Zero-Shot Edge Detection with SCESAME: Spectral Clustering-based Ensemble for Segment Anything Model Estimation

Hiroaki Yamagiwa[1,2], Yusuke Takase[1], Hiroyuki Kambe[2], Ryosuke Nakamoto[1,2]

[1]Kyoto University, [2]Rist Inc.

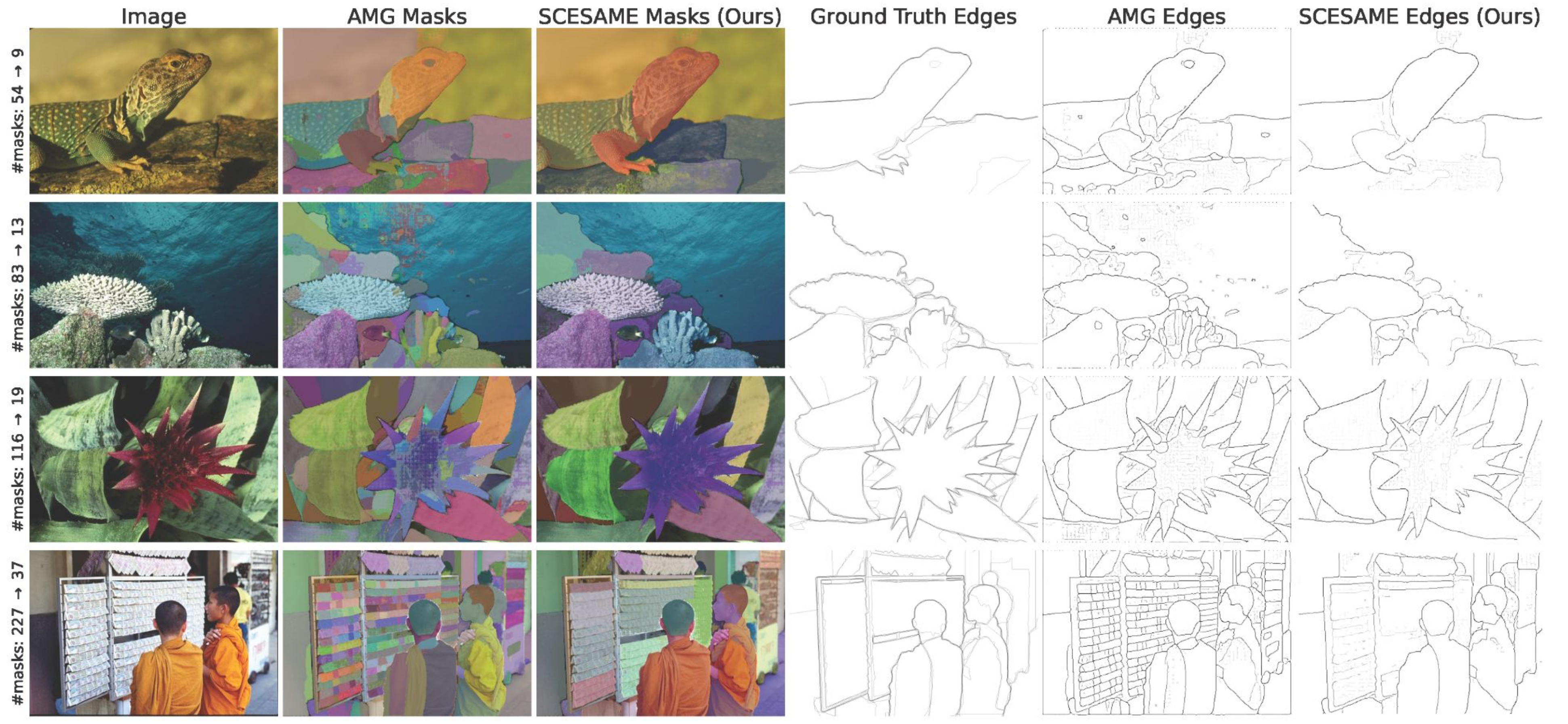
2nd Workshop on Pretraining - January 7, 2024 - at WACV 2024

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

Background — **SAM** can be used for edge detection, but suffers from the problem of **over-detecting edges**.

