Pakistan Journal of Scientific Research
Multidisciplinary & Peer Reviewed
Volume: 4, Number: 2(Suppl.), Pages: 92-101, Year:2025

Article

MASKYLO: Hybrid Deep Learning Framework for Detection and Segmentation of HE Stained Histology Image

Nimra Bukhari[1], Hassan Munir[1], Rafique Haider[2], Shabir Hussain[1,3*]

[1] Department of Computer Science, National College of Business Administration and Economics, Rahim Yar Khan, 64200, Pakistan
[2] Department of Computer Science, Khawaja Fareed University of Engineering & Information Technology, Rahim Yar Khan, Punjab, Pakistan
[3] Institute of Biopharmaceutical and Health Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

* Correspondence: Shabir Hussain (e-mail shabir.nicaas@gmail.com; shabir.hussain@sz.tsinghua.edu.cn)

Submitted: 02-03-2025**, Revised:** 20-05-2025, **Accepted:** 23-06-2025

Abstract

The analysis of pathology images is a crucial step of modern healthcare that helps in diagnosing tumours, determining the grade and type of tumours, planning to treat them and performing surgeries. In this study, we present MASKYLO, a novel hybrid deep learning framework that integrates YOLOv11 for real-time nuclei detection and using YOLOv8 and Mask R-CNN for precise pixel-level segmentation. MASKYLO ensures the speed and accuracy of YOLOv11 for detecting regions of interest, while the fused segmentation networks refine mask predictions to achieve high spatial fidelity. We evaluate our approach on the NuInsSeg dataset, demonstrating superior performance compared to existing methods. Our YOLOv11 detection branch achieves a precision of 0.89, a recall of 0.85, and an mAP@50 of 0.88, whereas the segmentation branch attains an IoU of 0.82 and a Dice score of 0.85, surpassing previously reported YOLOv8 and Mask R-CNN benchmarks with a 0.96 Dice Score and an IoU on the same dataset. These results highlight MASKYLO's ability to provide accurate, efficient, and reliable detection and segmentation of histopathological images, making it a promising tool for automated pathology workflows.

Keywords: HE image segmentation, hybrid detection-segmentation, medical image analysis, deep learning, pixel-level segmentation, histopathology.

1. Introduction

Digital Pathology [1] plays a dominant role in the diagnosis, prediction and grading of cancer disease by providing useful insights into cellular and tissue morphology. Traditionally, this technique has depended on visual inspection of glass slides under a microscope. But after the development of digital pathology, in which tissue slides are converted into ultra-high-resolution, H&E (or similar staining) preparations and whole-slide images (WSIs), this signifies a major shift in modern diagnostic workflows.

The primary method is WSI enabled through a hardware framework of robotic components, which converts the whole slide into a structured representation of spatial (x, y) and depth (z) pixel values. Pathologists can manipulate the digital slide through seamless zooming, panning, and region tiling across variable magnifications, facilitating histological view and diagnostic evaluation in a manner similar to glass-slide practice but augmented by 3D spatial image integrity and

navigation speed. In support of image acquisition, advanced image analysis algorithms [2], driven by neural nets and machine-learning pipelines, examine digital H&E stacks to segment and quantify histological context. Automated algorithms can save and measure nuclei extract quantitative descriptors of cytological architecture and tissue organization, thus minimizing observer-dependent variability. In the last decade, H&E image analysis has rapidly grown and is delivering high-end solutions for segmentation [3] with the help of deep learning techniques [4]. From fully convolutional neural network [5] to more refined and well-designed architectures for classification [6], segmentation [7], detection [8].H&E images are well-suited for computational stain normalization and transformation, their rich diagnostic and prognostic value allows AI models to identify fine-grained spatial and structural features which performs accurate classification, segmentation, and detection often without necessitating extra staining protocols. Digital pathology workflows became cost-effective and scalable automation of image analysis, which helped to reduce both human and laboratory resources. These reasons make H&E images the foundation for AI-driven applications [9] in digital pathology [10] and histopathology research. Pathology image analysis can be addressed at multiple levels, from the nuclear level to the cell level and tissue level features. At the cell level, along with their spatial relationships, the nuclei and cytoplasm of cells highlight crucial information about tumour type and grade. Moving forward to the tissue scale macroscale features, including the spatial arrangement and variations in tumour, provides significant diagnostic and prognostic information. Hematoxylin and eosin (H&E) images are the preferred choice for training AI models in pathology diagnostics due to their reliability, standardization, and information-rich content.

