This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Dynamic Knowledge Adapter with Probabilistic Calibration for Generalized Few-Shot Semantic Segmentation

Jintao Tong Haichen Zhou Yicong Liu Yiman Hu Yixiong Zou*
Huazhong University of Science and Technology
{jintaotong, m202273832, smnight, m202273659, yixiongz}@hust.edu.cn

Abstract

Generalized Few-shot Semantic Segmentation (GFSS) aims to use a few novel-class samples to enable the model trained on base classes to have the ability to segment for all classes (including base and novel classes). We analyze the three main reasons for the model's limited performance on GFSS: the lack of adaptability to learn novel classes, the instability that causes the catastrophic forgetting of base classes, and the biased prediction of imbalanced classes. To handle these issues, we design an auxiliary network (Dynamic Knowledge Adapter, DKA) for the GFSS task. Firstly, DKA handles the adaptability problem by selecting only efficient parameters for finetuning. Secondly, DKA addresses the stability problem by relabelling part of the training samples for iterative training, which alleviates the conflict between base and novel classes. Thirdly, it involves a probabilistic calibration module to help the model rectify the prediction bias caused by imbalanced data. Experimental results show that these designs can help the model to take into account the segmentation performance of base classes, novel classes, and the background class, that is, to perform well in all-class segmentation.

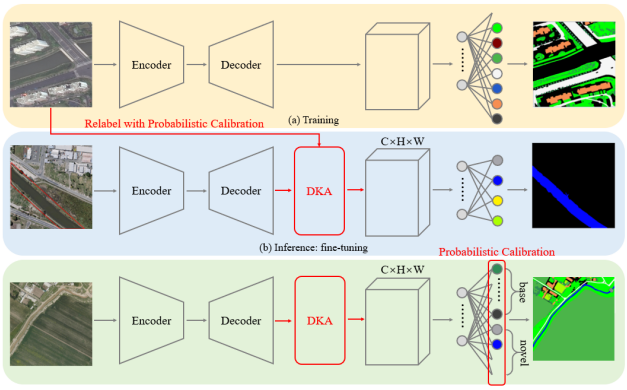

Figure 1. We design the Dynamic Knowledge Adapter (DKA) with probabilistic calibration for the generalized few-shot segmentation task, which addresses the GFSS challenge by efficient finetuning and relabelling. In the training phase (a), the model learns the segmentation of base classes according to the labeled training sample. In the inference stage (b), we fine-tune the DKA and the novel-class classifier with support samples of novel classes, allowing the DKA to acquire information about novel categories. Next, we integrate base and novel-class classifiers, relabeling a training image subset through probabilistic calibration. Finally, we use the DKA and classifier, both updated with appropriate weights, to make predictions via probabilistic calibration.

1. Introduction

Semantic segmentation is a typical computer vision problem that involves taking images as input and transforming them into masks with highlighted regions of interest, where each pixel in the image is assigned a category based on the highlighted regions [22]. With the development of deep learning techniques, there are more and more excellent solutions for semantic segmentation tasks [1, 25, 30], both CNN-based models and transformer-based models. However, traditional semantic segmentation tasks often require a large number of pixel-level annotation samples, which may not be satisfied in practical applications, such as tasks on remote sensing satellite images [10] and medical images of rare diseases [26]. Therefore, Few-shot Semantic Segmentation (FSS) has received wide attention [12], i.e., learning novel categories with only a few labeled training samples (a.k.a. support set), with knowledge transferred from non-overlapping base classes where sufficient data is available for pretraining.

FSS defaults that the test images (a.k.a. query set) contain only categories from the support. However, a more practical setting is to consider both the support-set classes (i.e., novel classes) and the base classes, which gives rise to the Generalized Few-shot Semantic Segmentation (GFSS) [12] task. Specifically, GFSS requires segmenting out all categories in images during the inference stage, including base classes and novel classes, as shown in Fig.1. This

*Corresponding author.

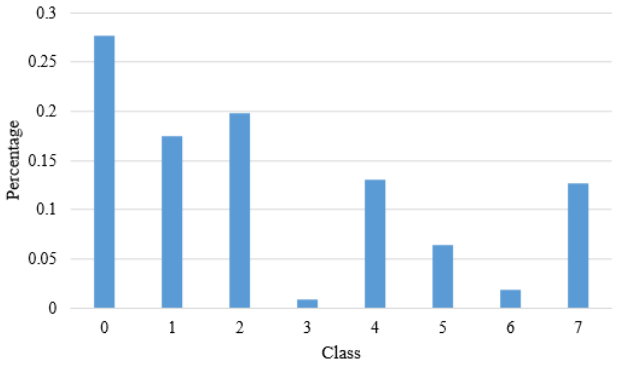

Figure 2. Proportion of all categories (including background classes) in the training stage. There is an imbalance between categories, especially the proportion of the background class is much higher than other categories.

task is even more challenging due to both the limited novel-class training samples and the catastrophic forgetting of base classes.

