Exemplar-Based
Open-Set Panoptic Segmentation Network




Jaedong Hwang
Seoul National Univ.

Seoung Wug Oh
Adobe Research

Joon-Young Lee
Adobe Research

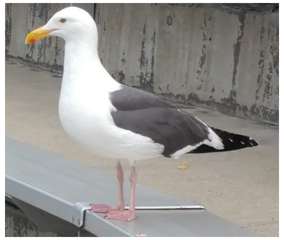
Bohyung Han
Seoul National Univ.

Open-Set Panoptic Segmentation

- **Closed-set:** In the test dataset, there only exists classes that are learned during training.

Training data

tortoise

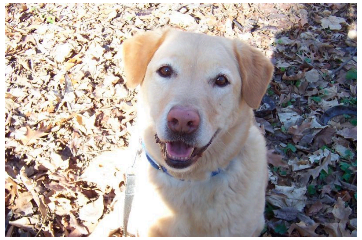


bird

horse

dog

Test data



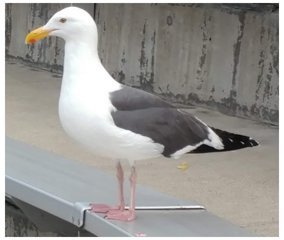
We know these classes!

Open-Set Panoptic Segmentation

- **Closed-set:** In the test dataset, there only exists classes that are learned during training.
- **Open-set:** In the test dataset, some classes do not appear during training.

Training data

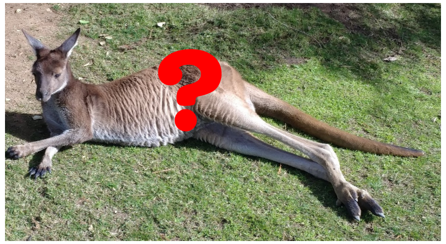

tortoise


bird


horse


dog

Test data



Who are they?

Open-Set Panoptic Segmentation

- Classes in the open-world can be divided.
  - **Known class:** class appears during training with labels and testing.
  - **Unknown class:**
    - **Seen Unknown class (Known Unknown class):** class appears during training without label and testing.
    - **Unseen Unknown class (Unknown Unknown class):** class only appears during testing.

Class

- **Known Class**
  person, car, motorbike
- **Unknown Class**
  armadillo, shoes
  - **Seen Unknown Class**
    shoes
  - **Unseen Unknown Class**
    armadillo

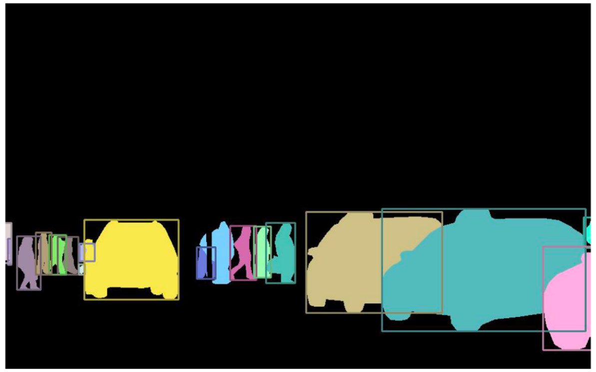
- As we gather more images, the number of seen unknown classes rapidly increases.

Open-Set **Panoptic Segmentation**

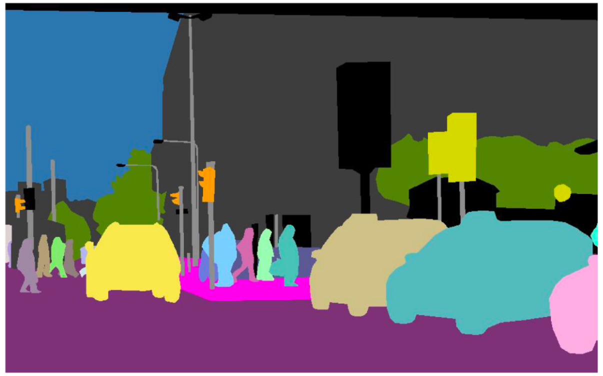
- **Semantic Segmentation**
  - Segment all semantic background and instances.
- **Instance Segmentation**
  - Segment all instance differentiating each instance (instance id).
- **Panoptic Segmentation**
  - Segment all semantic background and instances differentiating each instance.

Semantic Segmentation

Instance Segmentation



Panoptic Segmentation



Open-Set **Panoptic Segmentation**

- **Thing class**
  - Object (e.g., *car*, *person*, *motorbike*).
- **Stuff class**
  - Semantic background (e.g., *sky*, *road*).
  - Amortized instances (e.g., *building*, *grass*, *tree*).
- **Void class**
  - Remaining regions where are not associated with thing class and stuff class.

