Consistency-CAM
Towards Improved Weakly supervised Semantic Segmentation

Sai Rajeswar, Issam Laradji, Pau Rodriguez, David Vazquez, Aaron Courville

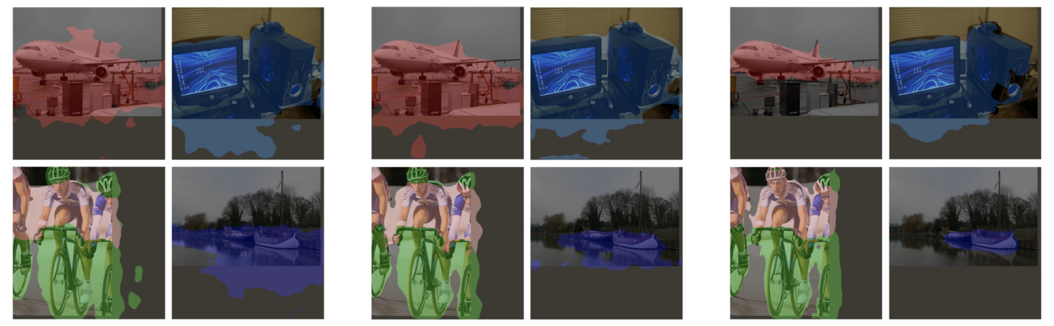
BMVC 2022

TL;DR

- We identify and propose three key improvements to high performing weakly supervised semantic segmentation (WSSS) tasks. The resulting Consistency-CAM framework attains superior performance on PascalVOC and MSCOCO datasets.

Introduction

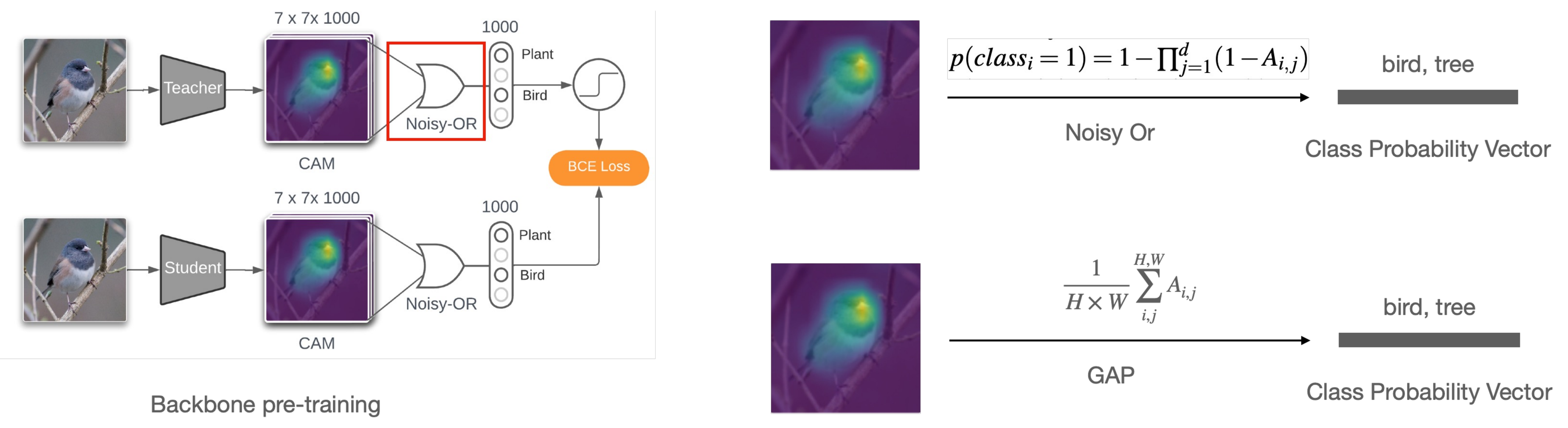
- Typical pipelines for WSSS are trained in two stages.
  - 1. Train a classification network with global average pooling to obtain 2d class activation maps (CAMs). 2. Train a segmentation network using CAMs as pixel-level supervision.
  - Issue: CAMs are noisier than real labels and needs refinement using some regularization.
- Puzzle-CAM splits the image into multiple tiles and ensures that the CAM for the image matches the CAM obtained after stitching the individual CAMs.
  - However, pre-training using single-label prediction has a negative effect since image segmentation datasets have more than one class.
  - Also, GAP enforces the network to overrepresent the labeled objects in the feature maps
  - Lastly, Puzzle-CAM Uses fixed tile sizes & the puzzle operation can be complemented with other transformation
- We address the above three issues highlighted using our Consistency-CAM pipeline.



