Separate and Conquer: Decoupling Co-occurrence via Decomposition and Representation for Weakly Supervised Semantic Segmentation

Zhiwei Yang *   Kexue Fu   Minghong Duan   Linhao Qu,
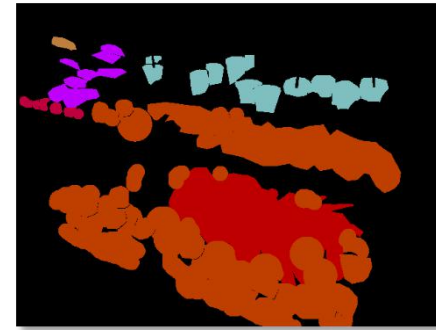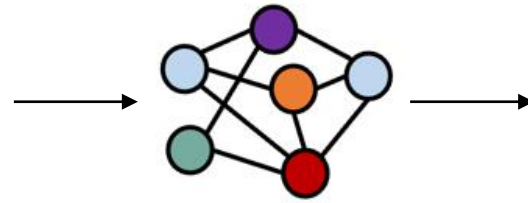Shuo Wang†   Zhijian Song†
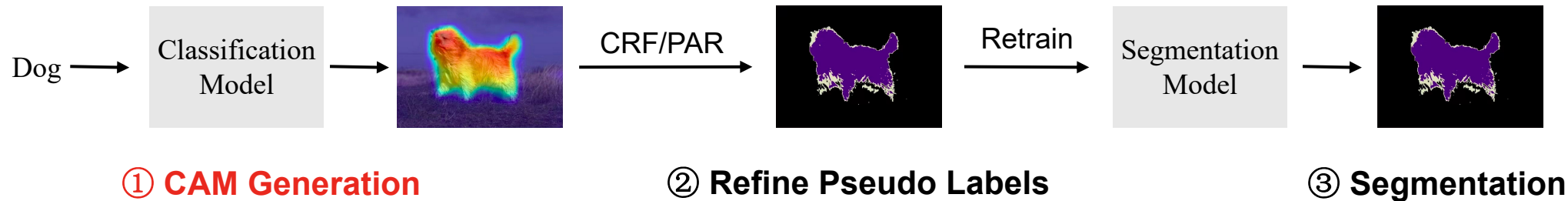
CODE

Contact: zwyang21@m.fudan.edu.cn

Weakly Supervised Semantic Segmentation

- Weakly supervised semantic segmentation (WSSS) with image-level labels aims to achieve segmentation tasks without dense annotations
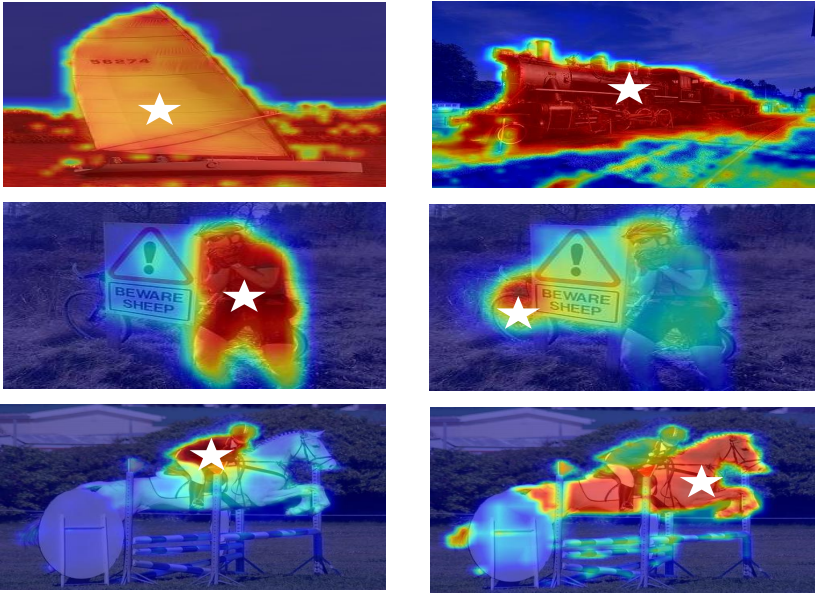

Banana, Orange, Pineapple, etc.

