

Edge-SAN: An Edge-Prompted Foundation Model for Accurate Nuclei Instance Segmentation in Histology Images

Xuening Wu¹, Yiqing Shen^{2*}, Qing Zhao¹, Yanlan Kang¹, Ruiqi Hu³, Wenqiang Zhang^{4*}

¹Shanghai Engineering Research Center of AI & Robotics, Academy for Engineering & Technology, Fudan University, Shanghai, China

Email: xueningwu@gmail.com, zhaoq19@fudan.edu.cn, 21110860035@m.fudan.edu.cn

²Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

Email: yshen92@jhu.edu

³Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong

Email: 23058044g@connect.polyu.hk

⁴Engineering Research Center of AI & Robotics, Ministry of Education,

Academy for Engineering & Technology, School of Computer Science, Fudan University, Shanghai, China

Email: wqzhang@fudan.edu.cn

Abstract—Accurate nuclei segmentation is fundamental in histology image analysis, playing an essential role in cancer grading and diagnosis. However, this task remains challenging due to variations in staining protocols, heterogeneity among nuclei types, and the densely clustered nature of nuclei. While SAM exhibits zero-shot generalization capabilities in natural image segmentation, its performance degrades when applied to nuclei segmentation in histology images. Existing adaptations of SAM for medical imaging primarily focus on organ or lesion segmentation, which differs substantially from nuclei segmentation due to the unique characteristics of nuclei—specifically, their sparse distribution combined with dense clustering. To address these challenges, we propose Edge-SAN (Segment Any Nuclei with Edge Prompting), an interactive segmentation foundation model specifically designed for nuclei segmentation. Edge-SAN introduces a novel edge prompting method that enhances the delineation of nuclei boundaries, particularly among densely clustered nuclei, by leveraging edge information to improve segmentation accuracy. We evaluate Edge-SAN on 12 diverse datasets in both few-shot and zero-shot scenarios, demonstrating its effectiveness as a foundation model for nuclei segmentation, achieving 66.81% AJI and 73.13% DSC—improvements of 16.33% and 15.48% over SAM-Med2D, respectively. The code is available at <https://github.com/deep-geo/Edge-SAN>.

Index Terms—Nuclei Segmentation, Deep Learning, Foundation Model, Segment Anything Model, Digital Pathology.

I. INTRODUCTION

Accurate instance segmentation of nuclei in digital pathology is fundamental for a range of downstream clinical analyses, including cancer diagnosis and prognosis, quantitative pathology, and biomarker analysis [1], [2]. However, this task remains challenging due to the variability in nuclei types, staining inconsistencies across different institutions, and the inherently clustered nature of nuclei [3], [4]. Existing instance

segmentation methods designed for natural images often struggle when applied to histopathological images, particularly in discriminating densely clustered nuclei. This results in imprecise segmentation of individual nuclei within clusters [4]. Fully supervised deep learning models have proven effective for nuclei segmentation, but they require extensive annotated datasets and often necessitate post-processing to isolate individual nuclei instances [5].

The recent Segment Anything Model (SAM) [6] has demonstrated zero-shot generalization capabilities in natural image segmentation, exhibiting robustness and adaptability across diverse context when provided with suitable user prompts. This potential has spurred interest in applying SAM to medical image segmentation [7]–[9]. However, adapting SAM for nuclei segmentation presents unique challenges distinct from broader medical image segmentation tasks. Unlike the larger, sparsely distributed anatomical structures common in medical images, nuclei segmentation involves smaller, densely packed objects.

Moreover, current adaptations of SAM in medical imaging rely heavily on precise manual prompts during inference, such as bounding boxes, which require considerable effort and expertise [9], [10], and may still capture unwanted regions. Additionally, multiple point prompts are often needed to achieve satisfactory segmentation results. Although SAM provides a prompt-free “automatic segment everything mode” mode, its performance is suboptimal for densely clustered nuclei. This reliance on manual prompting limits generalization across diverse nuclei types and disrupts clinical workflows, adding to the already intensive workload of pathologists. Consequently, there is a pressing need for improved prompting techniques tailored specifically to nuclei segmentation with SAM.

Previous work has emphasized the important role of both

*Corresponding authors: Yiqing Shen and Wenqiang Zhang

normal edges and clustered edges in nuclei segmentation [4], [11]. Recent advancements in multi-task learning, especially those focused on clustered nuclei edges [12], [13], underscore the potential of incorporating this specialized domain knowledge into SAM to enhance its nuclei instance segmentation capabilities.

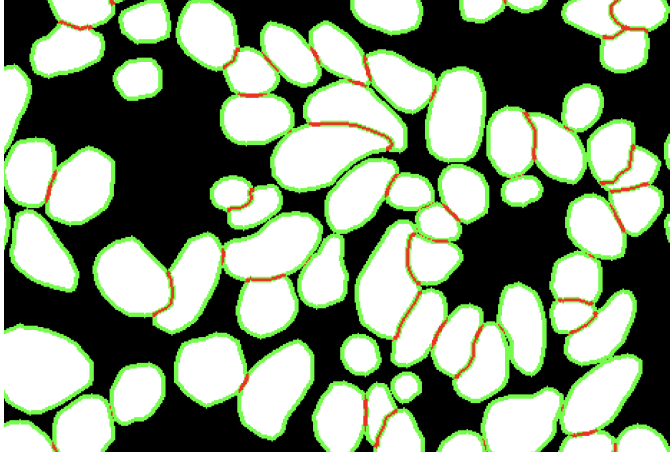


Fig. 1. The Cluster Edge Prompt is essential for accurate nuclei instance segmentation. By default, it automatically selects five points along the **clustered edges** of adjacent nuclei to ensure precise boundary delineation. When nuclei are not in contact, points are randomly selected from **normal edges**. This innovative approach enhances segmentation accuracy by focusing on critical boundary details, especially in densely clustered regions.