- Evaluation Metrics
  - PQ (Panoptic Quality), SQ (Segmentation Quality), RQ (Recognition Quality)

$$\text{PQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \text{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \cdot \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}$$

Open-Set Panoptic Segmentation

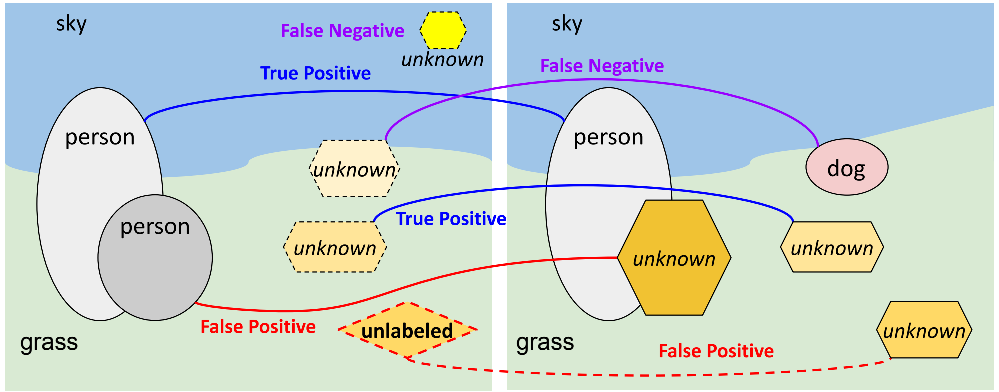
- **Open-Set + Panoptic Segmentation**
- The model should find and segment not only known class but also unknown class.

- **Class**
  - **Thing Class**
  - **Stuff Class**
  - **Void Class**

- **Class**
  - **Known Class**
  - **Unknown Class**
    - **Seen Unknown Class**
    - **Unseen Unknown Class**

Challenges

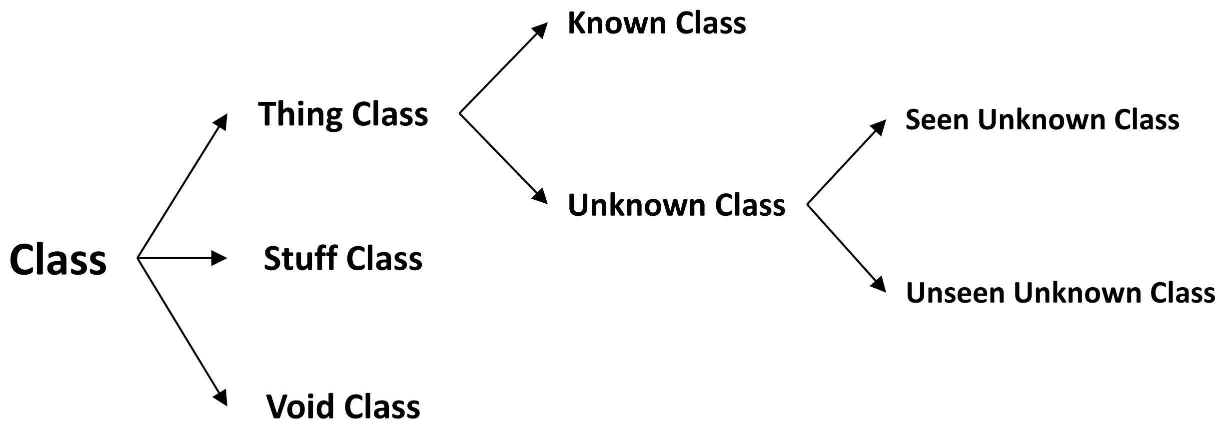
- Finding unknown instances are difficult than rejecting unknown instances
  - Unknown objects which might have been labeled as background during training.
- Labeling all objects for evaluation is almost impossible.
  - Some objects are composed of several parts.
  - The definition of 'object' varies by each person and it is hard to give one concrete guidance.
- Evaluation
  - Since ground-truths are incomplete PQ and RQ decrease as a model finds unlabeled unknown objects.
  - SQ cannot represent overall panoptic segmentation quality since it only consider true positives.



Ground-truth **Prediction**

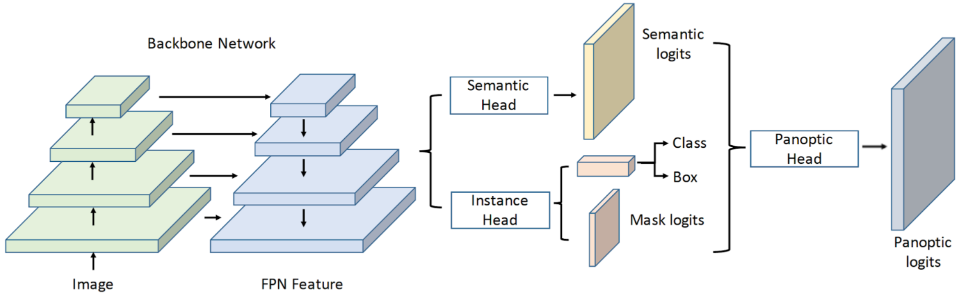
Assumptions for Tractable OPS

- All unknown classes are thing class.
  - stuff classes are defined as regions (e.g., water) or often ill-posed (e.g., *tree*).
- Parts of known classes cannot be unknown classes (e.g., head in *person*, tire in *car*)
  - If *tire* exists by itself, it can be an unknown object.
- Unknown class objects only appear void regions in the training data.
  - This is for preventing label conflict between known class and unknown class.