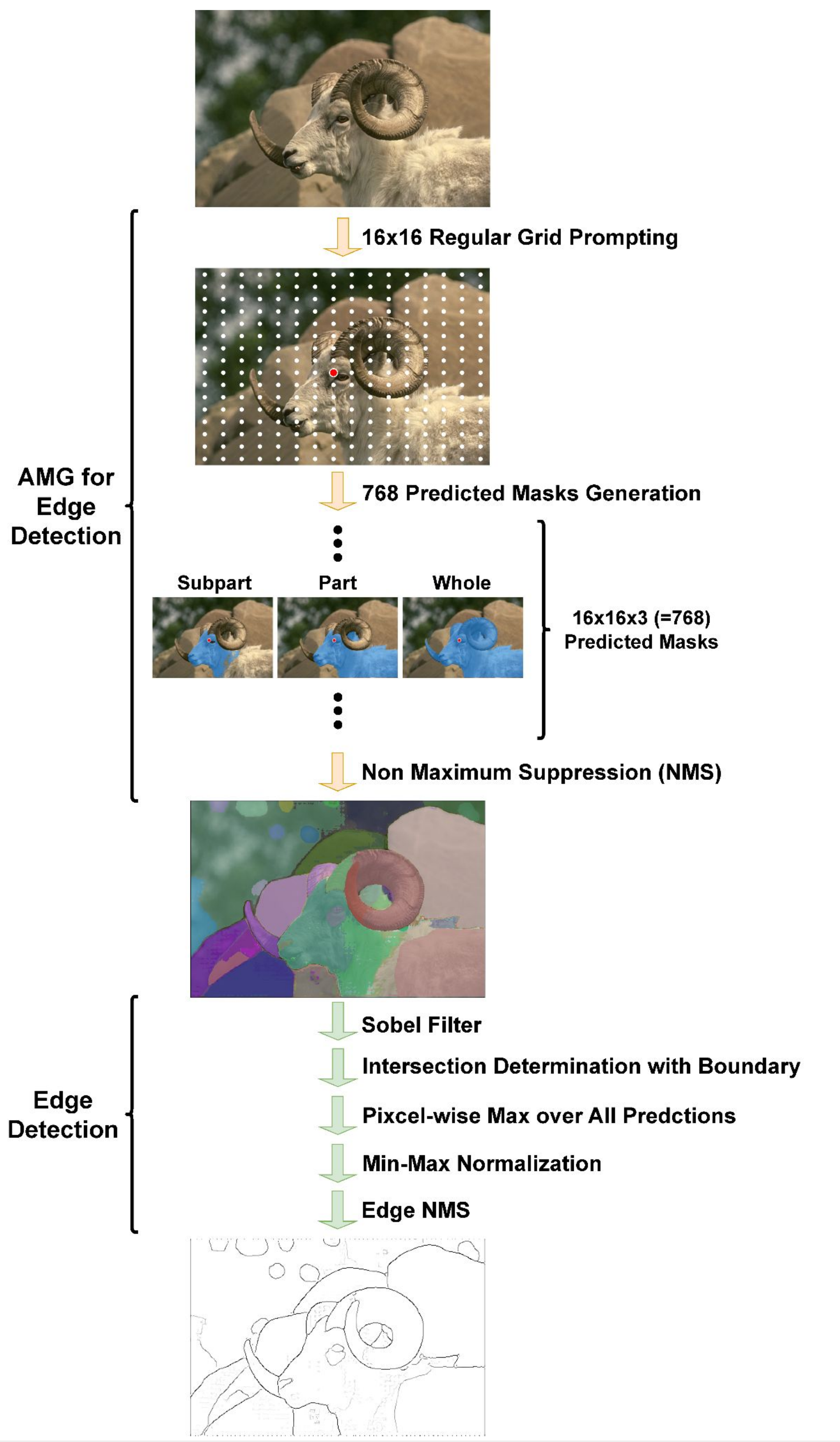
Method — (1) **eliminate small masks**, (2) **combine masks**, and (3) **remove artifacts** after edge detection.

Results
- Our method performs **better than SAM**.
- Our method comes **close to human performance**.



