NATIONAL UNIVERSITY OF SINGAPORE

Summary of Readings in Few Shot Learning

Xiao Junbin (xiaojunbin@u.nus.edu)

1 INTRODUCTION

Few shot learning aims to learn a model to predict the unseen categories, conditioned with very limited training samples corresponding to each category [15]. It is of increasing importance since it alleviates the annotation burdens embarrassing the vast majority of established deep models. It has also earned extensive attention in recent years. This article thus gives a brief summary to recent development in few-shot learning, and specifically in its application for image recognition. Meanwhile, some ideas are described at the end of the writing.

One promising approach to few-shot learning is the meta-learning paradigm where transferable knowledge is distilled from a collection of tasks to prevent over-fitting and to promote fast generalization on new tasks [1]. To my knowledge, existing meta-learning instances encompass: 1) model initialization based methods, 2) pseudo sample based methods, 3) metric learning based methods, and 4) others like [3, 4] which put emphasis on designing network architectures tailored for specific applications. Related works are summarized below.

Model initialization. Methods in this class aim to learn a good initial condition (set of neural network weights) that is capable of achieving rapid adaption to novel categories with inadequate labeled examples. E.g., Finn *et al.*[5] hypothesis that there should be a neural work that can learn latent knowledge shared by different tasks, and they take an explicit approach to this idea by sampling many distinct target problems from a multiple task training set, and optimizing the base neural network model in the process of solving these problems. Specifically, the success at each target problem after fine-tuning drives updates in the base model, and finally encourages the production of an easy to fine-tune model initialization. This kinds of learning methods are model-agnostic and thus appeal to extensive applications.

Pseudo samples. Data augmentation has already been shown a simple but effective approach to deep models [11]. By generating additional training samples, the over-fitting problem can be alleviated to a certain extent. In few-shot learning tasks, it is of particular significance to devise auxiliary data from the very limited supporting samples [2, 7, 9, 17]. In the work of [7], the authors propose to hallucinate data by learning category-invariant feature transformation, and the added data brings considerable boost in performance. More recently, Wang *et al.*[17] employ generators to learn model-specific data for online training, through which the data can also be optimized according to the backward gradients. Besides, other works using adversarial training like [2] also achieve promising results.

Metric learning based methods think outside the fine-tuning box by searching the supporting set to find the images mostly similar to the queries, and then propagate their categories to the query images. It follows such an intuition to learn feature representations that preserve the class neighbor structure (i.e., samples of the same category are closer than samples of different categories in the feature space). For example, Koch *et al.*[10] directly trains a Siamese neural network by employing the sigmoid cross-entropy loss on top of the L1-norm, to maximize the distances between samples from different categories and minimize those from same category. For testing, each sample will be fed into the learned model along with the supporting samples, to decide whether they belong to the same category or not. It is noteworthy that the samples in [10] are encoded independently, and share nothing among categories. In contrast, Vinyals *et al.*[16] propose Matching Networks which employs a differentiable nearest neighbor classifier implemented with an attention LSTM [8] over the representations of the training images. The training features share information among different categories through a bi-directional LSTM. By sharing information, the model is better to extract discriminative representation of a particular category. Specially, the authors also proposed "episodic training" strategy which mimics the few-shot problem by dividing the training on base categories into a collection of sub-tasks consisting of sub-classes as well as sub-samples. This learning strategy contributed to remarkable enhancement over naive training and was widely adopted in later research works [6, 13, 14, 17]. However, Match Network merely considers the one-shot scenario where one class contains only one training sample, and it is not hard to inference that its performance will be degraded when the size of the unseen category grows.

