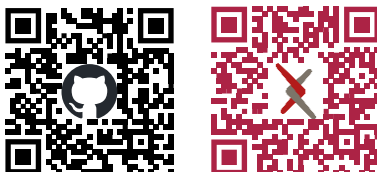


CorrCLIP: Reconstructing Patch Correlations in CLIP for Open-Vocabulary Semantic Segmentation

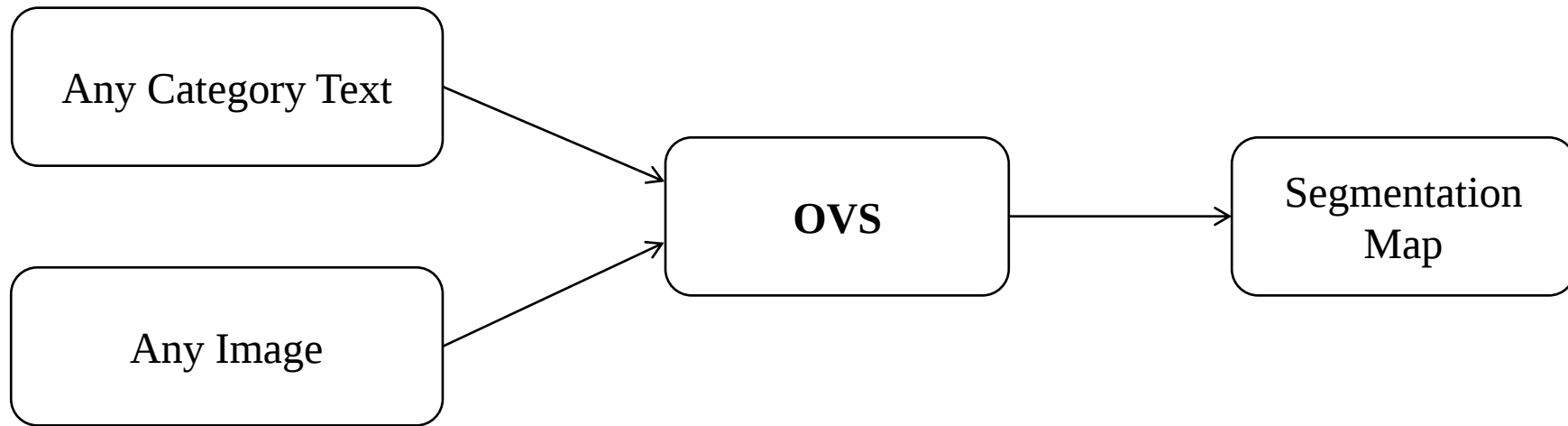
Dengke Zhang¹, Fagui Liu^{1,2}, Quan Tang²

¹South China University of Technology, ²Pengcheng Laboratory



Open-Vocabulary Semantic Segmentation (OVS)

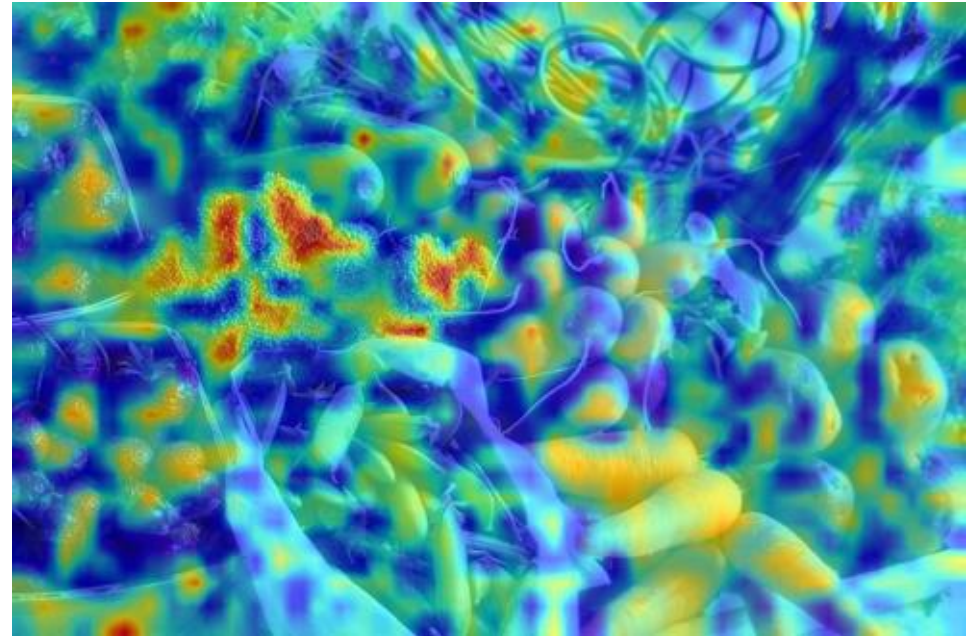
- Segment Arbitrary Categories.
- Built on Vision-Language Models.



Challenge

CLIP (Contrastive Language-Image Pre-Training)

