Dual Decision Improves Open-Set Panoptic Segmentation

Hai-Ming Xu[1]
hai-ming.xu@adelaide.edu.au
Hao Chen[2]
stanzju@gmail.com
Lingqiao Liu[1]
lingqiao.liu@adelaide.edu.au
Yufei Yin[3]
yinyufei@mail.ustc.edu.cn

[1] The University of Adelaide,
Adelaide, Australia
[2] Zhejiang University
Zhejiang, China
[3] University of Science and Technology of China
Anhui, China

Abstract

Open-set panoptic segmentation (OPS) problem is a new research direction aiming to perform segmentation for both *known* classes and *unknown* classes, i.e., the objects ("things") that are never annotated in the training set. The main challenges of OPS are twofold: (1) the infinite possibility of the *unknown* object appearances makes it difficult to model them from a limited number of training data. (2) at training time, we are only provided with the "void" category, which essentially mixes the "unknown thing" and "background" classes. We empirically find that directly using "void" category to supervise *known* class or "background" classifiers without screening will lead to an unsatisfied OPS result. In this paper, we propose a divide-and-conquer scheme to develop a dual decision process for OPS. We show that by properly combining a *known* class discriminator with an additional class-agnostic object prediction head, the OPS performance can be significantly improved. Specifically, we first propose to create a classifier with only *known* categories and let the "void" class proposals achieve low prediction probability from those categories. Then we distinguish the "unknown things" from the background by using the additional object prediction head. To further boost performance, we introduce "unknown things" pseudo-labels generated from up-to-date models to enrich the training set. Our extensive experimental evaluation shows that our approach significantly improves *unknown* class panoptic quality, with more than 30% relative improvements than the existing best-performed method. Code is available at: https://github.com/HeimingX/OPS_dual_decision.

1 Introduction

Panoptic segmentation (PS) is attracting growing attention from the vision community since it was proposed by Kirillov et al. [17]. Such hot trend attributes to its ambitious goal for accommodating both semantic segmentation and instance segmentation in an unified framework and producing holistic scene parsing results [8, 9, 10, 17, 18, 19, 20, 28, 29, 30, 32].

Figure 1: Demonstration of OPS task setting. At training phase, an natural image and its ground truth segments (i.e., *known* class *thing* and *stuff*) of interest are given for model learning. Some *thing* objects that are not interesting or difficult to label are left as *void*, i.e., blank area. While the model is required to be able to segment not only *known* class objects, but also *unknown* class objects (denoted by orange color in Testing GT) for testing images, e.g., bottles, fork and toothbrush.

Most of the researches are built under a common closed-set assumption, i.e., the model only needs to segment the same class of objects appeared in the training set. However, such kind of systems will not be competent for the complex open-set scenario. For example, automatic driving can not identify abnormal objects will lead to catastrophic danger [1] and possible problems can even not be predictable in medical diagnosis [2]. Therefore, PS systems for dealing with open-set challenge are urgently demanded.

Open-set problem [11, 25, 26, 27] has been well explored in classification tasks which refer to the scenario that when some new classes unseen in training appear in testing, the recognition model is required to not only accurately classify known classes given in training but also effectively deal with the unknown classes.

The recent research [15] extends PS to a realistic setting and firstly defines the *open-set panoptic segmentation* (OPS) task. OPS takes categories given during training as *known* classes and requires the model to produce segments for both *known* and *unknown* class objects ("things") at testing phase, where the *unknown* classes are never annotated or even appeared in the training set. As examples shown in Figure 1, many bottles in training image (above the closet) are hard to be labeled pixel by pixel and given as the *void* area in ground truth segments. While for testing images, it is required to predict segments for different kinds of bottles and even the fork and toothbrush which never appeared during training.

The OPS task is challenging because, on the one hand, the appearance of *unknown* class objects is diverse and it would be hard to directly modeling *unknown* classes from the given training images. On the other hand, although the "void" category is available at training phase, the training samples in "void" category are too noisy to provide effective supervisions since the "unknown thing" class and the "background" are confounded together.

To tackle these two challenges, we choose to recognize the *unknown* class objects in a dual decision process. By coupling the *known* class discriminator with a class-agnostic object prediction head, we can significantly improve the performance of the OPS task. Specifically, we build up a *known* class classifier and suppress its predictions for "void" class proposals to compact the decision boundaries of *known* classes and empower the *known* class classifier the ability to reject non-known classes. Then, we further create a class-agnostic object prediction head to further distinguish "unknown things" from the background. Moreover, we propose to use the pseudo-labeling method to further boost the generalization ability of the newly added object prediction head. Extensive experimental results show that our approach has successfully achieved a new state-of-the-art performance on various OPS tasks.

