PolarNeXt: Rethink Instance Segmentation with Polar Representation

Supplementary Material

Here we provide more details on the design principles and experimental results for our proposed method. An overview of this supplementary material is as follows:

- Section 1: More Details for Our Baseline PolarMask.
- Section 2: More Details for Representation Errors.
- Section 3: More Details for Our Proposed PolarNeXt.
- Section 4: What Makes for PolarNeXt?
- Section 5: Qualitative Results.

1. More Details for Our Baseline PolarMask

1.1. Polar Representation

PolarMask [7] formulates instance segmentation as starting point classification and dense distance regression. Given a starting point $\hat{s}(x, y)$ and a distance set of m rays $\hat{D} = \{\hat{d}_i | i = 1, 2, ..., m\}$, the coordinate of each vertex $v_i(x_i, y_i)$ of predicted polygon $\hat{P}$ can be formulated as follows:

$$\begin{cases} x_i = \cos\theta_i \times \hat{d}_i + x \\ y_i = \sin\theta_i \times \hat{d}_i + y \end{cases}, \tag{1}$$

where θ_i is the fixed angle between the i-th ray and polar axis. Starting from v_1 to v_m, these vertices are connected one by one to finally assemble predicted polygon $\hat{P}$. In this paper, our proposed PolarNeXt inherits this representation without any modification.

1.2. Network Architecture

PolarMask is directly constructed on one-stage object detector FCOS [6], with extended prediction heads. Specifically, the prediction heads consist of four parts: a classification head for starting point, a regression head for dense distance, a Centerness head for Polar Centerness, and an auxiliary box head. Experimental results indicate that removing the auxiliary box head results in a 0.1∼0.5% AP drop in PolarMask. In contrast, this operation has no impact on performance in our proposed PolarNeXt, while slightly reducing computational cost and inference time. Therefore, we have opted to remove this head for acceleration and simplification.

1.3. Center Prior

The center prior fully dictates the sample decisions of PolarMask during training. In label assignment, each instance is distributed to a specified FPN [5] layer based on its size (*e.g.*, the size range is [0, 64] for P3 layer and [64, 128] for P4 layer). Then, Center Sampling is applied on the corresponding FPN layer of each instance, assigning positive

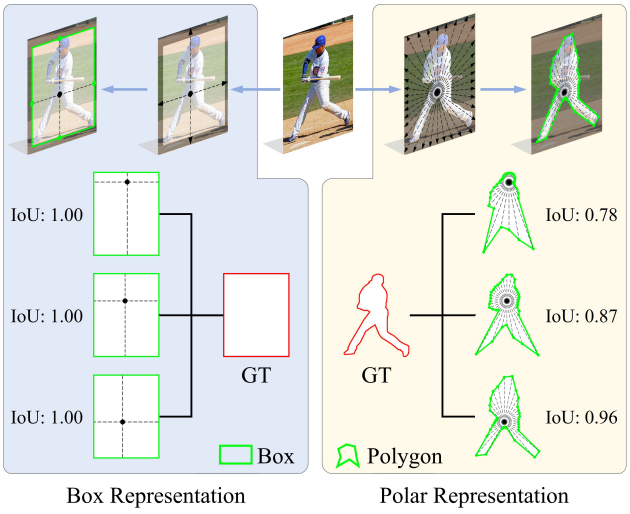


Figure 1. Illustration of the divergence between Box Representation and Polar Representation. The upper part compares the workflows of these two representations, and the lower part compares the boxes/polygons constructed from different starting points. "GT" is the abbreviation for Ground Truth.

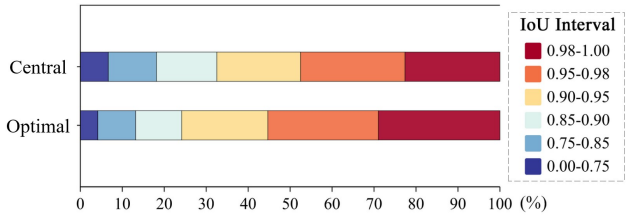


Figure 2. Comparison of two types of starting point selection. Each color bar corresponds to an IoU interval, with its length representing the proportion of instances within that interval.

labels to the samples around the center point[1]. In sample weighting, Polar Centerness serves as a quality score to modulate the weight of each positive sample. Assuming target distance set $D = \{d_i | i = 1, 2, ..., m\}$ and predicted distance set $\hat{D} = \{\hat{d}_i | i = 1, 2, ..., m\}$, Polar Centerness can be calculated as follows:

$$\text{Polar Centerness} = \sqrt{\frac{\min(\{d_1, d_2, ..., d_m\})}{\max(\{d_1, d_2, ..., d_m\})}}. \tag{2}$$

Notably, if a ray has no intersection with the contour, its distance will be set to 0. This rule enforces a Centerness

[1]Center Point: the center point typically refers to the instance mass center in Polar Representation, while it denotes the bounding box center in Box Representation.