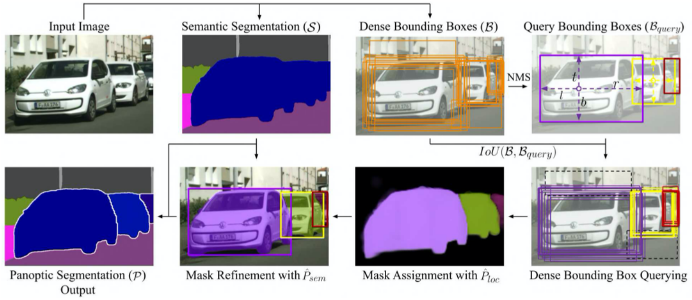


Previous Closed-Set Panoptic Segmentation Frameworks

- Generate RoIs and merge Semantic Segmentation and Instance Segmentation.

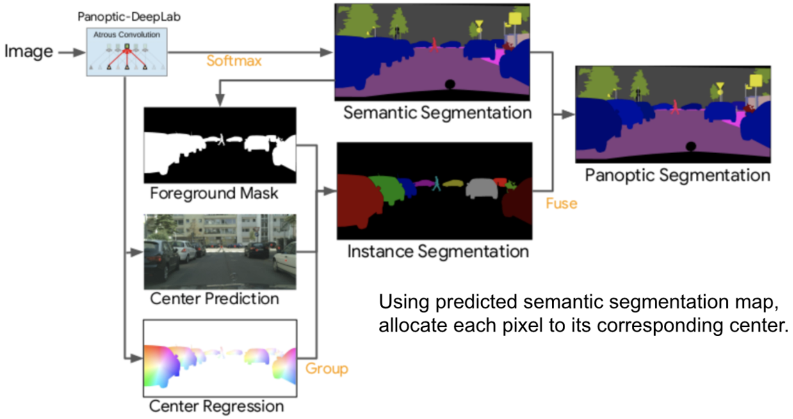


Xiong, Yuwen, et al. "UpsNet." *CVPR*. 2019.

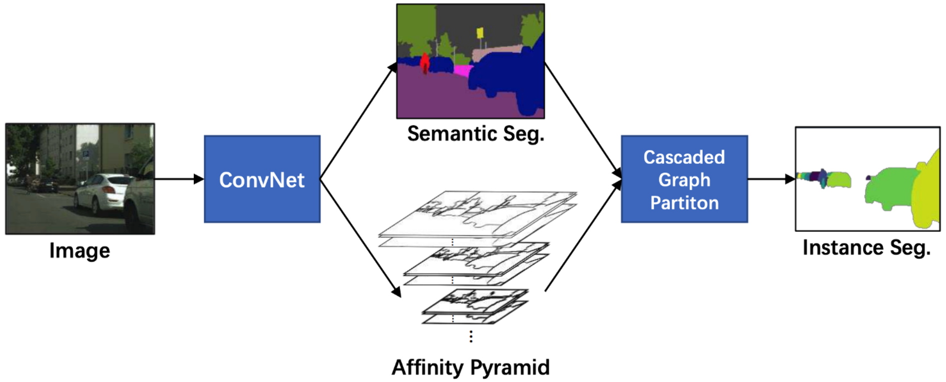


Hou, Rui, et al. "Real-Time Panoptic Segmentation from Dense Detections." *CVPR*. 2020.

- Allocate Instance ID after generating Semantic Segmentation.



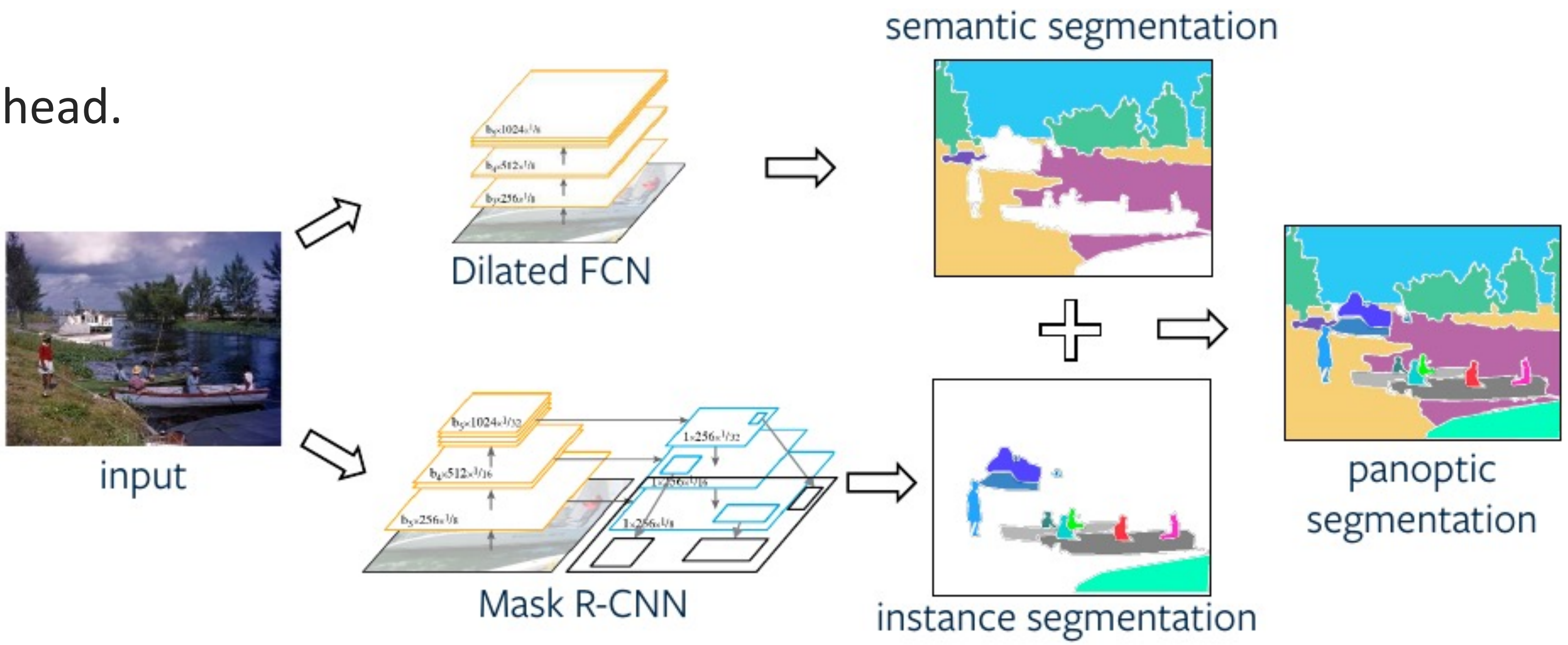
Cheng, Bowen, et al. "Panoptic-Deeplab." *CVPR*. 2020.



