Progress Report 4

groveh

April 2019

Problem Description

Training with MNIST dataset images is a trivial task compared to real animal images. The dataset is pre-processed to be upright, centered, cleaned, and comes in a variety of formats as well as even being integrated within numerous frameworks, PyTorch included. Loading images is not the difficult task here, but rather which part of the image to load. Images are also not guaranteed to be correctly annotated, so this is another issue that needs addressing. The region of interest within each image is the head of the turtle, which is defined by MSCOCO style annotation. With each annotation (there may be multiple in each image), important attributes include the segmentation points around the head, the viewpoint of the image (left side, top, ect.), and the angle of the segmentation (which translates to the angle of the head).

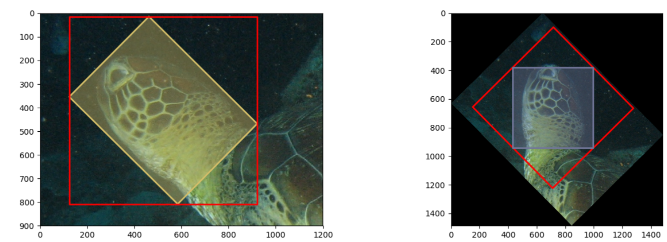
Problem Approach

The initial step is to pre-process the images and load them into memory. This is necessary because the images and the annotations are located in separate folders, therefore in order to avoid loading each image and annotation when collecting the current batch (this is equivalent to loading the dataset again at each epoch), they are stored as pairs within a list ready for data augmentation.

Data augmentation consists of cropping and rotating solely the head of the turtle so that it is facing upward at a universal $0°$ angle. It can then be augmented by some angle that is then used as the image label. The region of interest crop of the head is the maximal inscribed rectangle that results from the rotated minimal enclosing bounding box of the original annotation. An example is shown in figure 1.

Solution

During pre-processing, in order to reject images that are either poorly annotated or for some other reason cause an error during augmentation, a test augmentation is applied to each image found in the directory to test if it makes it through without error. This test augmentation ensures that during batch selection the images do not cause any issues during each unique augmentation.



(a) The annotated segmentation with surrounding minimal enclosing rectangle to be rotated with the image.

(b) The rotated minimal enclosing rectangle from Fig. 1a, with axis aligned maximal inscribed rectangle.

Figure 1: The segmentation and bounding box before rotation, and the rotated bounding box with final minimal inscribed crop.

The process of setting the image to its original orientation as well as cropping it to include only interest region pixels is not trivial. Theta is known from the annotation, so the first step is to rotate not only the image by this angle, but also the minimal enclosing rectangle of the segmentation. This minimal enclosing rectangle does not exceed the image dimensions, even if the segmentation does. This ensures that the maximal inscribed rectangle does not include any black background pixels generated during image rotation. The final crop is that of the maximal inscribed rectangle.

Challenges Faced

The pre-processing decision also avoids various errors down the road, such as those arising from lack of data cleaning. For example, the mean and standard deviation must be calculated in order to normalize the images as part of recommended transforms. For images that are labelled incorrectly or cause an error during test augmentation, they are not included in the dataset and therefore the final mean and standard deviation are not skewed since we only pre-load the images we intend to train on. Other challenges were the challenge of finding a reliable way to crop solely from the region of interest, which resulted in the process described in the Solution section.

Results

In terms of pre-processing, very few images are thrown away from errors during augmentation. There are a total of 2674 image in the training set, 940 in the validation set, and 904 in the test set. This means that 99.58% of images are able to be used of the total number of images in the complete dataset, $\frac{4{,}518}{4{,}537}$. The dataset itself is rather small, therefore data augmentation is extremely important. While the theoretical algorithm produces good training images, a significant portion of the dataset contains annotations that exceed the dimensions of the image. This limits the ability of the cropping and rotation algorithm to capture a crop that represents enough of the region of interest from the original annotation. If the annotation take up most of the original image, there is a chance that this interferes with the accuracy of final prediction because the network is trained on images with a very small window of the region of interest.

Future Work

More work needs to be done to filter out images where the annotation coordinates are severely outside the bounds of the image. When they are rotated and cropped they end up showing only a small part of the original annotation. This can potentially be fixed by making a more intelligent algorithm to go outside the rotated minimal enclosing rectangle of the original segmentation, but still focus on the turtle head as well as not include any of the black background pixels the appear from image rotation. If this intelligent crop is not possible, the image should not be loaded into the dataset, determined by the of areas of the final crop to the original segmentation.