To handle these challenges, a baseline model [11] has been proposed to consist of a pair of encoder-decoder and a classifier that learns to classify the base classes (including the background) during training. To make the model have the ability to segment novel classes, it uses the support set to fine-tune a classifier for novel classes in the inference stage. Finally, the two classifiers are concatenated together to segment full classes in the query set images: the base classes (including the background) and the novel classes.

However, the performance of this baseline model on GFSS tasks often falls short, primarily due to three reasons: Firstly, due to the ineffectiveness of finetuned parameters, it lacks the adaptability to efficiently learn from novel classes with scarce data. Secondly, due to the conflict between base and novel classes, it lacks the stability to maintain knowledge of base classes, leading to the catastrophic forgetting problem. Thirdly, we summarize the number of samples in each category (including background class) in the training stage, as shown in Fig.2 where class 0 indicates the background class. We find (1) there is an imbalance between the categories, and (2) the proportion of the background class is much larger than other classes. These will cause the model to be biased towards more frequent classes, resulting in prediction bias. (3) The novel classes in the inference stage appear in the background class in the training stage, and the model will mistakenly classify some of the novel classes into the background class, which is called the overconfidence of the model. We refer to the above three phenomena as the probabilistic bias of the model.

To address the first problem, we design an auxiliary network (Dynamic Knowledge Adapter, DKA) between the decoder and the classifier to help the model better adapt to novel classes. In the inference fine-tuning phase, only parameters in this network are finetuned based on support set samples. For the second problem, to help the model maintain performance on the base class and mitigate catastrophic forgetting, we randomly sample some training images and *relabel* them: the base-class (including background class) classifier and the novel-class classifier are connected to obtain the full-class classifier, and the full-class segmentation is performed on the training images. In this process, the DKA and classifier are updated to alleviate the conflict between base and novel classes. For the third problem, to rectify the prediction bias caused by the imbalanced training data, we design a probabilistic calibration module in the inference stage to scale the prediction of infrequent classes.

In summary, our contribution is as follows:

- We design a Dynamic Knowledge Adapter (DKA) between the decoder and classifier to enhance the model's adaptability to learning novel classes.
- We use partial training samples to relabel and fine-tune the DKA and the full-class classifier to improve the model's stability to mitigate catastrophic forgetting of base classes.
- We design a probabilistic calibration module to alleviate the probabilistic bias of the model and further improve the segmentation accuracy of all classes.

2. Related Works

2.1. Semantic Segmentation

Semantic segmentation is a challenging task that involves accurately assigning labels to each pixel. The first framework developed for semantic segmentation was FCN [22], which replaces the last fully connected layer of a classification network with convolution layers. Encoder-decoder style approaches [1, 25, 30] have been adopted to refine the output in multiple steps and achieve per-pixel predictions. To improve the performance of semantic segmentation, techniques such as dilated convolution [4, 50] have been introduced to increase the receptive field. Context modeling architectures, including global pooling [20] and pyramid pooling [4, 48, 54, 55], have also played a crucial role in incorporating context information. Attention models [14, 25, 35, 42, 51, 53, 56] have shown effectiveness in capturing long-range relations within scenes. Recently, the effectiveness of vision transformers for semantic segmentation has also been demonstrated [6, 34, 46]. However, despite the success of these advanced segmentation frameworks, they face challenges when it comes to adapting to unseen classes without sufficient annotated data and require fine-tuning.

2.2. Few-Shot Learning

Few-shot learning is a machine learning approach that aims to recognize new classes with only a few labeled samples [24]. Existing few-shot learning methods can be categorized into three main groups, as described in [18]: finetuning-based, meta-based, and metric-based methods. Finetuning-based strategies [5, 7, 23, 28, 39, 49] involve a transfer learning process where the model is pre-trained on base classes and then fine-tuned on novel classes. Meta-based approaches [3, 9, 27, 29, 31, 36, 37] adopt a meta-learning paradigm to learn cross-task knowledge by optimizing the interaction between a meta-learner and base-learner. This enables the model to quickly adapt to novel datasets. Metric-based techniques [2, 17, 33, 38, 41, 52] focus on learning transferable representations and making predictions based on the distance between feature representations. This approach eliminates the need for fine-tuning during test time.

While the combination of a supervised model (for base classes) and a prototype-based approach (for novel classes) has been explored in low-shot visual recognition [10, 26], it is important to note that dense pixel labeling in semantic segmentation is distinct from image-level classification. In image-level classification, contextual information for each target is not taken into account.

2.3. Few-Shot Segmentation

Few-shot semantic segmentation (FSS) [32] aims to perform pixel-wise labeling for new classes with only a limited number of support examples. It primarily focuses on the 1-way scenario, where binary maps are generated for query images to identify the pixels belonging to the class labeled in the support images. Approaches such as [8, 43] adapt prototype learning for FSS by calculating cosine similarities between pixels and prototypes derived from the support images. ASR [19] learns multiple orthogonal prototypes on the base data to represent novel categories. Furthermore, assigning multiple prototypes to each class has shown promise in improving FSS models [16, 21, 44, 47].