Gao, Naiyu, et al. ”SSAP." *ICCV*. 2019.

Motivation

- There is no methods that directly tackle open-set panoptic segmentation.
- However, we can utilize class-agnostic model (e.g., RPN)
- We use **Panoptic FPN** as our baseline employing
  - Class-agnostic regressor.
  - Class-agnostic mask prediction head.
  - Objectness score from RPN.



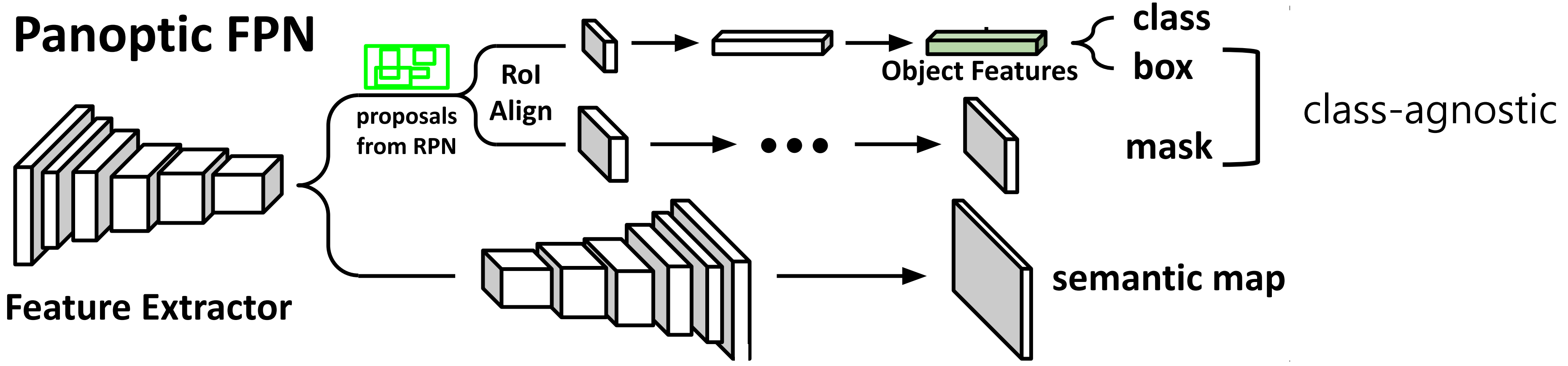
Kirillov, Alexander, et al. "Panoptic feature pyramid networks." *CVPR*. 2019.

Baseline (Panoptic FPN)

- We define **void box** of which more than a half of area is on the void regions.
- We have 4 different ways to use **void boxes**.
  - Train as a background (Void-background)
  - Ignore during training (Void-ignorance)
  - Give supervision not assigning as known class (Void-suppression)

$$\mathcal{L}_{\text{void}} = \sum_{c \in \mathcal{C}^{\text{Th}}} -\log(1 - p_c)$$

  - Train as a new class (Void-train)