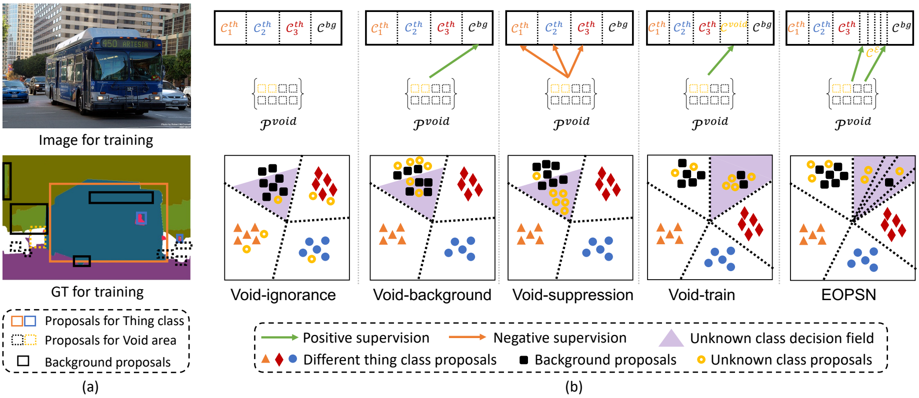


Figure 2: Illustration of existing OPS methods. (a) Proposal examples used in existing OPS methods. (b) Display of the usage of"void" class proposals in different OPS methods (top row) and how *unknown* class are predicted at testing phase (bottom row). $\mathcal{C}_i^{th}$, $\mathcal{C}^{bg}$, $\mathcal{C}^{void}$ and $\mathcal{C}^{\mathcal{E}}$ represent classifiers for thing classes, background, "void" class and exemplar-based classes, respectively. $\mathcal{P}^{void}$ means "void" class proposals. Positive supervision encourages model to produce a higher prediction score while negative supervision prefers a lower score.

2 Related Work

Panoptic Segmentation Pursuing a wholistic scene parsing, panoptic segmentation task (PS) is proposed to expect the generation of both semantic and instance segmentation simultaneously. Given fully annotations to the training images, different kinds of modeling targets have been explored for the PS problem. Specifically, unified end-to-end networks [17, 18, 30] are soon proposed after the initial release of baseline method with separate networks. DeepLab series methods [8, 28, 29] are deployed for fast inference speed. More recently, universal image segmentation [9, 10, 19, 20, 32] is pursued. While OPS shares a distinct target which demands the model to produce segments for *unknown* classes that are never acknowledged during training.

Open-Set Learning Open-set problem has been well explored in the recognition/classification task [3, 11, 14, 21, 25, 26, 27, 31]. The target of open-set recognition is to make the model successfully identify known classes and have the ability to identify unknown classes which are never exposed during training. OPS can be more challenging because *unknown* classes are not provided intact, but needs to be detected by the model itself. Other related works include open-world detection [13, 16]/segmentation [23] and class-agnostic detection [22]/ segmentation [5]. According to the problem definition, open-world object detection [16] contains a human labeling process after the *unknown* class detection while OPS does not require. And the open-set recognition procedure proposed in OW-DETR [13] significantly differs from our approach, e.g., the generation and usage of pseudo unknown objects varies greatly and the unknown class decision process is also different. While both open-world entity segmentation [23] and class-agnostic approaches [5, 22] aim to segment/detect visual entities without considering classes which is precisely the problem that OPS needs to solve urgently. Meanwhile, anomaly detection based open-set approaches either fail to produce instance-level unknown objects for the OPS task [6] or require a proxy out-of-distribution dataset for open-set recognition [4, 7].

Model	PQ	SQ	RQ	Recall	Precision
Void-ignorance	3.7	71.8	5.2	11.0	3.4
Void-background	4.3	70.1	6.2	11.2	4.2
Void-suppression	7.2	75.3	9.5	27.6	5.8
Void-train	7.5	72.9	10.3	21.8	6.7
EOPSN	11.3	73.8	15.3	11.8	21.9

Table 1: Comparisons of OPS results of *unknown* class for existing OPS methods.[1] Since the recognition quality (RQ) varies a lot among these algorithms, recall and precision statistics of *unknown* class are also reported for detailed inspection. All empirical numbers are obtained on COCO *val* set with 20% of *thing* classes are set as *unknown* class during model training.

3 Open-Set Panoptic Segmentation

According to the problem definition given in [15], *open-set panoptic segmentation* (OPS) has a similar definition to the standard closed-set panoptic segmentation except for the label space and targets of the task. Apart from the *known* label space (i.e., countable objects in *thing* class $\mathcal{C}^{\mathrm{Th}}$ and amorphous and uncountable regions in *stuff* class $\mathcal{C}^{\mathrm{St}}$) which has annotations at training phase and requires to be effectively segmented during testing, OPS also requires the model to be able to detect and generate instance masks for *unknown thing* class $\mathcal{C}_u^{\mathrm{Th}}$ in test set[2]. The *unknown thing* $\mathcal{C}_u^{\mathrm{Th}}$ are not annotated or even not appeared in the training images. For pixel areas in the ground truth of training images that are not manually annotated, a semantic label named *void* will be assigned to them.

