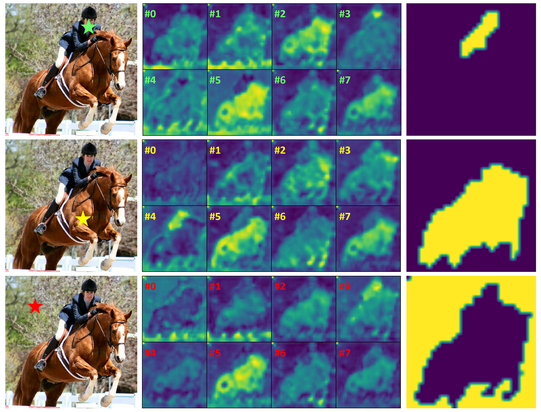
Learning Affinity from Attention: End-to-End Weakly-Supervised Semantic Segmentation with Transformers

Lixiang Ru, Yibing Zhan, Baosheng Yu, Bo Du

https://rulixiang.github.io/afa/

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Motivation: Deriving reliable sematic affinity from MHSA for label refinement.



(a) Images (b) Multi-head self-attention (c) Affinity

Contribution

- We propose an end-to-end Transformer-based framework for WSSS with image-level labels. To the best of our knowledge, this is the first work to explore Transformers for WSSS.
- We exploit the inherent virtue of Transformer and devise an Affinity from Attention (AFA) module. AFA learns reliable semantic affinity from MHSA and propagates the pseudo labels with the learned affinity.
- We propose an efficient Pixel-Adaptive Refinement (PAR) module, which incorporates the RGB and position information of local pixels for label refinement.

Method

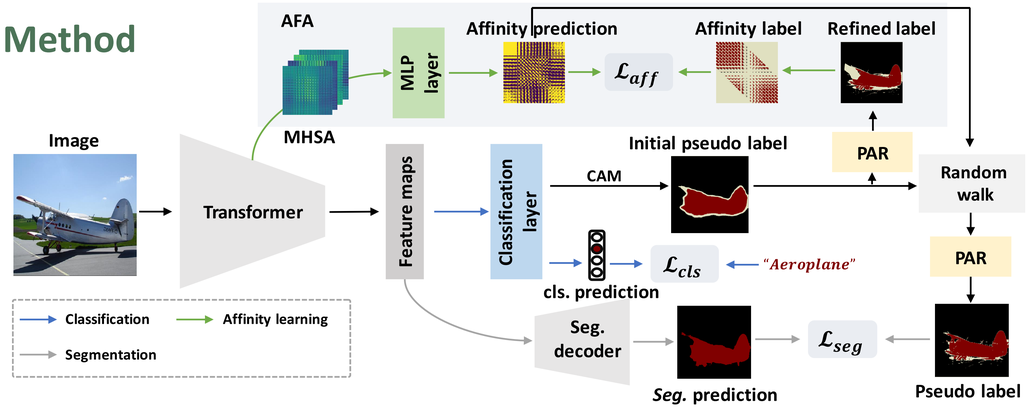
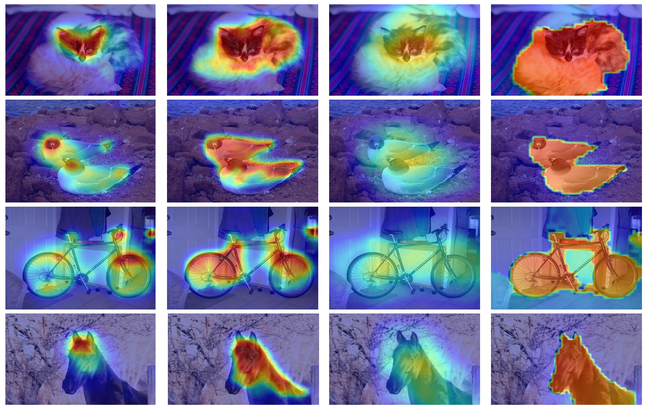


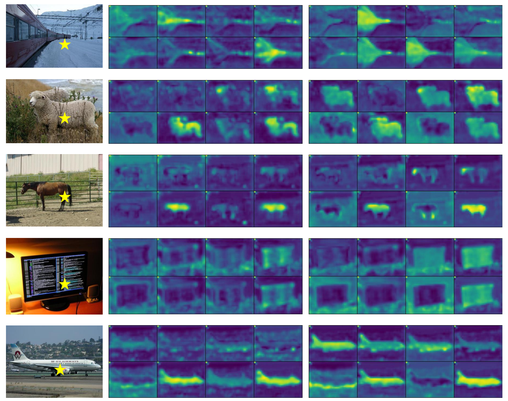
Fig 2: The proposed framework. The initial pseudo labels are generated with CAM and then refined with the proposed PAR. In AFA, we derive the semantic affinity from MHSA in Transformer blocks. Next, we employ the learned affinity to revise the pseudo labels via random walk propagation. The propagated labels are finally refined with PAR as the pseudo labels for the segmentation branch.

Fig 3: CAM & CAM refined with AFA.



(a) CNN CAM (b) Trans. CAM (c) Refine with MHSA (d) Ours

Fig 4: MHSA without/with AFA.

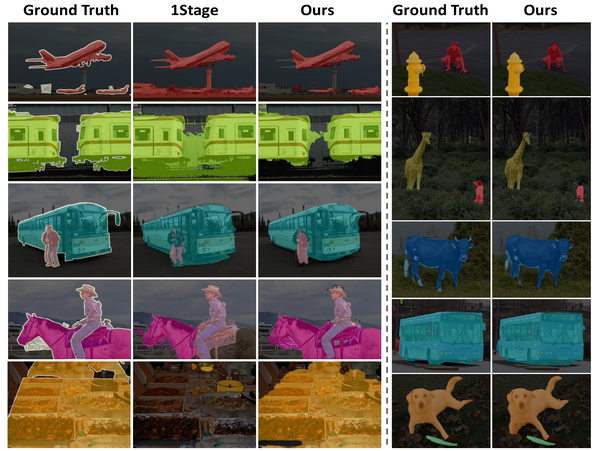


(a) Images (b) without AFA (c) with AFA

Results: Ablation, Pseudo Labels & Segmentation

Method	PAR	AFA	$\mathcal{L}_{reg}$	CRF	*val*
Our Baseline					46.7
Ours	✓				56.2
	✓	✓			62.6
	✓	✓	✓		63.8
	✓	✓	✓	✓	66.0

		train	*val*
PSA [2]		59.7	–
IRN [1]		66.5	–
1Stage [4]		66.9	65.3
Ours	*w/o* AFA	54.4	54.2
	AFA (*w/o prop.*)	66.3	64.4
	AFA (*prop.* with MHSA)	58.3	55.9
	AFA	**68.7**	**66.5**



PASCAL VOC MS COCO