UNICL-SAM: Uncertainty-Driven In-Context Segmentation with Part Prototype Discovery

Dianmo Sheng[1], Dongdong Chen[2], Zhentao Tan[1], Qiankun Liu[3], Qi Chu[1], Tao Gong[1*], Bin Liu[1], Jing Han[4], Wenbin Tu[4], Shengwei Xu[5], Nenghai Yu[1]

CVPR Nashville JUNE 11-15, 2025

Motivation

Simulated real-world degradation support examples ***(Our robustness simulation tests).***

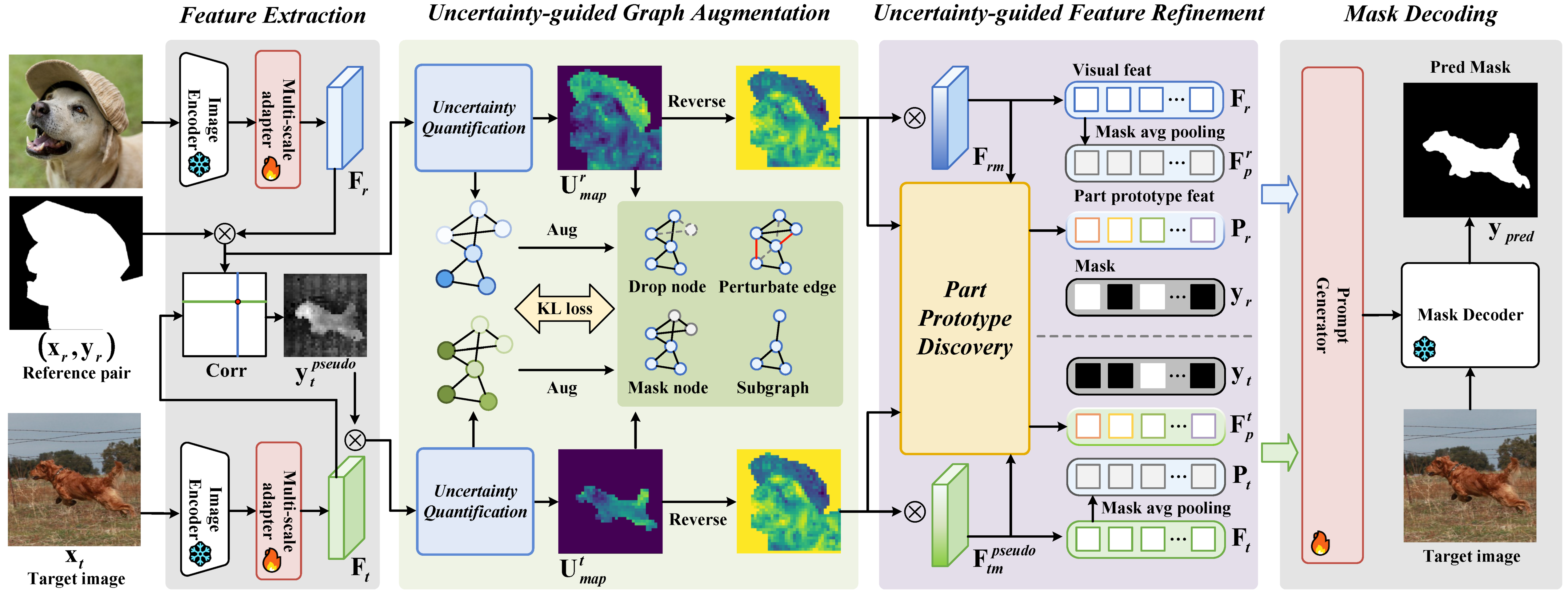
Real-world applications present challenges due to the variability of support examples, which often exhibit quality issues resulting from various sources and inaccurate labeling. Existing in-context segmentation methods will experience significant performance degradation when encountering such samples.

How to extract more robust representations from these examples has always been one of the goals of in-context visual learning.

Contributions

- We propose a ***probabilistic uncertainty module*** to model input data distributions, further using uncertainty-guided optimization strategies to enhance feature reliability.
- We introduce ***part prototypes*** to effectively extract knowledge from reference examples by aggregating local semantics.
- UNICL-SAM performs superior performance on traditional benchmarks and shows ***good generalization capabilities*** on conducted robustness simulation tests.