Existing OPS methods [15] are built upon the classic Panoptic FPN network [17] and this is due to that region proposal network [24] (RPN), an important part of the network, can generate class-agnostic proposals and enable the possible of finding various classes of objects in any image [12] and makes the OPS problem tractable.

Figure 2 (a) presents some proposal examples generated from RPN module where solid boxes in orange and blue denote the proposals are labeled as a specific *known thing* class. Dashed boxes in black and orange denote the "void" class proposals[3] and other black solid boxes are background proposals. Since the proposal labeling of *known* classes is based on the *known* class GT, the quality of selected proposals are guaranteed. However, the quality of proposals $\mathcal{P}^{void}$ varies greatly as the connected "void" area is not manually annotated and may contain multiple objects or just ambiguous pixels, therefore some of them should be labeled as background in the closed-set PS setting. Examples in Figure 2(a) show that few yellow dashed boxes are well aligned with an *unknown* instance in the "void" area, while a large number of black dashed boxes are not well aligned with a specific *unknown* instance which should have been labeled as background proposals in the closed-set setting but it is impossible for the open-set case.

The existing OPS methods differ in how to use "void" class proposals and top row of Figure 2(b) presents their usage ways: Void-ignorance baseline does not include "void" class proposals $\mathcal{P}_{void}$ into network training; Void-background takes $\mathcal{P}_{void}$ as background; Void-suppression alternatively utilizes $\mathcal{P}_{void}$ to do a suppression on *known* class classifiers[4]; Void-train treats all $\mathcal{P}_{void}$ as the same and adds an *void* class classifier during training; EOPSN

[1] Results of *known* class are comparable among these methods and can be found in Table 2

[2] Segmentation of *unknown stuff* class is not required in the current OPS definition [15].

[3] Proposals who have a half of the region is inside the "void" area.

[4] We empirically find that suppress background as well will deteriorate the recognition of *known* classes.

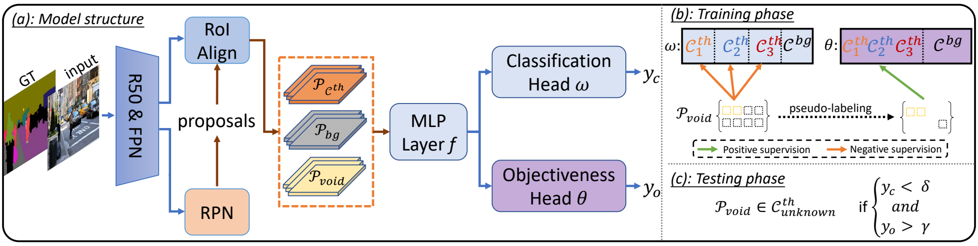


Figure 3: Demonstration of our method. We introduce an objectiveness head besides the default classification head for the prediction of objectiveness score of proposals. $\mathcal{P}_{\mathcal{C}^{th}}$, $\mathcal{P}_{bg}$ and $\mathcal{P}_{void}$ represent proposals in *known thing* classes, background and *void* class, respectively.

method can be seen as an enhanced version of Void-train and builds multiple representative exemplars from $\mathcal{P}_{void}$ through k-means clustering. During testing, proposals will be predicted as *unknown* class only when they are rejected by *known* classes with a pre-defined confidence threshold. Void-train and EOPSN further require the proposals to be predicted as "void" class or exemplar-based classes. Bottom row of Figure 2(b) gives a visualization of *unknown* class decision field for these methods and their *unknown* class recognition quality are presented in Table 1. We can find that neither Void-ignorance nor Void-background can produce a reasonable *unknown* class recognition result. Although both of Void-suppression and Void-train share a similar performance, i.e., relatively high recall and low precision, they may have different reasons. Void-suppression is due to the lack of ability for distinguish *unknown* class from background, while Void-train is because the supervision of $\mathcal{P}_{void}$ will make it overfit to the training set. EOPSN greatly improves the precision but heavily affects the recognition recall which means the exemplars obtained from proposals $\mathcal{P}_{void}$ are not representative enough.

4 Our Approach

In this section, we first present the necessary of constructing a two-stage decision structure for the OPS task. Then, we further propose a pseudo-labeling method to enhance the generalization ability of *unknown* class recognition.

4.1 Dual Decision Structure for the OPS Task

Based on the analysis in Sec. 3, we believe that *unknown* class cannot be well modeled at training phase without being aware of what kinds of *unknown* class will appear during testing. Therefore, both of Void-train and EOPSN may not be a promising direction for solving the OPS problem and the empirical results on *unseen* class [5] in Table 3 confirm our conclusion. However, other OPS methods can only rely on the *known* class classifier when making decisions on *unknown* classes and the empirical results show that such a decision procedure can not enable them to achieve a satisfactory recognition performance for unknown classes. Thus we build up a dual decision process for the effective recognition of *unknown* classes.

