SIA-MAML: A Novel Method for Few-Shot Classification

Xinrui Huang, Yuze Chen, Huxiao Ji

December 18, 2020

1 Introduction

In recent years, classical deep learning models have achieved remarkable success on many computer vision and image understanding tasks. Meanwhile, classical deep learning models often requires a large number of samples. Once the data set is small, the performance of the system will be greatly reduced. Unfortunately, if we want to further improve the performance by using deeper neural network, more labeled data in richer categories is required. Besides, the classical deep learning models have the problem of overfitting with limited labeled data. However, the labeled data is usually expensive. To resolve this problem, few-shot learning comes into being. Few-shot classification aims to identify new categories with only a few labeled images in each category. Existing metric-based few-shot classification algorithms predict the category by comparing the feature embedding of the query images with the feature embedding of a few labeled images. Although these methods have been proved to have good performance, due to the large differences in the feature distribution between domains, these methods often cannot be extended to invisible domains. In order to alleviate the problem of domain migration, many unsupervised domain adaptation techniques have been proposed. These methods focus on adapting classifiers of the same category from the source domain to the target domain. However, the unsupervised domain adaptation method assumes that there are a large number of unlabeled images in the target domain during the training process. In many cases, this assumption may be impractical. To resolve this problem, we propose a novel few-shot learning method termed as Semantic Information Assists Model-Agnostic Meta-Learning (SIA-MAML). Our method consists of an image feature extrator and a semanic information mapping module. During the training stage, we use Model-Agnostic Meta-Learning [2] (MAML) to obtain good pre-train parameters of the model. Unlike traditional MAML algorithm, we use not only the images in the support set, but also the embeddings of categories itself. The details of the method will be discussed in section 3. To evaluate the effectiveness of the proposed method, extensive experiments are conducted on two widely-used Mini-ImageNet benchmarks. In section 4, the results of experiment demonstrate the encouraging performance of proposed method compared with the state-of-the-art approaches. The main contributions of this study can be summarized as follows. (1) We propose a novel Semantic Information Assists Model-Agnostic Meta-Learning (SIA-MAML) method with feature extrator module and semanic information mapping module. (2) We demonstrate a solution to optimize MAML by adding the embeddings of categories as additional input.

2 Related work

Few-Shot Learning The study of few-shot learning has been of interest for some time, Earlier work on few-shot learning can be roughly divided into three categories: optimization-based (parametric) method [2, 6], metric based (non-parametric) method [14, 12, 16] and some other method [11, 13]. Optimization-based methods optimize a learner to perform wellafter fine-tuning on the task data done by a single (or few) step(s) of Gradient Descent. Metric based methods first use a deep embedding model to extract some high-level feature from an image, and compare the distances between samples from query set and samples from support set to classify the images.

Meta-Learning Meta-learning is used to learn the models, namely, learn to learn. Compared with making judgments on new samples from a task by learning samples, the goal of meta-learning can be seen as treating tasks as samples, enabling the meta-learner to learn on a large number of different tasks with appropriate speed and accuracy by learning several tasks.

Model-Agnostic Meta-Learning MAML is the method that can learn the parameters of any standard model via meta-learning in such a way as to prepare that model for fast adaptation. The intuition behind this approach is that some internal representations are more transferrable than others. We can train the model's initial parameters such that the model has maximal performance on a new task after the parameters have been updated through one or more gradient steps computed with a small amount of data from that new task. [2] discusses specific instantiations of MAML algorithm for supervised learning and reinforcement learning. These two domains use different loss function and data formats, but the same adaption mechanism can be applied in both cases.

Label Embedding In computer vision, a vast amount of work has been devoted to input embedding. This includes works on patch encoding, on kernel-based methods [10] with a recent focus on explicit embeddings [5], on dimensionality reduction [10] and on compression [4]. Comparatively, much less work has been devoted to label embedding. Provided that the embedding function ϕ is chosen correctly, label embedding can be an effective way to share parameters between classes. Consequently, the main applications have been multiclass classification with many classes [15] and zero-shot learning [7].

Meta-transfer Learning [11] propose a novel few-shot learning method called meta-transfer learning (MTL) which combines meta learning and transfer learning to adapt a deep NN for few shot learning tasks. MTL also introduces the hard task (HT) meta-batch scheme as an effective learning curriculum.

