Hindawi
Complexity
Volume 2022, Article ID 4901746, 7 pages
https://doi.org/10.1155/2022/4901746

Research Article

Few-Shot Segmentation via Capturing Interclass and Intraclass Cues Using Class Activation Map

Yan Zhao,[1] **Ganyun Lv**,[1] **and Gongyi Hong**[2]

[1]*School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China*
[2]*Nari Group Corporation, Nanjing 211000, China*

Correspondence should be addressed to Yan Zhao; zy_njit@163.com

Received 10 December 2021; Revised 23 March 2022; Accepted 22 April 2022; Published 4 July 2022

Academic Editor: Xuyun Zhang

Few-shot segmentation is a challenging task due to the limited class cues provided by a few of annotations. Discovering more class cues from known and unknown classes is the essential to few-shot segmentation. Existing method generates class cues mainly from common cues intra new classes where the similarity between support images and query images is measured to locate the foreground regions. However, the support images are not sufficient enough to measure the similarity since one or a few of support mask cannot describe the object of new class with large variations. In this paper, we capture the class cues by considering all images in the unknown classes, i.e., not only the support images but also the query images are used to capture the foreground regions. Moreover, the class-level labels in the known classes are also considered to capture the discriminative feature of new classes. The two aspects are achieved by class activation map which is used as attention map to improve the feature extraction. A new few-shot segmentation based on mask transferring and class activation map is proposed, and a new class activation map based on feature clustering is proposed to refine the class activation map. The proposed method is validated on Pascal Voc dataset. Experimental results demonstrate the effectiveness of the proposed method with larger mIoU values.

1. Introduction

Image segmentation [1] aims to segment object regions from images, which is fundamental to many computer vision tasks [2]. Based on the deep learning-based method [3–7], the existing segmentation models can segment object well when sufficient annotations are given [8]. However, the existing segmentation methods still have two drawbacks. Firstly, the annotation generation is time consuming. The number of annotations is usually so small that it is hard to train the segmentation models from a few of annotations. The other is that the segmentation models work badly on new classes, i.e., the segmentation models only recognize the objects in the training dataset and cannot segment regions of classes unknown.

To solve the drawbacks, few-shot segmentation [9–14] is proposed. Given a set of images of new classes, with a few of annotations (support images), the aim of few-shot segmentation is to segment region of query images efficiently. However, the intuitive method of refining the segmentation model by a few of annotations is proved to be ineffective. Few-shot segmentation faces the challenges of discovering object cues from limited annotations. To this end, researchers have proposed many methods to enhance few-shot segmentation [15–17]. These methods can be summarized to provide segmentation cues from existing annotations of known classes where the annotations are sufficient to train the model. Therefore, the class-agnostic guided model that transfers segmentation cues from support mask to query mask can be trained firstly and is then used in reference stage to locate the foreground regions in query image directly. Several strategies such as mask transferring and prototype feature are used. The few-shot segmentation has been improved obviously.

Meanwhile, few-shot segmentation still faces the lack of object priors although many existing annotation datasets are used. Two reasons caused such challenge. Firstly, there are large variation interclasses, which make the knowledge transferring between known class and new class very hard. Secondly, there is large variation intraclass. Therefore, a few of annotations cannot describe all the types of classes and leads to bad guidance. In other words, the foreground priors are still limited by current few-shot segmentation manner.

In this paper, we propose a new few-shot segmentation method that considers two aspects, namely, interclass cue and intraclass cue to capture more sufficient segmentation cues from known and unknown classes. The first one captures the semantic relationships between the existing classes and unknown classes and is used to capture the discriminative cue through comparing existing classes and unknown classes. The second one captures the common cues intraclasses, that is, the common features shared by the query and support images are captured to locate the object. The two aspects are achieved by class activation maps (CAMs). A classification model considering only class-level labels is first built. Then, class activation map is extracted based on the feedback analysis. Afterwards, since the discriminative regions are usually small, we expand the discriminative region using the feature clustering method guided by support masks. Finally, the CAM is introduced into the few-shot segmentation mask as an attention map to enhance the query image segmentation.

