Domain-Rectifying Adapter for Cross-Domain Few-Shot Segmentation

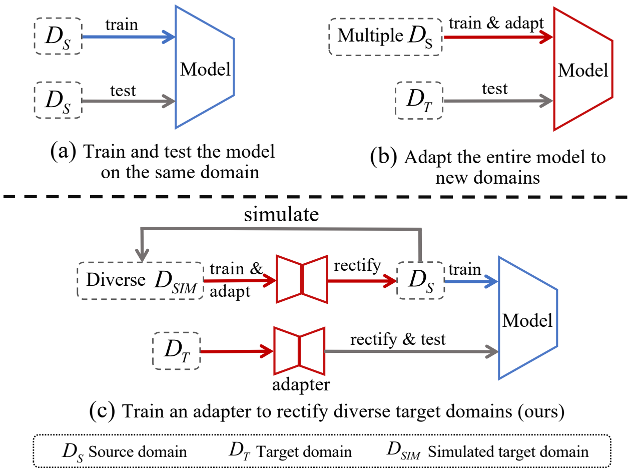
Jiapeng Su[1]*, Qi Fan[2]*, Wenjie Pei[1]†, Guangming Lu[1], Fanglin Chen[1]
[1]Harbin Institute of Technology, Shenzhen [2]Nanjing University
MattSu@163.com, fanqics@gmail.com, wenjiecoder@outlook.com,
luguangm@hit.edu.cn, chenfanglin@hit.edu.cn

Abstract

Few-shot semantic segmentation (FSS) has achieved great success on segmenting objects of novel classes, supported by only a few annotated samples. However, existing FSS methods often underperform in the presence of domain shifts, especially when encountering new domain styles that are unseen during training. It is suboptimal to directly adapt or generalize the entire model to new domains in the few-shot scenario. Instead, our key idea is to adapt a small adapter for rectifying diverse target domain styles to the source domain. Consequently, the rectified target domain features can fittingly benefit from the well-optimized source domain segmentation model, which is intently trained on sufficient source domain data. Training domain-rectifying adapter requires sufficiently diverse target domains. We thus propose a novel local-global style perturbation method to simulate diverse potential target domains by perturbating the feature channel statistics of the individual images and collective statistics of the entire source domain, respectively. Additionally, we propose a cyclic domain alignment module to facilitate the adapter effectively rectifying domains using a reverse domain rectification supervision. The adapter is trained to rectify the image features from diverse synthesized target domains to align with the source domain. During testing on target domains, we start by rectifying the image features and then conduct few-shot segmentation on the domain-rectified features. Extensive experiments demonstrate the effectiveness of our method, achieving promising results on cross-domain few-shot semantic segmentation tasks. Our code is available at https://github.com/Matt-Su/DR-Adapter.

1. Introduction

Benefiting from well-established large-scale datasets [1, 24], numerous semantic segmentation methods [5, 30, 36] have undergone rapid development in recent years. However, obtaining enough labeled data is still a challenging and resource-intensive process, particularly for tasks like instance and semantic segmentation. Unlike machine learning approaches, human capacity to recognize novel concepts from limited examples fuels considerable research interest. Hence, few-shot segmentation (FSS) is proposed to meet this challenge, developing a network that generalizes to new domains with limited annotated data.



(a) Train and test the model on the same domain

(b) Adapt the entire model to new domains

(c) Train an adapter to rectify diverse target domains (ours)

Figure 1. The comparison of our method with other approaches. (a) Traditional few-shot segmentation (FSS) methods train and test the model on the same domain. (b) Most domain generalization (DG) methods leverages multiple source domains to train and adapt the large-parameter model simultaneously. (c) In contrast to conventional DG methods, we propose using a lightweight adapter as a substitute. This adapter is designed to adapt to various domain data, thereby decoupling domain adaptation from the source domain training process.

Nonetheless, most existing few-shot segmentation methods [12, 21, 31, 32, 34, 41, 51, 53, 54] often exhibit subpar performance when confronted with domain shifts [25, 37, 50]. The cross-domain few-shot segmentation (CD-FSS) is thus proposed for generalizing few-shot segmentation models from the source domain to other domains [20, 47]. CD-

*Both authors contributed equally.
†Corresponding author.

FSS trains the model solely on the source domain, and generalizes the trained model to segment object of novel classes in a separate target domain, supported by few-shot samples.

Domain adaptation(DA) and domain generalization(DG) are closely related to cross-domain few-shot segmentation. However, DA methods require unlabeled training data from the target domain. DG aims to generalize models trained in the source domain to various unseen domains, often requiring extensive training data from multiple source domains. Consequently, DA/DG methods typically adapt the entire model to new domains, leveraging substantial domain-specific data. Similarly, most existing CD-FSS methods adapt the entire model to target domains. However, in few-shot learning, the scarcity of training data can lead to overfitting when directly adapting the entire model. Rather than generalizing the entire model, our approach focuses on adapting a compact adapter to rectify diverse target domain features to align with the source domain. Once rectified to the source domain, target domain features can effectively utilize the well-trained source domain segmentation model, which is intently optimized using extensive source domain data. Figure 1 shows the difference among our method and conventional FSS and domain generalization methods.

Training a domain-rectifying adapter requires extensive data of diverse target domains. The straightforward feature-level domain synthesis method can effectively generate diverse potential target domains by randomly perturbing feature channel statistics. We can diversify the synthesized domain styles by increasing the magnitude of perturbation noises. However, as shown in Figure 2, some feature channels in individual images exhibit very low activation values. These small feature channel statistic values result in the corresponding channels suffering from limited style synthesis. Merely increasing the perturbation noises may lead to model collapse, where highly activated channels are excessively perturbed. Consequently, we propose a novel local-global style perturbation method to generate diverse potential target domain styles. Our local style perturbation module generates new domains by perturbing the feature channel statistics of individual images, similar to DG methods. Our global style perturbation module effectively diversifies the synthesized styles by leveraging the collective feature statistics of the entire source domain. Dataset-level feature statistics are estimated through momentum updating on the entire source domain dataset. Our local and global style perturbation modules collaboratively generate diverse and meaningful domain styles.