Model's contribution

- We proposed a novel approach that integrates detection and segmentation models (YOLOv11, YOLOv8-seg, and Mask R-CNN via Detectron2), where each component can be updated or replaced independently without retraining the entire system.
- The framework combines object detection (YOLOv11) with pixel-level segmentation (YOLOv8-seg and Mask R-CNN) to resolve overlaps and segment small or complex nuclei in histopathology.
- The framework enables real-world applicability by supporting whole-slide analysis through a patching strategy for full-resolution processing.
- We propose a fusion mechanism combining model outputs through IoU voting or confidence weighting, adaptable to varied needs.

2. Related Work

The segmentation of medical images assists in marking the boundaries of affected features and disease. Image enhancement by focusing on methodologies, different datasets, and exceptional results in the analysis of H&E images, deep learning techniques are now considered more suitable than microscopic methods. Initial findings in semantic segmentation were introduced by Fully Convolutional Networks (FCNs) [11], which improved classification models such as GoogLeNet, AlexNet and VGGNet for solid prediction. Building on FCNs, U-Net became the most widely adopted architecture in biomedical segmentation [12], introducing skip connections to counteract information loss from down-sampling. To address U-Net's limitations in capturing fine structures such as retinal vessels, a Deep Guidance Network (DGN) [13] introduced a guided image filter module, enabling superior recovery of curvilinear structures. A deep learning–based interactive segmentation framework combined initial CNN outputs with user interactions, modelled through geodesic distance transforms [14] and integrated into a resolution-preserving network with Conditional Random Fields (CRFs). Validations on 2D placenta (fetal MRI) and 3D brain tumour (FLAIR) datasets. An ovarian cancer segmentation model fine-tuned from a breast cancer model [15] achieved an IoU of 0.74, a recall of 0.86, and a precision of 0.84, underscoring the efficiency of transfer learning for histopathology. Automated cancer detection in H&E WSIs [16] is challenged by the giga-pixel scale and scarce annotations. NAS-UNet [17], discovered via neural

architecture search, outperformed U-Net variants on MRI, CT, and ultrasound segmentation with only 0.8M parameters. Deep learning has advanced image reconstruction, where a GAN-based super-resolution framework with SE-embedded EDSR [18] outperformed SRGAN, EDSR, VDSR, and D-DBPN under high upscaling factors. In medical image colorization, a DNN with Y-loss and adaptive reference search [19] achieved 24% PSNR and 47% SSIM gains over baselines, producing clinically validated, realistic outputs. These advances highlight deep learning's potential to improve diagnostic interpretation through enhanced image quality and color.

3. Methodology

This research proposes a hybrid segmentation framework that integrates three state-of-the-art models for H&E image detection, YOLOv8 for segmentation, and Mask R-CNN for refinement into a combined multi-stage pipeline.

3.1. Dataset

In this study we worked on the NuInsSeg [20] dataset which has large-scale collection of fully manually annotated nuclei in hematoxylin and eosin (H&E)-stained histopathology images. The dataset shown in Fig.A consists of 665 image patches of 512x512 pixels in size, having more than 30,000 segmented nuclei from 31 human and mouse organs. This dataset is particularly useful for estimating robust nuclei segmentation methods.

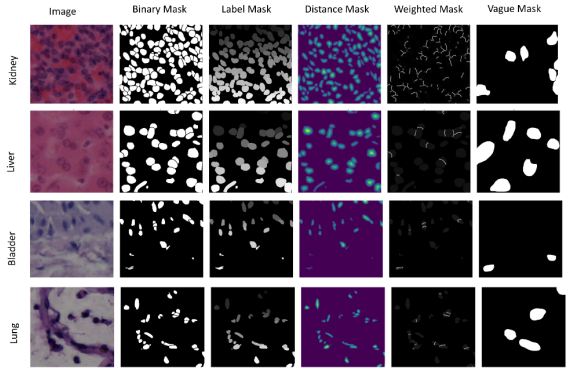

Figure A: Sample H&E patches with manually segmented nuclei masks from the NuInsSeg dataset

3.2. *Pre-Processing*

All H&E image patches from the NuInsSeg dataset were resized to 512 × 512 pixels to standardize the input across both models. To address stain variability, color normalization was applied prior to training. The nuclei masks were then reformatted YOLO compatible with nuclei boundary coordinates for YOLOv8 and COCO-style JSON annotations for Detectron2. The dataset was split into training and test sets, confirming a balanced distribution of organs and tissue types in each category.

