IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 25, 2019, accepted November 9, 2019, date of publication November 14, 2019,
date of current version November 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953465

Meta-Seg: A Generalized Meta-Learning Framework for Multi-Class Few-Shot Semantic Segmentation

ZHIYING CAO[1,2,4], **TENGFEI ZHANG**[1,2,4], **WENHUI DIAO**[1,2], **YUE ZHANG**[1,2], **(Member, IEEE), XIAODE LYU**[1,3], **KUN FU**[1,2,4], **AND XIAN SUN**[1,2]

[1]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100194, China
[2]Key Laboratory of Network Information System Technology (NIST), Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China
[3]Key Laboratory on Microwave Imaging Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China
[4]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China

Corresponding author: Xian Sun (sunxian@mail.ie.ac.cn)

This work was supported by the National Natural Science Foundation of China under Grant 41701508.

ABSTRACT Semantic segmentation performs pixel-wise classification for given images, which can be widely used in autonomous driving, robotics, medical diagnostics and etc. The recent advanced approaches have witnessed rapid progress in semantic segmentation. However, these supervised learning based methods rely heavily on large-scale datasets to acquire strong generalizing ability, such that they are coupled with some constraints. Firstly, human annotation of pixel-level segmentation masks is laborious and time-consuming, which causes relatively expensive training data and make it hard to deal with urgent tasks in dynamic environment. Secondly, the outstanding performance of the above data-hungry methods will decrease with few available training examples. In order to overcome the limitations of the supervised learning semantic segmentation methods, this paper proposes a generalized meta-learning framework, named Meta-Seg. It consists of a meta-learner and a base-learner. Specifically, the meta-learner learns a good initialization and a parameter update strategy from a distribution of few-shot semantic segmentation tasks. The base-learner can be any semantic segmentation models theoretically and can implement fast adaptation (that is updating parameters with few iterations) under the guidance of the meta-learner. In this work, the successful semantic segmentation model FCN8s is integrated into Meta-Seg. Experiments on the famous few-shot semantic segmentation dataset PASCAL5^i prove Meta-Seg is a promising framework for few-shot semantic segmentation. Besides, this method can provide with reference for the relevant researches of meta-learning semantic segmentation.

INDEX TERMS Meta-learning, few-shot, semantic segmentation.

I. INTRODUCTION

In recent years, deep learning, especially convolutional networks [1], have made significant breakthroughs in many visual understanding tasks including image classification [2]–[5], object detection [6]–[13] and semantic segmentation [14]–[21]. One crucial reason driving their development is the availability of large-scale datasets such as ImageNet [22] that enable the training of deep networks. Semantic segmentation aims to assign a class label to each pixel in an image. Deep convolutional network in semantic segmentation, as shown in Fig. 1 (a), requires a large amount of annotated data to ensure the robustness of the model. It still faces the challenges of overfitting in a few-shot regime. Nevertheless, data labeling is expensive and laborious, particularly for dense prediction tasks, *e.g*, semantic segmentation, instance segmentation and panoptic segmentation. Hence, weakly supervised semantic segmentation methods [23]–[27] are proposed to reduce the burden of data annotation. These methods merely solve the dependencies on annotated data, which still require plenty of training images. In addition to that, once the segmentation model is trained, it is difficult to

The associate editor coordinating the review of this manuscript and approving it for publication was Gulistan Raja.

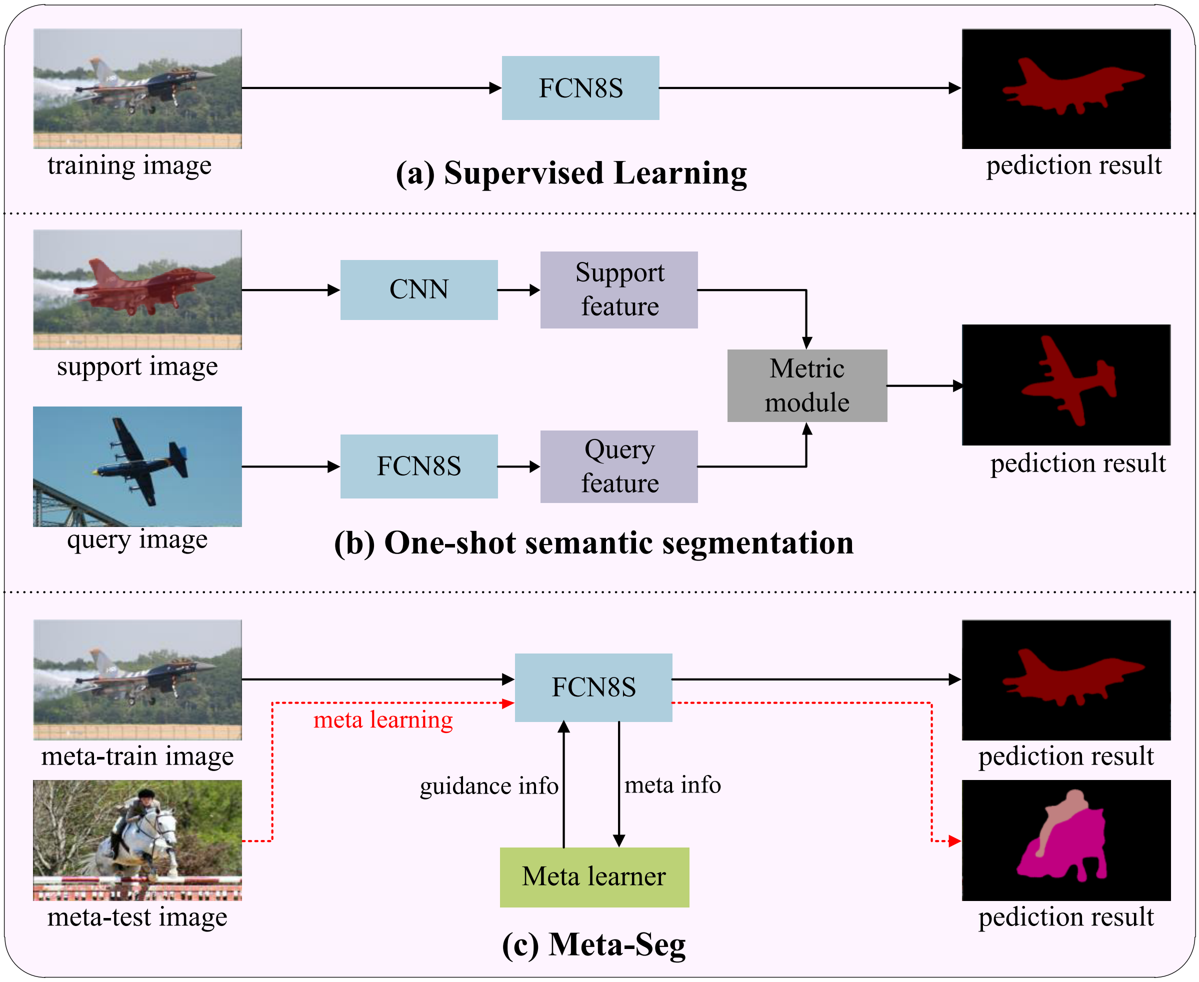


FIGURE 1. Comparison of three different semantic segmentation models. (a) The supervised learning semantic segmentation models rely on large-scale annotated training data of trained classes, e.g. aeroplane, but cannot work on unseen classes, e.g. person. (b) The Siamese structure based semantic segmentation models have the generalizing ability of few-shot semantic segmentation on unseen classes, but they mostly focus on single-class segmentation. (c) The proposed Meta-Seg can implement multi-class, few-shot semantic segmentation as well as fast parameter adaptation on unseen classes.

use existing model to predict new classes. In contrast, humans can segment a novel concept from the scene easily even with few samples.

The gap between humans and deep neural networks in learning ability with few samples motivates the study of few-shot learning method. Many researches have prompted related works in few-shot image classification [28]–[40] and detection [41], [42]. These remarkable methods, mainly based on transfer learning and meta-learning, have made a certain progress in avoiding overfitting with few training examples and alleviating the heavy burden of human annotation.