The perturbed feature channel statistics represent diverse potential styles, which are then input into the adapter to train the domain-rectifying adapter. The adapter predicts two rectification vectors to rectify the perturbed feature channel statistics to their original values. Additionally, we propose a cyclic domain alignment module to assist the adapter in learning to effectively rectify diverse domain styles to align with the source domain. Once rectified, the feature channel statistics will collaborate with the normalized feature map to train the segmentation model. During inference, we can directly use the domain-rectifying adapter to align the image features with the source domain and then input them into the well-trained source domain model for segmentation. In summary, our contributions are:

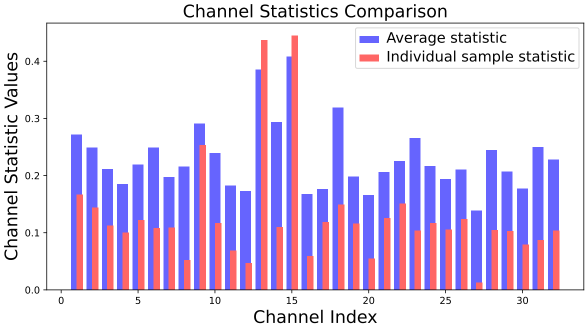


Figure 2. We show the feature channel statistic of an individual sample's statistic and the average statistic across the dataset on the pretrained backbone at stage 1. The average statistics exhibit a smoother profile compared to that of an individual sample, allowing for the application of more substantial noise to the feature with the smoother statistics.

- We introduce a novel domain-rectifying method for cross-domain few-shot segmentation, employing a compact adapter to align diverse target domain features with the source domain, mitigating overfitting in limited training data scenarios.
- We propose a unique local-global style perturbation module that generates diverse target domain styles by perturbing feature channel statistics at both local and global scales, enhancing model adaptability to various target domains.
- To enhance domain adaptation, we introduce a cyclic domain alignment loss that helps the domain-rectifying adapter align diverse domain styles with the source domain.

2. Related Work

2.1. Few-Shot Segmentation

Few-shot semantic segmentation [26–29, 41, 46, 49, 53, 57], using a limited number of labeled support images, predicts dense masks for query images. Previous methods primarily adopted a metric-based paradigm [8], improved in various ways, and fell into two main categories: prototype-based and matching-based approaches. Motivated by PrototypicalNet [40], prototype-based methods extract prototypes from support images to guide query object segmentation. Most studies concentrate on effectively utilizing limited support images to obtain more representative prototypes. Recent studies [52, 54] emphasize that a single prototype often fails to represent an entire object adequately. To address this, methods such as ASGNet [21] and PRMMs [41] explore using multiple prototypes to represent the overall target.

On the other hand, matching-based methods [31, 32, 43] concatenate support and query features, subsequently inputting the concatenated feature map into CNN or transformer networks. This process explores the dense correspondence between query images and support prototypes. Recently, researches [35, 39] has focused on leveraging pixel-to-pixel similarity maps for effective support prototype generation and query feature enhancement.

2.2. Domain Generalization

Domain Generalization (DG) targets at generalizing models to diverse target domains, particularly when target domain data is inaccessible during training. Existing domain generalization methods fall into two categories: learning domain-invariant feature representations from multiple source domains [9, 14, 33, 45] and generating diverse samples via data or feature augmentation [4, 38, 44, 58]. The core idea of learning domain-invariant features is to leverage various source domains to learn a robust feature representation. Data or feature augmentation aims to increase the diversity of training samples to simulate diverse new domains.

Domain generalization is particularly challenging in few-shot settings, as the target domain substantially differs from the source domains in both domain style and class content. Unlike popular DG methods generalizing the entire model, we train a small adapter to rectify the target domain data into the source domain style for model generalization.

2.3. Cross-domain Few-Shot segmentation

Recently, cross-domain few-shot segmentation has received increasing attention. PATNet [20] proposes a feature transformation layer to map query and support features from any domain into a domain-agnostic feature space. RestNet [17] addresses the intra-domain knowledge preservation problem in CD-FSS. RD [47] employs a memory bank to restore the meta-knowledge of the source domain to augment the target domain data. Unlike previous CD-FSS methhods, our method directly learns two rectification parameters for effective domain adaptation, eliminating the needs of restoring source domain styles.

3. Methodology

Problem Setting Cross-Domain Few-Shot Segmentation (CD-FSS) aims to apply the source domain trained few-shot segmentation models to diverse target domains. The CD-FSS model is typically trained using episode-based meta-learning paradigm [11, 43]. The training and testing data both consist of thousands of randomly sampled episodes, including K support samples and one query image. The model first extracts the support prototype and query feature from each training episode, and then performs pixel-wise feature matching between the support prototype and query feature to predict the query mask. The support prototype is typically a feature vector aggregating the object features of all support images. Once trained, the model is directly applied to various target domains.

Method Overview Our key idea is to train a adapter to rectify diverse target domain styles to the source domain, and leverage the well-trained source domain segmentation model to process the rectified target domain features for accurate few-shot segmentation. The crux is to align diverse potential target domain distributions to the source domain distribution. To train the domain-rectifying adapter, we thus synthesize various target domain styles by perturbing the feature channel statistics of the source domain training images. And the adapter is trained to rectify the synthesized feature styles to the source domain style. During inference, the adapter can be directly applied on the target domain features to rectify their domain styles, and the subsequent segmentation model can process the rectified support and query features for few-shot segmentation. he overall framework of our approach is illustrated in Figure 3.

