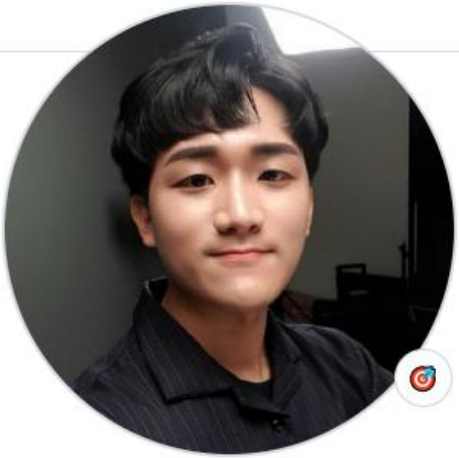
Vision AI 학습 데이터 라벨링 비용 20배 절약 방법

Weakly-supervised learning 산업 활용

조상현 (OGQ)

Self Introduction



Sanghyun Jo
shjo-april

Deep Learning Research Engineer

Work Experience: 회사 경험 (6년)

- ✓ GYNetworks - Research Engineer (17.07 ~ 22.12)
- ✓ OGQ (GYNetworks 인수) - Research Engineer (23.01 ~ Present)

Lecturer: 외부 강의/과외 경험 (5년, 약 150명)

- ✓ Tutoring for Developers or MS/PhD Students (18.06 ~ Present)
- ✓ NEXTPAGE - 비전공자를 위한 인공지능 강의 (18.10 ~ 20.09)
- ✓ EasyDeep - 비전공자를 위한 인공지능 강의 및 촬영 (20.11 ~ 21.02)
- ✓ 모두의연구소 – 논문 발표 및 비전공자를 위한 논문 작성 가이드 (22.10 ~ 22.12)

Reviewer: 국제 학회 리뷰어 경험 (4회, 17편)

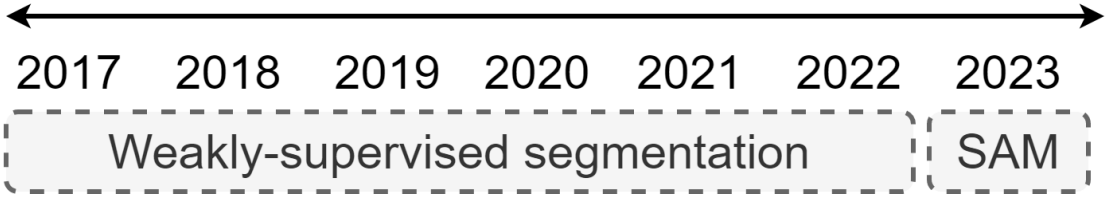
- ✓ ICIP 2022, ICIP 2023, IEEE TCSVT 2022, ICML 2022

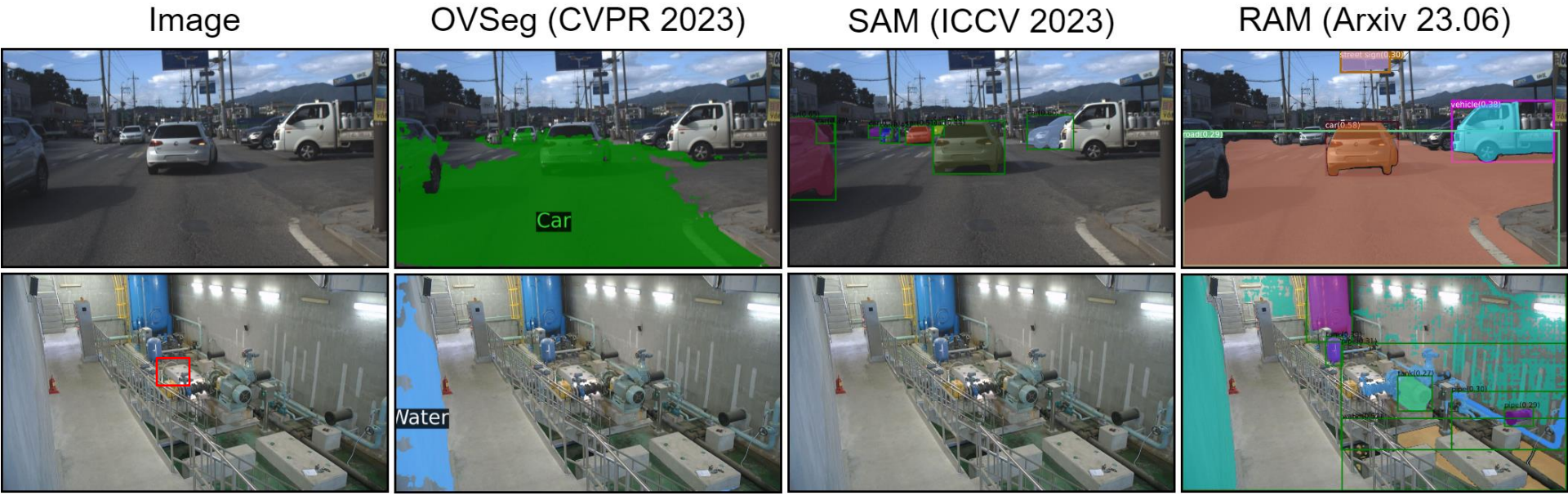
Papers: 인공지능 관련 저자 경험 (3회)

- ✓ **Puzzle-CAM: Improved localization via matching partial and full features** (☆: 161, Citations: 56, Rank: 10, ICIP 2021)
- ✓ **RecurSeed and EdgePredictMix: Single-stage learning is sufficient for Weakly-Supervised Semantic Segmentation** (☆: 41, Citations: 4, Rank: 3, Under Review)
- ✓ **MARS: Model-agnostic Biased Object Removal without Additional Supervision for Weakly-Supervised Semantic Segmentation** (☆: 5, Citations: 0, SOTA , ICCV 2023)

Future Direction of Weakly-supervised segmentation

- Before 2023: **Weakly-supervised segmentation**
- After 2023: **Open-vocabulary segmentation**





Background

What is the labeling for AI?

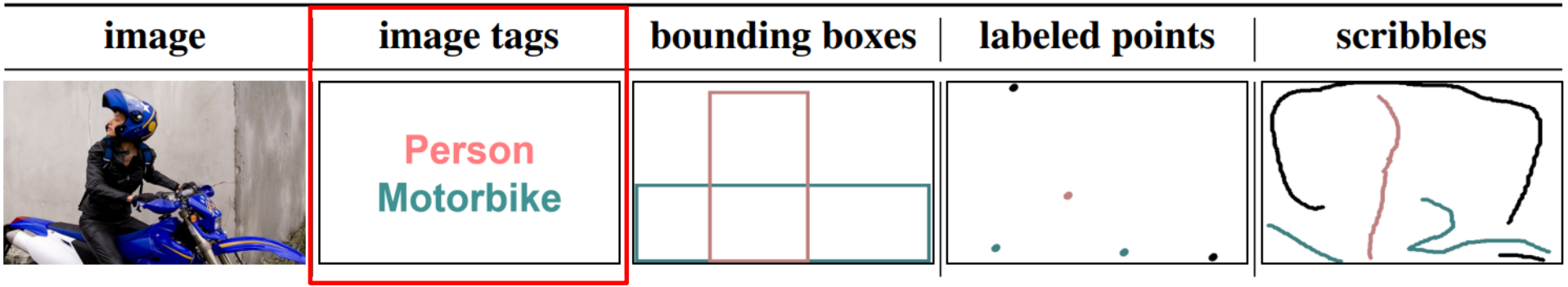
✓ In deep learning, data labeling is the process of identifying raw data (images, videos, etc.) and adding more informative labels.