2.4. Generalized Few-shot Semantic Segmentation

To address some of the limitations in few-shot semantic segmentation (FSS), a recent extension called generalized few-shot semantic segmentation (GFSS) was introduced [40]. GFSS approaches are designed to handle a single support set that contains images for each novel class, and they should be able to predict both base and novel classes in query images. Unlike standard FSS methods, GFSS models have no prior knowledge of the novel classes present in a query image. To tackle this challenge, CAPL [40] proposed a framework with two modules that dynamically adapt both base and novel prototypes. However, the presented results in CAPL are biased towards base classes, and this solution requires labeled base classes in the support samples.

Another model evaluated in the GFSS setting is the BAM model [15], initially proposed for FSS. The BAM model consists of two steps. First, a base learner is trained on base classes using the standard supervised learning paradigm, employing cross-entropy loss on the base training set. Then, a second meta-learning step is introduced, optimizing both the base-learner and a new meta-learner through episodic training. During inference, the output of the meta-learner is combined with the base-learner's output to make predictions on base classes and a single novel class. However, the limitation of the meta-learner being capable of distinguishing only background-foreground categories makes this method unsuitable for direct application to multi-class GFSS scenarios.

3. Backgroud

3.1. Notations

In the context of generalized few-shot semantic segmentation, we have a plentiful amount of annotated images for M base classes, denoted as $C^b = \{c_1, \cdots, c_M\}$, and only K labeled images per class for N novel classes, denoted as $C^n = \{c_{M+1}, \cdots, c_{M+N}\}$. Additionally, there is a background category, c_0, which represents pixels that do not belong to any target class. The objective of this task is to simultaneously differentiate between base and novel classes, as well as the background, resulting in a total of $M+N+1$ classes. In our paradigm, the training process consists of two phases. In the training phase, we utilize the images belonging to the base classes C^b to train our model. Subsequently, in the evaluation phase, we construct a support set S comprising N classes, where each class consists of K labeled images. We then fine-tune the model using the images from the support set S. Finally, we create a query set Q consisting of $M+N+1$ classes, where images are randomly selected from within these classes. We evaluate and test the performance of our model on this query set Q.

3.2. DIaM

Our method is modified from the baseline framework of GFSS proposed in [11]. We introduce DIaM first in this section. During the training stage, a segmentation model with an encoder f_ϕ and a classifier f_{θ_b} is trained on base classes C^b. At this stage, the model only can predict M+1 classes, i.e., the base classes and background. At the inference stage, the encoder is fixed, and the pre-trained classifier $\theta_b \in \mathbb{R}^{(M+1)\times d}$ is augmented with novel prototypes $\theta_n \in \mathbb{R}^{(N)\times d}$. The concatenation of $\theta = \theta_n + \theta_b$ forms the final classifier. We optimize the classifier θ for GFSS tasks. Note that d is the size of the feature space. The optimization

