Token Contrast for Weakly-Supervised Semantic Segmentation

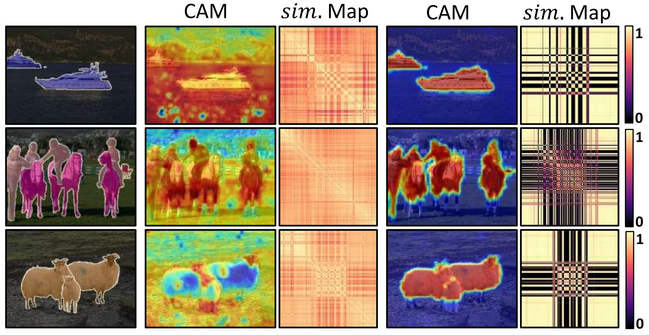
Lixiang Ru[1], Heliang Zheng[2], Yibing Zhan[2], Bo Du[1]

[1] School of Computer Science, Wuhan University [2] JD Explore Academy, JD.com

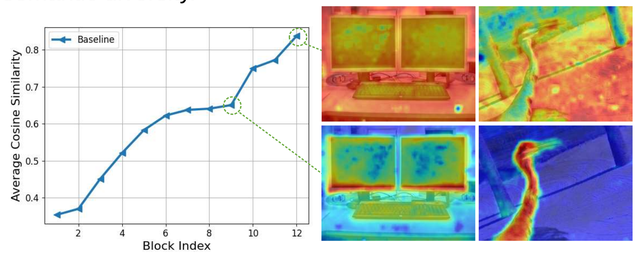
https://github.com/rulixiang/ToCo

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

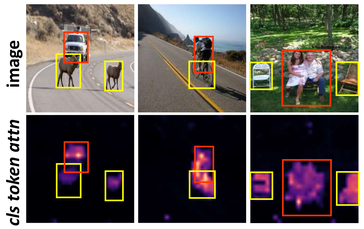
Goal: Addressing the over-smoothing issue of ViT and further leveraging its virtue for WSSS.



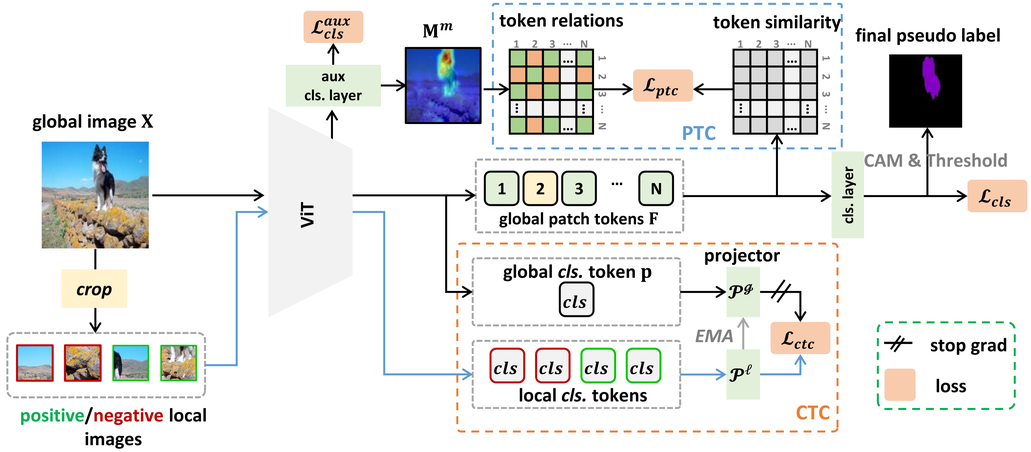
➢ **Motivation #1:** Intermediate layers can still retain the semantic diversity.



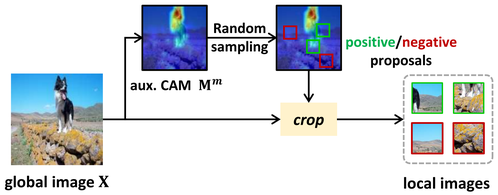
➢ **Motivation #2:** Class token can capture high-level foreground semantics.



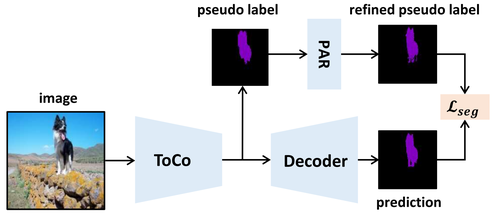
➢ **Method:** The overall framework of Token Contrast (ToCo).



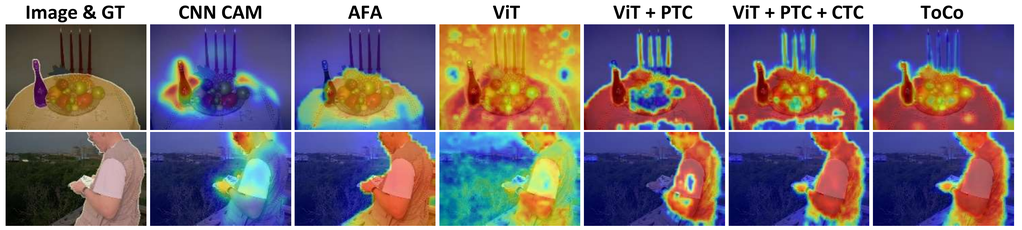
➢ **Random crop in ToCo:**



➢ **End-to-End WSSS based on ToCo:**



➢ **Class Activation Maps (CAM):**



➢ **Pseudo labels:**

Method	Backbone	train	val
RRM [50] AAAI'2020	WR38	–	65.4
1Stage [3] CVPR'2020	WR38	66.9	65.3
AA&LR [52] ACM MM'2021	WR38	68.2	65.8
SLRNet [28] IJCV'2022	WR38	67.1	66.2
AFA [34] CVPR'2022	MiT-B1	68.7	66.5
ViT-PCM [32] ECCV'2022	ViT-B†	67.7	66.0
ViT-PCM + CRF [32] ECCV'2022	ViT-B†	71.4	69.3
ToCo	ViT-B	72.2	70.5
ToCo†	ViT-B†	**73.6**	**72.3**

➢ **Semantic segmentation results:**

Single-stage WSSS methods.					
RRM [50] AAAI'2020	$\mathcal{I}$	WR38	62.6	62.9	–
1Stage [3] CVPR'2020	$\mathcal{I}$	WR38	62.7	64.3	–
AFA [33] CVPR'2022	$\mathcal{I}$	MiT-B1	66.0	66.3	38.9
SLRNet [28] IJCV'2022	$\mathcal{I}$	WR38	67.2	67.6	35.0
ToCo	$\mathcal{I}$	ViT-B	69.8	70.5[1]	41.3
ToCo†	$\mathcal{I}$	ViT-B†	**71.1**	**72.2**[2]	**42.3**

➢ **Analysis of PTC (left) and CTC (right):**