The contributions of the proposed method are listed as follows:

(1) A new few-shot segmentation method based on the segmentation cues interclass and intraclass is proposed

(2) Class activation map is used to capture the segmentation cues, and a new attention module is proposed to add the class activation map in to the few-shot segmentation network

(3) An extension method based on the clustering method is proposed to enlarging class activation map

2. Related Work

Few-shot segmentation aims to segment regions of new classes with a few annotated images given, which is a fundamental task in computer computing [18, 19]. The few-shot segmentation task is always formulated as an information guidance model, where the common knowledge that can be used in segmentation task is learned in the support branch and transferred between the support branch and the query branch. There are two key components in existing few-shot segmentation methods, of which the first component is a class prior extraction module in the support branch, and the second component is a guidance network to transfer the extracted knowledge between branches.

As for class prior extraction, multiple types of class prior have been proposed and can be further categorized into the weight-based methods and prototype-based methods. The weight-based methods consider the weight of a classifier as the class prior. The most representative work in the weight-based methods is OSLSM [10], which leverages a conditional branch to generate parameters for query branch.

The current state-of-the-art methods are prototype-based methods. The prototype-based methods can be further divided into the global prototype, the fusion of global and local prototype, and the prototype of background. The global prototype-based methods consider the converted deep features from the support branch into the class prototype, e.g., PANet [20] and CANet [21] learn a class-specific global prototype with a masked average pooling operation.

The second type of prototype-based methods takes the global and local prototypes into consideration simultaneously and has the ability to extract features with more semantic knowledge. The most representative methods are PPNet [22] and PMMs [22], where the first method decomposes the holistic class representation into a set of part-aware prototypes with k-means and the second correlates the diverse image regions with multiple prototypes to enforce the prototype-based representation with the aid of the EM algorithm.

The third type of prototype-based methods employs the background prototype to enhance the semantic knowledge of foreground. The most representative methods are MLCNet [23] and SCNet [24], where the first method introduces a mining branch that exploits latent novel classes via transferable subclusters and the second method generates self-contrastive background prototypes directly from the query image, enabling the construction of complete sample pairs to form a complementary and auxiliary segmentation task.

As for the design of guidance network between support branch and query branch, multiple types of guidance module have been proposed and can be further categorized into the feature-level guidance network and the parameter-level guidance network. The feature-level guidance conducts similarity propagation based on the extracted features by diverse branches. The representative methods include PFENet [25], which generates the prior mask based on the cosine similarity between features, and then employ the feature enrichment module to propagate this similarity in multiple resolutions. LTM [26] proposed a nonparametric and class-agnostic transformation method, where the relationship of the local features is calculated in a high-dimension metric embedding space based on cosine distance, and then are mapped from the low-level local relationships to high-level semantic cues with the generalized inverse matrix of the annotation matrix. The parameter-level guidance network considers the model parameter of the last specific layer as the class prior and uses the parameter transformation from support branch to query branch to achieve the guidance. The most representative work is CWT [27], where the guidance is conducted at the classification layer only; it proposed a Classifier Weight Transformer to dynamically adapt the support-set trained classifier's weights to each query image in an inductive way.

3. The Proposed Method

3.1. The Pipeline of the Proposed Method. The pipeline of the proposed method is shown in Figure 1, where the proposed method consists of four steps: the classification step, the