3.3. Proposed Framework

The proposed framework is designed to perform robust segmentation of Hematoxylin and Eosin (H&E)-stained histopathological images by leveraging both classical and deep learning-based approaches. The methodology is divided into sequential yet independent stages, starting with preprocessing and annotation, followed by segmentation using YOLOv8 and Mask R-CNN, and concluding with evaluation using quantitative metrics. The overall flow is structured as shown in Fig. B.

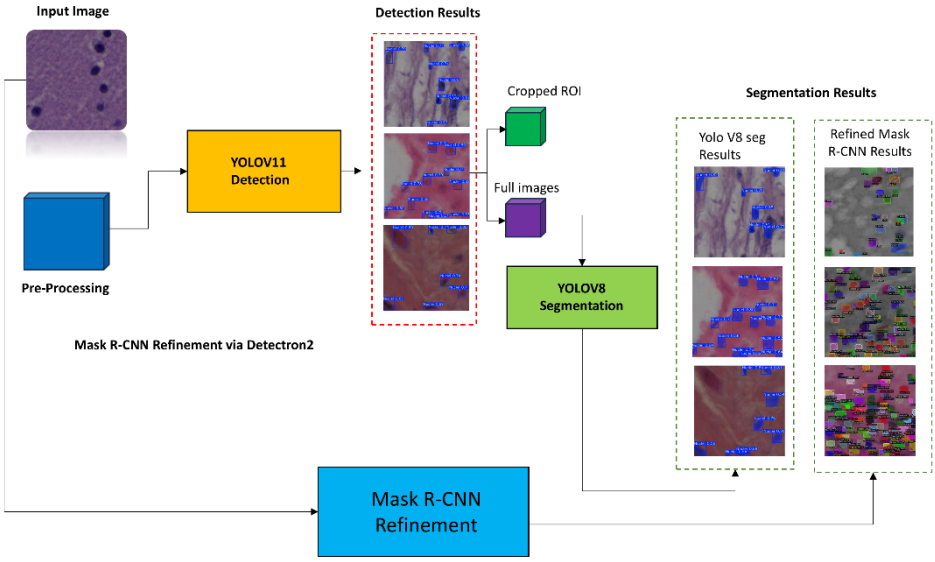

Figure B: Overview of the proposed MASKYLO framework. YOLOv11 performs real-time detection of regions of interest in H&E-stained histopathology images and YOLOv8 and Mask R-CNN branches for precise pixel-level segmentation.

All images are resized to 512 × 512 pixels for both segmentation models. To enable supervised learning, polygon-based annotations are employed for generating ground truth segmentation masks, which capture the precise boundaries of nuclei and tissue structures. These annotations are important for evaluating model performance. For segmentation, we employed two independent pipelines. The first used YOLOv8, a lightweight one-stage detector that performs real-time detection and instance segmentation with fast inference and competitive accuracy, making it well-suited for large-scale histopathology. The second utilized Detectron2's Mask R-CNN, a two-stage, region-based framework that generates proposals and refines them into high-resolution masks. YOLOv8 emphasizes speed and computational efficiency; Mask R-CNN offers more precise boundary refinement. This dual-model strategy highlights real-time feasibility and fine-grained accuracy in histopathological segmentation.

3.4. YOLO v8 Segmentation Architecture

A detailed architecture of YOLOv8 is shown in Fig. C for classification includes three main components: a **backbone** for feature extraction, a **neck** for aggregating multi-scale features, and a **head** that outputs class probabilities. This setup efficiently classifies images into correct mask, incorrect mask, or no mask categories.