[1] Amy et al., What's the Point: Semantic Segmentation with Point Supervision (ECCV 2016)

Background

Labeling for weakly-supervised semantic segmentation

✓ Tags, bounding boxes, points, and scribbles.

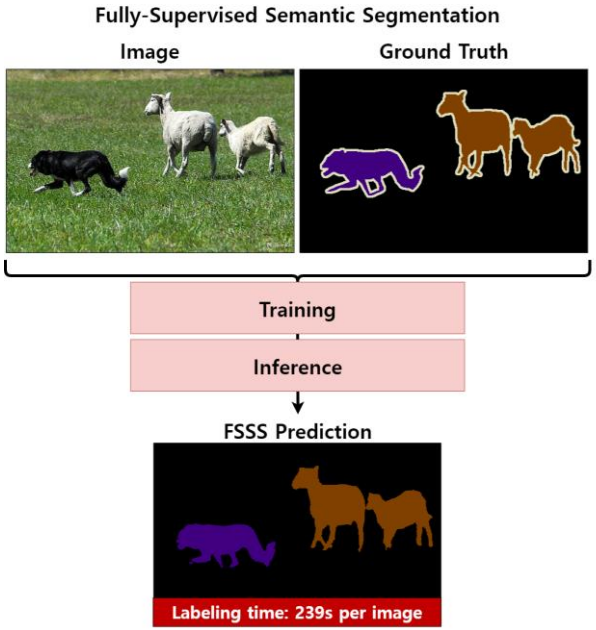


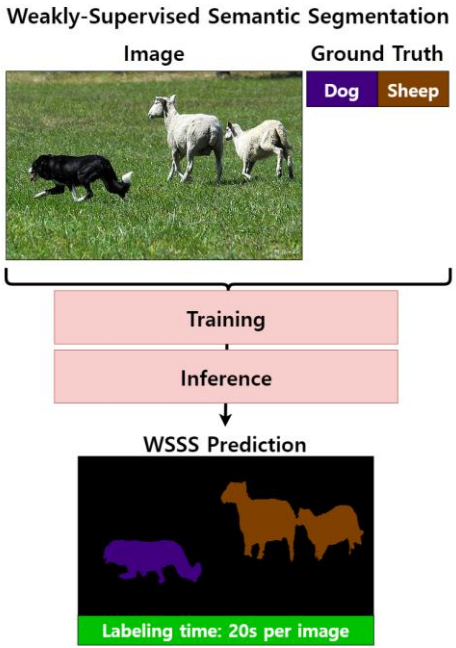
[1] Ke et al., Universal weakly supervised segmentation by pixel-to-segment contrastive learning (ICLR 2021)

Background

What is the difference between FSSS and WSSS?

- ✓ Fully-supervised semantic segmentation requires pixel-wise annotations.
- ✓ Weakly-supervised semantic segmentation requires image-wise annotations (i.e., tags).

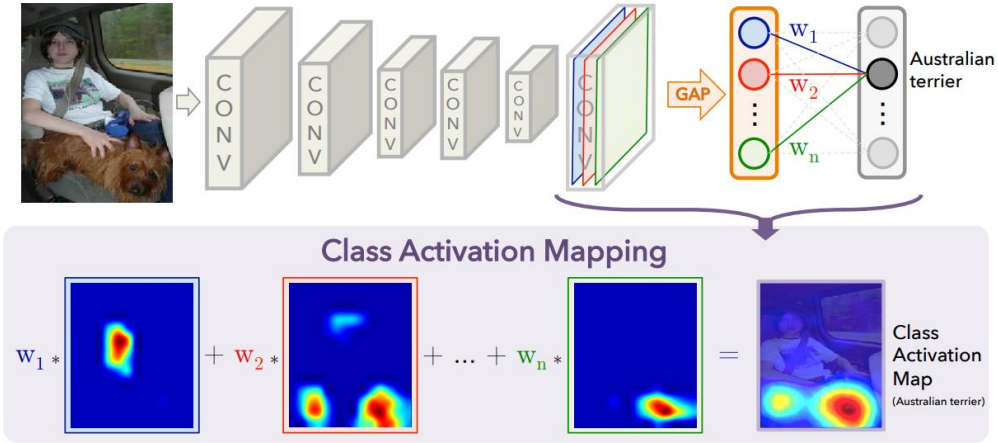


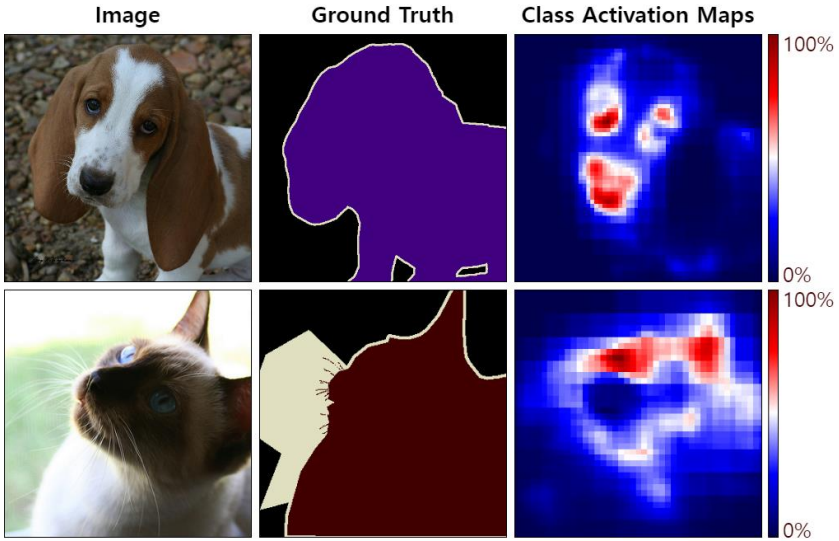


Background

The limitations of a class activation maps[1] (CAM)

✓ CAM is designed to visualize the most discriminative part of an object.



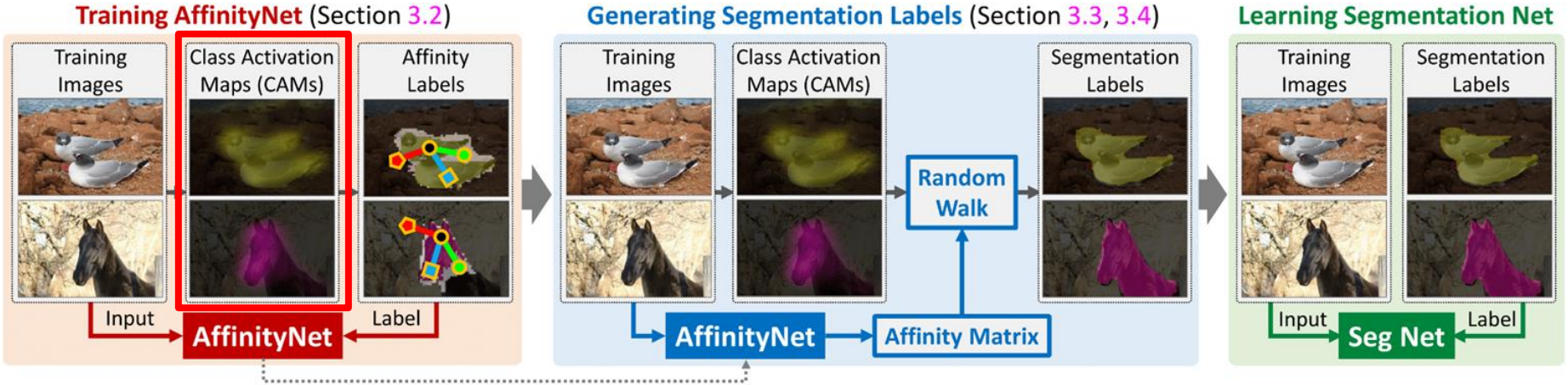


[1] Zhou et al., Learning Deep Features for Discriminative Localization (CVPR 2016)

Background

Most WSSS methods use multi-stage learning frameworks (MLFs).