To overcome these limitations, we propose Segment Any Nuclei with Edge Prompting (Edge-SAN), as illustrated in Fig. 1. Our contributions are three-fold: Firstly, we introduce a novel prompting type, the Cluster Edge Prompt (CEP), specifically designed to enhance nuclei instance segmentation in histology images with clustered nuclei. The CEP augments SAM’s existing point, bounding box, and mask prompts by improving edge delineation, particularly in densely packed regions. For non-clustered nuclei, it functions as a normal edge prompt, placing random points along the nuclei edge and thereby enabling precise edge delineation across diverse scenarios. Secondly, we propose a novel two-stage training strategy that effectively adapts SAM-Med2D for nuclei segmentation. The first stage focuses on semantic segmentation, fine-tuning only the adapter layers in the image encoder and mask decoder. This allows the model to learn general nuclei features without the complexity of instance-level segmentation. The second stage incorporates the prompt encoder and trains for instance segmentation, integrating the CEP and other prompts to refine the model’s ability to distinguish individual nuclei, especially in clustered scenarios. Thirdly, to complement the CEP and two-stage training strategy, we curated a comprehensive dataset for nuclei segmentation by combining 11 public datasets, creating a unified resource of over 28,000 images containing more than 1.6 million nuclei instance masks. Leveraging this dataset, we fine-tuned Edge-SAN by adapting SAM-Med2D’s encoder, prompt encoder, and decoder. To facilitate further research in this field, we

have also made this unified dataset publicly available.

II. METHODS

A. Architecture Overview

As illustrated in Fig. 2, Edge-SAN builds on SAM-Med2D [14], a medical variant of SAM [6], to create a specialized foundation model for nuclei segmentation in histology images. SAM-Med2D, following the architecture of SAM, includes an image encoder based on a Vision Transformer (ViT) [15], a prompt encoder, and a mask decoder. The image encoder extracts features from input images, while the prompt encoder converts user inputs (*e.g.*, points, boxes) into embeddings. These embeddings are then combined with image features and a dense prompt in the mask decoder to produce the final segmentation mask. Edge-SAN leverages SAM-Med2D’s pre-trained image encoder and mask decoder, utilizing their domain-specific knowledge acquired from diverse medical images. To balance computational efficiency with the benefit of SAM’s pre-existing knowledge, we fine-tune only the adapter layer in the image encoder. The key innovation in Edge-SAN is the introduction of the Cluster Edge Prompt (CEP) to the prompt encoder, specially designed to address the challenge of identifying edges in densely clustered nuclei. The CEP works by placing key points along the edges of clustered nuclei, which can be manually annotated by experts or generated automatically using an edge detector. The CEP can be represented as a set of 2D coordinates:

$$\text{CEP} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (1)$$

where (x_i, y_i) represents the coordinates of each edge point, and n denotes the number of points. These points are then embedded into a high-dimensional space:

$$E_{\text{CEP}} = \text{PromptEncoder}(\text{Embed}(\text{CEP})) \quad (2)$$

where Embed is an embedding function that transforms the 2D coordinates into a representation suitable for the prompt encoder. High-dimensional embedding of the CEP points enables the prompt encoder to retain rich spatial information that is crucial for accurate boundary delineation in nuclei segmentation. This edge-centric prompt enhances Edge-SAN’s ability to differentiate individual nuclei within clusters, thereby improving instance segmentation performance. When nuclei are not clustered, the CEP functions as a normal edge prompt, still providing valuable edge information for accurate instance segmentation. This flexibility allows Edge-SAN to perform well across diverse nuclei configurations, both clustered and non-clustered.

B. Cluster Edge Prompts

Building on the architecture overview, we now delve into the specifics of the Cluster Edge Prompt (CEP) and its role in Edge-SAN. The CEP is designed to address the challenge of instance segmentation in densely clustered nuclei, a task that can be decomposed into three interrelated subtasks: nuclei segmentation, edge segmentation, and clustered edge segmentation [13]. Among these, the identification of clustered

