ML2 Final Project Proposal Group 7

After you have selected a topic, a network, and a data set, submit a proposal of what you plan to do for the project. The proposal should be a few hundred words, and should address the following items.

• What problem did you select and why did you select it?
    - We want to create a deep learning network to classify pictures of food by dish
    - This could be useful for Yelp, instagram, etc. since they do not tag images automatically and rely on users to self-tag their images. Such a classification system could allow the company to tag the types of images automatically, allowing users to better search for images by type of food.
• What database/dataset will you use? Is it large enough to train a deep network?
    - Food-101 database: https://www.kaggle.com/kmader/food41 on kaggle
    - Also found here: https://www.vision.ee.ethz.ch/datasets_extra/food-101/
    - 1000 images per class - split into 850/150 training/test set
• What deep network will you use? Will it be a standard form of the network, or will you have to customize it?
    - Convolution Neural Network
    - Will require significant pre-processing - using OpenCV to pre-process data
• What framework will you use to implement the network? Why?
    - Pytorch / Caffe????
• What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?
    - OpenCV documentation, class notes,
• How will you judge the performance of the network? What metrics will you use?
    - Accuracy % for each class and overall accuracy
    - F1 and recall
• Provide a rough schedule for completing the project.
    - We will do it by the deadline

Food-101 dataset
101 classes of food, each with 1000 images
384x384 RGB image
Merge the different databases together and resize so uniform

Machine Learning II Final Project Proposal - Group 7
Joanne Chung, McLain Wilkinson, Ruonan Jia

We want to create a deep learning network capable of identifying the culinary dish pictured in an image. This network could have many applications for companies such as Yelp or Instagram, who have many images uploaded to their sites depicting different pictures of food. Yelp does not currently have a way to search for pictures of specific dishes at restaurants, relying only on user-tagged images in picture searches searches. Utilizing such a network could allow Yelp to automatically tag images to certain categories, allowing for users to easily search for pictures of the hamburger at a specific restaurant, for instance.

The dataset we have chosen to use for this project is the Food-101 dataset, found on kaggle at https://www.kaggle.com/kmader/food41. It contains five different datasets of food images, with varying resolutions and offered in color/grayscale. The data represents 101 classes of foods, with 1,000 images represented per class. Trying to train the network to learn 101 different classes will be difficult, so we will most likely combine many similar classes and corresponding images together. This will aid in reducing the number of target classes as well as increasing the number of training images per class. For instance, if two possible target classes of the dataset are “apple pie” and “chocolate pie”, we would combine these two classes into one class, “pie”.

We plan to use a convolution neural network (CNN) for this task. Much of the CNN architecture is well-designed for our purpose, but we will have to perform significant image pre-processing to achieve better results. The OpenCV library for python will allow us to crop, color scale, smooth, and resize each image if necessary. Pre-processing will be the most important aspect of achieving results for this project.

We plan to use either Pytorch or Caffe for this project. Pytorch would be a better choice at this time, while we don’t have experience with Caffe.

Significant research will need to be done for us to successfully complete the project. We will need to spend time reading the OpenCV documentation to learn about its capabilities and how to implement certain techniques to pre-process our images. It would also be helpful to look at other networks that have been created for the purpose of image classification to get a better understanding of the flow of the network and the steps we should have to build our network.

We can judge the success of our network by looking at classification accuracy, both for individual classes and overall accuracy. F1 score and recall would also be good metrics of success.

There are roughly 5 weeks to complete the project. The first 2 weeks will be spent doing research on Caffe, OpenCV, and image classification networks. During the next 2 weeks we will work on coding the network, aiming to have a working prototype complete by the November 27 class. This will give us 1 more week to tune the network and make improvements before the presentation on December 4.