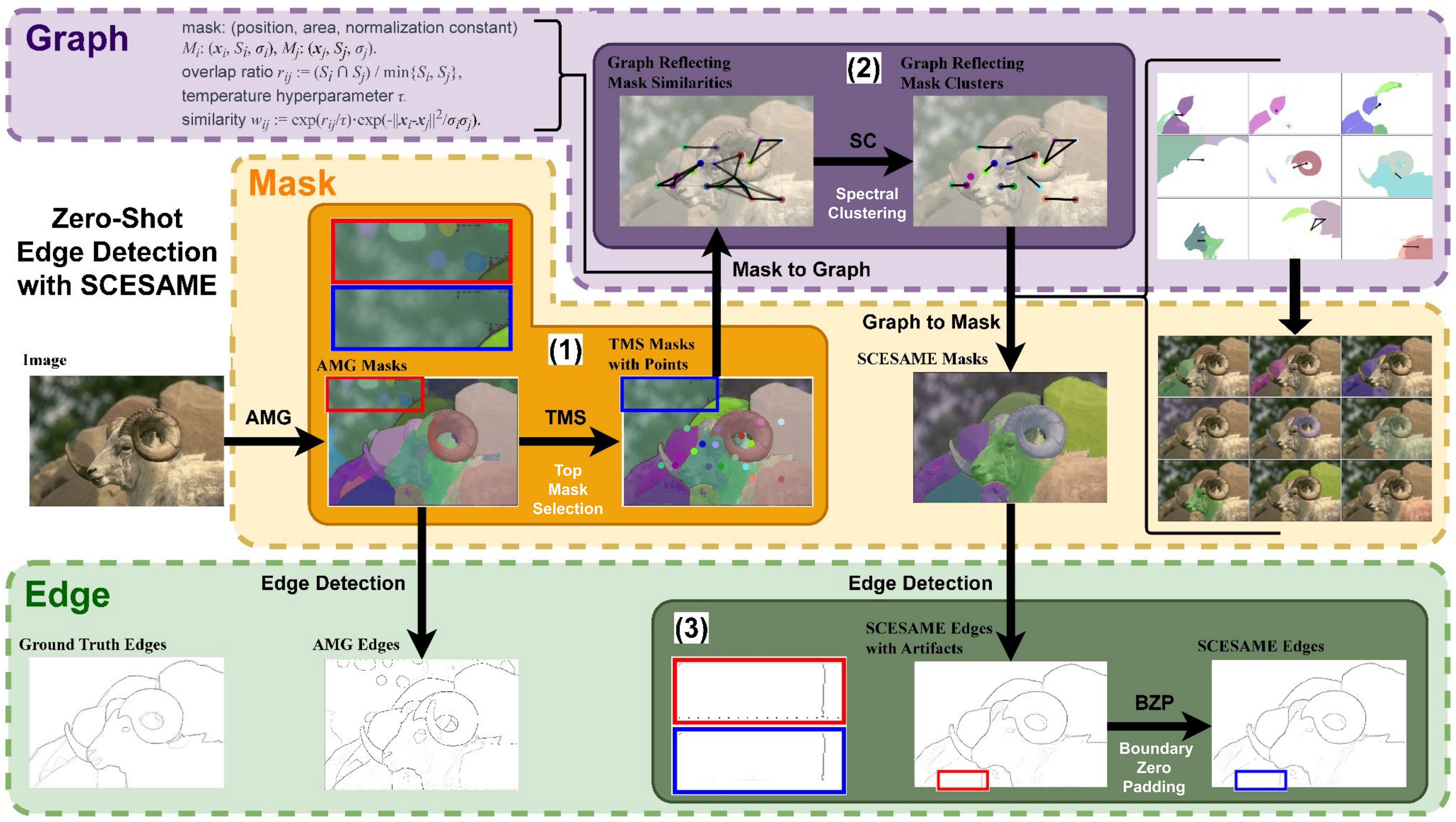
Background

- **Segment Anything Model** (**SAM**) can generate masks for the entire image in zero-shot using **Automatic Mask Generation** (**AMG**).
- Edge detection with AMG **over-detects edges**.



Method: Edge Detection with SCESAME Overcomes the Problem by Three Steps

1. **TMS**: Sort the AMG masks by size and **eliminate small masks**.
2. **SC**: Generate new masks by **combining the remaining masks**.
3. **BZP**: **Fill pixels** within a few pixels of the image boundary **with zeros**.



