Supplementary Material for Exemplar-FreeSOLO: Enhancing Unsupervised Instance Segmentation with Exemplars

Taoseef Ishtiak,[1] Qing En,[1] Yuhong Guo[1,2]
[1]Carleton University, Ottawa, Canada [2]CIFAR AI Chair, Amii, Canada

Abstract

The supplementary material is organized as follows: Section 1 presents additional parameter analysis results, including the impact of temperature parameter τ, the impact of parameters α and β, and the impact of the weight of the proposed loss function λ_{eec}. Section 2 provides additional design choices results. Section 3 presents additional supervised fine-tuning results. Section 4 provides additional qualitative evaluations to illustrate the effect of the proposed Exemplar-FreeSOLO. All experimental results are obtained on the MS COCO val2017 dataset.

1. Additional Parameter Analysis Results

1.1. Impact of the Temperature τ

To demonstrate the impact of the temperature τ of the exemplar embedding contrastive loss L_{eec}, we summarize the results in Figure 1. It demonstrates that τ plays a crucial role in regulating the strength of penalties on hard negative samples within our contrast learning paradigm. As shown in Figure 1, we present the performance analysis for object detection and instance segmentation tasks with varying values of τ. The results demonstrate that our proposed Exemplar-FreeSOLO achieves the best performance for each metric on both tasks when τ is set to 0.02. Moreover, when the value of τ is set too high or too low, the experimental results deteriorate. These findings suggest that an intermediate value of τ is required within a reasonable range to regulate the degree of penalty for hard negative samples in a balanced manner.

1.2. Impact of Parameters α and β

We summarize the experimental results of the impact of the parameters α and β on the MS COCO val2017 dataset in Figure 2. These parameters are utilized to balance the selection of positive and negative samples in the exemplar embedding contrastive module. In particular, different values of α and β represent the selection criterion for positive or negative samples with various degrees of confidence. In this section, we set α to [0.5, 0.6, 0.7, 0.8, 0.9] and β to [0.1, 0.2, 0.3, 0.4, 0.5]. We can see that when we set α to 0.8 and β to 0.3, the best values are achieved for all metrics except for AP_s and AP_m. Additionally, we observe that the experimental results decrease as the values of α and β are close to each other. This phenomenon is reasonable because as the values of α and β become closer, the choice of positive and negative samples becomes less discriminatory, thus impeding the effectiveness of our proposed loss function to learn the comparison information effectively. Consequently, the experimental phenomenon suggests that it is appropriate to choose two values with a large difference, while ensuring that the values are neither too large nor too small to enable a broader range of selection for positive and negative samples.

1.3. Impact of the weight of the proposed loss function λ_{eec}

We summarize the experimental results on the influence of the weight of the exemplar embedding contrastive loss λ_{eec} in Figure 3. The experiments are conducted by fixing the other hyperparameters and only changing the value of λ_{eec}. To justify our choice of the parameter, we report the performance of various λ_{eec} values in terms of AP_{75}, AP_s, AP_m and AP_l metrics. Figure 3 illustrates that we can obtain better experimental results by setting λ_{eec} to 0.9, 1.0 and 1.3. The experimental results demonstrate that setting λ_{eec} to less than 0.9 or more than 1.3 leads to a considerable drop in performance. For both the segmentation and detection tasks, the best results are achieved when λ_{eec} is set to 1.3 for all metrics. Compared to the other loss functions in our framework, which all have a weight of 1, the best results are achieved when λ_{eec} is similar to the weights of other loss functions. These findings highlight the importance of the exemplar embedding contrastive loss function in our proposed framework.

2. Additional Design Choices Results

We summarize the experimental results on the impact of different design choices in terms of AP value in Table 1. It