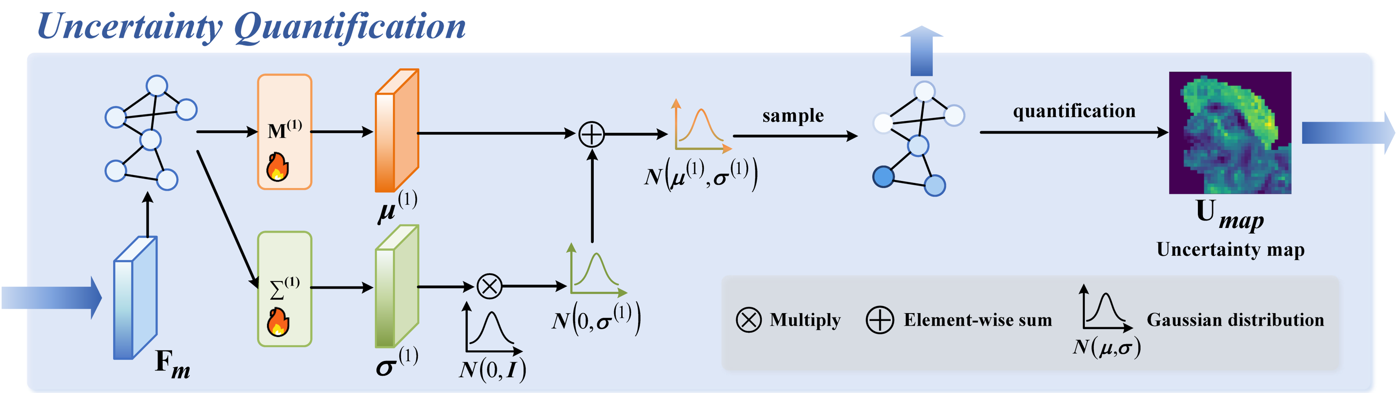
Method



Overview of UNICL-SAM. With the fine-grained features, UNICL-SAM performs ***uncertainty quantification*** *to model the input distribution and produce the estimated uncertainty map. The corresponding* ***uncertainty-guided graph augmentation and feature refinement stages*** *tend to learn robust graph representations by mitigating the impact of high uncertainty areas. Further, UNICL-SAM conducts* ***part prototype discovery*** *to get the part prototypes with the other three kinds of in-context instructions. Finally, the in-context instructions are fed into the query prompt generator to get the in-context prompt for guiding the mask decoder to predict segmentation results.*

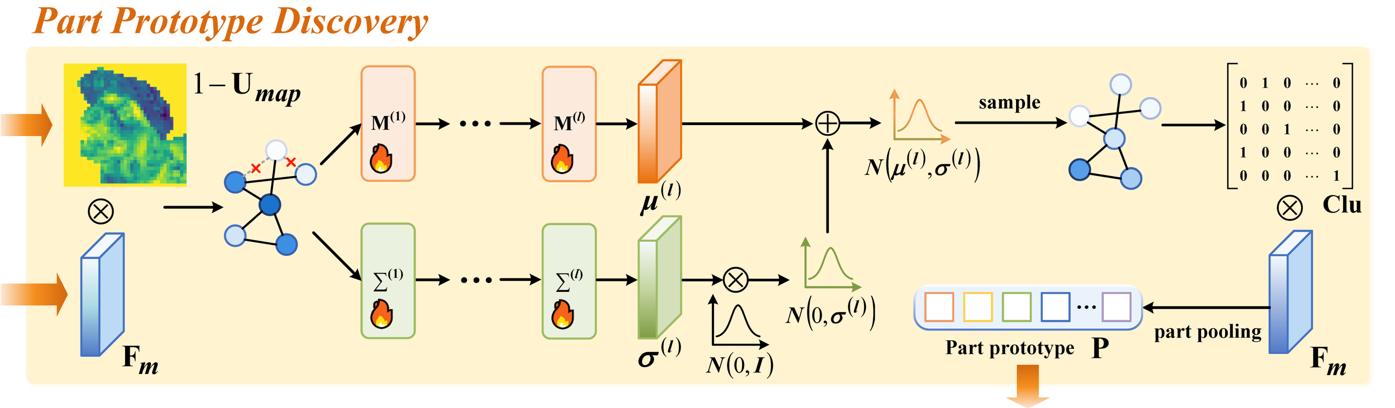
Uncertainty Quantification

- Our proposed uncertainty Gaussian graph network (UGGN) maps the foreground graph representation to a high-dimensional Gaussian distribution and obtains the uncertainty estimate map through parameter resampling and quantification.