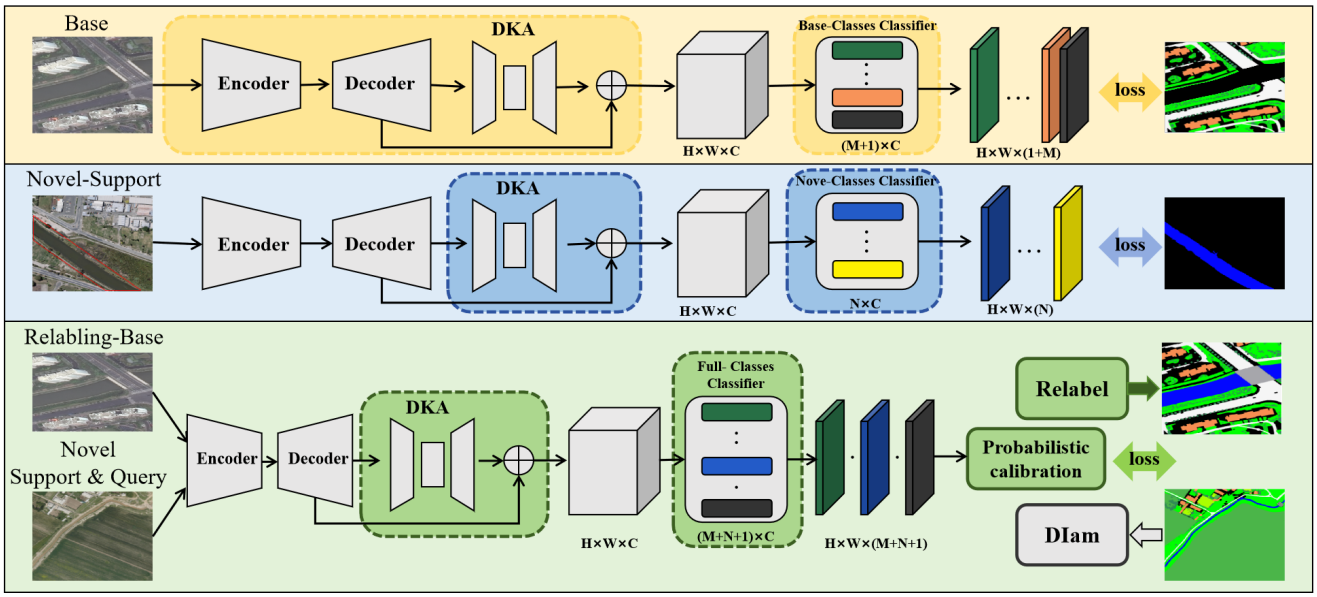

Figure 3. Our method framework: Our method builds upon the two-stage training paradigm (training-inference) to enhance the learning of novel classes. We introduce a Dynamic Knowledge Adapter (DKA) to the existing segmentation model. Initially, utilizing the base class training set, we train an encoder, decoder, DKA, and base-class classifier using cross-entropy loss. Then, fine-tuning of the DKA is performed based on the support set of novel classes, concurrently training a proficient novel-classes classifier. Concatenating the novel-classes and base-classes classifiers initializes a full-classes classifier. Subsequently, we randomly sample and relabel base class training samples, followed by another round of fine-tuning for both DKA and the full-classes classifier. Finally, leveraging support and query samples within the DIAM framework, we perform the final adjustments to the full-classes classifier. Throughout the fine-tuning process, we introduced a Probabilistic Calibration (PC) module to alleviate the probabilistic bias caused by data imbalance.

objective is based on the InfoMax framework:

$$\max_{\theta} I(X;P) = H(P) - H(P|X), \tag{1}$$

where X and P are random variables respectively associated with the pixel distribution and model's predictions.

The DIaM baseline is composed of three loss terms: $\mathcal{L}_{cond-ent}$, $\mathcal{L}_{marg-ent}$, and $\mathcal{L}_{KD}$. The conditional entropy term reads as:

$$\mathcal{L}_{cond-ent} = \alpha \sum_{i=1}^{|S|} H(y_i; \pi_S(p_i)) + H(p_{|S|+1}), \tag{2}$$

where α controls the reliance on the labeled support set.

To account for the misalignment between predictions p and labels y the model's predictions are projected as below, the j denotes the pixel index:

$$\pi_S(p_i)(j) = \left[\sum_{k=0}^{M} p_k, 0, \ldots, 0, p_{M+1}, \ldots, p_{M+N} \right]^{\mathrm{T}} \tag{3}$$

The marginal entropy term is calculated as:

$$\mathcal{L}_{marg-ent} = H(P;\Pi) = Cste - KL(\hat{p}||\Pi), \tag{4}$$

where $KL(\cdot||\cdot)$ denotes the Kullback-Leibler divergence. Π is estimated from the model's initial marginal distribution and re-updated during optimization.

The knowledge-distillation term is expressed as:

$$\mathcal{L}_{KD} = KL(\pi_{new2old}(p_{|S|+1})||p^{old}_{|S|+1}), \tag{5}$$

where the predictions p is projected as:

$$\pi_{new2old}(p)(j) = \left[p_0 + \sum_{i=1}^{N} p_{M+i}, p_1, p_2, \ldots, p_M \right]^{\mathrm{T}} \tag{6}$$

The final objective of DIaM is represented as:

$$\min_{\theta} \mathcal{L}_{DIaM} = \mathcal{L}_{cond-ent} - \mathcal{L}_{marg-ent} + \beta \mathcal{L}_{KD} \tag{7}$$

4. Method

To address the challenges presented by the Generalized Few-shot Semantic Segmentation (GFSS) task, our primary focus is on refining the inference stage within the existing two-stage training paradigm (training-inference). The overview of our framework is shown in Fig.3. Specifically, to better capture the distinctive features of novel classes with limited samples, we introduce a Dynamic Knowledge Adapter (DKA) module. During the inference stage, we fine-tune this module using samples from the support set of novel classes, while simultaneously training to obtain a robust classifier for novel classes.

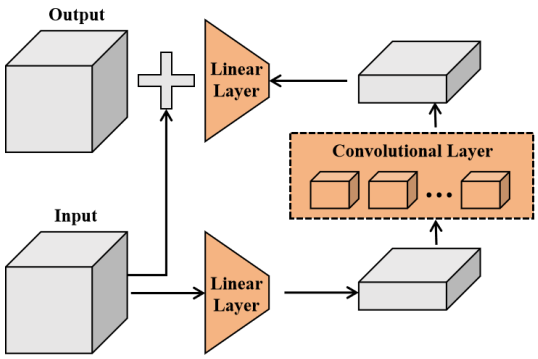

Figure 4. The structure of Dynamic Knowledge Adapter (DKA).

Subsequently, to mitigate catastrophic forgetting and rectify misclassifications of novel classes as background during training, we randomly selected a subset of training samples from the base classes and re-labeled them. The objective was to identify and correct instances where novel classes were erroneously labeled as background during training. This process facilitated fine-tuning of both the DKA and the full-class classifier.

Finally, to address the phenomenon of probabilistic bias occurring in the model's predictions, we integrate a Probabilistic Calibration (PC) module. This module serves to refine both the DKA and the full-class classifier while simultaneously adjusting prediction outcomes.

