Project proposal: EXPERIMENTALLY IMPROVING A YOLO-INSPIRED OBJECT DETECTION MODEL

Herluf Baggesen
201800258@post.au.dk

Mads Buchmann Frederiksen
202003444@post.au.dk

The purpose of the project is to experiment with how a simple baseline model for object detection can be improved. The goal is not to effectively solve the problem of object detection but rather gain insight and understanding of why different changes have the outcome they have. We believe this is easier done if experiments are first carried out on a simple model and which can then gradually be made more complex. Therefore as a baseline model, we plan to employ the simple model presented in Lab 3, which is a naive implementation inspired by the YOLO architecture [3].

Task & Dataset

For the object detection task we have chosen digit detection using the Street View House Number Dataset (SVHN). The main advantages of this dataset is that it is easily available, of good quality and of reasonable size, while being small enough to feasibly train models on the hardware available to us.

Experiments

At the time of writing we plan to experiment with:

- Using different loss functions. This is motivated by the fact that object detection consist of both regression, when finding a bounding box, and classification. The naive model uses only mean squared error loss, even though it is unsuited for classification tasks. We expect that this change will make the model converge faster during training.
- Adding depth to both the encoding and decoding steps. We expect to find that adding layers with some motivated combination of activations will improve the models performance dramatically. This is motivated by empirical results from [2] showing that deeper networks generalize better when used to transcribe multi-digit numbers from photos of addresses.
- Adding skip connections. This is a known technique to address the issue of "exploding gradients" that we would like to investigate. This item is motivated by [1].

We expect, as the project and course progresses, that we should identify more techniques that will motivate further experimentation. We leave some room in our planning to pursue these angles as well.

To compare the base model with a modified model during a particular experiment, we plan to:

- Compute intersection over union to measure bounding box accuracy.
- Compute mean average precision for classification of digits.
- Investigate precision/recall curves
- Investigate confusion matrices to detect misclassification trends

Plan

We plan to implement our project using the pytorch API, as it allows for the granular level of tweaking that we want to pursue. We plan to begin working on these steps first:

1. Implementing and training the base model in pytorch

2. Implement a reusable way to investigate model performance via the metrics mentioned earlier. This will allow for faster experimentation.
3. Begin performing an experiment, presumably with the loss function.

References

[1] David Balduzzi, Marcus Frean, Lennox Leary, JP Lewis, Kurt Wan-Duo Ma, and Brian McWilliams. The shattered gradients problem: If resnets are the answer, then what is the question? 2017.

[2] Julian Ibarz Sacha Arnoud Vinay Shet Ian J. Goodfellow, Yaroslav Bulatov. Multi-digit number recognition from street view imagery using deep convolutional neural networks. 2013.

[3] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.