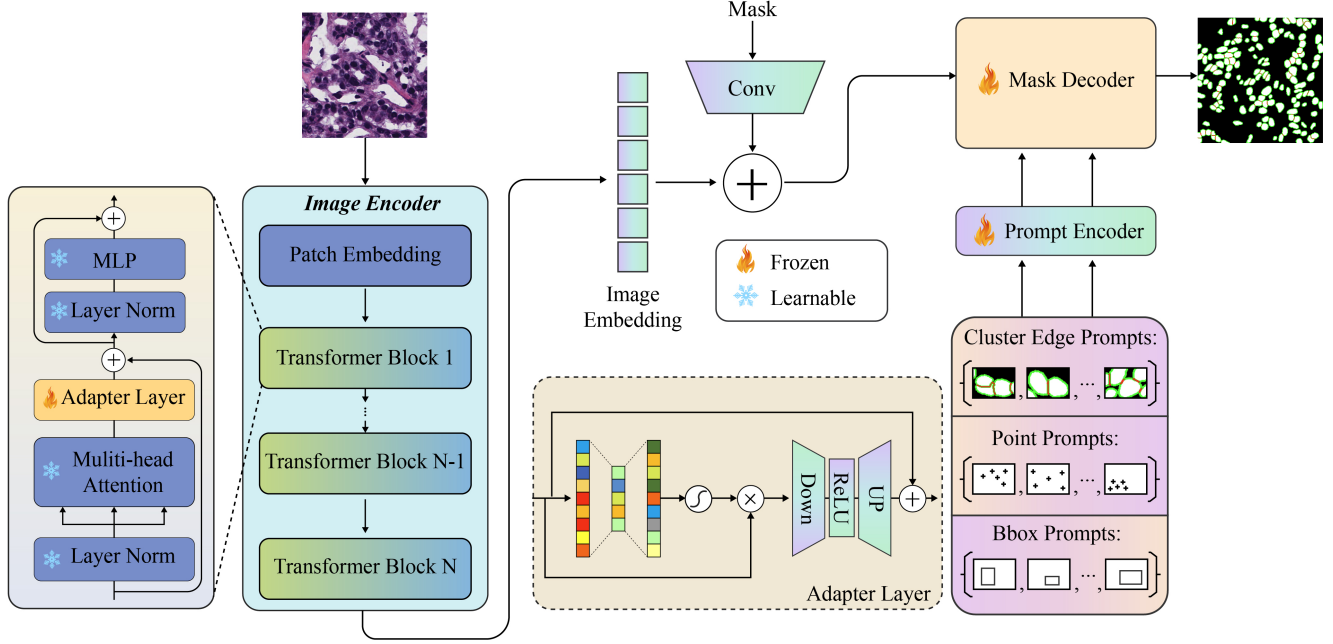


Fig. 2. The overall architecture of Edge-SAN for nuclei instance segmentation. Built on SAM-Med2D, Edge-SAN introduces a cluster edge prompt method to enhance the segmentation of densely clustered nuclei. The image encoder, featuring adaptive layers, processes the input histology image. The prompt encoder integrates the CEP alongside traditional point and bounding box prompts. The mask decoder combines image features and prompt embeddings to generate the final segmentation mask. Trainable components are highlighted in orange, while frozen pre-trained components are in blue and marked with snowflakes. The adapter layer in the image encoder and the entire prompt encoder are fine-tuned in a two-stage training process to optimize performance.

edges is important for accurately separating tightly packed nuclei [11], [13], [16]. The CEP is processed by the prompt encoder as described in Eq. (2), where resulting embeddings are concatenated with other sparse prompts (if available) to form the final sparse prompt embedding:

$$E_{sparse_prompt} = \text{Concat}(E_{CEP}, E_{point}, E_{bbox}) \quad (3)$$

This combined embedding guides the segmentation process in the mask decoder:

$$E = E_{img} + E_{dense_prompt} \quad (4)$$

$$M = \text{MaskDecoder}(E, E_{sparse_prompt}) \quad (5)$$

where E is the combined embedding, E_{img} represents the image embeddings from the image encoder, E_{dense_prompt} is the embedding of the dense mask, and M is the output segmentation mask.

The CEP enhances the prompt encoder by providing additional spatial context for delineating individual nuclei within clusters. When nuclei are not clustered, the CEP functions as a standard edge prompt. To streamline nuclei segmentation, we made several modifications to the original SAM-Med2D architecture. We disabled the text prompt, as it is not applicable in this context, and reduced the number of output masks from three to one, as nuclei images typically lack the complex

hierarchical structures found in natural images processed by SAM. These modifications can be represented as:

$$M = \text{SingleMaskDecoder}(E, E_{sparse_prompt}) \quad (6)$$

where SingleMaskDecoder is the modified mask decoder that outputs a single predicted mask, M .

C. Two-Stage Training Strategy

Edge-SAN employs a two-stage training strategy to effectively leverage the pre-trained SAM-Med2D model while adapting it for nuclei segmentation with CEP. This approach allows for gradual fine-tuning, initially focusing on nuclei semantic segmentation before progressing to the more challenging task of interactive nuclei instance segmentation. In the first stage, only the adapters in the image encoder and mask decoder are trained, framing the task as nuclei semantic segmentation. This stage enables the model to learn general nuclei features in histology images without the added complexity of instance-level segmentation.

In the second stage, we continue training the adapters in the image encoder and mask decoder, now incorporating the prompt encoder as well. At this stage, Edge-SAN is trained for instance segmentation, utilizing the cluster edge prompt along with other prompts. The combined loss \mathcal{L} includes an additional IoU loss term:

$$\mathcal{L} = \mathcal{L}_{sem} + \gamma \mathcal{L}_{IoU} \quad (7)$$

where \mathcal{L}_{sem} and \mathcal{L}_{IoU} represent the semantic segmentation loss and IoU loss, respectively, with the weight γ set to 1.0.

For both stages, \mathcal{L}_{sem} is a linear combination of the focal loss \mathcal{L}_{Focal} and the dice loss \mathcal{L}_{dice} :

$$\mathcal{L}_{sem} = \alpha \mathcal{L}_{Focal}(M, y) + \beta \mathcal{L}_{dice}(M, y) \quad (8)$$

$$\mathcal{L}_{IoU} = 1 - \frac{\text{Intersection}}{\text{Union}} = 1 - \frac{\sum(M \cdot y)}{\sum(M + y - M \cdot y)} \quad (9)$$

where y represents the ground truth, and α and β are set to 20.0 and 1.0, respectively, as in the SAM default. This two-stage approach allows Edge-SAN to first establish a foundation of general features related to nuclei and edge detection, which is then refined to achieve instance-level segmentation accuracy in the second stage. By incorporating the CEP and additional prompts in the second stage, Edge-SAN enhances its ability to accurately segment individual nuclei within clustered regions, making it well-suited for instance segmentation in histology.

III. EXPERIMENTS

A. Datasets

TABLE I
SUMMARY OF THE DATA SET DETAILS.

Index	Dataset	Dims	# of Images	# of Nuclei
1	Fluorescence [17]	2D	524	19,987
2	Histology [18] [16]	2D	462	43,143
3	Thyroid [16]	2D	487	12,464
4	CoNIC [19]	2D	4,981	569,861
5	CoNSeP [20]	2D	41	24,332
6	TNBC [21]	2D	50	4,022
7	GlandSeg [22]	2D	165	1,530
8	CPM15 [23]	2D	15	2,905
9	CPM17 [23]	2D	32	3,766
10	DynNuclearNet [24]	2D+time	7084	558,591*
11a	NucMM-Mouse [25]	3D	4,608	52,996*
11b	NucMM-Zebrafish [25]	3D	10,368	350,566*
Sum			28,817	1,644,163

To evaluate Edge-SAN, we compiled a comprehensive dataset from 11 distinct nuclei image collections, as detailed in Table I. This unified dataset includes a diverse array of tissue types, staining protocols, and imaging modalities, offering a robust foundation for training and evaluation.