4.1. Training

Traditional segmentation models can be divided into base model f_ϕ (i.e. encoder-decoder) and a linear classifier $\theta_b \in \mathbb{R}^{(M+1)\times d}$. During the training process of GFSS, the model is trained using conventional cross-entropy loss based on training samples from the base categories. At this stage, the classifier can only predict one class among $1+M$ classes, namely the background class and the base classes.

The existing segmentation models demonstrate proficiency in tasks involving base class segmentation. However, their applicability to Generalized Few-shot Semantic Segmentation (GFSS) tasks is limited. On one hand, directly fine-tuning the model with a small subset of samples from novel classes often leads to catastrophic forgetting. Conversely, keeping the model fixed without fine-tuning restricts its performance on novel classes. To address these challenges inherent in both generalization settings, we propose integrating a Dynamic Knowledge Adapter (DKA) to facilitate fine-tuning on novel classes, thereby enhancing the model's capacity for learning novel classes while maintaining proficiency in base class segmentation.

Taking inspiration from the LoRA [13] architecture, we formulate the DKA f_{DKA} as a composite structure comprising a fully connected layer $W_{\text{in}} \in \mathbb{R}^{C\times r}$, a convolutional layer $W_{\text{tran}} \in \mathbb{R}^{r\times r\times 1\times 1}$, and a fully connected layer $W_{\text{out}} \in \mathbb{R}^{r\times C}$, as shown in Fig. 4. Here, C represents the number of feature channels, and r is the low-rank used for dimension reduction. The rationale behind the design of this structure is to compress essential information through dimension reduction, then apply 1×1 convolutions to assign different weights to this compressed information, and finally restore the original dimensions. This approach steers the model to focus on features with higher discriminative capability. This DKA module is positioned between the decoder and the classifier, engaging in the training phase for image segmentation of the base classes.

4.2. Inference

During the inference phase, we choose DIaM as our baseline. Firstly, to mitigate catastrophic forgetting and retain the model's ability to recognize base classes, we initially freeze the encoder-decoder f_ϕ and base-classes classifier θ_b. Subsequently, to enhance the model's capability to learn novel classes, we fine-tune the DKA using only the support set containing samples from the novel classes, simultaneously obtaining a classifier tailored to the novel classes $\theta_n \in \mathbb{R}^{d\times N}$. Next, to prevent the model from forgetting base classes while rectifying the misclassification of novel classes as background during training, we sample a portion of training samples from the base classes and relabel them for fine-tuning DKA and the full-classes classifier $\theta_{\text{full}} \in \mathbb{R}^{d\times M+N+1}$. Finally, we further fine-tune the full-classes classifiers θ_{full} using the DIaM framework, leveraging the support set of labeled data for the novel classes and unlabeled query set. During the fine-tuning process, to alleviate the probabilistic bias of the model, we propose a probabilistic calibration module to assist the model in achieving better generalization, thereby improving its ability to distinguish among base, novel, and background classes.

4.2.1 Finetuning DKA

Firstly, we fine-tune DKA using samples from the novel classes in the Support set, thereby obtaining a classifier θ_n specialized in classifying the novel classes. Specifically, as we focus on enhancing the model's ability to learn novel classes at this stage, we initially process the novel-class samples from the support set. We mark their backgrounds as "ignore" and obtain a new label $y_{\text{new}} \in [0,1]^{N\times(H\times W)}$, ensuring that the classifier only needs to distinguish among the novel classes. Next, we compute the prototypes for novel classes to initialize the novel-classes classifier θ_n. Finally, we use Eq. 8 to calculate the loss for fine-tuning DKA f_{DKA} and the novel-classes classifier θ_n.

$$\mathcal{L}_{novel\_ft} = CE(P, y_{new}), \tag{8}$$

Here, $CE(\cdot)$ denotes the cross-entropy loss function for segmentation tasks, where $P \in \mathbb{R}^{N\times H\times W}$ represents the

predicted values outputted by the model and can be calculated using the following equation:

$$P = softmax(f_{DKA}(f_\phi(X))\theta_n) \tag{9}$$

Where X represents samples from the support set.

4.2.2 Relabling Training Samples

After fine-tuning DKA with the support set containing novel classes, there is a risk of the model being biased toward the novel classes, potentially harming the performance of the base classes. Simultaneously, the labels in the training set might mark potential novel classes as background, leading to the possibility that new classes are mistakenly identified as background during final predictions. Hence, we employ a relabeling strategy that not only recovers some segmentation ability for the base class but also alleviates the situation where novel classes are mistakenly recognized as background. Specifically, we sample $k \times 10$ images from the training set (where k is the total number of categories) and have the model predict their masks to serve as pseudo masks. Then iteratively optimize DKA and the classifier through n iterations.

4.3. Results

Specifically, we first randomly sample a training subset D^{sub}_{train} containing N samples belonging to the base classes C^b from the training dataset D_{train}. Next, we concatenate the classifiers θ_b obtained during the training phase with the fine-tuned classifier θ_n, resulting in the initialization of a full-classes classifier θ_{full}. Next, we treat the samples from D^{sub}_{train} as unlabeled query samples Then based on the DIaM framework, we utilize Eq.7 to calculate the loss for both the support set S and the training subset D^{sub}_{train} and then fine-tune DKA and the full-classes classifier θ_{full} accordingly. Finally, we further fine-tune the full-classes classifier θ_{full} based on the DIaM framework, using the support set S and the actual query set Q.