3.1. Local Domain Perturbation

Previous works [13, 59] show that perturbing feature channel statistics can effectively synthesize diverse domain styles and meanwhile preserves the image contents. We thus synthesize various domain styles by injecting gaussian noises into feature channel statistics of source domain images.

Given a feature map $F_o \in \mathcal{R}^{B\times C\times H\times W}$, we first compute the feature channel statistics, *i.e.*, mean μ_o and variance σ_o along each channel dimension:

$$\mu_o(F_o) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} F_o, \tag{1}$$

$$\sigma_o(F_o) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} (F_o - \mu_o(F_o))^2 + \epsilon}, \tag{2}$$

where $\mu_o, \sigma_o \in \mathcal{R}^{B\times C}$, ϵ is a small constant for numerical stability, B, C, H, and W represent the batch size, channel dimension, height, and width of the feature map.

Then, we leverage two perturbation factors α and β to control the gaussian noise injection process for μ_o and σ_o. The noise vectors, sharing the same dimension as μ_o and σ_o, are used to compute the perturbed mean μ_p and variance σ_p:

$$\begin{aligned} \mu_p &= (1+\alpha)\mu_o, \\ \sigma_p &= (1+\beta)\sigma_o. \end{aligned} \tag{3}$$

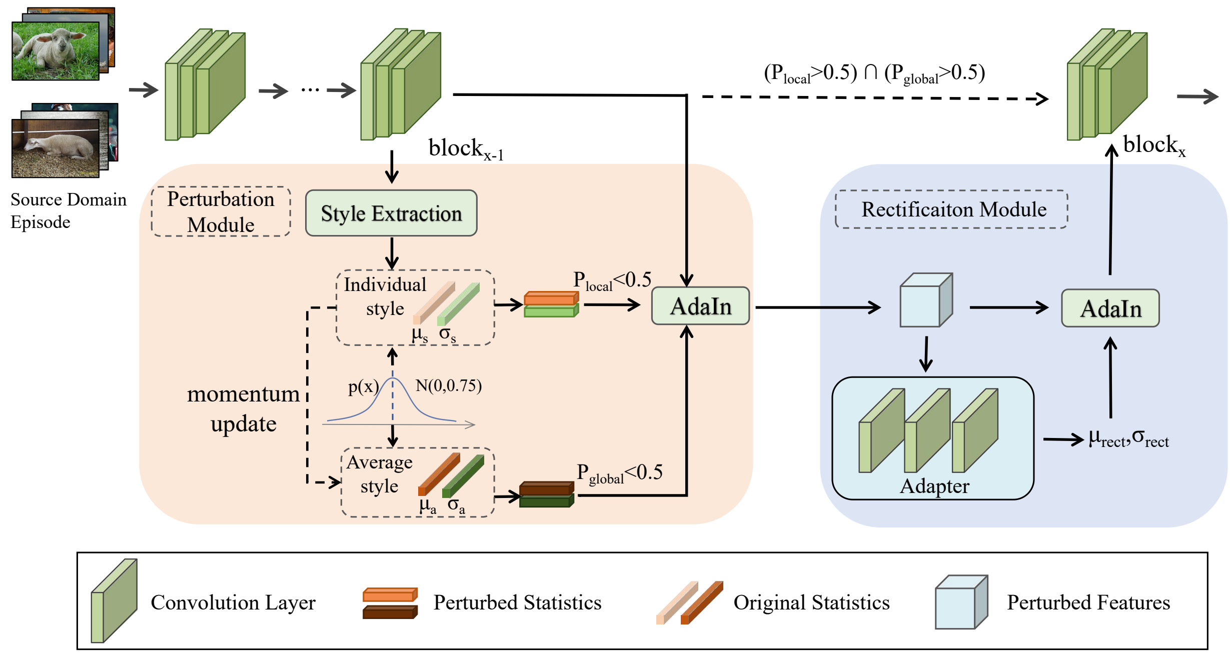


Figure 3. Overview of our cross-domain few-shot segmentation approach. Our method consists of two modules: a feature perturbation module and a feature rectification module. The former is used to generate simulated domain features, while the latter trains the adapter by restoring the features to their original states. During the perturbation process, we employ both local and global perturbations, controlled by two different probabilities P to decide if a feature is perturbed. Note that when both probabilities exceed 0.5, the entire backbone undergoes standard training. During testing, we treat target domain features as perturbed features and directly rectify them using the adapter.

We can obtain the perturbed feature map F_p by replacing the feature channel statistics $\{\mu_o, \sigma_o\}$ of the original feature map F_o with the perturbed channel statistics $\{\mu_p, \sigma_p\}$ using the Adaptive Instance Normalization formula [16]:

$$F_p = \sigma_p \frac{F_o - \mu_o}{\sigma_o} + \mu_p. \tag{4}$$

Within each episode, the support and query features share the same perturbation factors. The above equations can be further simplified to the following expression:

$$F_p = (1+\beta)F_o + (\alpha - \beta)\mu_o. \tag{5}$$

We call this feature channel statistic perturbation method as local domain perturbation, as it is enabled on individual images with probability P_{local}.

3.2. Global Domain Perturbation

We need to bound the local domain perturbation to prevent potential training collapse caused by the aggressive perturbation noises. However, insufficient domain perturbation may lead the domain-rectifying adapter to underperform when encountering new domain styles. The local domain perturbation method is trapped in the stability and performance dilemma. We thus propose a novel global domain perturbation by leveraging the global style statistics of the entire dataset to facilitate the domain style synthesis. The dataset's global style statistics exhibit better perturbation stability when leveraging aggressive perturbation noises to synthesize meaningful target domain styles for sufficient style diversity.