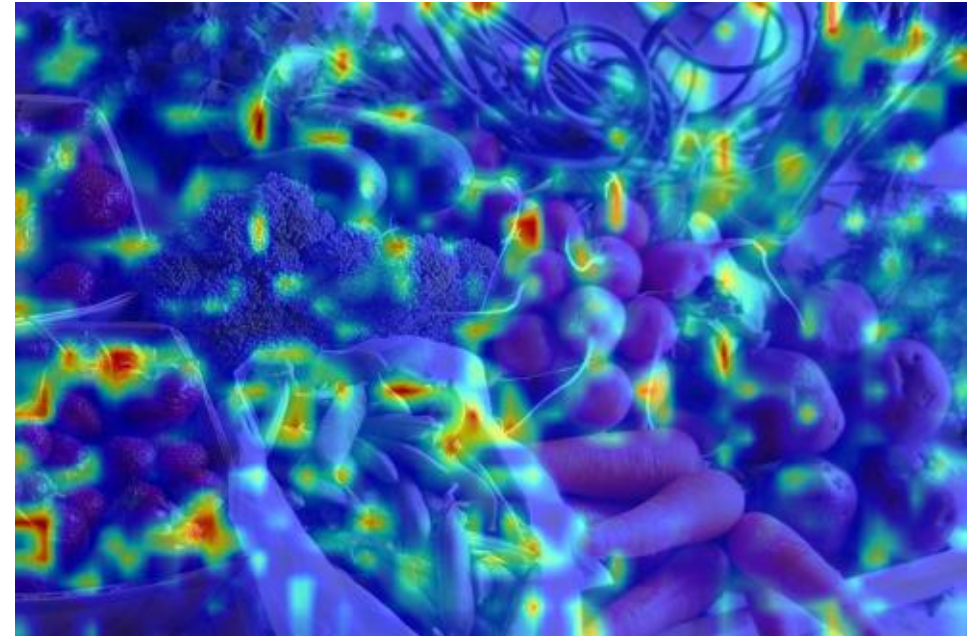
- Zero-shot classification.
- Image-text to patch-text.
- Global semantics.



Motivation

Modifying the attention in the final layer of ViT

- Query-Key $\rightarrow$ Query-Query (Key-Key).



Motivation

Modifying the attention in the final layer of ViT

- Query-Key $\rightarrow$ Query-Query (Key-Key).
- CLIP Features $\rightarrow$ Vision Foundation Model Features.



