Precision matters: Precision-aware ensemble for weakly supervised semantic segmentation

AAAI-24 / IAAI-24 / EAAI-24

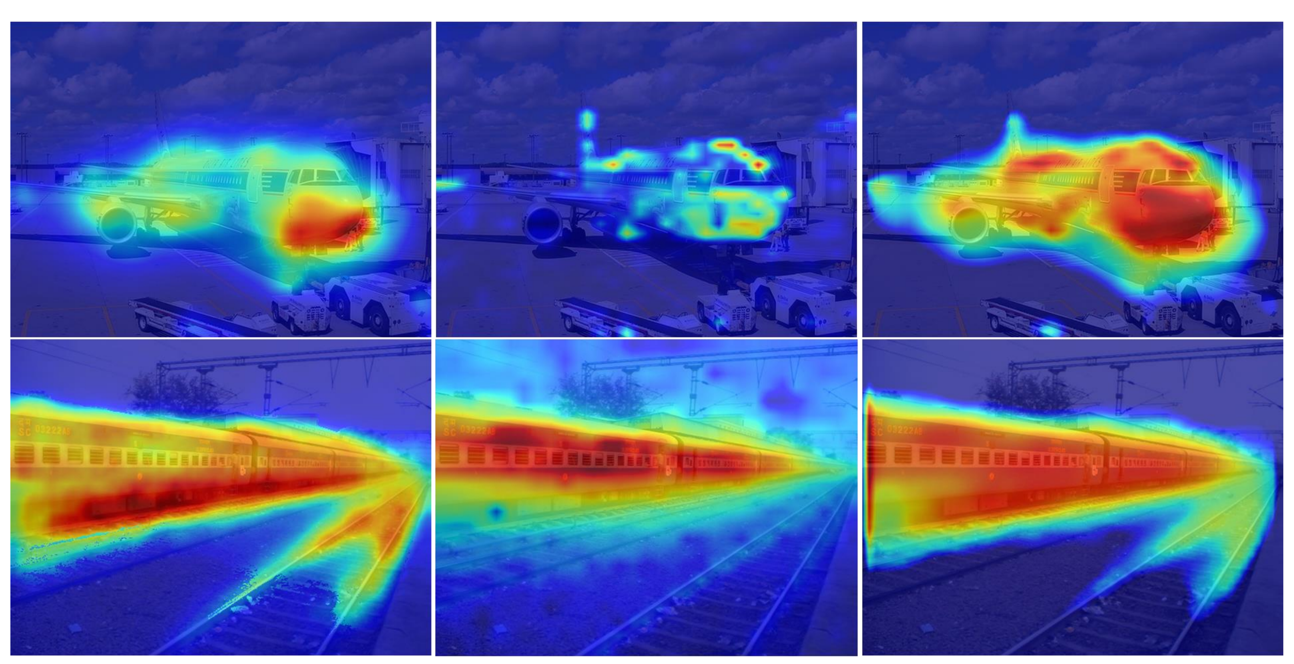
Junsung Park, Hyunjung Shim

Korea Advanced Institute of Science & Technology, The Kim Jaechul Graduate School of AI

