University of Münster
Department of Information Systems

Title

Seminar Thesis

in the context of the seminar

More than meets the A-eye: Reflecting Human Vision in Artificial Intelligence

submitted by

Annika Terhörst, Dominik Eitner, Simon Luttmann

Principal Supervisor	Prof. Dr. Christian Grimme
Supervisor	Lucas Stampe, M.Sc.
	Chair for Data Science: Computational Social Science & Systems Analysis
Student Candidate	Annika Terhörst, Dominik Eitner, Simon Luttmann
Matriculation Number	527050, 503468, ..
Field of Study	Information Systems
Contact Details	annika.terhoerst@uni-muenster.de, deitner@uni-muenster.de, sluttman@uni-muenster.de
Submission Date	4.12.2025

Contents

1 Introduction

1.1 Motivation and Context

1.2 Research Gap

2 Theoretical Background

2.1 Image Segmentation

Image segmentation is the process of dividing an image into different regions by grouping pixels and assigning each pixel a label. This step is an important part of many computer vision applications, such as detecting tumors in medical images or identifying pedestrians in autonomous driving. According to human visual perception, the identified regions are non-overlapping and meaningful - however, defining what exactly counts as a “meaningful” region can be difficult, as human perception is subjective and object boundaries are not always clear (Yu et al., 2023).

There are three common types of segmentation:

Semantic segmentation assigns every pixel in an image a semantic label, such as “car” or “sky”. *Instance segmentation* separates individual objects within the same class, for example distinguishing several people in one image. *Panoptic segmentation* combines both approaches by providing pixel-wise class labels and also identifying individual object instances.

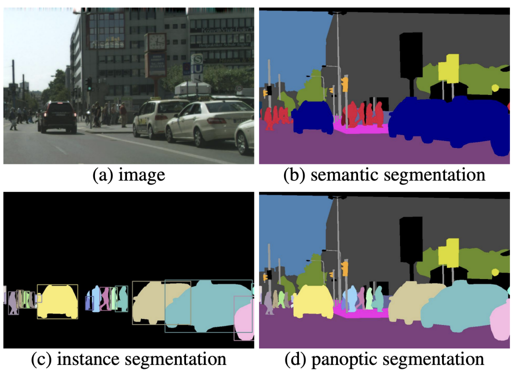


Figure 1 Types of image segmentation by Kirillov et al., 2019

Earlier approaches to image segmentation include algorithms such as k-means-clustering (Dhanachandra et al., 2015). Yet in recent years, deep learning models have significantly improved the segmentation effect and performance, therefore becoming the dominant method for solving segmentation tasks in complex environments (Minaee et al., 2022).

According to Zhou et al., 2024, the above-described image segmentation methods fall into the category of generic image segmentation (GIS). The category of promptable image segmentation (PIS) extends GIS by specifying the target to segment through a prompt. This prompt can have various forms such as text, box or points.

2.2 Salient Object Detection

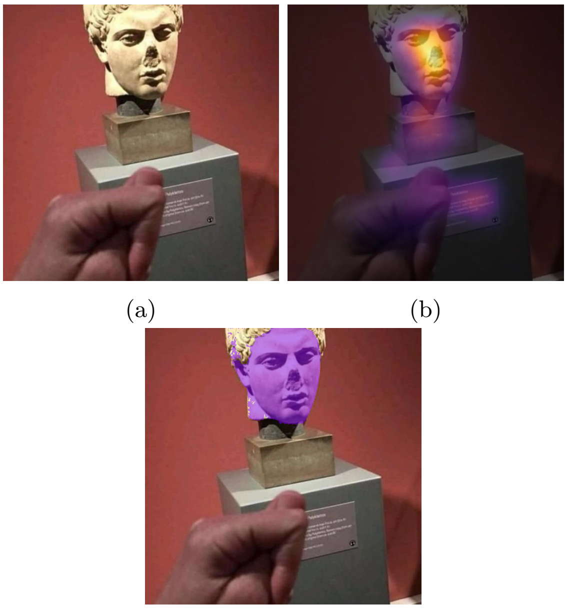
The human visual system pays more attention to certain parts in an image, a property known as saliency. Inspired by this mechanism, *saliency detection* models aim to predict which regions in an image are most likely to attract human visual attention. These models typically provide saliency maps in form of heat maps, in which higher intensity values indicate regions detected to be more important (Ahmadi et al., 2018).

Salient Object Detection (SOD) – also referred to as salient object segmentation (Borji et al., 2019) or saliency segmentation (Kakanopas & Woraratpanya, 2021) – goes one step further by segmenting the most salient object(s) of an image. SOD can be interpreted as a two-stage process: 1) Detection of the most salient object and 2) Accurate segmentation of the region of that object. In contrast to general image segmentation, SOD focuses on segmenting only those objects that are (or that are predicted to be) most salient (Borji et al., 2019; T. Liu et al., 2011). Figure 2 illustrates the difference between saliency detection and salient object detection.