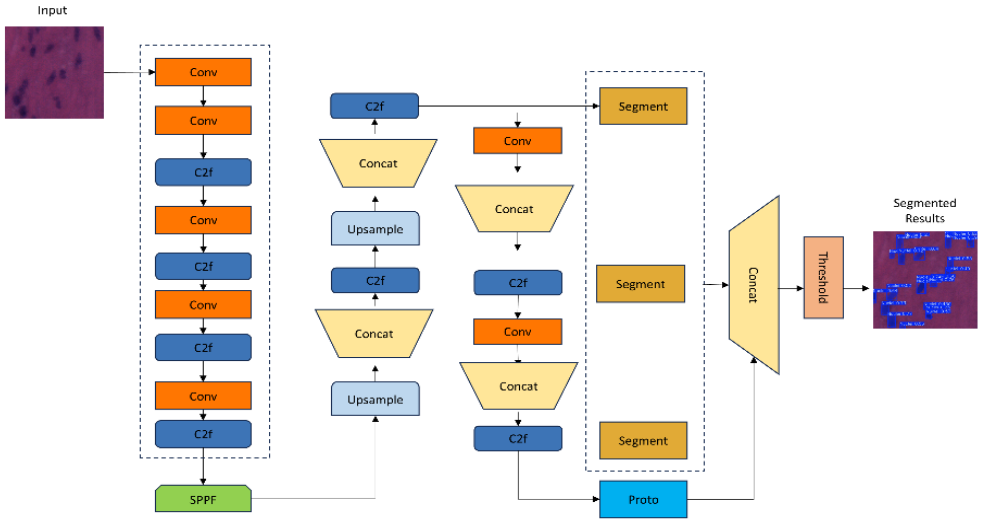

Figure C: Detailed architecture of the YOLOv8 segmentation model for H&E images.

1) Input Image Representation

The Equation.1 An H&E image with height H, width W, and RGB channels and prediction of bounding boxes B and segmentation masks M can be shown in Equation. 2:

$$I \in R^{H \times W \times 3} \quad (1)$$

$$\hat{Y} = \{(B_i, C_i, M_i)\} i^N = 1 \quad (2)$$

where Ci is the predicted class label, Bi bounding box, Mi segmentation mask.

2) Backbone (Feature Extraction)

YOLOv8 backbone extracts multi-scale features using **C2f modules** (an efficient variant of CSP) in Equation 3:

$$F = f_\theta(I) = \{F_1, F_2, F_3, \ldots, F_k\} \quad (3)$$

3) Neck (Feature Fusion with FPN/PAN)

Feature pyramid combines high- and low-level features in Equation 4,5 and 6 using upsampling (↑) and downsampling (↓):

$$F_j^* = \phi(F_j, F_j+1, \ldots, F_k) \quad (4)$$

$$F^* = \{F_1^*, F_2^*, \ldots, F_k^*\} \quad (5)$$

$$F_j^* = \text{Conv}(F_j + \uparrow(F_j+1) + \downarrow(F_j-1)) \quad (6)$$

4) Detection Head and Segmentation Head

Bounding box regression predicts (x,y,w,h) in Equation 7. Mask is generated using prototype masks P and mask coefficients αi in Equation 8

$$B_i = (\sigma(t_x) + c_x, \sigma(t_y) + c_y, e^{tw} a^w, e^{th} a_h) \quad (7)$$

$$P \in R^{Hp \times Wp \times k}, \alpha i \in R^k \quad (8)$$

3.5. Mask R-CNN model via Detectron2

For H&E patch segmentation, we employed Detectron2 with a Mask R-CNN (ResNet-50 + FPN) pretrained on ImageNet. Whole slide images (WSIs) were tiled into 512 × 512 patches.The segmentation head comprised four 3 × 3 conv layers (256 channels), a deconvolution layer, and sigmoid activation, whereas the classification branch used two 1024-d fully connected layers. Performance was evaluated using precision, recall, mAP@50, mAP@[50–95], Dice, and IoU. A detailed configuration is described in Table I.

