Multimedia Project

SoSe 23

Multimedia Computing Lab
Prof. Dr. Rainer Lienhart, Katja Ludwig, Julian Lorenz

Assignment Submission 1

Jan Claar

Exercise 1.3

a) See `report_a.json`

The model identifies the Koala, the Golden Retriever and the Ox in the zoo image correctly. Since the classes the network is trained on don't contain classes that fit the contents of the other images exactly, they are instead classified as similar objects or concepts. Some of these are fitting, e.g. 'toyshop' for the image containing a lot of rubber ducks, others only vaguely resemble their actual content if at all ('Petri Dish' for the pacifier). It still seems like the model has recognized some broad concepts (flat round object) instead of choosing a class at random.

b) See `report_b_{128, 512}.json`

Making the images smaller results in the model not being able to identify any of them correctly, probably due to lack of information. However, the model performance also deteriorates with larger images, which is probably due to it being trained on 224 by 224 crops of images resized to 256 by 256 (as seen in the documentation for the pretrained weights). The added information only confuses the model and the filters learned by the convolutional layers aren't the correct size for the larger images' bigger features.

c) See `report_c.json`

Flipping the images vertically results in a different (wrong) classification for almost all images besides the koala and the golden retriever. The augmentations used during the training of the network likely don't contain any flips, so the network hasn't learned any invariances concerning orientation of objects. As to why the golden retriever and koala were classified correctly anyway, even though the images don't contain any strong vertical symmetry, I can only suspect that the training data perhaps naturally contains

images of golden retrievers and koalas on their heads. Considering the nature of these animals, this guess might not be too unlikely.