(a) Puzzle-CAM (b) Puzzle-CAM+aug (c) Consistency-CAM

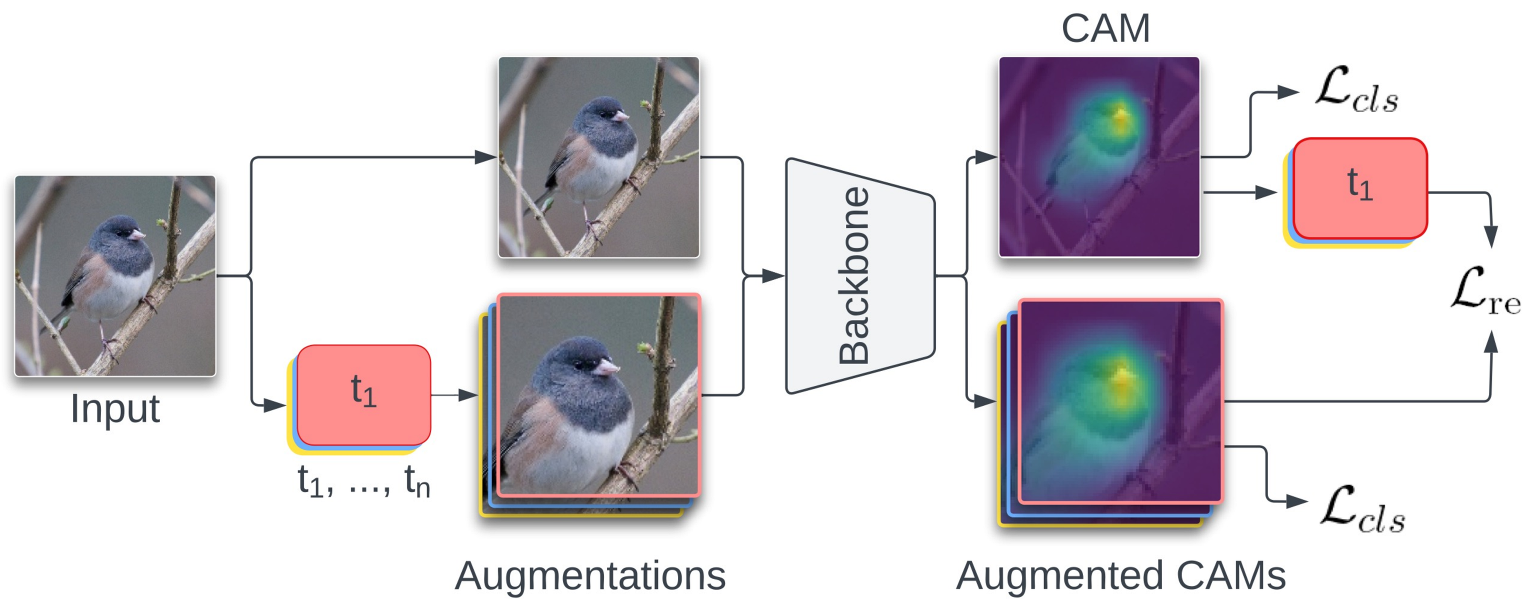
Method

1. we train the backbone with MILe, which learns multi-label representations from singly labeled images.
   - Ensures the backbone is able to predict multiple classes per image,
2. we replace GAP with a differentiable noisy-or operation. This marks the presence of a class in an image independent to the number of pixels that belong to that class.
   - With noisy-or a class will be active with high probability even if only one pixel is activated for that class.
3. we propose a more general set of transformations.
   - we use a consistency or reconstruction loss to ensure that the CAM is robust to a diverse set of augmentations as well as the puzzle operation in with different tile sizes.