Feature-wise Transformation [13] forces on the issue that existing metric-based approaches often do not generalize well to categories from different domains. They design a learned feature-wise transformation layer to simulate various distributions of image features during training stage, in order to improve the generalization ability of the metric function in the testing phase.

3 Method

Our Semantic Information Assists Model-Agnostic Meta-Learning (SIA-MAML) method has three key points which we describe in the following subsections. First, we introduce the feature extractor used in our method to obtain image features. Second, we elaborate the semantic information mapping module which can transform the semantic space to hidden space and combine image features and categories features to get hidden features. Third, we describe the optimization procedure that we use to optimize parameters in feature extractor and semantic information mapping module. The overview of SIA-MAML is shown in Figure 1.

3.1 Preliminary

The goal of few-shot classification is to learn an image classification model by using the support set $D_{support} = \{(x_i, y_i)\ldots\}$ and generalize to the query set $D_{query} = \{(x_i, y_i)\ldots\}$ well. We set the few-shot learning problem as N-way K-shot learning, which means each support set consists of N categories and K labeled examples. We call a support set and a query set as a task $T_i = \langle D_{support}, D_{query}\rangle$.
MAML is an optimization based algorithm. It aims to obtain good pre-train parameters of the model. In the meta-learning scenario, we consider a distribution over tasks $p(\mathcal{T})$ that we want our model to be able to adapt to. During meta-training, a task T_i is sampled from $p(\mathcal{T})$, the model is trained with $N \times K$ samples in support set and feedback from the corresponding loss $\mathcal{L}_{T_i}$ from T_i and then tested

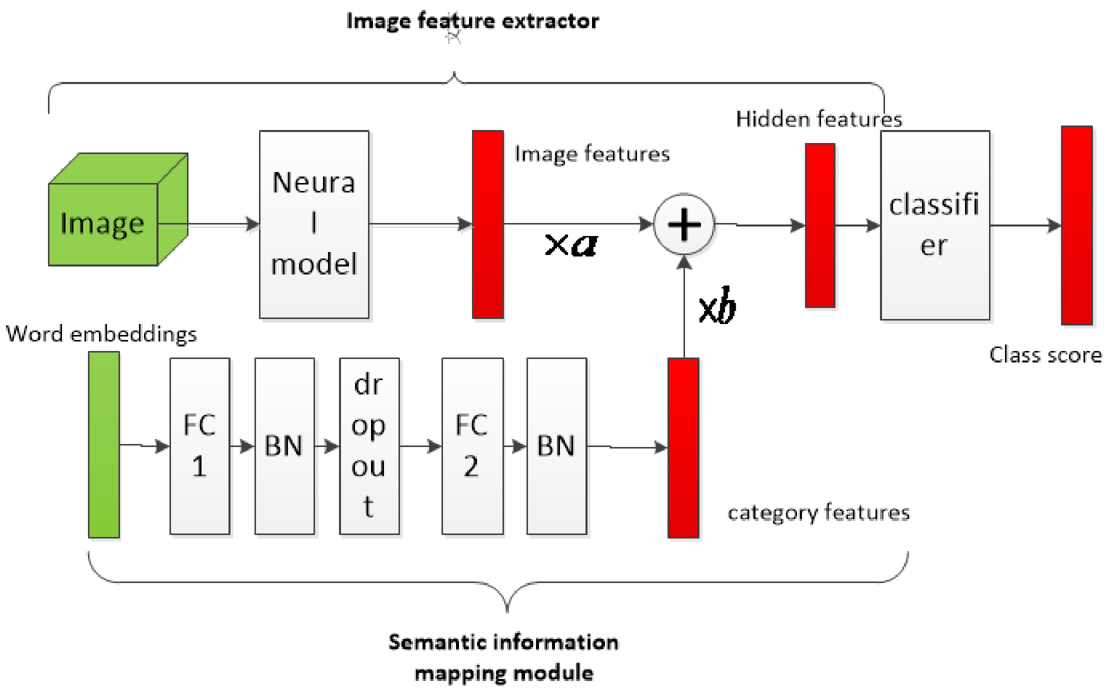


Figure 1: SIA-MAML

on new samples in query set from T_i. In effect, the test error on sampled tasks T_i serves as the training error of the meta-learning process. While our method (SIA-MAML, denoted as$M_{\phi,\theta}(X,C)$) not only use images in support set, but also fuse the embeddings of categories itself during meta-training. The training error is only calculated with images in T_i as MAML. At the end of meta-training, new tasks are sampled from $p(\mathcal{T})$, and meta-performance is measured by the model's performance after learning from K samples. Generally, tasks used for meta-testing are held out during meta-training.