Motivation

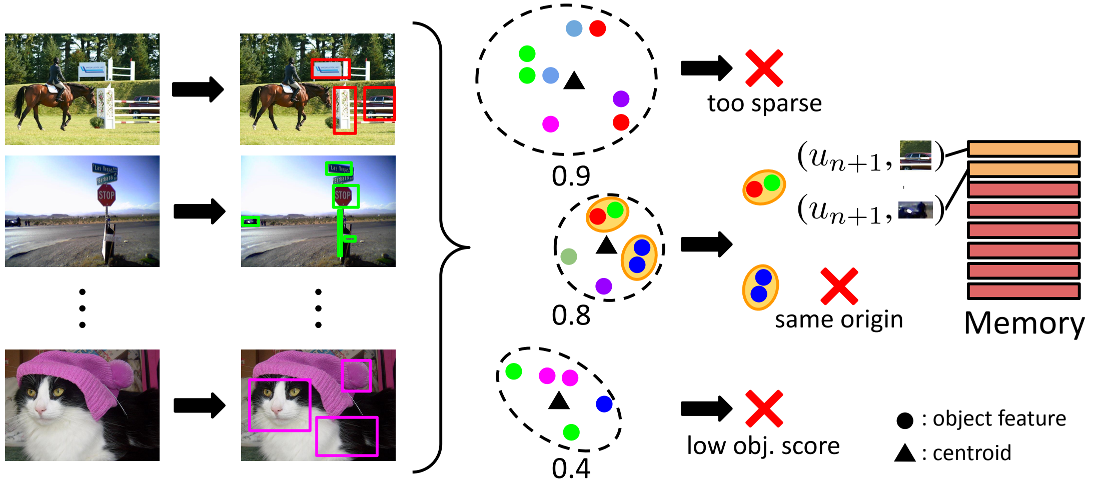
- Class-agnostic approach assumes unknown class has similar semantics with known ones.
  - It is hard to find unknown class having different semantics.

Known classes

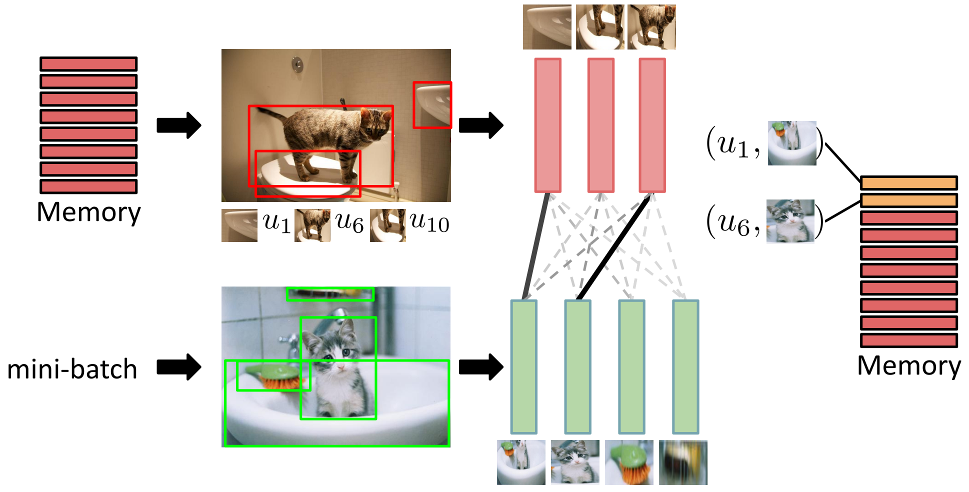
Unknown classes

Overview

- We propose exemplar-based open-set panoptic segmentation network (**EOPSN**).
- In exemplar theory, people categorize new objects by comparing exemplars of each class.
- We first find new unknown class with exemplars by clustering
- Then, we mine additional exemplars by comparing them with object proposals.