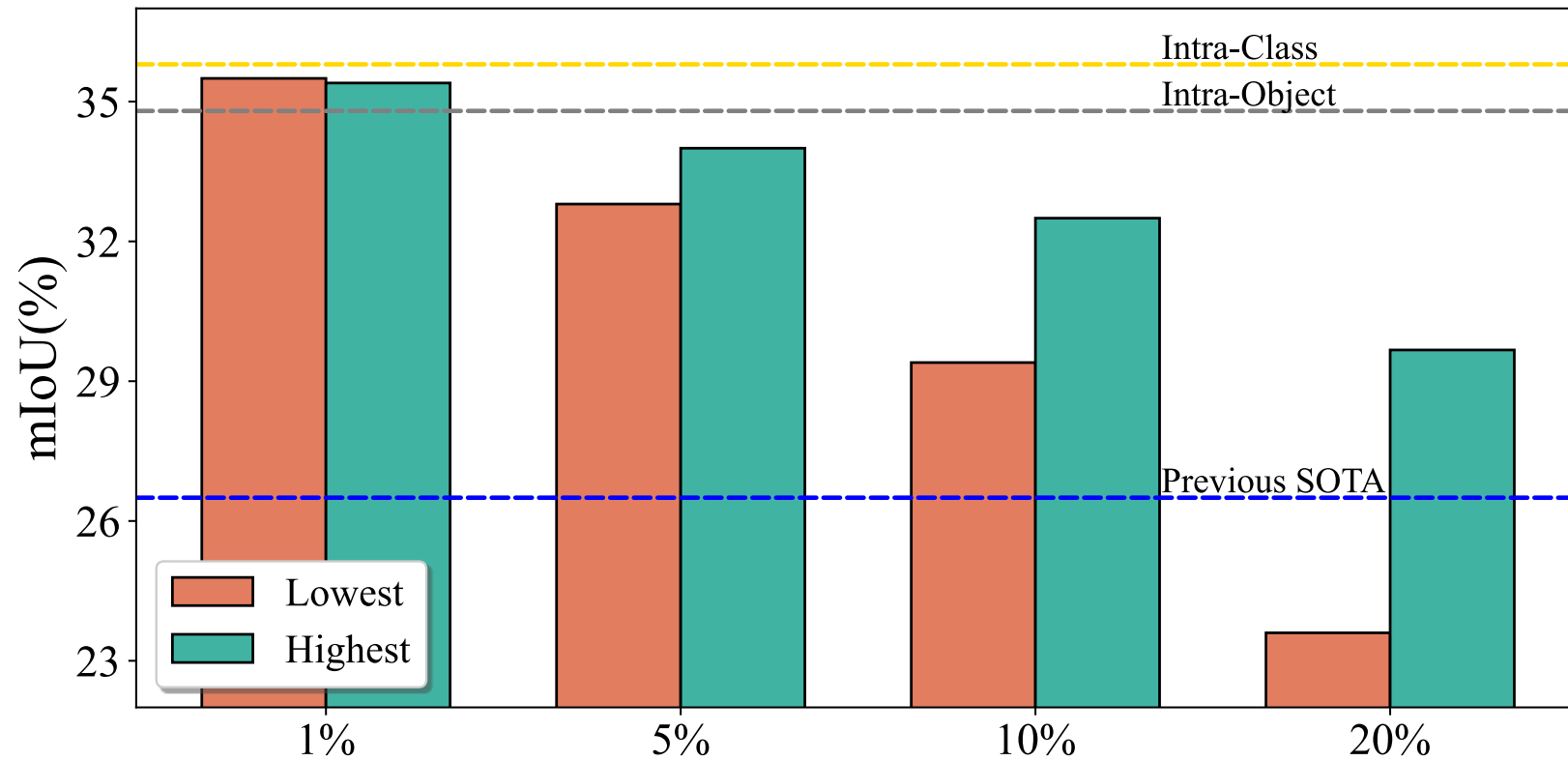
Motivation

Modifying the attention in the final layer of ViT

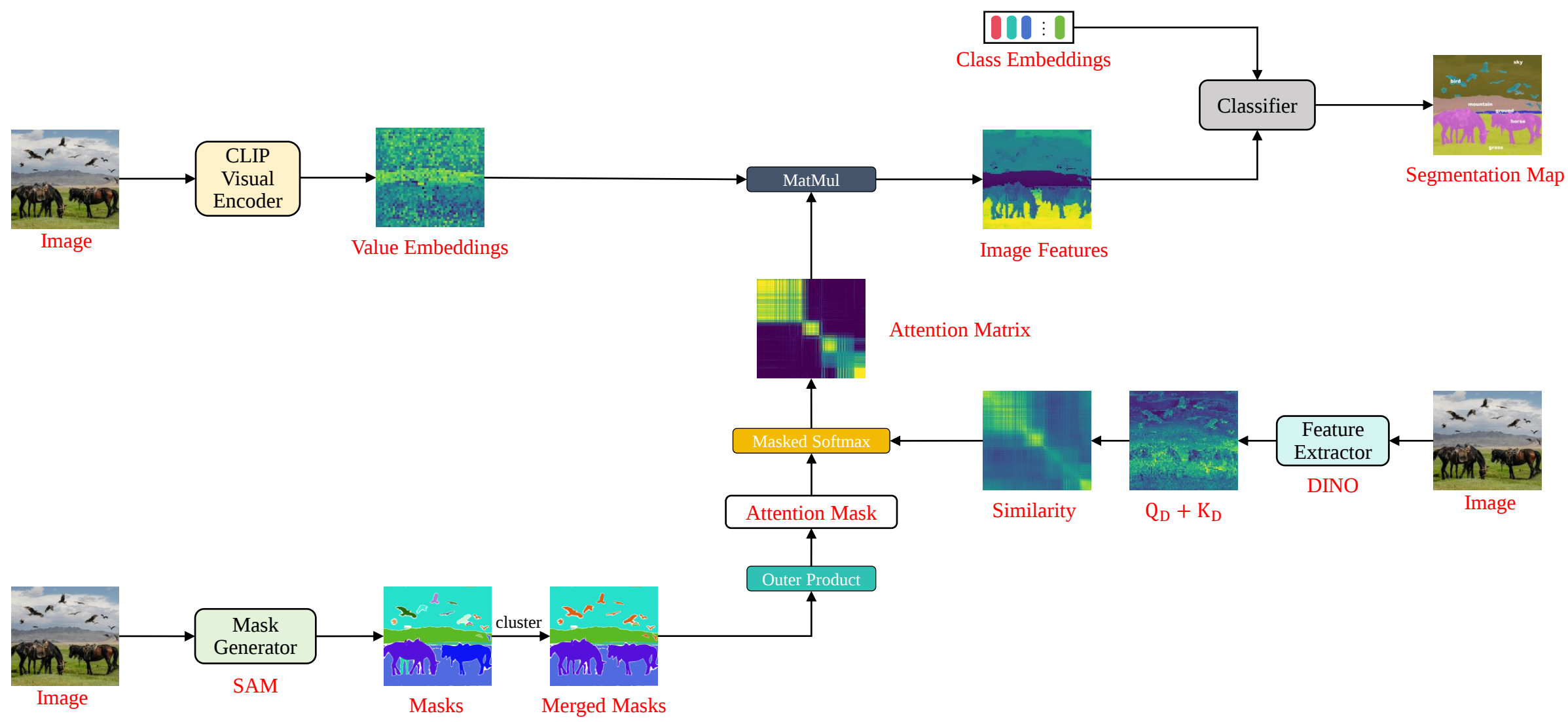
- Query-Key $\rightarrow$ Query-Query (Key-Key).
- CLIP Features $\rightarrow$ Vision Foundation Model Features.
- **What Patch Correlations?**

