Online Continual Learning via Multiple Metric Learning and Uncertainty-guided Progressive Memory Sampling

Muhammad Rifki Kurniawan*

School of Electronic and Information Engineering, Xi'an Jiaotong University, China

October 11, 2021

1 Method Summary

Learning over sequential data points and continuously involving distribution shifts as this SSLAD competition is a very challenging task. Therefore, in order to avoid the massive forgetting due to overfitting to certain data points, we proposed multiple deep metric learning in order to build generalized representation learning along with replay method with uncertainty-based progressive episodic memory in an online way. In the online setting, we can not revisit the whole data stream while we should select the samples among batch samples at a time. We employ uncertainty-aware sampling strategy with entropy measure as the metric. As a result, given a batch of samples the sampler will select the top-k most uncertain samples with highest entropy to be stored in episodic memory in a stream. Moreover, we executed periodic memory sweeps up via uncertainty approximation with Monte-Carlo method as proposed by Rainbow Memory [1] and keep the top-k most uncertain while detach the rest since the memory is limited to 1000 objects while the steam data points continuously coming.

*email: `mrifkikurniawan17@gmail.com`; corresponding author

In addition, for the purpose of training generalized representation that is optimum for the whole distribution, we train the networks using deep metric learning and learning with soft labels targets strategy along with classification loss. While training the networks, instead of minimizing the common cross entropy loss, we find that focal loss [2] works better since this loss is aware of class imbalance problems as the stream data setting issue and less consider the overconfidence predictions while strongly consider the hardly-classified examples. In addition, in deep metric learning, we employ both contrastive learning [3] and supervised contrastive learning [4] to build better and more general representation that potentially avoids task-specific optimization which intensively leads to overfitting. Aligning with the general representation, we also employ the soft labels learning via retrospecting the soft predictions of the former trained weight that are stored in episodic memory.

For the technical implementation, we use Imagenet pre-trained resnet50d backbone with 23.5 millions parameters including the fully-connected classifier head. Moreover, we set the episodic memory capacity to 1000 objects and execute sweeping up periodically if the memory capacity is full or every 1000 training steps to default capacity which is set to 500. While the online stream batch data points are coming in 10 mini batch sizes, we concatenated those with 5 samples obtained from the sorted-by-uncertainty episodic memory samples and augmenting the memory samples with common image-level augmentations including random rotation, horizontal flip, and color jitter to improve model generalization. Then, we train the networks with SGD optimizer with learning rate 0.0111 along with a multi-step lr scheduler.

References

[1] J. Bang, H. Kim, Y. Yoo, J.-W. Ha, and J. Choi, "Rainbow memory: Continual learning with a memory of diverse samples," in *CVPR*, June 2021. 1

[2] T.-Y. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, 2017. 2

[3] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE Com-*

puter Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, 2005, pp. 539–546 vol. 1. 2

[4] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 18 661–18 673. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/d89a66c7c80a29b1bdbab0f2a1a94af8-Paper.pdf 2