工源三仟

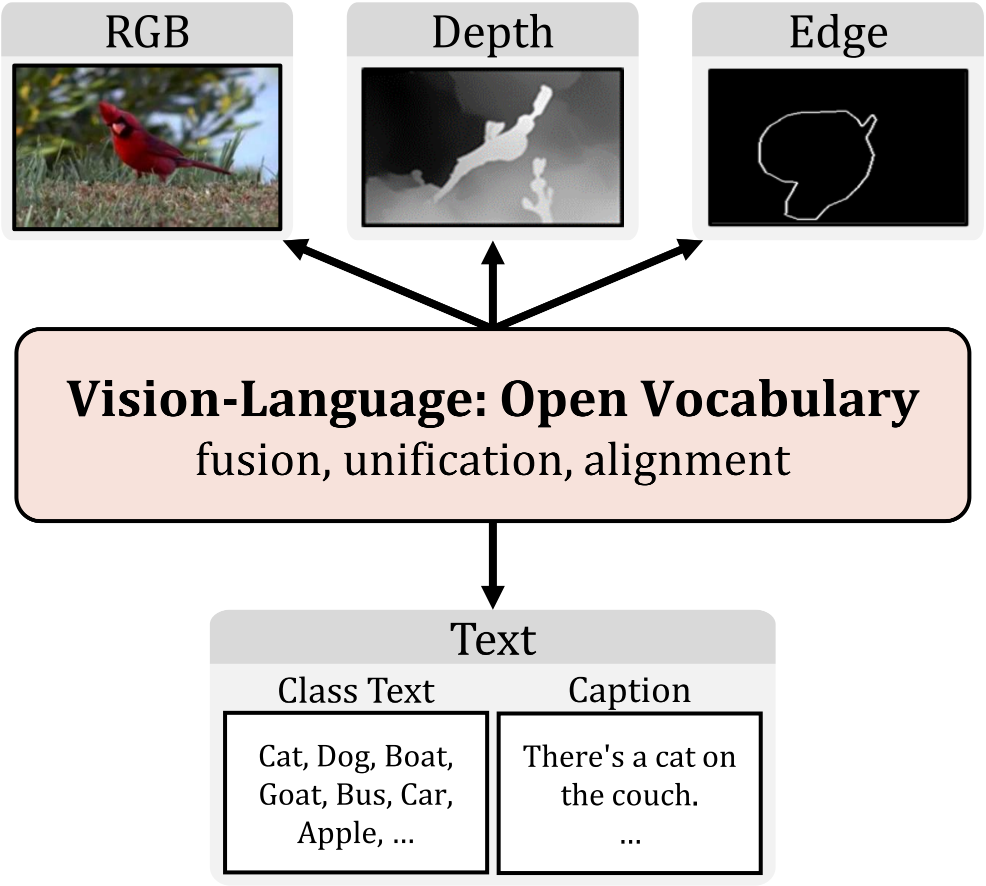
Open-Vocabulary Camouflaged Object Segmentation

Youwei Pang[1,2], Xiaoqi Zhao[1,2], Jiaming Zuo[2], Lihe Zhang[1], Huchuan Lu[1]
[1]Dalian University of Technology [2]X3000 Inspection Co., Ltd

ECCV
EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO
2024

Vision-Language: Open Vocabulary

- ➢ **Multi-Sensor/View: RGB + Depth + Edge**
- ➢ **Language: Class Text**



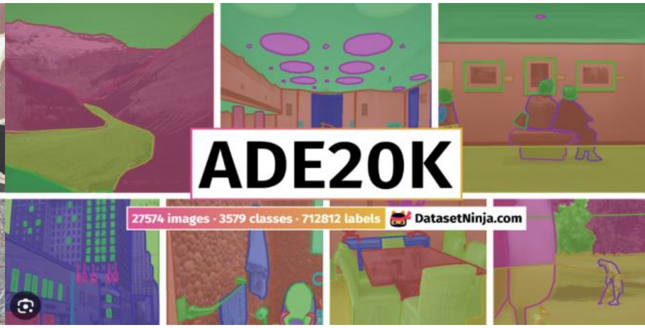
VCOS

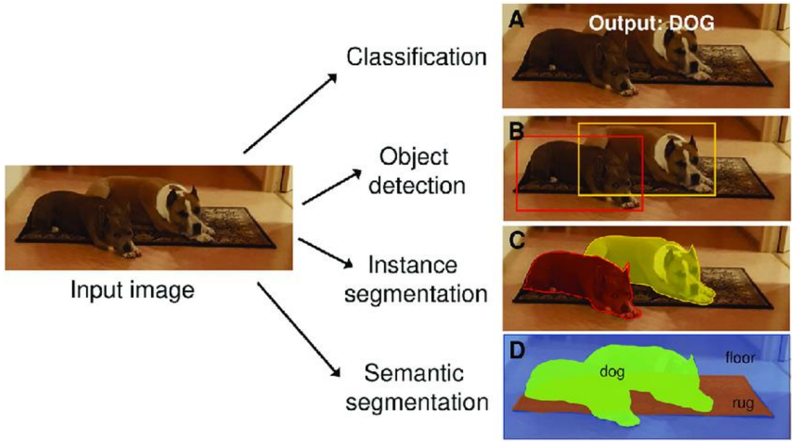
- ☐ RGB + Depth + Edge + Text
- ☐ New Challenge: OVCOS
- ☐ New Benchmark: OVCamo
- ☐ Strong Baseline: OVCoser