Motivation

- Intra-class correlations 
- Inter-class correlations 



CorrCLIP: Correlation Reconstruction

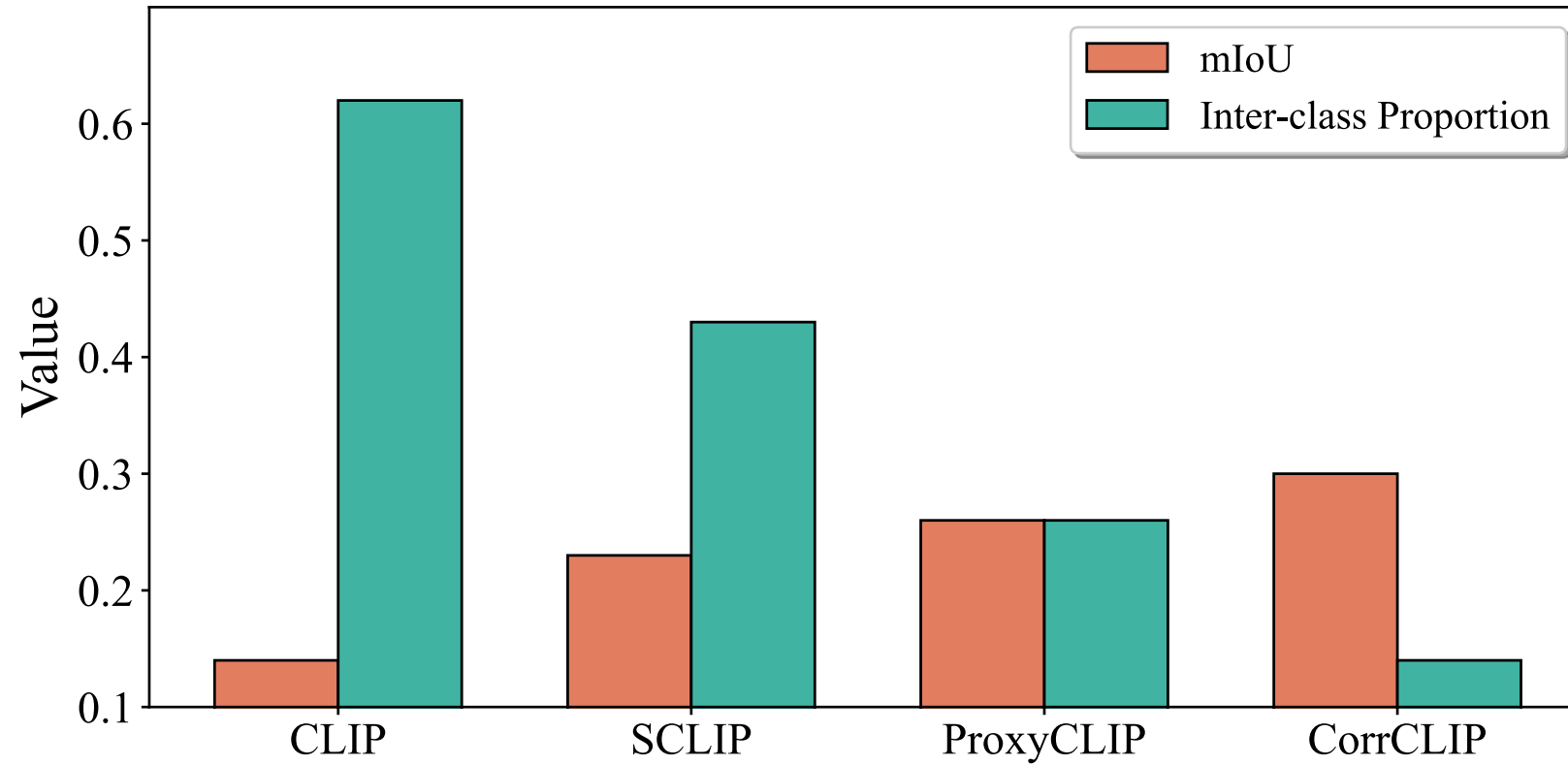


CorrCLIP: Correlation Reconstruction



CorrCLIP: Correlation Reconstruction

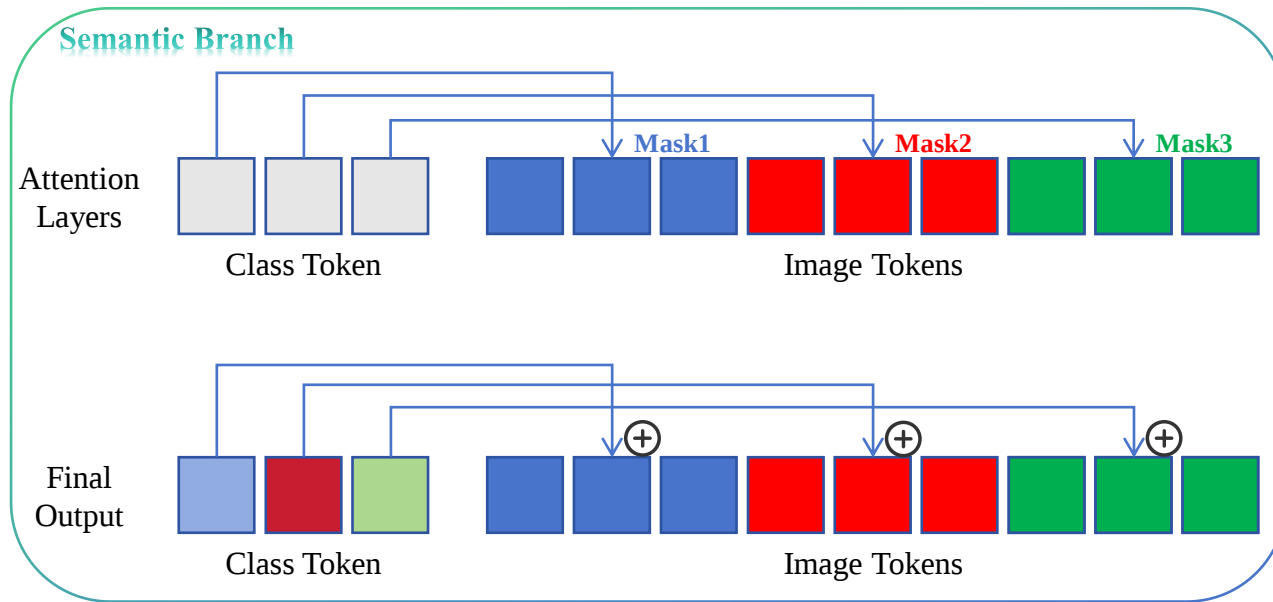
- Reduce inter-class correlations.
- Improve segmentation performance.