3.2 Image features extractor

SIA-MAML is a model-agnostic algorithm as MAML. Any neural type model which can extract features from input images and construct hidden features can be used in our SIA-MAML. In the paper, we construct a simple convolutional model for image features extracting and follows the same architecture as the embedding function used by [14], which has 4 modules with a 3×3 convolutions and 64 filters, followed by batch normalization [3], a ReLU nonlinearity, and 2×2 max-pooling. All images are resize to $84 \times 84 \times 3$ when they are sent to the image features extractor. The image features extractor is denoted as f_θ. The image features is denoted as $f_\theta(X)$.

3.3 Semantic information mapping

In few-shot image classification task, the category names can usually be known in advance. These are very important prior knowledge which is helpful to improve the accuracy of image classification. We use [8] to obtain 50-dim word embeddings ($E(C_i)$) of categories in each task T_i. The origin embeddings exist in semantic space while image features are in the hidden space which can be used as classifier's input. The semantic information mapping module can transform semantic space to hidden space the same as image features. Following formula shows this process.

$$f_\phi(C) = f_\phi(E(c_1), E(c_2)\ldots, E(c_n))$$

The module consists of two fully connected layers followed by batch normalization and a ReLU non-linearity. In order to prevent overfitting, a dropout layer is added between first fully connected layer

and second fully connected layer. The construct of semantic information mapping module is shown in figure.
In the fusion of image features and semantic features to obtain hidden features ($H(X, C)$), two weight coefficients (a, b) are added to original features. We expect semantic information can be the auxiliary of image information, so the coefficient of semantic features needs to be reduced. Following formula shows the fusing process.

$$H(X, C) = af_{\theta}(X) + bf_{\phi}(C)$$

$$a + b = 1.0$$

3.4 Optimization algorithm

During training process, we should optimize the parameters in both image features extractor and semantic information mapping module. When we train on $D_{support}$, the input of SIA-MAML has two parts including image x_i and category embedding$E(c_i)$. The updating process is as follows:

$$\left\langle \theta_i^{'}, \phi_i^{'} \right\rangle = \langle \theta_i, \phi_i \rangle - \alpha \nabla \phi, \theta \mathcal{L}_{T_i}\left(M_{\phi,\theta}\left(X_i, C_i\right)\right)$$

The addition of semantic information can help us get a better initial image feature extractor. However, when we evaluate or test on D_{query}, the input only consists of images. The meta-objective is as follows:

$$\min_{\theta,\phi} \sum_{T_i \sim p(\mathcal{T})} \mathcal{L}_{T_i}\left(M_{\theta_i^{'}}(X_i)\right)$$

For discrete classification tasks with a cross-entropy loss, the loss takes the form:

$$\mathcal{L}_{T_i}\left(M_{\phi,\theta}\left(X_i\right)\right) = Y_i \log\left(M_{\phi,\theta}\left(X_i\right)\right) + \left(1 - Y_i\right) \log\left(1 - M_{\phi,\theta}\left(X_i\right)\right)$$

$$\mathcal{L}_{T_i}\left(M_{\theta}\left(X_i\right)\right) = Y_i \log\left(M_{\theta}\left(X_i\right)\right) + \left(1 - Y_i\right) \log\left(1 - M_{\theta}\left(X_i\right)\right)$$

The optimization algorithm details in Algorithm 1.

