Final Project Group 6 Proposal

Qizhen(Kurtis) Shen, Changhong Zhang, Hui Yang

- In our final project, we want to apply our Neural Networks knowledge learnt from the ML2 class on the image classification problem of leaves. It is a relatively new competition from the Kaggle website Classify Leaves with not so many baseline codes that we can borrow from. We want to do a Image classification problem again since it is the most fundamental problem in the neural network field and we want to use some simple models as well as more advanced models, including some models that we never covered in the class so we can learn something new.

- We are using the leave classification dataset which has a total number of 18, 353 training images and 8, 800 test images. The image size is 224 x 224 pixels. One complexity part of the leaves dataset is that it has a total number of 176 classes. But the imbalance of the dataset is not a big issue where each category has at least 50 images. The dataset is definitely large to train a deep network and is not that large so that it requires a lot of computational power. It is another reason that we choose this project/dataset that the size of the dataset is perfect for a class project.

- We will use the convolutional neural networks (including a simple CNN baseline, VGG-16 and ResNet-18) and more advanced frameworks like “Vision Transformer”. We will build the CNN baseline model ourselves with not many numbers of convolutional layers. It will also be designed for better visualization purposes since we are planning to visualize the baseline model with its filters and feature maps. For the VGG-16 and ResNet-18 models, it will be in a more standard form without much modifications. For the Vision Transformer, it was modified from the NLP framework to accommodate our image classification problem.

- We are using the Pytorch framework since Pytorch is more friendly to customized neural networks and provides more flexibility to access low-level functions. It also helps us to understand the modeling better as the programming is more intuitive and mimicking the real modeling process.

- We will be reading the initial papers of related works (VGG Simonyan and Zisserman, 2014, ResNet He et al., 2015, etc.) as well as the online blogs and online classes e.g., CS231n: Deep Learning for Computer Vision from Stanford.)

- We will use the accuracy, Cohen's kappa score as well as the f1 macro average score for our metrics and the sum of these three metrics as the final metric.

- We will spend about one week doing EDA and getting ourselves familiar with the leaves dataset. In the meantime we will do background reading and find out the most popular models and more advanced models in the market which we would like to try. Then we will work on our CNN baseline model as well as some other deeper models starting in the second week. One of our group-mates will focus on more advanced models like Vision Transformer since we don't have prior knowledge about that and may require more time to understand and to implement the model. At the end of the second week we will be able to run all the models that we plan to run on the leaves dataset and the comparison of the model performance will be done by then. In the last week we will collect all the modeling results and apply visualization on the CNN baseline model to understand our results. We may have done some augmentation at the beginning of the project but we may reflect on the modeling comparison results and think about further augmentation to mitigate any problems we may encounter during the modeling part. We will also be wrapping up the report writing and collecting and assembling the github repo from the individual branches. The project is estimated to be finished in about 3 weeks.

References

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv e-prints, Article arXiv:1409.1556, arXiv:1409.1556.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. arXiv e-prints, Article arXiv:1512.03385, arXiv:1512.03385.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.