We first compute the feature channel statistics μ_o for individual images and then progressively update the global style statistics through momentum updating:

$$\mu_{\text{datum}} = \lambda \mu_{\text{datum}} + (1-\lambda)\mu_o, \tag{6}$$

where λ is the momentum updating factor. Then we can perform the global domain perturbation by replacing the image feature channel statistics μ_o in equation 5 with the global style statistics μ_{datum}. This global domain perturbation is randomly enabled with probability P_{global}.

3.3. Domain Rectification Module

The domain rectification module leverages an domain-rectifying adapter to rectify the target domain feature channel statistics to the source domain. The adapter takes as input the perturbed features and predicts two rectification vectors $\{\alpha_{rect}, \beta_{rect}\}$ to rectify the feature channel statistics of the perturbed feature map F_p as the rectified feature channel statistics $\{\mu_{rect}, \sigma_{rect}\}$.

$$\begin{aligned} \mu_{rect} &= (1+\alpha_{rect})\mu_p, \\ \sigma_{rect} &= (1+\beta_{rect})\sigma_p. \end{aligned} \tag{7}$$

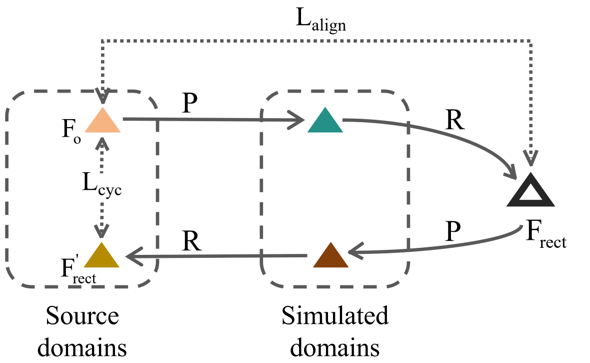


Figure 4. The process of cycle alignment, where 'P' denotes perturbation and 'R' stands for rectification.

Then we leverage the AdaIN function to generate the rectified feature map F_{rect} based on the perturbed feature map F_p and the rectified feature channel statistics $\{\mu_{rect}, \sigma_{rect}\}$:

$$\mathrm{F}_{\text{rect}} = (1 + \beta_{rect})\, \sigma_p \frac{\mathrm{F}_{\mathrm{p}} - \mu_{\mathrm{p}}}{\sigma_p} + (1 + \alpha_{rect})\, \mu_p, \tag{8}$$

which can be further simplified as:

$$\mathrm{F}_{\text{rect}} = (1 + \beta_{rect})\, \mathrm{F}_{\mathrm{p}} + (\alpha_{rect} - \beta_{rect})\, \mu_p. \tag{9}$$

We expect the adapter can adaptively predict the rectification factors $\{\alpha_{rect}, \beta_{rect}\}$ to rectify the perturbed features corresponding to diverse potential target domains. Consequently, during inference, we can leverage the adapter to rectify the target domain features to the source domain, and the rectified features can fittingly benefit from the well-trained source domain model for satisfactory few-shot segmentation results.

3.4. Cyclic Domain Alignment

Our goal is enabling the adapter to rectify the perturbed features back to the source domain space. Insufficient supervision during this process may lead the adapter to rectify the features into an unknown space. Therefore, in addition to utilizing the standard Binary Cross-Entropy (BCE) loss for supervision, we propose incorporation of a cyclic alignment loss to constrain the adapter.

After obtaining the rectified feature F_{rect}, we further perturb the F_{rect} with the same noise α and β to get a new perturbed feature F_{rect}^{p}. This perturbed image F_{rect}^{p} is then input into the adapter for a reverse rectification, resulting in $F_{rect}^{'}$. If the adapter can map features back to the source domain space, the style of $F_{rect}^{'}$ should closely match that of F_o. The cycle process is shown in figure 4. Consequently, we align the statistics between original feature and the cyclically rectified feature:

$$\begin{aligned} L_{\text{cyc}} = \frac{1}{C} \sum_c & (|\mu(F_o) - \mu(F'_{\text{rect}})| \\ & + |\sigma(F_o) - \sigma(F'_{\text{rect}})|) . \end{aligned} \tag{10}$$

We add constraint to the statistics between F_o and F_{rect}:

$$\begin{aligned} L_{\text{align}} = \frac{1}{C} \sum_c & (|\mu(F_o) - \mu(F_{\text{rect}})| \\ & + |\sigma(F_o) - \sigma(F_{\text{rect}})|) . \end{aligned} \tag{11}$$

We optimize the model with the final loss L:

$$L = L_{BCE} + L_{cyc} + L_{align} \tag{12}$$

4. Experiments

4.1. Datasets

Following [20], we validate our methods on the cross-domain few-shot segmentation (CD-FSS) benchmark. This benchmark includes images and pixel-level annotations from the FSS-1000 [22], DeepGlobe [7], ISIC2018 [6, 42], and Chest X-ray datasets [3, 18]. These datasets range from natural to medical images, providing sufficient domain diversity. We train models on the natural image dataset PASCAL VOC 2012 [10] with SBD [15] augmentation and evaluate models on the CD-FSS benchmark.

FSS-1000 [22] is a dataset designed for few-shot segmentation, containing 1,000 different categories of natural objects and scenes, with each category comprising 10 annotated images. We evaluate models on the official test set with 2,400 images.

Deepglobe [7] is a complex Geographic Information System (GIS) dataset, containing satellite images with categories of urban, agriculture, rangeland, forest, water, barren, and unknown. For testing, we follow the processing in [20] to divide each image into six patches and filtering out the 'unknown' category, and evaluate models on the resulting 5,666 test images and their corresponding masks.

ISIC2018 [6, 42] is used for skin lesion analysis, containing numerous skin images with associated segmentation labels. We evaluate models on the official training set following the common practice [6], using a uniform resolution of 512×512 pixels, comprising a total of 2,596 test images.

Chest X-ray [3, 18] is an X-ray image dataset for tuberculosis detection, containing X-ray images of Tuberculosis cases as well as images from normal cases. We downsample the original image resolution to 1024×1024 pixels for testing.