Algorithm 1: SIA-MAML for Few-Shot Supervised Learning

Input: $p(\mathcal{T})$: distribution over tasks
Input: α, β: step size hyperparameters
Randomly initialize θ and ϕ;
while *not done* **do**
    Sample batch of tasks $T_i \sim p(\mathcal{T})$ **for** *all* T_i **do**
        Sample K datapoints for each category in the sampled task, form the support set $D^i_{support} = \{(x_1, y_1) \ldots (x_k, y_1)(x_1, y_2) \ldots (x_k, y_2) \ldots\}$ from T_i;
        Evaluate $\nabla_{\phi,\theta}\mathcal{L}_{T_i}(M_{\phi,\theta}(X_i, C_i))$ using $D_{support}$ and cross-entropy $\mathcal{L}_{T_i}$;
        Compute adapted parameters with gradient decent:

$$\left\langle \theta_i^{'}, \phi_i^{'} \right\rangle = \langle \theta_i, \phi_i \rangle - \alpha \nabla \phi, \theta \mathcal{L}_{T_i}\left(M_{\phi,\theta}\left(X_i, C_i\right)\right)$$

        Sample datapoints $D^i_{query} = \{(x_1, y_1) \ldots\}$ from T_i for the meta-update
    end
    Update $\langle \theta, \phi \rangle = \langle \theta, \phi \rangle - \alpha \nabla \phi, \theta \sum_{T_i \sim p(\mathcal{T})} \mathcal{L}_{T_i}\left(M_{\theta_i^{'},\phi_i^{'}}(X_i)\right)$ using each D^i_{query} and cross-entropy $\mathcal{L}_{T_i}$;
end

Table 1: Comparison of different obfuscations in terms of their transformation capabilities

Method	5-Way 1-Shot Acc.	5-Way 5-Shot Acc.
Matching Nets [14]	43.56%	55.31%
Meta-learn Lstm [9]	43.44%	60.60%
Maml[1] [2]	45.16%	59.94%
Sia-Maml	**46.53%**	**60.86%**

4 Experiment

All of the experiments were performed using Tensorflow [1], which allows for automatic differentiation through the gradient updates during meta-learning. The hard ware setting is a NVIDIA RTX 2080 GPU.

4.1 Dataset

We use MiniImagenet dataset for training and testing. The MiniImagenet dataset was proposed by [9] and involves 64 training classes, 12 validation classes, and 24 test classes. This is a most common used few-shot learning benchmarks. We perform 5-way 1-shot and 5-way 5-shot learning on the MiniImagenet dataset. We use the Glove as word embeddings generator model to generate 50-dim word embeddings for each category.

4.2 Baselines

To compare the related works, we choose some famous model as baselines, for instance, fine-tuning baseline, nearest neighbor baseline, matching nets [14], meta-learner LSTM [9] and MAML [2].

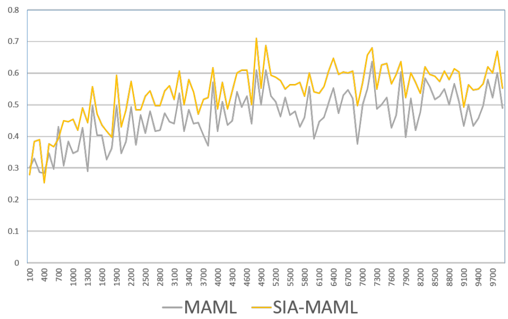
4.3 Evaluation metrics

We mainly evaluate models according to their accuracy. Due to the limitation of computation source, our training data cannot be set to the same amount as the baseline method, but our model still achieves impressive results. We present the training data, iteration number and parameter number of each model in experiment. At the same time, we also show the relationship between the iteration times and the verification accuracy of the model in the training process, which is used to illustrate the convergence rate of the model.

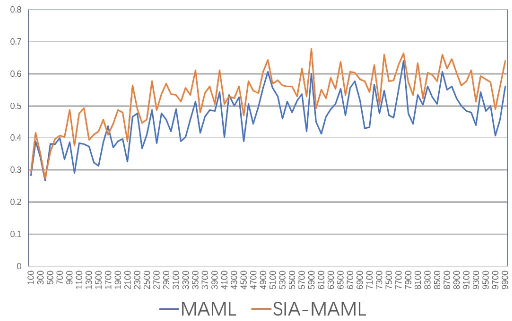
4.4 Experimental results

This section discusses the results of the proposed method.
First, experiments are conducted on the Mini-ImageNet dataset. The proposed SIA-MAML is compared with Matching Nets [14], Meta-learn Lstm [9] and Maml [2]. The compared results are demonstrated in Table 3. Experiments are independently repeated 600 times and the testing data is randomly sampled, the average results are reported. It can be seen that the proposed SIA-MAML achieves the best performance for 5-way 1-shot and 5-way 5-shot image recognition tasks. Specifically, it improved the performance of traditional MAML. This result indicates that it is effective in Semantic information mapping, especially when the number of examples in the support set is very small. It can well verify the motivation of the proposed method.