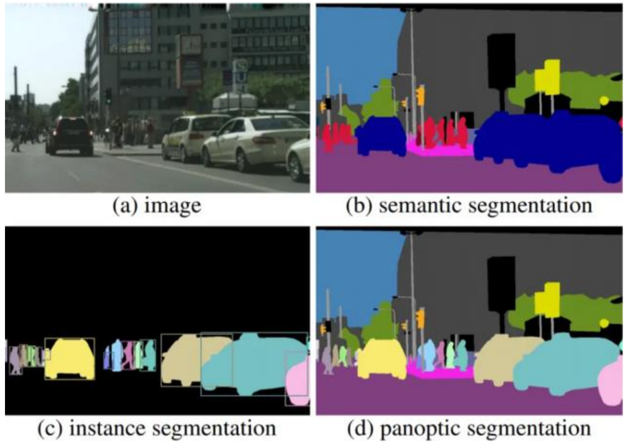
Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024

Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

Data Requirements

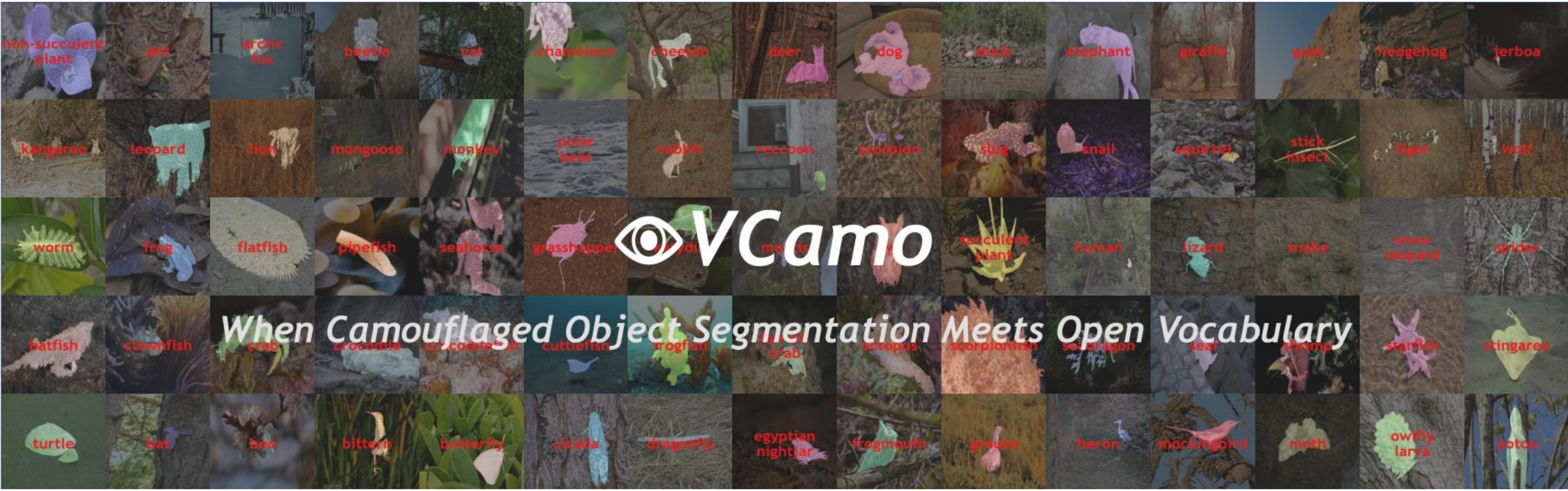






- Existing open-vocabulary methods focus only on normal scenes.
- Based on publicly datasets which are not tailored for open vocabulary.
- Rarely involve imperceptible objects camouflaged in complex scenes.
- Lack of exploration due to data collection bias and annotation costs.