Following existing OPS methods, our structure is also adapted from the Panoptic FPN framework [17] and the core structure is presented in Figure 3(a). Specifically, for a given

[5] *unseen* means the corresponding thing never appears in the training images. Sec. 5.1 gives a detail definition.

image, we first use the ResNet50 network and feature pyramid network to extract multi-scale feature representations. Then a region proposal network is used to generate class-agnostic proposals and their features can be obtained through the RoI align module. Given the ground truth segmentation annotations at training stage, these proposals can be assigned labels according to their positional relationship with the annotations. For example, the proposals will be labeled as one *known thing* class $\mathcal{C}_i^{th} \in \mathcal{C}^{th}$ when it has a large overlap to any *known thing* class instance. Similarly, the "void" areas are also utilized for defining "void" class proposals $\mathcal{P}_{void}$. Other proposals are labeled as background class samples $\mathcal{P}_{bg}$.

In order to identify *known* classes, the classification head is supervised by the proposals

$$\min -\frac{1}{N_{\mathcal{P}_{\overline{void}}}} \sum_{i \in \{\mathcal{C}^{th}, bg\}} \sum_{k=1}^{N_{\mathcal{P}_i}} \log \frac{\exp\left(w_i^T f(\mathcal{P}_i^k)\right)}{\sum_{j \in \{\mathcal{C}^{\mathrm{Th}}, bg\}} \exp\left(w_j^T f(\mathcal{P}_i^k)\right)} \tag{1}$$

where w is the weight of classification head. $N_{\mathcal{P}_{\overline{void}}}$ is the number of proposals except for those belonging to "void" class. $N_{\mathcal{P}_i}$ is the number of proposals in any specific *thing* class. In order to separate *known* and *unknown* class effectively, we follow the Void-suppression baseline to do a suppression on *known* class classifiers with "void" class proposals

$$\min -\frac{1}{N_{\mathcal{P}_{void}}} \sum_{i=1}^{N_{\mathcal{P}_{void}}} \sum_{k \in \mathcal{C}^{\mathrm{th}}} \log\left(1 - \frac{\exp\left(w_k^T f(\mathcal{P}_{void}^i)\right)}{\sum_{\{\mathcal{C}^{\mathrm{th}}, bg\}} \exp\left(w_j^T f(\mathcal{P}_{void}^i)\right)}\right). \tag{2}$$

where $N_{\mathcal{P}_{void}}$ means the number of "void" class proposals.

Considering the modeling of Eqs. 2 and 1 can only improve the discriminative ability of *known* classes, *unknown* classes may still mix with background ones. In order to mitigate this drawback, we introduce a class-agnostic object prediction head (a.k.a. objectiveness head) parallel to the *known* class classification head and optimize it as follows

$$\min \frac{-1}{N_{\mathcal{P}_{\overline{void}}}} \left(\sum_{i \in \mathcal{C}^{th}} \sum_{k=1}^{N_{\mathcal{P}_i}} \log \frac{\exp\left(\theta^T f(\mathcal{P}_i^k)\right)}{1 + \exp\left(\theta^T f(\mathcal{P}_i^k)\right)} + \sum_{l=1}^{N_{\mathcal{P}_{bg}}} \log \frac{1}{1 + \exp\left(\theta^T f(\mathcal{P}_{bg}^l)\right)} \right) \tag{3}$$

where θ is the weight of objectiveness head and $N_{\mathcal{P}_{bg}}$ is the number of background proposals.

At the testing stage, the recognition of *unknown* class will be made in a dual decision process based on the predictions on both *known* class classification head and class-agnostic object prediction head, i.e., only proposals who are rejected by the *known* class classification head and accepted by the objectiveness head simultaneously will be predicted as *unknown* class. Empirical results in Table 2 shows that such kind of dual decision process significantly boosts the *unknown* class recognition performance on all kinds of OPS settings.

Rationale of design: The key feature of the above design is that we will treat all known class proposals as training samples for *a single class-agnostic* "object" class. In contrast, the methods described in Figure 2 will treat each class separately. The class-agnostic classification head will encourage the network identify patterns that are shared across class rather than focusing on (known-)class specific patterns. The former can generalize well to unseen thing while the latter may overfit to things only seen at the training stage.

4.2 Improve Object Recognition Generalization with Pseudo-labeling

Currently, the newly added class-agnostic object prediction head is only optimized on proposals belonging to *known thing* class or background ones and the "void" class proposals [6]

[6]We take any connected "void" area in ground truth of training images as "void" class proposals.

