

Master Thesis Defense Arun Rajendra Prabhu

Date: Monday, 31.08.2020

Time: 13:00

Room: Online

Title: An investigation of regression as an avenue to find precision-runtime trade-off for object segmentation.

Abstract:

The ability to finely segment different instances of various objects in an environment forms a critical tool in the perception tool-box of any autonomous agent. Traditionally instance segmentation is treated as a multi-label pixel-wise classification problem. This formulation has resulted in networks that are capable of producing high-quality instance masks but are extremely slow for real-world usage, especially on platforms with limited computational capabilities. This thesis investigates an alternate regression-based formulation of instance segmentation to achieve a good trade-off between mask precision and run-time. Particularly the instance masks are parameterized and a CNN is trained to regress to these parameters, analogous to bounding box regression performed by an object detection network.

In this investigation, the instance segmentation masks in the Cityscape dataset are approximated using irregular octagons and an existing object detector network (i.e., SqueezeDet) is modified to regress to the parameters of these octagonal approximations. The resulting network is referred to as SqueezeDetOcta. At the image boundaries, object instances are only partially visible. Due to the convolutional nature of most object detection networks, special handling of the boundary adhering object instances is warranted. However, the current object detection techniques seem to be unaffected by this and handle all the object instances alike. To this end, this work proposes selectively learning only partial, untainted parameters of the bounding box approximation of the boundary adhering object instances. Anchor-based object detection networks like SqueezeDet and YOLOv2 have a discrepancy between the ground-truth encoding/decoding scheme and the coordinate space used for clustering, to generate the prior anchor shapes. To resolve this disagreement, this work proposes clustering in a space defined by two coordinate axes representing the natural log transformations of the width and height of the ground-truth bounding boxes.

When both SqueezeDet and SqueezeDetOcta were trained from scratch, SqueezeDetOcta lagged behind the SqueezeDet network by a massive ≈ 6.19 mAP. Further analysis revealed that the sparsity of the annotated data was the reason for this lackluster performance of the SqueezeDetOcta network. To mitigate this issue transfer-learning was used to fine-tune the SqueezeDetOcta network starting from the trained weights of the SqueezeDet network. When all the layers of the SqueezeDetOcta were fine-tuned, it outperformed the SqueezeDet network paired with logarithmically extracted anchors by ≈ 0.77 mAP. In addition to this, the forward pass latencies of both SqueezeDet and SqueezeDetOcta are close to ≈ 19 ms. Boundary adhesion considerations, during training, resulted in an improvement of ≈ 2.62 mAP of the baseline SqueezeDet network. A SqueezeDet network paired with logarithmically extracted anchors improved the performance of the baseline SqueezeDet network by ≈ 1.85 mAP.

In summary, this work demonstrates that if given sufficient fine instance annotated data, an existing object detection network can be modified to predict much finer approximations (i.e., irregular octagons) of the instance annotations, whilst having the same forward pass latency as that of the bounding box predicting network. The results justify the merits of logarithmically extracted anchors to boost the performance of any anchor-based object detection network. The results also showed that the special handling of image boundary adhering object instances produces more performant object detectors.