1. Training the classification model to produce the CAM.
2. Training the AffinityNet[1] or IRNet[2] to generate pseudo label using a random walk (RW).
3. Training the segmentation model with pseudo labels.



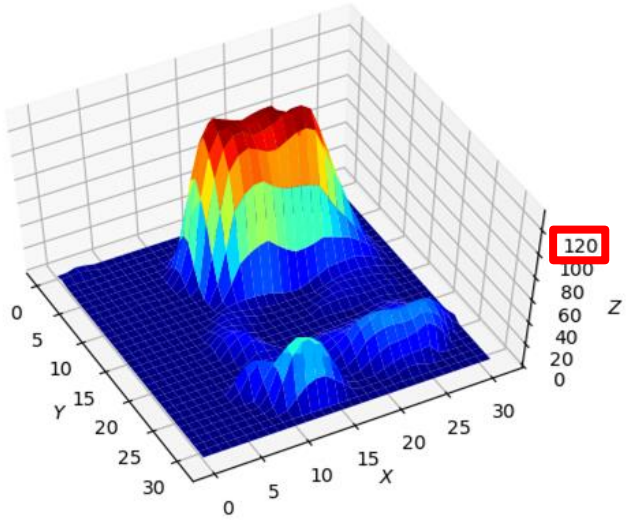
[1] Ahn et al., Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation (CVPR 2018)
[2] Ahn et al., Weakly supervised learning of instance segmentation with inter-pixel relations (CVPR 2019)

Puzzle-CAM / Motivation

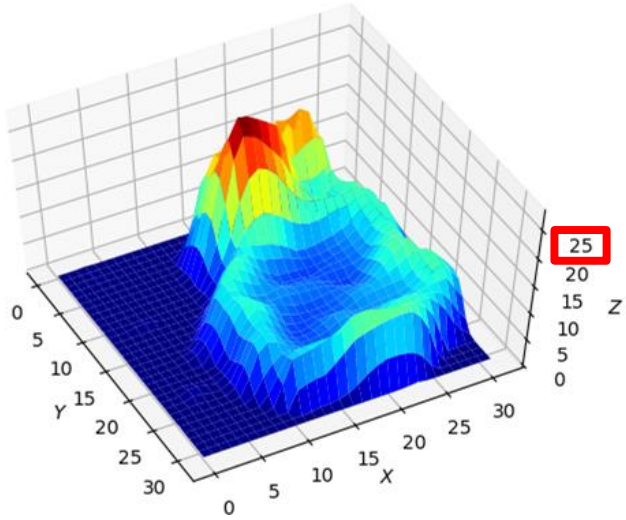
How to avoid detecting the discriminative region of the object?

✓ Puzzle-CAM suppresses the attention on the discriminative region of the object.

CAM



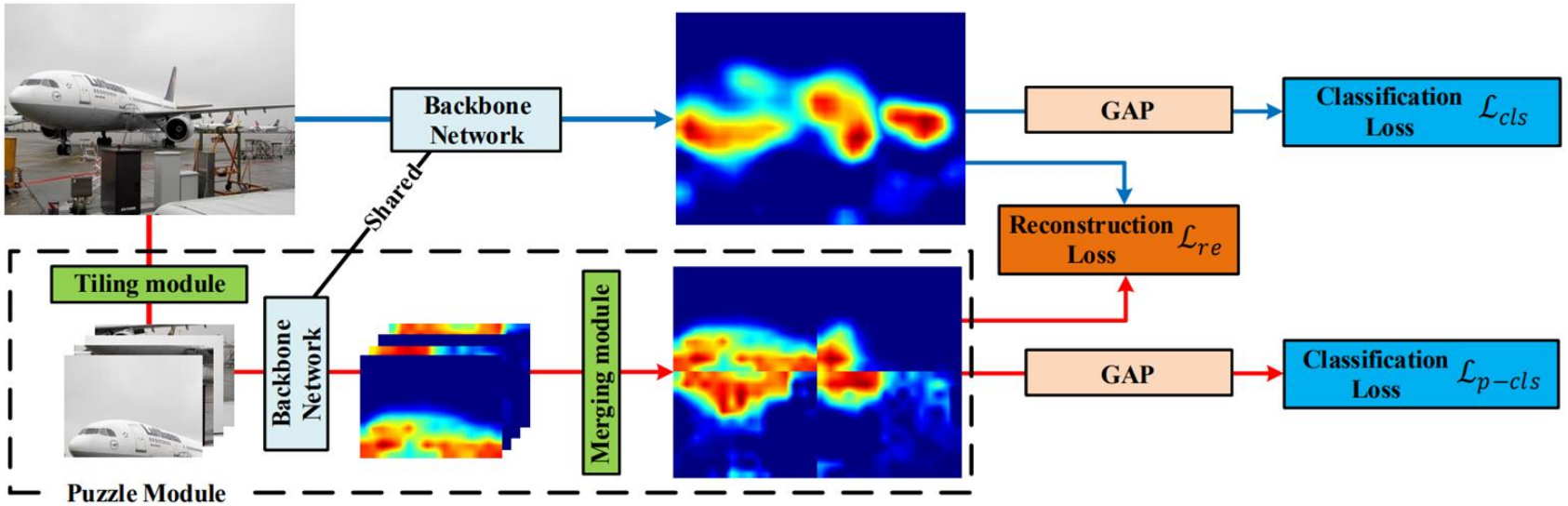
Puzzle-CAM (Ours)



Puzzle-CAM / Method

How to learn the integral region of the object by using the tags?

1. Tilling an image to image patches to divide into the attention.
2. Merging the feature maps from the network to produce the reconstructed features.
3. Matching partial and full features with reconstructing regularization.

