Appendix for "Regression-Oriented Knowledge Distillation for Lightweight Ship Orientation Angle Prediction with Optical Remote Sensing Images"

December 7, 2024

S.1 GitHub repository

Please find the codes for this paper at

https://github.com/UBCDingXin/SOAP-KD

S.2 Detailed Experimental Setups

S.2.1 The Setups for Mobile-SOAP and Four Tiny Students

The convolutional blocks utilized in Mobile-SOAP are derived from a MobileNetV2 checkpoint that has been pre-trained on ImageNet. Similarly, the convolutional blocks for the ShuffleNet-based students are sourced from a ShuffleNet checkpoint pre-trained on ImageNet. However, the ResNet8 and WRN16×1-based students are trained from scratch.

For training, we set the number of epochs to 200 and the batch size to 128. The initial learning rate is 0.01, and it undergoes decay at the 80th and 150th epochs, respectively.

S.2.2 The Setups of ASD [3] and AMEFRN [6]

When implementing ASD, we start by binning the range of angles, specifically $[0, 180]$, into 60 disjoint intervals, with each interval representing a distinct class. The ASD model is then trained to categorize samples into these 60 classes. During the inference phase, any given image is assigned to one of these intervals (or classes), and the predicted angle for the image is calculated as the average of the starting and ending points of that interval. We experimented with dividing the angles into different numbers of classes, but found that using 60 classes yielded the best results. For training, we set the number of epochs to 200 and the batch size to 128. The initial learning rate is set at 0.01, and it is reduced at the 80th and 150th epochs.

When implementing AMEFRN, we made a modification to the original design by retaining only the attribute-guided branch and removing the classification branch, as depicted in Fig. 5 of [6]. The attribute-guided branch is then utilized as the SOAP model. For training, we set the number of epochs to 200 and the batch size to 128. The initial learning rate is 0.01, and it decreases at the 80th and 150th epochs.

Both the networks for ASD and AMEFRN are built upon the VGG16 architecture, which has been pre-trained on ImageNet. This ensures that the models start with a strong feature extraction capability, leveraging the knowledge learned from the large-scale ImageNet dataset.

S.2.3 The Setups of FitNet [4], RKD [7], DKD [2], and NDF [31]

The implementations of FitNet and RKD are primarily based on the code provided by [1].

For DKD [2], it proposed a DA module and a SSG module. However, the DA module requires modifications to both the student and teacher architectures to define an additional KD loss, which is not ideal for our setting. These modifications would prevent us from leveraging the pre-trained MobileNetV2 checkpoint and would introduce additional parameters to the student model. Consequently, we did not implement the DA module when applying DKD.

Regarding NDF, [31] assumes that the teacher and student features share the same dimension, which is not the case in our experiment. Therefore, we utilize the proposed adapter network to adjust the channel size of the student features prior to applying the NDF method.

Please refer to our code repository for more detailed setup information.

S.2.4 The Setups of SOAP-KD

When training the CcGAN in the optimized cGAN-KD framework, we adopt the SAGAN architecture as described in [5]. For the regression CNN used for label embedding, we designed it based on the VGG8 architecture and trained it for a total of 10 epochs. The parameters σ and κ for CcGAN are set to 0.051 and 8100, respectively. The majority of the other settings follow those outlined in [1].

For the subsampling mechanism in cGAN-KD, we utilize a classification-oriented CNN to extract features. This CNN is built upon the ResNet50 backbone, which has been pre-trained on ImageNet. The ResNet50 is fine-tuned to predict the ship type of input images, with a total of 23 classes. It is trained for 200 epochs using a batch size of 128. The initial learning rate is set to 0.1 and is reduced by a factor of 0.1 at the 80th and 150th epochs. After training the ResNet50, we proceed with the conditional density ratio estimation. For this purpose, we design a simple 5-layer CNN as the density ratio model. This DRE model is trained for 100 epochs with a batch size of 100 and a regularization parameter of 0.05. The remaining settings are similar to those described in [1]. Finally, we generate a total of 25,340 fake images using the CcGAN for training the student models.

For a more detailed implementation of SOAP-KD, please refer to our provided code.

References

[1] Xin Ding, Yongwei Wang, Zuheng Xu, Z Jane Wang, and William J Welch. Distilling and transferring knowledge via cGAN-generated samples for image classification and regression. *Expert Systems with Applications*, 213:119060, 2023.

[2] Daxiang Li, Yixuan Nan, and Ying Liu. Remote sensing image scene classification model based on dual knowledge distillation. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[3] Jinlei Ma, Zhiqiang Zhou, Bo Wang, Hua Zong, and Fei Wu. Ship detection in optical satellite images via directional bounding boxes based on ship center and orientation prediction. *Remote Sensing*, 11(18):2173, 2019.

[4] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. FitNet: Hints for thin deep nets. In *International Conference on Learning Representations*, 2015.

[5] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *International Conference on Machine Learning*, pages 7354–7363, 2019.

[6] Xiaohan Zhang, Yafei Lv, Libo Yao, Wei Xiong, and Chunlong Fu. A new benchmark and an attribute-guided multilevel feature representation network for fine-grained ship classification in optical remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:1271–1285, 2020.

[7] Qilu Zhao, Junyu Dong, Hui Yu, and Sheng Chen. Distilling ordinal relation and dark knowledge for facial age estimation. *IEEE Transactions on Neural Networks and Learning Systems*, 32(7):3108–3121, 2020.