Different from [10] and [16] for one-shot task, Prototypical Network [15] addresses the few shot problem by maximizing the Softmax probability over the Euclidean distances between the testing sample to the prototype representations of the novel categories. The prototype feature of each category is obtained by averaging the feature vectors corresponding to the training examples of that category. Despite its simplicity, Prototypical Network is able to achieve state-of-the-art performance on benchmark datasets, and has promoted several descending works [6, 13, 14, 17]. For instance, Qiao *et al.*[14] proposes to average the activations of training samples in a category during episodic training, to learn a transformation between these activations and the weights of the corresponding classifier. During testing, the average activations of the support samples in a category will be transformed to the weights of the classifier, in which the obtained weight vector can be regarded as prototype representation. Distances between the test images and the supporting ones are thus obtained by dot-product between activations and weights (prototype features). Meanwhile, its peer works [6, 13] show that cosine similarity is much superior to dot-product under similar condition, and it always gives rise to models of better generalization ability.

While the works above focus on learning image representations by fixing the distance metrics (*e.g.*, L1, Euclidean, Dot-product and Cosine), Yang *et al.*[18] propose Relation Network to jointly learn the metric and image representation in a data-driven way (*i.e.*, there is no explicit metric defined, and the model directly learns the similarity grounded on the pairwise feature), through which it can better identify matching/mismatching pairs under different data distributions. Relation Network shows a promising approach towards metric learning, and has been further improved in recent work [12] by mainly proposing complete comparison of object-level features between image pairs.

2 INSIGHT

Few-shot learning is a promising solution for the problem of data starvation. Nevertheless, most of the current works solely focus on single label recognition where one image/text contains one category, and few works tackle the multi-label problem where one image/text is associated with multiple categories (labels), which is considered more practical in real-world applications. Hence, I propose few-shot multi-label learning tailored for multi-label scenario. In this problem, the methods that cultivate pseudo training samples is not so feasible and effective. However, we can still draw inspiration from metric learning.

While typical approaches for multi-label learning transfer it to multiple binary classification problems, in few shot learning, these binary models may be extremely relevant to the base categories and thus result in lower generalization capability on new tasks. Instead, we can design a label-ranking objective function (together with feature regularization) which maximizes the margin between positive and negative labels associated with an image, and meta-learn a model on base datasets. Note that the sub-tasks formulated from base datasets require being carefully organized to balance the label distribution. Then, we can adopt the learned model to initialize the neural network for novel categories similar to [6, 13]. Given a test image, we can combine the results predicted by the fine-tuned model and the results determined according to KNNs found on the support sets. Particularly, the KNN can also be sophisticatedly implemented with attention LSTM similar to Matching network [16], so as to jointly optimized with the label-ranking loss function.

REFERENCES

[1] Anonymous. A closer look at few-shot classification. In *Submitted to International Conference on Learning Representations*, 2019. under review.

[2] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. In *ICLR workshop*, 2018.

[3] Hao Chen, Yali Wang, Guoyou Wang, and Yu Qiao. Lstd: A low-shot transfer detector for object detection. In *AAAI*, 2018.

[4] Xuanyi Dong, Linchao Zhu, De Zhang, Yi Yang, and Fei Wu. Fast parameter adaptation for few-shot image captioning and visual question answering. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM '18, pages 54–62, New York, NY, USA, 2018. ACM.

[5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017.

[6] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In *CVPR*, pages 4367–4375, 2018.

[7] Bharath Hariharan and Ross B Girshick. Low-shot visual recognition by shrinking and hallucinating features. In *ICCV*, pages 3037–3046, 2017.

[8] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[9] Nakamasa Inoue and Koichi Shinoda. Few-shot adaptation for multimedia semantic indexing. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 1110–1118. ACM, 2018.

[10] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2, 2015.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[12] Liangqu Long, Wei Wang, Jun Wen, Meihui Zhang, Qian Lin, and Beng Chin Ooi. Object-level representation learning for few-shot image classification. *arXiv preprint arXiv:1805.10777*, 2018.

[13] Hang Qi, Matthew Brown, and David G Lowe. Low-shot learning with imprinted weights. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5822–5830, 2018.

[14] Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan Yuille. Few-shot image recognition by predicting parameters from activations. In *CVPR*, 2018.

[15] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087, 2017.

[16] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, pages 3630–3638, 2016.

[17] Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan. Low-shot learning from imaginary data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[18] Flood Sung Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales.

Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.