors for image labeling,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 2019–2026.

[2] R. Zhang et al., “Cognition-driven structural prior for instance-dependent label transition matrix estimation,” IEEE Trans. Neural Netw. Learn. Syst., 2024, early access, doi: 10.1109/TNNLS.2023.3347633.

[3] R. Zhang, L. Xu, Z. Yu, Y. Shi, C. Mu, and M. Xu, “Deep-IRTarget: An automatic target detector in

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

infrared imagery using dual-domain feature ex- traction and allocation," IEEE Trans. Multimedia, vol. 24, pp. 1735–1749, 2022.
[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3431–3440.
[5] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets,atrous convolution, and fully connected CRFs," IEEE Trans. Pattern Anal.Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr. 2018.
[6] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, "Semantic segmen-tation of remote sensing images with sparse annotations," IEEE Geosci. Remote Sens. Lett., vol. 19, 2022, Art. no. 8006305.
[7] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," Expert Syst. Appl., vol. 169, 2021, Art. no. 114417.
[8] X. He, Y. Zhou, J. Zhao, D. Zhang, R. Yao, and Y. Xue, "Swin transformer embedding UNet for remote sensing image semantic segmentation," IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 4408715.
[9] Y. Mao et al., "Beyond single receptive field: A receptive field fusion- and-stratification network for airborne laser scanning point cloud classi-fication," ISPRS J. Photogrammetry Remote Sens., vol. 188, pp. 45–61, 2022.
[10] X. Qi, Y. Mao, Y. Zhang, Y. Deng, H. Wei, and L. Wang, "PICS: Paradigms integration and contrastive selection for semisupervised remote sens-ing images semantic segmentation," IEEE Trans. Geosci. Remote Sens.,vol. 61, 2023, Art. no. 5602119.
[11] Y. Song, T. Wang, S. K. Mondal, and J. P. Sahoo, "A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportuni-ties," 2022, arXiv:2205.06743.
[12] A. Parnami and M. Lee, "Learning from few examples: A summary of approaches to few-shot learning," 2022, arXiv:2203.04291.
[13] R. Zhang et al., "Part-aware correlation networks for few-shot learning," IEEE Trans. Multimedia, vol. 26, pp. 9527–9538, 2024.
[14] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018,pp. 1199–1208.
[15] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "PANet: Few-shot im- age semantic segmentation with prototype alignment," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 9197–9206.
[16] W. Jiang, K. Huang, J. Geng, and X. Deng, "Multi-scale metric learning for few-shot learning," IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 3, pp. 1091–1102, Mar. 2020.
[17] O. Vinyals et al., "Matching networks for one shot learning," in Proc. Int. Conf. Adv. Neural Inf. Process. Syst., 2016, vol. 29, pp. 3637–3645.
[18] Y. Liu, X. Zhang, S. Zhang, and X. He, "Part-aware prototype network for few-shot semantic segmentation," in Proc. Int. Conf. 16th Eur. Conf. Comput. Vis., Glasgow, U.K., 2020, pp. 142–158.
[19] B. Hui, P. Zhu, Q. Hu, and Q. Wang, "Self-attention relation network for few-shot learning," in Proc. IEEE Int. Conf. Multimedia Expo Workshops, 2019, pp. 198–203.
[20] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," in Proc. Int. Conf. Mach. Learn., 2016, pp. 1842–1850.
[21] R. Li, H. Liu, Y. Zhu, and Z. Bai, "ARNet: Attention-based refinement network for few-shot semantic segmentation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2020, pp. 2238–2242.
[22] W. Liu, C. Zhang, G. Lin, and F. Liu, "CRCNet: Few-shot segmenta-tion with cross-reference and region–global conditional networks," Int. J. Comput. Vis., vol. 130, no. 12, pp. 3140–3157, 2022.
[23] B. Yang, C. Liu, B. Li, J. Jiao, and Q. Ye, "Prototype mixture models for few-shot semantic segmentation," in Proc. 16th Eur. Conf. Comput. Vis., Glasgow, U.K., 2020, pp. 763–778.
[24] X. Yao, Q. Cao, X. Feng, G. Cheng, and J. Han, "Scale-aware detailed matching for few-shot aerial image semantic segmentation," IEEE Trans. Geosci. Remote Sens., vol. 60, 2021, Art. no. 5611711.
[25] B. Wang, Z. Wang, X. Sun, H. Wang, and K. Fu, "DMML-Net: Deep metametric learning for few-shot geographic object segmentation in re- mote sensing imagery," IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5611118.
[26] F. Hao, F. He, J. Cheng, L. Wang, J. Cao, and D. Tao, "Collect and select: Semantic alignment metric learning for few-shot learning," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 8460–8469.
[27] G. Li, M. Muller, A. Thabet, and B. Ghanem, "DeepGCNs: Can GCNs go as deep as CNNs?," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 9267–9276.