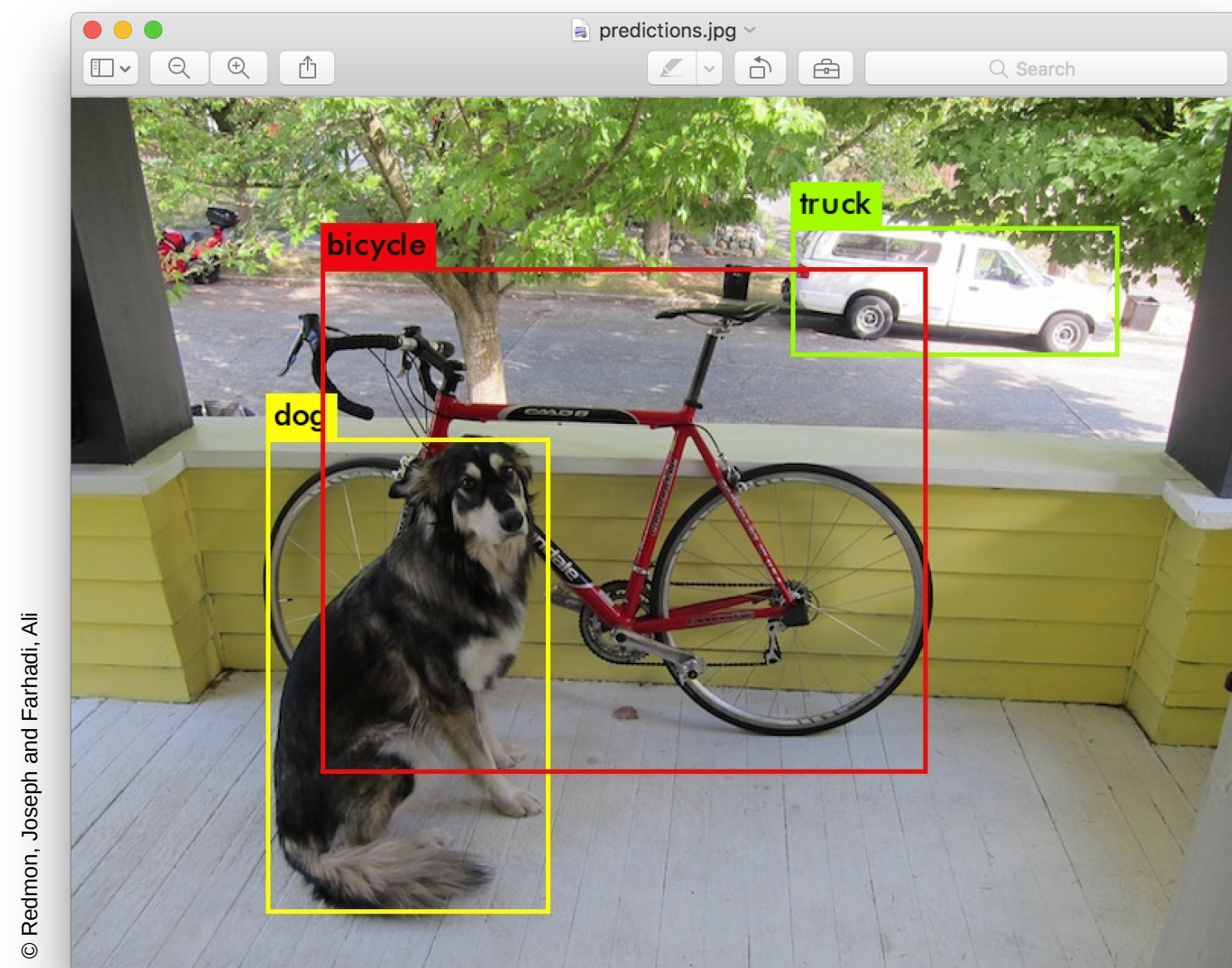


How to train your own CNN for Object Localization using Transfer Learning on the ImageNet dataset?

Dieter Ladenhauf, dieter.ladenhauf@student.tugraz.at

Transfer Learning

The task was to use transfer learning to train the tiny yolo nn-architecture (this is a convolutional neural network (CNN)) with a set of own labels. Transfer learning is a technique of using an already pre-trained neural network for a different set of labels. This approach works surprisingly well because the first layers of the neural network detect low-level features, like edges, lines, or shapes, which can be used to detect different objects as well.



Object localization using the YOLO¹ algorithm

„You only look once“¹ (YOLO)

The term YOLO stands for "You Only Look Once" and it depicts an algorithm where a fixed set of bounding boxes is used to detect objects. Additionally, it subdivides the training images into a grid (usually 19x19) and each object is assigned to a single grid cell during training. The different grid cells are then represented by a specific part of the output volume of the neural network.

At the end of a prediction, YOLO returns a set of bounding boxes with different confidence levels assigned to them. Now the algorithm non-max suppression is used to filter out unnecessary or overlapping bounding boxes. To detect overlapping bounding boxes, intersection over union is used. In the end only the most promising bounding boxes (highest confidence and no overlapping) survive the postprocessing done in non-max suppression.

Purpose

The results (obtained models) of this project will be used for a smartphone game which contains a game element, where the user is required to take a photo of certain objects from time to time. If this action is performed, a machine learning server analyzes the photo if it contains the required object and localizes it. The plan is to release the game by the end of the year.