4.3.1 Calibrating Probabilistic

To address the probabilistic bias introduced by the task setup, we introduce a probabilistic calibration module. This module consists of three parts, each designed to address one of the three reasons leading to the probabilistic bias.

Firstly, to alleviate the probabilistic bias caused by the imbalance of samples between classes, we introduce a temperature coefficient during the fine-tuning process. Specifically, as shown in Eq, we manipulate the logit values of the model output to ensure a smoother probability distribution.

$$P = softmax(\tau f_{DKA}(f_\phi(X))\theta_n) \tag{10}$$

where τ is a hyper-parameter.

Table 1. MIoU compared to baseline methods. We have significant advantages in both base classes, novel classes, and the final weighted result.

Method	Base	Novel	Weighted average
Baseline	29.89401	3.15314	13.84949
Ours	**42.00045**	**20.51581**	**29.10967**

Furthermore, to mitigate the model's bias towards the background, we propose utilizing the model's predictions of the background during fine-tuning to identify regions where the model is prone to misclassification and force the model to focus on these error-prone regions. Specifically, we first set a threshold β to locate the areas where the model makes classification errors. Then, by calculating the probability of each pixel being classified as background by the model and comparing it with the threshold, we obtain a mask $M \in [0,1]^{H\times W}$ representing the regions where the model is prone to misclassification. Subsequently, we set the classification probabilities of all classes in these error-prone regions to the same value γ, thereby compelling the model to focus on these areas. β and γ are hyperparameters. The calculation process can be represented as:

$$P_{new} = (1 - M) \odot P + \gamma M \tag{11}$$

Finally, during the training phase, backgrounds may include new classes, leading the model to mistakenly classify new classes as background during the inference phase. Based on this, we propose a post-processing method for the prediction of query samples. Specifically, we argue that due to the model's bias towards the background, the model's prediction of a pixel as background is unreliable. Thus, we introduce a threshold α to filter out low-confidence background predictions using mask $M_p \in [0,1]^{H\times W}$. This implies that if the model is not confident in predicting a pixel as background, we manually set its probability of being predicted as background to a small fixed value η. α and η are hyperparameters.

5. Experiments

5.1. Datasets

The OpenEarthMap few-shot learning challenge dataset is derived from the original OpenEarthMap benchmark dataset[45] for remote sensing image semantic segmentation. This challenge dataset comprises only 408 samples, which is a subset of the larger benchmark dataset. The challenge dataset expands the original 8 semantic classes of OpenEarthMap to 15 classes. It is divided into three disjointed sets: train_base_class, val_novel_class, and test_novel_class, with a split ratio of 7:4:4, respectively. Out of the 408 samples, 258 are allocated for the trainset, 50 for

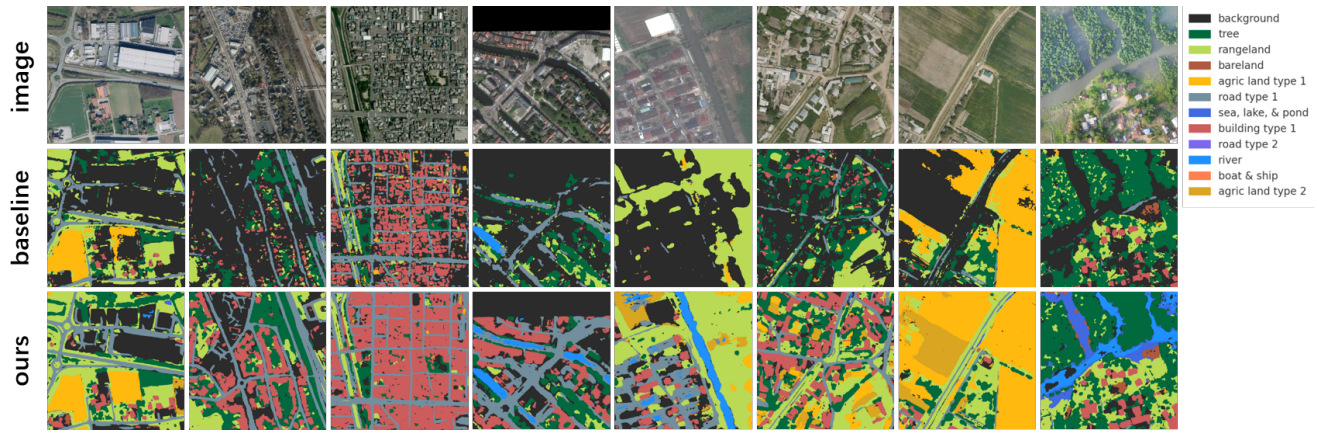

Figure 5. Visual experiments show that our method effectively distinguishes between base classes, novel classes, and backgrounds.