K	Model	Known									Unknown				
		PQ	SQ	RQ	PQ^{Th}	SQ^{Th}	RQ^{Th}	PQ^{St}	SQ^{St}	RQ^{St}	PQ	SQ	RQ	R	P
	Supervised	39.4	77.7	48.4	45.8	80.7	55.4	29.7	73.1	38.0	-	-	-	-	-
5	Void-supp.	38.0	77.0	46.7	44.8	80.6	54.1	28.3	71.7	36.1	6.7	76.2	8.8	39.9	4.9
	Void-train	37.3	76.7	45.9	43.6	80.4	52.8	28.2	71.5	36.0	8.6	72.7	11.8	29.8	7.3
	EOPSN	38.0	76.9	46.8	44.8	80.5	54.2	28.3	71.9	36.2	23.1	74.7	30.9	25.9	38.3
	Ours	38.1	77.7	46.6	45.1	80.9	54.3	28.1	73.1	35.7	**30.2**	**80.0**	**37.8**	32.8	44.5
10	Void-supp.	37.6	76.8	46.3	44.3	80.5	53.5	28.5	71.7	36.4	6.5	76.0	8.6	32.7	5.0
	Void-train	37.1	77.1	45.8	43.7	80.1	53.1	28.1	73.0	35.9	8.1	72.6	11.2	25.7	7.2
	EOPSN	37.7	76.8	46.3	44.5	80.6	53.8	28.4	71.8	36.2	17.9	76.8	23.3	19.0	30.2
	Ours	37.7	77.1	46.3	45.0	80.7	54.3	27.8	72.2	35.4	**24.5**	**79.9**	**30.7**	24.7	40.6
20	Void-supp.	37.5	75.9	46.1	45.1	80.6	54.5	28.2	70.2	36.1	7.2	75.3	9.5	27.6	5.8
	Void-train	36.8	76.3	45.4	44.1	80.1	53.5	27.9	71.6	35.6	7.5	72.9	10.3	21.8	6.7
	EOPSN	37.4	76.2	46.2	45.0	80.3	54.5	28.2	71.2	36.2	11.3	73.8	15.3	11.8	21.9
	Ours	37.1	75.8	45.7	45.0	80.6	54.3	27.6	70.1	35.3	**21.4**	**79.1**	**27.1**	21.9	35.4

Table 2: Comparisons of open-set panoptic segmentation performance against the state-of-the-art methods on MS-COCO *val* set with three *known-unknown* splits $K(\%)$ which denotes the ratio of *unknown* classes to all classes. Recall (R) and precision (P) of *unknown* objects are also presented for reference. The best results on *unknown* classes are bold highlighted.

Model	Known									Unknown			Unseen		
	PQ	SQ	RQ	PQ^{Th}	SQ^{Th}	RQ^{Th}	PQ^{St}	SQ^{St}	RQ^{St}	PQ	SQ	RQ	PQ	SQ	RQ
Void-supp.	35.8	76.7	44.5	43.0	81.2	52.5	27.7	71.6	35.4	7.6	75.5	10.1	4.5	75.9	6.0
Void-train	35.4	77.2	43.9	42.2	81.0	51.6	27.7	72.8	35.3	8.8	73.8	15.7	4.4	74.8	5.9
EOPSN	35.7	76.6	44.7	43.2	81.1	52.7	27.8	71.4	35.6	23.0	74.6	30.8	0.4	80.3	0.5
Ours	35.8	76.6	44.5	43.0	81.1	52.5	27.6	71.4	35.3	**30.2**	**80.2**	**37.7**	**9.3**	**82.5**	**11.2**

Table 3: Comparisons of OPS performance on MS-COCO *val* set under the newly proposed *zero-shot* setting. The best results on *unknown* and *unseen* classes are bold highlighted.

are not fully utilized. Since the proposals of "void" class may contain many novel objects which does not belong to the annotated *known thing* classes, we assume the properly exploiting of "void" class proposals can be helpful for the recognition generalization of objectiveness head. One straightforward way is to directly take all the "void" class proposals as potential *unknown* ones to supervise the objectiveness head. However, results in Figure 6 shows that this strategy will heavily deteriorate the recognition quality. It may because the proposals of *void* class are not precise and contain much noise which is not suitable for the immediate exploiting. Therefore, we propose to use the pseudo-labeling technique to filter out invalid "void" class proposals.

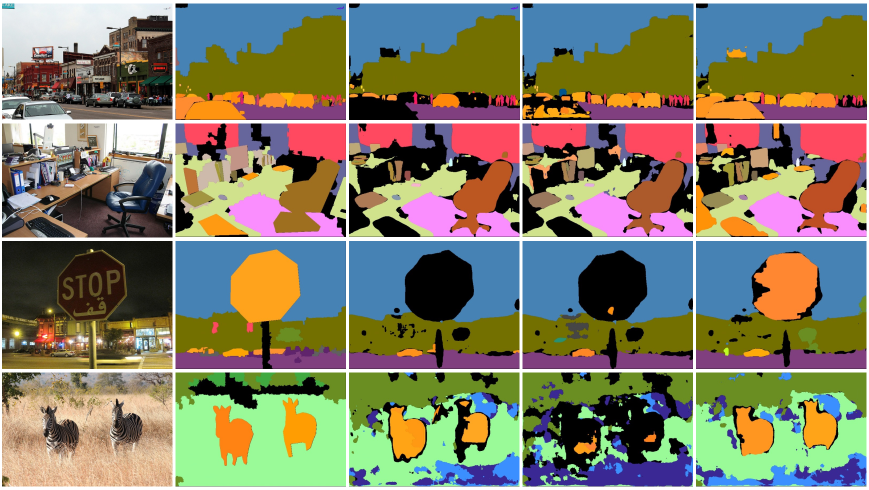
Since the newly added objectiveness head is designed in a class-agnostic fashion, the quality of "void" class proposals can be predicted by the up-to-date objectiveness head and we can select those high confident ones to further supervise the objectiveness head