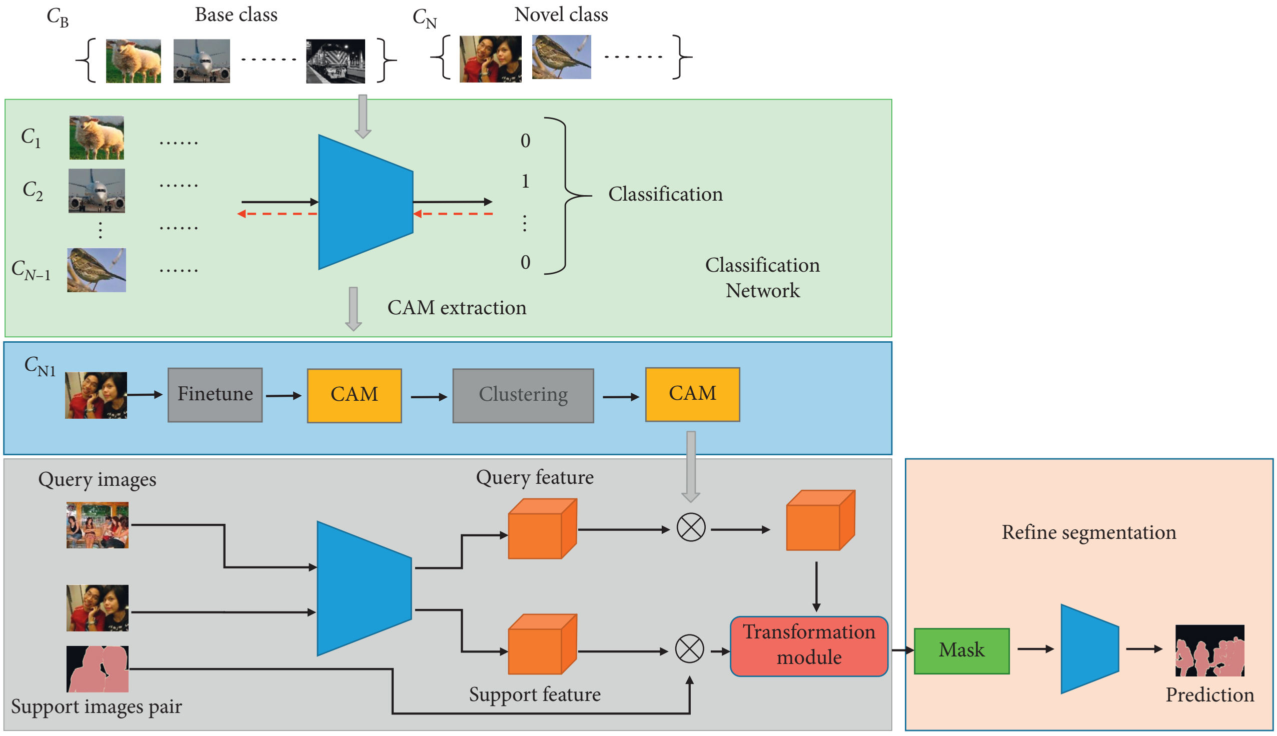

Figure 1: The pipeline of the proposed method.

CAM generation step, the mask generation step, and the mask refinement step. The classification step is to train a classification network by considering all the existing classes and the new classes based on image-level labels only and output the class activation map that represents the discriminative regions of the unknown classes via gradient feedback forward. Then, since the initial CAM is usually very small, the CAM generation step expands the CAM using the clustering strategy. Afterwards, the mask generation step generates the segmentation mask in terms of soft values based on mask transferring strategy where the CAM generated in the second step is used as attention map to enhance the features of the query image. Finally, the mask refinement step is to improve the segmentation mask based on classical segmentation framework. We next detail the four steps.

3.2. Classification Step. The aim of the classification step is to train a classification model by considering the known classes and new classes and extracts the discriminative regions of new classes that distinct the new class from the existing classes. Therefore, the rough location of new classes can be obtained in the query image.

Specifically, a training dataset $\{C_B, C_N\}$ consisting of the existing classes and the new classes is constructed firstly. Here, $C_B = \{C_1, \ldots, C_{N-1}\}$ is composed of existing classes with number $N-1$. $C_N = \{C_{N1}, \ldots C_{Nk}\}$ is the image set of new classes. Based on all classes, a classification network is trained, and the classification map C_0 is extracted using Grad-CAM methods.

Meanwhile, the regions are usually small due to the fact that the rough image-level labels cannot obtain the whole region of the object but a small area. The next step expands the highlighted region using feature clustering.