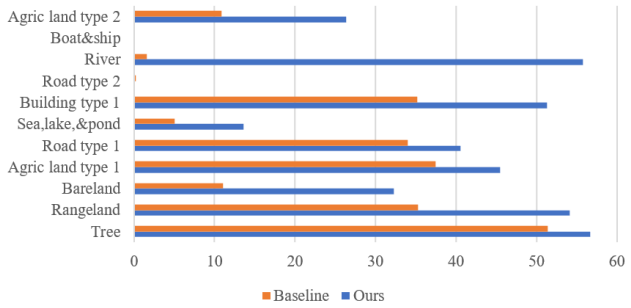

Figure 6. Miou of each class. Our method performs better on almost every class.

the valset, and 100 for the testset. The trainset is intended for pre-training a backbone network and contains only the images and labels from the train_base_class split. Both the valset and the testset are composed of a support set and a query set. The valset includes images and labels from the val_novel_class split, while the testset includes images and labels from the test_novel_class split.

5.2. Comparison With Baseline.

As described in Section 4, our final scheme includes fine-tuning the Dynamic Knowledge Adapter (DKA), relabelling training samples, and calibration probabilities. The final performance is shown in Tab.1. Compared with the baseline method, we can see that we have improved the segmentation performance of all classes, that is, on both base classes and novel classes. Specifically, Fig.6 shows the respective performance of eleven classes, and our method helps to segment better on almost all classes.

5.3. Visualization.

As shown in Fig.5, we visualize the predicted results of query samples. It can be observed that our approach not only improves the learning of both base and novel classes but also effectively alleviates the phenomenon of misclassifying other semantic categories as background.

5.4. Ablation Study.

As shown in Tab.2, we empirically demonstrated the effectiveness of the three modules through ablation experiments. Firstly, compared to the baseline method, introducing fine-tuning of Dynamic Knowledge Adapter (DKA) improves the model's ability to learn new classes. Next, using the re-labeling strategy can restore the model's ability to learn base classes while enhancing the model's ability to distinguish between novel, base, and background classes. Finally, the probabilistic calibration module effectively alleviates probability bias, thereby improving the model's ability to learn both new and base classes.

Table 2. Ablation Study. Our method's three modules are designed to effectively enhance the model's ability to learn both novel and base classes.

DKA	Relabeling	Probabilistic Calibration	base	novel	Weighted Average
x	x	x	29.89	3.15	13.85
y	x	x	37.20	8.49	19.97
y	y	x	37.93	10.93	21.73
y	y	y	42.00	20.52	29.11

5.5. Sensitivity Analysis

We have modified the threshold β and set it to range from 0 to 1 with a step size of 0.1. The mIoU (mean Intersection over Union) values for both base classes and novel classes at different threshold values are shown in 7. From the trend observed, we can see that after a threshold value of 0.2, the mIoU values start to decrease. Therefore, we have chosen a threshold value of 0.15 as the final value.

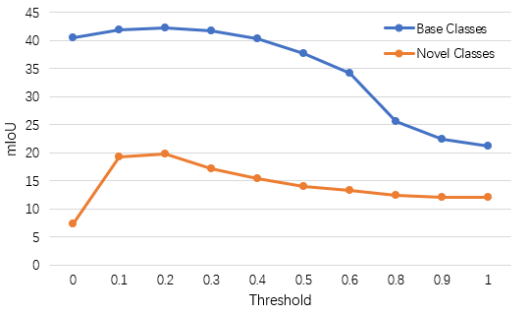

Figure 7. mIoU under different thresholds.

6. Conclusion

To handle the challenging GFSS problem, we design the Dynamic Knowledge Adapter (DKA), which handles the adaptability by finetuning efficient parameters, addresses the instability problem by sample relabeling, and rectifies the biased prediction by probabilistic calibration. Extensive experiments validate the rationale and effectiveness of our methods.

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 1, 2

[2] Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood, and Leonid Sigal. Improved few-shot visual classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14493–14502, 2020. 3

[3] Luca Bertinetto, Joao F Henriques, Philip HS Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. *arXiv preprint arXiv:1805.08136*, 2018. 3

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 2

[5] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019. 3

[6] Bowen Cheng, Alex Schwing, and Alexander Kirillov. Per-pixel classification is not all you need for semantic segmentation. *Advances in neural information processing systems*, 34:17864–17875, 2021. 2

[7] Guneet S Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A baseline for few-shot image classification. *arXiv preprint arXiv:1909.02729*, 2019. 3

[8] Nanqing Dong and Eric P Xing. Few-shot semantic segmentation with prototype learning. In *BMVC*, page 4, 2018. 3

[9] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017. 3

[10] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4367–4375, 2018. 1, 3

[11] Sina Hajimiri, Malik Boudiaf, Ismail Ben Ayed, and Jose Dolz. A strong baseline for generalized few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11269–11278, 2023. 2, 3

[12] Sina Hajimiri, Malik Boudiaf, Ismail Ben Ayed, and Jose Dolz. A strong baseline for generalized few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11269–11278, 2023. 1

[13] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 5

[14] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross attention for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 603–612, 2019. 2

[15] Chunbo Lang, Gong Cheng, Binfei Tu, and Junwei Han. Learning what not to segment: A new perspective on few-shot segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8057–8067, 2022. 3

