Survey Report: Self-Training on Image Classification

김한비 2021712203, 성균관대학교 데이터사이언스융합과

Abstract

While deep learning had been making progress in the field of computer vision, there was huge success in image procssing. Convolutional Neural Newtork (CNN) is one of the most widely used neural networks which has shown remarkable performance on various fields. Especially, image classification is one of the mian issues in comuter vision problems and CNN has already proven its ability to solve image classification problem efficiently. Using a large data set for image classification generally increases its accuracy, however there exists a trade-off between accuracy and time. Using a large data set also means that it takes more time to train model, not to mention of more time for labelling data before modelling. Semi-Supervised Learning adopts unlabelled data and it helps reducing time for labelilng data and also increasing its accuracy by using more data for training. From this survey report, we will look through the ideas of Semi-supervised learning accepted in image classification.

Conference Information and Reason for Journal Selection and

CNN was one of the main topics during the class, especially image classification was taught brief history about ImageNet from AlexNet to ResNet. Hence, I tried to study more about ImageNet and Computer Vision. ICCV(international Conference on Computer Vision) is one of the most well-known conferences in the world about computer vision for the past 32 years. It has been established on 1987 under the supervision and co-operation of IEEE(Institute of Electrical and Electronics Engineers) and CVF(computer vision foundation) [7]. “Self-Training” discussed as a trend during recent conference on 2021. Therefore, I selected the first paper “Self-training with Noisy Student improves ImageNet classification”, since it proposed idea about self-training used in image classification. For now, it is cited by 773 times according to GOOGLE Scholar. Reading journal critically make me re-search other journals related on the same field but suggest new idea to solve some problems or issues arisen from the original journal. The second journal “Meta pseudo labels” was also published the same conference later to solve the problem of pseudo labeling.

2. Self-training with Noisy Student improves ImageNet classification

2-1 Paper Information

Qizhe Xie, Minh-Thang Luong, Eduard Hovy, Quoc V. Le; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10687-10698

PDF Link: https://arxiv.org/abs/1911.04252

2-2. ImageNet classification

ImageNet is an image database organized according to the WorldNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images [4]. They also provide some challenges to motivate computer vision problems such as image classification. One of the challenges called “The ImageNet Large Scale Visual Recognition Challenge (ILSVRC)” evaluates algorithms for object detection and image classification at large scale [5]. For the past challenges, there were numerous attempts to increase accuracy and also reduce computing time for image classification. From this survey report, we concentrated on recent researches regarding self-training and semi-supervised learning used in image classification. “Self-training with Noisy student improves ImageNet classification”[1] suggested making both teacher models and student models recurcively and allow teacher models to teach noise-added data to their student models.