Introduction

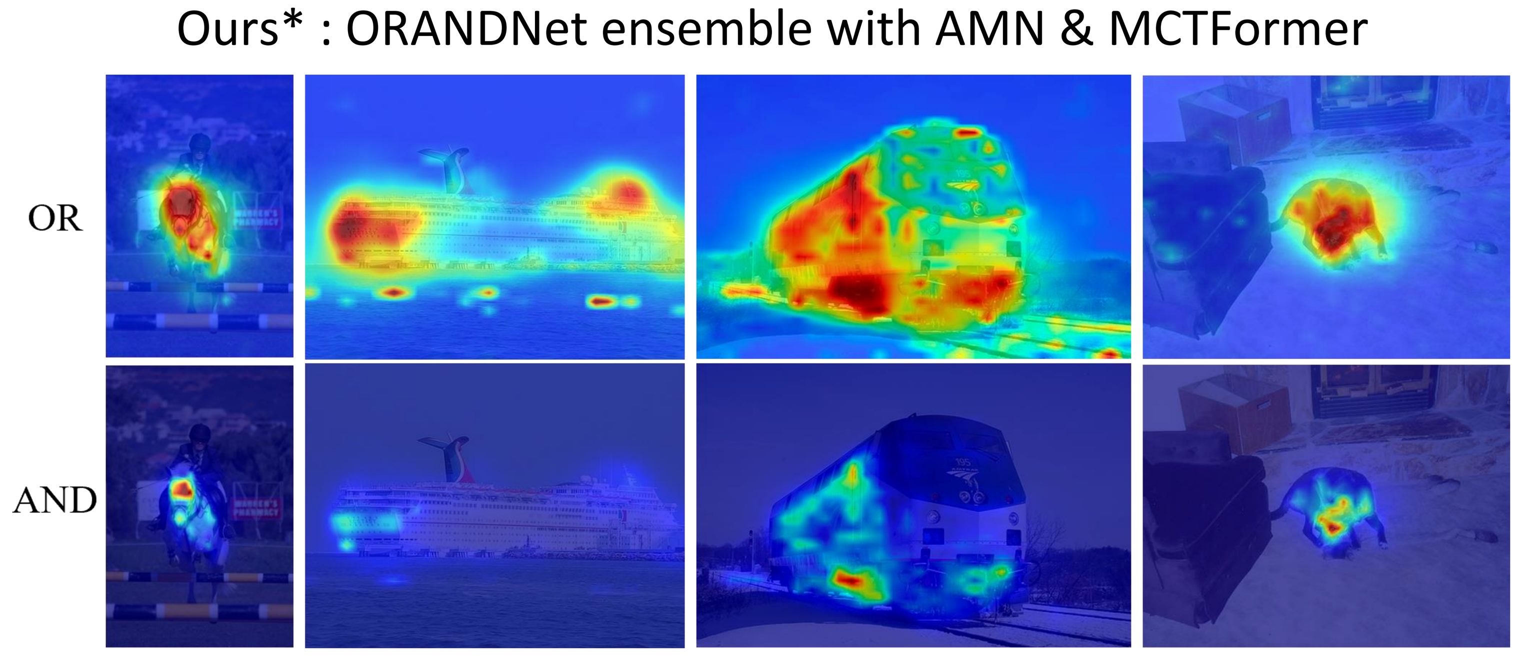
- **Problem**: In WSSS scenarios, pseudo labels with high mean Intersection of Union (mIoU) does not guarantee high segmentation performance.

- **Key observation:** CAMs from ResNet show concentrated activations at key object points. Meanwhile, ViT's are denser but weaker, covering more of the object. This inconsistency in activation maps also appears in various models with ResNet / ViT backbone (e.g. AMN / MCTFormer).