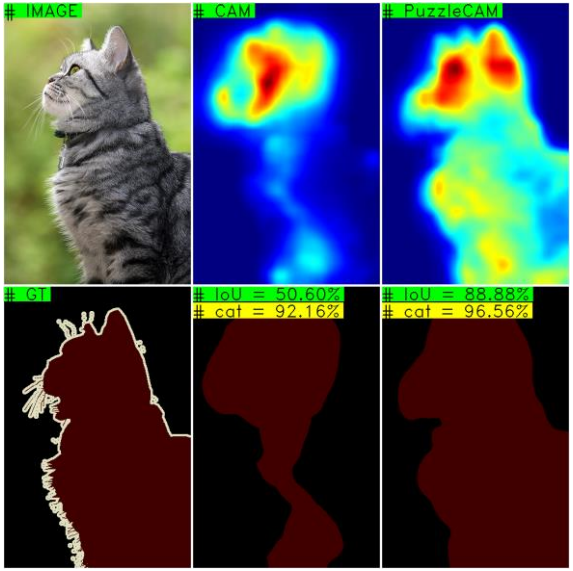


Puzzle-CAM / Limitation

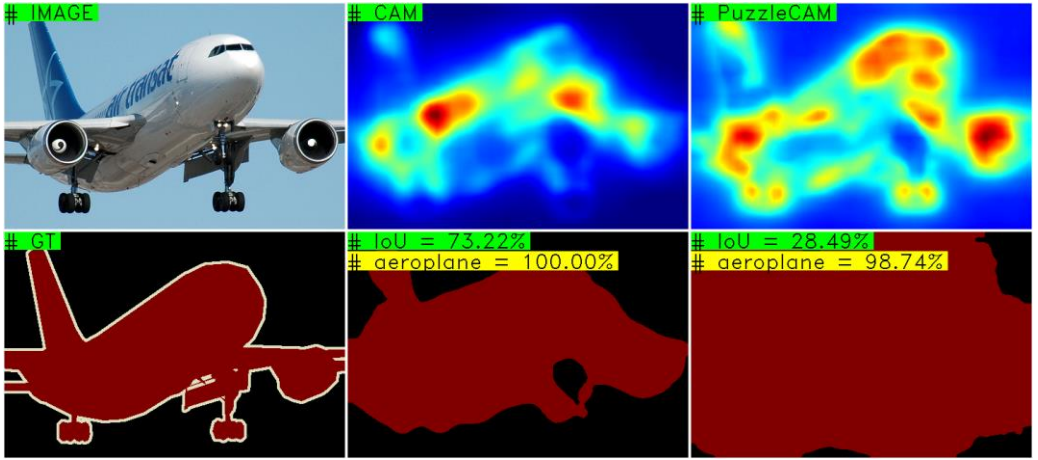
Qualitative results

✓ Our method causes over-activated CAMs on similar background or small objects.

Successful case



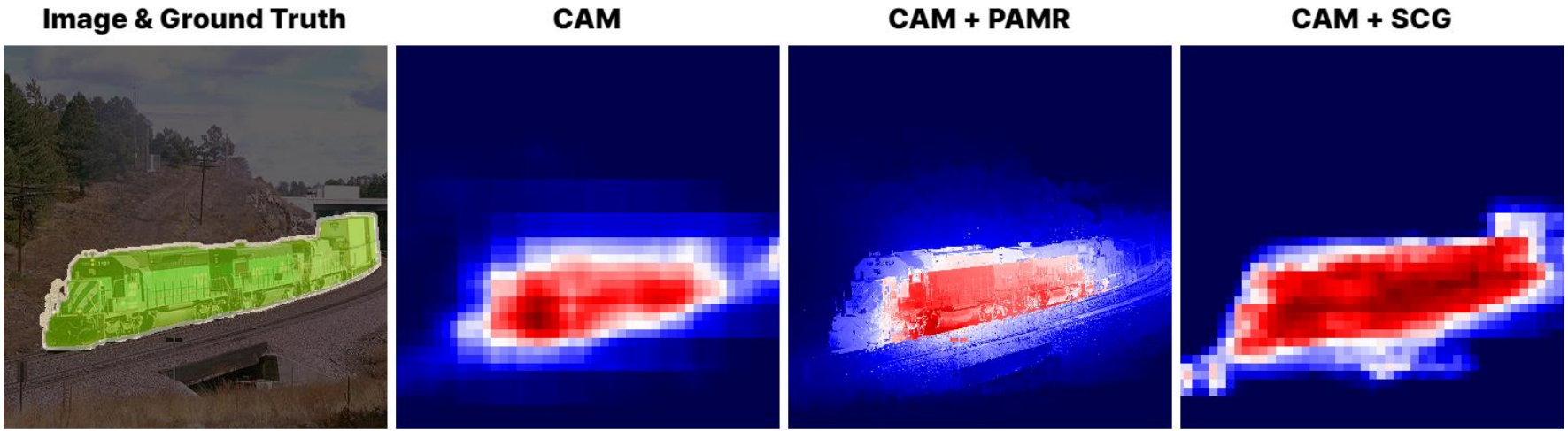
Typical failure case



RS+EPM / Motivation

The drawbacks of SCG[1] and PAMR[2]

- ✓ SCG decreases the false negative of the CAM but increases the false positive of the CAM.
- ✓ PAMR decreases the false positive of the CAM but increases the false negative of the CAM.



[1] Pan et al., Unveiling the Potential of Structure Preserving for Weakly Supervised Object Localization (CVPR 2021)
[2] Araslanov and Roth, Single-Stage Semantic Segmentation from Image Labels (CVPR 2021)

RS+EPM / Motivation

The shortcoming of existing DA approaches in the WSSS setup

- ✓ The simple synthesis without any refinements using predicted masks accelerates the ambiguity of mixed results due to insufficient quality of the mixed masks.

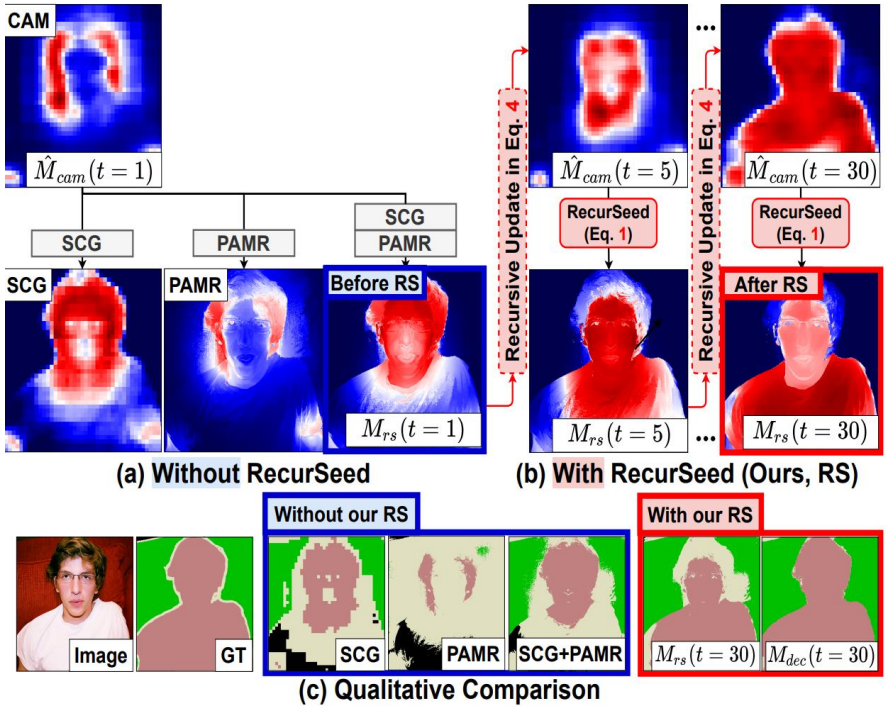
ClassMix [1]

[1] Olsson et al., Classmix: Segmentation-based data augmentation for semi-supervised learning. (WACV 2021)

RS+EPM / Method

The details of the proposed RecurSeed (RS).

