

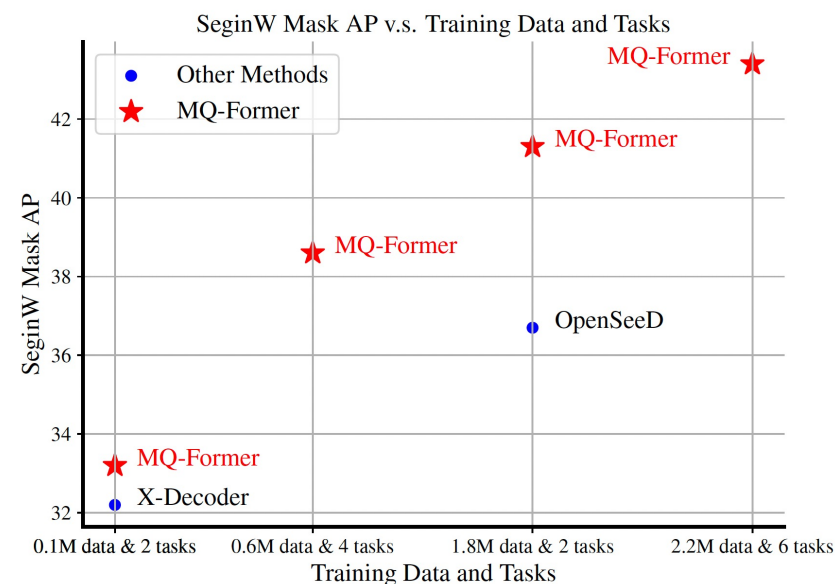
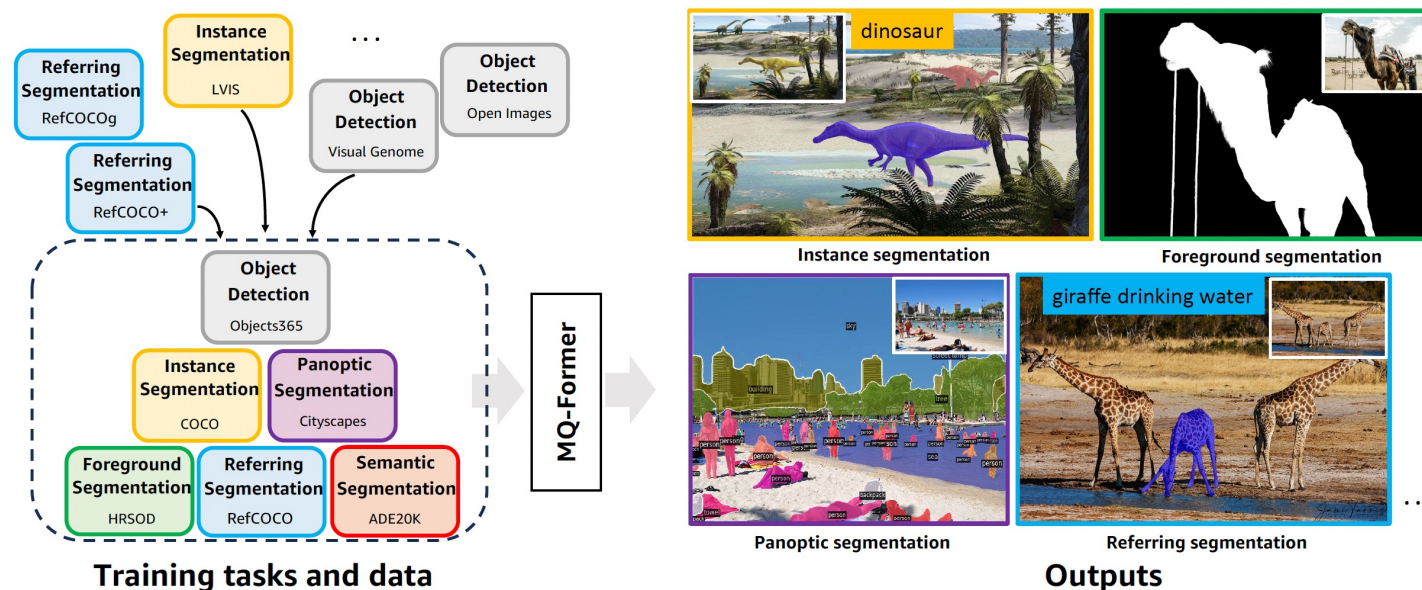
Scaling up Image Segmentation across Data and Tasks