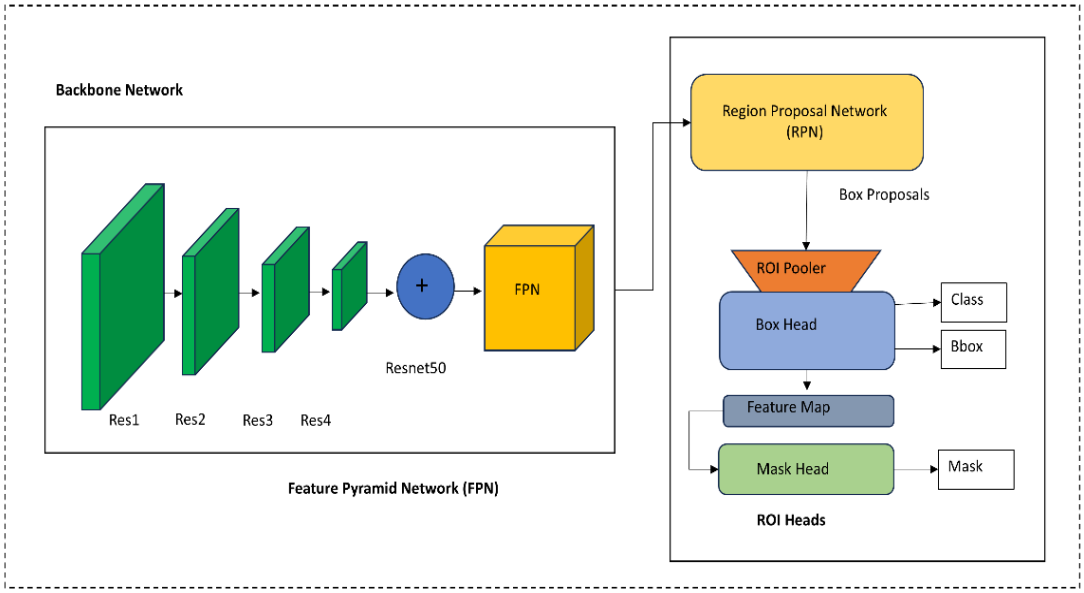

Figure D. Mask R-CNN configuration in Detectron2 for H&E segmentation, covering backbone, anchors, input, training, optimization, and evaluation.

Table I: Detailed configuration of the Mask R-CNN model implemented in Detectron2 for H&E image segmentation. The framework leverages a ResNet-50 backbone with FPN.

Component	**Configuration**
Framework	Detectron2
Model	Mask R-CNN
Backbone	ResNet-50 (pretrained on ImageNet)
Feature Extractor	FPN (Feature Pyramid Network)
Input Size	512 × 512 (H&E patches, tiled from WSI)
ROI Pooling	RoIAlign (7 × 7)
Segmentation Head	4 conv layers (3×3), 256 channels each, followed by deconvolution + sigmoid
Classification Head	2 fully connected layers (1024-d) for bounding box classification + regression
Optimizer	SGD with Momentum (0.9)
Learning Rate	0.0025 (warmup, step decay schedule)
Batch Size	16
Loss Functions	- Classification: Cross-Entropy - Box Regression: Smooth L1 - Mask: BCE
Evaluation Metrics	Precision, Recall, mAP@50, mAP@[50–95], Dice, IoU

4. Results and Discussion

In task of nuclei detection and segmentation, the Box Branch is responsible solely for predicting nuclei bounding boxes, and its performance is quantified using detection-oriented metrics such as precision, recall, and mAP, as observed with YOLOv11 (Precision: 0.89, Recall: 0.85, mAP@50: 0.88) and YOLOv8 (Precision: 0.92, Recall: 0.88, mAP@50: 0.91). Pixel-level overlap metrics like IoU and Dice are integrally unsuitable to bounding boxes, as they do not capture the spatial shape of the tissues. YOLOv11 Mask Branch (IoU: 0.78, Dice: 0.81) and YOLOv8 Mask Branch (IoU: 0.82, Dice: 0.85), highlighting superior mask-level performance in H&E image analysis using YOLOv8 in Table II.

Table II: Performance evaluation of YOLOv11 and YOLOv8 on box and mask branches.

Metric	**YOLOv11** **Box** Branch	**YOLOv11** **Mask** Branch	**YOLOv8** **Box** Branch	**YOLOv8** **Mask** Branch
Precision	0.89	0.86	0.92	0.90
Recall	0.85	0.83	0.88	0.87
mAP@50	0.88	0.84	0.91	0.89
mAP@50–95	0.74	0.70	0.77	0.75
IoU	—	0.78	—	0.82
Dice	—	0.81	—	0.85

Table III: Performance metrics of Mask R-CNN on nuclei segmentation and detection tasks.

Metric	**Mask R-CNN (bbox)**	**Mask R-CNN (segm)**
Precision	0.66	0.64
Recall	0.42	0.48
Dice Coefficient	—	0.96
IoU (Jaccard)	—	0.96
AP (IoU=0.50:0.95)	0.344	0.313
AP50	0.663	0.635
AP75	0.330	0.278
AR@100	0.423	0.384

The table III presents a detailed comparison of Mask R-CNN performance for nuclei detection and segmentation, achieving a Dice coefficient and IoU of 96%. COCO-style metrics, including AP, AP50, AP75, and AR@100, highlighting the model's robust performance in both instance-level localization and pixel-level segmentation. Figure E illustrates the performance of the proposed MASKYLO framework on H&E-stained test images. Panel (a) shows YOLOv11 bounding boxes, highlighting rapid nuclei localization, while panel (b) depicts YOLOv8 segmentation masks with refined boundary delineation. Panel (c) presents Mask R-CNN outputs, achieving precise nuclei segmentation with clear background separation.