Data Engineering

Aquiring the necessary training data was definitely the hardest part of this project. It required several iterations, until the acquired data could be used for training the neural network. The data was downloaded from the ImageNet² API in an automated fashion.

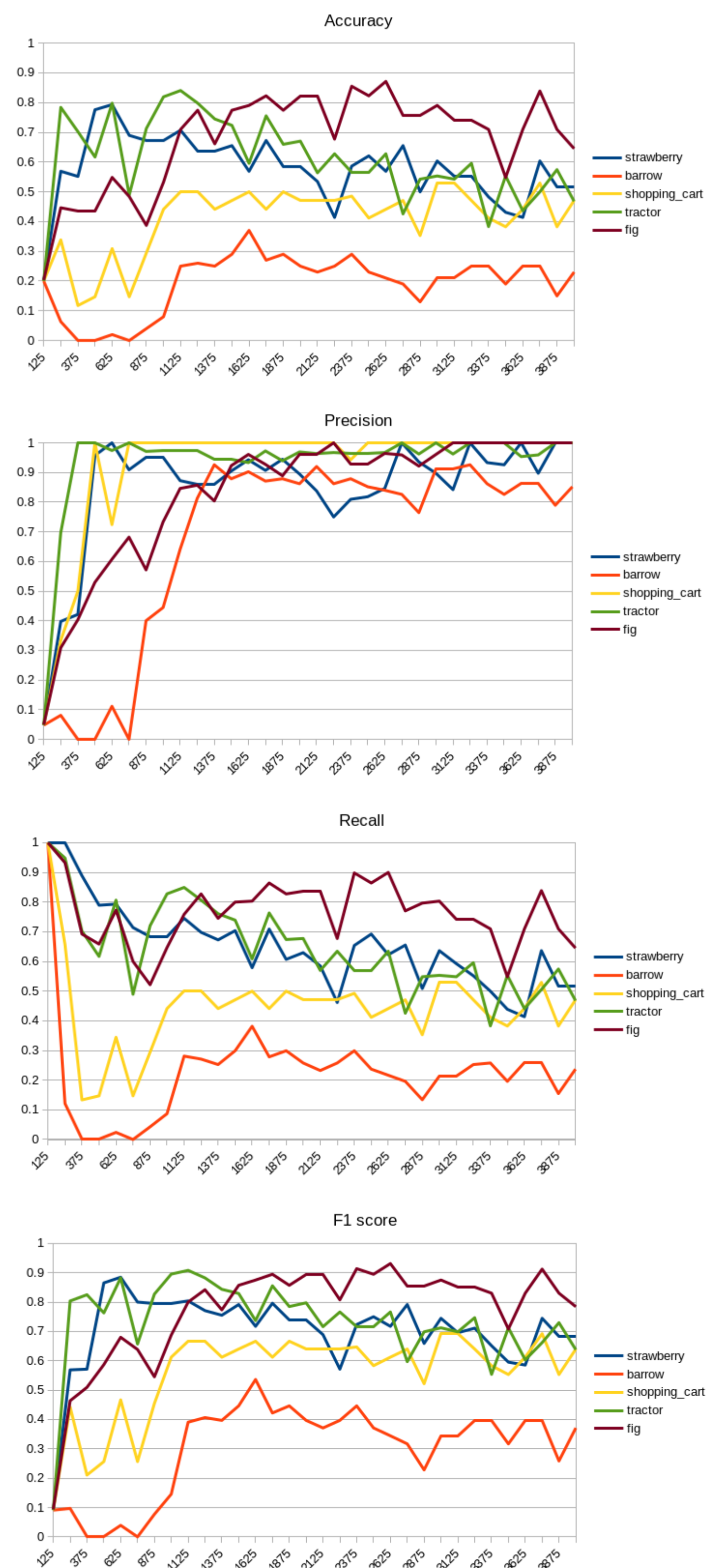
Training

For training the neural network, the open source project darkflow³ was used. It ports the original YOLO implementation darknet⁴ to Tensorflow. Training was performed on the cloud on a machine with just CPUs. GPUs would have been nice, but since time was not a limiting factor, it didn't really matter that it took quite a long time to produce the needed machine learning models.

Evaluation and Results

The data was divided into 80% training data, 10% dev (validation) data, and 10% test data. The dev data was used to evaluate the hyperparameters and to decide on how to split the labels for training (how many models). After this phase was finished, the dev (validation) set was added to the training data in order to maximize the amount of used training data.

Training with darkflow automatically produced checkpoint files from time to time. An evaluation script was used for evaluating these checkpoints in an automated fashion. For each ckeckpoint it took all the test data and performed a prediction with the given model for each image. It then produced a score between 0 and 1 to depict the overall accuracy. 0 means that the object never appeared within the test data of the given label, 1 means that the abject was always correctly located within the test data. Additionally, the evaluation script calculated precision, recall and F1 score:



Future Work

- Use Google's Open Images dataset⁵ for training since the training ran into overfitting quite fast. Overfitting means that we memorize the training data, which decreases the overall accuracy on the test set.

Literature

- Redmon et al (2015). You Only Look Once: Unified, Real-Time Object Detection.
- <http://www.image-net.org/>
- <https://github.com/thtrieu/darkflow>
- <https://pjreddie.com/darknet/>
- <https://storage.googleapis.com/openimages/web/index.html>