4.2. Implementation Details

We utilize the train set of PASCAL VOC dataset as the source domain training set. During training, we employ SSP [12] with the ResNet-50 backbone as the baseline model. We first train the baseline model on the whole training set, and then train our method with additional 5 epochs using a batch size of 8. We use SGD to optimize our model, with a 0.9 momentum and an initial learning rate of 1e-3. To reduce memory consumption and accelerate the training

Table 1. Mean-IoU of 1-way 1-shot and 5-shot results of tradictional few-shot approaches and cross-domain few-shot method on the four CD-FSS benchmark.Bold denotes the best performance among all methods.

Methods	Deepglobe		ISIC		Chest X-ray		FSS-1000		Average	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Few-shot Segmentation Methods										
PGNet [52]	10.73	12.36	21.86	21.25	33.95	27.96	62.42	62.74	32.24	31.08
PANet [46]	36.55	45.43	25.29	33.99	57.75	69.31	69.15	71.68	47.19	55.10
CaNet [53]	22.32	23.07	25.16	28.22	28.35	28.62	70.67	72.03	36.63	37.99
RPMMs [48]	12.99	13.47	18.02	20.04	30.11	30.82	65.12	67.06	31.56	32.85
PFENet [41]	16.88	18.01	23.50	23.83	27.22	27.57	70.87	70.52	34.62	34.98
RePRI [2]	25.03	27.41	23.27	26.23	65.08	65.48	70.96	74.23	46.09	48.34
HSNet [32]	29.65	35.08	31.20	35.10	51.88	54.36	77.53	80.99	47.57	51.38
SSP [12]	40.48	49.66	35.09	44.96	74.23	80.51	79.03	80.56	57.20	63.92
Cross-domain Few-shot Segmentation Methods										
PATNet [20]	37.89	42.97	**41.16**	**53.58**	66.61	70.20	78.59	**81.23**	56.06	61.99
Ours	**41.29**	**50.12**	40.77	48.87	**82.35**	**82.31**	**79.05**	80.40	**60.86**	**65.42**

process, we resize both query and support images to 400 × 400. We apply our two domain perturbation modules into the first three layers of ResNet. For local perturbations, we use the Gaussian noise with a mean of zero and a standard deviation of 0.75, while for global perturbations, we used the Gaussian noise with a mean of zero and a standard deviation of one. All models are evaluated using the mean Intersection Over Union (mIOU).

4.3. Comparison Experiments

In Table 1, we present a comparison between our method and other approaches, including traditional few-shot segmentation methods and existing cross-domain few-shot segmentation methods. Traditional few-shot segmentation methods usually underperform in cross-domain scenarios due to the large domain gap between the train and test data. While our approach effectively reduces the domain gap and improves the segmentation performance. This performance improvement is particularly notable in Chest X-rays, where our 1-shot and 5-shot performance surpasses the PATNet [20] by 15.74% and 12.11%, respectively. In Deepglobe, the improvement is 3.4%(1-shot) and 7.15%(5-shot). For FSS-1000, our model achieves comparable performance to PATNet, because the domain gap is small.

We also follow the same setting of RD [47] to train our model on VOC and evaluate models on SUIM. Table 2 shows our method performs much better than RD.

We present some qualitative results of our proposed model for 1-way 1-shot segmentation in Fig. 5. These results indicate that our method improves the generalization ability of traditional few-shot models, attributing to its capability of aligning various domains to the source domain.

Table 2. Mean-IoU of 1-way 1-shot results of our method following the same setting of RD[47].

	split-0	split-1	split-2	split-3	Average
RD[47]	35.20	33.40	34.30	36.00	34.70
Ours	40.60	38.18	41.53	40.72	40.25

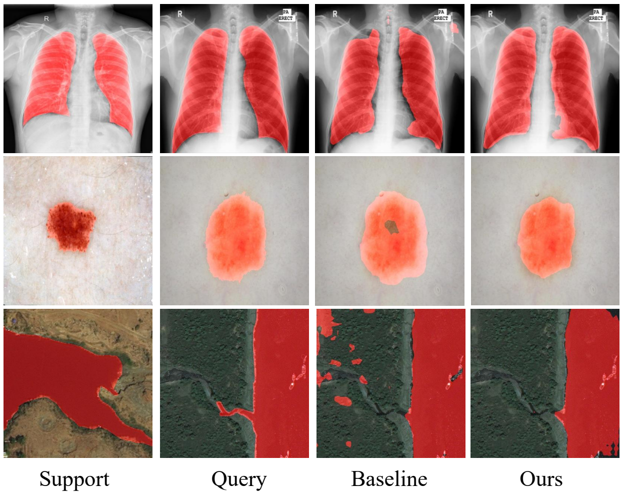


Figure 5. Qualitative results of our model and baseline in 1-way 1-shot setting on challenging scenarios with large domain gap.

4.4. More Analysis

We conduct extensive ablation experiments to demonstrate and analyze the effectiveness of our approach.

4.4.1 Ablation Studies

We conduct comprehensive ablation experiments to evaluate the effectiveness of our proposed components.

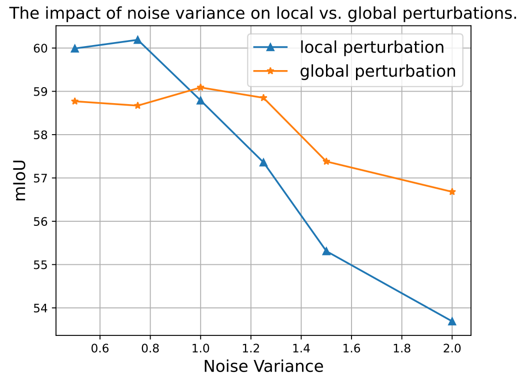


Figure 6. We demonstrate the trend of global and local perturbations under different Gaussian noise variances.