- **Fluorescence Microscopy Dataset** [17]: This dataset integrates three data sources, capturing sample heterogeneity through 524 fluorescence images at a resolution of 512×512 pixels.¹
- **Histology Dataset**: This dataset combines the publicly available MoNuSAC dataset [18] with an in-house histology dataset from previous research [16], resulting in a collection of 462 histology images.²
- **Thyroid Dataset** [16]: This dataset includes 487 thyroid cytology image patches with 12,464 nuclei, each at a

resolution of 224×224 pixels. Derived from 98 needle biopsies scanned at 20x magnification ($0.5 \mu\text{m}/\text{pixel}$), these images were captured with a digital slide scanner using a 3CCD linear camera, achieving 50 nm localization accuracy. H&E staining enhances nuclei cluster visualization for detailed analysis.

- **CoNIC** [19]: The Colon Nuclei Identification and Counting dataset also known as Lizard, includes 4,981 colon histology image patches, each 256×256 pixels, annotated for epithelial and connective tissue nuclei. It provides RGB images with detailed instance segmentation and classification maps, supporting research on the colonic tumor microenvironment.
- **CoNSeP** [20]: The Colorectal Nuclei Segmentation and Phenotype (CoNSeP) dataset comprises 41 H&E-stained image tiles, each sized $1,000 \times 1,000$ pixels at 40x magnification. These images were extracted from 16 whole slide images (WSIs) of colorectal adenocarcinoma, each from a different patient, and scanned with the Omnyx VL120 scanner.
- **TNBC** [21]: The TNBC dataset includes a diverse range of annotated cell types, including normal epithelial and myoepithelial breast cells, invasive carcinomatous cells, fibroblasts, endothelial cells, adipocytes, macrophages, and inflammatory cells like lymphocytes and plasmocytes. It comprises 50 images with a total of 4,022 annotated cells. Cell counts per image range from 5 to 293 (mean: 80, SD: 58). Annotation was conducted by an expert pathologist and two research fellows, with discrepancies resolved through consensus.
- **GlandSeg** [22]: This dataset, designed for colorectal cancer research, consists of 165 BMP images captured with a Zeiss MIRAX MIDI scanner at 20x magnification ($0.62 \mu\text{m}/\text{pixel}$). It is divided into training and test sets: 37 benign and 48 malignant images for training, and 37 benign and 43 malignant images for testing.
- **CPM15** [23]: The CPM15 dataset comprises 15 images featuring two cancer types, with a total of 2,905 cell nuclei. Images were scanned at 40x and 20x magnifications, with resolutions of 400×400 and 1000×600 pixels.
- **CPM17** [23]: The CPM17 dataset comprises 64 images covering four types of brain cancer: non-small cell lung cancer (NSCLC), head and neck squamous cell carcinoma (HNSCC), glioblastoma multiforme (GBM), and lower-grade glioma (LGG). It contains 7,570 cell nuclei with resolutions of 500×500 or 600×600 pixels, scanned at 40x and 20x magnifications.
- **DynamicNuclearNet** [24]: We use the segmentation subset of the DynamicNuclearNet dataset, a 2D+time series with approximately 600,000 unique nuclear segmentations. A 10% sample is taken across the timeline supports downstream analyses, including signaling reporter quantification and nuclear morphology assessment.
- **Zebrafish** [25]: This dataset consists of a 0.1 mm^3 scanning electron microscopy (SEM) volume of zebrafish brain, containing approximately 170,000 nuclei. The

¹<https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BSST265>

²<https://github.com/lu-yizhou/ClusterSeg>

TABLE II

PERFORMANCE COMPARISON ON THE NUCLEI INSTANCE SEGMENTATION, INCLUDING OUR PROPOSED EDGE-SAN, ON SIX DIVERSE HISTOLOGY DATASETS: CoNIC, TNBC, HISTOLOGY, THYROID, CPM17, AND GLANDSEG. EDGE-SAN CONSISTENTLY OUTPERFORMS EXISTING METHODS ACROSS ALL DATASETS AND METRICS, DEMONSTRATING ITS EFFECTIVENESS AND ROBUSTNESS IN NUCLEI INSTANCE SEGMENTATION (%).

Methods	CoNIC				TNBC				Histology			
	AJI	DQ	SQ	PQ	AJI	DQ	SQ	PQ	AJI	DQ	SQ	PQ
U-Net [26]	41.79	74.95	76.33	57.21	54.96	81.82	76.10	62.27	38.65	69.70	74.54	51.95
Mask-RCNN [27]	45.58	57.46	57.20	32.87	52.68	82.35	63.03	51.90	60.46	92.49	63.08	58.34
HoVer-Net [20]	59.69	86.90	61.79	53.69	63.79	75.00	66.47	49.85	70.54	98.90	69.98	69.21
Vanilla SAM [6]	39.07	37.32	61.26	22.87	49.74	63.60	64.66	41.12	56.12	75.68	65.91	49.88
SAM-Med2D [14]	36.99	56.96	62.09	35.36	56.68	79.55	62.62	53.79	56.32	85.25	69.82	59.53
MedSAM [28]	60.93	86.54	66.34	57.41	68.90	91.63	71.41	65.43	69.59	95.22	71.21	67.80
Med-SA [29]	58.03	87.64	64.24	56.30	62.57	86.89	66.78	58.02	65.61	93.09	71.00	66.10
Edge-SAN (ours)	78.16	97.83	78.98	77.26	79.67	98.77	78.58	77.61	84.05	99.34	83.23	82.68
	Thyroid				CPM17				GlandSeg			
	AJI	DQ	SQ	PQ	AJI	DQ	SQ	PQ	AJI	DQ	SQ	PQ
U-Net [26]	48.93	84.38	84.63	69.49	47.04	68.40	75.57	50.45	37.10	65.02	83.51	52.11
Mask-RCNN [27]	82.31	99.49	82.71	82.28	57.81	72.73	66.19	48.14	69.71	93.55	74.79	69.96
HoVer-Net [20]	84.50	99.99	84.39	84.39	70.01	96.19	67.96	67.96	30.19	40.00	67.09	26.83
Vanilla SAM [6]	67.39	94.34	70.02	66.06	54.17	59.15	67.91	40.17	67.20	95.62	77.76	74.35
SAM-Med2D [14]	70.77	94.07	74.22	69.82	56.99	64.90	68.51	44.47	58.61	90.84	71.78	65.20
MedSAM [28]	71.51	97.73	72.09	70.45	66.93	80.94	69.39	56.17	85.49	99.62	89.31	88.97
Med-SA [29]	74.40	96.58	77.66	75.00	72.77	92.80	74.65	69.27	56.23	88.39	78.94	69.78
Edge-SAN (ours)	90.72	99.99	90.07	90.07	79.03	96.50	76.02	73.36	89.16	97.86	93.47	91.46