$$\min \frac{-1}{N_{\mathcal{P}_{void}}} \sum_{i=1}^{N_{\mathcal{P}_{void}}} \mathbb{1}\left(\frac{\exp\left(\theta^T f(\mathcal{P}^i_{void})\right)}{1+\exp\left(\theta^T f(\mathcal{P}^i_{void})\right)} \geq \delta \right) \log \frac{\exp\left(\theta^T f(\mathcal{P}^i_{void})\right)}{1+\exp\left(\theta^T f(\mathcal{P}^i_{void})\right)} \tag{4}$$

where δ is the confidence threshold.

5 Experimental Results

In this section, we conduct experiments to evaluate the proposed approach and existing OPS methods on open-set panoptic segmentation task.



(a) Input (b) GT (c) Void-supp. (d) EOPSN (e) Ours

Figure 4: Visual results on COCO *val* set with K=20%. Compared to Void-suppression and EOPSN, our algorithm can detect more novel objects and generates better instance masks. The most prominent *unknown* class objects in row 1-4 are car, keyboard, stop sign and zebra, respectively. (a) Input image (b) Ground-truth (c), (d) and (e) are panoptic segmentation results of Void-suppression, EOPSN and our method, respectively. Predicted instances in the *unknown* class are denoted by orange color and the black areas represent the areas that are fail annotated (i.e., (b)) or segmented (i.e., (c)-(e)).

5.1 Experimental Details

To make a fair comparison, we directly build up our experiments based on the released codebase[7]. Some experimental details are as follows

Datasets: Following the protocol of [15], all experiments are conducted on MS-COCO 2017 dataset whose default annotations are constructed by 80 *thing* classes and 53 *stuff* classes. [15] manually removes a subset of *known thing* classes (i.e., K% of 80 classes) in the training dataset and takes them as *unknown* classes for evaluating on open-set task (*stuff* classes are all kept). Three *known-unknown* splits of K are considered:**5%**, **10%** and **20%**.

In order to evaluate the object recognition generalization ability of OPS methods, we further construct a more realistic OPS setting named ***zero-shot*** which is built up from the 5% split setting mentioned above and further removes training images that contains instances belonging to the 20% tail *thing* class of MS-COCO. These classes are {hair drier, toaster, parking meter, bear, scissors, microwave, fire hydrant, toothbrush, stop sign, mouse, refrigerator, snowboard, frisbee, keyboard, hot dog, baseball bat}. To distinguish from *unknown* classes, we call these classes ***unseen* classes**.

Methods: Two strong baselines and the state-of-the-art OPS method are included for comparison, i.e., Void-train, Void-suppression and EOPSN. Meanwhile, Panoptic FPN trained on full 80 *thing* classes are also reported for a reference baseline (denoted as supervised).

Evaluation Metric: The standard panoptic segmentation metrics (i.e., PQ, SQ, RQ) are reported for *known*, *unknown* and *unseen* classes (see detail formulations in the appendix).