Table 3. The effects of each module within the baseline, namely the Perturbation module, Rectification module, and Cyclic Alignment Loss, are demonstrated.

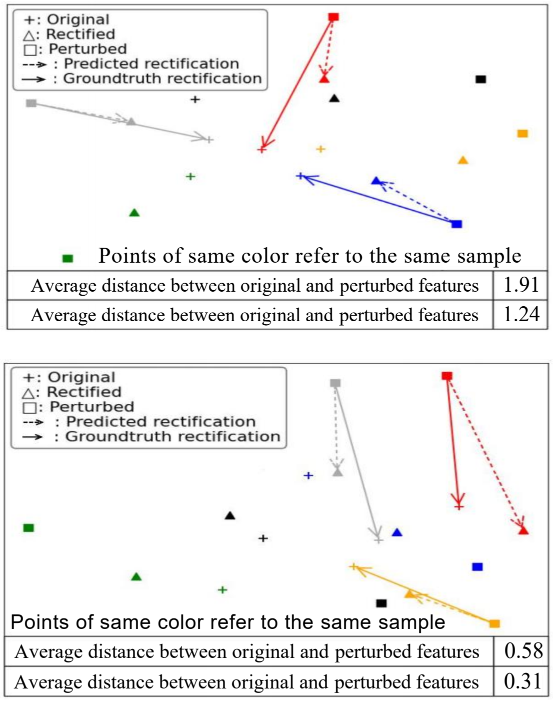
Perturbation	Rectification	Cyclic Alignment	mean-IoU
			57.20
✓			58.45
	✓		57.80
✓	✓		59.17
✓	✓	✓	60.86

Table 4. Results of using our Cyclic Alignment Loss.

BCE loss	+ cyclic loss	+ align loss	+ cyclic & align loss
57.65	58.56	59.12	60.86

Impact of Noise Variance on Perturbations. Figure 6 illustrates the effects of Gaussian noise with varying variances in local and global perturbations. Local perturbations suffer from performance degradation with slightly higher noise levels, whereas global perturbations withstand larger noise levels with minimal performance impact, suggesting greater stability. Thus, in our method, we set Gaussian noise variances at 0.75 for local and 1 for global perturbations to broaden the simulated domain range and improve the domain generalizability.

Impact of Each Component. Table 3 illustrates the effectiveness of each module in the model. Integration of all modules results in 3.66% performance improvement compared to the SSP [12] baseline. Importantly, the feature perturbation and rectification processes complement each other: perturbation simulates features across domains, and rectification aligns these features back to the source domain space. Solely using feature perturbation degrades the model to the domain generalization approach similar to NP [13]. Additionally, our cyclic alignment loss is indispensable as it ensures the unification of images from various domains to the source domain.



Average distance between original and perturbed features	1.91
Average distance between original and perturbed features	1.24

Average distance between original and perturbed features	0.58
Average distance between original and perturbed features	0.31

Figure 7. Visual analysis (t-SNE) of channel-wise means(top) and variations(bottom).

Table 5. Results of feature perturbation methods.

	local style	global style	Both perturbation
mean-IoU	59.81	59.17	60.86

Table 4 shows the ablation analysis on Cyclic Alignment Loss. Both the alignment loss and cyclic loss can improve the performance.

Table 5 compares the impact of local and global styles in feature perturbation, showing that their combination improves model performance attributing to a wider range of domain simulation. Using the channel-wise means and variances as features, the t-SNE(Figure 7) shows that the perturbed features are rectified to be closer to the original features, demonstrating our model's effectiveness.

Impact of Noise Types. We choose the popular Gaussian distribution to generate random noises, which has been widely used by other works (*e.g.*, Mixstyle, DSU and NP). Perturbing feature statistics with random noises can effectively synthesize diverse domain styles, while the noise type is not essential. Table 6 shows that our method is insensitive to the noise types, performing well with Beta, and Uniform noises. Note that our novel Local-Global Domain Perturbation and Cyclic Domain Alignment can largely improve the domain style synthesis diversity for all kinds of noise.

More adapters. Table 7 shows that applying multiple adapters can further improve performance.

Table 6. Results of using different types of noise.

Noise Type	Gaussian (1,0.75)	Beta (3,4)	Uniform (-1,1)
mIoU	60.86	60.78	60.00

Table 7. Results of using one/two adapters within a single stage.

	FSS	Chest	Deepglobe	ISIC	Average
One adapter	79.05	82.35	41.29	40.77	60.86
Two adapters	79.25	83.04	41.74	41.63	61.41

Table 8. Comparison to domain adaption and domain generalization approaches under 1-shot setting. We use same baseline with different methods to ensure fair comparison.

Method	FSS	Chest	Deepglobe	ISIC	Average
Baseline(SSP [12])	79.03	74.23	40.48	35.09	57.20
AdaIN [16]	78.89	74.23	**41.85**	34.36	57.33
Mixstyle [59]	**79.24**	76.63	41.05	35.98	58.21
DSU [23]	78.99	77.83	41.19	36.64	58.66
NP [13]	78.98	76.44	41.83	37.87	58.78
Ours	79.05	**82.35**	41.29	**40.77**	**60.86**

4.4.2 Comparion with Domain Transfer Methods

We compare our method against traditional domain adaptation (DA) and domain generalization (DG) approaches to validate our method's effectiveness. For a fair comparison, all categories in the PASCAL VOC were used for training in both DA and DG methods. We evaluate models in the 1-shot setting on the CD-FSS benchmark.

Domain Adaptation. We adopt the classical AdaIN [16] method to train four models for the four test datasets. During training, we randomly sample images from the test dataset and extract their feature channel statistics in the low-level feature map. And then the AdaIN is applied to replace the feature channel statistics of the train image with the extracted statistics from the test dataset.

