

MERIT Corporate Internship (Domestic) Report

Department of Physics, Graduate School of Science
5th generation student of MERIT course
Daiki Adachi

【Date】

2018/11/1～2018/12/27

【Host corporation】

Data Science Research Laboratories, NEC Corporation

【Outline】

I completed an internship at Data Science Research Laboratories of NEC Corporation at Kawasaki-shi, Kanagawa-ken, Japan. The theme of this internship was “Development for image processing technology by deep learning.” I researched the performance and approaches to enhance the performance of deep-learning “semi-supervised” method, which is in recently researched actively in the field of image-processing machine learning field.

【Research】

Image classification is one of the most researched problems in machine learning. For example, classifying MNIST dataset (which has images of handwriting numbers from 0 to 9) to 10 classes from 0 to 9 is very famous as a benchmark of image classification. However, such classification problems have practical problems as follows:

1. Training data should be labeled manually.
2. Correct answer rate of classification would decrease when training dataset is biased (ex. the number of each number is very different)

To overcome these problems, many approaches have been developed. In this internship, I researched using a network called Domain Separation Network (DSN), which is one of the famous deep-learning networks for semi-supervised learning.

Semi-supervised learning is used for the problem to classify an unlabeled domain (target datasets) by the use of another labeled domain (source datasets). For example, the network that is optimized to classify the handwriting numbers of one person could not classify the handwriting numbers of another person. By introducing the information that “the labeled dataset comes from a different domain than the unlabeled dataset” and learning both datasets at the same time, the correct answer rate of classification for unlabeled dataset can extremely increase. This is very important theme for classification to apply social problems because producing labeled dataset automatically (ex. taking images of 3D CG models from various directions) would eliminate human labor.

In this internship, I applied DSN network to various datasets to evaluate its performance and improved its network to increase the success rate of classifications. This knowledge is important for not only image processing, the theme of this internship, but also the general classification problems and the researches of deep learning for classifications.

【Acknowledgements】

I'm deeply grateful to NEC Corporation for accepting the internship. Especially Mr. Shibata, who accepted me as an intern student and gave me guidance on administrative procedures and research, Mr. Namiki, who had a in-depth discussion about research and setups of instruments for my research environment, and Mr. Yachida, who lectured me how to perform basic research at a company, and other members in Data Science Research Laboratories. In addition, Mr. Ando and Mr. Taira, who were the fellow of the internship from another university, helped me to understand the general topics in machine learning. I also appreciate the help of PCoMS, especially Ms. Arima, for coordination with the NEC. Finally, I would like to show my gratitude to Professors Synge Todo and Hajime Tanaka for permitting to join the internship program, and the MERIT program for giving me this valuable opportunity.