While in semantic segmentation, there are few researches focus on the few-shot semantic segmentation problem, especially meta-learning semantic segmentation. The previous works [43]–[47] mostly employ the Siamese structure to implement one- or few-shot semantic segmentation by giving annotated images as a condition, shown in Fig. 1 (b). Although these Siamese structure based methods have got some achievements, they mostly focus on single-class semantic segmentation in each forward propagation. The network structure and the pair-wise data organization may be inefficient for multi-class semantic segmentation at the same time.

This work aims to overcome the limitations of the supervised learning based methods as well as implement multi-class semantic segmentation in few-shot regime. We follow the excellent meta-learning methods [28], [40], [42] in few-shot image classification and object detection and propose a generalized meta-learning framework Meta-Seg for few-shot semantic segmentation, as shown in Fig. 1 (c). Meta-Seg, consists of a meta-learner and a base-learner, is characterized with multi-class, few-shot semantic segmentation and fast parameter adaptation. To the best of our knowledge, this is the first work to implement a generalized meta-learning framework for few-shot semantic segmentation. In addition to multi-class semantic segmentation, another advantage of Meta-Seg is its performance can be further improved with the development of supervised learning based semantic

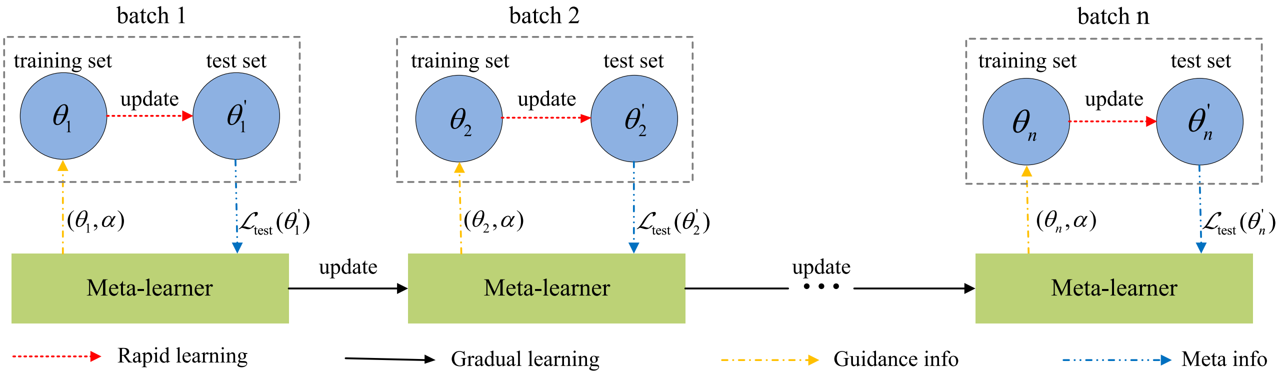


FIGURE 2. Meta-learning process of the generalized Meta-Seg framework. θ denotes the parameters of the semantic segmentation network. Meta-learner guides the semantic segmentation model to adjust its parameters θ according to the feedback from the training set in each task, then is updated by leveraging the meta-info from the test-sets of a batch of tasks.

segmentation models, because Meta-Seg is a generalized framework and the base-learner can be replaced with different deep-learning based semantic segmentation models.

As shown in Fig. 2, the meta-learning pipeline of Meta-Seg can be divided into meta-training and meta-test. In meta-training phase, Meta-Seg is trained on a series of few-shot tasks. Within each K-way N-shot task, there are K randomly selected classes and N annotated training images for each class. The meta-learner is responsible to guide the base-learner to learn quickly from few training examples by learning a good initialization and a parameter update strategy. After trained on many meta-training tasks, meta-learner can acquire strong prior knowledge to extend the capacities of few-shot multi-class semantic segmentation and fast parameter adaptation to the meta-test tasks with different categories from meta-training tasks.

In this work, FCN8s is used as the base-learner. It should be noted that other semantic segmentation models can also be integrated into the generalized framework Meta-Seg in theory. We evaluate Meta-Seg on the PASCAL5^i [43] dataset devised for few-shot semantic segmentation.

Our contributions are as follows:

(1) A generalized meta-learning framework named Meta-Seg is proposed for few-shot semantic segmentation. It is easy to integrate existing arbitrary semantic segmentation models in Meta-Seg theoretically.

(2) Compared with the previous few-shot semantic segmentation methods, Meta-Seg can implement multi-class semantic segmentation and fast parameter adaptation simultaneously with few training examples.

(3) The experimental results on the PASCAL5^i dataset with classical FCN8s base-learner have indicated that Meta-Seg is an effective framework and meta-learning is a promising method for multi-class few-shot semantic segmentation.

II. RELATED WORKS

Our work aims to address the few-shot semantic segmentation problem by combining the meta-learning approach with supervised-learning semantic segmentation method. In this section we will expatiate the related works of semantic segmentation, meta-learning and few-shot semantic segmentation respectively.

A. SEMANTIC SEGMENTATION

Semantic segmentation is an active research area where deep learning are used to classify each pixel in the image individually, especially since the introduction of fully convolutional networks (FCN) [16]. FCN replaces the fully connected layers with convolutional layers to fit the task of dense prediction. Plenty of state-of-art methods are proposed based on the architecture of FCN which often employ a convolutional neural network (CNN) pretrained for classification as the backbone networks. At present, the research direction in semantic segmentation can be roughly divided into two directions, namely dilated-based model and encoder-decoder model. Dilated-based model utilizes dilated convolutions [48] to obtain large receptive field of view. Besides, multi-scale context modules are often used to obtain high-level semantic features. Encoder-decoder model utilizes the encoder to extract feature maps and utilizes the decoder to combine the feature maps into the final predictions.

Dilated-Based Model: In order to extract multi-scale context information, PSPNet [49] proposes spatial pyramid pooling (SPP) at several grid scales. DeeplabV3plus [50] performs atrous spatial pyramid pooling (ASPP) with several parallel atrous convolution in different rates. Inspired by atrous spatial pyramid pooling (ASPP), DenseASPP [51] is proposed to further capture dense context information. FastFCN [52] proposes a joint pyramid upsampling (JPU) module to extract high-resolution feature maps, which can reduce the computation complexity without performance loss.

Encoder-Decoder Model: In order to gradually recover the spatial resolution, U-Net [53] proposes skip connections as the decoder module. SegNet [54] constructs a typical encoder-decoder model architecture which consists an encoder network, a corresponding decoder network followed by a dense prediction classification layer. RefineNet [55] utilizes all the features available along the down-sampling process with a multi-path refinement network. DeeplabV3plus [50] attempts to combine the advantages from both dilated method and encoder-decoder method, which employs a simple but efficient decoder module to recover spatial information.

B. META-LEARNING

Meta Learning [56], also known as learning to learn, is the science of systematically learning the experience or meta-data which is learned by different machine learning approaches in a wide range of learning tasks. Few-shot learning can be regarded as an application of meta-learning in the field of supervised learning. As a general approach to few-shot learning, meta-learning aims to train a robust model using only a few training data, given prior experience with similar tasks for which we have a large amount of training data available. In general, the meta-learning architecture usually contains two major components, a meta-learner and a learner. Meta-learner can be regarded as a teacher, imparting prior experience to learner. Learner uses prior experience to learn a common feature representation of tasks and is trained on a distribution of similar tasks with a better model parameter initialization and acquire an inductive bias which helps guide the optimization of parameters. Thus, new tasks can be trained much faster in such regime.