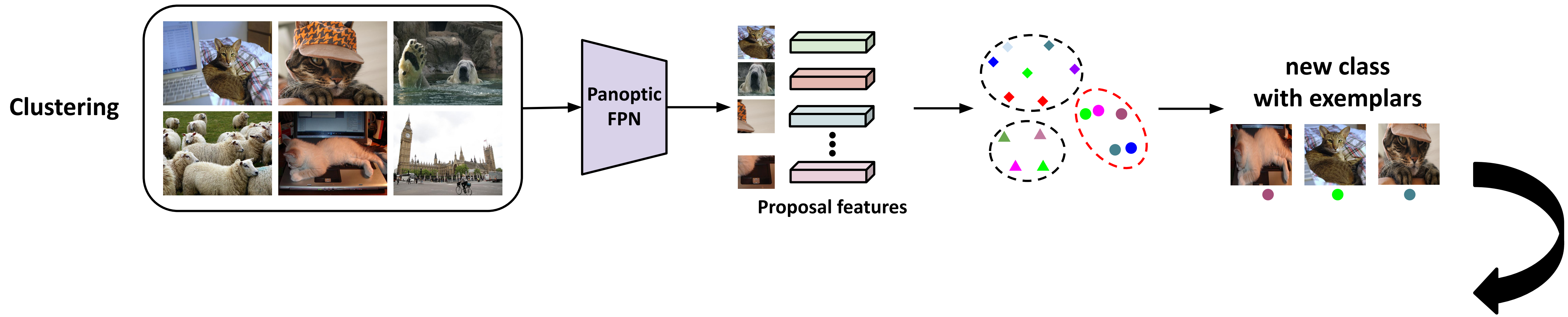


Clustering



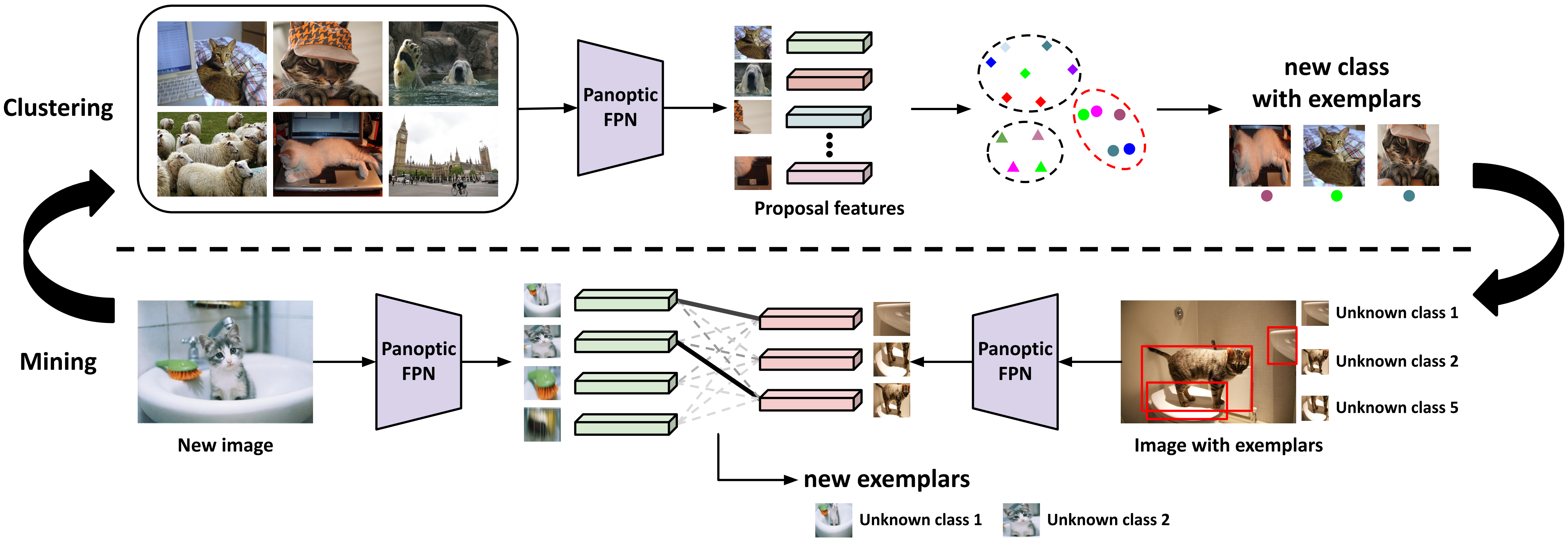
Mining

Overall Architecture



- Proposal features are the inputs of last classification layer.

Overall Architecture



- Proposal features are the inputs of last classification layer.

Dataset

- Our open-set panoptic segmentation (OPS) dataset is based on Microsoft MS COCO 2017.
  - 118K training image, 5K validation images (80 thing classes, 53 stuff classes).
- We construct different splits with different ratio, *K* of unknown classes among thing classes.
  - **5%:** *car, cow, pizza, toilet*
  - **10%:** (5%) + *boat, tie, zebra, stop sign*
  - **20%:** (10%) + *dining table, banana, bicycle, cake, sink, cat, keyboard, bear*
- We remove annotations for unknown classes in the training dataset.

Image

Original annotation

OPS annotation

Performance of Baseline Methods

- Open-set panoptic segmentation results on COCO val set with K = 20% of the baselines.

Utilization of *void* regions	Known									Unknown		
	PQ	SQ	RQ	PQ^{Th}	SQ^{Th}	RQ^{Th}	PQ^{St}	SQ^{St}	RQ^{St}	PQ	SQ	RQ
Void-background	**37.7**	76.3	**46.6**	44.8	79.3	54.1	**29.2**	72.8	**37.5**	4.0	71.1	5.7
Void-ignorance	37.2	76.3	45.9	43.9	79.0	53.1	29.1	**73.0**	37.3	3.7	71.8	5.2
Void-suppression	37.5	75.9	46.1	**45.1**	**80.6**	**54.5**	28.2	70.2	36.1	7.2	**75.3**	9.6
Void-train	36.9	**76.4**	45.5	44.0	80.3	53.3	28.2	71.7	36.0	**7.8**	73.4	**10.7**

Comparison with Baseline Method

- Open-set panoptic segmentation results on COCO val set with various splits, *K*.