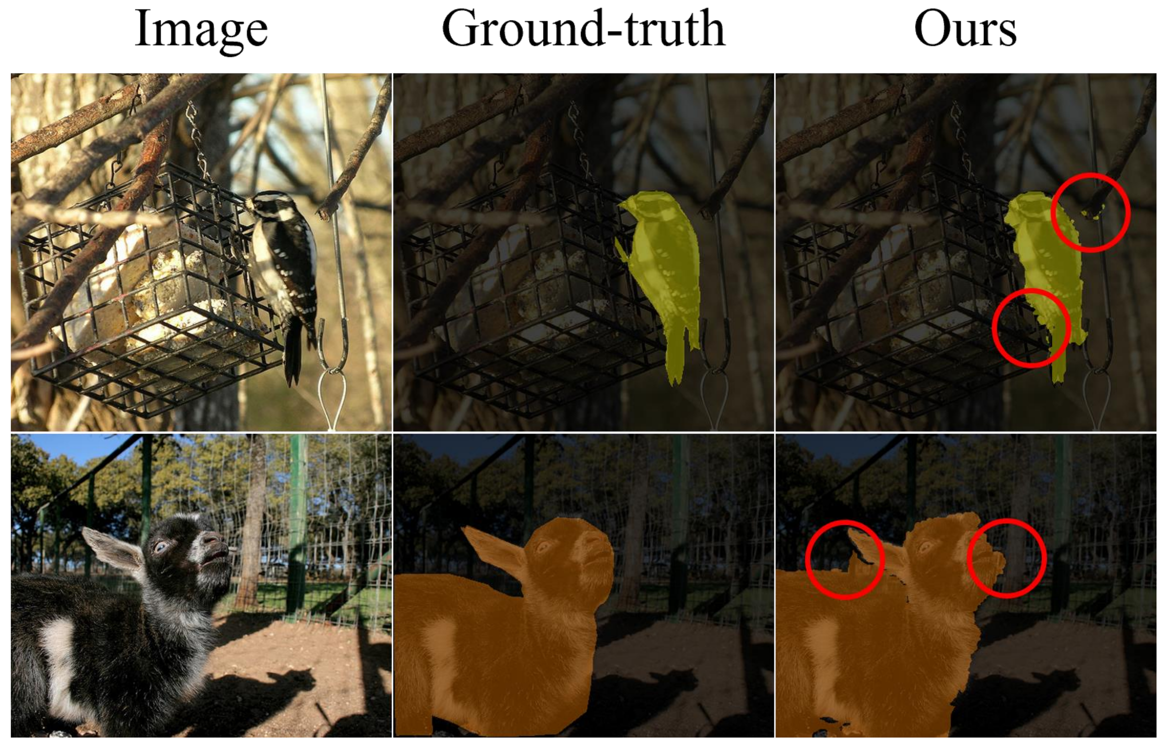


Ours : ORANDNet ensemble with ResNet-50 & ViT

Ours* : ORANDNet ensemble with AMN & MCTFormer



Result of a probabilistic AND on the activation maps of ResNet & ViT.

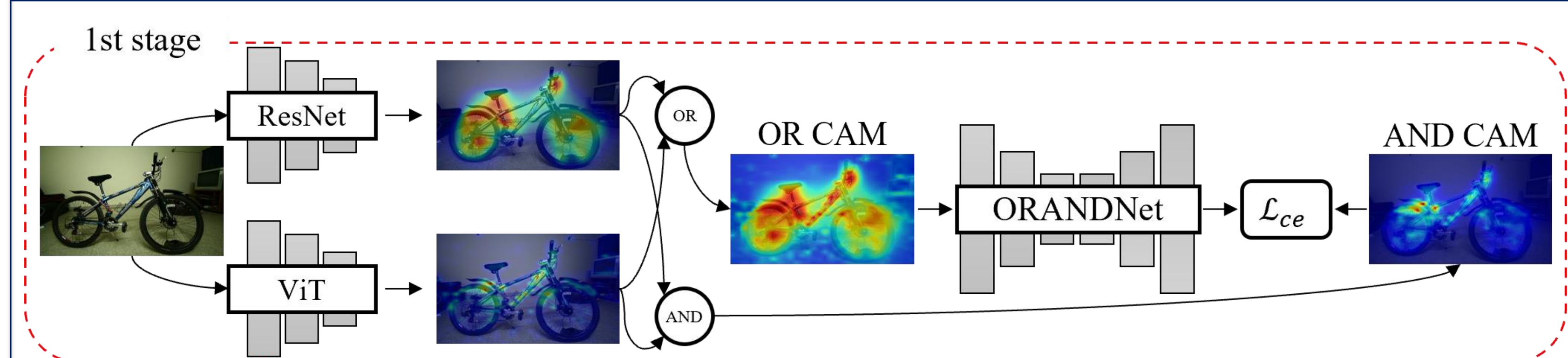


Pseudo masks from ORANDNet.

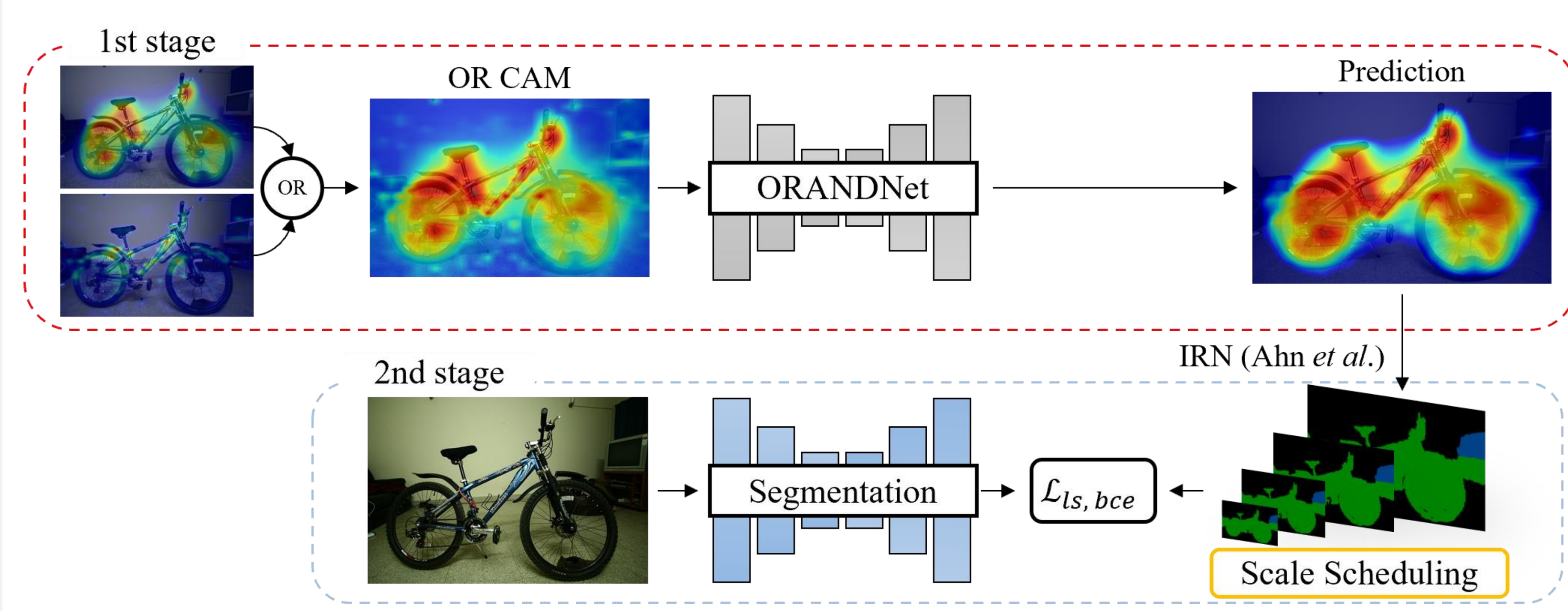
Red circles indicate small-sized error in the pseudo mask.