Meta-Learning has made significant breakthroughs in the filed of computer vision. There exist many formulations including recurrent neural network with memories [31], learning to fine-tune models [28], [40], network parameter prediction [30] and metric learning [33]. Reference [31] utilizes a memory-augmented model for rapid generalization on new tasks. Model-agnostic meta-learning (MAML) [28] learns a model parameter initialization that generalizes better to similar tasks. Based on MAML [28], Meta-SGD [40] proposes a method to learn a set of model parameters as well as a learning rate for each parameter. In [30], experience with already learned samples is used to facilitate the learning of novel samples. Relation network [33] meta-learns a distance metric and computes the similarity score for classification. Our work is most related to Meta-SGD [40], which is a SGD-like meta-learner to learn initialization, update direction and learning rate via meta-learning in an end-to-end manner. The Meta-Seg proposed in this work can be regarded as an extension of meta-SGD in a dense form to tackle the task of semantic segmentation.

C. FEW-SHOT SEMANTIC SEGMENTATION

At present, the related researches in few-shot semantic segmentation [43]–[47] are relatively less than those in few-shot image classification. Reference [43] is the first work of few-shot semantic segmentation, which is based on a Siamese structure. A support branch processes the annotated object as the condition to predict the weights of the query branch which extracts the feature of a test image for semantic segmentation. In [44], the feature extracted from the support branch can be viewed as a condition of the query branch for few-shot semantic segmentation. The two methods mainly focus on one-shot semantic segmentation.

These methods are based on Siamese structure and focus on single-class semantic segmentation for each forward propagation. We argue that the two-branch structure is inconvenient and ineffective for multi-class few-shot semantic segmentation. Therefore, we address the multi-class semantic segmentation from the perspective of meta-learning.

III. METHODOLOGY

The goal of this work is to implement multi-class few-shot semantic segmentation and fast parameter adaptation via meta-learning. The proposed Meta-Seg is a generalized meta-learning framework which can be combined with any supervised learning based semantic segmentation model in theory. In this section, we first introduce the related concepts about meta-learning and give the problem formulation of few-shot semantic segmentation, followed by a brief description of the base-learner in Meta-Seg. Then, the whole meta-learning pipeline is illustrated in detail.

A. PRELIMINARY FOR META-LEARNING

Meta-learning, also known as learning to learn, has been widely used in few-shot image classification [28], [40], regression and object detection [42]. As shown in Fig. 2, a common meta-learning framework, consists of a meta-learner and a base-learner, aims to learn a good initialization and a parameter update strategy for few-shot semantic segmentation. Different from supervised learning, meta-learning is trained on a series of **tasks**. The whole meta-learning pipeline is composed of **meta-training** and **meta-test** phases. Herein, we introduce some concepts in meta-learning.

Task: The basic training unit of meta-learning is "task" $\mathcal{T}$. For each task $\mathcal{T}$, there are two sub parts: meta-training $\mathcal{T}^{tr}$ and meta-test $\mathcal{T}^{te}$. A K-way N-shot task is defined by the number of categories and the number of training images per category in $\mathcal{T}^{tr}$. That is, there are K categories and N training images per category in $\mathcal{T}^{tr}$ of a K-way N-shot task. The categories and training images are randomly selected and usually vary from task to task.

Data Organization: As shown in Fig. 3, the meta-learning dataset $\mathcal{D}$ can be divided into two sets: meta-training set $\mathcal{D}^{tr}$ and meta-test set $\mathcal{D}^{te}$. The categories and images of the two sets are non-overlapping. In general, the categories on $\mathcal{D}^{tr}$ can be called meta-training classes or seen classes, similarly, the categories on $\mathcal{D}^{te}$ can be called meta-test classes or unseen classes. This setting can mimic the fact that there may not always have sufficient training images for some categories and can ensure the effective evaluation for meta-learning methods.

Meta-Training: In meta-training phase, the meta-learning model is trained on meta-training set $\mathcal{D}^{tr}$. We sample categories and training images to build a task $\mathcal{T}$. In each K-way N-shot task $\mathcal{T}$, for $\mathcal{T}^{tr}$, there are K categories and N training images per category, and for $\mathcal{T}^{te}$, there are same K categories but different N training images per category from $\mathcal{T}^{tr}$. Within

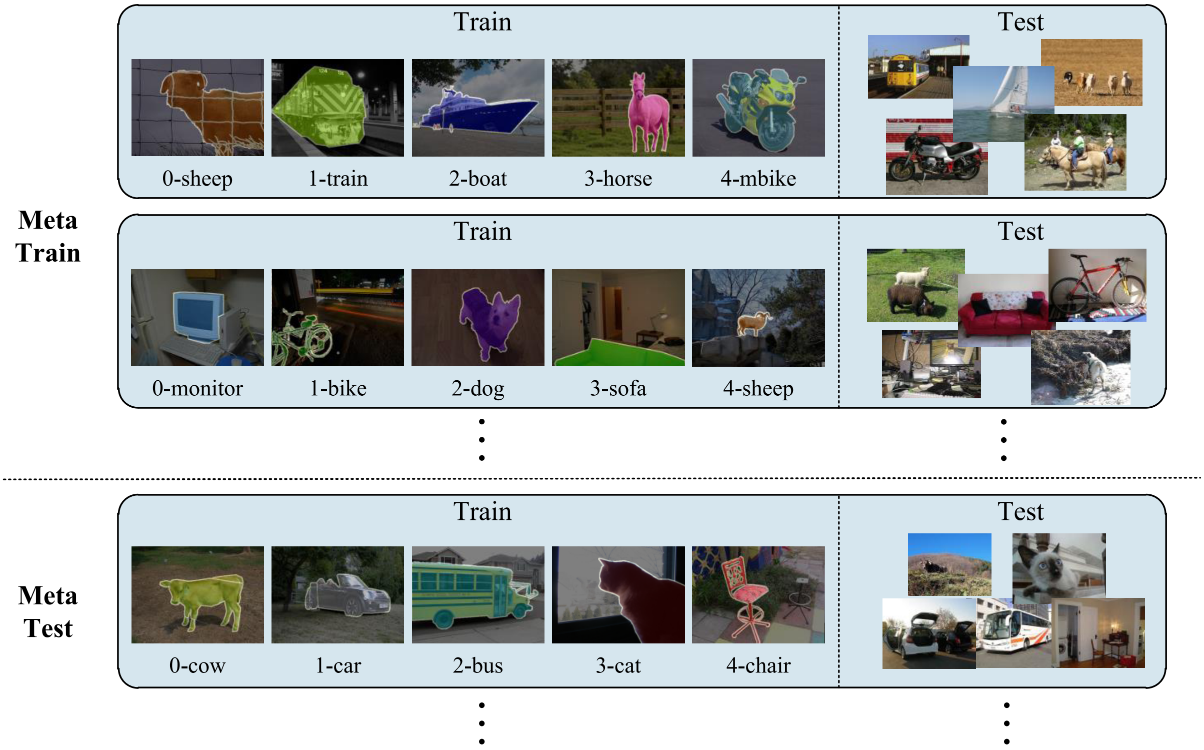


FIGURE 3. The data organization of meta-learning semantic segmentation of 5-way 1-shot. The dataset is composed of meta-training and meta-test sets. In each task (denoted by blue box), there are training and test subsets sampled from meta-training or meta-test set, and they have same classes. One image per class is randomly selected for training, test subsets of meta-training and training subset of meta-test, while fifteen images per class are randomly selected for test subset of meta-test. Note that the classes of meta-training set are not present in meta-test set.

each task, the base-learner is trained on $\mathcal{T}^{tr}$, then update its parameters under the guidance of meta-learner by leveraging the feedback from $\mathcal{T}^{tr}$. Hence, we can get a temporary semantic segmentation model for the current task. The performance of the temporary segmentation model is evaluated on $\mathcal{T}^{te}$, then the loss from $\mathcal{T}^{te}$ is collected to update the meta-learner.

Meta-Test: In meta-test phase, the trained meta-learning model is evaluated on meta-test set $\mathcal{D}^{te}$. The sampling of categories and images is same as that in meta-training phase. But it should be noted that the number of images per category on $\mathcal{T}^{te}$ is usually more than that in meta-training phase for effective evaluation. The meta-learning model update the parameters of the base-learner according to the loss from $\mathcal{T}^{tr}$, then the updated base-learner is evaluated on $\mathcal{T}^{te}$. The accuracies from all $\mathcal{T}^{te}$ in meta-test phase are averaged as the final evaluated result.

