

# Project Proposal: Digits Classification using CNN

Team 5\* - Shiwei Chen, Shixiang Wang

March 2022

## I Introduction

In this project, we will research on classifying handwritten digits, a subtask under the field of Contextual Image Classification in Computer Vision. The task is provided by the famous Modified National Institute of Standard and Technology datasets [1] (Hereinafter referred as *the MNIST dataset*). In the early days of machine learning, the dataset was widely used for benchmarking various image processing systems. While the original datasets contains 60000 samples for training and 10000 samples for testing, our project uses only a fraction of them, which are 2000 samples for training and 2000 samples for testing. Additionally, the project will be tested on a hidden dataset at the end of submission deadline to determine the final performance. In the subsequent sections, we will describe our method for the task, the basic structure of the experiment, and the expected outcome.

## II Methodology

### II.(a) Data Augmentation

Due to the restricted size of the data, we deem that Data Augmentation, a technique to increase the diversity of the training data without actual collection and labeling, is necessary. Large amount of data is essential for supervised learning methods to achieve high performance [2]. We plan on using random noise, random rotation and translation, flipping (for symmetric digits), cropping, and resizing, to at least increase the size of the restricted dataset to the same magnitude as the original MNIST dataset.

### II.(b) Network Structure

According to the leaderboard [3], Convolution Neural Network (CNN)[4] is still the dominant component in the state of the arts of this task. Therefore, we plan on construct our own network structure on the basis of CNN. More precisely, we want to construct an ensemble of slightly different CNN-based networks that will together vote for the final classification decision [5]. The difference can be kernel sizes, number of layers, with or without normalization, etc.

## III Experiment and Expected Result

The experiment will be conducted in a Jupyter notebook for convenience, with PyTorch being the primary network building tool. The experiment will follow general data science workflow and thus roughly consist of four parts:

- Perform Exploratory Data Analysis (EDA) to get a general view of the data and see if there are some noteworthy characteristics.
- Conduct Data Augmentation base on the result of EDA.
- Construct the designed structure and implement the proposed method.
- Compare and analyze the final results and select the best one achieved.

---

\*Contact: {shiwechen6-c, sxwang6-c}@my.cityu.edu.hk

The expected outcome would be our best model saved in a .pt or .pth file accompanied by a python script to reproduce the results, a notebook for visualizations, a final report, and a presentation video.

## References

- [1] LeCun, Y., Bottou L., Bengio, Y., and Haffner, P. “Gradient-based learning applied to document recognition.” Proceedings of the IEEE, 86(11):2278-2324, November 1998.
- [2] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, “AutoAugment: Learning augmentation policies from data”, in Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR), 113-123 (2019)
- [3] Meta AI, 2022. Access Date: March 21st, 2022. Url: <https://paperswithcode.com/sota/image-classification-on-mnist>.
- [4] LeCun, Y., Haffner, P., Bottou, L., Bengio, Y. (1999). “Object Recognition with Gradient-Based Learning”. Shape, Contour and Grouping in Computer Vision.
- [5] An, S., Lee, M., Park, S., Yang, H., So, J. (2020). “An Ensemble of Simple Convolutional Neural Network Models for MNIST Digit Recognition”.