$K(\%)$	Model	Known									Unknown		
		PQ	SQ	RQ	PQ^{Th}	SQ^{Th}	RQ^{Th}	PQ^{St}	SQ^{St}	RQ^{St}	PQ	SQ	RQ
	Supervised	39.4	77.7	48.4	45.8	80.7	55.4	29.7	73.1	38.0	-	-	-
5	Baseline (*Void-train*)	37.7	76.7	46.4	44.2	80.4	53.5	28.3	71.3	36.2	10.0	73.8	13.5
	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**
10	Baseline (*Void-train*)	36.9	75.4	45.5	43.2	79.0	52.4	28.3	70.4	36.2	8.5	73.2	11.6
	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**
20	Baseline (*Void-train*)	36.9	**76.4**	45.5	44.0	80.3	53.3	28.2	**71.7**	36.0	7.8	73.4	10.7
	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**

Sensitivity Analysis

- Size of input proposals for clustering on COCO val set (*K* =10%)

Proposal size	Known			Unknown		
	PQ	SQ	RQ	PQ	SQ	RQ
Large	37.7	77.5	46.4	13.5	**78.1**	17.3
Medium	37.7	77.5	46.4	12.5	74.8	16.7
Small	37.6	76.7	46.3	0.3	64.1	0.4
Large + Medium	37.7	76.8	46.3	**17.9**	76.8	**23.3**
Large + Medium + Small	37.8	77.1	46.6	6.9	69.8	9.9

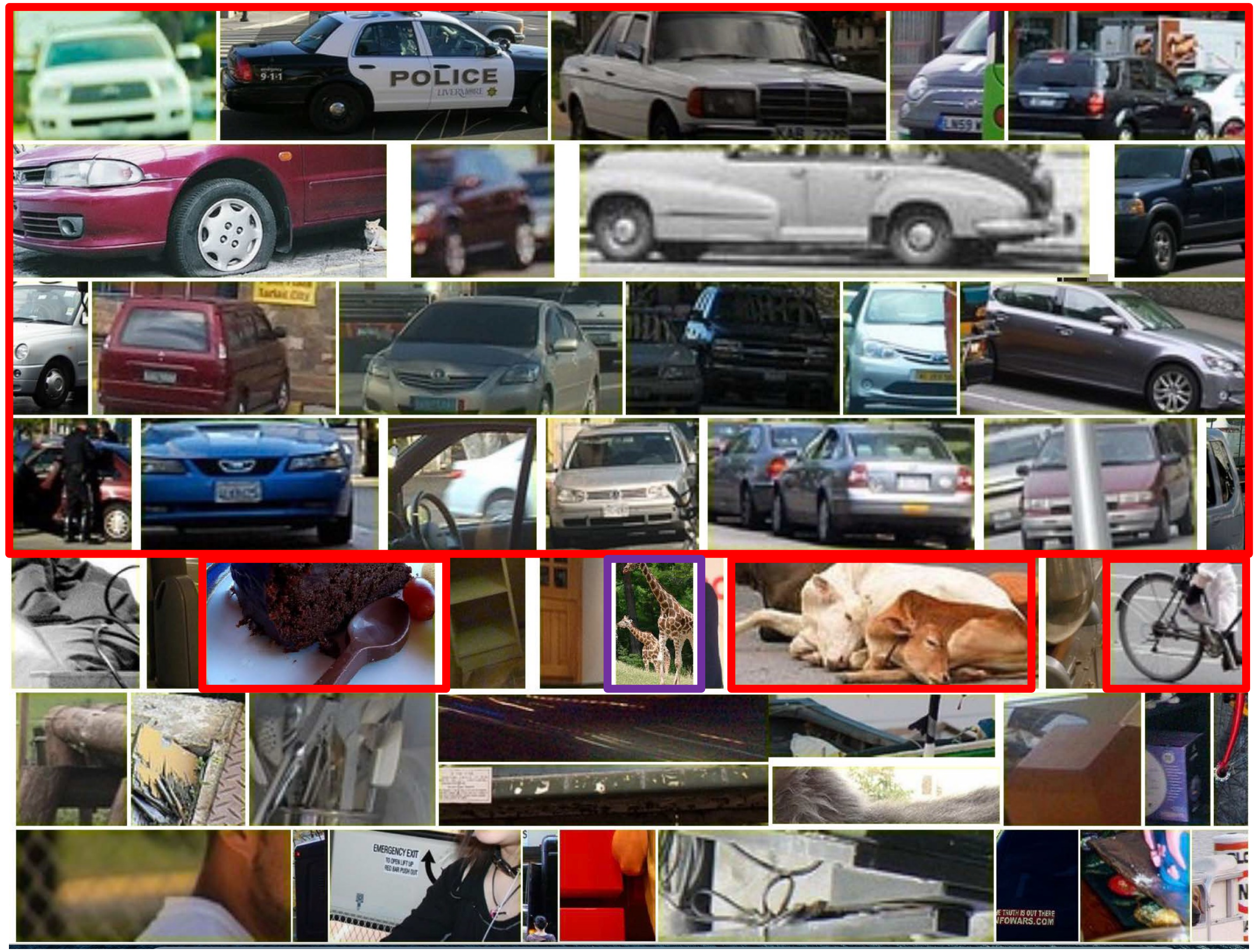
- The number of clusters and clustering interval on COCO val set (*K* =10%)

The number of clusters	Known			Unknown		
	PQ	SQ	RQ	PQ	SQ	RQ
64	37.3	76.2	45.8	12.9	76.5	16.8
128 (ours)	37.7	76.8	46.3	**17.9**	76.8	**23.3**
256	37.2	76.8	45.7	12.6	**78.8**	16.0