CorrCLIP: Feature Refinement

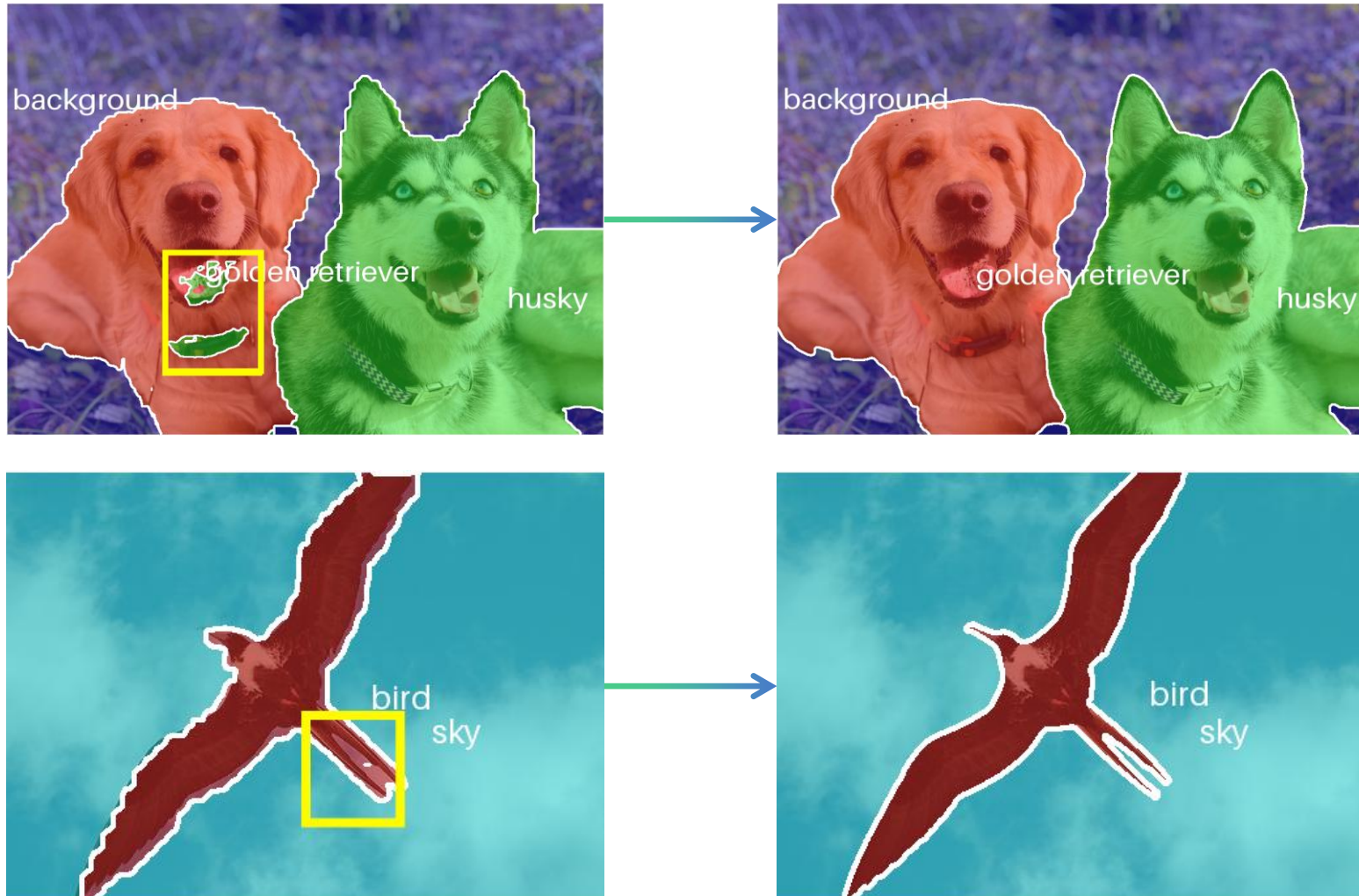
Feature Refinement

- Semantic Branch: Mask class tokens.
- Spatial Branch: Lower layer features.

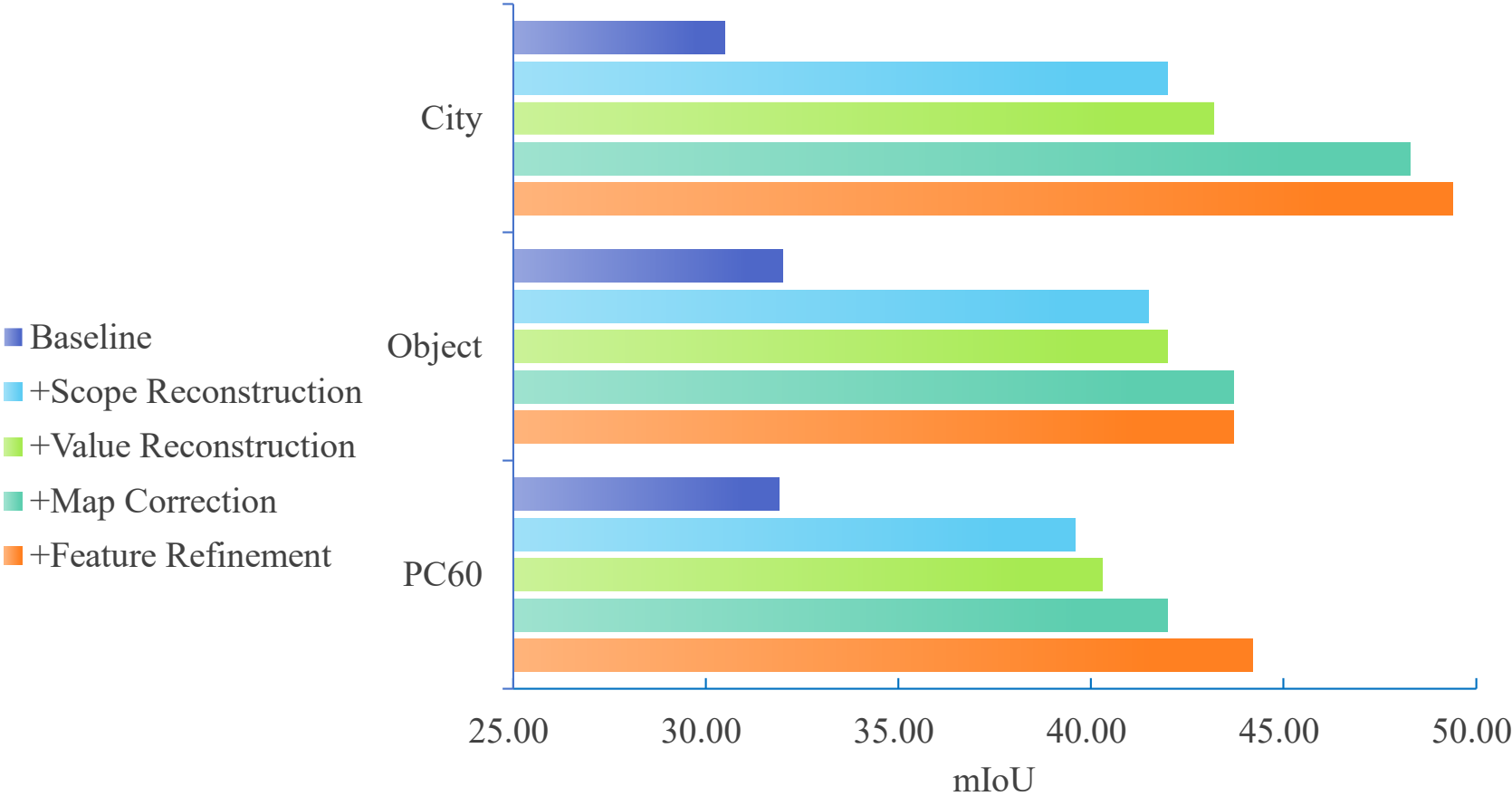


CorrCLIP: Map Correction

Change the category inside the mask to the most frequent one.