SEM captures details at a resolution of $0.48 \times 0.51 \times 0.51 \mu\text{m}$, spanning a total volume 0.14 mm^3 . We extracted 2D slices from the 3D dataset, spaced 10 pixels apart.

We preprocessed the 11 supervised datasets using OpenCV for normalization. For semantic segmentation datasets lacking instance-level annotations, nuclei instance information was extracted by detecting edges with OpenCV. All images were resized or cropped to a uniform resolution of 256×256 pixels for consistency across datasets. Input images are in RGB format, with pixel values normalized to the range $[0, 1]$. Annotation pixel values range from 0 to 65,535, with each distinct value representing a unique nucleus instance. To improve data loading efficiency, all annotations were converted to NumPy array format. Fig. 3 displays representative samples and processed masks generated by different models.

B. Evaluation Metrics

We use the following metrics for evaluation [30]. For instance segmentation, we measure performance with the Aggregated Jaccard Index (AJI), Detection Quality (DQ), Segmentation Quality (SQ), and Panoptic Quality (PQ). For nuclei semantic segmentation, we assess performance using the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision, and Recall. Higher values in these metrics indicate improved segmentation quality.

C. Implementation Details

All experiments were conducted using PyTorch on a single NVIDIA RTX 4090 GPU with 24GB memory, with a random seed set to 42. All images were standardized to a size of 256×256 pixels. The unified dataset was divided into training (90%)

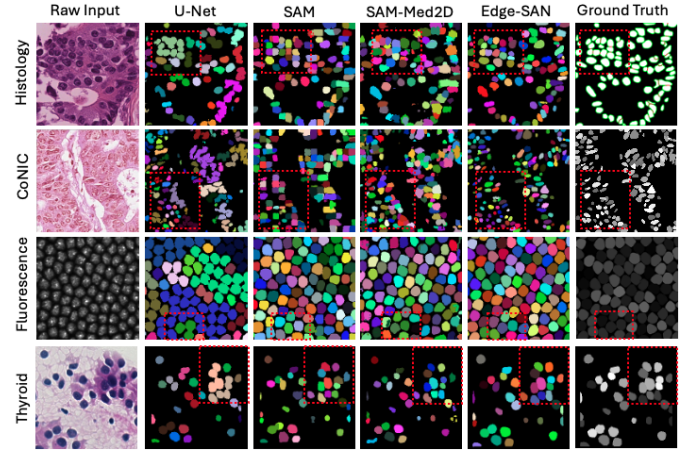


Fig. 3. Comparison of model outputs, highlighting Edge-SAN's improved performance in accurately delineating edges of densely clustered nuclei.

and test (10%) subsets. In the first training stage, we fine-tuned the adapters in SAM-Med2D's encoder and decoder on the unified dataset, initializing from the ViT-b checkpoint. The second stage introduced the edge prompt, continuing from the checkpoint saved at the end of the first stage. Throughout both stages, the text prompt and *multimask* options were disabled. We employed the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 16.

D. Instance Segmentation Performance Analysis

The performance comparisons presented in Table II and Fig. 4 demonstrate the exceptional performance of our proposed Edge-SAN across diverse histology datasets. We analyze these

TABLE III
PERFORMANCE COMPARISON ON THE NUCLEI SEMANTIC SEGMENTATION (%).

