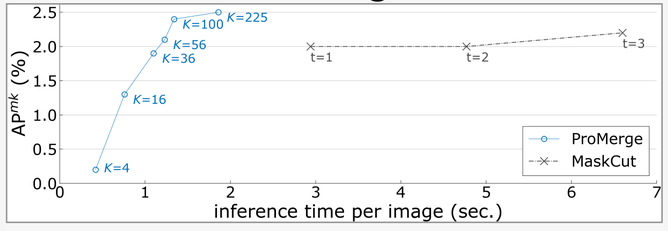
ProMerge: Prompt and Merge for Unsupervised Instance Segmentation

Dylan Li[1,*] and Gyungin Shin[2,*†]
[1]Meta Reality Labs [2]Visual Geometry Group, University of Oxford
*Equal contribution †Corresponding author

Problem statement

Instance segmentation aims to segment multiple foreground objects in an image. However, training a model typically requires human labels, which can be a costly to obtain. We approach instance segmentation and object detection without such annotations by leveraging self-supervised pre-trained features.

Our solution: ProMerge



Recently, state-of-the-art methods tackle this task by solving computationally intensive generalized eigenvalue systems. Unlike these approaches, we propose a simple yet effective strategy, called ProMerge that is:

Conceptually simple.
No need to solve computationally demanding eigenvalue systems

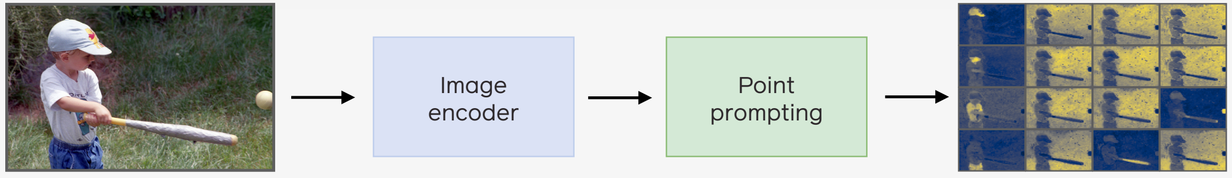
3.6 times faster.
Up to 4.9 times faster at a comparable AP on COCO2017.

Leading in AP and AR

	LVIS (AR/AP)	KITTI (AR/AP)	Objects 365 (AR/AP)	SA-1B (AR/AP)
MaskCut	2.6 / 0.9	2.2 / 0.2	4.0 / 1.7	0.6 / 0.8
ProMerge	3.3 / 1.3	1.9 / 0.3	6.0 / 2.2	0.8 / 1.2
CutLER	24.9 / 8.7	23.3 / 3.9	34.3 / 11.5	13.5 / 5.5
ProMerge+	25.1 / 8.9	25.7 / 5.4	35.8 / 12.2	16.3 / 7.8

Step 1. Prompting

We obtain mask proposals for an image with point-prompting.

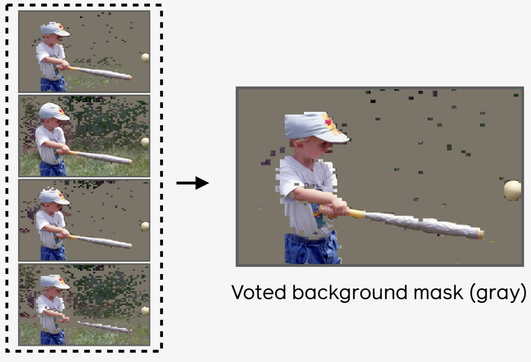


Step 2. Mask pruning

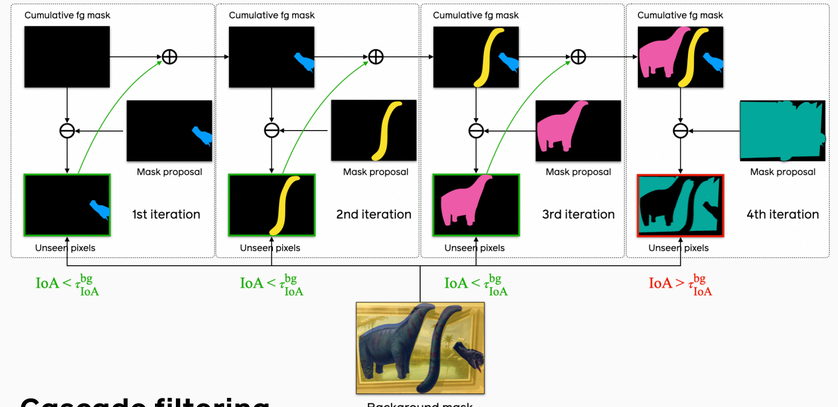
Not all mask proposals are foreground masks. To filter background masks among the proposals, we employ a two-step strategy:

(i) Background aggregation

(ii) Cascade filtering.



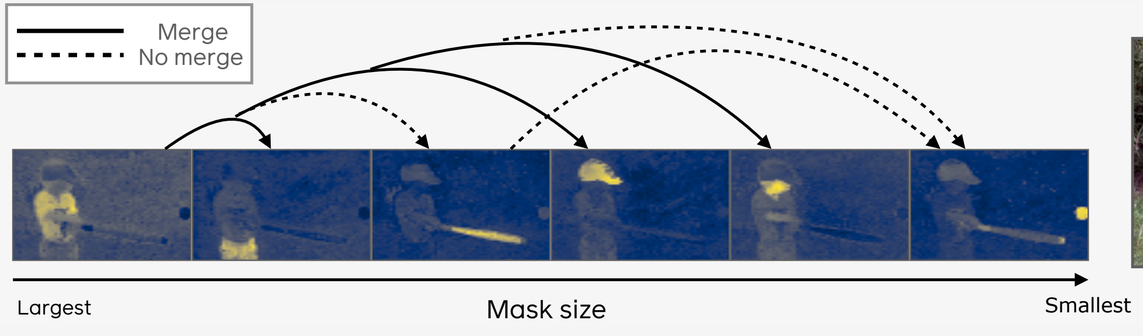
Background aggregation



Cascade filtering

Step 3. Merging

After the mask pruning step, we iteratively merge filtered masks to produce object-level masks.



Final predictions