*Pei Wang¹, Zhaowei Cai¹, Hao Yang¹, Ashwin Swaminathan¹,
R. Manmatha², Stefano Soatto²*

¹Amazon AGI ²AWS AI Labs

Scalable Segmentation Models

- A model is **scalable** if can effectively improve with the increase in both dataset size and task diversity.



Scalable Segmentation Models

- Key advantages:
 - **Continuous Evolution**: Learns and improves over time by leveraging both existing and future datasets, without the need for frequent redesign.
 - **Cross-Task Integration**: Incorporates knowledge across diverse tasks and datasets, enhancing generalization.
 - **Universal Applicability**: Enables a single, robust model capable of generalizing across a wide range of applications, including open-vocabulary, free-form, and in-the-wild segmentation.

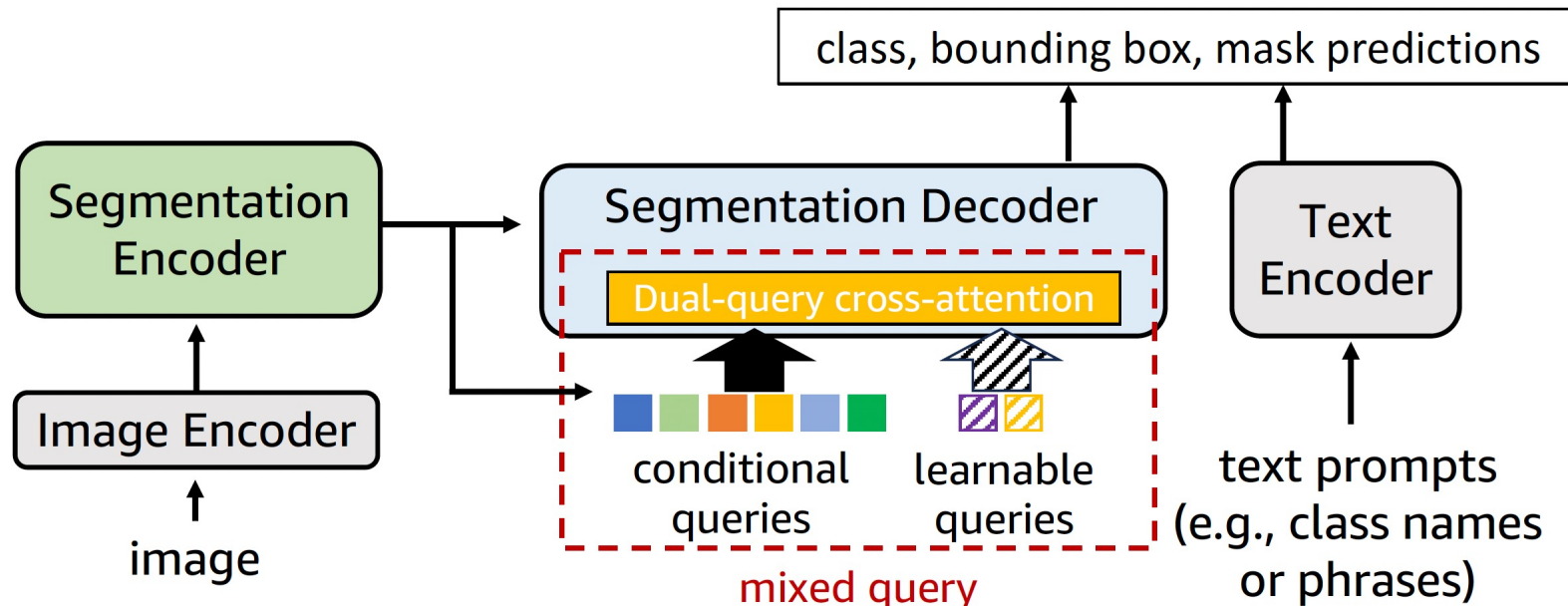
Scalable Segmentation Models

- Existing models have **limited scalability**.

