Sapienza University of Rome

Vision and Perception
Mask R-CNN

Ivan Bergonzani, Michele Cipriano,
Ibis Prevedello, Jean-Pierre Richa

June 13, 2018

1 Introduction

The aim of the project is to train a model based on Mask R-CNN[1] using an extended version of COCO that includes the dataset created on Labelbox. The new dataset consists on a bunch of images that shows the Gymnastic activities of ActivityNet. All the images have been downloaded from Google using a Python tool called `google-images-download`. The idea is to have a working model that will be later used to classify videos that show Gymnastic activities.

The project has been developed in Python and it has been tested using Google Compute Engine. The final training has been performed at Alcor lab.

Results. Problem of the network. Peaks. How to improve.

2 Training

Mask R-CNN has a set of losses that are used to check the performances of the classification, the RPN, the regression on the bounding boxes and the instance segmentation:

- `smooth_l1_loss`: the smooth-L1 loss on the classification of the objects.
- `rpn_class_loss`: the loss on the classification of the object contained in the region proposals, they can either be foreground if there is an object inside or background otherwise.
- `rpn_bbox_loss`: the loss on the bounding box returned by the RPN.
- `mrcnn_class_loss`: the loss for the classifier head of Mask R-CNN.
- `mrcnn_bbox_loss`: the loss for the bounding box refinement at the end of the network.

- `mrcnn_mask_loss`: the binary cross-entropy loss for the masks.

It is possible to see the graphs of these losses using tensorboard once the training is complete.

3 Conclusion

Conclusion.

References

[1] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2017.