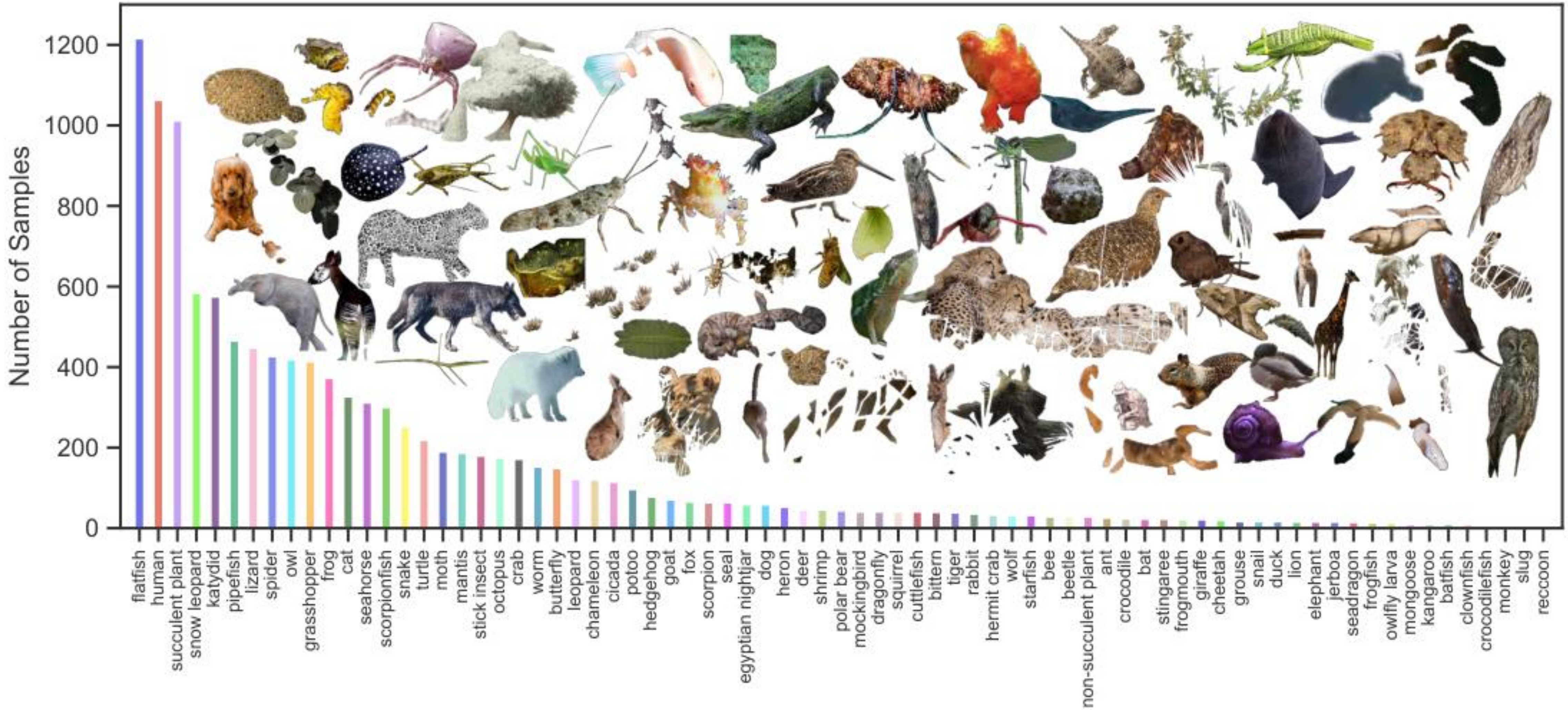
Main Contributions



- **New Challenge.** In view of the limitations of the existing OVSIS, we introduce a more challenging OVCOS task for open-vocabulary segmentation of camouflaged objects.
- **New Benchmark.** A new large-scale benchmark OVCamo with diverse samples carefully collected from existing publicly available data is proposed to better evaluate and analyze algorithms.
- **Strong Baseline.** A robust single-stage baseline is equipped with iterative semantic guidance and structure enhancement and benefits from the joint optimization of multi-source information.

Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024
Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

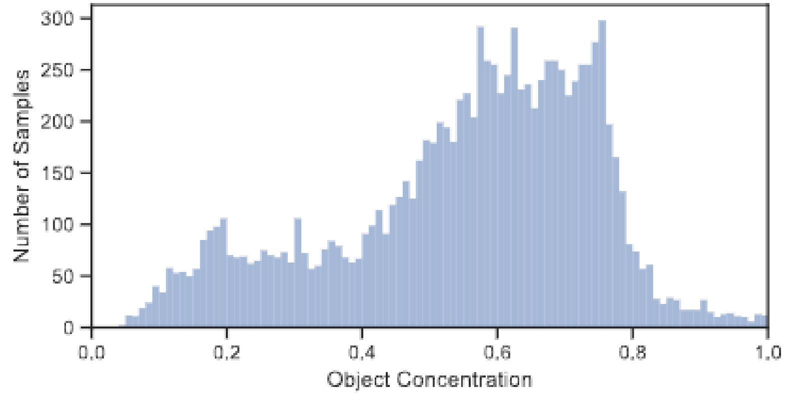
New Benchmark—OVCamo: Class Distribution



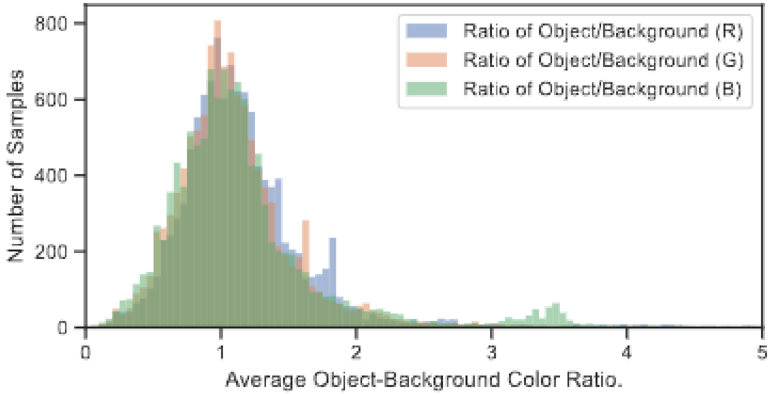
OVCamo: A large-scale complex scene dataset containing 11483 hand-selected images with fine annotations and corresponding 75 object classes.

Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024
Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

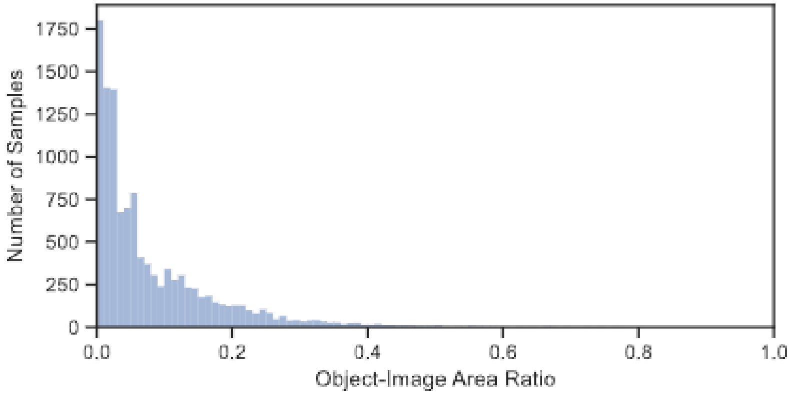
New Benchmark—OVCamo: Data Attributes



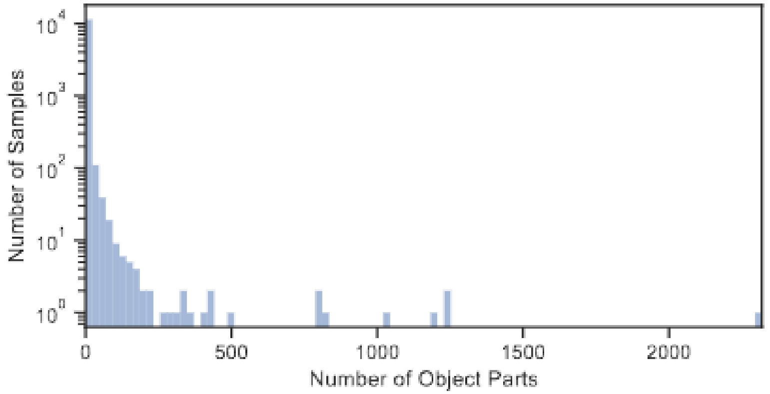
(a) Object Concentration.



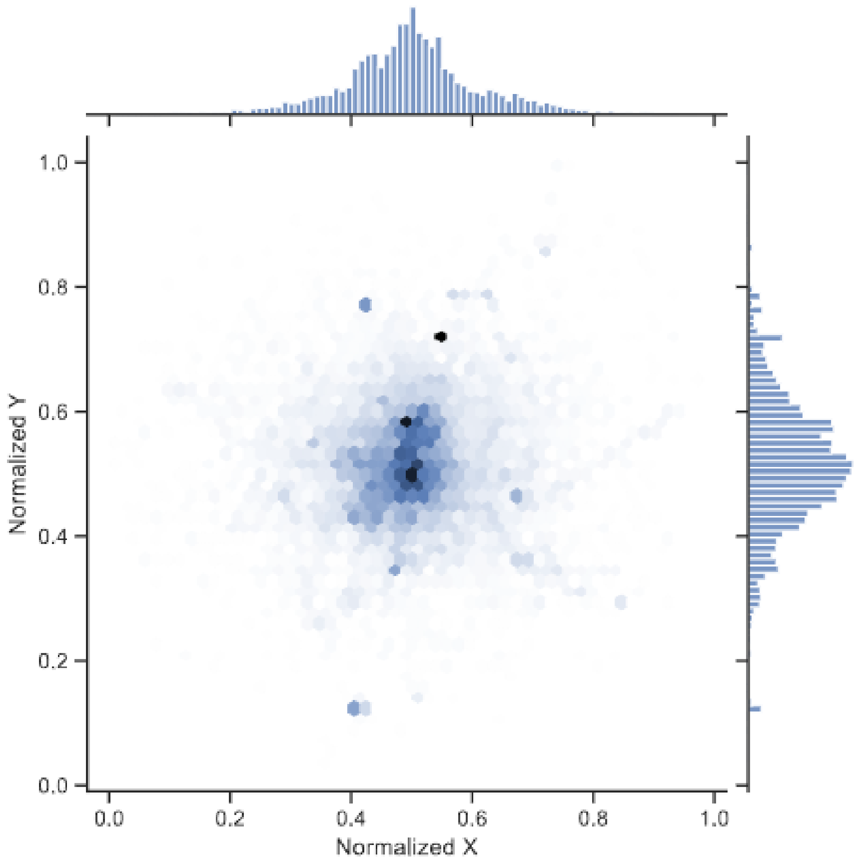
(b) Average Color Ratio.



(c) Object-Image Area Ratio.



(d) Number of Object Parts.