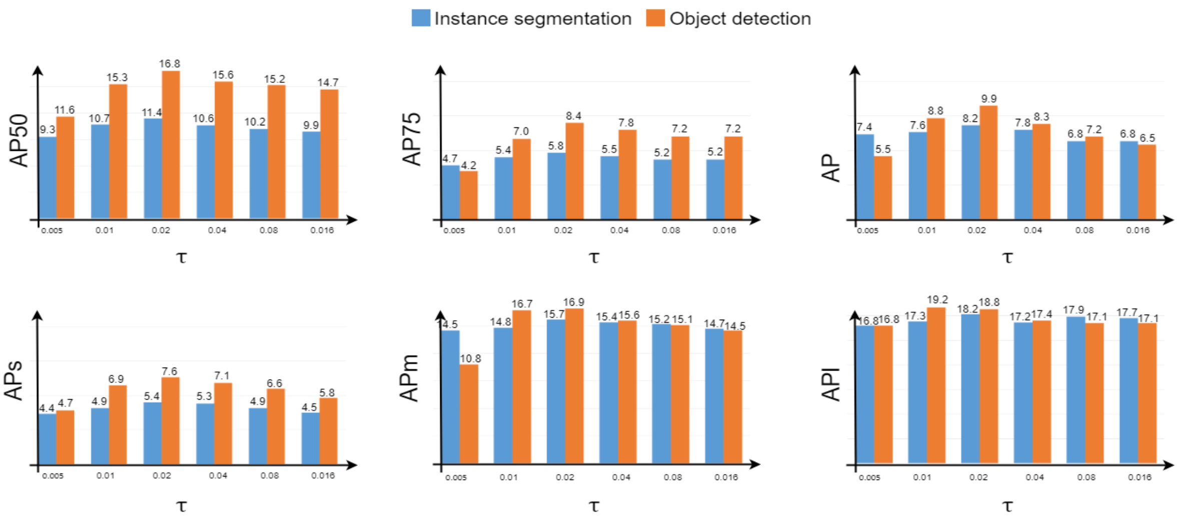

Figure 1. Impact of temperature τ in terms of AP_{50}, AP_{75}, AP, AP_s, AP_m and AP_l. We report unsupervised class-agnostic instance segmentation and unsupervised class-agnostic object detection results on MS COCO val2017.

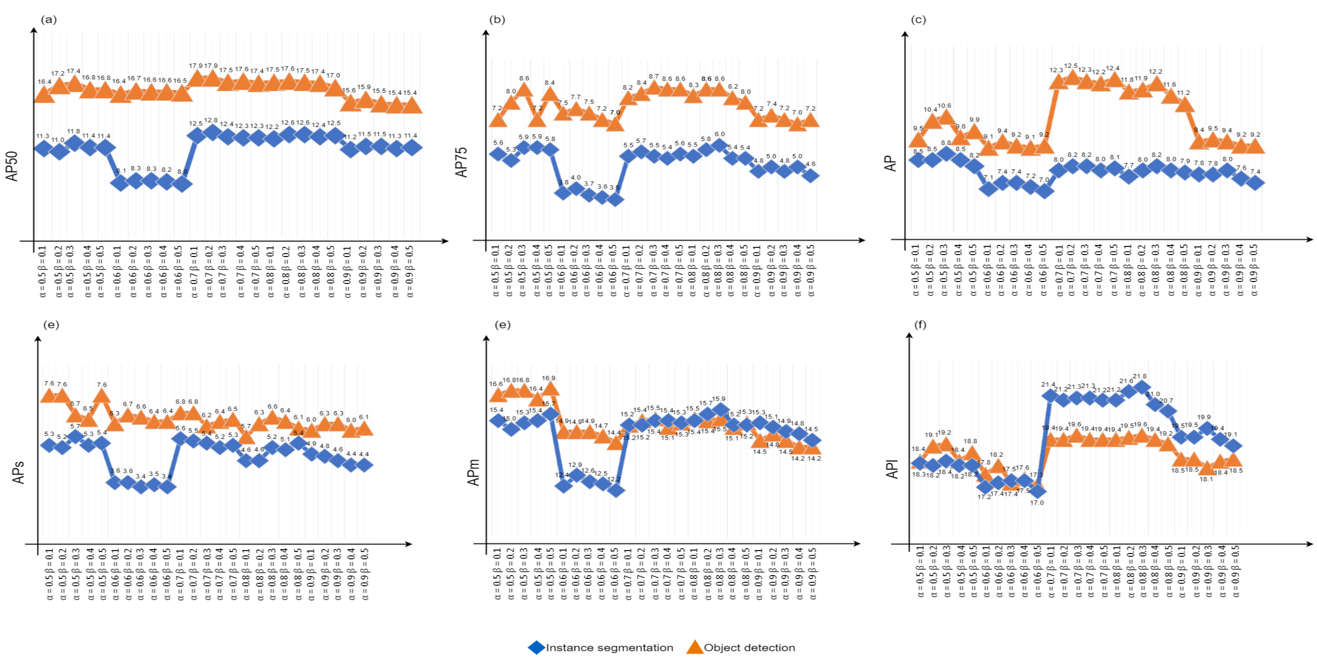

Figure 2. Impact of parameters α and β in terms of AP_{50}, AP_{75}, AP, AP_s, AP_m and AP_l. We report unsupervised class-agnostic instance segmentation and Unsupervised class-agnostic object detection results on MS COCO val2017.

can be observed that the AP values for instance segmentation and instance detection decrease to 7.8 and 10.7, respectively, when FreeSOLO is removed from the EKA module. On the other hand, removing Grabcut barely affects the instance segmentation results, with an AP of 8.5, but the instance detection results show a more significant drop to an AP of 11.1. Moreover, both segmentation and detection results experience a more substantial decrease when the EKA module is removed, with AP values of 6.7 and 9.9, respectively. In this case, the exemplar images are directly used to construct the exemplar pool. The experimental results demonstrate the effectiveness of the proposed EKA module. To validate our choice of using Grabcut in the EKA module, we experimented with replacing it with Graphcut [2] and Objectness [1] methods. Our results showed a certain degree of degradation with these alternatives, which further supports our decision to use Grabcut for exemplar object extraction.

