Dense Contrastive Loss for Instance Segmentation (*Supplementary Material*)

Hang Chen[1]
chenhang20@mails.tsinghua.edu.cn
Chufeng Tang[1]
tcf18@mails.tsinghua.edu.cn
Xiaolin Hu[†1,2]
xlhu@mail.tsinghua.edu.cn

[1] Depart. of Comp. Sci. & Tech., State Key Lab of Intell. Tech. & Sys., THU-Bosch JCML Center, BNRist, Institute for AI, THBI, IDG/McGovern Institute for Brain Research, Tsinghua University, Beijing, China

[2] Chinese Institute for Brain Research (CIBR), Beijing, China

1 Comparison of Error Analysis Tools

We compare several error analysis tools in detail in Table S1. The main difference between our tool and previous works is that we focus on segmentation errors in the instance segmentation task. The previous works are mainly designed for object detection and adapted for instance segmentation by simply replacing box IoU with mask IoU. Consequently, segmentation errors are coupled with other types of errors (*e.g.*, localization or classification errors) in these works, making it impossible to analyze the segmentation quality in isolation and to draw finer-grained conclusions.

Feature	COCO [1]	UpperBound [3]	TIDE [2]	Ours
Compact Summary of Error Types	✘	✔	✔	✔
Isolates Error Contribution	✘	✘	✔	✔
Dataset Agnostic	✘	✔	✔	✔
Segmentation Errors	✘	✘	✘	✔
Designed for Instance Segmentation	✘	✘	✘	✔

Table S1: Comparison of error analysis tools. ✔ indicates the toolkit has the specific feature, and ✘ indicates that it does not. Table extended from TIDE [2]. Our error analysis tool focuses on segmentation errors, complementing previous works.

Alternatively, our tool can also output compact summary of error types (*i.e.*, one value per error type) and isolate error contribution, making it easier to compare different types of errors. It does not rely on additional annotations and therefore can be applied to any instance segmentation dataset.

[†] Corresponding Author.
© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

2 Derivation of Segmentation Errors

We provide a detailed derivation on how to decouple segmentation error pixels into three orthogonal sets. The segmentation error pixels are defined as the symmetric difference of the prediction D_i and its corresponding ground-truth mask G_i^*:

$$
\begin{aligned}
\mathcal{P}_i &= D_i \ominus G_i^* \\
&= (D_i - G_i^*) \cup (G_i^* - D_i) \\
&= \big((D_i - G_i^*) \cap (G - G_i^*)\big) \cup \big((D_i - G_i^*) \cap \overline{(G - G_i^*)}\big) \cup (G_i^* - D_i) \\
&= \big(D_i \cap \overline{G_i^*} \cap G \cap \overline{G_i^*}\big) \cup \big(D_i \cap \overline{G_i^*} \cap \overline{(G \cap \overline{G_i^*})}\big) \cup (G_i^* - D_i) \\
&= \big(D_i \cap (G \cap \overline{G_i^*})\big) \cup \big(D_i \cap \overline{G_i^*} \cap (\overline{G} \cup G_i^*)\big) \cup (G_i^* - D_i) \\
&= \big(D_i \cap (G - G_i^*)\big) \cup \big((D_i \cap \overline{G_i^*} \cap \overline{G}) \cup \emptyset\big) \cup (G_i^* - D_i) \\
&= \underbrace{(D_i \cap (G - G_i^*))}_{\text{fFP}} \cup \underbrace{(D_i \cap \overline{G})}_{\text{bFP}} \cup \underbrace{(G_i^* - D_i)}_{\text{FN}},
\end{aligned} \tag{1}
$$

where $G = \bigcup_{i=1}^{M} G_i$, $\overline{X}$ means the complement set of X in the set of all image pixels. These three sets (denoted as $\mathcal{P}_{\text{fFP}}, \mathcal{P}_{\text{bFP}}, \mathcal{P}_{\text{FN}}$, respectively) are orthogonal, since

$$
\begin{aligned}
\mathcal{P}_{\text{fFP}} \cap \mathcal{P}_{\text{bFP}} &= \big(D_i \cap (G - G_i^*)\big) \cap (D_i \cap \overline{G}) \\
&= \big(D_i \cap G \cap \overline{G_i^*}\big) \cap (D_i \cap \overline{G}) \\
&= \emptyset,
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
\mathcal{P}_{\text{fFP}} \cap \mathcal{P}_{\text{FN}} &= \big(D_i \cap (G - G_i^*)\big) \cap (G_i^* \cap \overline{D_i}) \\
&= \emptyset,
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
\mathcal{P}_{\text{bFP}} \cap \mathcal{P}_{\text{FN}} &= (D_i \cap \overline{G}) \cap (G_i^* \cap \overline{D_i}) \\
&= \emptyset.
\end{aligned} \tag{4}
$$

3 More Error Analysis Results

We analyze the segmentation errors of more models on the COCO dataset in Figure S1. The error distribution varies between models. Models with similar structures (*e.g.*, HTC [4], Mask Scoring R-CNN [7], PointRend [8]) have similar error distributions, while those with different structures (*e.g.*, one-stage vs. multi-stage) have different error distributions.

