Review for CenterMask: single shot instance segmentation with point representation

Irem Arpag

May 26, 2021

Since requiring localization, classification, and segmentation of each instance in an image all in one, and although it shows the characteristic of both, instance segmentation is thought to be more complex and harder than object detection and semantic segmentation, as one of the challenging computer vision tasks. Despite the popularity of two-stage object detection in recent years, it is not preferred too much by the instance segmentation methods mainly because of two outstanding problems that are differentiating object instances and preserving pixel-wise location information. To overcome these issues, the authors of this article present a model composed of two branches that are "local shape prediction" by separating objects locally "global saliency generation" by segmenting the full image in a pixel-to-pixel manner and then combined them to form the final instance mask that completed each other in a satisfying way that is coarse but instance-aware local shape overlaps precise but instance unaware global saliency map.

Because of adapting local shape information from object center points, the model called shortly CenterMask and accomplished 34,5 mask AP with a speed of 12.3 fps on COCO dataset, by indicating a perfect speed-accuracy consistency. With its scratch only training, anchor-box free and one-stage modeling, it is simple but fast and efficient at the same time. Adapting to other one-stage object detectors and achieving good results demonstrate its generation capacity. Another significant contribution of the method to the domain is to present novel local shape branch that is separating objects even in overlapping positions and global saliency branch that is separating the foreground from the background at pixel level. Both branches showing good results separately is a good indicator of their concurrent success even in more complex and hard over-lapping conditions which is confirmed with extensive ablation studies.

Although two-stage "Mask R-CNN" and one-stage "Tensor Mask" achieve a higher AP than "CenterMask" on SOTA comparison, because of their 5 times slowness, it might not be accepted as a weakness.

The CenterMask is evaluated on the challenging MS COCO instance segmentation dataset. To analyze the core factors effect on sensitivity and accuracy of the model, some ablation studies are performed. Size of local shape displays that larger shape brings more gains, but not in large amount, and another factor larger backbone provides 1.4 gains compared with the smaller one. Testing the model with and without local shape branch, it gains 10 points with it, while the same practice with global saliency branch brings 5 points. The class-specific setting of global saliency brings 2.4 extra points comparing to class-agnostic setting, and a final ablation study on direct supervision brings 0.5 point. Visualization results with and without proposed two branches clearly confirm that CenterMask by using two branches together, prevents the weaknesses such as artifacts on the overlapping area completely. When comparing with state-of-art YOLACT, CenterMask has both a faster speed and a higher AP.

Since global saliency branch resembles to semantic segmentation in some extend, it may be a motivation of some one-stage panoptic segmentation research for future works. Besides its easily embedding mechanism, CenterMask is also a simple, fast and accurate model with its two parallel, crucial components; local shape and global saliency branches that make instance level recognition more convenient and robust.