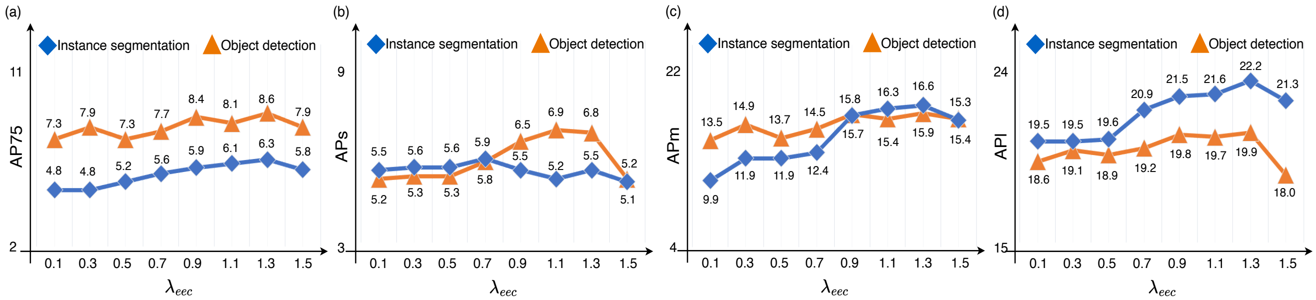

Figure 3. Impact of parameter λ_{eec} in terms of AP_{75}, AP_s, AP_m and AP_l. We report unsupervised class-agnostic instance segmentation and unsupervised class-agnostic object detection results on MS COCO val2017.

Figure 4. Visualized examples of our experimental results on unsupervised instance segmentation and object detection task.

3. Additional Supervised Fine-tuning Results

We present a summary of the experimental results by fine-tuning the segmentation model using a limited number of fully annotated images and segmentation masks. The AP values are presented in Table 2. Our framework outperforms FreeSOLO by 2.0 and 2.4 AP values, achieving AP values of 31.9 and 33.5, respectively, when fine-tuning with 5% and 10% masks. Additionally, Exemplar-FreeSOLO achieves higher AP values than FreeSOLO when fine-tuning with 5% and 10% COCO training images, achieving AP values of 23.8 and 26.1, respectively. Specifically, Exemplar-FreeSOLO achieves 1.8 and 0.5 AP values higher than FreeSOLO for the two respective scenarios. Our experimental results demonstrate that Exemplar-FreeSOLO, leveraging its proposed exemplar mechanism, can be an effective pre-trained instance segmentation model and yield better fine-tuning results compared to FreeSOLO.

4. Additional Qualitative Results

To validate the segmentation and detection performance of the proposed Exemplar-FreeSOLO, some visualization examples from the MS COCO val2017 dataset are shown in Figure 4. The results demonstrate that our proposed framework can achieve relatively good segmentation and detec-

	w/o FreeSOLO	w/o Grabcut	w/o EKA	Graphcut [2]	Objectness [1]	Ours
Segmentation	7.8	8.5	6.7	7.6	8.2	8.4
Detection	10.7	11.1	9.9	11.4	12.5	12.6

Table 1. Comparison of different design choices in the EKA module in terms of AP. **w/o FreeSOLO** and **w/o Grabcut** indicate that FreeSOLO or Grabcut is removed. **w/o EKA** indicate that both FreeSOLO and Grabcut are removed. **Graphcut** and **Objectness** indicate that using these two methods to replace Grabcut, respectively in our proposed approach. **Ours** indicates the choice in our paper, *i.e*. Grabcut+FreeSOLO.

	5% masks	10% masks	5% images	10% images
FreeSOLO	29.9	31.1	22.0	25.6
Exemplar-FreeSOLO	31.9	33.5	23.8	26.1

Table 2. Supervised instance segmentation with limited fully annotated images (5% and 10% training images) and limited segmentation masks (5% and 10% training masks) on MS COCO dataset in terms of AP.

tion results for targets with different contours and sizes. Our framework is also capable of accurately segmenting multiple occluded targets. We attribute this capability to the proposed exemplar mechanism, which provides top-down knowledge guidance and improves the segmentation model's discriminability.

References

[1] Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, and Philip Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. In *IEEE/CVF conference on computer vision and pattern recognition*, 2014. 2, 4

[2] Sudipta N Sinha. Graph cut algorithms in vision, graphics and machine learning an integrative paper. *UNC Chapel Hill*, 2004. 2, 4