B. PROBLEM FORMULATION

With the introduction for meta-learning mentioned above, the proposed Meta-Seg is to learn a good initialization and a parameter update strategy from the meta-training tasks. Once trained, the meta-learner should be capable of guiding the base-learner to update its parameters with few available training images and few update iterations for multi-class semantic segmentation. Generally, the meta-training and meta-test tasks are from a same distribution $p(\mathcal{T}_i)$ of few-shot tasks, but they have different categories.

The proposed Meta-Seg can combine with any deep learning semantic segmentation models theoretically, which can bring two advantages. Firstly, it is easier and more efficient to implement multi-class semantic segmentation than the previous Siamese structure based models. Secondly, the performance of Meta-Seg will be further improved with the development of supervised learning based semantic segmentation models.

C. FCN8s REVISIT

As illustrated in Fig. 4, the proposed meta-learning semantic segmentation model utilizes FCN as the learner to obtain the segmentation results. In this section, we will not go into the details of FCN. Readers can refer to [16] for more details about architecture design and experimental setting. FCN is a semantic segmentation network based on a convolutional neural network, which has demonstrated significant improvement than conventional methods. FCN employs classification network VGG [3] as the base network for feature extraction. In order to apply classification network to pixel-level prediction task, FCN replaces the fully connected layers with the fully convolutional layers. In the FCN framework, several stages of strided convolutional and spatial pooling reduce the predictions by a factor of 32, which leads to

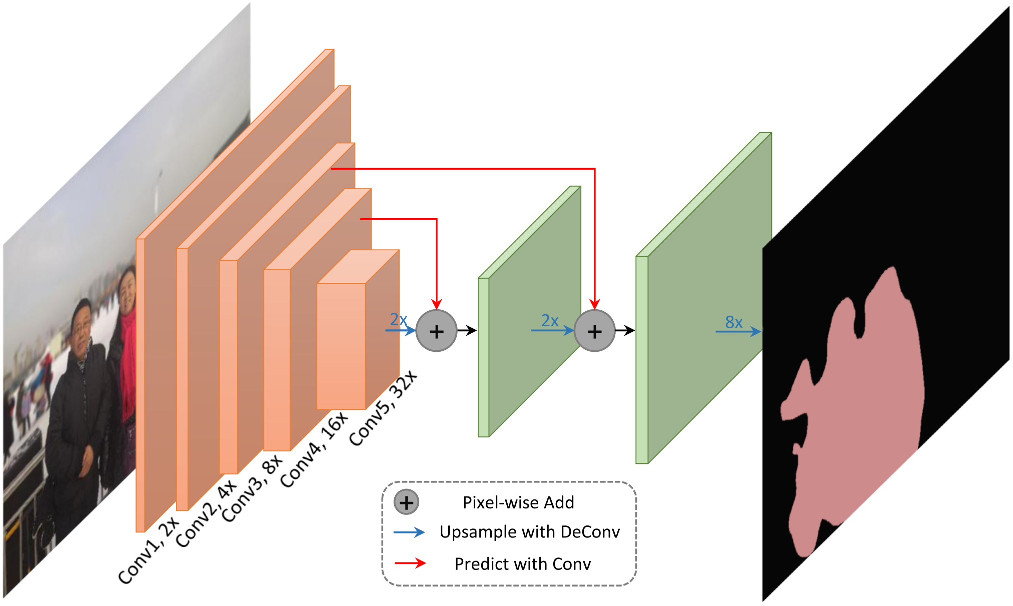


FIGURE 4. The architecture of semantic segmentation network FCN8s [16]. FCN8s employs VGG16 as the base network to extract features, then employs deconvolution layers for upsampling. Besides, FCN8s utilizes features extracted from different layers for dense prediction.

inaccurate predictions for losing fine spatial details. FCN employs deconvolution layers to connect coarse outputs to dense pixels as the interpolation methods. Deconvolution operation simply reverse the forward and backward passes of convolution, thus upsampling is performed for an end-to-end learning. Another way to refine the output predictions is combining coarse, high layer information with fine, low layer information. FCN32s directly upsamples the final prediction with stride=32 back to original resolution. FCN16s combines final prediction with prediction from low layer and upsamples with stride=16. Based on FCN16S, additional predictions is utilized in FCN8s as illustrated in Fig. 4. Compared with original FCN network, FCN8s combines more spatial information from different layers and further improves the segmentation results.

This work is devoted to solving the problem of semantic segmentation from the perspective of meta-learning. Thus a simple semantic segmentation model is needed to evaluate the feasibility of the meta-learning framework. FCN8s has become the prototype of the mainstream segmentation model because of its efficient and simple architecture. In this work, we select FCN8s to implement meta-learning semantic segmentation, called Meta-Seg. It should be noted that our goal of this work is to verify the feasibility of Meta-Seg framework, rather than pursue the accuracy of the model. Meta-Seg can embed other complex and well-designed semantic segmentation models flexibly, which will be left for future work.

D. META-LEARNING SEMANTIC SEGMENTATION

In this subsection, we introduce how to meta-learn a few-shot semantic segmentation model.

For a standard supervised learning pipeline, we train the semantic segmentation model for hundreds of thousands of iterations on a large-scale dataset as the following:

$$\theta' = \theta - \alpha \nabla \mathcal{L}(\theta) \tag{1}$$

here, α denotes the learning rate and $\mathcal{L}$ is the cross-entropy loss for semantic segmentation.

While the meta-learning pipeline is different. An effective approach of meta-learning is to learn a good initialization as well as a parameter update strategy. This is based on the following reasons.

(1) The meta-learning model can only acquire limited knowledge from the few training images on meta-test classes. It is very difficult to converge well from scratch. Thus, a good initialization is crucial.

(2) The parameter update methods in supervised learning can not work well in the few-shot regime, because few training images will cause overfitting and poor performance. Therefore, a good parameter update strategy is necessary.

With the above analysis, the proposed meta-learning framework Meta-Seg is designed elaborately to solve the few-shot semantic segmentation problem, inspired by the famous meta-learning methods [28], [40]. Meta-Seg can learn a good initialization and a good parameter update strategy to implement few-shot semantic segmentation and fast parameter adaptation. Once trained on many meta-training tasks, Meta-Seg can extend the few-shot semantic segmentation ability to the meta-test tasks of unseen classes.

In concretely, Meta-Seg consists of a meta-learner and a base-learner. Within each task, Meta-Seg works in the supervised learning manner. The base-learner is trained on $\mathcal{T}^{tr}$, then update network with the guidance of the meta-learner.

This process can be formulated as follows:

$$\theta' = \theta - \alpha^* \nabla \mathcal{L}_{\mathcal{T}}(\theta) \tag{2}$$

Note that the learning rate α^* is provided by the meta-learner rather than an optimizer such as stochastic gradient descent (SGD). Obviously, α^* is the representative of the parameter update strategy and reflects the role of the meta-learner.

After training the base-learner on $\mathcal{T}^{tr}$, the next step is how to learn the learnable learning rate α^* and a good initialization θ. We test the updated base-learner on $\mathcal{T}^{te}$. The objective of meta-learning is to get a good temporary base-learner for each task and can be formulated as the following:

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\phi_{\theta'}) = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\phi_{\theta - \alpha^* \nabla \mathcal{L}_{\mathcal{T}_i}(\phi_\theta)}) \tag{3}$$

Herein, we follow [40] and update the base-learner's parameters with one step in each task, which is significantly different from supervised learning. This setting can avoid overfitting and speed up the meta-learning speed. The one-step update corresponding to the fast parameter adaptation mentioned above.