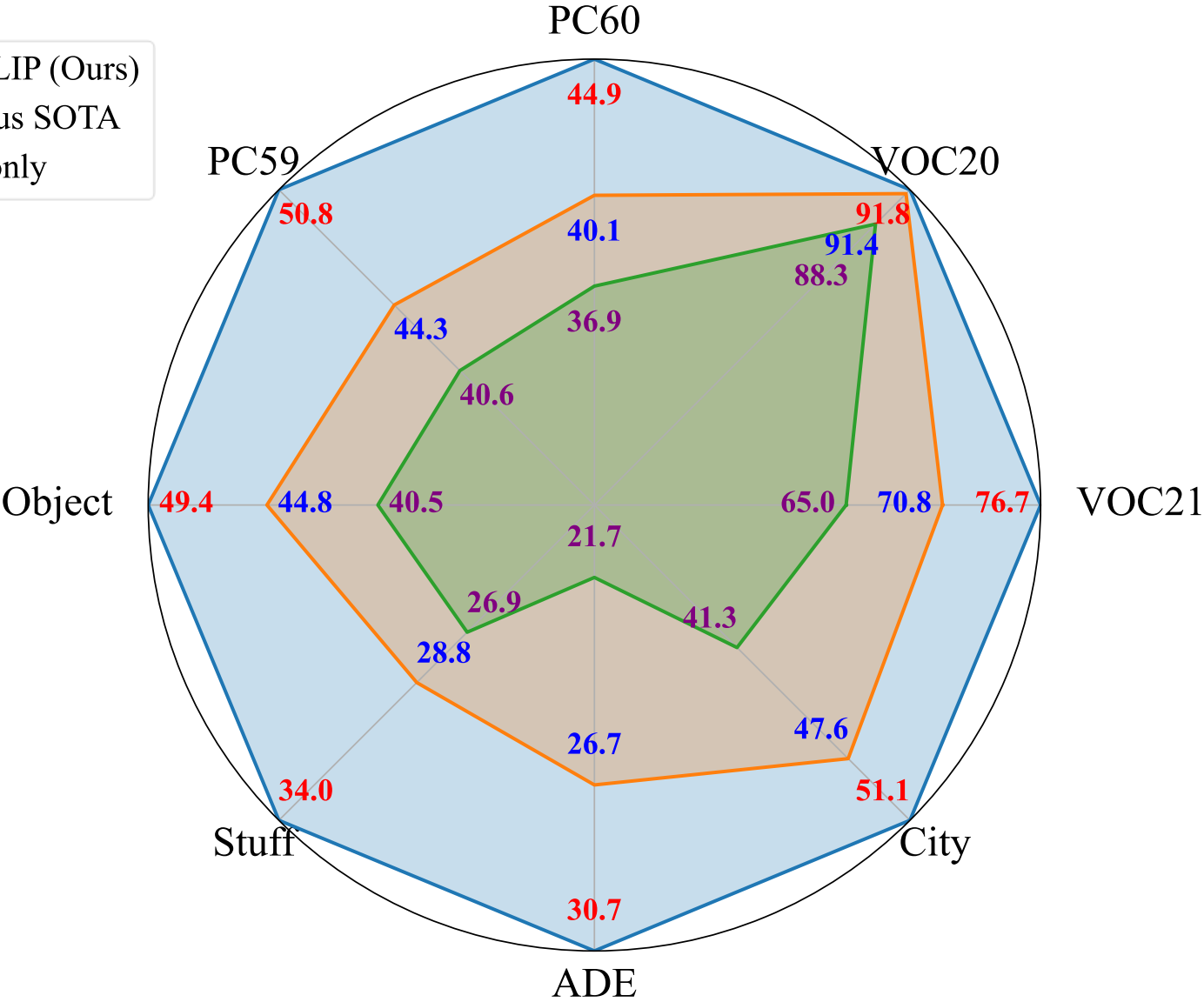
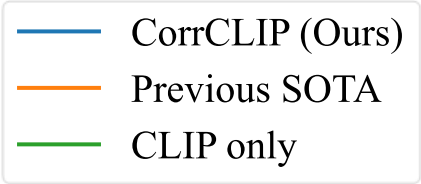