- WSSS Pipeline, Single-stage or Multi-stage

Dog → Classification Model → CRF/PAR → Retrain → Segmentation Model →

① **CAM Generation**   ② **Refine Pseudo Labels**   ③ **Segmentation**

Co–occurrence issue

- **Problem:** Co-occurrence widely exists in WSSS and intends to inccur false activation.

- **Analysis:** Objects frequently appear together while the limited supervision from image-level labels is incompetent to decouple such dependence.

Motivation

- **Solution:** Cropping images into patches separates the coupled objects and naturally avoids the false bias.

- Challenges remained:

How to generate patches containing only one category?

How to recognize the category at patch level with only image-level labels?
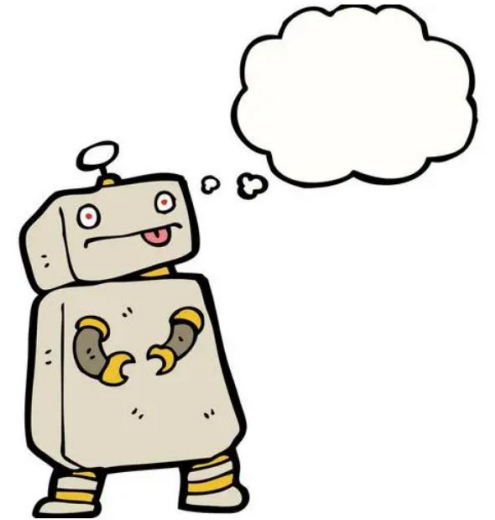
Our approach

- **Separate and Conquer**
- **Essential idea:** Initially separating the co-occurring objects before feature extraction and then enhancing category-specific representations to promote the discrepancy among co-contexts.
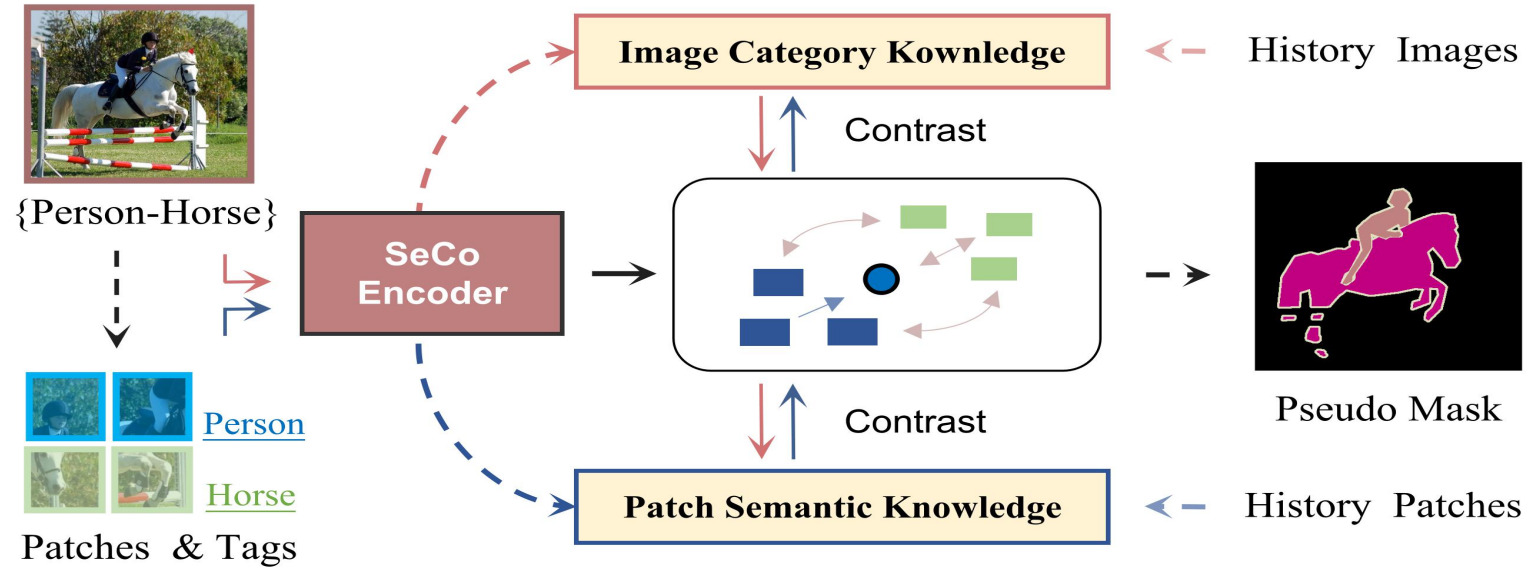
1) **Separate Co-contexts (Image Space)**   2) **Conquer False Activation (Feature Space)**