We collect the losses from $\mathcal{T}^{te}$ of a batch of tasks as the meta-info to update the meta-learner (that is, updating α^* and θ). This parameter update for meta-learner can be implemented by stochastic gradient descent (SGD) as follows:

$$(\theta, \alpha^*) = (\theta, \alpha^*) - \beta \nabla \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\phi_{\theta'}) \tag{4}$$

Here, β is the learning rate for the meta-learner just like in supervised learning. So far, the whole meta-learning framework can be trained in the supervised learning manner and is easy to implemented based on the common deep learning framework, e.g. Pytorch [57].

Algorithm 1 Meta-Learning for Few-Shot Semantic Segmentation

```
Input: Few-shot semantic segmentation task distribution
    $p(\mathcal{T})$, learning rate $\beta$ for meta-learner
Output: Few-shot semantic segmentation model's parameters $\theta$, learnable learning rate $\alpha^*$ for base-learner
Initialize $\theta, \alpha^*$
while not end do
    Sample $n$ tasks from $p(\mathcal{T})$
    for all $j = 1; j \leq n$ do
        $\mathcal{L}_{\mathcal{T}_j^{train}} = \frac{1}{|\mathcal{T}_j^{train}|} \sum_{i \in \mathcal{T}_j^{train}} \ell(\phi_\theta(i))$
        $\theta' = \theta - \alpha^* \nabla \mathcal{L}_{\mathcal{T}_{j_{train}}}$
        $\mathcal{L}_{\mathcal{T}_j^{test}} = \frac{1}{|\mathcal{T}_j^{test}|} \sum_{i \in \mathcal{T}_j^{test}} \ell(\phi_{\theta'}(i))$
    end for;
    $(\theta, \alpha^*) = (\theta, \alpha^*) - \beta \nabla \sum_{j \in (1,n)} \mathcal{L}_{\mathcal{T}_j^{test}}$
end while
```

TABLE 1. The settings of training and test sets for baseline Seg-JT and Seg-FT. S indicates all examples of meta-training classes and n/U indicates n examples per meta-test classes. Note that: $S1 \subseteq S$, $S2 \subseteq S$, $U1 \subseteq U$, $U2 \subseteq U$, $S1 \cap S2 = \varnothing$,$U1 \cap U2 = \varnothing$.

model	train	test
Seg-JT	S + 1/U	5/U
Seg-FT	S1	S2
	1/U1	5/U2

TABLE 2. Meta-test classes for each fold of PASCAL5^i.

i=0	i=1	i=2	i=3
aeroplane, bicycle, bird, boat, bottle	bus, car, cat, chair, cow	diningtable, dog, horse, motorbike, person	potted plant, sheep, sofa, train, tv/monitor

E. IMPLEMENTATION

In this work, Meta-Seg is implemented by integrating FCN8s as the base-learner, which is a simple and effective semantic segmentation model. It is an end-to-end framework. The learnable learning rate α^* is set for every parameter of FCN8s, such that the meta-learner can guide it to update parameters quickly with few training images. Besides, in each task, FCN8s works in the supervised learning manner and can implement multi-class semantic segmentation effectively.

The whole system is trained on a series of meta-training tasks and can learn a temporary few-shot multi-class semantic segmentation model for each task. The meta-learner is updated by leveraging the feedback from each $\mathcal{T}^{te}$. The meta-training task $\mathcal{T}^{tr}$ is consistent with the meta-test task $\mathcal{T}^{te}$, which ensures the successful generalization to meta-test classes. This meta-learning framework is so generic that other semantic segmentation models can be integrated in the future research.

IV. BENCHMARK

We train and evaluate Meta-Seg on the famous few-shot semantic segmentation dataset PASCAL5^i, which is commonly used in related few-shot semantic segmentation methods. It has 4 random class partitions for effective evaluation of few-shot models, there are 15 classes as meta-training classes and 5 classes as meta-test classes in each partition, the detail is shown in Table 2.

V. EXPERIMENTS

A. BASELINE

1) BASELINES FOR FAIR COMPARISON

In order to highlight the effectiveness of Meta-Seg, we compare it with two baselines. The first baseline jointly train segmentation network (FCN8s) on meta-training classes with sufficient labeled examples and meta-test classes with one labeled example per class. For the sake of simplicity, we will refer to the baseline network as Seg-JT. The second baseline is a two-stage training process. The baseline train segmentation network only on meta-training classes with sufficient labeled examples and then fine-tune it on meta-test classes with few examples. we will refer to the baseline network as Seg-FT.

2) BASELINES FOR QUALITATIVE COMPARISON

We note that some relevant researches [43], [44] implement the one-shot semantic segmentation based on two-branch structure, which is different from this work. The two-branch based frameworks just focus on single class during each forward computation. This is inconvenient and ineffective for multi-class few-shot semantic segmentation. Besides the two-branch based models are trained and evaluated based on the support-query image pairs. Therefore, we can not compare our model with them fairly and just use them for qualitative comparison. We also provide the performances of three baselines (1-NN, LogReg and Siamese) in OSLSM [43], the detailed instructions can be found in [43].

3) WHY NOT COMPARE WITH THE SEMANTIC SEGMENTATION MODELS USING LABELS OF ALL TRAINING IMAGES?

We don't compare with the supervised learning based semantic segmentation models which are using labels of all training images. The reasons are as follows. Firstly, the few-shot image segmentation models still lag behind the state-of-the-art deep learning based semantic segmentation models in performance, i.e., 48.6% mIOU of 1-shot Meta-Seg vs. 62.7% mIOU of FCN8S. Secondly, the comparison between the few-shot image segmentation models and the state-of-the-art deep learning based models is not fair. For the test classes, the deep learning models are trained with all training images, while the few-shot learning models are trained with a few training images. Besides, the training and evaluation of the deep learning based models are on same classes, while the training and evaluation of the few-shot segmentation models are on different classes (seen and unseen classes). Thirdly, the relevant few-shot semantic segmentation models also didn't compare with the state-of-the-art deep learning based models. Finally, the goal of this work is to implement a few-shot image segmentation model. Therefore, the comparison with the deep learning based image segmentation models is not necessary.

B. EXPERIMENTAL SETTING

For Meta-Seg and two baselines, we fine-tune the model weights of the Imagenet-pretrained VGG16 network to adapt them for the segmentation task. Truncated normal distribution initializes the rest parameters which are not included in the pretrained model. All of our networks are implemented on Tesla P100 GPU. The basic learning rate β for meta-learner is set to 10^{-3}. The learnable learning rate α^* for base-learner is initialized to 10^{-3}. All models are optimized by the SGD optimizer. The meta-learning model is trained with 30000 episodes and the inner task is 4 (that is, every episode has 4 tasks). Seg-JT is trained with 20 epoches and 12 batch size with the learning rate of 10^{-3} for fair comparison. Seg-FT is pre-trained in the same setting as Seg-JT, then is fine-tuned with 1000 epoches and 5 batch size. Table 1 shows the training and test data organization of two baselines.

TABLE 3. Semantic segmentation results of Meta-seg and baseline methods. The numbers in the table indicate the mIOU (%) on different folds.

Method	split-0	split-1	split-2	split-3	mean
Seg-JT-1	4.5	4.7	4.4	3.9	4.4
Seg-FT-1	19.5	31.9	27.7	21.5	25.2
Meta-Seg-1	**42.2**	**59.6**	**48.1**	**44.4**	**48.6**
Seg-JT-5	6.9	9.5	5.4	4.2	6.5
Seg-FT-5	29.9	44.4	34.8	25.5	33.7
Meta-Seg-5	**43.1**	**62.5**	**49.9**	**45.3**	**50.2**

We just train Meta-Seg in 1-shot setting due to the limitation of graphic memory. This is a challenge for Meta-Seg and we will solve this problem in the future research. However, we still evaluate the 5-shot performance of Meta-Seg using the trained 1-shot model. It should be noted that the 5-shot performance of Meta-Seg will be better theoretically.