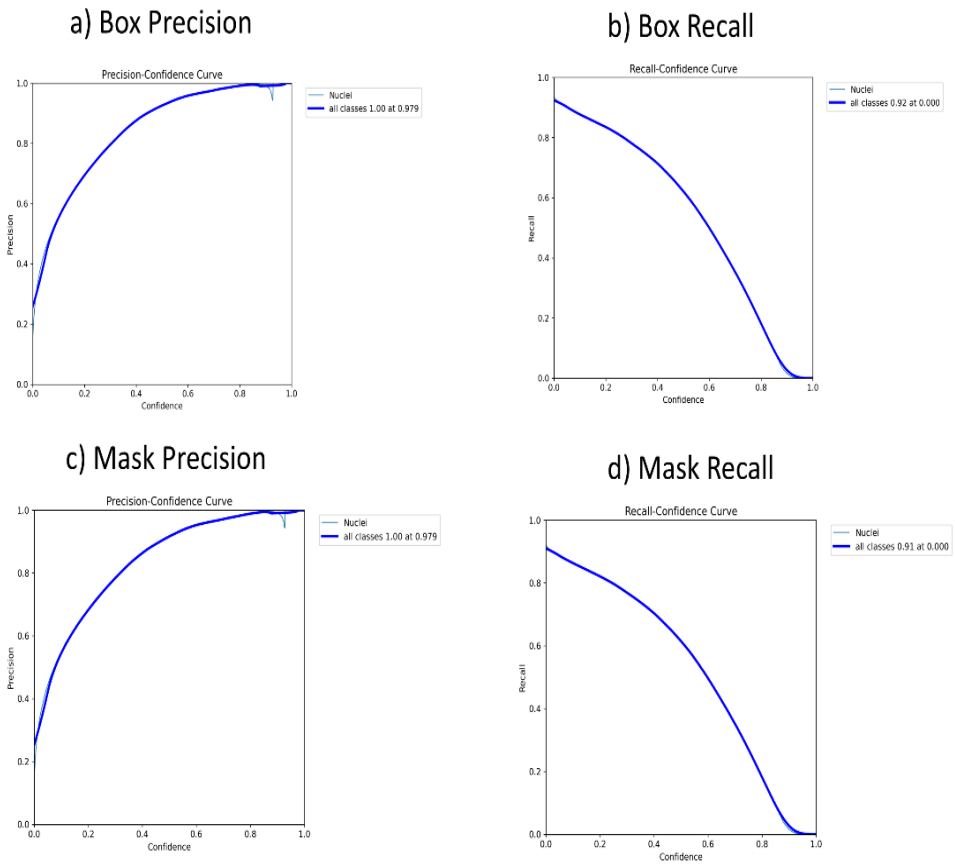

Figure E: Box-level Precision and Recall curves of the YOLOv8 segmentation model plotted over training epochs.

The figure F presents a comparison between the ground truth labels and the model's predicted labels for background and nuclei. The table IV summarizes the comparative performance of YOLOv11, YOLOv8, and Mask R-CNN on H&E-stained histopathology images.

