EECE 7370 Advanced Computer Vision
Project Abstract
Instance segmentation

Zhiyang Liu, Yaofei Wang, Huaiyu Zheng

To make the images easier to understand for computer and extract useful information of the image. Image segmentation is a good way to give a visual scene of where and what are in the image. It shows the location of the objects in the image and their boundaries. For now, image segmentation has two main research direction, semantic segmentation and instance segmentation. The figure shows the differences between semantic segmentation and instance segmentation. The objection localization only provides the bounding box of the object. Semantic segmentation can tell the location and the shape of the object. For instance segmentation, it can not only detect the pixel-level segmentation but also segment individual object even in the same class. We can think of that instance segmentation is the combination of semantic segmentation and object detection.

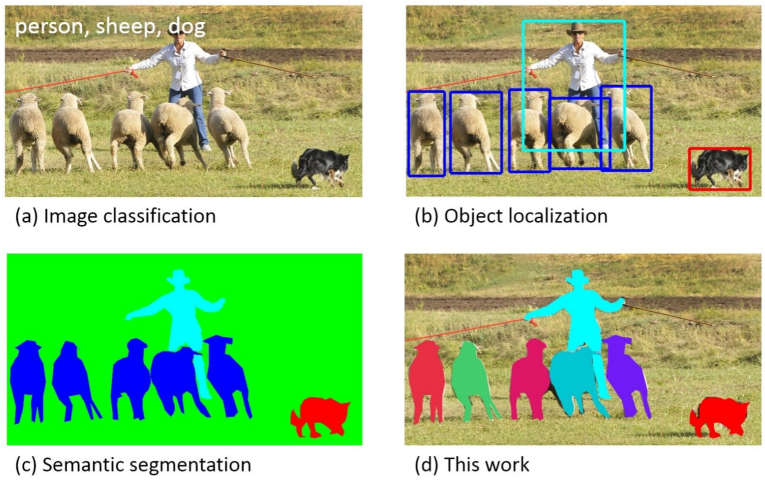

Figure 1: While previous object recognition datasets have focused on (a) image classification, (b) object bounding box localization or (c) semantic pixel-level segmentation, we focus on (d) segmenting individual object instances.[1]

In our project, we will compare the performance of three different methods for instance segmentation and analyse their results.

Mask R-CNN[2] is based on Faster R-CNN. The output of Faster R-CNN are class labels and bounding box. And Mask R-CNN add another branch to estimate the mask of the object after purposed the region of interest (ROI). There are also some improvement on network architecture to make the model more suitable for instance segmentation task. RoIAlign layer improved the accuracy of the ROI and let the mask cover on the right location. The AP score increased 1.1 due to the applied of RoIAlign layer. Rather using FCIS loss function that some time attempt to cut the pixel of the same class object into a different instance, Mask R-CNN purposed a new loss function that can prevent the competition between the same class when applying the mask on them.

Based on Mask R-CNN, the PANet[3] proposes three features to improve the performance. First, they create bottom-up path augmentation, to shorten information path and enhance feature pyramid in low-levels, which is really useful for localization.
Second, they develop adaptive feature pooling, to recover broken information path between each proposal and all feature levels. It is a simple component to aggregate features from all feature levels for each proposal, avoiding arbitrarily assigned results.
Finally, they add tiny fully-connected layers for mask prediction, which possess complementary properties to

FCN. It is helpful to differentiate instances and recognize separate parts belonging to the same object.

GMIS[4] generate pixel affinities based on CNN for semantic segmentation. GMIS splits the task of instance segmentation into two sequential steps. The first step utilizes CNN, which specifically is Deeplabv3, to obtain class information and pixel affinity of the input image, while the second step applies the graph merge algorithm on those results to generate the pixel-level masks for each instance, which is a novel proposal-free instance segmentation scheme, where semantic information and pixel affinity information are both used to derive instance segmentation results.
The method demonstrated that even with a simple graph merge algorithm, we can outperform other methods, including proposal-based ones. It clearly shows that proposal- free methods can have comparable or even better performance than proposal- based methods.
The method also shows that semantic segmentation network is reasonably suitable for pixel affinity prediction with only the meaning of the output changed.

We expect to use the following 6 metrics will be used for characterizing the performance in our experiments:
Average Precision:

AP : AP at IoU=.50:.05:.95 (primary challenge metric)
AP_{50} : AP at IoU=.50
AP_{75} : AP at IoU=.75

AP Across Scales:

AP_S : AP for small objects: area < 32^2
AP_M : AP for medium objects: 32^2 < area < 96^2
AP_L : AP for large objects: area > 96^2

[1] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European conference on computer vision*. Springer, Cham, 2014.
[2] He, Kaiming, et al. "Mask r-cnn." *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017.
[3] Liu, Shu, et al. "Path aggregation network for instance segmentation." *Proceedings of the IEEE Conference on Computer* Vision *and Pattern Recognition*. 2018.
[4] Liu, Yiding, et al. "Affinity derivation and graph merge for instance segmentation." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.