Domain Generalization. We employ the Mixstyle [59], DSU [23] and NP [13] methods for comparison. These approaches also involves perturbing feature statistics, but they only perform local perturbations and lack a feature rectification process.

Table 8 shows that our method performs much better than DA and DG methods in cross-domain few-shot segmentation.

4.4.3 Applying SAM in CD-FSS

The recent released large-scale SAM [19] model has greatly advanced image segmentation, demonstrating remarkable zero-shot segmentation capabilities. However, SAM cannot be directly applied to cross-domain few-shot segmentation. Thus we evaluate PerSAM [56] to compare our method to the SAM-based method in cross-domain few-shot segmentation. PerSAM is a training-free method. It adapts SAM into the one-shot setting by using support images as the prompt input to segment target objects in query images. Table 9 shows that our method performs much better than PerSAM in cross-domain few-shot segmentation.

Table 9. The result of directly applying PerSAM to cross-domain few-shot segmentation.

	FSS	Chest	Deepglobe	ISIC	Average
PerSAM [56]	79.65	31.12	33.39	21.27	41.35
Ours	79.05	82.35	41.29	40.77	60.86

Table 10. Applying our method to transformers can further enhance the model's performance in cross-domain tasks.

	FSS	Chest	Deepglobe	ISIC	Average
FPTrans [55]	78.92	80.49	39.21	47.79	61.60
FPTrans + ours	78.63	82.74	40.32	49.43	62.78

4.4.4 Extension to Transformer

In Table 10, we show the results of applying our method within FPTrans[55], which leverages support sample prototypes as prompts and Vision Transformer (ViT) as the backbone. Applying our method to the lower-level blocks of ViT improves performance in cross-domain datasets.

5. Conclusion

In this paper, we propose a method to effectively bridge the domain gap between different datasets by aligning the target domain with the source domain space. During training, we train a unified adapter by using simulated perturbed features. In the inference stage, we consider target domain images as a form of perturbed images for the direct rectification. Furthermore,we introduce both local and global perturbations to ensure significant style changes, not only based on individual sample but also on the overall style of the dataset. We utilize a cyclic alignment loss to ensure the alignment between the source and target domains for model optimization. We conduct extensive experiments to validate the effectiveness of the proposed framework on various cross-domain segmentation tasks and achieve state-of-the-art (SOTA) results on multiple benchmarks.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China (U2013210, 62372133), Guangdong Basic and Applied Basic Research Foundation under Grant (Grant No. 2022A1515010306, 2024A1515011706), Shenzhen Fundamental Research Program (Grant No. JCYJ20220818102415032), and Shenzhen Key Technical Project (Grant NO. KJZD20230923115117033).

References

[1] Rodrigo Benenson, Stefan Popov, and Vittorio Ferrari. Large-scale interactive object segmentation with human annotators. In *CVPR*, 2019. 1

[2] Malik Boudiaf, Hoel Kervadec, Ziko Imtiaz Masud, Pablo Piantanida, Ismail Ben Ayed, and Jose Dolz. Few-shot segmentation without meta-learning: A good transductive inference is all you need? In *CVPR*, 2021. 6

[3] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, Jonathan P Musco, Rahul K Singh, Zhiyun Xue, Alexandros Karargyris, Sameer Antani, George Thoma, and Clement J McDonald. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Transactions on Medical Imaging*, 2013. 5

[4] Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *CVPR*, 2019. 3

[5] 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. *Transactions on Pattern Analysis and Machine Intelligence*, 2017. 1

[6] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 5

[7] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *CVPR*, 2018. 5

[8] Nanqing Dong and Eric P Xing. Few-shot semantic segmentation with prototype learning. In *BMVC*, 2018. 2

[9] Yingjun Du, Jun Xu, Huan Xiong, Qiang Qiu, Xiantong Zhen, Cees GM Snoek, and Ling Shao. Learning to learn with variational information bottleneck for domain generalization. In *ECCV*, 2020. 3

[10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 2010. 5

[11] Qi Fan, Wei Zhuo, Chi-Keung Tang, and Yu-Wing Tai. Few-shot object detection with attention-rpn and multi-relation detector. In *CVPR*, 2020. 3

[12] Qi Fan, Wenjie Pei, Yu-Wing Tai, and Chi-Keung Tang. Self-support few-shot semantic segmentation. In *ECCV*, 2022. 1, 5, 6, 7, 8

[13] Qi Fan, Mattia Segu, Yu-Wing Tai, Fisher Yu, Chi-Keung Tang, Bernt Schiele, and Dengxin Dai. Towards robust object detection invariant to real-world domain shifts. In *ICLR*, 2022. 3, 7, 8

[14] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, and David Balduzzi. Domain generalization for object recognition with multi-task autoencoders. In *ICCV*, 2015. 3

[15] Bharath Hariharan, Pablo Arbeláez, Lubomir Bourdev, Subhransu Maji, and Jitendra Malik. Semantic contours from inverse detectors. In *ICCV*, 2011. 5

[16] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *CVPR*, 2017. 4, 8

[17] Xinyang Huang, Chuang Zhu, and Wenkai Chen. Restnet: Boosting cross-domain few-shot segmentation with residual transformation network. In *BMVC*, 2023. 3

[18] Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue, Kannappan Palaniappan, Rahul K Singh, Sameer Antani, et al. Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging*, 2013. 5

[19] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *ICCV*, 2023. 8

[20] Shuo Lei, Xuchao Zhang, Jianfeng He, Fanglan Chen, Bowen Du, and Chang-Tien Lu. Cross-domain few-shot semantic segmentation. In *ECCV*, 2022. 1, 3, 5, 6

[21] Gen Li, Varun Jampani, Laura Sevilla-Lara, Deqing Sun, Jonghyun Kim, and Joongkyu Kim. Adaptive prototype learning and allocation for few-shot segmentation. In *CVPR*, 2021. 1, 3