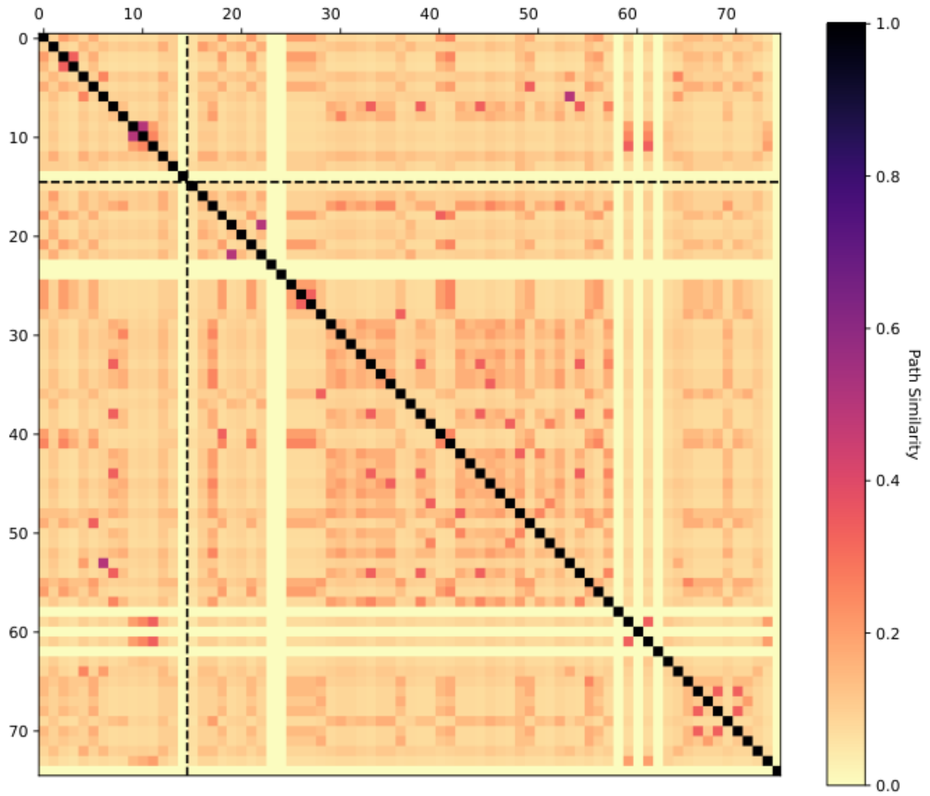


(e) Normalized Centroid.

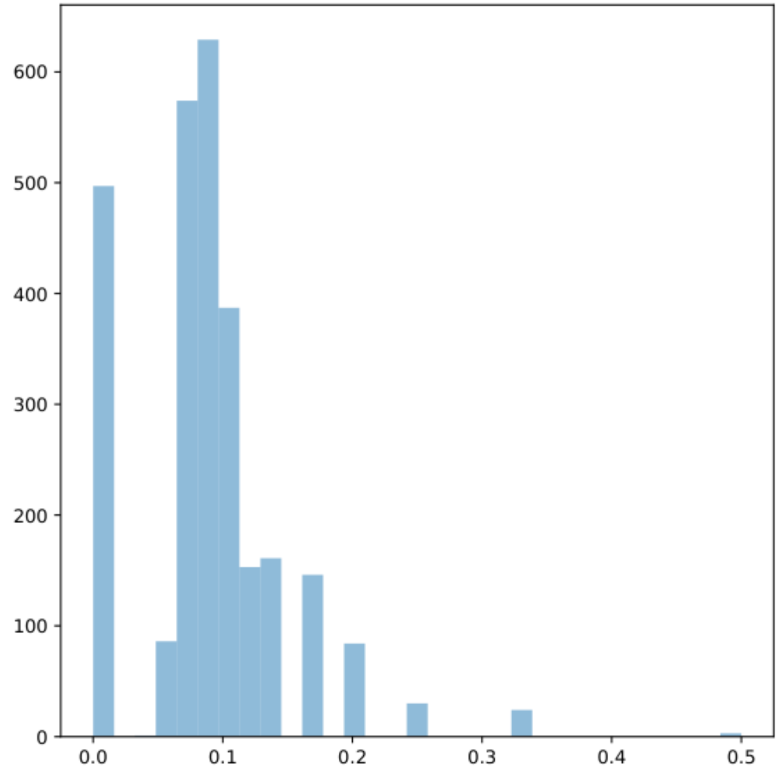
The camouflaged objects of interest usually have ***complex shape***, ***high similarity to the background***, ***small size***, ***multiple camouflaged objects or sub-regions***, and ***central biases***.

Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024
Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

New Benchmark—OVCamo: Semantic Similarity



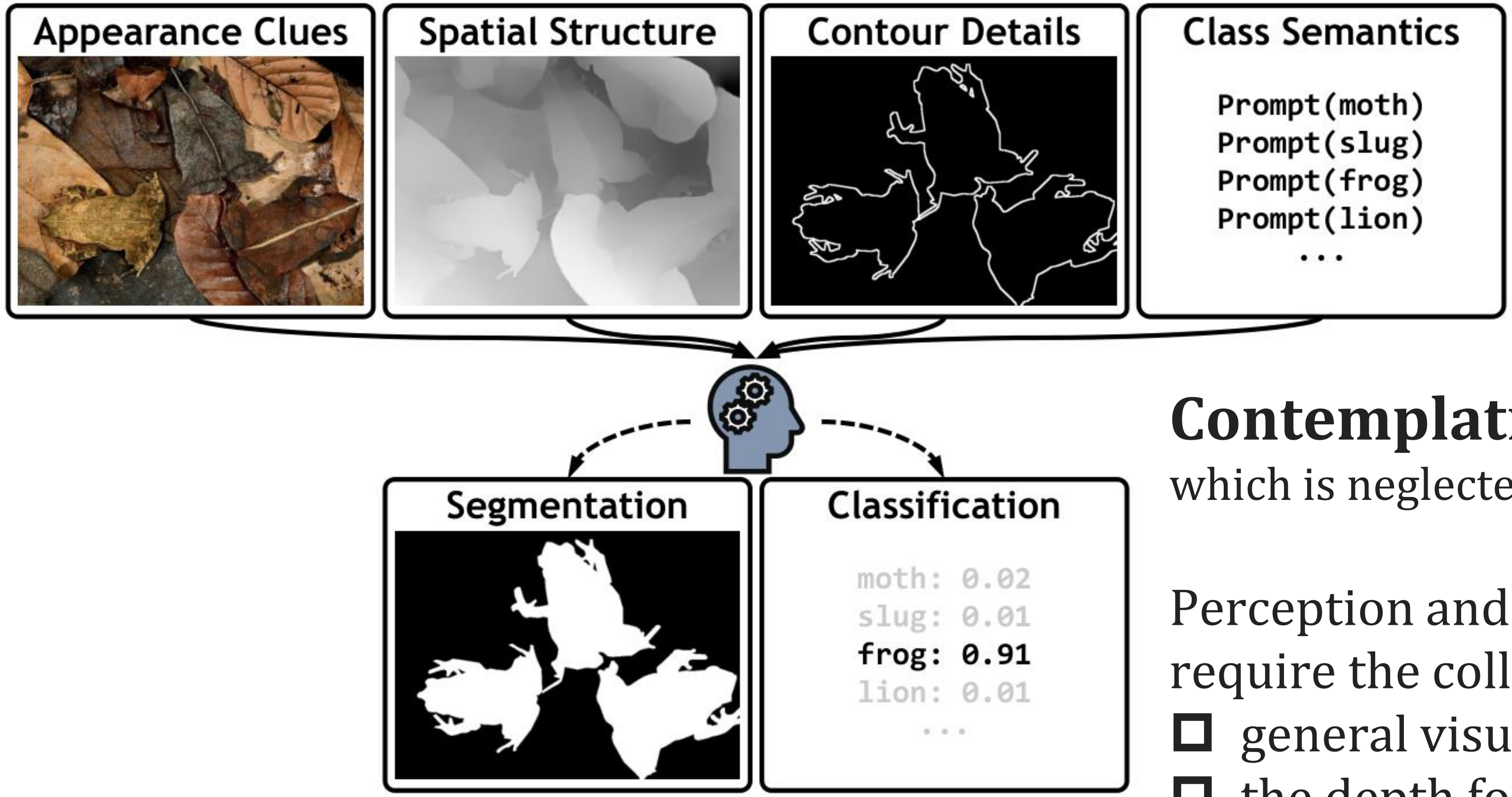
(a) Semantic similarity score map.



(b) Semantic similarity score histogram.

- Class semantic similarity of OVCamo based on the Open English WordNet.
- Class semantic similarity in our class set is very low, which can better alleviate the complexity due to class semantic similarity during open vocabulary evaluation.