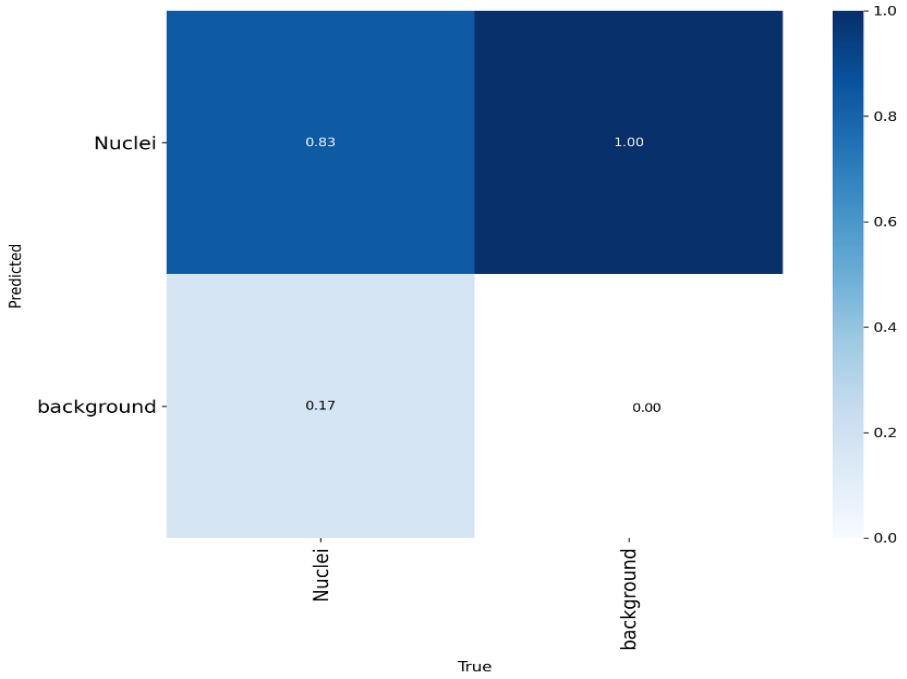

Figure F: Visualization of true versus predicted labels for background and nuclei.

Table IV: Comparative performance metrics on different models showing YOLOv11, YOLOv8, Mask R-CNN remarkable improvement by the proposed MASKYLO framework.

Study	Model	mAP@50	IoU	Dice	Precision	Recall
Juhong,.A et al., 2022 [21]	SRGAN-ResNeXt	0.86	0.89	0.76	0.83	0.81
Mostafa et al., 2022	U-net	0.79	0.71	75.5	0.81	0.78
Our Study (MASKYLO)	YOLOv11 Mask Branch	**0.88**	**0.78**	**0.81**	**0.86**	**0.83**
Our Study (MASKYLO)	YOLOv8 Mask Branch	**0.91**	**0.82**	**0.85**	**0.90**	**0.87**
Our Study (MASKYLO)	Mask R-CNN (segm)	**0.66**	**0.96**	**0.96**	**0.64**	**0.48**

5. Conclusions

In this study, we proposed MASKYLO, a hybrid deep learning framework combining YOLOv11 for fast nuclei detection with YOLOv8 and Mask R-CNN for precise pixel-level segmentation. Our approach leverages the real-time detection capability of YOLOv11 while refining mask predictions through fused segmentation networks, achieving high spatial fidelity. Evaluated on the NuInsSeg dataset, MASKYLO demonstrated superior performance, with the detection branch achieving a precision of 0.89, a recall of 0.85, and an mAP@50 of 0.88, and the segmentation branch attaining an IoU and Dice score of 0.96. These results highlight MASKYLO's effectiveness in delivering accurate, efficient, and reliable detection and segmentation of histopathological images, underscoring its potential as a robust tool for automated pathology workflows and improved clinical decision-making.