Methods	CoNIC				TNBC				Histology			
	DSC	IoU	Prec.	Recall	DSC	IoU	Prec.	Recall	DSC	IoU	Prec.	Recall
U-Net [26]	63.81	54.17	65.19	70.83	71.73	61.06	72.11	79.59	58.01	48.47	61.23	64.15
Mask-RCNN [27]	63.01	46.86	90.51	49.57	67.19	52.98	81.49	62.80	75.28	60.79	91.30	65.04
HoVer-Net [20]	70.14	55.60	77.97	65.74	67.77	53.19	71.35	65.82	81.70	69.45	81.62	83.44
Vanilla SAM [6]	48.10	35.03	41.53	71.58	59.17	45.88	52.26	78.73	68.01	54.13	62.32	82.04
SAM-Med2D [14]	58.37	43.92	46.52	88.47	56.68	70.60	58.88	94.54	73.50	60.74	65.82	88.72
MedSAM [28]	73.95	60.08	67.25	87.15	79.20	66.92	74.20	88.93	80.61	68.53	83.69	80.21
Med-SA [29]	72.85	58.60	70.36	79.81	73.30	59.72	69.19	81.38	78.57	66.34	77.97	82.48
Edge-SAN (ours)	85.47	76.53	87.26	84.80	86.80	77.57	87.35	87.36	90.10	82.61	92.12	88.75
	Thyroid				CPM17				GlandSeg			
	DSC	IoU	Prec.	Recall	DSC	IoU	Prec.	Recall	DSC	IoU	Prec.	Recall
U-Net [26]	78.09	68.75	75.67	89.46	62.18	50.93	60.12	74.78	49.79	43.30	58.63	52.61
Mask-RCNN [27]	90.09	82.31	96.62	84.82	72.32	57.55	89.53	61.68	81.08	70.02	78.35	88.22
HoVer-Net [20]	91.41	84.39	94.04	89.53	80.74	67.96	82.04	79.86	40.39	29.21	93.76	30.00
Vanilla SAM [6]	79.16	66.76	82.76	79.08	57.30	44.58	50.96	77.72	84.59	74.71	79.83	93.18
SAM-Med2D [14]	81.56	70.31	84.41	81.49	63.32	49.64	52.38	88.86	76.72	65.07	74.18	82.57
MedSAM [28]	82.50	70.84	92.18	75.95	71.94	58.72	67.27	82.60	93.87	88.90	94.08	94.23
Med-SA [29]	84.62	74.69	82.46	88.50	81.00	69.70	79.05	86.12	77.48	67.92	83.06	80.39
Edge-SAN (ours)	94.70	90.07	95.14	94.46	82.79	72.74	82.85	84.34	94.71	91.12	95.34	94.55

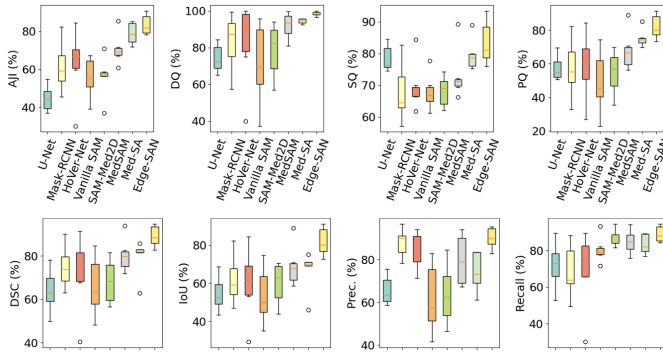


Fig. 4. Edge-SAN outperforms all other models in both instance and semantic segmentation metrics.

results in the context of nuclei instance segmentation, comparing Edge-SAN with leading methods including traditional deep learning approaches and recent SAM-based models. Edge-SAN consistently outperforms all other methods across the six evaluated datasets: CoNIC, TNBC, Histology, Thyroid, CPM17, and GlandSeg. This performance is particularly noteworthy given the varied nature of these datasets, which represent different tissue types and staining protocols, underscoring the robustness of our approach. In terms of the AJI, Edge-SAN yields significant improvements. For instance, on the challenging CoNIC dataset, Edge-SAN attains an AJI of 78.16%, surpassing the next best method (MedSAM) by 17.23%. This pattern is consistent across all datasets. The DQ scores further underscore Edge-SAN's effectiveness in accurately identifying individual nuclei. Our model achieves near-perfect DQ scores across several datasets, with 99.34% for Histology, indicating its high accuracy in locating nuclei

instances. SQ results reflect Edge-SAN's proficiency in delineating precise nuclear boundaries. The model shows notable improvements, particularly evident in the GlandSeg dataset where it achieves an SQ of 93.47%, outperforming other methods. The PQ metric, which integrates both detection and segmentation accuracy, further emphasizes Edge-SAN's balanced performance. For instance, in the GlandSeg dataset, Edge-SAN achieves a PQ of 91.46%, surpassing the next best method by 2.49%. Notably, Edge-SAN demonstrates substantial gains over its predecessors in the SAM family, including Vanilla SAM, SAM-Med2D, Med-SAM and Med-SA. This highlights the effectiveness of our edge prompting strategy and two-stage training approach in adapting SAM for nuclei segmentation tasks.

E. Semantic Segmentation Performance Analysis

The results presented in Table III and Fig. 4 demonstrate Edge-SAN's well-balanced performance in nuclei semantic segmentation across diverse histology datasets. Edge-SAN achieves the highest DSC scores across all datasets, with notable gains in challenging cases. For instance, in the CoNIC dataset, Edge-SAN attains a DSC of 85.47%, surpassing the next best method (MedSAM) by 11.53%. This consistent improvement trend is observed across datasets, with the largest gain in the Thyroid dataset, where Edge-SAN outperforms the next best method by 12.20%. The IoU scores further underscore Edge-SAN's effectiveness, with our model achieving the highest scores across all datasets. In terms of Precision and Recall, Edge-SAN demonstrates a well-balanced performance, achieving the highest Precision across four of the six datasets while maintaining competitive Recall scores. While SAM-Med2D occasionally records the highest Recall, it often exhibits significantly lower Precision compared to

Edge-SAN. For example, on the CoNIC dataset, although SAM-Med2D achieves a slightly higher Recall (88.47% vs. 84.80%), its Precision is substantially lower (46.52% vs. 87.26%), underscoring Edge-SAN’s more balanced and robust segmentation performance.

TABLE IV
PERFORMANCE COMPARISON OF ZERO-SHOT INSTANCE SEGMENTATION ON THE MoNuSAC DATASET (%).

Methods	AJI	DQ	SQ	PQ
Mask-RCNN [27]	39.81	55.56	65.80	36.43
HoVer-Net [20]	44.76	57.82	62.43	36.10
U-Net [26]	44.58	74.62	71.55	54.88
Vanilla SAM [6]	44.72	36.76	63.89	23.48
SAM Med2D [14]	50.48	59.42	67.80	40.29
Med-SA [29]	52.38	77.84	62.74	48.84
MedSAM [28]	66.69	84.81	69.15	58.65
Edge-SAN (ours)	66.81	88.84	73.09	64.94

TABLE V
PERFORMANCE COMPARISON OF ZERO-SHOT SEMANTIC SEGMENTATION ON THE MoNuSAC DATASET (%).