[16] Gen Li, Varun Jampani, Laura Sevilla-Lara, Deqing Sun, Jonghyun Kim, and Joongkyu Kim. Adaptive prototype learning and allocation for few-shot segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8334–8343, 2021. 3

[17] Wenbin Li, Lei Wang, Jinglin Xu, Jing Huo, Yang Gao, and Jiebo Luo. Revisiting local descriptor based image-to-class measure for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7260–7268, 2019. 3

[18] Wenbin Li, Ziyi Wang, Xuesong Yang, Chuanqi Dong, Pinzhuo Tian, Tiexin Qin, Jing Huo, Yinghuan Shi, Lei Wang, Yang Gao, et al. Libfewshot: A comprehensive library for few-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 3

[19] Binghao Liu, Yao Ding, Jianbin Jiao, Xiangyang Ji, and Qixiang Ye. Anti-aliasing semantic reconstruction for few-shot semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9747–9756, 2021. 3

[20] Wei Liu, Andrew Rabinovich, and Alexander C Berg. Parsenet: Looking wider to see better. *arXiv preprint arXiv:1506.04579*, 2015. 2

[21] Yongfei Liu, Xiangyi Zhang, Songyang Zhang, and Xuming He. Part-aware prototype network for few-shot semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 142–158. Springer, 2020. 3

[22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1, 2

[23] Puneet Mangla, Nupur Kumari, Abhishek Sinha, Mayank Singh, Balaji Krishnamurthy, and Vineeth N Balasubramanian. Charting the right manifold: Manifold mixup for few-shot learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2218–2227, 2020. 3

[24] Erik G Miller, Nicholas E Matsakis, and Paul A Viola. Learning from one example through shared densities on transforms. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, pages 464–471. IEEE, 2000. 3

[25] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015. 1, 2

[26] Hang Qi, Matthew Brown, and David G Lowe. Low-shot learning with imprinted weights. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5822–5830, 2018. 1, 3

[27] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. *arXiv preprint arXiv:1909.09157*, 2019. 3

[28] Jathushan Rajasegaran, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Mubarak Shah. Self-supervised knowledge distillation for few-shot learning. *arXiv preprint arXiv:2006.09785*, 2020. 3

[29] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International conference on learning representations*, 2017. 3

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 1, 2

[31] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850. PMLR, 2016. 3

[32] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots. One-shot learning for semantic segmentation. *arXiv preprint arXiv:1709.03410*, 2017. 3

[33] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017. 3

[34] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7262–7272, 2021. 2

[35] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5693–5703, 2019. 2

[36] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 403–412, 2019. 3

[37] Qianru Sun, Yaoyao Liu, Zhaozheng Chen, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning through hard tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1443–1456, 2020. 3

[38] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018. 3

[39] Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B Tenenbaum, and Phillip Isola. Rethinking few-shot image classification: a good embedding is all you need? In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, pages 266–282. Springer, 2020. 3

[40] Zhuotao Tian, Xin Lai, Li Jiang, Shu Liu, Michelle Shu, Hengshuang Zhao, and Jiaya Jia. Generalized few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11563–11572, 2022. 3

[41] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016. 3

[42] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3349–3364, 2020. 2

[43] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *proceedings of the IEEE/CVF international conference on computer vision*, pages 9197–9206, 2019. 3

[44] Zhonghua Wu, Xiangxi Shi, Guosheng Lin, and Jianfei Cai. Learning meta-class memory for few-shot semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 517–526, 2021. 3

[45] Junshi Xia, Naoto Yokoya, Bruno Adriano, and Clifford Broni-Bediako. Openearthmap: A benchmark dataset for global high-resolution land cover mapping. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6254–6264, 2023. 6

[46] 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. 2

[47] Boyu Yang, Chang Liu, Bohao Li, Jianbin Jiao, and Qixiang Ye. Prototype mixture models for few-shot semantic

segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 763–778. Springer, 2020. 3

[48] Maoke Yang, Kun Yu, Chi Zhang, Zhiwei Li, and Kuiyuan Yang. Denseaspp for semantic segmentation in street scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3684–3692, 2018. 2

[49] Shuo Yang, Lu Liu, and Min Xu. Free lunch for few-shot learning: Distribution calibration. *arXiv preprint arXiv:2101.06395*, 2021. 3

[50] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015. 2

[51] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 173–190. Springer, 2020. 2

[52] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. Deepemd: Few-shot image classification with differentiable earth mover's distance and structured classifiers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12203–12213, 2020. 3

[53] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. Context encoding for semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7151–7160, 2018. 2

[54] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 2

[55] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. Icnet for real-time semantic segmentation on high-resolution images. In *Proceedings of the European conference on computer vision (ECCV)*, pages 405–420, 2018. 2

[56] Hengshuang Zhao, Yi Zhang, Shu Liu, Jianping Shi, Chen Change Loy, Dahua Lin, and Jiaya Jia. Psanet: Point-wise spatial attention network for scene parsing. In *Proceedings of the European conference on computer vision (ECCV)*, pages 267–283, 2018. 2