3.3. CAM Generation Step. The CAM generation step expands the class activation maps based on the idea that the regions located by the initial step can be treated as the class center, and the rest pixels similar with the region highlighted can be treated as the object regions. Therefore, we use the clustering method to obtain the similar pixels.

Specifically, the CAM generation step consists of three substeps: pixel clustering, cluster selection, and CAM generation. In the first step, the K-means clustering [28] is used to cluster the pixels into n_g clusters based on the deep features obtained in the classification step. For each cluster, each pixel is given the activation value in the class activation map, and the mean value of the cluster is obtained through averaging the activation values. The mean value represents the important for the pixel to the class, and the mean value is used as the activation value for all the pixels in the cluster. Thus, a new class activation map M is obtained.

3.4. Mask Generation Step. Mask generation step segments foreground regions of query image based on the class activation map M. Here, a few-shot segmentation network based on transferring is used, and the class activation map M is embedded into the network to enhance the guidance.

3.4.1. Few-Shot Segmentation Network. The few-shot segmentation network is constructed by the method in [26], of which the idea is to obtain the query mask M_q (with size $n \times n$) based on the relationships as follows:

$$M_q * M_s^T = R, \tag{1}$$

where M_q and M_s are the query mask and support mask, respectively, and the two masks are reshaped into column vector. R is the matrix product of M_q and M_s, with size $n^2 \times n^2$. It is seen that value one in R means that the values in M_q and M_s are all value one. Otherwise, the value in R is zero.

Once R is known, the query mask can be obtained by

$$M_q = R * \left(M_s^T\right)^{-1}. \tag{2}$$

Thus, the few-shot segmentation problem changes to obtain the Matrix product R, which can be estimated by the feature similarity of the pixels in the support and query images, i.e., the foreground pixels have the similar features, and have similarity distance of value one. Otherwise, the distance is value zero.

Based on the formulation above, the few-shot segmentation network can be constructed as a two-branch based network, with a guidance model by formula (2). The network is shown in Figure 1.

Specifically, given a support image I_s with support mask M_s and query image I_q, a two-branch based network is used to extract the pixel features. One is the support branch that extracts the features of support image F_s, and the other is the query branch that extracts the features of query image F_q. Then, the similarity matrix M_{sq} of F_s and F_q is calculated via calculating the discrete cosine distance, where

$$M_{sq}(i, j) = d\left(F_s\{i\}, F_q\{j\}\right), \tag{3}$$

where $M_{sq}(i, j)$ is the value at location (i, j). $F_s\{i\}$ and $F_q\{j\}$ are the i th feature and j th feature in F_s and F_q. d is the discrete cosine distance. Therefore, M_{sq} refers to the similarity of pixels, which is similar with the similarity relationships of masks, and can be used to estimate the matrix product R.

Then, M_{sq} is used to estimate the matrix product R and is used to obtain the query mask via (2).

Note that estimating R using the pixel feature is challenging. Thus, we use the support mask to filter the foreground regions via element-wise production.

3.4.2. Feature Enhancement via CAM. Different from the few-shot segmentation method in [26], we introduce the class activation map which carries the discriminative cues through all classes to enhance the features of query image. Specifically, as shown in Figure 1, the query image is sent into the classification network to form the initial classification map. Then, the clustering algorithm is used to refine the class activation map. The class activation map is then used as attention map to refine the deep features of query image, and the refined features guide the segmentation of query image.

3.5. Mask Refinement Step. The output of few-shot segmentation branch is the soft mask of query image. To obtain the binary mask, a threshold can be used to obtain the hard mask from the soft mask. However, the results are sensitive to the selection of threshold. Therefore, a segmentation mask is used to segment the final mask from the soft mask, where the soft mask is used as foreground probability map, and the segmentation network is performed to obtain the final hard segmentation mask. We use the method in [8] to implement the mask refinement.