	Data Scalability	Task Scalability					
		Instance	Semantic	Panoptic	Referring	Foreground	Detection
MSeg [26]	✓		✓				
UniSeg [21]	✓		✓				
OneFormer [18]		✓	✓	✓			
OpenSeeD [70]		✓	✓	✓			✓
OMG-Seg [31]	✓	✓	✓	✓			
X-Decoder [75]	✓	✓	✓	✓	✓		
DaTaSeg [15]	✓		✓	✓			✓
Our MQ-Former	✓	✓	✓	✓	✓	✓	✓

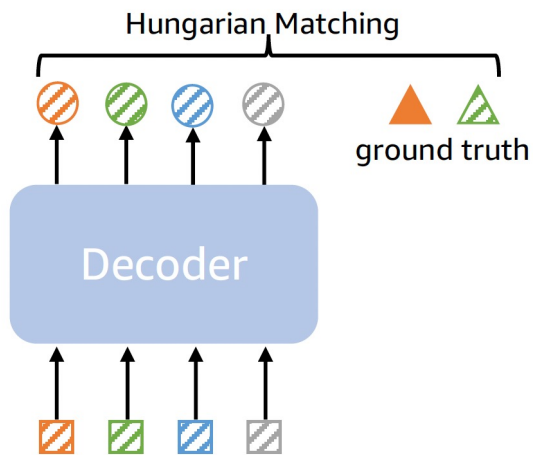
Query Design for Scalable Segmentation

- The design of **object queries** is the key limitation preventing effective scalability.
- Mixed-Query Transformer (**MQ-Former**), a novel scalable segmentation framework, is introduced.