Compared to the multi-stage models, CondInst [9] and SOLO [10] generated fewer FN errors and more bFP errors, while Mask2Former [5] and QueryInst [6] yielded more bFP errors. The results are as expected since CondInst and SOLO do not rely on the box proposals and can therefore predict more complete masks. The bFP error can be attributed to that instance kernels predicted by these models easily respond to the background features.

4 More Qualitative Results

We present some qualitative results on the Cityscapes validation set in Figure S2. Compared to the Mask R-CNN baseline, our method yielded higher quality and more complete masks,

Figure S1: Error analysis of different models on the COCO validation set. The error distribution varies between models.

#Anchor	#Positive	#Negative	AP	AP^{50}
32	32	32	**37.1**	63.2
64	32	32	36.7	62.6
32	64	32	36.8	63.4
32	32	64	35.5	61.1
16	32	32	37.0	**63.8**
32	16	32	35.7	62.3
32	32	16	36.0	62.8

Table S2: Results of different number of samples. Cityscapes-style segmentation AP is reported.

especially in complex scenarios, such as occlusion (Figure S2a,c), low illumination (Figure S2d). Our method also produced better results for the extreme crowded cases (see the yellow box in Figure S2f). Beyond that, we found that our predictions retain more details (see blue boxes in Figure S2b,d).

5 Different Number of Samples

Table S2 shows the results using different numbers of samples. AP decreases slightly when the number of samples varies. We attribute this to the trade-off between the number of samples and the number of valid RoIs.

References

[1] Coco analysis toolkit. http://cocodataset.org/#detection-eval. Accessed: 2022-06-28.

[2] Daniel Bolya, Sean Foley, James Hays, and Judy Hoffman. TIDE: A general toolbox for identifying object detection errors. In Andrea Vedaldi, Horst Bischof,

Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part III*, volume 12348 of *Lecture Notes in Computer Science*, pages 558–573. Springer, 2020. doi: 10.1007/978-3-030-58580-8_33. URL https://doi.org/10.1007/978-3-030-58580-8_33.

[3] Ali Borji and Seyed Mehdi Iranmanesh. Empirical upper bound in object detection and more. *CoRR*, abs/1911.12451, 2019. URL http://arxiv.org/abs/1911.12451.

[4] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. Hybrid task cascade for instance segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 4974–4983. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.00511. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Chen_Hybrid_Task_Cascade_for_Instance_Segmentation_CVPR_2019_paper.html.

[5] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. *CoRR*, abs/2112.01527, 2021. URL https://arxiv.org/abs/2112.01527.

[6] Yuxin Fang, Shusheng Yang, Xinggang Wang, Yu Li, Chen Fang, Ying Shan, Bin Feng, and Wenyu Liu. Instances as queries. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pages 6890–6899. IEEE, 2021. doi: 10.1109/ICCV48922.2021.00683. URL https://doi.org/10.1109/ICCV48922.2021.00683.

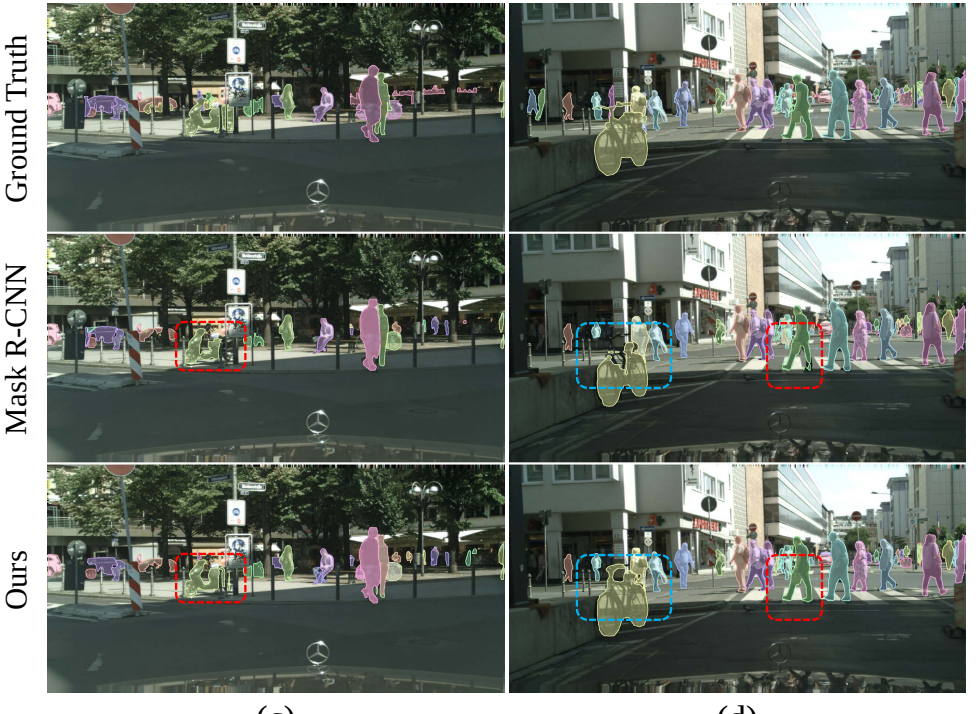
[7] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask scoring R-CNN. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 6409–6418. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.00657. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Huang_Mask_Scoring_R-CNN_CVPR_2019_paper.html.