4. Experimental Results

4.1. Dataset. We next verify the proposed method based on the Pascal Voc dataset which consists of 20 classes. Similar with the existing few-shot segmentation method, the 20 classes are split into two class set. One is training set that trains the few-shot segmentation network. The other is the test set that validates the segmentation quality of the network. To fully validate the few-shot segmentation model, four splits are used. The details are found in Table 1.

4.2. Implementation Details. We implement our method on Titan-XP GPU. Pytorch is used to realize our method. The network is optimized by Adam optimizer with the initial learning rate $1e-4$. Several backbones such as VGG16, ResNet50, and ResNet 101 are used for sufficient evaluation. The pretrained backbone network based on ImageNet [29] is used for training.

4.3. Subjective Results. We first display some subjective results in Figure 2, where the input image, the prediction results, and the ground truth results are displayed. It is seen that the prediction results are similar with the ground truth results, which demonstrates the fact that our method can segment these new classes of images successfully.

4.4. Objective Results. We objectively evaluate the proposed method by mIoU and FB-IoU values that are usually used for few-shot segmentation evaluation. The results are shown in Table 2, where 1-shot and 5-shot mean the few-shot segmentation with one and five support annotations, respectively. Three backbones such as VGG16, ResNet 50, and ResNet 101 are considered. We can see that ResNet 101 obtains the best results due to the deeper layers in the networks. The results by ResNet 50 are better than VGG 16, which is also caused by the deeper network that captures more semantic features.

4.5. Comparison with Existing Methods. We also compare our method with the existing state-of-the-art methods. The comparison methods are displayed in Table 2. It is seen that our method outperforms these comparison methods, which demonstrate the effectiveness of the proposed method, especially for the comparison with the method in [26], our method can be considered as a improvement of the method [26]. It is seen that our method is better than the method in

Table 1: The detailed splitting of Pascal Voc 2012 dataset.

Subdataset	Corresponding classes
PASCAL-5^0	Aeroplane, bicycle, bird, boat, bottle
PASCAL-5^1	Bus, car, cat, chair, cow
PASCAL-5^2	Dining table, dog, horse, motorbike, person
PASCAL-5^3	Potted plant, sheep, sofa, train, tv/monitor