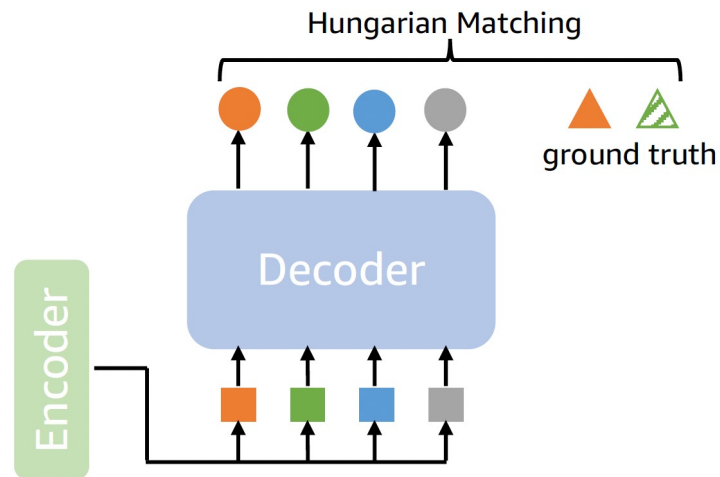


Query Design for Scalable Segmentation

- Learnable query vs Conditional query



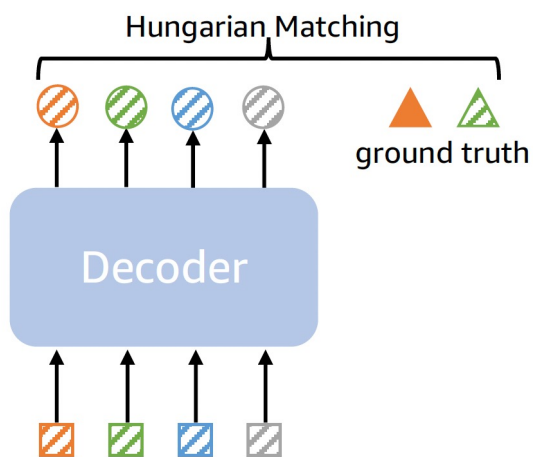
(a) Learnable query



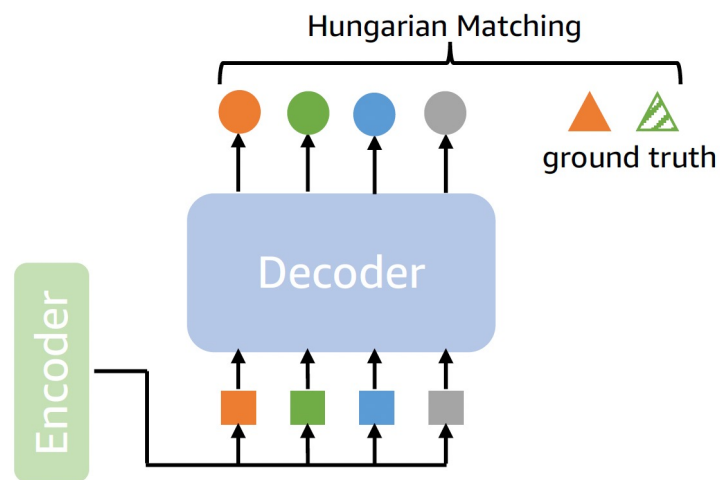
(b) Conditional query

Query Design for Scalable Segmentation

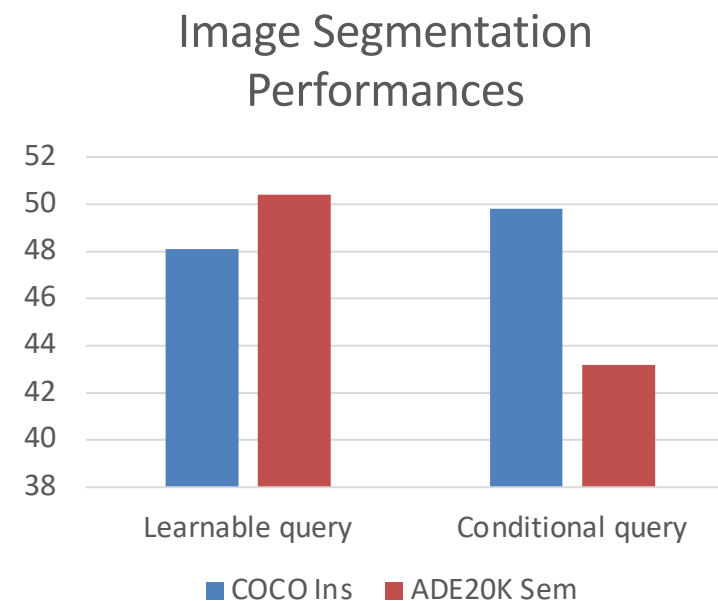
- Learnable query vs Conditional query



(a) Learnable query



(b) Conditional query



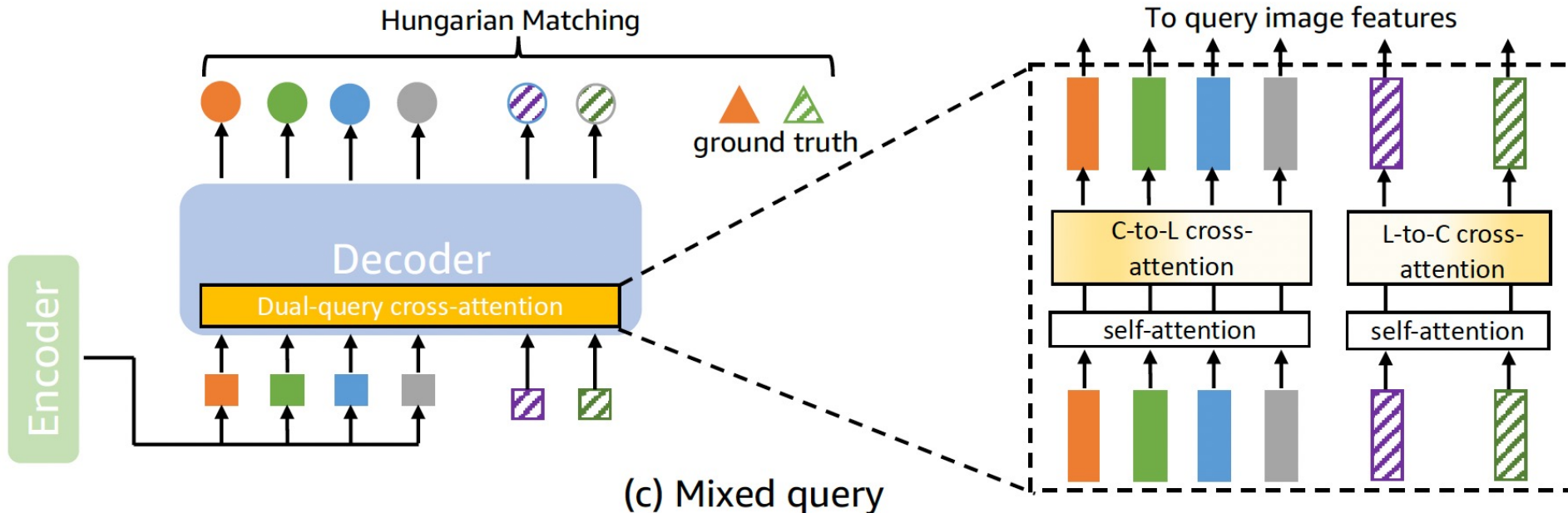
Learnable queries perform well on semantic segmentation but struggle with instance-level tasks, while conditional queries perform significantly worse on semantic segmentation.