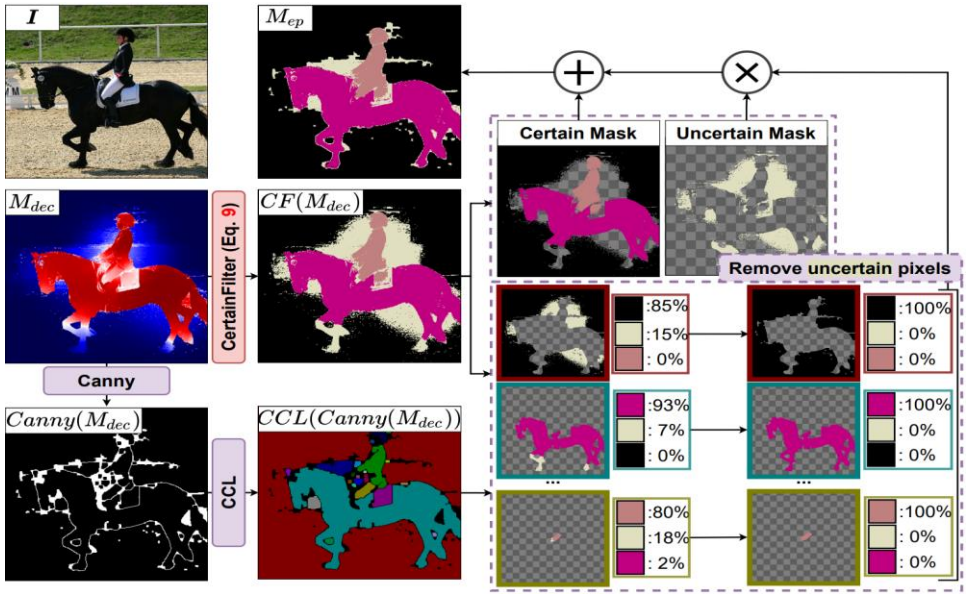
✓ We propose RecurSeed that recursively rectifies the CAM by leveraging SCG and PAMR, making a seed that minimizes both the false positive and false negative.

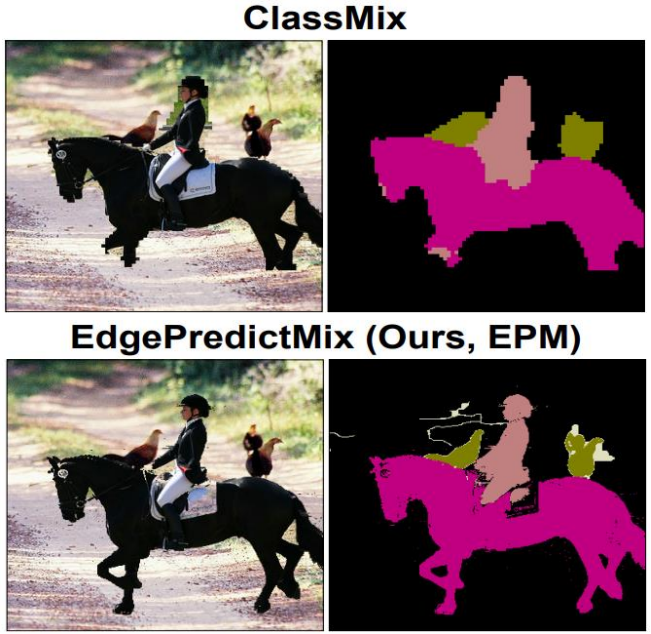


RS+EPM / Method

The details of the proposed EdgePredictMix (EPM).

- ✓ We propose EdgePredictMix that remedies the uncertainty of mixed results by reviving edge information of target objects.
- ✓ Our EP refines predicted masks by utilizing the absolute (using CF) and relative (using Canny) per-pixel probability information.

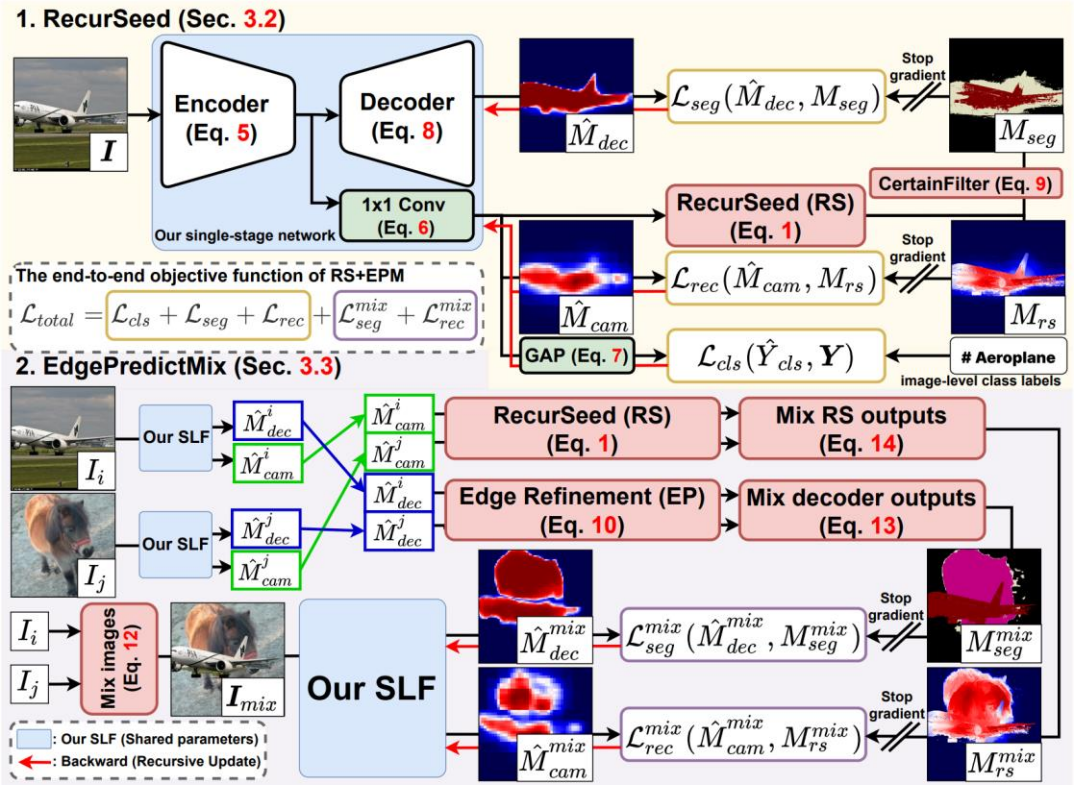




RS+EPM / Method

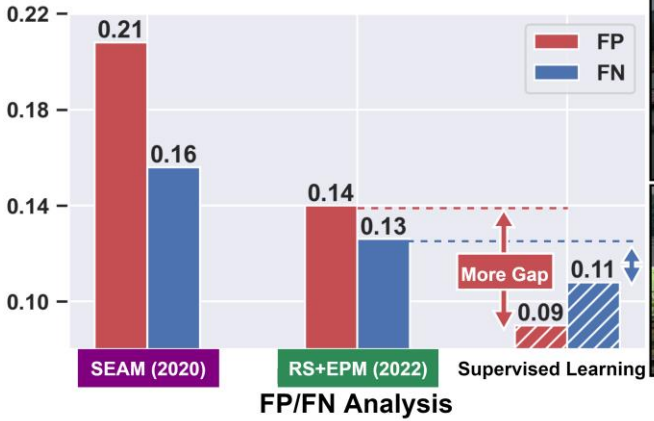
Overview of the single-stage learning framework with RS and EPM.