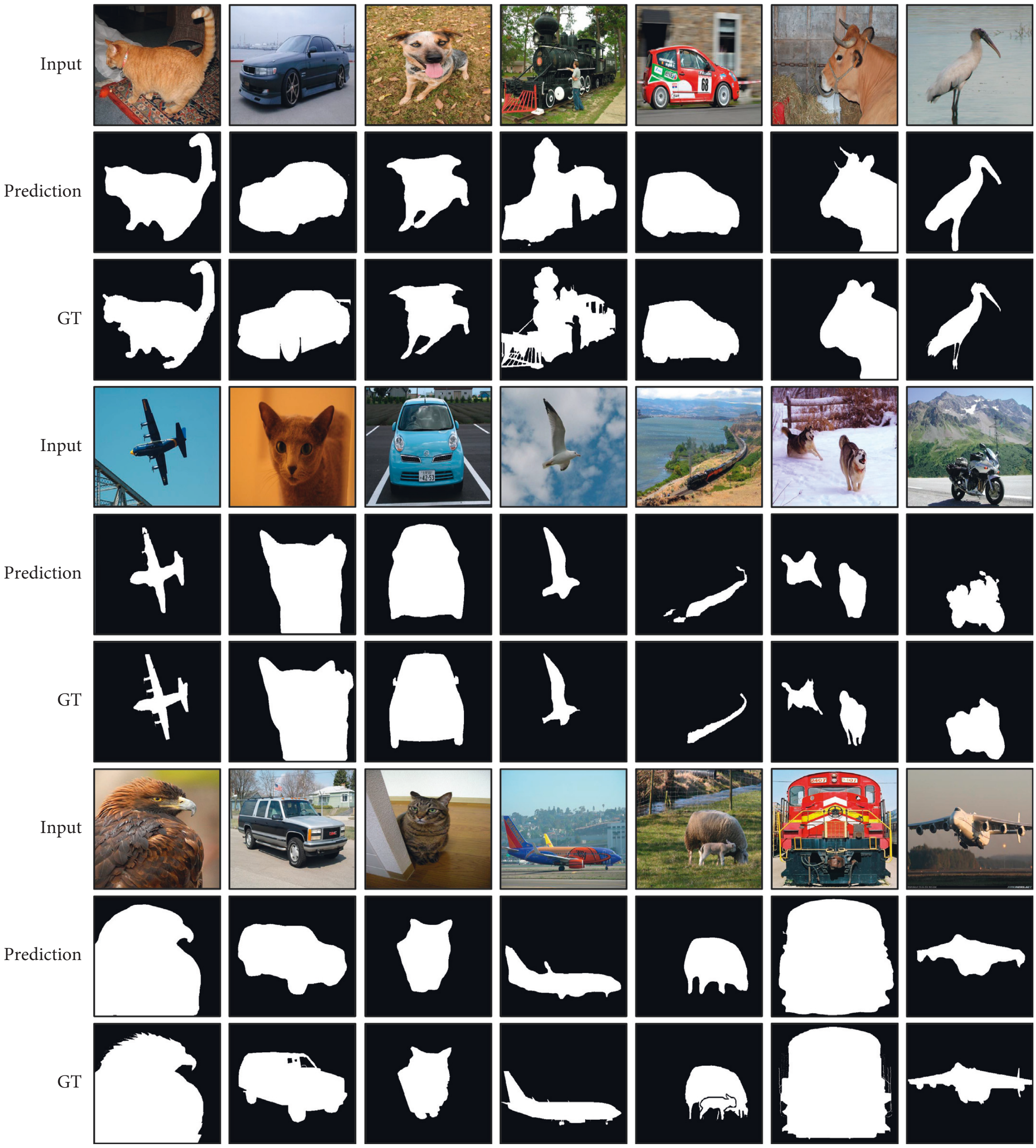

Figure 2: The segmentation results by the proposed method.

[26] which demonstrates the effectiveness of our strategy that introduces the class activation map to capture both the cues interclass and intraclass.

4.6. The Ablation Study. We next show the ablation results. The initial CAM and improved CAM are considered for the ablation study. The backbone ResNet 50 is used. The results

Table 2: Comparison with SOTA on the PASCAL-5^i dataset.

Backbone	Method	1-shot		5-shot	
		mIoU	FB-IoU	mIoU	FB-IoU
VGG 16	OSLSM [10]	40.8	61.3	43.9	61.5
	Co-FCN [30]	41.0	60.1	41.4	60.2
	SG-one [11]	46.3	63.1	47.1	65.9
	PANet [20]	48.1	66.5	55.7	70.7
	FWB [31]	51.9	—	55.1	—
	RPMM [22]	53.0	—	54.0	—
	Ours	**56.1**	**71.9**	**58.1**	**73.2**
ResNet 50	A-MCG	—	61.2	—	62.2
	CANet [21]	55.4	66.2	57.1	69.6
	PGNet [13]	56.0	66.9	58.5	70.5
	CRNet [32]	55.7	66.8	58.8	71.5
	RPMM [22]	56.3	—	57.3	—
	LTM [26]	57	—	60.6	—
	Ours	**58.3**	**73.7**	**60.9**	**74.4**
ResNet 101	FWB [31]	56.19	—	59.92	—
	DAN [33]	58.2	71.9	**60.5**	72.3
	LTM [26]	60	74	61.5	74.5
	Ours	**60.4**	**74.3**	**61.8**	**75.0**

Table 3: The ablation study on the PASCAL-5^i dataset and ResNet 50 backbone.

CAM	Our CAM	1-shot (mIoU)	5-shot (mIoU)
		57	60.6
✓		57.4	60.7
	✓	58.3	60.9

are shown in Table 3, where mIoU values are shown. It is seen that original CAM can also lead to the improvement. Meanwhile, our improved CAM can enhance the results further, which demonstrates that fact that clustering strategy is a useful method to enhance CAM regions.

