論 文 の 内 容 の 要 旨

論文題目 Rapid and Robust Learning for Homogeneous Image Data
（同質画像に対する高速・堅牢な学習）

氏　　名　　ミシュラ　ソーラヴ

This thesis investigates the performance of fine-grained classification models on homogeneous images and attempts to interpret and improve the outcomes. Additionally, we have explored the effects of adversarial perturbations and out-of-distribution conditions and whether we can establish confidence in such models given these scenarios. Deep learning has improved discriminative tasks and fine-grained analysis of visually homogeneous image data has been a long-standing challenge. Model learning is computationally expensive and difficult for many applications. The thesis has explored some of the pertinent questions and attempts to provide a few solutions & explanations.

The introductory chapter gives a capsule overview of the existing methods for different topics discussed in the thesis, intended to bring the reader up-to-speed with the state of the art. In subsequent chapters, we explore the attributes of homogeneous data and briefly discuss datasets that could be beneficial for our experiments. We have investigated designing baseline measurements and shown a combination scheme involving logarithmically interpolated layer-wise rates and cycle length multiplication can immensely improve the model accuracy and learning speed as compared to commonly used practices. We have carefully experimented the scheme over three different types of data - standard vision (CIFAR-10), standard fine-grained dataset (Oxford-IIIT Pets) and extremely homogeneous skin data (Exmedio Skin lesions). In each case, we have observed the performance to be better than the baseline.

Following good model learning schemes, we have tried to understand the reason behind the gap between error-free human level performance, and our peak model Top-1 accuracy. The choices that Deep learning models make in classification is not well understood. An established line of research (Geirhos et al. 2018) has proposed

that image texture weighs heavily in decision choices. New line of research (Kornblith et al., 2020) have given counter examples regarding shape instead of texture. We have investigated the reasons behind classification on visually similar images and found that models have different constraints and choices depending on the type of labels under question. It is also probable that in absence of distinguishing high-level features, models depend further on earlier layers for decision making. We have discussed data curation strategies including effects of weighted sampling, confounder removal, etc. to be used in combination with learning improvements in presenting the best outcomes.

The effect of adversarial perturbation has been discussed for Exmedio Skin images. The thesis has investigated effect of various levels of impulse noise and presence of soft-blur (isotropic and anisotropic) by various ablation studies. We have found that homogeneous images are far more susceptible to impulse noise than standard data such as CIFAR-10. This is probably due to the lack of strong visual attributes to begin with, which are hard to learn, as compared to common images found in datasets such as ImageNet, CIFAR-10, STL-10, etc. Both isotropic and anisotropic blur have been found to have smaller effect on model accuracy in similar comparison to CIFAR-10. We have also investigated the effect of dataset shift and compared the performance between previously established results (Recht et al. 2018) with dataset shift effects in homogeneous skin images. In both adversarial inputs and dataset shifts, we have observed that it is difficult to establish confidence bounds in predictions.