- **Idea 1) ORANDNet combines Class Activation Maps (CAMs) from two different classifiers to increase the precision of pseudo-masks (PMs).**
- **Idea 2)** Training a segmentation model **initially with pairs of smaller-sized images & PMs, gradually transitioning to the original-sized pairs** to further mitigate the size error obtained from ORANDNet's & postprocessing module's incorrect PM predictions.

Method



- A probabilistic AND operation between two distinct CAMs can increase precision of PMs but reduce the amount of available labeled data.
- **ORANDNet is trained with the AND CAM as the pseudo-mask and the OR CAM as the input to achieve high mIoU and precision simultaneously.**



- When generating the PM in 1st stage, the OR CAM is utilized as the input for ORANDNet.
- During training in 2nd stage, We adopt **scale scheduling that downsamples training data (image and pseudo-mask pairs) in early training.**

Conclusion

- ORANDNet improves mIoU and precision, acheiving performance on par with state-of-the-art even with basic classifiers like ResNet-50 and ViT.
- Scale scheduling further prevents the segmentation model from learning the noise in the pseudo mask.

Experiments

- Pseudo-mask quality on PASCAL VOC 2012

Method	mIoU	Precision	Recall
ResNet-50	48.3	66.9	64.8
ViT	53.4	67.7	72.4
Naïve ensemble	49.0	61.5	72.2
Ours	**54.3** (+5.0)	**71.0** (+9.5)	70.9 (-1.3)
Ours w/IRN	**70.9**	**82.9**	82.6
AMN	62.8	74.0	80.5
MCTformer	62.1	74.7	78.8
Ours*	**64.3** (+1.5, +2.2)	**78.9** (+4.9, +4.2)	77.9 (-2.6, -0.9)
Ours* w/IRN	**74.3**	**85.5**	84.3

- mIoU on PASCAL VOC 2012

Method	Stg	Sup.	Val	Test
EDAM (Wu et al. 2021)	1st	I+S	70.9	70.6
EPS (Lee et al. 2021)	1st	I+S	71.0	71.8
SANCE (Li, Fan, and Zhang 2022)	1st	I+S	72.0	72.9
OC-CSE (Kweon et al. 2021)	1st	I	68.4	68.2
CDA (Su et al. 2021)	1st	I	66.1	66.8
PPC (Du et al. 2022)	1st	I	72.6	**73.6**
URN (Li et al. 2022)	2nd	I	69.5	69.7
ADELE (Liu et al. 2022)	2nd	I	69.3	68.9
BECO (Rong et al. 2023)	2nd	I	**73.7**	73.5
IRN (Ahn, Cho, and Kwak 2019)	1st	I	63.5	64.8
Ours	1st	I	70.3	72.1
Relative to IRN			+6.8	+7.3
AMN (Lee, Kim, and Shim 2022)	1st	I	70.7	70.6
MCTformer (Xu et al. 2022)	1st	I	71.9	71.6
Ours*	1st	I	72.2	72.9
Relative to AMN, MCTFormer			+1.5, +0.3	+2.3, +1.3

- Effect of scale scheduling

Method	Baseline	Scale scheduling
Deeplabv1(FSSS)	76.4	76.3
Naïve ensemble	68.5	68.4
Ours	**70.1** (+1.6)	**70.3** (+1.9)
Ours*	**71.3** (+2.8)	**72.2** (+3.8)

- Qualitative results