Impact of Each Component

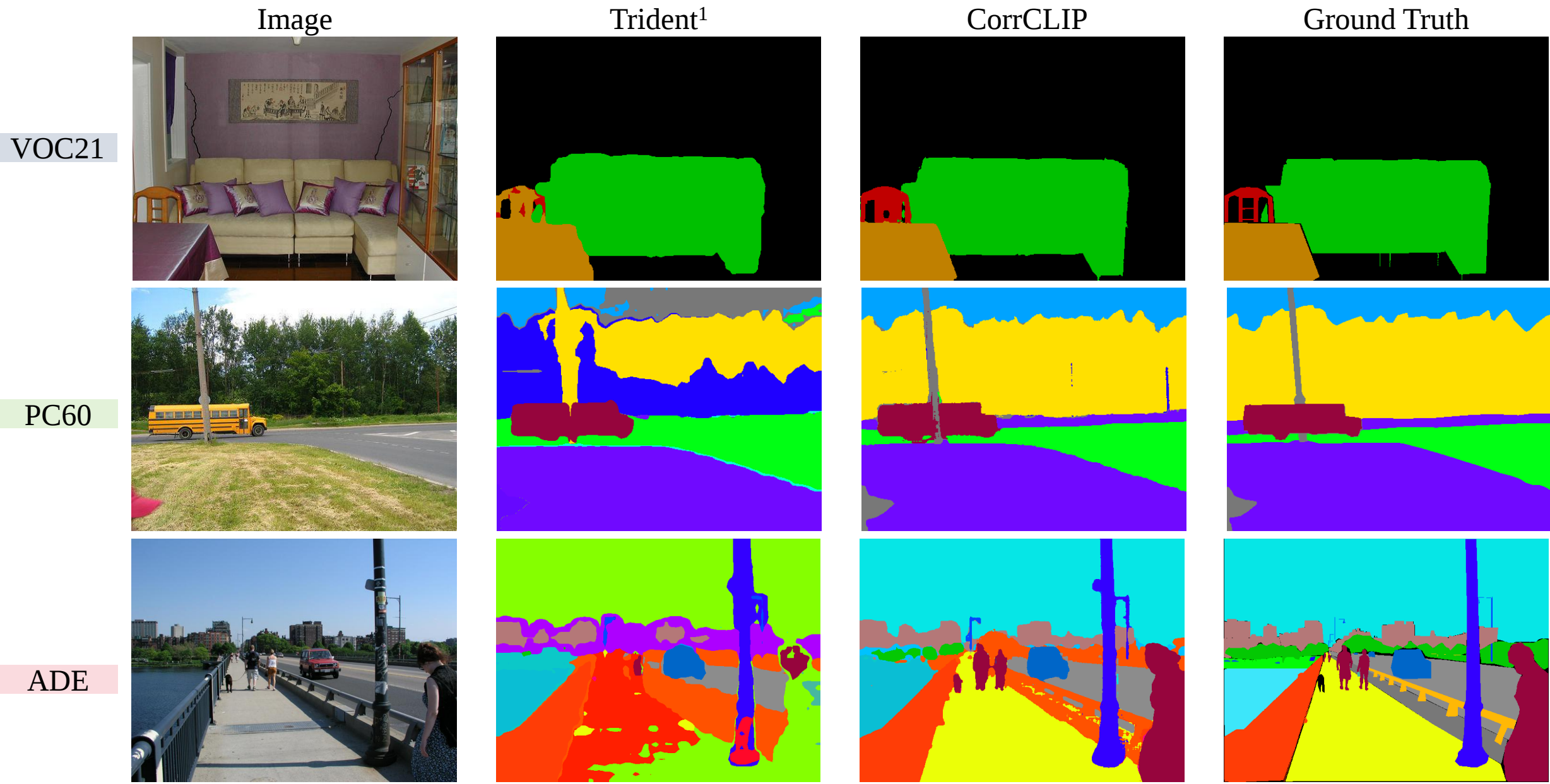


Performance on Standard Benchmarks

Benchmark	Dataset
1. PC60	Pascal Context
2. PC59	
3. Object	COCO 2017
4. Stuff	
5. ADE	ADE20k
6. City	Cityscapes
7. VOC21	Pascal VOC 2012
8. VOC20	



Performance on Standard Benchmarks



[1] Harnessing Vision Foundation Models for High-Performance, Training-Free Open Vocabulary Segmentation (ICCV 2025)