Framework overview

- Dual-teacher-single-student Architecture to Extract Knowledge
- Tag guided Multi-granularity Contrast to Transfer Knowledge
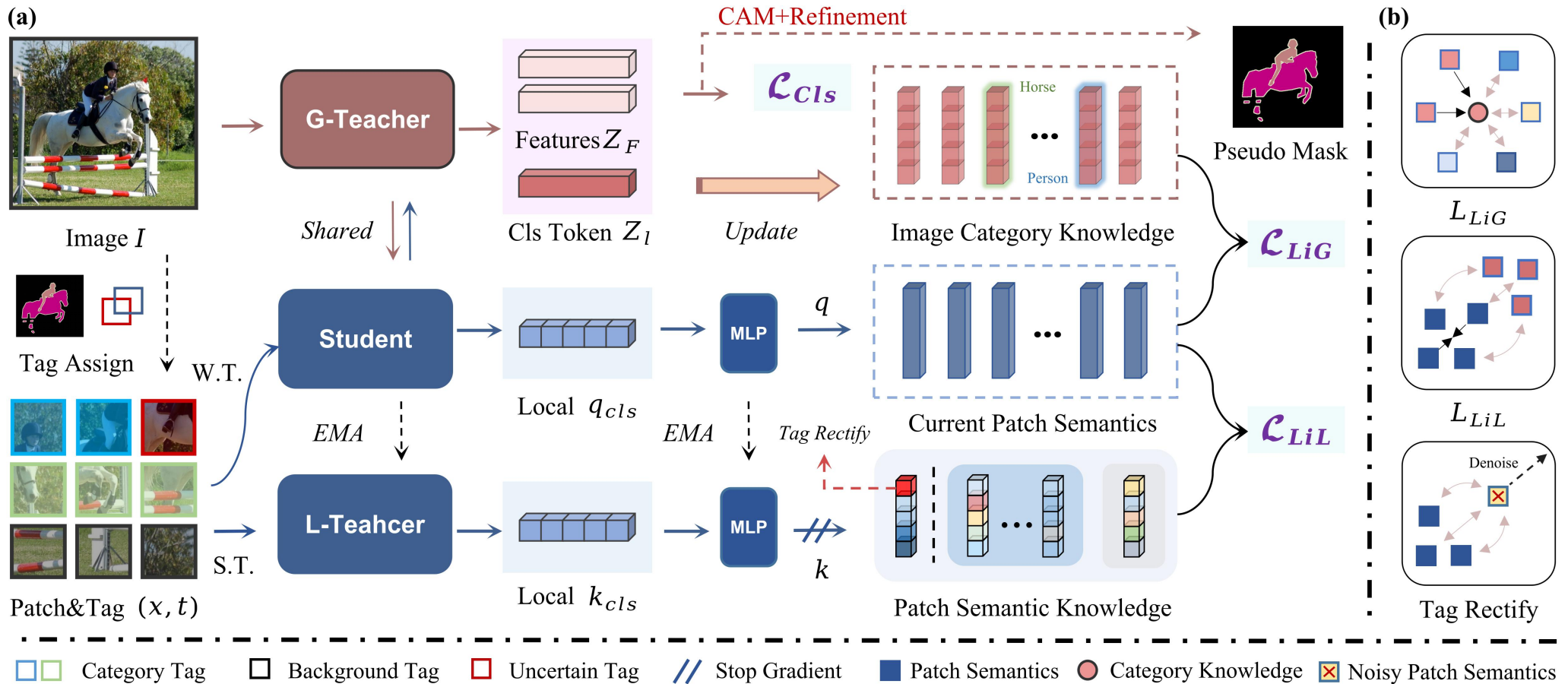
Results