2-3. Self-Training

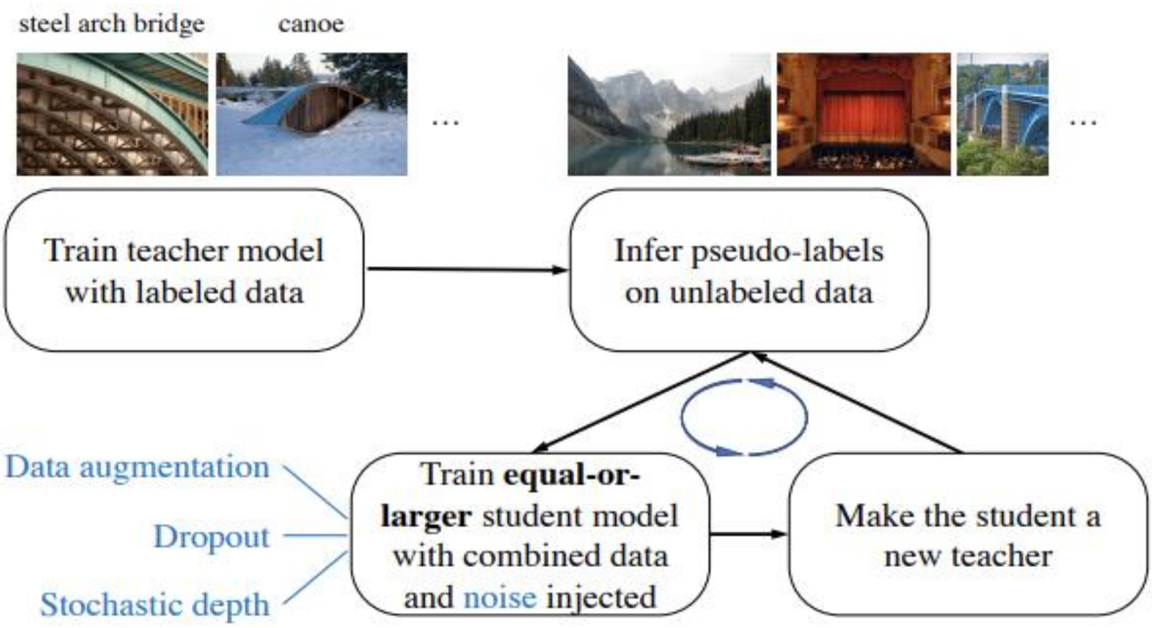

Figure 1: Illustrarion of the NoisyStudent Method

1) Use labeled image to train model by using standard cross entropy loss and the model named as teacher model. According to the paper, the best result was obtained as the first teacher model was trained by an EfficientNet-B7.
2) Use the teacher model to predict unlabeled image and the results then used for the label. In this case, predicted labels are not actual answer for the image. Therefore, labels are named as pseudo-labels.

3) Noise then added to the combined data. Noise is generated by Data augmentation, Dropout and Stochastic depth.

4) Train the noised added data by using combined cross entropy. Therefore, model is trained to reduce two losses on labeled image data and also pseudo-labeled image data. Since the new model is trained by pseudo-labeled image data generated by the teacher model, it is named as a student model.

5) Since our student model has been trained noise added data, its perforamce is at least equal or mostly larger than the teacher model. Therefore, we make the student model as a new teacher model to make another pseudo-labels on unlabeled data.

6) Iterate the process to minimize the cross entropy loss. During Iteration, the best result was obtained when the student model was trained by EfficientNet-L2 model.

2-4. Noisy Student

According to the original paper, they emphaszied continually on noising data before training student model. As we can see from the step (3), three different types of noising data applied. Data augmentation is an input noise which adds noise on data. Dropout and Stochastic depth are model noises which add noises on model.

For data augmentation, they referred “RandAugment: Practical automated data augmentation with a reduced search space”[3]. Data augmentation is the method of transforming and adjusting image data to generate noise and sometimes to increase the number of image data when input images are not sufficient. Accroding to RandAugment, they attempted to reduce the number of parameters controlled in data augmentation since taking all the parameter is expensive. Therefore, they suggested 14 different parameters in Figure 2 below. Also, they selected N, number of parameters randomly among 14 parameters below and M, for their magnitudes respectively.

• identity	• posterize	• autoContrast	• equalize
• rotate	• solarize	• color	• contrast
• sharpness	• brightness	• shear-x	• shear-y
• translate-x	• translate-y		

Figure 2: Lists of Parmeters

2-5. Conclusion and Dicsussion

In conclusion, self-training with Noisy student model is based on self-training and semi-supervised learning. They have shown that adding noise to data before training can achieve higher accuracy than the model trained without adding noise. Also, noise added model has high chance of avoiding overfiiting. Their experiment on ImageNet Data also achieved higher accuracy than state or the-art (SOTA) vision model.

As they already mentioned the importance of noising data, it will be expected to achieve better result if they will be able to control the size of nosing data for different stages. For instance, in reinforcement learning, there are artificially settled value, “Epsillon” for exploration at the earliest steps during learning since beginning model is not trained sufficiently to perform better compared to later model. Thus, the exploration occrus at the beginning steps and Epsillon value decreases gradually to reduce the number of exploration, but increase the chance of exploitation. In the same manner, we will be able to control the size of unlabled images by taking extra variable depending on stage.

Also, we need to realize that every model is not perfect. Using pre-trained model may allow us to save training time and to generate new model with better performance. However, that is only feasible under the condition where we already have relatively good pre-trained model. Even though our pre-trained model shows remarkable performance, it is a fact that we try to create new model since our pre-trained model is not good enough. If pre-trained model predict inaccurately, then new trained model might be worse than the previous model by overfitting.

3. Meta Pseudo Labels

3-1. Paper Information

Hieu Pham, Zihang Dai, Qizhe Xie, Quoc V. Le; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 11557-11568

PDF Link: https://arxiv.org/abs/2003.10580

3-2. Confirmation Bias

As we mentioned earlier, Pseudo Labeled data are generated by Pre-trained Teacher model on unlabeled data. Since we also have labeled data, the combined data of labeled data and pseudo labeled data then used to train the New Student model. From the earlier paper and other papers related to self-training with using Pseudo labeled data, they successfully improved accuracy compared to SOTA vision model. However, we need to focus on the fact that Pseudo-label is not actual classification label. That is, the newly trained student model is not always better than the previous teacher model. There are still high chance of creating pseudo labels inaccurately by teacher model, which will lead student model much more inaccurate. Such problem is also known as confirmation bias in pseudo-labelling [6].