Results: Edge Detection Experiments on BSDS500 and NYUDv2

BSDS500 (left)
- It **outperforms most CNN-based methods from 7-8 years ago**.
- It also comes **close to human performance**.

NYUDv2 (right)

It performs **almost as well as recent CNN-based methods**.

- Our method performs **better than SAM**.
- There is still a gap compared to SOTA.

	Method	Pub.'Year	ODS	OIS	AP
	Human [25]	ICLR'16	**0.803**	-	-
Traditional	Canny [7]	PAMI'86	0.600	0.640	0.580
	Felz-Hutt [15]	IJCV'04	0.610	0.640	0.560
	gPb-owt-ucm [1]	PAMI'10	0.726	0.757	0.696
	SCG [41]	NeurIPS'12	0.739	0.758	0.773
	Sketch Tokens [27]	CVPR'13	0.727	0.746	0.780
	PMI [21]	ECCV'14	0.741	0.769	0.799
	SE [12]	PAMI'14	0.746	0.767	0.803
	OEF [18]	CVPR'15	0.746	0.770	0.820
	MES [44]	ICCV'15	0.756	0.776	0.756
7 to 8-Year-Old CNN	DeepEdge [2]	CVPR'15	0.753	0.772	0.807
	CSCNN [20]	ArXiv'15	0.756	0.775	0.798
	MSC [45]	PAMI'15	0.756	0.776	0.787
	DeepContour [42]	CVPR'15	0.757	0.776	0.800
	HFL [3]	ICCV'15	0.767	0.788	0.795
	HED [50]	ICCV'15	0.788	0.808	**0.840**
	Deep Boundary [25]	ICLR'16	**0.813**	**0.831**	**0.866**
	CEDN [53]	CVPR'16	0.788	0.804	-
	RDS [31]	CVPR'16	0.792	0.810	0.818
	COB [32]	ECCV'16	0.793	**0.820**	**0.859**
SAM	SAM [23]	ICCV'23	0.768	0.786	0.794
	SAM [23] (Recalc.)	ICCV'23	0.730	0.754	0.729
	SAM-p5 (Our Baseline)	-	0.754	0.779	0.763
Ours	SCESAME-t2c2p5	-	0.796	0.812	0.780
	SCESAME-t2c3p5		0.797	0.811	0.768
	SCESAME-t3c2p5		**0.800**	**0.814**	0.773
	SCESAME-t3c3p5		0.796	0.809	0.753
SOTA	EDTER-MS [37]	CVPR'22	0.840	0.858	0.896
	EDTER-MS-VOC [37]	CVPR'22	**0.848**	**0.865**	0.903
	UAED-MS [58]	CVPR'23	0.837	0.855	0.897
	UAED-MS-VOC [58]	CVPR'23	0.844	0.864	**0.905**

	Method	Pub.'Year	ODS	OIS	AP
Traditional	gPb-ucm [1]	PAMI'11	0.632	0.661	0.562
	Silberman *et al.* [43]	ECCV'12	0.658	0.661	-
	gPb+NG [16]	CVPR'13	0.687	0.716	0.629
	SE [12]	PAMI'14	0.695	0.708	0.679
	SE+NG+ [17]	ECCV'14	0.706	0.734	**0.738**
	OEF [18]	CVPR'15	0.651	0.667	-
	SemiContour [57]	CVPR'16	0.680	0.700	0.690
CNN-based	HED [50]	ICCV'15	0.720	0.734	0.734
	RCF [30]	CVPR'17	0.729	0.742	-
	AMH-Net [51]	NeurIPS'17	**0.744**	**0.758**	**0.765**
	LPCB [10]	ECCV'18	0.739	**0.754**	-
	BDCN [19]	CVPR'19	**0.748**	**0.763**	**0.770**
	PiDiNet [46]	ICCV'21	0.733	0.747	-
	SAM-p5 (Our Baseline)	-	0.699	0.719	0.707
	SCESAME-t3c2p5 (Ours)	-	**0.742**	**0.754**	0.707
	EDTER [37] (SOTA)	CVPR'22	**0.774**	**0.789**	**0.797**

The best three results, excluding **SOTA** methods, are highlighted in **red**, **blue**, and **purple**.