Query Design for Scalable Segmentation

- Both learnable and conditional queries have their respective strengths.
- **Learnable queries** excel at handling large, amorphous background regions.
- **Conditional queries** specialize in capturing local, instance-level features.
- However, their individual limitations hinder scalability across diverse datasets and tasks.

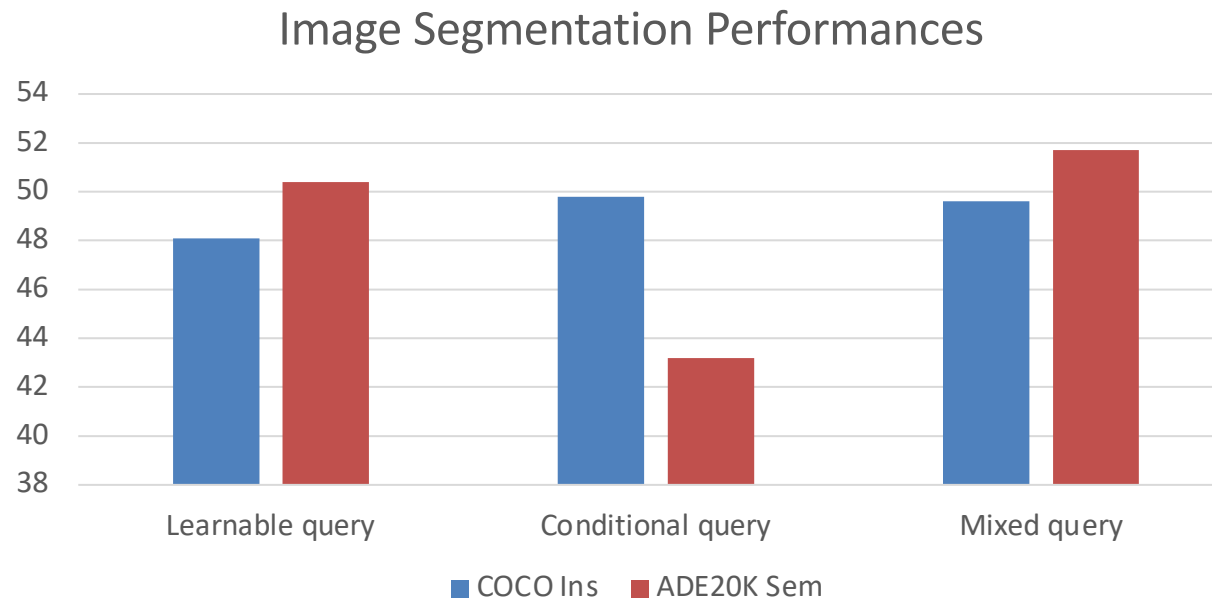
Query Design for Scalable Segmentation

- Mixed query
 - combines learnable and conditional queries
 - refines their representation through mutual exchange
 - dynamically chooses preferred objects