[22] Xiang Li, Tianhan Wei, Yau Pun Chen, Yu-Wing Tai, and Chi-Keung Tang. Fss-1000: A 1000-class dataset for few-shot segmentation. In *CVPR*, 2020. 5

[23] Xiaotong Li, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan, and Ling-Yu Duan. Uncertainty modeling for out-of-distribution generalization. In *ICLR*, 2022. 8

[24] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 1

[25] Bingyu Liu, Zhen Zhao, Zhenpeng Li, Jianan Jiang, Yuhong Guo, and Jieping Ye. Feature transformation ensemble model with batch spectral regularization for cross-domain few-shot classification. *arXiv preprint arXiv:2005.08463*, 2020. 1

[26] Lizhao Liu, Junyi Cao, Minqian Liu, Yong Guo, Qi Chen, and Mingkui Tan. Dynamic extension nets for few-shot semantic segmentation. In *ACMMM*, 2020. 2

[27] Weide Liu, Chi Zhang, Guosheng Lin, and Fayao Liu. Crnet: Cross-reference networks for few-shot segmentation. In *CVPR*, 2020.

[28] Yongfei Liu, Xiangyi Zhang, Songyang Zhang, and Xuming He. Part-aware prototype network for few-shot semantic segmentation. In *ECCV*, 2020.

[29] Yuanwei Liu, Nian Liu, Qinglong Cao, Xiwen Yao, Junwei Han, and Ling Shao. Learning non-target knowledge for few-shot semantic segmentation. In *CVPR*, 2022. 2

[30] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015. 1

[31] Zhihe Lu, Sen He, Xiatian Zhu, Li Zhang, Yi-Zhe Song, and Tao Xiang. Simpler is better: Few-shot semantic segmentation with classifier weight transformer. In *ICCV*, 2021. 1, 3

[32] Juhong Min, Dahyun Kang, and Minsu Cho. Hypercorrelation squeeze for few-shot segmentation. In *ICCV*, 2021. 1, 3, 6

[33] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *ICCV*, 2017. 3

[34] Khoi Nguyen and Sinisa Todorovic. Feature weighting and boosting for few-shot segmentation. In *ICCV*, 2019. 1

[35] Bohao Peng, Zhuotao Tian, Xiaoyang Wu, Chengyao Wang, Shu Liu, Jingyong Su, and Jiaya Jia. Hierarchical dense correlation distillation for few-shot segmentation. In *CVPR*, 2023. 3

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 1

[37] Kuniaki Saito, Donghyun Kim, Piotr Teterwak, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Tune it the right way: Unsupervised validation of domain adaptation via soft neighborhood density. In *ICCV*, 2021. 1

[38] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. *ICLR*, 2018. 3

[39] Xinyu Shi, Dong Wei, Yu Zhang, Donghuan Lu, Munan Ning, Jiashun Chen, Kai Ma, and Yefeng Zheng. Dense cross-query-and-support attention weighted mask aggregation for few-shot segmentation. In *ECCV*, 2022. 3

[40] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *NeurIPS*, 2017. 2

[41] Zhuotao Tian, Hengshuang Zhao, Michelle Shu, Zhicheng Yang, Ruiyu Li, and Jiaya Jia. Prior guided feature enrichment network for few-shot segmentation. *TPAMI*, 2020. 1, 2, 3, 6

[42] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 2018. 5

[43] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *NeurIPS*, 2016. 3

[44] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. *NeurIPS*, 2018. 3

[45] Haohan Wang, Zexue He, Zachary C Lipton, and Eric P Xing. Learning robust representations by projecting superficial statistics out. In *ICLR*, 2019. 3

[46] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *ICCV*, 2019. 2, 6

[47] Wenjian Wang, Lijuan Duan, Yuxi Wang, Qing En, Junsong Fan, and Zhaoxiang Zhang. Remember the difference: Cross-domain few-shot semantic segmentation via meta-memory transfer. In *CVPR*, 2022. 1, 3, 6

[48] Boyu Yang, Chang Liu, Bohao Li, Jianbin Jiao, and Qixiang Ye. Prototype mixture models for few-shot semantic segmentation. In *ECCV*, 2020. 6

[49] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. Mining latent classes for few-shot segmentation. In *ICCV*, 2021. 2

[50] Xiangyu Yue, Zangwei Zheng, Shanghang Zhang, Yang Gao, Trevor Darrell, Kurt Keutzer, and Alberto Sangiovanni Vincentelli. Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation. In *CVPR*, 2021. 1

[51] Bingfeng Zhang, Jimin Xiao, and Terry Qin. Self-guided and cross-guided learning for few-shot segmentation. In *CVPR*, 2021. 1

[52] Chi Zhang, Guosheng Lin, Fayao Liu, Jiushuang Guo, Qingyao Wu, and Rui Yao. Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation. In *ICCV*, 2019. 3, 6

[53] Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In *CVPR*, 2019. 1, 2, 6

[54] Gengwei Zhang, Guoliang Kang, Yi Yang, and Yunchao Wei. Few-shot segmentation via cycle-consistent transformer. *NeurIPS*, 2021. 1, 3

[55] Jian-Wei Zhang, Yifan Sun, Yi Yang, and Wei Chen. Feature-proxy transformer for few-shot segmentation. *NeurIPS*, 2022. 8

[56] Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Hao Dong, Peng Gao, and Hongsheng Li. Personalize segment anything model with one shot. *arXiv preprint arXiv:2305.03048*, 2023. 8

[57] Xiaolin Zhang, Yunchao Wei, Yi Yang, and Thomas S Huang. Sg-one: Similarity guidance network for one-shot semantic segmentation. *Transactions on Cybernetics*, 2020. 2

[58] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *ECCV*, 2020. 3

[59] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *ICLR*, 2021. 3, 8