5. Discussion

The existing few-shot segmentation methods usually focus on the learning class-agnostic model, which is based on the level interclasses only. Such class-agnostic model can lead to good generalization on new classes, which however also lacks the class cues of new classes. Based on the existing class-agnostic model, we try to add new segmentation cues through the discriminative cues interclass and the common cues intraclasses, which is the level of both interclass and intraclass. Therefore, better segmentation results can be obtained by our method.

It is seen that our method is based on the method in [26] (LTM), which proposed a few-shot segmentation method via estimating the relationship matrix of masks that is an interesting idea. However, our method is different from LTM [26]. Firstly, our main contribution is using the class activation map to capture the segmentation cues interclass and intraclass, which is not considered in [26]. Secondly, an attention module is added in LTM, which can add the CAM segmentation cues to enhance the segmentation. Therefore, our method can be considered as an extension to LTM [26] with better segmentation results."

6. Conclusion

This paper proposed a new few-shot segmentation method that uses the class activation map to enhance the generation of object priors by considering the common cues intraclass and the discriminative cues interclass. The proposed network consists of four steps: classification step, CAM generation step, mask generation step, and mask refinement step, which are used to generate the initial CAM via class classification, to generate the CAM via feature clustering, to generate the segmentation mask, and to refine the segmentation mask, respectively. The proposed method is validated on Pascal Voc dataset. The experimental results demonstrate that the consideration of common cues intraclass and the discriminative cues interclasses can enhance the few-shot segmentation in terms of large IoU values.

Data Availability

The datasets used for validation are available from https://host.robots.ox.ac.uk/pascal/VOC/. The detailed results are listed in the paper. More results can be found from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was partially supported by The National Natural Science Foundation of China (51577086); Jiangsu Six Talent Peaks (TD-XNY004); and Jiangsu Major Scientific Research Projects in Colleges and Universities (19KJA510012).

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 39, no. 4, pp. 3431–3440, Boston, MA, USA, June 2015.

[2] X. Xu, H. Li, W. Xu, Z. Liu, L. Yao, and F. Dai, "Artificial intelligence for edge service optimization in internet of vehicles: a survey," *Tsinghua Science and Technology*, vol. 27, no. 2, pp. 270–287, 2022.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representation (ICLR)*, San Diego, CA, USA, April 2015.

[5] Z. Hu, X. Xu, Y. Zhang et al., "Cloud-edge Cooperation for Meteorological Radar Big Data: A Review of Data Quality Control," *Complex & Intelligent Systems*, pp. 1–15, 2021.

[6] Q. Wang, C. Yuan, and Y. Liu, "Learning deep conditional neural network for image segmentation," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1839–1852, 2019.

[7] X. Xu, Z. Fang, J. Zhang et al., "Edge content caching with deep spatiotemporal residual network for iov in smart city," *ACM Transactions on Sensor Networks*, vol. 17, no. 3, pp. 1–33, 2021.

[8] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[9] Y. Liu, N. Liu, Q. Cao, X. Yao, J. Han, and L. Shao, "Learning non-target knowledge for few-shot semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, Louisiana, May 2022.

[10] A. Shaban, S. Bansal, Z. Liu, I. Essa, and B. Boots, "One-shot learning for semantic segmentation," in *Proceedings of the British Machine Vision Conference 2017, BMVC*, p. 167, London, UK, September 2017.

[11] X. Zhang, Y. Wei, Y. Yang, and T. S. Huang, "Sg-one: similarity guidance network for one-shot semantic segmentation," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3855–3865, 2020.

[12] N. Dong and E. Xing, "Few-shot semantic segmentation with prototype learning," in *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 79–91, Newcastle, UK, September 2018.

[13] C. Zhang, G. Lin, F. Liu, J. Guo, Q. Wu, and R. Yao, "Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9587–9595, Seoul, Korea (South), November 2019.

[14] Y. Yang, F. Meng, H. Li, K. N. Ngan, and Q. Wu, "A new few-shot segmentation network based on class representation," in *Proceedings of the 2019 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, IEEE, Sydney, NSW, Australia, December 2019.