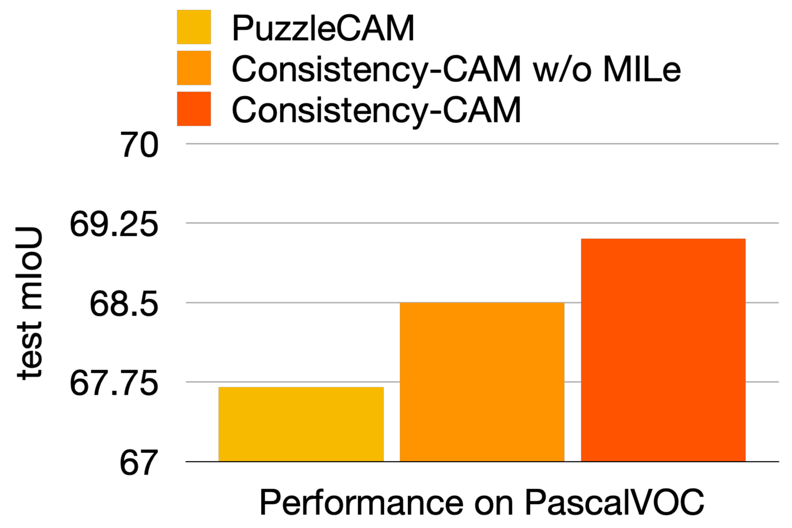
Multi-label iterated learning Pre-training with Noisy-OR

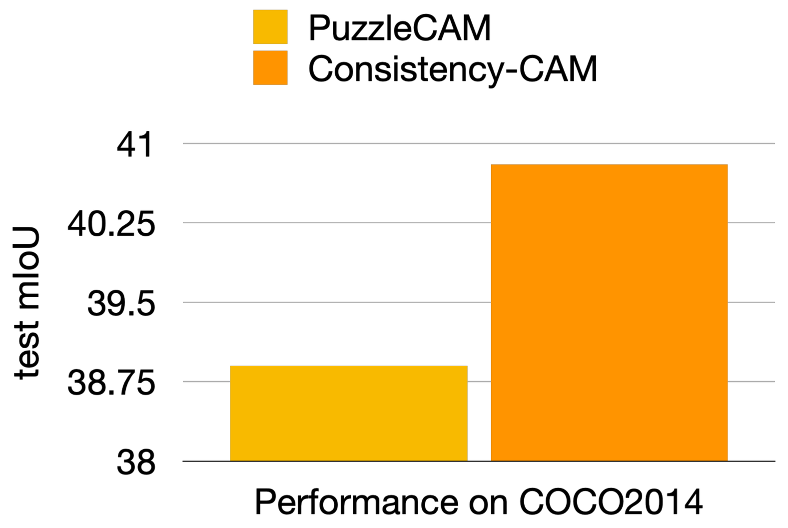
Qualitative Results



CAM Refinement

Quantitative Results





Conclusions

- We pretrain the backbone for multi-label classification.
- We change the GAP operation by a noisy-or operation
- We propose a more general set of augmentations for CAM refinement.
- Finally, these three improvements result in better performance on COCO and Pascal.
- Our method improves Puzzle-CAM by several points.
- We also see that training the backbone with multi-label learning is beneficial.

References

Iterated learning
- Kirby, Simon. "Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity." IEEE Transactions on Evolutionary Computation 5.2 (2001): 102-110.
- Rajeswar, Sai, et al. "Multi-label iterated learning for Image Classification with Label Ambiguity." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2021.

Datasets
- Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009.
- Everingham, M., et al. "PascalVOC 2012"

WSSS Models
- He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Jo, Sanghyun, et al "Puzzle-CAM: Improved localization via matching partial and full features." ICIP, 2021.

servicenow · Mila · Université de Montréal