- SeCo can precisely localize the co-categories and generate better CAMs with much fewer false activations.
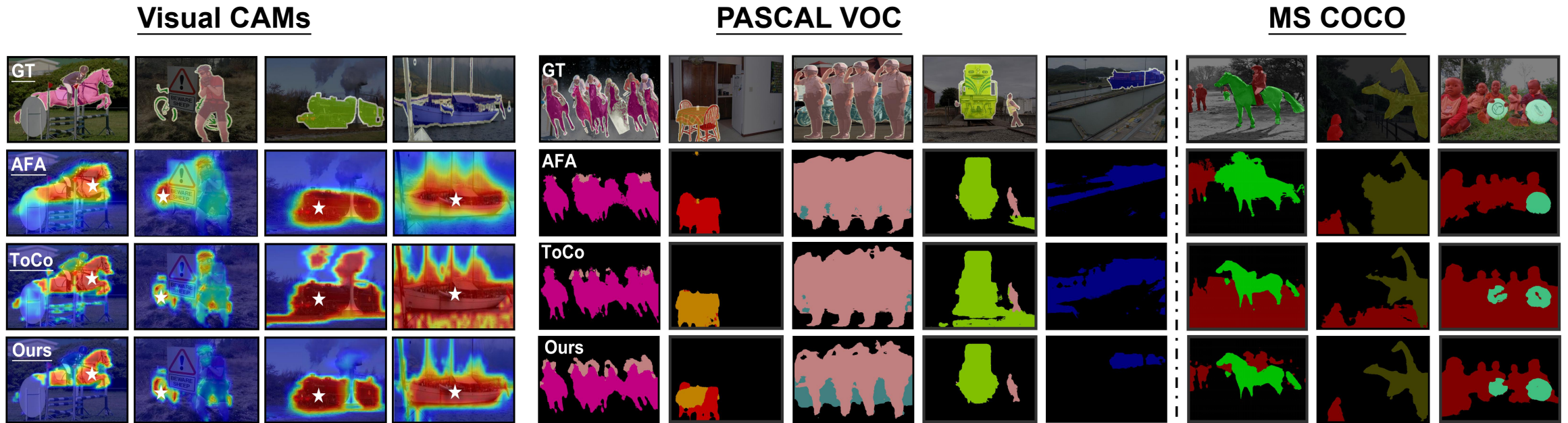
- It improves the semantic performance by reducing the false positives from co-occurrence issue.

AFA [CVPR 2022] ToCo[CVPR 2023]

Conclusion

- **SeCo,** a novel single-stage WSSS scheme to tackle co-occurrence issue without any extra supervision.

- **Tag-guided multi-granularity Contrast**, a new strategy to separate co-occurrence in both image- and feature space.