Performance on Out-of-Distribution Benchmarks

Standard benchmarks: person, dog, road, car, chair

Out-of-distribution benchmarks: niche, specific

FoodSeg103



CUB-200



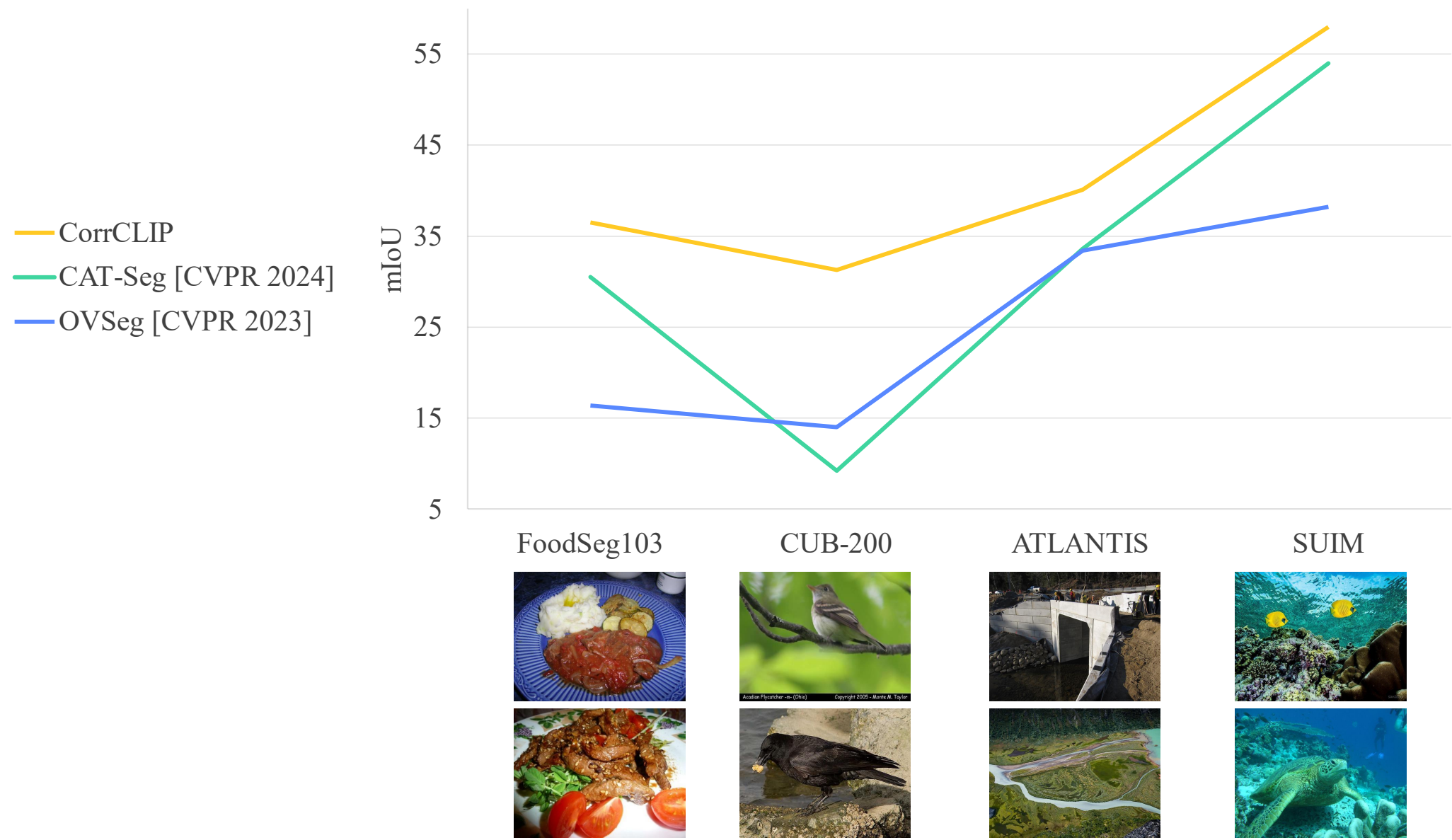
ATLANTIS



SUIM



Performance on Out-of-Distribution Benchmarks



Open-Vocabulary Examples

Diverse Domains



Vast Categories

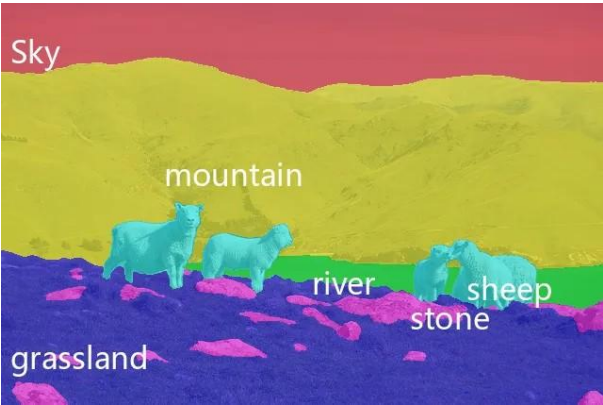


Fine-grained Categories



Multilingual OVS

English



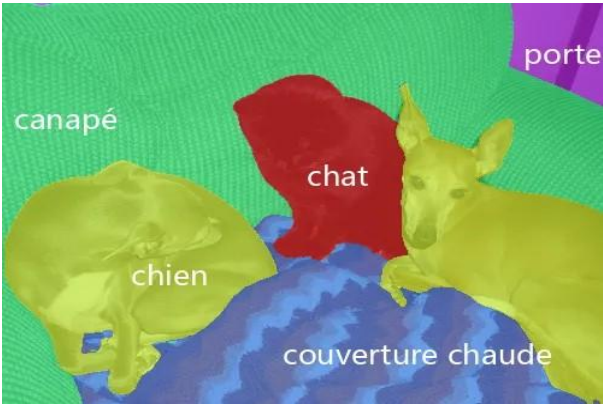
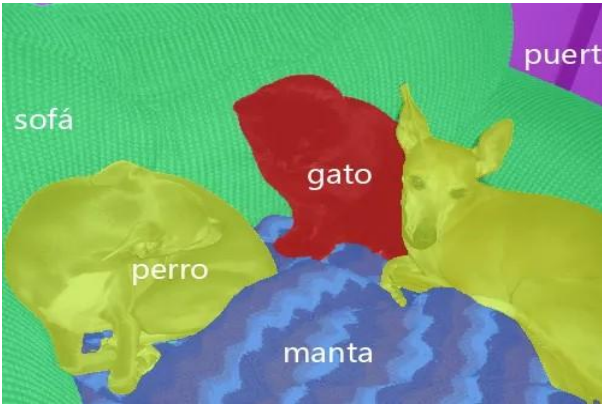
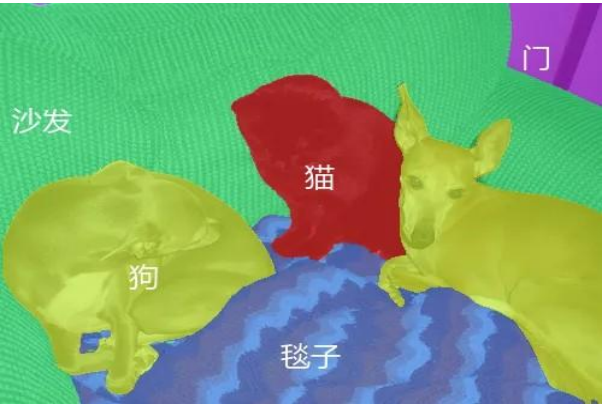
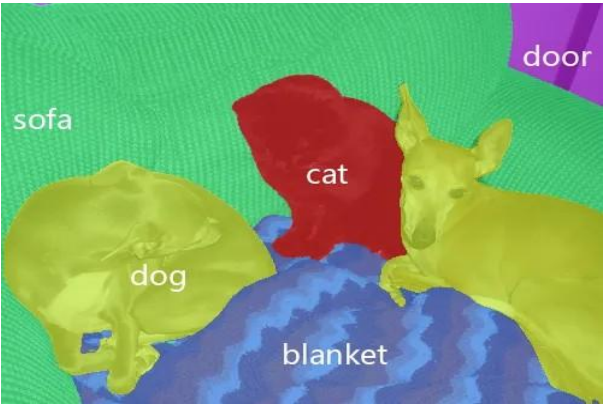
Chinese



Spanish



French



Summary

CorrCLIP: An effective training-free OVS method

- Inter-class correlations impairs segmentation performance.
- Scope and value reconstruction.
- Feature refinement and map correction.
- Superior segmentation performance.

Thank you