Part Prototype Discovery

- Through systematic classification of nodes within the uncertainty-optimized foreground graph representation, each node is assigned to its corresponding cluster category. Then we obtain the part prototypes by averaging the node features in the same cluster.



[1] Hong *et al.*, Cost aggregation with 4d convolutional swin transformer for few-shot segmentation, in *ECCV 2022*. [2] Liu *et al.*, Matcher: Segment anything with one shot using all-purpose feature matching, in *ICLR 2024*. [3] Liu *et al.*, A Simple Image Segmentation Framework via In-Context Examples, in *NeurIPS 2024*. [4] Min *et al.*, Hypercorrelation squeeze for few-shot segmentation, in *ICCV 2021*. [5] Sheng *et al.*, Towards more unified in-context visual understanding, in *CVPR 2024*. [6] Sun *et al.*, VRP-SAM: SAM with visual reference prompt, in *CVPR 2024*. [7] Wang *et al.*, Images speak in images: A generalist painter for in-context visual learning, in *CVPR 2023*. [8] Wang *et al.*, SegGPT: Segmenting Everything In Context, in *ICCV 2023*. [9] Zhang *et al.*, Feature- proxy transformer for few-shot segmentation, in *NeurIPS 2022*. [10] Zhang *et al.*, Personalize segment anything model with one shot, in *ICLR 2024*.

Experimental Results

Main results & Ablation studies

Methods	Resolution	Trainable Params	COCO-20^i one-shot	COCO-20^i few-shot	FSS-1000 one-shot	FSS-1000 few-shot	LVIS-92^i one-shot	LVIS-92^i few-shot
specialist model								
HSNet [4]	480	28M	41.7*	50.7*	86.5*	88.5*	17.4	22.9
VAT [1]	417	52M	42.9*	49.4*	90.3*	90.8*	18.5	22.7
FPTrans [9]	480	139M	56.5*	65.5*	-	-	-	-
generalist model								
Painter [7]	448	354M	33.1	32.6	61.7	62.3	10.5	10.9
SegGPT [8]	448	354M	56.1	67.9	85.6	89.3	18.6	25.4
PerSAM [10]	-	2M	23.5	-	75.6	-	18.4	-
Matcher [2]	896	-	52.7	60.7	87.0	89.6	33.0	40.0
VRP-SAM [6]	512	1.6M	53.9	-	-	-	-	-
ICVU [5]	256	309M	58.0	-	-	-	-	-
SINE [3]	896	19M	64.5	66.1	-	-	31.2	35.5
UNICL-SAM	518	55M	**77.8**	**78.7**	84.0	86.3	**34.1**	37.4

Comparison with state-of-the-art specialist and generalist models on few-shot semantic segmentation benchmarks.

Multi-scale Feature Adapter	Part Prototype Discovery	Uncertainty Probabilistic Modeling	COCO-20^i MIoU ↑	COCO-20^i MAE ↓
✗	✗	✗	71.77	0.039
✗	✗	✓	72.99	0.037
✗	✓	✗	72.41	0.038
✓	✗	✓	73.28	0.036
✓	✓	✗	73.78	0.036
✗	✓	✓	74.61	0.034
✓	✓	✓	**74.64**	**0.034**

Ablation of proposed components.

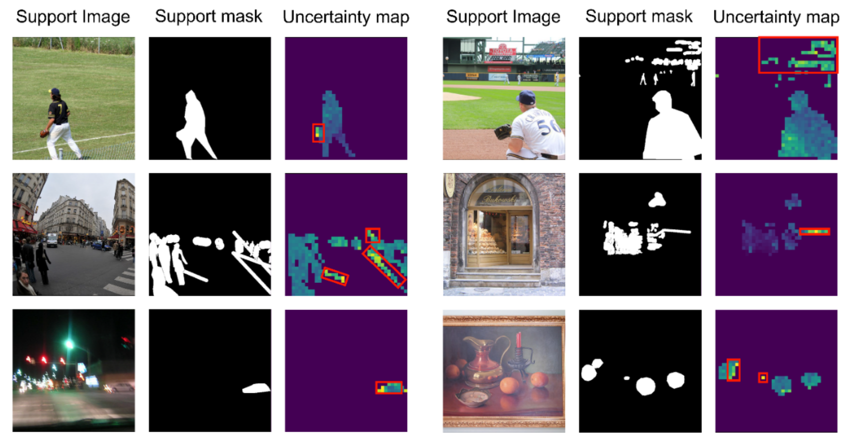
Un-guided Graph Augmentation	Un-guided Feature Refinement	COCO-20^i MIoU ↑	COCO-20^i MAE ↓
✗	✗	73.78	0.036
✓	✗	73.82	0.036
✗	✓	73.99	0.035
✓	✓	**74.64**	**0.034**

