QMUL-UK

Authors: Jingxiong Li, Qianni Zhang, Yaqi Wang

Email: jingxiong.li@se18.qmul.ac.uk

Platform: Windows

Prerequisites: Python

QMUL-UK: SUMMARY

Focusing on the Cell Segmentation Benchmark only, we trained a Mask R-CNN network. Its advantage is in generating instance segmentation. In theory, this network can check all objects in the input image and generate segmentation for each detected instance with a decent accuracy, simple extensible structure and high training speed [1].

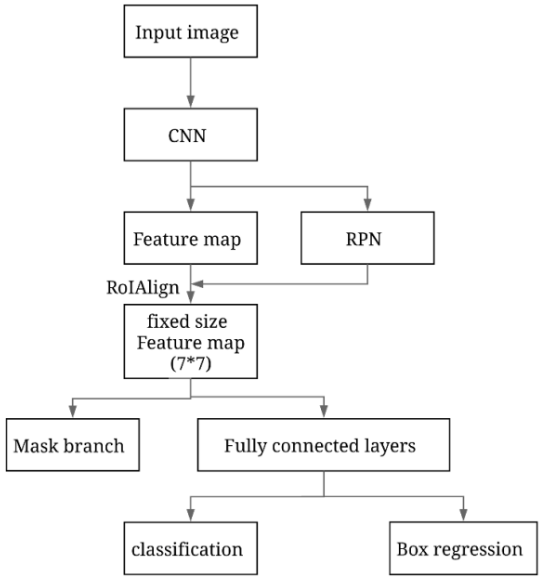
QMUL-UK: PREPROCESSING

No preprocessing step is performed.

QMUL-UK: SEGMENTATION

The architecture of the network is shown below, extending the output of a Faster R-CNN [2] by adding a branch to generate mask for the target. The input image is processed by a pre-trained network (e.g., ResNet) to get the feature map. Then, pre-set candidate bounding boxes (i.e., candidate Region of interest (RoI)) are distributed for each point in the map, called a Region Proposal Network (RPN). Then the candidate RoIs are distinguished into foreground and background then screening out some false RoIs. The next stage, features are refined by RoIAlign, corresponding the feature map and the RoIs. Finally, the model parallelly classify RoIs, do bounding box regression and generate mask inside each RoI. Formally, the multi-task loss has been defined as $L = L_{cls} + L_{box} + L_{mask}$ where L_{cls} and L_{box} are defined by a Faster R-CNN [3], and L_{mask} is identified in [1].

The input images and their segmentation are used for training the network. Some of the configurations are inherited from Faster R-CNN. A positive RoI must be at least 0.5 in Intersection over Union (IoU), if not, RoI will be considered as negative. L_{mask} is defined on positive RoIs only. The training is image-oriented [2]. All the images are resized to the same size. The mini-batch has 2 images per GPU. For each image, there is a fixed number N for their RoIs, the ratio between positive and negative is 1:3. For FPN-based model, N = 512 [5]. The model is trained on 1 GPU with 20K iterations. The learning rate is 0.001, with a weight decay of 0.0001 and the momentum is 0.9. The anchors of RPN also succeed from Faster R-CNN, with 5 scales and 3 aspect ratios. The pre-trained weight of this model is 'Usiigaci_01' from [3]. When testing the trained model, the number of proposals is 2000. The proposals are processed by box prediction then non-maximum suppression [4]. The model will pick up top 100 detection boxes, then pick out boxes according to the mask branch predicted K masks.



QMUL-UK: POST-PROCESSING

No post-processing step is performed.

REFERENCES

1. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In *Proceedings of the 16th IEEE International Conference on Computer Vision*, 2980-2988 (2017).
2. Girshick R. Fast R-CNN. In *Proceedings of the 15th IEEE International Conference on Computer Vision*, 1440-1448 (2015).
3. https://github.com/oist/usiigaci
4. Girshick R, Iandola F, Darrell T, Malik J. Deformable part models are convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 437-446 (2015).
5. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 936-944 (2017).