Strong Baseline—OVCoser: Task-Inspired Design



Contemplation on OVCOS

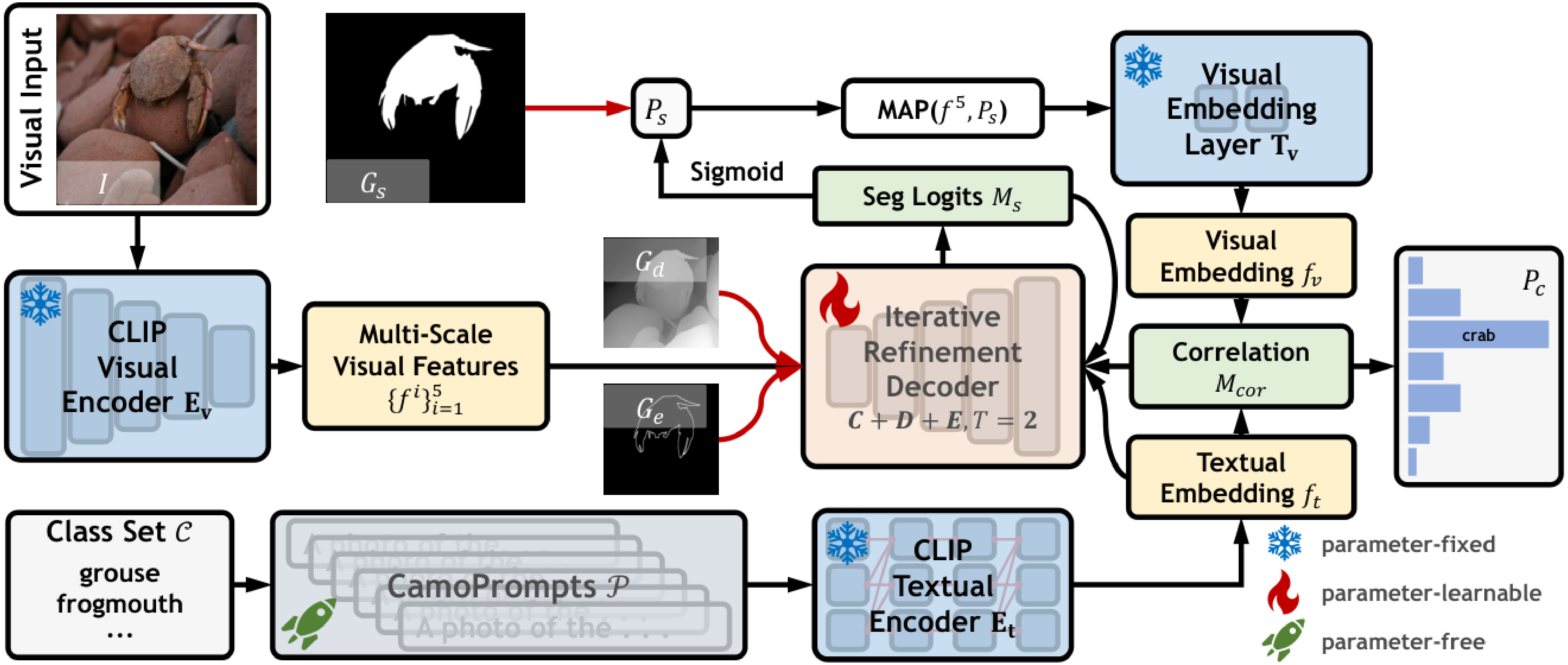
which is neglected by existing methods.

Perception and recognition of camouflaged objects require the collaboration of multi-source information:

- ☐ general visual appearance cues,
- ☐ the depth for the spatial structure of the scene,
- ☐ the edge for the regional changes about objects,
- ☐ the text for the context-aware class semantics.

Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024
Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

Strong Baseline—OVCoser: Overall Architecture



- Richer information: semantic guidance and structural enhancements.
- More applicable architecture: iterative refinement in decoding.
- More targeted prompt design: CamoPrompts.
- More efficient architecture: built on the frozen CLIP.

Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024

Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

Strong Baseline—OVCoser: Experiments

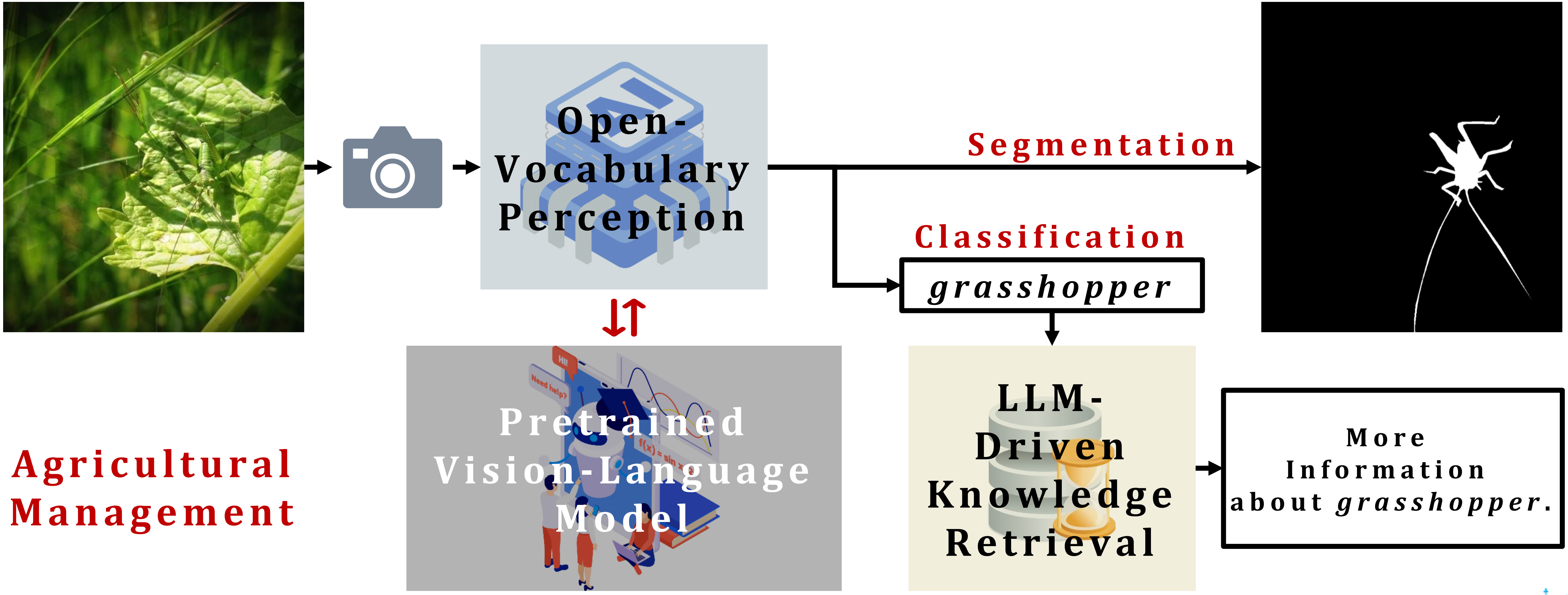
Model	VLM	Feature Backbone	Text Prompt	$cS_m \uparrow$	$cF^{\omega}_{\beta} \uparrow$	cMAE $\downarrow$	$cF_{\beta} \uparrow$	$cE_m \uparrow$	cIoU $\uparrow$
Test on OVCamo with the weight trained on COCO.									
SimSeg^{21} [49]	CLIP-ViT-B/16 [37]	ResNet-101 [20]	Learnable [60]	0.128	0.105	0.838	0.112	0.143	0.094
OVSeg^{22} [27]	CLIP-ViT-L/14 [37]	Swin-B [30]	[18]	0.341	0.306	0.584	0.325	0.384	0.273
ODISE^{23} [47]	CLIP-ViT-L/14 [37]	StableDiffusionv1.3 [40]	[17]	0.409	0.339	0.500	0.341	0.421	0.302
SAN^{23} [48]	CLIP-ViT-L/14 [37]	ViT Adapter	[18]	0.414	0.343	0.489	0.357	0.456	0.319
CAT-Seg^{23} [10]	CLIP-ViT-L/14 [37]	Swin-B [30]	[37]	0.430	0.344	0.448	0.366	0.459	0.310
FC-CLIP^{23} [52]	CLIP-ConvNeXt-L [9]	—	[18]	0.374	0.306	0.539	0.320	0.409	0.285
Finetune on OVCamo with the weight trained on COCO.									
SimSeg^{21} [49]	CLIP-ViT-B/16 [37]	ResNet-101 [20]	Learnable [60]	0.098	0.071	0.852	0.081	0.128	0.066
OVSeg^{22} [27]	CLIP-ViT-L/14 [37]	Swin-B [30]	[18]	0.164	0.131	0.763	0.147	0.208	0.123
ODISE^{23} [47]	CLIP-ViT-L/14 [37]	StableDiffusionv1.3 [40]	[17]	0.182	0.125	0.691	0.219	0.309	0.189
SAN^{23} [48]	CLIP-ViT-L/14 [37]	ViT Adapter	[18]	0.321	0.216	0.550	0.236	0.331	0.204
CAT-Seg^{23} [10]	CLIP-ViT-L/14 [37]	Swin-B [30]	[37]	0.185	0.094	0.702	0.110	0.185	0.088
FC-CLIP^{23} [52]	CLIP-ConvNeXt-L [9]	—	[18]	0.124	0.074	0.798	0.088	0.162	0.072
Train on OVCamo.									
SimSeg^{21} [49]	CLIP-ViT-B/16 [37]	ResNet-101 [20]	Learnable [60]	0.053	0.049	0.921	0.056	0.098	0.047
OVSeg^{22} [27]	CLIP-ViT-L/14 [37]	Swin-B [30]	[18]	0.024	0.046	0.954	0.056	0.130	0.046
ODISE^{23} [47]	CLIP-ViT-L/14 [37]	StableDiffusionv1.3 [40]	[17]	0.187	0.119	0.700	0.211	0.298	0.167
SAN^{23} [48]	CLIP-ViT-L/14 [37]	ViT Adapter	[18]	0.275	0.202	0.612	0.220	0.318	0.189
CAT-Seg^{23} [10]	CLIP-ViT-L/14 [37]	Swin-B [30]	[37]	0.181	0.106	0.719	0.123	0.196	0.094
FC-CLIP^{23} [52]	CLIP-ConvNeXt-L [9]	—	[18]	0.080	0.076	0.872	0.090	0.191	0.072
Ours	CLIP-ConvNeXt-L [9]	—	CamoPrompts	0.579	0.490	0.337	0.520	0.615	0.443

Model	Trainable Param.	Total Param.	FLOPs
SimSeg^{21} [49]	61M (28.91%)	211M	1.9T
OVSeg^{22} [27]	531M (100.00%)	531M	8.0T
ODISE^{23} [47]	28M (1.80%)	1522M	5.5T
SAN^{23} [48]	9M (2.06%)	437M	0.4T
CAT-Seg^{23} [10]	104M (21.22%)	490M	0.3T
FC-CLIP^{23} [52]	20M (5.38%)	372M	0.8T
Ours	7M (1.95%)	359M	0.2T

Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024
Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

Potential Applications

- ➢ Species Identification
- ➢ Medical Image Analysis
- ➢ Agricultural Management



Open-Vocabulary Camouflaged Object Segmentation, ECCV 2024
Youwei Pang, Xiaoqi Zhao*, Jiaming Zuo, Lihe Zhang, Huchuan Lu*

Thanks!

工源三仟

ECCV

EUROPEAN CONFERENCE ON COMPUTER VISION