1. The classifier and decoder produce the CAMs and fine-grained segmentation masks

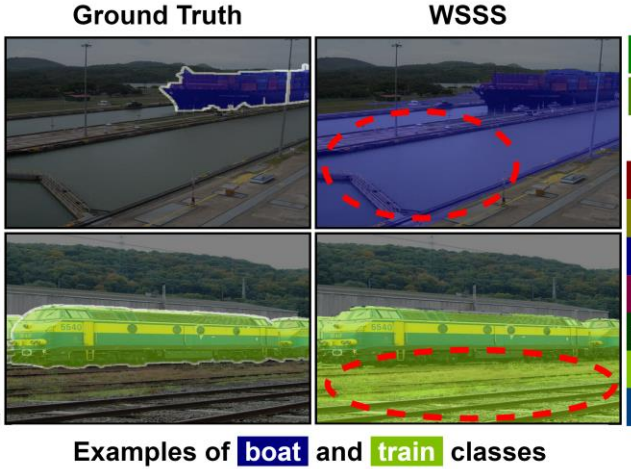


WSSS / Limitation

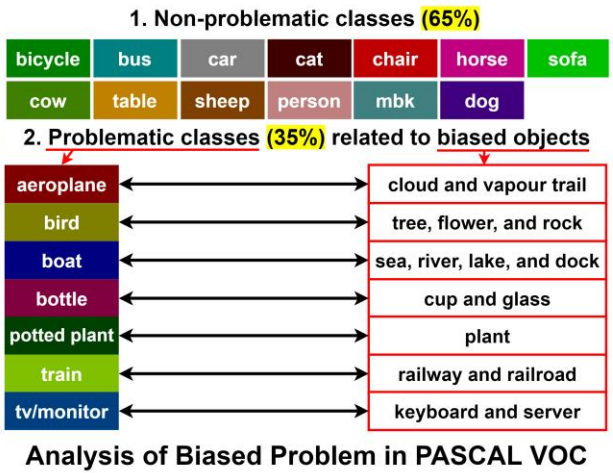
1. The FP is the most crucial bottleneck for WSSS methods.
2. Predicting class-related objects with target classes are factored into increasing FP.
3. 35% of classes in the VOC dataset have biased objects.



FP/FN Analysis



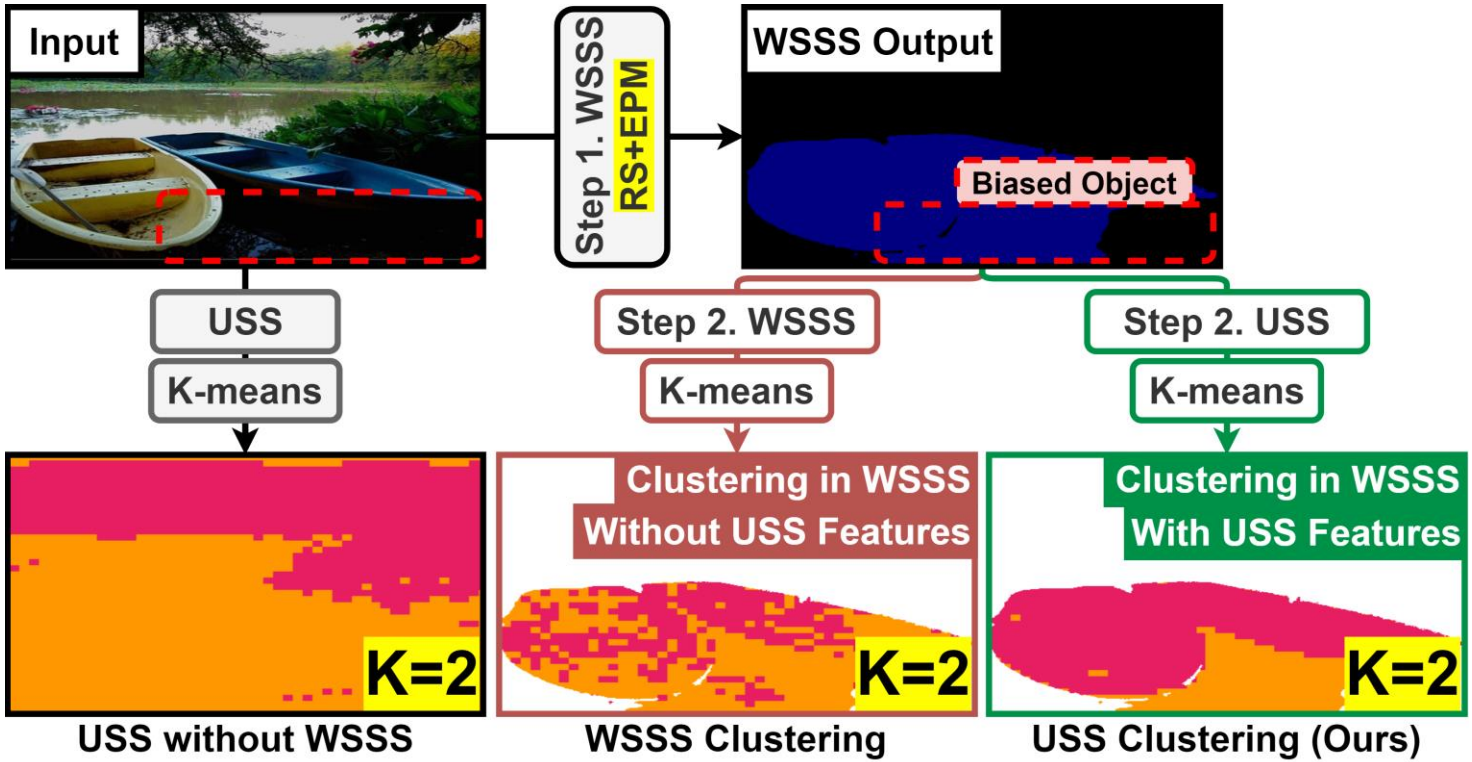
Examples of boat and train classes



Analysis of Biased Problem in PASCAL VOC

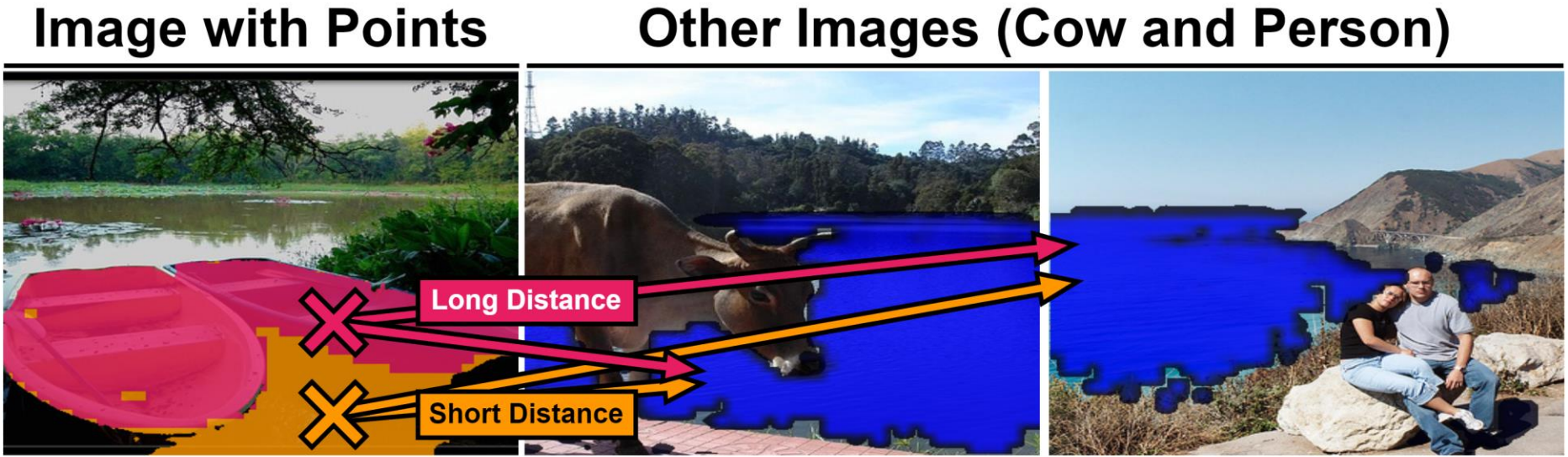
MARS / Motivation

1. USS feature clustering separates biased and target objects.
2. The proposed distance metric selects the biased object among two isolated objects.