3-3. Pseudo-Labels VS Meta Pseudo-Labels

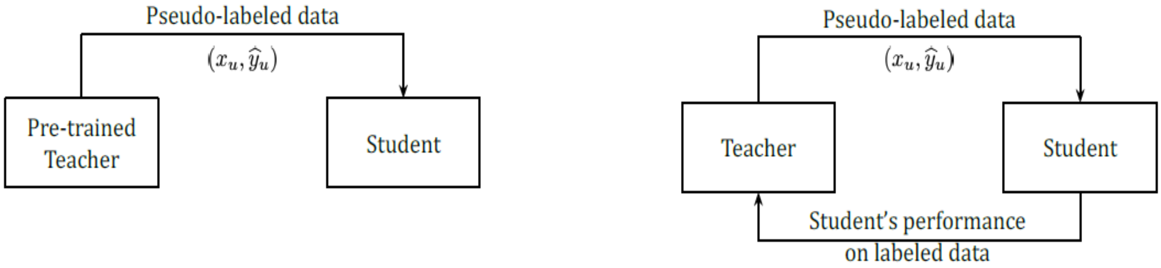

Figure 3: The difference between Pseudo Labels(Left) and Meta Pseudo Labels(Right)

The main difference between Pseudo Labels and Meta Pseudo Labels are shown in Figure 3 above. In the case of Pseudo-Labels, there will be problem if the teacher model predict inaccurate labels on unlabeled data, because parameters of teacher model are fixed. Therefore, Meta Pseudo Labels suggest that sending feedback from Student model to Teacher model to update and optimize teacher model parameters. The word "Meta" is derived from meta learning which is based on Bi-level learning. Both teacher model's and student model's parameters are updated interchangably by sending their feedbacks each other. Assume that teacher model predicted inaccurate pseudo-label on unlabeled data, and student model predicted differently, which is not certain to be accurate as well. However, teacher model paramter will be updated to reduce the loss by student model's feedback. Even though both of their predictions on unlabeled data is inaccurate and different each other, it is important that both of teacher model and student model attempted to update parameter and change their predictions. According to the paper, it was shown that meta pseudo label method performs better than SOTA vision model without meta pseudo labelling and any other semi-supervised learning.

3-4. Conclusion and Discussion

		Student Model	
		Accurate	Inaccurate
Teacher Model	Accurate	Best – No update	Student Model Update
	Inaccurate	Teacher Model Update	Still Problem

Figure 4. Teacher Model and Student Model update

Conclusively, we will save time and improve accuracy by using well-trained teacher model to generate pseudo label on unlabeled data. Self-training occurs between teacher model and student model. One of the main problems was that the model's performance might be decreasing in the case of teacher model generates inaccurate pseudo labels. Meta pseudo label prevents those cases by updating teacher model's parameters by feedback from student model. In those cases where one of teacher model or student model generates inaccurate pseudo labels, there will be still chance to generate accurate psudo label if one of them have accurate labels by sending feedback interchangably. However, if both teacher model and student model predict inaccuately on unlabeled data, the problem can not be fixed.

One of the other problem of Meta pseudo labels is modelling time. Since the parameer update and optimization between teacher model and student model occurs more with given learning rate, it is much slower than Pseudo labeling. According to the paper, they mentioned "A lite version of Meta Pseudo Labels" termed "Reduced Meta Pseudo Labels" at Appendix, but it just suggests reducing time by loading data and model efficiently to avoid large memory issues. Hence, there will be still not enough solution for reducing time.

References:

[1] *Qizhe Xie, Minh-Thang Luong, Eduard Hovy, Quoc V. Le*; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10687-10698

[2] *Hieu Pham, Zihang Dai, Qizhe Xie, Quoc V. Le*; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 11557-11568

[3] *Ekin D Cubuk, Barret Zoph, Jonathon Shlens, Quoc V Le;* Randaugment: Practical data augmentation with no separate search. *arXiv preprint arXiv:1909.13719*, 2019.

[4] ImageNet Wegsite Introduction https://www.image-net.org/index.php

[5] ImageNet Challenges https://www.image-net.org/challenges/LSVRC/index.php

[6] *Eric Arazo, Diego Ortego, Paul Albert, Noel E. O'Connor,Kevin McGuinness*. Pseudo-labeling and confirmation bias in deep semi-supervised learning. *Arxiv, 1908.02983*,2019.

[7] ICCV https://now.snu.ac.kr/49/2/1501