Method	AP	AP_S	AP_M	AP_L
Baseline	29.1	13.3	31.3	42.6
Image-aligned	31.3	13.9	33.9	47.1
Union-aligned	**33.9**	**16.4**	**35.9**	**49.4**

Table 1. Analysis of different strategies for coordinate alignment. Image-aligned: alignment using the entire image; Union-aligned: alignment using the Union Box.

Quality Score	AP	AP_{50}	AP_{75}
Centerness	29.1	49.6	29.7
Polar IoU	29.4	50.6	30.0
1 - Cost	29.8	51.8	30.1
RMask IoU	**30.3**	**51.2**	**31.2**

Table 2. Analysis of different quality scores for sample weighting.

of 0 for the samples outside. In this paper, we apply Center Sampling across all FPN layers and replace Centerness with our proposed RMask IoU as the quality score.

1.4. Polar IoU

Polar IoU is a metric for polygonal assessment, which compares the consistency between distance sets D and $\hat{D}$. Polar IoU and its variation Polar IoU Loss can be formulated as follows:

$$\text{Polar IoU} = \frac{\sum_{i=1}^{m} \min(d_i, \hat{d}_i)}{\sum_{i=1}^{m} \max(d_i, \hat{d}_i)}, \tag{3}$$

$$\text{Polar IoU Loss} = \log \frac{\sum_{i=1}^{m} \max(d_i, \hat{d}_i)}{\sum_{i=1}^{m} \min(d_i, \hat{d}_i)}. \tag{4}$$

Obviously, these two formulas are independent of the instance contour, which exhibits significant assessment blindness to representation errors. In this paper, we propose RMask IoU and RMask IoU Loss for more reliable polygonal assessment.

2. More Details for Representation Errors

2.1. The Source of Representation Errors

As shown in Fig. 1, Polar Representation is conceptually a generalized form of Box Representation, using more concentric rays. However, a noteworthy observation is that Box Representation does not exhibit representation errors, whereas Polar Representation does. In light of this, we investigate the divergence between these two representations to explain the source of representation errors. In Box Representation, the position of starting points has no impact on the quality of constructed boxes, as the ground truth is regular. Conversely, in Polar Representation, the position of starting points is closely related to the capacity of polygons

s.r./ k	1.5/6	1.5/12	**1.5/9**	1.2/9	1.8/9
AP	33.8	33.7	**33.9**	33.1	33.8

Table 3. Performance impact of varying radius of Center Sampling (*s.r.*) and number of positive samples (*k*).

Ratio	1:1:1:0.5	**1:1:1:1**	2:1:1:1	1:5:2:1
AP	33.7	**33.9**	33.3	33.4

Table 4. Effectiveness of different combinations of loss coefficients.

to capture boundary details of the intricate ground truth. If an inappropriate starting point is selected, the deviation between its bounding polygon and instance contour, *i.e.*, the representation error, will be amplified. As a result, selecting the starting points with minimal representation errors may positively impact the segmentation accuracy.

2.2. The Motivation of the APSD Strategy

In this section, a statistical experiment is conducted to further demonstrate the effectiveness of selecting the points with minimal representation errors. We take the 36334 instances in MS COCO val [4] as study cases and record their IoU between instance contours and bounding polygons constructed from two types of starting points: the center point (Central) and the optimal point (Optimal). According to the IoU values, we divide these instances into six intervals and calculate their proportion within each interval. As shown in Fig. 2, the polygons constructed from the optimal points are significantly superior to those constructed from the center points, with a greater proportion in high-IoU intervals. These observations motivate our proposed Adaptive Polygonal Sample Decision strategy to enable optimal starting point selection during inference.