[15] S. Zhang, T. Wu, S. Wu, and G. Guo, "Catrans: context and affinity transformer for few-shot segmentation," in *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, Vienna, Austria, July 2022.

[16] S. Gairola, M. Hemani, A. Chopra, and B. Krishnamurthy, "Simpropnet: Improved Similarity Propagation for Few-Shot Image Segmentation," 2020, https://arxiv.org/abs/2004.15014.

[17] J. Liu, Y. Bao, G. Xie, H. Xiong, J. Sonke, and E. Gavves, "Dynamic prototype convolution network for few-shot semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, April 2022.

[18] X. Xu, Z. Fang, L. Qi, X. Zhang, Q. He, and X. Zhou, "TripRes," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 2, pp. 1–21, 2021.

[19] B. Shen, X. Xu, L. Qi, X. Zhang, and G. Srivastava, "Dynamic server placement in edge computing toward internet of vehicles," *Computer Communications*, vol. 178, pp. 114–123, 2021.

[20] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: few-shot image semantic segmentation with prototype alignment," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9197–9206, Seoul, Korea, October 2019.

[21] C. Zhang, G. Lin, F. Liu, R. Yao, and C. Shen, "Canet: class-agnostic segmentation networks with iterative refinement and attentive few-shot learning," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5217–5226, Long Beach, CA, USA, June 2019.

[22] Y. Liu, X. Zhang, S. Zhang, and X. He, "Part-aware prototype network for few-shot semantic segmentation," in *Proceedings of the Computer Vision – ECCV 2020, European Conference on Computer Vision*, pp. 142–158, Springer, Glasgow, UK, August 2020.

[23] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "Mining latent classes for few-shot segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8721–8730, Montreal, BC, Canada, October 2021.

[24] J. Chen, B.-B. Gao, Z. Lu, J.-H. Xue, C. Wang, and Q. Liao, "Scnet: Enhancing Few-Shot Semantic Segmentation by Self-Contrastive Background Prototypes," 2021, https://arxiv.org/abs/2104.09216.

[25] Z. Tian, H. Zhao, M. Shu, Z. Yang, R. Li, and J. Jia, "Prior Guided Feature Enrichment Network for Few-Shot Segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, 2020.

[26] Y. Yang, F. Meng, H. Li, Q. Wu, X. Xu, and S. Chen, "A new local transformation module for few-shot segmentation," in *Proceedings of the International Conference on Multimedia Modeling*, pp. 76–87, Springer, Daejeon, South Korea, January 2020.

[27] Z. Lu, S. He, X. Zhu, L. Zhang, Y.-Z. Song, and T. Xiang, "Simpler is better: few-shot semantic segmentation with classifier weight transformer," in *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8741–8750, Montreal, BC, Canada, October 2021.

[28] J. A. Hartigan and M. A. Wong, "Algorithm as 136: a k-means clustering algorithm," *Applied Statistics*, vol. 28, no. 1, p. 100, 1979.

[29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Kai Li, and L. Li Fei-Fei, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, Miami, FL, USA, June 2009.

[30] K. Rakelly, E. Shelhamer, T. Darrell, A. Efros, and S. Levine, "Conditional networks for few-shot semantic segmentation," in *Proceedings of the International Conference on Learning Representation workshop (ICLRW)*, Vancouver, BC, Canada, April 2018.

[31] K. Nguyen and S. Todorovic, "Feature weighting and boosting for few-shot segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 622–631, Seoul, South Korea, September 2019.

[32] W. Liu, C. Zhang, G. Lin, and F. Liu, "Crnet: cross-reference networks for few-shot segmentation," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4165–4173, Seattle, WA, USA, June 2020.

[33] H. Wang, X. Zhang, Y. Hu, Y. Yang, X. Cao, and X. Zhen, "Few-shot semantic segmentation with democratic attention networks," in *Proceedings of the Computer Vision – ECCV 2020, Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 730–746, Springer, Glasgow, UK, August 2020.