ROTATION-INVARIANT REPRESENTATION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep learning based models fail in generalizing well to test instances sampled from outside the training distribution. To approach this issue, ongoing research efforts in the field of invariant risk minimization aim at formulating & optimizing objectives that offer better prospects on the model's ability to generalize beyond the domain of the training set. Our work focuses on learning image representations invariant to rotational transformations—a prevalent impediment in several modern computer vision applications as models often fail to produce good results when objects & entities in the test images are rotated versions of those in the train images (Azulay & Weiss, 2019). We draw inspirations from the algorithmic fairness literature wherein prior works (Baharlouei et al., 2020; Cho et al., 2020; Jiang et al., 2020) have developed *in-processing* (or train-time) mitigation strategies to promote statistical independence between the predictor's output and a certain predetermined *sensitive feature* in the dataset. We encode rotational transformation as discrete quantities $\{0,90,180,270\}$ and consider it a hand-crafted sensitive feature for our use-case. We will extend the FERMI framework (Lowy et al., 2021) for our task to penalize any dependency between the model's output and the input image's rotation feature, and hypothesize that a model trained through this adjusted risk minimization strategy will generalize better to test samples. For experimental evaluation, we pre-process the dataset so that the train images' rotation angle strongly correlates with the train labels and the test images uncorrelated (example visualized in figure 1). In this scenario, a model trained using classical ERM obtains poor accuracy scores on the test data due to its inability to generalize. Using this setup, we will perform a systematic evaluation of the efficiency and generalization ability of commonly used CNN-based architectures like LeNet, Resnet-18, etc., on rotated versions of several popular datasets like MNIST, CIFAR10, etc. In addressing a specific computer vision issue, we believe this study will also help researchers & practitioners gain more insights into generalization in deep learning.

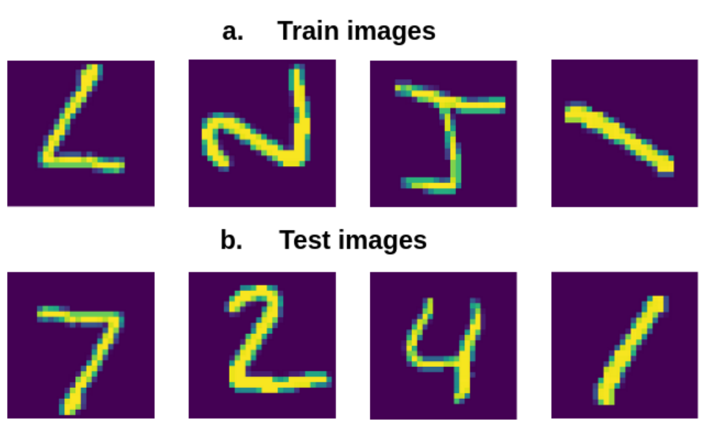

Figure 1: Rotated MNIST

References

Aharon Azulay and Yair Weiss. Why do deep convolutional networks generalize so poorly to small image transformations? *Journal of Machine Learning Research*, 20(184):1–25, 2019.

Sina Baharlouei, Maher Nouiehed, Ahmad Beirami, and Meisam Razaviyayn. Rényi fair inference, 2020.

J. Cho, G. Hwang, and C. Suh. A fair classifier using mutual information. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 2521–2526, 2020. doi: 10.1109/ISIT44484.2020.9174293.

Ray Jiang, Aldo Pacchiano, Tom Stepleton, Heinrich Jiang, and Silvia Chiappa. Wasserstein fair classification. In Ryan P. Adams and Vibhav Gogate (eds.), *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pp. 862–872, Tel Aviv, Israel, 22–25 Jul 2020. PMLR.

Andrew Lowy, Rakesh Pavan, Sina Baharlouei, Meisam Razaviyayn, and Ahmad Beirami. Fermi: Fair empirical risk minimization via exponential rényi mutual information, 2021.