3. More Details for Our Proposed PolarNeXt

3.1. Ablation Studies

Strategies for Coordinate Alignment. As introduced in Section 3.3 of the main paper, coordinate alignment is necessary before rasterization, given that no pre-predicted RoIs are available in one-stage detectors. Besides our proposed Union-aligned strategy, another option is to use the entire image as the window for alignment. As shown in Tab. 1, although the Image-aligned strategy brings some performance improvement, it is less effective than the Union-aligned strategy. Notably, the performance improvement decreases progressively across large, middle, and small objects (+4.5%/+2.6%/+0.6% AP). This may arise from the fact that as the window size increases, the alignment operation results in more missed details in smaller objects.

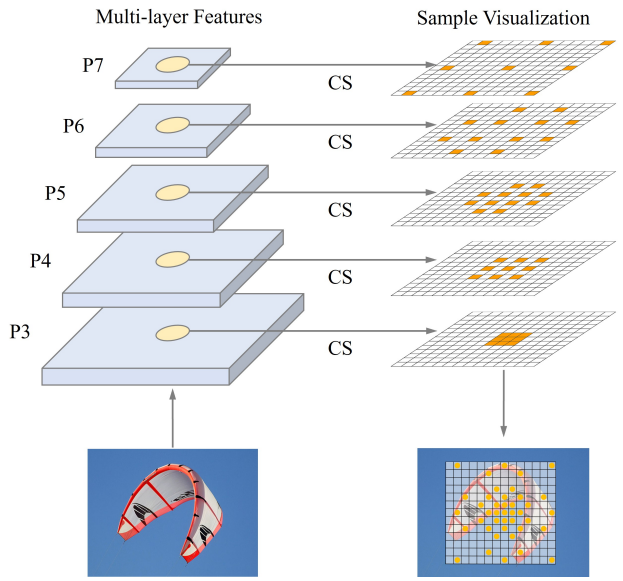


Figure 3. Visualization of Multi-layer Center Sampling (MCS). The multi-layer features are extracted on the input image by FPN. On each layer, Center Sampling (CS) is applied for a group of candidate samples. These samples are visualized on the right and combined onto a single plane for clearer understanding.

Quality Scores for Sample Weighting. As shown in Tab. 2, we compare more feasible quality scores for sample weighting. Among them, the strategy that directly replaces Centerness with RMask IoU yields the largest performance improvement, with an increase of 1.2% in AP.

3.2. Hyperparameters Tuning

Radius of Center Sampling and Number of Positive Samples. In Tab. 3, we further investigate the hyperparameters of our proposed APSD, *i.e.*, the radius of Center Sampling (*s.r.*) and the number of positive samples (k). Experimental results indicate that the effectiveness of {*s.r.*=1.5 / k=9} surpasses other versions.

Loss coefficients. In Tab. 4, we explore some commonly used combinations of loss coefficients λ_{cls}, λ_{reg}, λ_{poly}, and λ_{sre}. Experimental results indicate that the {1:1:1:1} ratio achieves the best performance.

3.3. Why Center Prior still Works?

As demonstrated in Section 4.3 of the main paper, the center prior still has a positive influence on sample decisions, even though its role has diminished. Here, we provide two additional experiments to further support this point:

Visualization of Multi-layer Center Sampling. As illustrated in Fig. 3, we perform a visualization of Multi-layer Center Sampling (MCS) for a clear understanding of the candidate selection in APSD. It is evident that MCS selects samples with stronger receptive fields on each FPN layer, while also increasing coverage over instances to achieve more candidates.

APSD	AP	AP_{50}	AP_S	AP_M	AP_L
w/o.	37.7	57.3	19.9	40.8	50.5
w/.	**38.9**	**57.9**	**22.2**	**42.0**	**51.1**

Table 5. Effectiveness of APSD on the object detector FCOS. "w/o." and "w/." denote the absence or presence of APSD, respectively.

(a) Classification Confidence (b) Polar Centerness

Figure 4. Visualization of Classification Confidence and Polar Centerness.

Correlation between Classification and Center Prior. Furthermore, we compare the heatmaps of Classification Confidence and Polar Centerness in Fig. 4. Within the red circles, high-confidence predictions locate near the peak areas of Centerness. We believe that features closer to the instance center tend to contain more complete instance information, which advances the network to recognize the objects.