Ablation of uncertainty-guided optimization strategies.

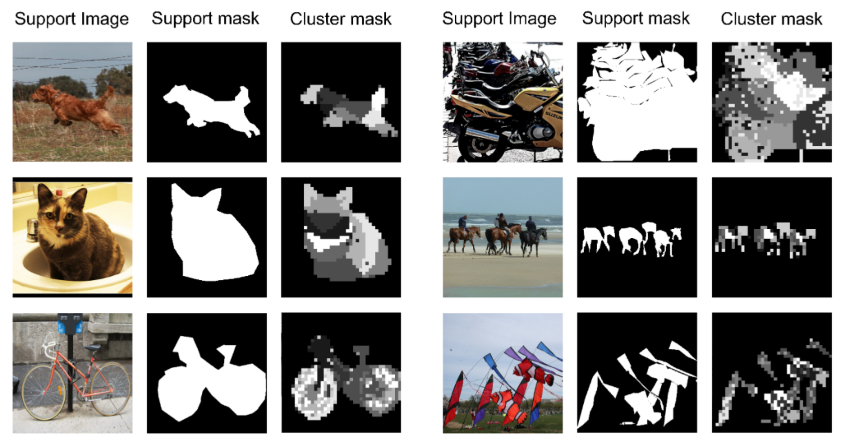
Robustness comparison with advanced in-context segmentation generalists

Methods	Resolution	Trainable Params	Clean	Support Image: Color	Support Image: Blurness	Support Image: Compression	Support Image: Space	Support Image: Domain shift	Support Label: Deformation	Avg. Decline	Avg. Decline Ratio
SegGPT [8]	448	354M	62.1	−2.7	−3.0	−2.7	−7.8	−6.2	−8.8	−5.2	−8.4%
Matcher [2]	896	-	41.6	−0.5	−2.5	−6.5	−3.8	−5.3	−6.2	−4.1	−9.9%
SINE [3]	896	19M	70.0	−0.7	−1.7	−5.9	−2.8	−6.5	−12.4	−5.0	−7.1%
UNICL-SAM	518	55M	**79.8**	−1.0	−1.8	−3.0	−3.3	−4.8	−5.0	**−3.1**	**−3.9%**

Visualization of intermediate outputs

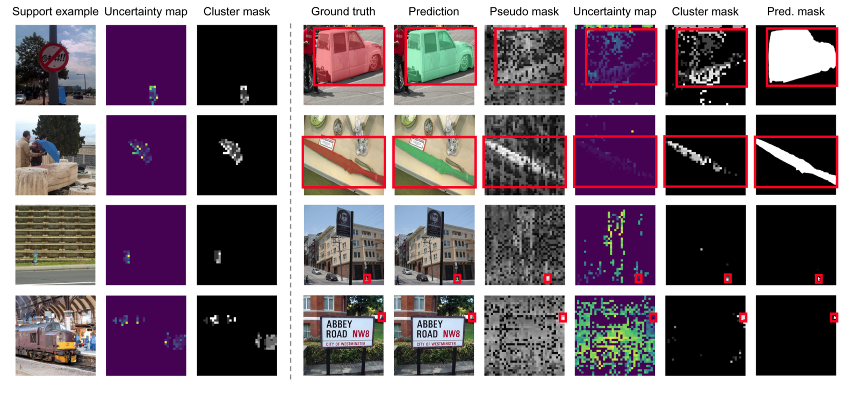


Visualization of uncertainty maps.



Visualization of cluster masks.

Qualitative results and analysis



Detailed visualization of the network's intermediate outputs, including pseudo masks, uncertainty maps, and clustered masks.

Qualitative results on robustness tests. UNICL-SAM demonstrates strong robustness across diverse degradation scenarios.