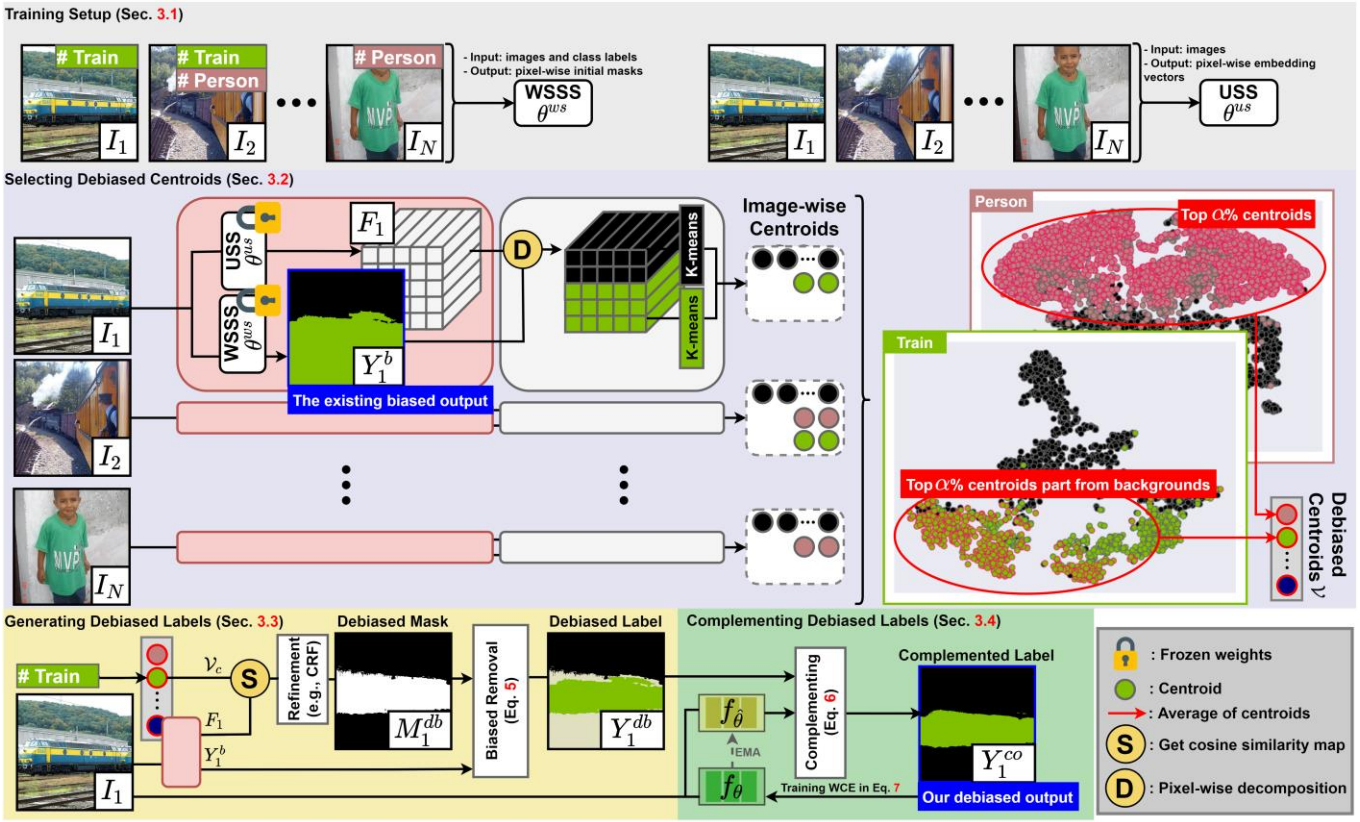
MARS / Motivation

1. USS feature clustering separates biased and target objects.
2. The proposed distance metric selects the biased object among two isolated objects.



MARS / Method

- Overview of MARS. Our method consists of three stages:



MARS / Experiments

- SOTA performance on WSSS benchmarks.

Method	Backbone	Sup.	VOC		COCO
			val	*test*	*val*
DSRG CVPR'18 [16]	R101	$\mathcal{I}$+$\mathcal{S}$	61.4	63.2	26.0*
W-OoD CVPR'22 [29]	R101	$\mathcal{I}$+$\mathcal{D}$	69.8	69.9	-
L2G CVPR'22 [19]	R101	$\mathcal{I}$+$\mathcal{S}$	72.1	71.7	44.2
RCA CVPR'22 [58]	R101	$\mathcal{I}$+$\mathcal{S}$	72.2	72.8	36.8*
PPC CVPR'22 [12]	R101	$\mathcal{I}$+$\mathcal{S}$	72.6	73.6	-
SSDD ICCV'19 [43]	R101	$\mathcal{I}$	64.9	65.5	-
OAA ICCV'19 [18]	R101	$\mathcal{I}$	63.9	65.6	-
CONTA [56]	R101	$\mathcal{I}$	66.1	66.7	32.8
AdvCAM CVPR'21 [28]	R101	$\mathcal{I}$	68.1	68.0	-
RIB NeurIPS'21 [26]	R101	$\mathcal{I}$	68.3	68.6	43.8
AMN CVPR'22 [30]	R101	$\mathcal{I}$	69.5	69.6	44.7
SANCE CVPR'22 [32]	R101	$\mathcal{I}$	70.9	72.2	44.7†
RS+EPM Arxiv'22 [21]	R101	$\mathcal{I}$	74.4	73.6	46.4
MARS (Ours)	R101	$\mathcal{I}$	**77.7**	**77.2**	**49.4**
FSSS	R101	$\mathcal{F}$	80.6	81.0	61.8