4. What Makes for PolarNeXt?

4.1. TIDE Error Analysis

Fig. 6 shows the error analysis through TIDE [1] and comparisons between PolarMask and PolarNeXt. In detail, our proposed PolarNeXt exhibits lower *Miss* error and *Loc* error. First, the reduction of *Miss* error by 1.3% and FN by 4.7% indicates that PolarNeXt can identify foreground objects more effectively than PolarMask, reducing the probability of missing detections. Second, the 2.0% decrease in *Loc* error shows that the bounding polygons predicted by PolarNeXt are more accurate than those predicted by PolarMask, enabling improved instance localization.

4.2. Mitigation of Miss Error

To further investigate the source of the mitigation of *Miss* error, our proposed APSD strategy is migrated to the classical object detector FCOS. Specifically, the matching cost is obtained by a weighted summation of Focal Loss, Box IoU Loss, and L1 Loss. Meanwhile, Box IoU is used as the quality score for sample weighting. As shown in Tab. 5, APSD

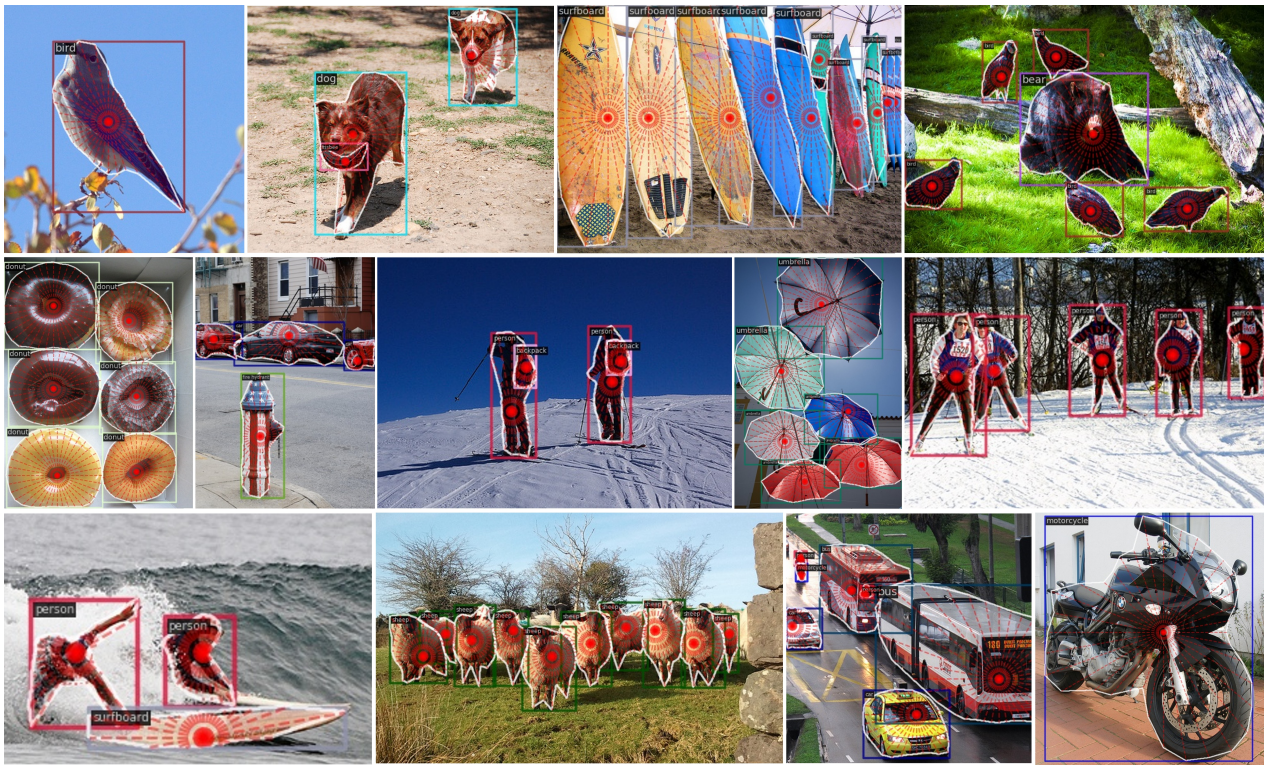
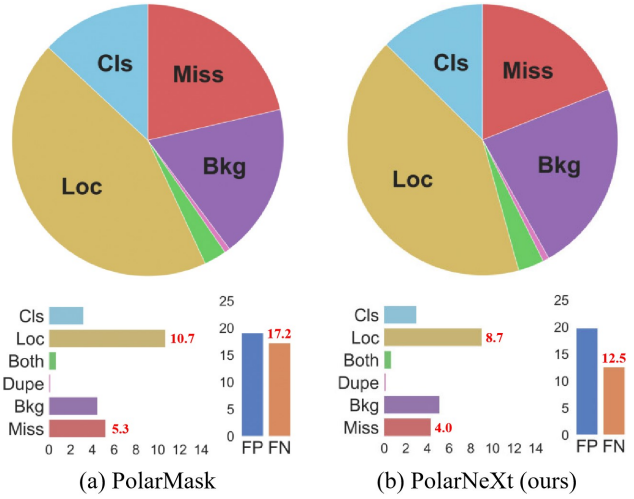