- State-of-the-art performance over recent works on PASCAL VOC 2012 and MS COCO 2014.
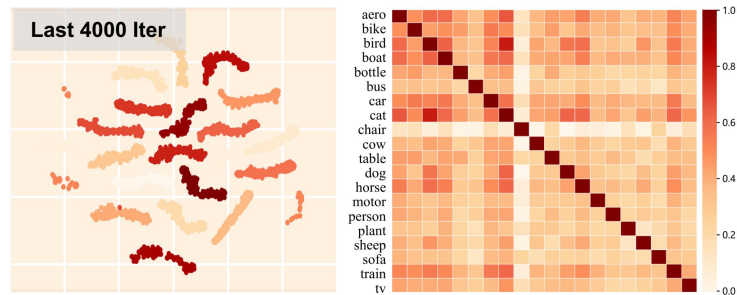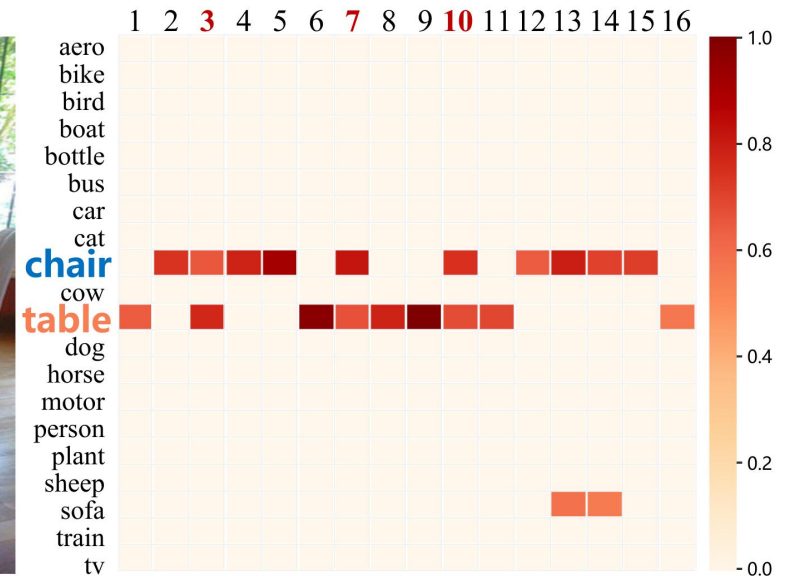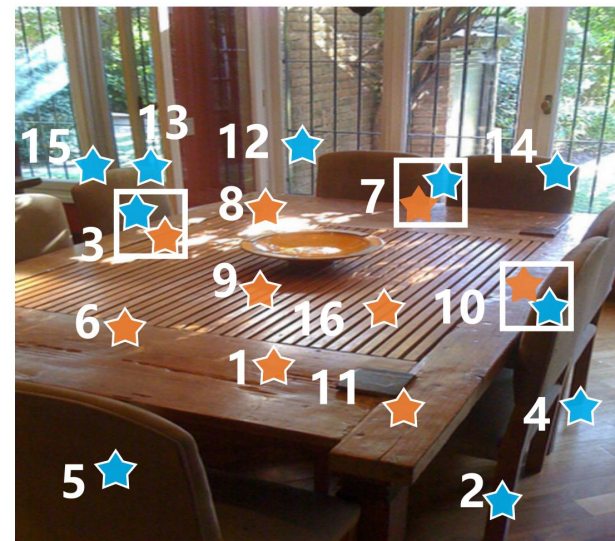

Table 4. Efficiency performance of SeCo compared to others. The experiment is conducted on PASCAL VOC with RTX 3090.

$\mathcal{M}$	CAM	Refine	Decoder	Val	Test
CLIMS [46]	101 mins	332 mins	635 mins	70.4	70.0
$\mathcal{S}$					
AFA [36]		554 mins		66.0	66.3
ToCo [37]		506 mins		71.1	72.2
SeCo(Ours)		**417 mins**		**74.0**	**73.8**

Shanghai Key Lab of Medical Image
Computing and Computer-assisted Intervention

THANK YOU

Zhiwei Yang *   Kexue Fu   Minghong Duan   Linhao Qu,
Shuo Wang†   Zhijian Song†

CODE

Contact: zwyang21@m.fudan.edu.cn