C. PERFORMANCE

In order to prove the effective of the proposed Meta-Seg, we evaluate it with 4 different class partitions. The experiment results are shown in Table 3. Meta-Seg achieves promising results for 4 different class partitions. For split-1, Meta-Seg yields the best performance of 59.6% than other class partitions. This is because the objects in split-1 are relatively larger and easier for segmentation. While the mean IOU 42.2% in split-0 is lowest due to the small objects such as bird and bottle. As mentioned in experimental setting, the 5-shot performance of Meta-Seg is obtained by evaluating the trained 1-shot Meta-Seg in 5-shot setting. The experiment results are shown in Table 3, there is a slight performance gain (1.6% mean IoU).

The experiment results indicate that Meta-Seg is an effective multi-class few-shot semantic segmentation framework. Once trained on the distribution of few-shot semantic segmentation tasks on meta-training classes, Meta-Seg can successfully classify the pixels of meta-test class objects with only few training examples and iterations.

D. COMPARISON WITH BASELINES

As shown in Table 3, both the two baselines yield worse performance than Meta-Seg, which can highlight the advantages of Meta-Seg. We visualize the results of Meta-Seg and the two baselines in Fig. 5 for better comparison. Although Seg-JT, shown in the third row in Fig. 5, can distinguish objects from background, it cannot even recognize the correct category of the meta-test class objects. For example, Seg-JT classifies bus as train. It means that Seg-JT is disturbed by meta-training classes with more training examples. This can explain why the performance of Seg-JT is so poor. In contrast, Seg-FT can focus on the meta-test classes after fine-tuning. While

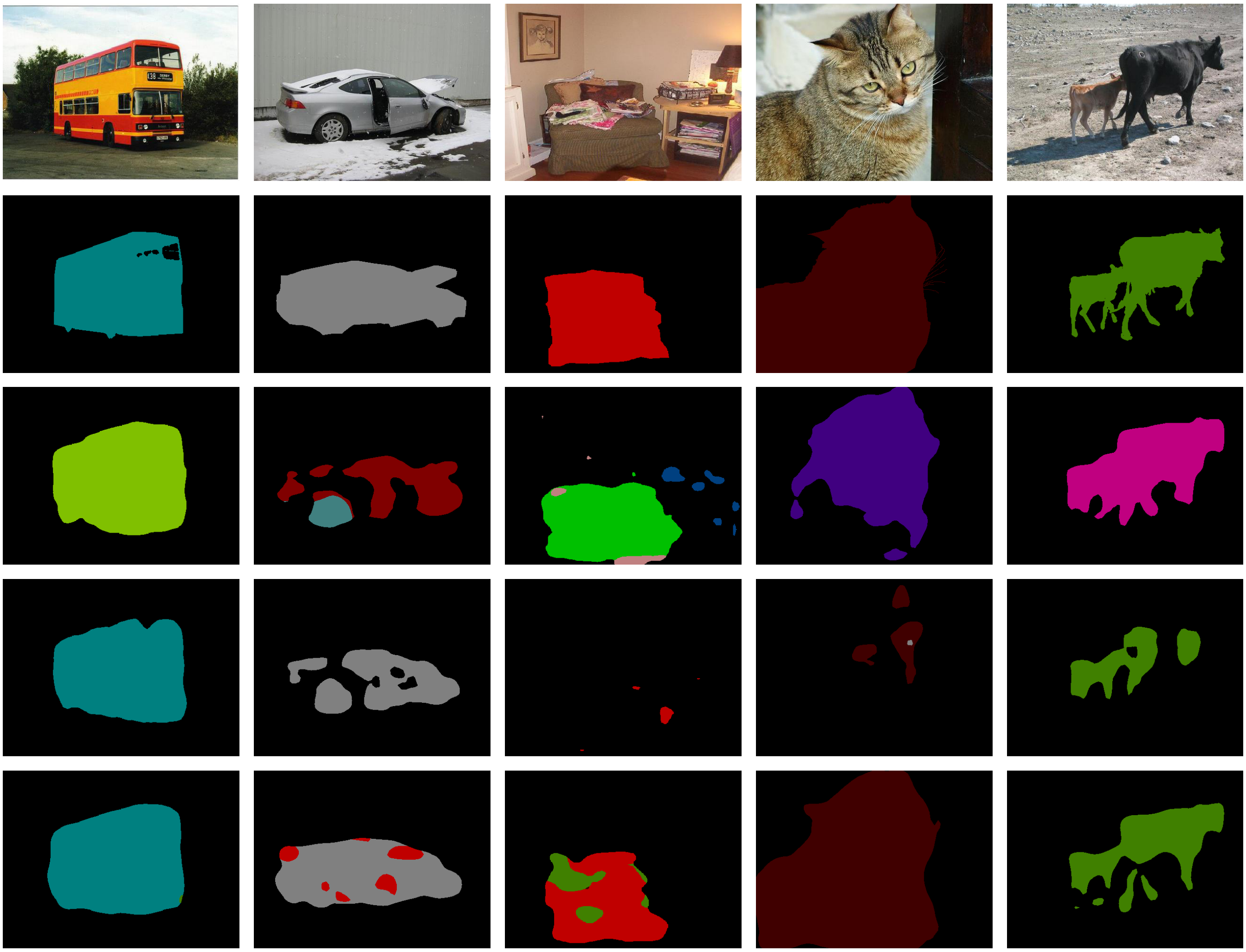


FIGURE 5. Schematic diagram of semantic segmentation results on split-1.The first row shows the original input images.The second row shows the ground truths. The third and fourth rows show the predicted results of Seg-JT and Seg-FT, respectively. The last row shows the predicted results of Meta-seg.

Seg-FT can not perform well sometimes. As shown in the fourth row in Fig 5, Seg-FT fails to distinguish chair and cat from background. This indicates the difficulty of few-shot segmentation. Although Seg-JT and Seg-FT can get strong prior knowledge from the meta-training classes, they can only acquire limited information from few meta-test examples. The proposed Meta-Seg can learn good initialization and update strategy from meta-training classes. Therefore, Meta-Seg works better on meta-test classes.

The 5-shot performance of Meta-Seg is evaluated using the trained 1-shot Meta-Seg, shown in Table 3. Although slight performance gain is obtained by Meta-Seg in 5-shot test setting, it also performs better than the two baselines.

The experimental results indicate that the performance gap is due to the intrinsic characteristics of the parameter optimization of Meta-Seg.

E. QUALITATIVE COMPARISON

There are some relevant works based on two-branch structure, which mainly focus on single-class for each forward computation. This is inconvenient and ineffective for multi-class few-shot semantic segmentation. Besides, the evaluation of single-class semantic segmentation is easier than our multi-class semantic segmentation in principle. As mentioned in baseline (Section V-A), the data organization is different between Meta-Seg and the two-branch structure based models. Herein, we provide the results of the two-branch structure based models in Table 4 just for qualitative comparison. Meta-Seg can also yield a competitive results with the single-class semantic segmentation approaches.

F. COMPUTATIONAL OVERHEAD

The number of parameters, run time and memory usage of Meta-Seg and FCN8s, shown in Table 5, are obtained by experiments to analyze the computational overhead of Meta-Seg. The training and test times are the average of 100 runs for every row in Table 5. We set batch size as 1 for FCN8S. We set the batch size of task as 1 and sample 1 image for training and test respectively in each task for Meta-Seg.

TABLE 4. Semantic segmentation results of Meta-seg and other state-of-art methods. The numbers in the table indicate the mIOU (%) on different folds.

Method	split-0	split-1	split-2	split-3	mean
1-NN-1	25.3	44.9	41.7	18.4	32.6
LogReg-1	26.9	42.9	37.1	18.4	31.4
Siamese-1	28.1	39.9	31.8	25.8	31.4
OSLSM-1	33.6	55.3	40.9	33.5	40.8
Meta-Seg-1	**42.2**	**59.6**	**48.1**	**44.4**	**48.6**
1-NN-5	34.5	53.0	46.9	25.6	40.0
LogReg-5	35.9	51.6	44.5	25.6	39.3
OSLSM-5	35.9	58.1	42.7	39.1	43.9
Meta-Seg-5	**43.1**	**62.5**	**49.9**	**45.3**	**50.2**

TABLE 5. The comparison of computational overhead for Meta-Seg and FCN8S.

