Project Proposal: Digits Classification

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 contain 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. The fewer data makes our project even more challenging. Additionally, the project will be tested on a hidden dataset at the end of the 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

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. A 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. After that, we will walk through the entire machine learning progress. First, we will try to use dimension reduction algorithms such as PCA to preprocess the feature vectors. Then we plan to try traditional classification algorithms such as logistic regression, SVM, and KNN. As this is a multi-class classification problem, we will try to use *one vs. all* or *softmax* to extend the binary classification to multi-class classification. 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 constructing our network structure based on 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, number of nodes, with or without normalization, etc. Besides, we will also try to apply residual blocks [6] to alleviate the vanishing gradient problem in deep neural networks. We will test our implemented algorithms on a separate test data set to ensure that the training set and test set are independent. The accuracy will be chosen as the evaluation criteria.

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.

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.

---

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

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".

[6] He, K., Zhang, X., Ren, S., and Sun J. "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016 pp. 770-778. doi: 10.1109/CVPR.2016.90