2.3 Image Segmentation Models

Table 1 offers a structured overview of prominent and state-of-the-art image segmentation models, organized according to the segmentation tasks for which they are most suitable.

For our use case, only promptable segmentation models are suitable, as the eye-tracking data provides point inputs that will be used to im-



(a) (b)

(c)

Figure 2 (a) the original image, (b) saliency map (Alexander Kroner, 2025) (c) salient object detection mask generated using SAM3 guided by the eye-tracking data

plement SOD. Among the identified promptable models, SAM and its successors are the only widely adopted models that natively support point-based prompts (Kirillov et al., 2023). In contrast, Grounded-SAM (Ren et al., 2024) and Florence-2 (Xiao et al., 2023) are limited to text prompts, and SEEM, whose last update dates back to 2023, is less actively maintained and less commonly used than SAM (Zou et al., 2023).

The Segment Anything Model (SAM) was developed by Meta AI and first introduced in mid-2023 (Kirillov et al., 2023). SAM performs object segmentation based on prompts, including points and bounding boxes. With more than 15,000 citations, SAM has become one of the de facto standards for domain-specific applications and is already employed in several specialized salient object detection settings, such as camouflage object segmentation and medical image segmentation (Chen et al., 2025), RGB-T SOD (Z. Liu et al., 2025), and text-driven SOD (Yuan et al., 2026). The most recent version, SAM 3, was released on 20 November 2025. Its rapid adoption - reaching over 5.1k GitHub stars within two weeks - indicates strong community interest. Compared to previous versions, SAM 3 introduces the ability to detect and segment instances that match a given text description, and to further refine detections using visual examples (Carion et al., 2025).

2.4 Evaluation of Image Segmentation Models

- Subjective & Objective (Image segmentation evaluation: a survey of methods)

- SAM: IoU (SAM paper itself)

- Metrics in SOD: PR, F-measure, MAE (Salient object detection: A survey)

Generic Image Segmentation			
Instance segmentation	**Semantic segmentation**	**Panoptic segmentation**	**Promptable Image Segmentation**
Mask R-CNN	DeepLabV3	Mask2Former	SAM2, SAM3
YOLOv11-seg	FCN	Panoptic-DeepLab	Grounded-SAM
YOLACT	U-Net	Panoptic FPN	SEEM
Mask2Former	SegFormer	...	Florence-2
Mask DINO	LR-ASPP		...
...	...		

Table 1 Overview of prominent image segmentation models categorized by segmentation task.

3 Methodology

3.1 Experimental Design

3.2 Planned Practical Steps

4 Results

4.1 Model Selection

4.2 Expected Results

A Appendix

TODO: Add result pictures and/or our code here

Bibliography

Ahmadi, M., Hajabdollahi, M., Karimi, N., & Samavi, S. (2018, January). Context aware saliency map generation using semantic segmentation [arXiv:1801.00256 [cs]]. https://doi.org/10.48550/arXiv.1801.00256

Alexander Kroner. (2025, March). Visual Saliency Prediction - a Hugging Face Space by alexanderkroner. Retrieved December 2, 2025, from https://huggingface.co/spaces/alexanderkroner/saliency

Borji, A., Cheng, M.-M., Hou, Q., Jiang, H., & Li, J. (2019). Salient object detection: A survey. *Computational Visual Media*, *5*(2), 117–150. https://doi.org/10.1007/s41095-019-0149-9

Carion, N., Gustafson, L., Hu, Y.-T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K. V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., ... Feichtenhofer, C. (2025, November). SAM 3: Segment Anything with Concepts [arXiv:2511.16719 [cs]]. https://doi.org/10.48550/arXiv.2511.16719

Chen, T., Cao, R., Yu, X., Zhu, L., Ding, C., Ji, D., Chen, C., Zhu, Q., Xu, C., Mao, P., & Zang, Y. (2025, November). SAM3-Adapter: Efficient Adaptation of Segment Anything 3 for Camouflage Object Segmentation, Shadow Detection, and Medical Image Segmentation [arXiv:2511.19425 [cs]]. https://doi.org/10.48550/arXiv.2511.19425

Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, *54*, 764–771. https://doi.org/10.1016/j.procs.2015.06.090

Kakanopas, D., & Woraratpanya, K. (2021). A Unified Framework for Saliency Segmentation. In P. Meesad, D. S. Sodsee, W. Jitsakul, & S. Tangwannawit (Eds.), *Recent Advances in Information and Communication Technology 2021* (pp. 191–200). Springer International Publishing. https://doi.org/10.1007/978-3-030-79757-7_19

Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic Segmentation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9396–9405. https://doi.org/10.1109/CVPR.2019.00963

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023, April). Segment Anything [arXiv:2304.02643 [cs]]. https://doi.org/10.48550/arXiv.2304.02643

Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., & Shum, H.-Y. (2011). Learning to Detect a Salient Object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(2), 353–367. https://doi.org/10.1109/TPAMI.2010.70

Liu, Z., Wang, X., Fang, X., Tu, Z., & Wang, L. (2025, October). SAMSOD: Rethinking SAM Optimization for RGB-T Salient Object Detection [arXiv:2510.03689 [cs]]. https://doi.org/10.48550/arXiv.2510.03689

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968

Ren, T., Liu, S., Zeng, A., Lin, J., Li, K., Cao, H., Chen, J., Huang, X., Chen, Y., Yan, F., Zeng, Z., Zhang, H., Li, F., Yang, J., Li, H., Jiang, Q., & Zhang, L. (2024, January). Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks [arXiv:2401.14159 [cs]]. https://doi.org/10.48550/arXiv.2401.14159

Xiao, B., Wu, H., Xu, W., Dai, X., Hu, H., Lu, Y., Zeng, M., Liu, C., & Yuan, L.

(2023, November). Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks [arXiv:2311.06242 [cs]]. https://doi.org/10.48550/arXiv.2311.06242

Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., Yang, T., & Gao, M. (2023). Techniques and Challenges of Image Segmentation: A Review [Publisher: Multidisciplinary Digital Publishing Institute]. *Electronics*, *12*(5), 1199. https://doi.org/10.3390/electronics12051199

Yuan, Y., Zhang, Y., Zhang, S., & Wang, H. (2026). SaliencyCLIP-SAM: Bridging Text and Image Towards Text-Driven Salient Object Detection. In Z. Lin, L. Wang, Y. Jiang, X. Wang, S. Liao, S. Shan, R. Liu, J. Dong, & X. Yu (Eds.), *Image and Graphics* (pp. 29–41). Springer Nature. https://doi.org/10.1007/978-981-95-3393-0_3

Zhou, T., Xia, W., Zhang, F., Chang, B., Wang, W., Yuan, Y., Konukoglu, E., & Cremers, D. (2024, November). Image Segmentation in Foundation Model Era: A Survey [arXiv:2408.12957 [cs]]. https://doi.org/10.48550/arXiv.2408.12957

Zou, X., Yang, J., Zhang, H., Li, F., Li, L., Wang, J., Wang, L., Gao, J., & Lee, Y. J. (2023, July). Segment Everything Everywhere All at Once [arXiv:2304.06718 [cs]]. https://doi.org/10.48550/arXiv.2304.06718

Declaration of Authorship

We hereby declare that, to the best of our knowledge and belief, this thesis titled *Title* is our own, independent work. We confirm that each significant contribution to and quotation in this thesis that originates from the work or works of others is indicated by proper use of citation and references; this also holds for tables and graphical works.

Münster, 4.12.2025

Signatures

Consent Form

Name: Annika Terhörst, Dominik Eitner, Simon Luttmann
Title of Thesis: Title

What is plagiarism? Plagiarism is defined as submitting someone else's work or ideas as your own without a complete indication of the source. It is hereby irrelevant whether the work of others is copied word by word without acknowledgment of the source, text structures (e.g. line of argumentation or outline) are borrowed or texts are translated from a foreign language.

Use of plagiarism detection software. The examination office uses plagiarism software to check each submitted bachelor and master thesis for plagiarism. For that purpose the thesis is electronically forwarded to a software service provider where the software checks for potential matches between the submitted work and work from other sources. For future comparisons with other theses, your thesis will be permanently stored in a database. Only the School of Business and Economics of the University of Münster is allowed to access your stored thesis. The student agrees that his or her thesis may be stored and reproduced only for the purpose of plagiarism assessment. The first examiner of the thesis will be advised on the outcome of the plagiarism assessment.

Sanctions Each case of plagiarism constitutes an attempt to deceive in terms of the examination regulations and will lead to the thesis being graded as "failed". This will be communicated to the examination office where your case will be documented. In the event of a serious case of deception the examinee can be generally excluded from any further examination. This can lead to the exmatriculation of the student. Even after completion of the examination procedure and graduation from university, plagiarism can result in a withdrawal of the awarded academic degree.

We confirm that we have read and understood the information in this document. We agree to the outlined procedure for plagiarism assessment and potential sanctioning.

Münster, 4.12.2025

Signatures