Author Contributions: Conceptualization, N.B., H.M., and S.H.; methodology, N.B. and H.M.; software, N.B.; validation, H.M. and S.H.; formal analysis, N.B.; investigation, N.B.; resources, R.H. and S.H.; data curation, N.B.; writing—original draft preparation, N.B.; writing—review and editing, H.M. and S.H.; visualization, N.B.; supervision, S.H. and R.H.; project administration, S.H.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available on reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Madabhushi, A., & Lee, G. (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, *33*, 170-175.
2. Hussain, S., et al., IoT and deep learning based approach for rapid screening and face mask detection for infection spread control of COVID-19. Applied Sciences, 2021. **11**(8): p. 3495.
3. Kleczek, P., J. Jaworek-Korjakowska, and M. Gorgon, *A novel method for tissue segmentation in high-resolution H&E-stained histopathological whole-slide images.* Computerized Medical Imaging and Graphics, 2020. **79**: p. 101686.
4. Bukhari, N., Hussain, S., Ayoub, M., Yu, Y., & Khan, A. Deep learning based framework for emotion recognition using facial expression. *Pakistan Journal of Engineering and Technology*, 2022, *5*(3), 51-57.
5. Husssain, S., Ayoub, M., Wahid, J. A., Khan, A., Alabrah, A., & Amran, G. A. Cough2COVID-19 detection using an enhanced multi layer ensemble deep learning framework and CoughFeatureRanker. *Scientific Reports 2024, 14*(1), 25207
6. Hussain, S., et al., *Ensemble Deep Learning Framework for Situational Aspects-Based Annotation and Classification of International Student's Tweets during COVID-19.* Computers, Materials & Continua, 2023. **75**(3).
7. Hussain, S., Wahid, J. A., Ayoub, M., Tong, H., & Rehman, R. Automated segmentation of coronary arteries using attention-gated unet for precise diagnosis. *Pakistan Journal of Scientific Research (PJOSR)*, 2024, *3*(1), 124-129.
8. Hussain, S., Amran, G. A., Alabrah, A., Alkhalil, L., & AL-Bakhran, A. A. C19-MLE: A Multi-Layer Ensemble Deep Learning Approach for COVID-19 Detection Using Cough Sounds and X-ray Imaging. *IEEE Access, 2024*.
9. Naseer F, Khan MN, Tahir M, Addas A, Kashif H. "Enhancing Elderly Care with Socially Assistive Robots: A Holistic Framework for Mobility, Interaction, and Well-Being. IEEE ACCESS. 2025; 15(3):301.
10. A. Addas, Khan MN, M. Tahir, Naseer F, Gulzar Y, Onn C.W. Integrating sensor data and GAN-based models to optimize medical university distribution: a data-driven approach for sustainable regional growth in Saudi Arabia. *Frontiers in Education*. 2025; 10:1527337.
11. Naseer F, **Khan MN**, Addas A, Awais Q, Ayub N. Game Mechanics and Artificial Intelligence Personalization: A Framework for Adaptive Learning Systems. *Education Sciences*. **2025**; 15(3):301.
12. Naseer F, Addas A, M. Tahir, **Khan MN**, Sattar N. Integrating generative adversarial networks with IoT for adaptive AI-powered personalized elderly care in smart homes. *Frontiers in Artificial Intelligence*. **2025**; 8:1520592.
13. A. Addas, F. Naseer, M. Tahir and **Muhammad Nasir Khan**, "Enhancing Higher Education Governance through Telepresence Robots and Gamification: Strategies for Sustainable Practices in the AI-Driven Digital," *Educ. Sci.* **2024**, *14*(12), 1324.
14. Wang, G., et al., *DeepIGeoS: a deep interactive geodesic framework for medical image segmentation.* IEEE transactions on pattern analysis and machine intelligence, 2018. **41**(7): p. 1559-1572.
15. Ho, D.J., et al., *Deep interactive learning-based ovarian cancer segmentation of H&E-stained whole slide images to study morphological patterns of BRCA mutation.* Journal of Pathology Informatics, 2023. **14**: p. 100160.
16. Pinckaers, H., et al., *Detection of prostate cancer in whole-slide images through end-to-end training with image-level labels.* IEEE Transactions on medical imaging, 2021. **40**(7): p. 1817-1826.

17. Abdullah Addas, Muhammad Nasir Khan and Fawad Naseer, " Waste management 2.0 leveraging internet of things for an efficient and eco-friendly smart city solution" *PLOS ONE, 2024,* 19, no. 7: e0307608.
18. Bing, X., et al., *Medical image super resolution using improved generative adversarial networks.* IEEE Access, 2019. **7**: p. 145030-145038.
19. Fawad Naseer, Khan, Muhammad Nasir, Muhammad Tahir, Abdullah Addas and SM Haider Aejaz. "Integrating deep learning techniques for personalized learning pathways in higher education" *Heliyon, 2024,* 10, no. 15: e32628.
20. Khan, Muhammad Nasir, Ali Altalbe, Fawad Naseer, and Qasim Awais. "Telehealth-Enabled In-Home Elbow Rehabilitation for Brachial Plexus Injuries Using Deep-Reinforcement-Learning-Assisted Telepresence Robots" *Sensors, 2024,* 24, no. 4: 1273
21. Juhong, A., Li, B., Yao, C. Y., Yang, C. W., Agnew, D. W., Lei, Y. L., ... & Qiu, Z. (2022). Super-resolution and segmentation deep learning for breast cancer histopathology image analysis. *Biomedical Optics Express*, *14*(1), 18-36.
22. Mostafa, A. M., Alaerjan, A. S., Aldughayfiq, B., Allahem, H., Mahmoud, A. A., Said, W., ... & Ezz, M. (2025). Optimized YOLOv8 for enhanced breast tumor segmentation in ultrasound imaging. *Discover Oncology*, *16*(1), 1152.