Figure 5. Qualitative Results of PolarNeXt on MS COCO val images.



(a) PolarMask (b) PolarNeXt (ours)

Figure 6. TIDE Error Analysis. We adopt TIDE [1] to compare the errors of PolarMask and our proposed PolarNeXt. *Cls*: classification error; *Loc*: localization error; *Miss*: missing detections; *Bkg*: background detections; *Dupe*: duplicated detections; *Both*: *Cls*+*Loc* error.

continues to have a positive influence on object detection tasks, slightly increasing the detection accuracy by 1.2% AP. Therefore, we speculate that the mitigation of *Miss* error may stem from the improvement in detection performance.

class (higher)	airplane	fork	umbrella	dog	cycle
PolarMask	31.4	7.1	39.7	46.9	26.6
PolarNeXt	39.7	13.8	45.8	52.0	31.2
impr.	**+8.3**	**+6.7**	**+6.1**	**+5.1**	**+4.6**
class (lower)	hydrant	parkmeter	broccoli	mouse	hotdog
PolarMask	59.1	45.4	21.0	58.6	26.1
PolarNeXt	59.5	45.9	22.0	59.7	27.9
impr.	**+0.4**	**+0.5**	**+1.0**	**+1.1**	**+1.8**

Table 6. Class-wise comparison on MS COCO val. All values refer to mask AP in this table. "*impr.*" is the abbreviation for improvement.

4.3. Mitigation of Loc Error

The mitigation of *Loc* error is relatively easier to explain. As discussed in Section 4.4 of the main paper, PolarNeXt significantly improves the boundary quality over PolarMask, achieving a 3.5% increase in Boundary AP [2]. Moreover, we conduct a class-wise comparison experiment on MS COCO val, where the top five classes with higher and lower improvements are presented in Tab. 6. Obviously, the classes with higher improvements feature complex boundaries with more disconnected and concave regions, while the classes with lower improvements exhibit regular boundaries. This suggests that our proposed ap-

proach is particularly effective for some instances with complex contours.

5. Qualitative Results

Fig. 5 shows the qualitative results of PolarNeXt. Here, the network uses R50-FPN [3] as the backbone, trained with a 3x schedule. Experimental results show that our proposed PolarNeXt can generate precise bounding polygons for instance segmentation. For crowd and dense scenes, PolarNeXt can also distinguish different instances well.

References

[1] Daniel Bolya, Sean Foley, James Hays, and Judy Hoffman. Tide: A general toolbox for identifying object detection errors. In *ECCV*, pages 558–573, 2020. 3, 4

[2] Bowen Cheng, Ross Girshick, Piotr Dollár, Alexander C Berg, and Alexander Kirillov. Boundary iou: Improving object-centric image segmentation evaluation. In *CVPR*, pages 15334–15342, 2021. 4

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 5

[4] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755, 2014. 2

[5] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, pages 2117–2125, 2017. 1

[6] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: A simple and strong anchor-free object detector. *IEEE TPAMI*, 44(4):1922–1933, 2020. 1

[7] Enze Xie, Peize Sun, Xiaoge Song, Wenhai Wang, Xuebo Liu, Ding Liang, Chunhua Shen, and Ping Luo. Polarmask: Single shot instance segmentation with polar representation. In *CVPR*, pages 12193–12202, 2020. 1