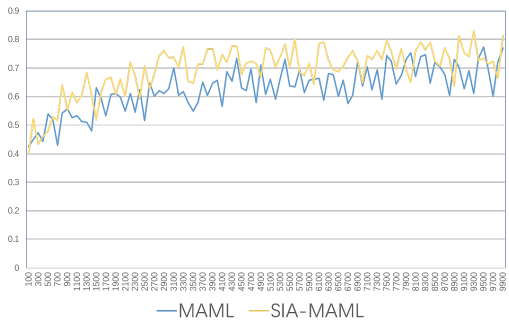
[1]Experiments are performed on our own machine, setting the task count 200000 in the origin public code as 20000 and the iteration times 60000 in the origin public code as 10000. Meawhile, our SIA-MAML method is also performed under same task count and iteration times.



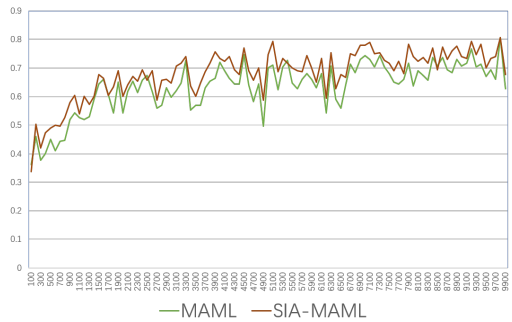
(a) 5-Way, 1-Shot with 2000 tasks



(b) 5-Way, 1-Shot with 20000 tasks



(c) 5-Way, 5-Shot with 2000 tasks



(d) 5-Way, 5-Shot with 20000 tasks

Figure 2: Training accuracy of MAML and SIA-MAML during the training process

4.5 Experimental observations

This section we show the training accuracy of MAML and SIA-MAML during the training process. In Figure 2 we display the result in the training iteration, which shows that the SIA-MAML outperforms MAML significantly in the query set from training set.

5 Conclusion

In order to obtain better generalization ability in few-shot image classification, we introduce a meta-learning method which combines prior semantic knowledge of category names in support set when meta-learning to get better initial parameters in model. We also propose a new optimization process to optimize two parts of parameters in image features extractor and semantic information mapping module. Our method has a number of benefits. It is model-agnostic and can be used in any optimization-based neural model. It makes full use of the semantic information of categories and provides a priori knowledge for few-shot learning or even zero-shot learning. For the involvement of semantic information, we can use smaller training set during meta-learning and achieve same results faster. This is useful for tasks where adequate training data is difficult to obtain. We believe that this work is one step toward a simple and general-purpose meta-learning technique that can be applied to any problem and any model.

References

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *CoRR*, abs/1703.03400, 2017.

[3] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[4] Herve Jegou, Matthijs Douze, and Cordelia Schmid. Product quantization for nearest neighbor search. *IEEE transactions on pattern analysis and machine intelligence*, 33(1):117–128, 2010.

[5] Subhransu Maji and Alexander C Berg. Max-margin additive classifiers for detection. In *2009 IEEE 12th International Conference on Computer Vision*, pages 40–47. IEEE, 2009.

[6] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *CoRR*, abs/1803.02999, 2018.

[7] Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. Zero-shot learning with semantic output codes. *Advances in neural information processing systems*, 22:1410–1418, 2009.

[8] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[9] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. 2016.

[10] John Shawe-Taylor, Nello Cristianini, et al. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[11] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[12] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H.S. Torr, and Timothy M. Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[13] Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation, 2020.

[14] Oriol Vinyals, Charles Blundell, Timothy P. Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. *CoRR*, abs/1606.04080, 2016.

[15] Kilian Q Weinberger and Olivier Chapelle. Large margin taxonomy embedding for document categorization. *Advances in Neural Information Processing Systems*, 21:1737–1744, 2008.

[16] Ling Yang, Liangliang Li, Zilun Zhang, Xinyu Zhou, Erjin Zhou, and Yu Liu. Dpgn: Distribution propagation graph network for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.