Project B: Knowledge Distillation for Building Lightweight Deep Learning Models in Visual Classification Tasks

Abstract—This paper introduce the implementation and evaluation of KD based on two couples of Teacher-Student Models.

I. Task 1: Knowledge Distillation in MNIST dataset

This document is a model and instructions for LaTeX. Please observe the conference page limits.

II. Task 2: Knowledge Distillation in MHIST dataset

References

[1] Cristian Bucilu˘a, Rich Caruana, and Alexandru Niculescu-Mizil. Model compression. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 535–541, 2006. https://dl.acm.org/doi/10.1145/1150402.1150464.

[2] Jang Hyun Cho and Bharath Hariharan. On the efficacy of knowledge distillation. In Proceedings of the IEEE/CVF international conference on computer vision, pages 4794–4802, 2019. http://openaccess.thecvf.com/content_ICCV_2019/papers/Cho__On_the_Efficacy_of_Knowledge_Distillation_ICCV_2019_paper.pdf.

[3] Xiang Deng and Zhongfei Zhang. Comprehensive knowledge distillation with causal intervention. Advances in Neural Information Processing Systems, 34, 2021. https://openreview.net/forum?id=ch9qlCdrHD7.

[4] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. International Journal of Computer Vision, 129(6):1789–1819, 2021. https://link.springer.com/article/10.1007/s11263-021-01453-z.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In European conference on computer vision, pages 630–645. Springer, 2016. https://arxiv.org/abs/1603.05027.

[6] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2(7), 2015. https://arxiv.org/abs/1503.02531.

[7] Takumi Kobayashi. Extractive knowledge distillation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 3511–3520, 2022. https://openaccess.thecvf.com/content/WACV2022/papers/Kobayashi__Extractive_Knowledge_Distillation_WACV_2022_paper.pdf.

[8] Yann LeCun, L´eon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998. http://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf.

[9] Seyed Iman Mirzadeh, Mehrdad Farajtabar, Ang Li, Nir Levine, Akihiro Matsukawa, and Hassan Ghasemzadeh. Improved knowledge distillation via teacher assistant. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 5191–5198, 2020. https://ojs.aaai.org/index.php/AAAI/article/view/5963/5819.

[10] Rafael Mu¨ller, Simon Kornblith, and Geoffrey Hinton. Subclass distillation. arXiv preprint arXiv:2002.03936, 2020. https://arxiv.org/abs/2002.03936.

[11] Dae Young Park, Moon-Hyun Cha, Daesin Kim, Bohyung Han, et al. Learning student-friendly teacher networks for knowledge distillation. Advances in Neural Information Processing Systems, 34, 2021. https://proceedings.neurips.cc/paper/2021/file/6e7d2da6d3953058db75714ac400b584-Paper.pdf.

[12] Wonpyo Park, Dongju Kim, Yan Lu, and Minsu Cho. Relational knowledge distillation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3967–3976, 2019. https://openaccess.thecvf.com/content_CVPR_2019/html/Park_Relational_Knowledge_Distillation_CVPR_2019_paper.html.

[13] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. Fitnets: Hints for thin deep nets. arXiv preprint arXiv:1412.6550, 2014. https://arxiv.org/abs/1412.6550.

[14] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015. https://arxiv.org/abs/1409.0575.

[15] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4510–4520, 2018. https://openaccess.thecvf.com/content_cvpr_2018/papers/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.pdf.

[16] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive representation distillation. arXiv preprint arXiv:1910.10699, 2019. https://openreview.net/pdf?id=SkgpBJrtvS.

[17] Jerry Wei, Arief Suriawinata, Bing Ren, Xiaoying Liu, Mikhail Lisovsky, Louis Vaickus, Charles Brown, Michael Baker, Naofumi Tomita, Lorenzo Torresani, et al. A petri dish for histopathology image analysis. In International Conference on Artificial Intelligence in Medicine, pages 11–24. Springer, 2021. https://arxiv.org/abs/2101.12355.