Clustering interval	Known			Unknown		
	PQ	SQ	RQ	PQ	SQ	RQ
100	37.6	77.2	46.3	8.2	**77.5**	10.6
200 (ours)	37.7	76.8	46.3	**17.9**	76.8	**23.3**
400	37.7	77.5	46.3	14.6	76.4	19.1

Exemplars in an Identified Unknown Class by First Clustering

- Most exemplars represent *car.*



Qualitative Results on COCO val set (*K*=10%)

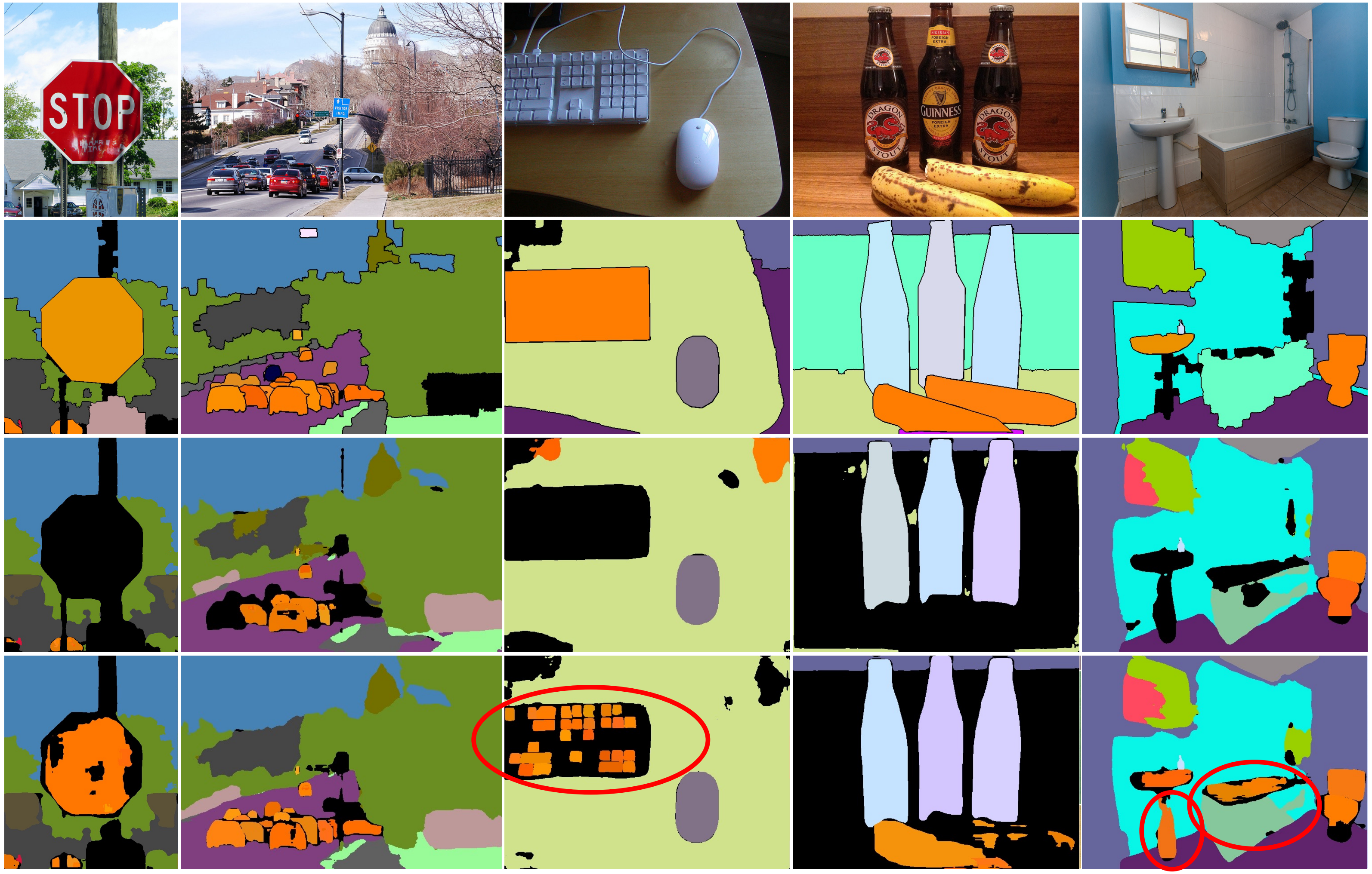
- **Orange** color presents unknown classes

Image

Ground-Truth

Baseline

EOPSN



Limitations

- EOPSN mainly focuses on recognition.
- EOPSN cannot deal with unseen unknown classes.
  - However, as we collect more data, the number of seen unknown classes will grow rapidly.
- We employ existing metrics that have limitations in OPS.
  - A new metric should be defined for OPS.
  - One possible way is to modify SQ by properly considering false positive of unknown classes.

Conclusion

- We propose a novel task, open-set panoptic segmentation.
- We also propose a novel framework for the task called EOPSN that
  - is based on top-down panoptic segmentation network, Panoptic FPN.
  - finds unknown classes with exemplars by clustering.
  - collects more exemplars by comparing found exemplars and object proposals in the mini-batch.
- Several limitations still remain.
  - We hope that this work draw community's attention on open-set problem in many other areas.

Thank You!

Questions?