MARS / Experiments

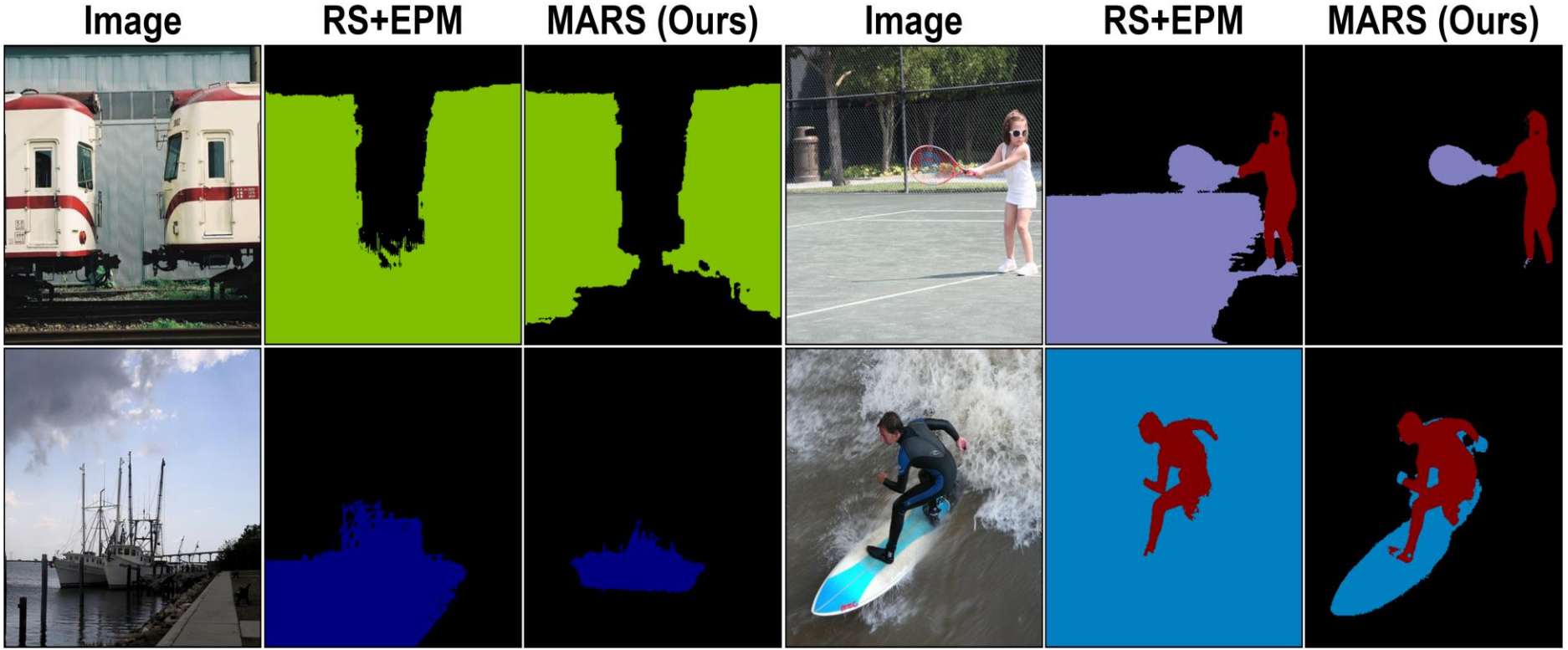
- Verify the flexibility of MARS by integrating it with various WSSS and USS methods.

Method	USS	Backbone	mIoU (*val*)	mIoU (*test*)
RS+EPM [21]	✗	R101	74.4	73.6
+ Ours	Leopart [59]	R101	75.4	75.8
+ Ours	STEGO [14]	R101	**77.7**	**77.2**

Method	Backbone	Segmentation	mIoU (*val*)	mIoU (*test*)
IRNet [1]	R50	DeepLabv2	63.5	64.8
+ Ours	R50	DeepLabv2	**69.8 (49%)**	**70.9 (52%)**
FSSS	R50	DeepLabv2	76.3	76.5
RS+EPM [21]	R101	DeepLabv3+	74.4	73.6
+ Ours	R101	DeepLabv3+	**77.7 (53%)**	**77.2 (49%)**
FSSS	R101	DeepLabv3+	80.6	81.0

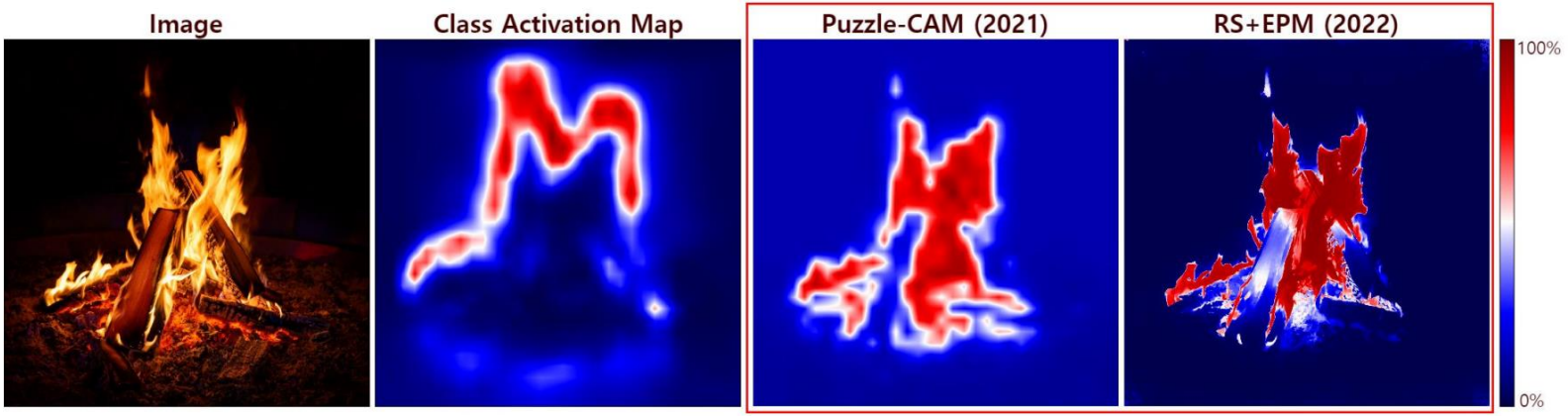
MARS / Experiments

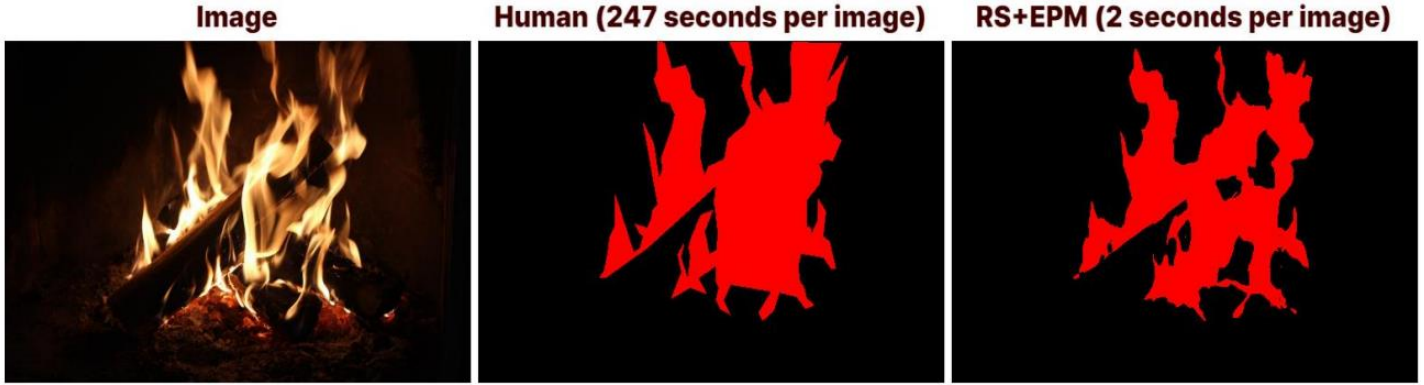
- Visualization of our segmentation results.



Results (Industry)

✓ Demonstrate the state-of-the-art performance of the proposed method.





Q & A

shjo.april@gmail.com

ICCV23 PARIS

MARS

Puzzle-CAM

RS+EPM