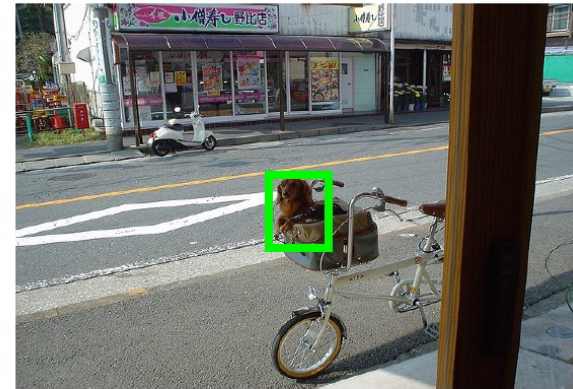
Query Design for Scalable Segmentation

- Mixed query
 - combines learnable and conditional queries
 - refines their representation through mutual exchange
 - dynamically chooses preferred objects



Scalability to More Data and Tasks

- To further enhance scalability
 - Synthesizing masks and object captions on bbox-annotated datasets











the dog is brown; this is a dog;
a dog in a basket; a dog in the
basket; a brown dog



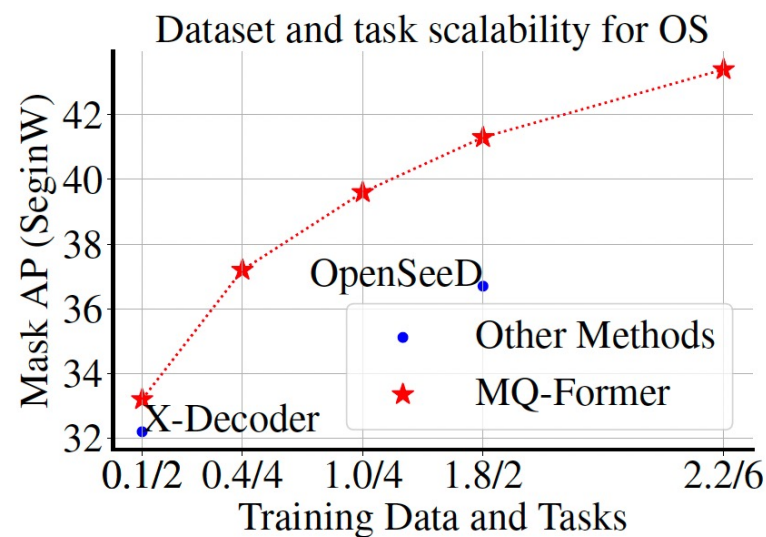
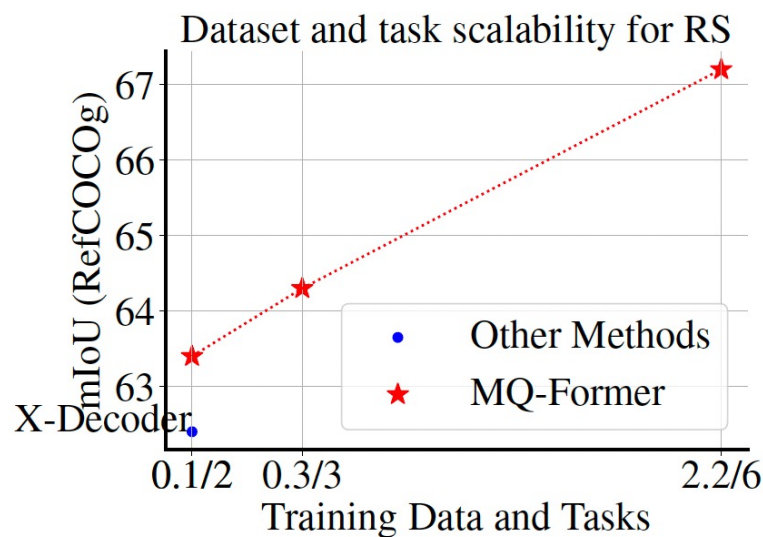
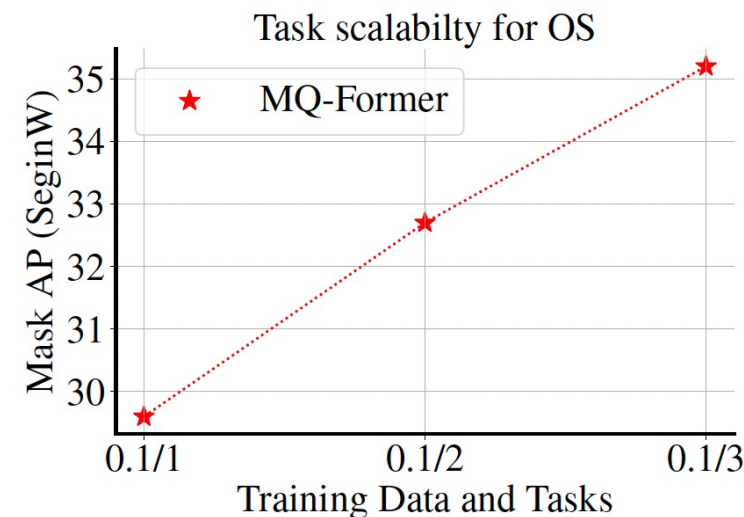
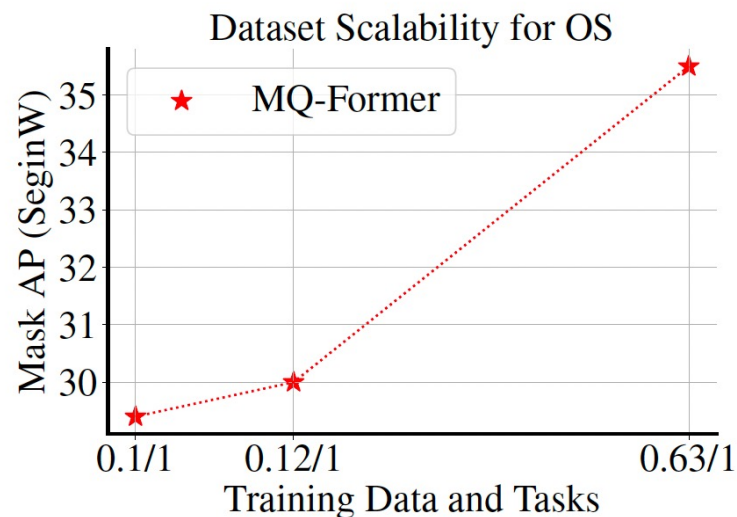