[8] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross B. Girshick. Pointrend: Image segmentation as rendering. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 9796–9805. Computer Vision Foundation / IEEE, 2020. doi: 10.1109/CVPR42600.2020.00982. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Kirillov_PointRend_Image_Segmentation_As_Rendering_CVPR_2020_paper.html.

[9] Zhi Tian, Chunhua Shen, and Hao Chen. Conditional convolutions for instance segmentation. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I*, volume 12346 of *Lecture Notes in Computer Science*, pages 282–298. Springer, 2020. doi: 10.1007/978-3-030-58452-8_17. URL https://doi.org/10.1007/978-3-030-58452-8_17.

[10] Xinlong Wang, Tao Kong, Chunhua Shen, Yuning Jiang, and Lei Li. SOLO: segmenting objects by locations. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XVIII*, volume 12363 of *Lecture Notes in Computer Science*, pages 649–665. Springer, 2020. doi: 10.1007/978-3-030-58523-5_38. URL https://doi.org/10.1007/978-3-030-58523-5_38.

(a) (b)



(c) (d)

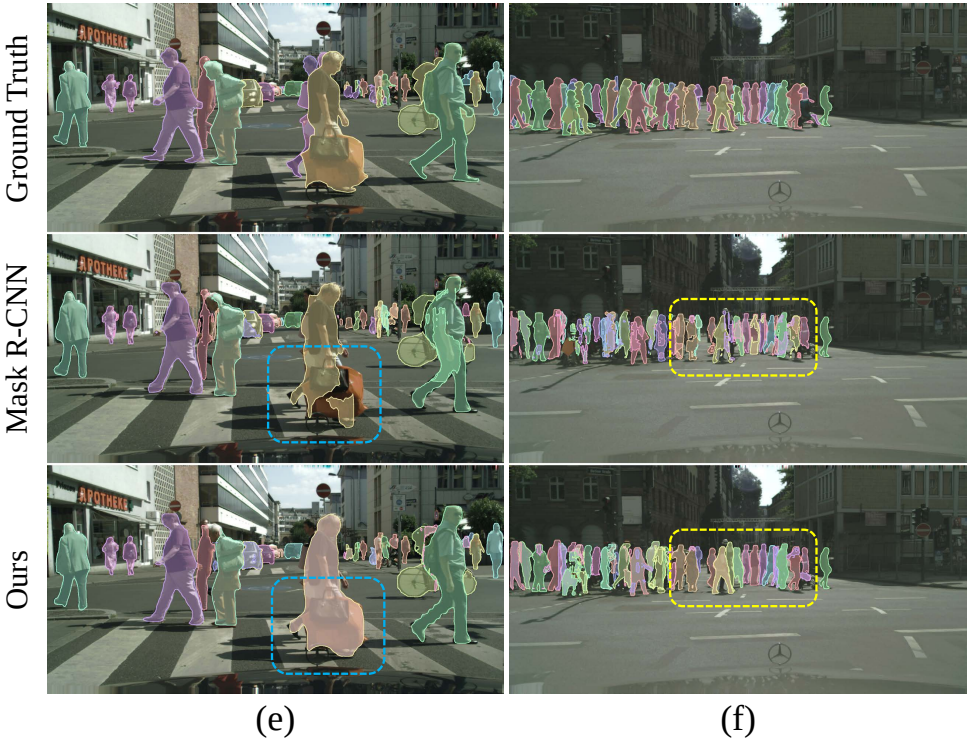


Figure S2: Qualitative Results on the Cityscapes validation set. Our method yielded higher quality and more complete masks compared to the Mask R-CNN baseline. Best viewed in color with zoom-in.