Few-shot semantic segmentation using REPTILE meta-learning approach

Matúš Tuna, Igor Farkaš

Department of Applied Informatics, Comenius University
Mlynská dolina, 84248 Bratislava
Email: tuna1@uniba.sk

Abstract

Semantic image segmentation task is important in many machine learning applications, but it usually requires a large number of training samples with pixel level annotations which is time consuming. As a remedy, we explore object segmentation using only a few training examples, by adapting a few-shot learning method known as REPTILE. Our preliminary results show comparable accuracy to the previous work, except one case, despite a smaller number of trainable parameters in our model.

1 Few-shot semantic segmentation

Semantic segmentation task is inherently more difficult than category classification, because instead of one label, it requires a 2D map of labels, separating a target objects from the background. Few-shot classification models such as Siamese neural networks (Koch et al., 2015), Prototypical networks (Snell et al., 2017) or MAML (Finn el al., 2017) achieve classification accuracy approaching fully supervised models, using only a fraction of labelled data. Due to inherently higher difficulty of semantic segmentation and limited availability of training data, few-shot approaches to semantic segmentation seem viable, but they have not yet been sufficiently explored, compared to their use for categorization. Among the first attempts to few-shot semantic segmentation is Shaban et al. (2017) who combined a segmentation model based on a fully convolutional network (Shelhamer et al., 2016) with a conditional model trained to generate a set of parameters conditioning the segmentation model to segment objects of particular category using only a few examples. Dong et al. (2018) introduced a metric learning approach (similar to Prototypical networks) that relies on creating per class prototypes from a limited amount of training data that can be used for final segmentation. Guided networks (Rakelly et al., 2018) use a small number of training images with limited annotation (one point per object) to perform semantic segmentation. This model also uses two networks, one for generating latent representation of a task from a small number of sparsely annotated images and one that uses this representation to generate a segmentation map.

2 Semantic segmentation using REPTILE

REPTILE algorithm (Nichol et al., 2018) belongs, together with MAML, to a family of meta-learning based few-shot learning methods. They propose a specific learning procedure that results in a network that can be adapted to a new task using only a small number of training data. These methods are universal with respect to model that is used, lending themselves for any (stochastic) gradient descent (SGD) learning. Unlike MAML, REPTILE algorithm does not calculate the second-order derivatives of a cost function which results in a strong advantage without sacrificing performance. REPTILE uses two optimization steps: (1) A task is sampled from a distribution of tasks (e.g. 5 random categories from a set of possible categories), then n steps of stochastic gradient descent are performed on this task resulting in a new set of parameters $\tilde{\theta}$. (2) Starting parameters θ are moved towards the new set $\tilde{\theta}$. The intuition behind REPTILE is that eventually the parameter vector θ should converge to a state that is close (in a Euclidian sense) to manifolds of optimal solutions of all possible tasks. In this state, only a small number of gradient steps using new training examples are required to update the network to a new task. REPTILE algorithm can be summarized as follows:

- Initialize network parameters θ,
- for i iterations do:
  - sample task τ and compute loss L_τ
  - perform n steps of SGD: $\tilde{\theta} = SGD(\theta, L_\tau, n)$
  - REPTILE update with learning rate $\varepsilon : \theta \leftarrow \theta + \varepsilon(\tilde{\theta} - \theta)$
- end for.

In our experiments, to be consistent with the setup used by Shaban et al. (2017), we used the network architecture FCN32 (Shelhamer et al., 2016). The only difference is that we chose ADAM optimizer (Kingma et al., 2014). We also adapted the ADAM optimizer to the REPTILE update by treating the term $(\tilde{\theta} - \theta)$ as a gradient. The convolutional feature extractor was pretrained using the ImageNet dataset (Deng et al., 2009). We used PASCAL VOC 2012 image segmentation dataset (Everingham et al., 2015) and applied the same category splits as Shaban et al. (2017). The performance was evaluated on a one way 5-shot tasks, which means that during training every iteration a random category is sampled from a set of 15 training categories from the training subset. Then every SGD iteration five supports images that contain at least one pixel from that category are sampled from training data. All pixels belonging to categories that are different from the training category are relabeled as background. During testing the similar procedure is performed, although the categories are taken from the validation subset and the network is retrained using only the 5 support examples selected at the beginning of the test episode. After each test episode the network parameters and ADAM optimizer statistics are reset to the state before the test. We used the same performance metric as Shaban et al. (2017), namely the per class mean intersection over union computed across all 5 test classes (excluding background). The learning rates for ADAM optimizer and for REPTILE were set to 0.0001. For every REPTILE update during training we performed 24 ADAM updates, and during testing 64 ADAM updates. The network was trained for 1000 REPTILE updates.

Figure 1: Examples of segmentation from our model.

3 Results and conclusion

In Table 1 we compare our model with Shaban et al. using the same 4 test splits (named PASC5-0, PASC5-1, PASC5-2 and PASC5-3). We have not compared our model to Dong et al. and Rakelly et al. due to a different metric used in their work. Accuracy of our model for most test splits is comparable to the model by Shaban et al. except for the PASC5-1 split. There is a large difference between the number of parameters between the two models. Our model does not contain the conditional model that contains similar number of parameters as the segmentation model which might explain the difference in segmentation accuracy. We plan to test this hypothesis in the future. We also plan to verify our model on larger semantic segmentation dataset which will enable us to test the accuracy on 5-way tasks.

Method (5-shot)	PASC5-0	PASC5-1	PASC5-2	PASC5-3	Mean
Shaban et al. 2017	35,9	58,1	42,7	39,1	43,9
Ours	37,1	48,5	43,8	38,5	42,0

Table 1: Performance comparison of the models. We used the mean intersection over union (MIOU) metric from Shaban et al. (2017).

Acknowledgement

The work was supported by Slovak Grant Agency for Science (VEGA), project 1/0796/18.

Literature

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248-255).

Dong, N., & Xing, E. P. (2018). Few-shot semantic segmentation with prototype learning. In *BMVC*.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2), 303-338.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, pp. 1126-1135.

Kingma, D. P. a Ba, J. (2014). Adam: A method for stochastic optimization. *Preprint arXiv 1412.6980*.

Koch, G., Zemel, R. a Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431-3440).

Nichol, A., Achiam, J., & Schulman, J. (2018). On first-order meta-learning algorithms. *arXiv preprint 1803.02999*.

Rakelly, K., Shelhamer, E., Darrell, T., Efros, A., & Levine, S. (2018). Few-shot segmentation propagation with guided networks. *ArXiv preprint 1806.07373*.

Shaban, A., Bansal, S., Liu, Z., Essa, I., & Boots, B. (2017). One-shot learning for semantic segmentation. *arXiv preprint 1709.03410*.

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. In *Advances in NIPS*, pp. 4080–4090.