[7] https://github.com/jd730/EOPSN.git

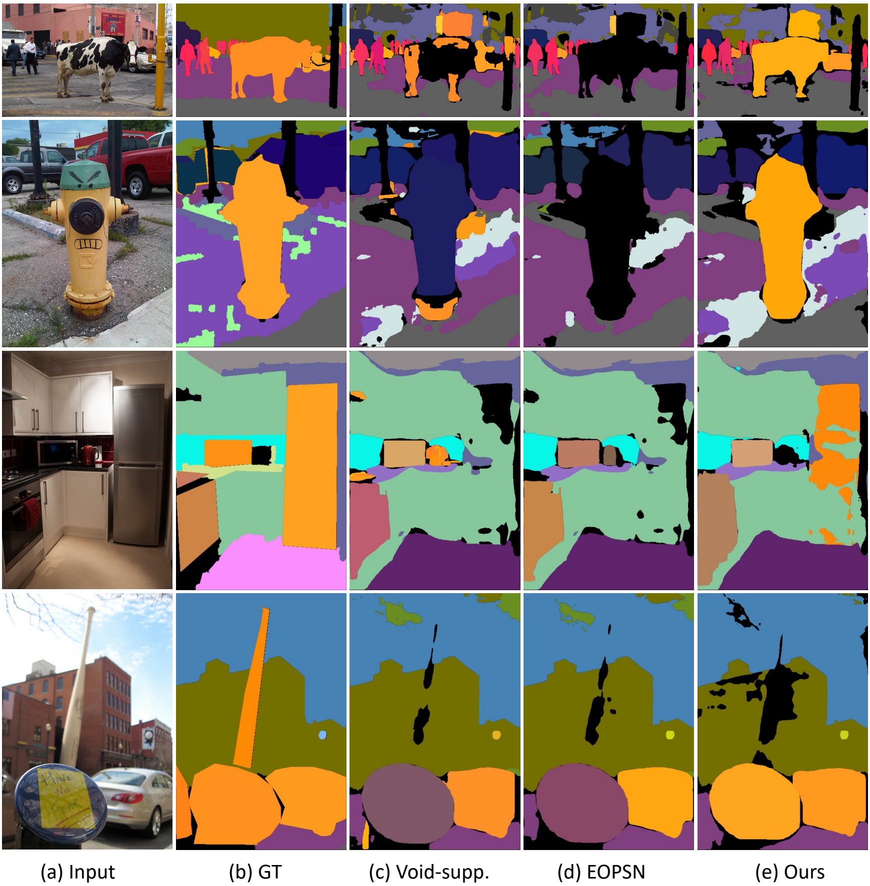


Figure 5: Visual results on COCO *val* set with *unknown* and *unseen* class. The first row shows that our method can generate better instance mask for *unknown* class, e.g., cow. Row 2-4 present that the proposed approach can successfully detect some *unseen* classes (tail classes on COCO), e.g., fire hydrant, refrigerator and parking meter.

5.2 Results on *known-unknown* Setting

Table 2 shows the quantitative results of comparing methods. It is clear that our proposed method significantly improves the panoptic quality of *unknown* class objects than the Void-suppression baseline across all kinds of splits. Meanwhile, compared with the SOTA method EOPSN, our approach excels on both of the recall and precision of *unknown* objects recognition and therefore achieves much better PQ values. Figure 4 illustrates the qualitative results. We find that our approach can successively detect more *unknown* class objects and generate more precise instance masks than both of Void-suppression baseline and EOPSN method.

5.3 Results on *zero-shot* Setting

Our approach has been verified to be effective on *known-unknown* setting in Sec. 5.2, we also want to know its novel object recognition ability in a *zero-shot* setting. Table 3 presents that the proposed methods are superior than the comparing ones on both *unknown* class and

Setting	Obj.	PL	Unknown PQ	Unknown SQ	Unknown RQ	Unseen PQ	Unseen SQ	Unseen RQ
K=20%	✗	✗	7.2	75.3	9.5	-	-	-
	✓	✗	19.5	**79.5**	24.5	-	-	-
	✓	✓	**21.4**	79.1	**27.1**	-	-	-
zero-shot	✗	✗	7.6	75.5	10.1	4.5	75.9	6.0
	✓	✗	29.6	**80.5**	36.8	6.9	81.3	8.5
	✓	✓	**30.2**	80.2	**37.7**	**9.3**	**82.5**	**11.2**

Table 4: Ablation study to the effectiveness of each component in our method.

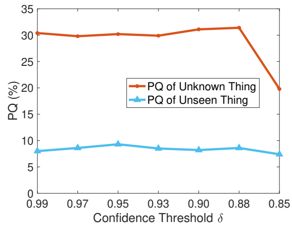


Figure 6: Ablation study of confidence threshold of pseudo labeling on *zero-shot* setting.

unseen class objects. It is interesting that EOPSN performs well on *unknown* class but almost fails on *unseen* class. This may be due to the fact that the exemplars obtained in EOPSN are completely derived from the training set and cannot be generalized to unseen class objects. Qualitative results for the *zero-shot* setting shown in Figure 5 present that our approach can always detect salient objects in the image and produce overall best instance masks.

5.4 Ablation Study

We are interested in ablating our approach from the following perspective views:

Effective of each component in our method: Our approach is mainly composed of two components (i.e., objectiveness head and pseudo labeling) and Table 4 shows the performance contribution of each component on two kinds of settings. It is obvious that simply adding the objectiveness head significantly improves the unknown segmentation performance and incorporating the pseudo labeling trick further boost the overall performance.

Sensitivity analysis: Our method only has one hyper-parameter, i.e., the confidence threshold δ in pseudo labeling mechanism. As shown in Figure 6, the performance of our approach is stable when the confidence value falls into $\delta \in [0.88, 0.99]$.

6 Conclusion

Open-set panoptic segmentation (OPS) is a newly proposed research task which aims to perform segmentation for both *known* classes and *unknown* classes. In order to solve the challenges of OPS, we propose a dual decision mechanism for *unknown* class recognition. We implement this mechanism through coupling a *known* class classification head and a class-agnostic object prediction head and make them corporate together for final *unknown* class prediction. To further improve the recognition generalization ability of the objectiveness head, we use the pseudo-labeling technique to boost the performance of our approach. Extensive experimental results verify the effectiveness of the proposed approach on various kinds of OPS tasks.