Methods	DSC	IoU	Prec.	Recall	F1
Mask-RCNN [27]	55.97	42.28	83.42	50.24	62.72
HoVer-Net [20]	55.24	41.64	75.51	50.58	60.49
U-Net [26]	63.80	53.14	66.35	68.98	67.64
Vanilla SAM [6]	39.41	29.14	39.08	22.52	28.31
SAM-Med2D [14]	57.65	44.50	51.96	87.58	65.16
Med-SA [29]	68.11	53.37	70.77	69.87	68.11
MedSAM [28]	72.67	60.52	69.13	83.77	73.67
Edge-SAN (ours)	73.13	65.52	82.62	79.93	81.26

F. Zero-shot Evaluation

The zero-shot instance and semantic segmentation results on the MoNuSAC 2020 dataset [18], summarized in Tables IV and V, illustrate the advancements Edge-SAN achieves over existing methods. The MoNuSAC 2020 dataset was intentionally excluded from training to assess Edge-SAN’s zero-shot capabilities, ensuring no prior exposure for any model. This setup underscores Edge-SAN’s effectiveness in performing robustly without fine-tuning on the MoNuSAC dataset. In instance segmentation (Table IV), Edge-SAN surpasses SAM and its variants, achieving higher scores in AJI (66.81%), DQ (88.84%), SQ (73.09%), and PQ (64.94%), outperforming models that generally rely on fine-tuning and struggle in zero-shot scenarios. Similarly, in semantic segmentation (Table V), where traditional methods like HoVer-Net and Vanilla SAM exhibit limited effectiveness, Edge-SAN demonstrates notable gains with DSC (73.13%), IoU (65.52%), and Precision (82.62%). Although Edge-SAN’s Recall is lower than that of SAM-Med2D, its overall performance is stronger, reflected in a higher F1-score (81.26% vs. 65.16%) as previously discussed. This zero-shot evaluation underscores Edge-SAN’s robustness and ability to handle nuclei imaging tasks without

prior task-specific data annotations, showcasing its potential in both instance and semantic segmentation applications.

G. Ablation Study

TABLE VI
ABLATION STUDY ON THE IMPACT OF CLUSTER EDGE POINT PROMPTS ON EDGE-SAN PERFORMANCE (0 POINTS INDICATES CEP IS DISABLED). METRICS FOR INSTANCE SEGMENTATION (AJI, DQ, SQ, PQ) AND SEMANTIC SEGMENTATION (DSC, IoU, PRECISION, RECALL) SHOW CONSISTENT IMPROVEMENTS WITH INCREASING CEP POINTS.

# of CEP Points	0	1	2	3	4	5
AJI	75.51	79.54	80.19	81.53	80.43	81.70
DQ	95.98	96.92	97.08	97.26	97.47	97.62
SQ	77.26	79.20	80.31	80.34	81.22	81.56
PQ	74.16	76.76	77.96	78.14	79.17	79.62
DSC	82.90	84.71	85.35	85.60	86.30	86.65
IoU	73.28	75.83	76.90	77.16	78.18	78.66
Prec.	83.13	84.68	86.28	85.04	85.70	87.51
Recall	84.38	86.01	85.48	87.28	87.90	86.73

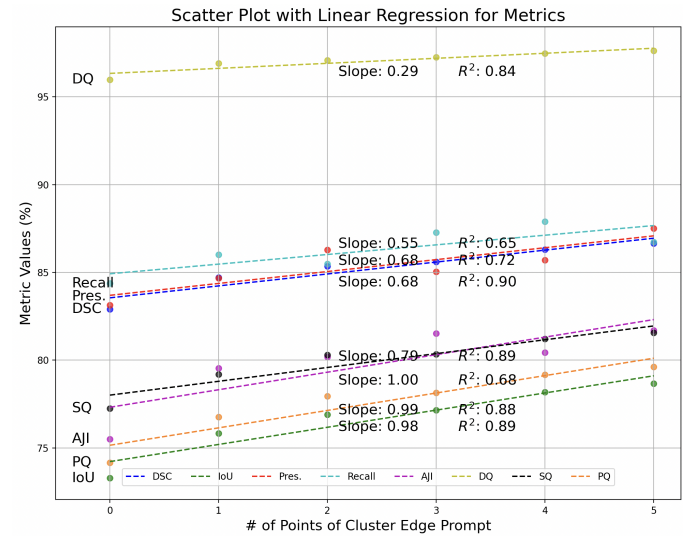


Fig. 5. Linear regression analysis of Edge-SAN performance metrics relative to the number of CEP points. This scatter plot illustrates the trends in instance segmentation metrics (AJI, DQ, SQ, PQ) and semantic segmentation metrics (DSC, IoU, Precision, Recall) as the number of CEP points increases from 0 to 5. The slopes and R^2 values are shown for each metric, underscoring a strong positive correlation (with $R^2 > 0.7$) between the number of CEP points and segmentation performance.

To evaluate the impact of CEPs on Edge-SAN’s performance, we conducted an ablation study varying the number of CEP points from 0 to 5. Table VI presents the detailed results across various metrics for both instance and semantic segmentation tasks. The findings reveal a clear positive correlation between the number of CEP points and segmentation performance. For instance segmentation, as we increase CEP points from 0 to 5, notable improvements are observed: AJI increases from 75.51% to 81.70%, DQ improves from 95.98% to 97.62%, SQ rises from 77.26% to 81.56%, and PQ enhances from 74.16% to 79.62%. A similar trend is evident in semantic segmentation metrics. The DSC increases from 82.90% to

86.65%, IoU improves from 73.28% to 78.66%, precision rises from 83.13% to 87.51%, and recall enhances from 84.38% to 86.73%.