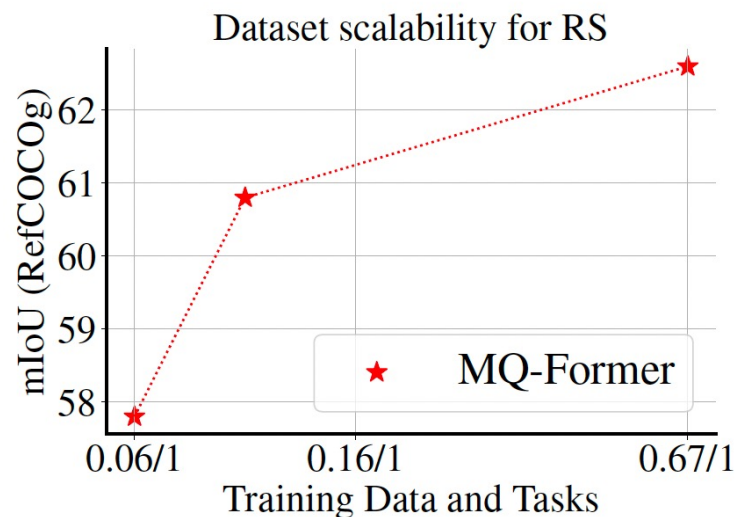
the stool is wooden; the
stool is brown; the stool is
wood; a wooden bar stool;
a wooden stool

Advancement of Mixed Query

- The counter prediction of examples by mixed query.

Objects predicted by conditional queries			Objects predicted by learnable queries		
Input image	Prompt	Output mask	Input image	Prompt	Output mask
	'sky'			'table'	
	'floor'			'car'	

Scalability of MQ-Former



Segmentation Output Visualization for MQ-Former



(a) Open-vocabulary panoptic segmentation



(b) Open-vocabulary instance segmentation

(c) Free-form referring segmentation



(d) Free-form referring matting