Model	Para (M)	Times (ms)	Memory Usage (MiB)
Meta-Seg Training	268.54	143.27	5721
Meta-Seg Test	268.54	24.63	1971
FCN8S Training	134.30	80.11	3445
FCN8S Test	134.30	23.70	1575

The number of parameters of Meta-Seg is twice as much as that of FCN8S for training and test, because the learnable learning rate α^* is set for each parameter of FCN8S. Therefore, Meta-Seg requires more memory. Specifically, for the training of Meta-Seg, it requires 5217MiB memory and more time to update α^* and the parameters of FCN8S. While for the test of Meta-Seg, it just requires 1971MiB memory and the same time as the test of FCN8S whose α^* needn't to be updated. From Table 5, we can conclude that the training of Meta-Seg has a higher computational overhead due to the update of α^*, which should be solved in the future. But Meta-Seg can perform few-shot semantic segmentation as efficient as FCN8S.

G. FAST ADAPTATION

In this subsection, we expatiate the fast adaptation ability of Meta-Seg, which is another advantage in addition to the multi-class few-shot semantic segmentation. With the help of the learned parameter update strategy, Meta-Seg can converge quickly. As shown in Fig. 6, Meta-Seg can yield better performance with just one update iteration. While both Seg-JT and Seg-FT have worse performance even trained for thousands of iterations. The very fast adaptation ability of Meta-Seg means that it can receive scarcely labeled novel classes in any time and process the test examples of novel classes in real time, which makes Meta-Seg become a perfect few-shot semantic segmentation model in practice. Furthermore, the fast adaptation ability enables Meta-Seg to learn knowledge in a life-long learning manner.

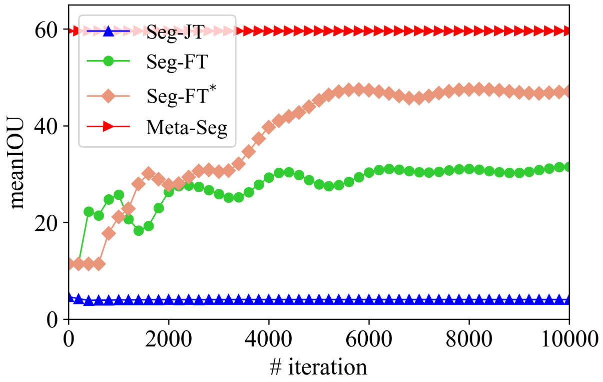


FIGURE 6. Adaptation speed comparison between the proposed Meta-Seg and the two baselines (Seg-JT, Seg-FT). Besides, to further analyze the effectiveness of the initialization learned by Meta-Seg, we provide the adaptation process of Seg-FT*, which means that fine-tuning the segmentation network on meta-test classes with the initialization parameters of Meta-Seg learned on meta-training classes.

H. FINE-TUNING WITH THE INITIALIZATION OF META-SEG

Meta-Seg can learn a good initialization from the distribution of few-shot semantic segmentation tasks. To further verify the effectiveness of the learned good initialization on meta-training classes, we fine-tune the semantic segmentation model on meta-test classes based on the learned good initialization. This fine-tuning process is shown as Seg-FT* in Fig. 6. Obviously, Seg-FT* yields better performance than Seg-FT, indicating the effectiveness of the learned good initialization. This also opens up Meta-Seg's potential in supervised-learning semantic segmentation, which means that the supervised-learning models may converge better and faster by leveraging the good initialization of Meta-Seg.

I. VISUALIZATION OF UPDATE STRATEGY

To further analyze the update strategy, e.g. the learnable learning rate α^*, learned by Meta-Seg, we compute the arithmetic mean of α^* along the output channel, then visualize them from the first 64 input channels in each layer in Fig. 7 (a). Besides, the values of α^* in layer 0 to 13 and layer 15 to 16 are visualized in smaller ranges respectively in Fig. 7 (b), (c) for better visualization.

Obviously, the learnable learning rate α^* vary from channel to channel. The parameters of the base feature extractor learn the general knowledge from different tasks, such that they have small learnable learning rate values (see the light green areas in Fig. 7 (a)). Layer 14 increases the dimensions, which may be different for different tasks. Similarly, the dimensionality reduction and classification are executed in layer 17, which are sensitive to the change of tasks. Hence, the values of α^* in layer 14 and layer 17 change tempestuously in larger range. While the values of α^* in layer 15 and 16 are small, the underlying reason may be that the two layers execute linear calculation in the same dimensions. Even though the values of α^* are smaller in the light green areas in Fig. 7, they have complicated distributions, which are difficult to set

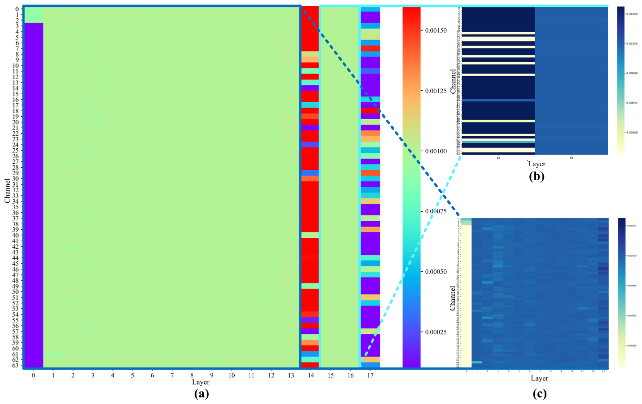


FIGURE 7. Visualization of update strategy. In Meta-Seg, the meta-learner learns an update strategy (that is the learnable learning rate α^*) for each parameter of the base-learner. The learnable learning rates of the first 64 channels from each layer are visualized.

TABLE 6. The performance of Meta-Seg with different initialized learnable learning rate α^* and learning rate β.

α^*	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Meta-Seg	29.4	**59.6**	56.6	56.9
β	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Meta-Seg	57.3	**59.6**	58.1	58.4

manually for different few-shot tasks. This can further prove the effectiveness of the learned update strategy in Meta-Seg.

J. THE INITIALIZATION OF LEARNING RATE

In this subsection, we analyze the effect of different initialized α^* and β on the performance of Meta-Seg respectively, as shown in Table 6. The experiment result indicates that the performance of Meta-Seg are sensitive to the initialized value of α^*. However, the experiment results demonstrate relatively little effect of different leaning rate β on the performance of Meta-Seg. This is because that the values of α^* can only be adjusted around the initialized value. The small values of α^* may cause under-fitting, such as 10^{-4} and 10^{-5} in Table 6. While larger value of α^*, e.g. 10^{-2} can cause over-fitting. Therefore, an appropriate initialized value of α^* is important for Meta-Seg and this should be set as the case may be.

K. ANALYSIS

Few-shot semantic segmentation has attracted much attention, while most of the previous works are based on two-branch structure based framework and execute single-class semantic segmentation. These methods train and test models with many support-query image pairs, this is inconvenient and ineffective for multi-class segmentation. This work solves the few-shot semantic segmentation problem from the perspective of meta-learning and opens a new door for this research direction. The experiment results prove the effectiveness of Meta-Seg and meta-learning is a promising approach for few-shot semantic segmentation. In addition, we just implement Meta-Seg by combining the meta-learning pipeline with FCN8s in this work. It should be noted that any successful semantic segmentation models can be integrated into Meta-Seg. This means that the performance of Meta-Seg can be further improved with the development of supervised-learning semantic segmentation.

VI. CONCLUSION

This work proposes to solve the few-shot semantic segmentation problem via meta-learning. We design a generalized few-shot semantic segmentation framework named Meta-Seg, which consists of a meta-learner and a base-leaner. In this work, FCN8s is integrated into Meta-Seg. In theory, any supervised-learning semantic segmentation models can be embedded into Meta-Seg. The proposed Meta-Seg can learn a good initialization and a parameter update strategy from the distribution of few-shot semantic segmentation tasks on meta-training classes. After trained, Meta-Seg can implement fast parameter adaptation with few training examples on meta-test classes. The experiment results prove the effectiveness of Meta-Seg. Besides, Meta-Seg can segment multi-class objects efficiently than the previous two-branch structure based models. Although this work proposes a new method of few-shot semantic segmentation, more work should be done to improve the semantic segmentation network architecture and solve the problem of high graphic memory requirement.

