Boundary IoU: Improving Object-Centric Image Segmentation Evaluation

Bowen Cheng[1*] Ross Girshick[2] Piotr Dollár[2] Alexander C. Berg[2] Alexander Kirillov[2]

[1]UIUC [2]Facebook AI Research (FAIR)

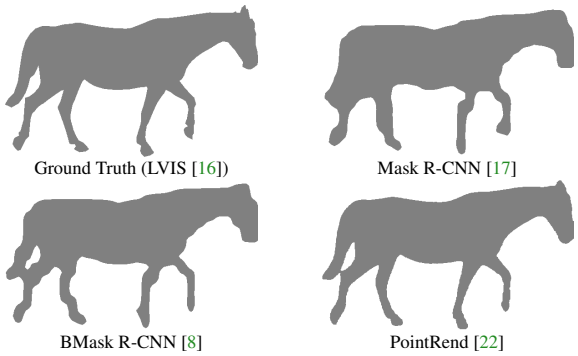
Abstract

We present Boundary IoU (Intersection-over-Union), a new segmentation evaluation measure focused on boundary quality. We perform an extensive analysis across different error types and object sizes and show that Boundary IoU is significantly more sensitive than the standard Mask IoU measure to boundary errors for large objects and does not over-penalize errors on smaller objects. The new quality measure displays several desirable characteristics like symmetry w.r.t. prediction/ground truth pairs and balanced responsiveness across scales, which makes it more suitable for segmentation evaluation than other boundary-focused measures like Trimap IoU and F-measure. Based on Boundary IoU, we update the standard evaluation protocols for instance and panoptic segmentation tasks by proposing the Boundary AP (Average Precision) and Boundary PQ (Panoptic Quality) metrics, respectively. Our experiments show that the new evaluation metrics track boundary quality improvements that are generally overlooked by current Mask IoU-based evaluation metrics. We hope that the adoption of the new boundary-sensitive evaluation metrics will lead to rapid progress in segmentation methods that improve boundary quality.[1]

1. Introduction

The Common Task Framework [27], in which standardized tasks, datasets, and evaluation metrics are used to track research progress, yields impressive results. For example, researchers working on the instance segmentation task, which requires an algorithm to delineate objects with pixel-level binary masks, have improved the standard Average Precision (AP) metric on COCO [28] by an astonishing 86% (relative) from 2015 [12] to 2019 [24].

However, this progress is not equal across all error modes, because different evaluation metrics are sensitive (or insensitive) to different types of errors. If a metric is used for a prolonged time, as in the Common Task Framework, then the corresponding sub-field most rapidly resolves the types of errors to which this metric is sensitive. Research directions that improve other error types typically advance more slowly, as such progress is harder to quantify.



	Mask R-CNN	BMask R-CNN	PointRend
Mask IoU	89%	92% (+3%)	97% (+8%)
Boundary IoU	69%	78% **(+9%)**	91% **(+22%)**

Figure 1: Given the bounding box for a horse, the mask predicted by Mask R-CNN scores a high Mask IoU value (89%) relative to the ground truth despite having low-fidelity, blobby boundaries. The recently proposed BMask R-CNN [8] and PointRend [22] methods predict masks with higher fidelity boundaries, yet these clear visual improvements only marginally improve Mask IoU (+3% and +8%, respectively). In contrast, our proposed **Boundary IoU** measure demonstrates greater sensitivity to boundary errors, and thus provides a clear, quantitative gradient that rewards improvements to boundary segmentation quality.

This phenomenon is at play in instance segmentation, where, among the multitude of papers contributing to the impressive 86% relative improvement in AP (*e.g.*, [41, 4, 1, 19, 25]), only a few address mask boundary quality.

Note that mask boundary quality is an essential aspect of image segmentation, as various downstream applications directly benefit from more precise object segmentations [39, 33, 34]. However, the dominant family of Mask R-CNN-based methods [17] are well-known to predict low-fidelity, blobby masks (see Figure 1). This observation suggests that the current evaluation metrics may have limited sensitivity to mask prediction errors near object boundaries.

To understand why, we start by analyzing Mask Intersection-over-Union (Mask IoU), the underlying mea-

*Work done during an internship at Facebook AI Research.

[1]Project page: https://bowenc0221.github.io/boundary-iou