Acknowledgements. We thank Dr. Yutong Dai for her kind proofreading and the anonymous reviewers for their insightful comments and suggestions on our manuscript. We also thank Huawei Noah's Ark Lab gratefully for sponsoring large-scale GPUs.

References

[1] Hillary Abraham, Chaiwoo Lee, Samantha Brady, Craig Fitzgerald, Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. Autonomous vehicles, trust, and driving alternatives: A survey of consumer preferences. *Massachusetts Inst. Technol, AgeLab, Cambridge*, 1(16):2018–12, 2016.

[2] Mihalj Bakator and Dragica Radosav. Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies and Interaction*, 2(3):47, 2018.

[3] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016.

[4] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Dense open-set recognition based on training with noisy negative images. *Image and Vision Computing*, page 104490, 2022.

[5] Vighnesh Birodkar, Zhichao Lu, Siyang Li, Vivek Rathod, and Jonathan Huang. The surprising impact of mask-head architecture on novel class segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7015–7025, 2021.

[6] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *International Journal of Computer Vision*, 129(11):3119–3135, 2021.

[7] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proceedings of the ieee/cvf international conference on computer vision*, pages 5128–5137, 2021.

[8] Bowen Cheng, Maxwell D Collins, Yukun Zhu, Ting Liu, Thomas S Huang, Hartwig Adam, and Liang-Chieh Chen. Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12475–12485, 2020.

[9] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. *arXiv preprint arXiv:2112.01527*, 2021.

[10] Bowen Cheng, Alex Schwing, and Alexander Kirillov. Per-pixel classification is not all you need for semantic segmentation. *Advances in Neural Information Processing Systems*, 34, 2021.

[11] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3614–3631, 2020.

[12] Xiuye Gu, Tsung-Yi Lin, Weicheng Kuo, and Yin Cui. Open-vocabulary object detection via vision and language knowledge distillation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=lL3lnMbR4WU.

[13] Akshita Gupta, Sanath Narayan, KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Ow-detr: Open-world detection transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9235–9244, 2022.

[14] Jiaming Han, Yuqiang Ren, Jian Ding, Xingjia Pan, Ke Yan, and Gui-Song Xia. Expanding low-density latent regions for open-set object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9591–9600, 2022.

[15] Jaedong Hwang, Seoung Wug Oh, Joon-Young Lee, and Bohyung Han. Exemplar-based open-set panoptic segmentation network. In *CVPR*, 2021.

[16] KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5830–5840, 2021.

[17] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollár. Panoptic feature pyramid networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6399–6408, 2019.

[18] Yanwei Li, Xinze Chen, Zheng Zhu, Lingxi Xie, Guan Huang, Dalong Du, and Xingang Wang. Attention-guided unified network for panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7026–7035, 2019.

[19] Yanwei Li, Hengshuang Zhao, Xiaojuan Qi, Liwei Wang, Zeming Li, Jian Sun, and Jiaya Jia. Fully convolutional networks for panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 214–223, 2021.

[20] Zhiqi Li, Wenhai Wang, Enze Xie, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, Tong Lu, and Ping Luo. Panoptic segformer. *arXiv preprint arXiv:2109.03814*, 2021.

[21] Poojan Oza and Vishal M Patel. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2307–2316, 2019.

[22] Lu Qi, Jason Kuen, Zhe Lin, Jiuxiang Gu, Fengyun Rao, Dian Li, Weidong Guo, Zhen Wen, Ming-Hsuan Yang, and Jiaya Jia. Ca-ssl: Class-agnostic semi-supervised learning for detection & segmentation. *arXiv preprint arXiv:2112.04966*, 2021.

[23] Lu Qi, Jason Kuen, Yi Wang, Jiuxiang Gu, Hengshuang Zhao, Zhe Lin, Philip Torr, and Jiaya Jia. Open-world entity segmentation. *arXiv preprint arXiv:2107.14228*, 2021.

[24] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[25] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012.

[26] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 36(11): 2317–2324, 2014.

[27] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. *arXiv preprint arXiv:2110.06207*, 2021.

[28] Huiyu Wang, Yukun Zhu, Bradley Green, Hartwig Adam, Alan Yuille, and Liang-Chieh Chen. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. In *European Conference on Computer Vision*, pages 108–126. Springer, 2020.

[29] Huiyu Wang, Yukun Zhu, Hartwig Adam, Alan Yuille, and Liang-Chieh Chen. Max-deeplab: End-to-end panoptic segmentation with mask transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5463–5474, 2021.

[30] Yuwen Xiong, Renjie Liao, Hengshuang Zhao, Rui Hu, Min Bai, Ersin Yumer, and Raquel Urtasun. Upsnet: A unified panoptic segmentation network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8818–8826, 2019.

[31] Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. Classification-reconstruction learning for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4016–4025, 2019.

[32] Wenwei Zhang, Jiangmiao Pang, Kai Chen, and Chen Change Loy. K-net: Towards unified image segmentation. *Advances in Neural Information Processing Systems*, 34, 2021.