ACKNOWLEDGMENT

(Tengfei Zhang and Xian Sun contributed equally to this work.)

REFERENCES

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[3] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," 2016, *arXiv:1606.01781*. [Online]. Available: https://arxiv.org/abs/1606.01781

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1–9.

[6] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2014, pp. 1–16.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21–37.

[10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.

[11] R. B. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[13] W. Siyu, G. Xin, S. Hao, Z. Xinwei, and S. Xian, "An aircraft detection method based on convolutional neural networks in high-resolution SAR images," *J. Radars*, vol. 6, no. 2, pp. 195–203, 2017.

[14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

[15] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4151–4160.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[17] Z. Yan, M. Yan, S. Hao, K. Fu, and S. Xian, "Cloud and cloud shadow detection using multilevel feature fused segmentation network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1600–1604, Oct. 2018.

[18] X. Gao, X. Sun, Y. Zhang, M. Yan, G. Xu, H. Sun, J. Jiao, and K. Fu, "An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network," *IEEE Access*, vol. 6, pp. 39401–39414, 2018.

[19] S. Minaee and Y. Wang, "Masked signal decomposition using subspace representation and its applications," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017.

[20] S. Minaee and Y. Wang, "Screen content image segmentation using sparse decomposition and total variation minimization," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 3882–3886.

[21] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1520–1528.

[22] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[23] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei, "What's the point: Semantic segmentation with point supervision," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 549–565.

[24] J. Dai, K. He, and J. Sun, "BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proc. Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1635–1643.

[25] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3159–3167.

[26] G. Papandreou, L.-C. Chen, K. Murphy, and A. L. Yuille, "Weakly- and semi-supervised learning of a DCNN for semantic image segmentation," 2015, *arXiv:1502.02734*. [Online]. Available: https://arxiv.org/abs/1502.02734

[27] J. Xu, A. G. Schwing, and R. Urtasun, "Learning to segment under various forms of weak supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3781–3790.

[28] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, vol. 70, Aug. 2017, pp. 1126–1135.

[29] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2017, pp. 1–11.

[30] Y.-X. Wang and M. Hebert, "Learning to Learn: Model regression networks for easy small sample learning," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 616–634.

[31] T. Munkhdalai and H. Yu, "Meta networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, vol. 70, Aug. 2017, pp. 2554–2563.

[32] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 3630–3638.

[33] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 1199–1208.

[34] M. Ren, S. Ravi, E. Triantafillou, J. Snell, K. Swersky, J. B. Tenenbaum, H. Larochelle, and R. S. Zemel, "Meta-learning for semi-supervised few-shot classification," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2018, pp. 1–15.

[35] A. A. Rusu, D. Rao, J. Sygnowski, O. Vinyals, R. Pascanu, S. Osindero, and R. Hadsell, "Meta-learning with latent embedding optimization," 2018, *arXiv:1807.05960*. [Online]. Available: https://arxiv.org/abs/1807.05960

[36] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," 2018, *arXiv:1803.02999*. [Online]. Available: https://arxiv.org/abs/1803.02999

[37] T. Munkhdalai and A. Trischler, "Metalearning with Hebbian fast weights," 2018, *arXiv:1807.05076*. [Online]. Available: https://arxiv.org/abs/1807.05076

[38] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 4077–4087.

[39] Y. Wang, X.-M. Wu, Q. Li, J. Gu, W. Xiang, L. Zhang, and V. O. K. Li, "Large margin few-shot learning," 2018, *arXiv:1807.02872*. [Online]. Available: https://arxiv.org/abs/1807.02872

[40] Z. Li, F. Zhou, F. Chen, and H. Li, "Meta-SGD: Learning to learn quickly for few-shot learning," 2017, *arXiv:1707.09835*. [Online]. Available: https://arxiv.org/abs/1707.09835

[41] H. Chen, Y. Wang, G. Wang, and Y. Qiao, "LSTD: A low-shot transfer detector for object detection," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2836–2843.

[42] K. Fu, T. Zhang, Y. Zhang, M. Yan, Z. Chang, Z. Zhang, and X. Sun, "Meta-SSD: Towards fast adaptation for few-shot object detection with meta-learning," *IEEE Access*, vol. 7, pp. 77597–77606, 2019.

[43] A. Shaban, S. Bansal, Z. Liu, I. Essa, and B. Boots, "One-shot learning for semantic segmentation," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2017, pp. 1–17.

[44] K. Rakelly, E. Shelhamer, T. Darrell, A. Efros, and S. Levine, "Conditional networks for few-shot semantic segmentation," in *Proc. ICLR*, Feb. 2018, pp. 1–4.

[45] T. Hu, P. Yang, Z. Chiliang, G. Yu, Y. Mu, and C. Snoek, "Attention-based multi-context guiding for few-shot semantic segmentation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 8441–8448.

[46] K. Rakelly, E. Shelhamer, T. Darrell, A. A. Efros, and S. Levine, "Few-shot segmentation propagation with guided networks," 2018, *arXiv:1806.07373*. [Online]. Available: https://arxiv.org/abs/1806.07373

[47] N. Dong and E. P. Xing, "Few-shot semantic segmentation with prototype learning," in *Proc. BMVC*, 2018, p. 6.

[48] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Representations*, 2016, pp. 1–13.

[49] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2881–2890.

[50] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with Atrous separable convolution for semantic image segmentation.," in *Proc. Eur. Comput. Vis. Pattern Recognit.*, Sep. 2018, pp. 801–818.

[51] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3684–3692.

[52] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation," 2019, *arXiv:1903.11816*. [Online]. Available: https://arxiv.org/abs/1903.11816

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2015, pp. 234–241.

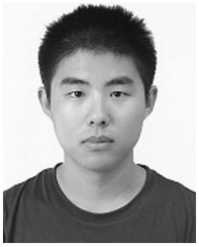
[54] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[55] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1925–1934.

[56] S. Thrun and L. Pratt, "Learning to learn: Introduction and overview," in *Learning to Learn*. Springer, 1998, pp. 3–17.

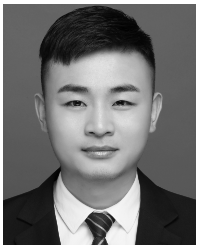
[57] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pyTorch," in *Proc. NIPS Workshop*, 2017, pp. 1–4.

ZHIYING CAO received the B.Sc. degree from Xidian University, Xi'an, China, in 2016. He is currently pursuing the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision, pattern recognition, and remote sensing image processing, especially on semantic segmentation.



TENGFEI ZHANG received the B.Sc. degree from the Ocean University of China, Qingdao, China, in 2016. He is currently pursuing the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision, pattern recognition, and remote sensing image processing, especially on object detection.

WENHUI DIAO received the B.Sc. degree from Xidian University, Xi'an, China, in 2011, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2016. He is currently an Assistant Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include computer vision and remote sensing image analysis.



YUE ZHANG (M'18) received the B.E. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 2012, and the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2017. He is currently an Assistant Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include the analysis of optical and synthetic aperture radar remote sensing images.

XIAODE LYU received the B.Sc. degree from the Hebei University of Technology, Tianjin, China, in 1991, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 1997. He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include array signal processing and radar systems.

KUN FU received the B.Sc., M.Sc., and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 1995, 1999, and 2002, respectively.

He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision, remote sensing image understanding, geospatial data mining, and visualization.

XIAN SUN received the B.Sc. degree from Beihang University, Beijing, China, in 2004, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2006 and 2009, respectively. He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include computer vision and remote sensing image understanding.