Fig. 5 further illustrates this trend through a linear regression analysis, showing strong positive correlations with R^2 values exceeding 0.7 for most metrics. This consistent performance enhancement with additional CEP points is particularly notable for metrics like AJI, DSC, and IoU, which display steeper slopes, indicating a higher sensitivity to the number of CEP points. The most substantial gains occur with the introduction of the first CEP point, followed by continued gradual enhancements as additional points are added. This suggests that even a single CEP point significantly boosts the model's ability to delineate nuclei boundaries, especially in clustered regions.

IV. CONCLUSION

In this study, we introduced Edge-SAN, a foundation model for nuclei instance segmentation that incorporates a novel Cluster Edge Prompt (CEP) to enhance segmentation accuracy by leveraging domain-specific knowledge. This innovation substantially improves SAM's ability to accurately delineate clustered nuclei, achieving state-of-the-art performance across 11 histology datasets. By delivering high accuracy with reduced annotation demands, Edge-SAN marks a substantial step forward in foundation model-based nuclei segmentation, making it well-suited for both research and clinical applications. Future optimizations, such as streamlined training and inference procedures with knowledge distillation, could further support Edge-SAN's deployment in resource-constrained settings or real-time scenarios.

REFERENCES

- [1] N. R. Gudhe, V.-M. Kosma, H. Behravan, and A. Mannermaa, "Nuclei instance segmentation from histopathology images using bayesian dropout based deep learning," *BMC Medical Imaging*, vol. 23, no. 1, p. 162, 2023.
- [2] A. H. Song, G. Jaume, D. F. Williamson, M. Y. Lu, A. Vaidya, T. R. Miller, and F. Mahmood, "Artificial intelligence for digital and computational pathology," *Nature Reviews Bioengineering*, vol. 1, no. 12, pp. 930–949, 2023.
- [3] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P.-A. Heng, J. Li, Z. Hu *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1380–1391, 2019.
- [4] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "Dcan: Deep contour-aware networks for accurate gland segmentation," *Cornell University - arXiv, Cornell University - arXiv*, Apr 2016.
- [5] Z. Chen, Q. Xu, X. Liu, and Y. Yuan, "Un-sam: Universal prompt-free segmentation for generalized nuclei images," *arXiv preprint arXiv:2402.16663*, 2024.
- [6] A. Kirillov and *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.
- [7] G. Deng *et al.*, "Sam-u: Multi-box prompts triggered uncertainty estimation for reliable sam in medical image," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 368–377.
- [8] K. Zhang and D. Liu, "Customized segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.13785*, 2023.
- [9] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, p. 654, 2024.
- [10] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen *et al.*, "Segment anything model for medical images?" *Medical Image Analysis*, vol. 92, p. 103061, 2024.
- [11] J. Huang, Y. Shen, D. Shen, and J. Ke, "Ca 2.5-net nuclei segmentation framework with a microscopy cell benchmark collection," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. Springer, 2021, pp. 445–454.
- [12] X. Wu, Y. Shen, Q. Zhao, Y. Kang, and W. Zhang, "Moe-nuseg: Enhancing nuclei segmentation in histology images with a two-stage mixture of experts network," *Alexandria Engineering Journal*, vol. 110, pp. 557–566, 2025.
- [13] Z. He, M. Unberath, J. Ke, and Y. Shen, "Transnuseg: A lightweight multi-task transformer for nuclei segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 206–215.
- [14] J. Cheng and *et al.*, "Sam-med2d," *arXiv preprint arXiv:2308.16184*, 2023.
- [15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [16] J. Ke, Y. Lu, Y. Shen, J. Zhu, Y. Zhou, J. Huang, J. Yao, X. Liang, Y. Guo, Z. Wei *et al.*, "Clusterseg: A crowd cluster pinpointed nucleus segmentation framework with cross-modality datasets," *Medical Image Analysis*, vol. 85, p. 102758, 2023.
- [17] F. Kromp and *et al.*, "An annotated fluorescence image dataset for training nuclear segmentation methods," *Scientific Data*, vol. 7, no. 1, p. 262, 2020.
- [18] R. Verma, N. Kumar, A. Patil, N. C. Kurian, S. Rane, S. Graham, Q. D. Vu, M. Zwager, S. E. A. Raza, N. Rajpoot *et al.*, "Monusac2020: A multi-organ nuclei segmentation and classification challenge," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3413–3423, 2021.
- [19] S. Graham, Q. D. Vu, M. Jahanifar, M. Weigert, U. Schmidt, W. Zhang, J. Zhang, S. Yang, J. Xiang, X. Wang *et al.*, "Conic challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting," *Medical image analysis*, vol. 92, p. 103047, 2024.
- [20] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical image analysis*, vol. 58, p. 101563, 2019.
- [21] P. Naylor, M. Laé, F. Reyat, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 448–459, 2018.
- [22] K. Sirinukunwattana and *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," *Medical image analysis*, vol. 35, pp. 489–502, 2017.
- [23] Q. D. Vu and *et al.*, "Methods for segmentation and classification of digital microscopy tissue images," *Frontiers in bioengineering and biotechnology*, vol. 7, p. 433738, 2019.
- [24] E. Moen and *et al.*, "Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning," *Biorxiv*, p. 803205, 2019.
- [25] Z. Lin and *et al.*, "Nucmm dataset: 3d neuronal nuclei instance segmentation at sub-cubic millimeter scale," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer, 2021, pp. 164–174.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science, Lecture Notes in Computer Science*, Jan 2015.
- [27] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [28] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, p. 654, 2024.
- [29] J. Wu, W. Ji, Y. Liu, H. Fu, M. Xu, Y. Xu, and Y. Jin, "Medical sam adapter: Adapting segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.12620*, 2023.
- [30] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022.