

1. AUFGABENSTELLUNG

Object detection in fine-art photography

Themenbereiche:	Artificial Intelligence, Visual Computing
Studierender:	Fabian Meyer
Betreuungsperson:	Dr. Simone Lionetti
Experte:	Roman Bachmann, Swisscom
Keywords:	Object detection, Machine learning, Computer vision, Photography

1 Aufgabenstellung

Object detection is an important computer vision and machine learning task that consists in the localisation and classification of items within digital images. The business and industry applications of this technology are growing in number and relevance, especially because of the tremendous progress, largely driven by deep learning, that has been made in recent years.

The success of object detection models is usually measured with benchmark metrics on standard datasets such as MNIST, PASCAL VOC, ImageNet, COCO and Open Images. These reference scores summarise a complex set of performance parameters and are essential for a meaningful comparison of different approaches. Precisely because they are extremely synthetic, however, it is important to assess the quality of results with other, more ingenuous criteria.

2 Ergebnisse

The following results have been made:

- A dataset from different artists with different styles has been gathered
- Multiple object detection models have been tested on images of fine-art photography
- The most promising model (Mask R-CNN) has been selected
- An program that takes in an image and puts out a tidied up version with all objects found has been developed
- The program has been converted into a webapp
- ...

The webapplication can be viewed at: <http://bdaf20-iameyer.enterpriselab.ch>.

Performance of the Mask R-CNN model depends on:

- Class of objects in the image (included in COCO-dataset or not)
- Size of objects in the image
- Degree to which an object is visible and is shown in a natural style

When applied to images of fine-art photography we can observe that performance is very variable because different photographers use a different style to depict their objects in their images. For example when using an image with a lot of small objects, performance is poor compared to an image with less number and bigger objects.

3. LÖSUNGSKONZEPT



Abbildung 1: Die Storage-Lösung für das nächste Jahrtausend

3 Lösungskonzept

The webapplication takes in images either from a selection or via upload or via a url. Then the model starts inference on the chosen image. The chosen model will detect potential objects in the image and returns them as masks, bounding boxes, name of class and a confidence score.

The program takes in the four results from the model and cuts out all objects from the original image. The found objects will be depicted on a new image in a grid-like fashion.

Cutting out the objects is possible with the bounding box and the object mask. One can cut out an object with the bounding box and colour all pixels for example white which do not belong to the mask.

Die erneuerte Storage-Lösung des Enterprise Labs ist in Abbildung 1, Seite 2 dargestellt.

4 Spezielle Herausforderungen

The biggest challenge was to get a model run on a standard CPU, for development but also for deployment on a server. As most deep learning networks are trained with GPUs, the easiest way to deploy an app would be to use a GPU environment. But there are some frameworks that can accept GPU-trained networks and can use CPU for inference.

5 Ausblick

Die bestehende Storage-Lösung des Enterprise Labs ist in Abbildung 2, Seite 3 dargestellt.

5. AUSBLICK



Abbildung 2: Die bestehende Storage-Lösung im Enterprise Lab