NUCLEUS SEGMENTATION

in Heterogeneous Microscopy Images on the Data Science Bowl 2018 Dataset

A Comparative Study of YOLOv11-seg and YOLOv12-seg for Instance-Level Nuclei Detection

Sercan Külcü

Computer Engineering Department, Giresun University, Türkiye

3rd International Conference on Recent and Innovative Results in Engineering and Technology (ICRIRET 2025)
November 15–16, 2025 · Konya, Turkey

Introduction & Motivation

1

A fundamental task

Accurate nucleus segmentation in microscopy images underpins computational pathology, cancer diagnosis, and biomedical research.

2

Beyond medicine

The same computer-vision techniques support maritime navigation, sonar-based tracking, and MRI-based tumor detection — showing their versatility on noisy, multi-modal data.

3

A growing toolbox

Recent work introduces wavelet transformers, boundary-aware modules, transformers, graph networks, and real-time YOLO-based detectors for nuclei segmentation.

4

An open question

Do newer, lighter architectures reliably outperform mature, well-optimized ones in demanding biomedical imaging tasks?

Related Work

Four main research directions in nucleus segmentation

Improved U-Net Designs

Wavelet-guided attention (BAWGNet), contourlet-driven attention (ConDANet), and boundary-aware blocks improve edge preservation (Dice up to 90.8%).

Transformer & Graph Models

Swin Transformer + graph convolutional fusion and multi-task transformer heads capture long-range context, gaining 2–5% over CNN baselines.

Real-Time One-Stage Detectors

ASF-YOLO, YOLACT and BlendMask variants prioritize speed, reaching up to 47.3 FPS while keeping competitive accuracy.

Clinical Deployment Tools

GUI-based platforms and genetic-algorithm–optimized U-Nets (GA-UNet) focus on usability and lightweight, real-world integration.

A common limitation persists: high computational overhead and limited generalization to unseen staining or imaging modalities.

Research Objective

This study offers the first direct comparison of YOLOv11-seg and YOLOv12-seg for nucleus instance segmentation on the heterogeneous DSB2018 dataset.

- Compare a mature, pretrained architecture (YOLOv11-seg) against a lighter, attention-centric successor (YOLOv12-seg)
- Evaluate both models under identical training conditions on 512×512 microscopy images
- Assess detection and segmentation quality via mAP, precision, recall, F1-score, and inference speed
- Identify whether architectural maturity outweighs novel lightweight design in biomedical imaging

Dataset: Data Science Bowl 2018 (DSB2018)

A widely recognized benchmark for nuclei instance segmentation

670

Total labelled microscopy images

80 / 20

Train / validation split ratio

10–300+

Nuclei per image (density range)

Key characteristics

- