Multimedia Project

SoSe 23

Multimedia Computing Lab

Prof. Dr. Rainer Lienhart, Katja Ludwig, Julian Lorenz

Assignment Submission 3

Jan Claar

Exercise 3.1

The maximum number of annotations for a single image in the training dataset is 5. There are 5 images with this number of annotations, one of which is shown in Figure 1. The chosen image is simply the first one encountered when iterating over all images in the directory.
Since the filenames are obtained via the `glob.glob` function, the order is arbitrary and the image encountered by `get_maximum_annotation_image()` may be different for different environments.

Figure 1: One of the Images with the maximum number of Annotations (5). Bounding boxes are drawn in red.

Exercise 3.3

The final accuracy after 50 Epochs of Training was 0.86 with the best accuracy being 0.868.

The accuracy is not a good measurment for the model performance in this particular setup. The reason for this is that the training and validation set are highly imbalanced, when only differentiating between images containing one or multiple people. For both datasets, around 77% the images contain only one person (see output of `distribution.py`), so a network labeling all images as such would already have a 77% accuracy. The trained network in this is case is only a little better than that. A class wise accuracy or anything that accounts for this imbalance would be a more relevant performance metric.