# Progress Report 3

groveh

March 2019

## Problem Description

Becoming familiar with the orientation problem is a crucial step, and the MNIST dataset provided a simple set of images to begin with. When training reaches an acceptable stopping point, it is more beneficial to start working with real data. This is important because black and white numbers may present unseen artifacts that interfere with the angle estimation, and results need to be verified with non gray scale images. Another reason is that rotation with real animal images is different than black and white. MNIST is rotated at the outer edges of the frame, and therefore to preserve true pixels, a maximal inscribed rectangle operation must be implemented. On animal images, the area of interest is specified by a bounding box and this is the only area that should be evaluated, therefore the maximal inscribed rectangle is less important because we have more true pixels than are solely within the bounding box. Instead, we need the opposite, a minimal enclosing rectangle operation.

## Problem Approach

The data is formatted in MSCOCO style with folders of images for each training, validation, and testing, as well as a folder for json files of annotations for each. Using the official python API, the annotations were verified to be accurate for almost all of the images. The API is necessary for this problem because it will be how the bounding box is extracted to be cropped and fed into the orientation estimation network. Using the bounding box and a labelled theta in radians, these will act as the input image and corresponding image label to train the network.

## Solution

The MNIST dataset is syntactically different from real animal images because it is not a built in dataset. It took a bit more complexity to load the animal images using the API to find the corresponding filename to load. From here the issues became trivial, using the API to get the bounding box and thus minimal enclosing rectangle from the min and max points around the bounding box. The cropped region of interest is then input into the network along with the annotated radians of rotation.

## Challenges Faced

When the images are selected for each batch, the API must be used to locate each filename and load the specific annotation and related image. This computation is now happening during training rather than as a pre-processing measure. This isn't a significant tradeoff, however it is equivalent to loading the dataset repeatedly at the beginning of each epoch, rather than solely during the first. The other challenge is a logical decision about what to do when the minimal enclosing rectangle is located beyond one of the edges of the image. Currently the region of interest is cropped to preserve pixel correctness, as this seemed like the answer most similar to the maximal inscribed rectangle tactic used during use with the MNIST dataset to avoid additional black background from being added to the image input.

## Results

Pre-processing images as well as finalizing the cropping of the region of interest enables the network to begin training on these images. The initial results were less ideal than expected, most likely due to lack of augmentation. As the network trains with one angle per image, it learns the training set to the point where it is memorizing and mapping images to angles. During testing the network performs poorly without adequate dataset size and random manipulation. While the challenges of utilizing the COCO API and loading animal images has been solved, augmentation is a necessary step in reaching satisfactory results.

## Future Work

In the near future, a better way of loading images through the MSCOCO python API will be implemented. This is to avoid computation of loading the image during each batch selection and contributing to the increase in training time. Additionally, the dataset is not augmented currently. The rotation measure used is the annotated value, and therefore there is only one version of each training input. This can be improved by essentially "zeroing" the image at a rotation of zero radians then augmenting with an angle from a distribution of angles. Also there is also potential for an investigation into possible vanishing gradient when radians are used as opposed to degrees because these label values are a factor of 60 times smaller.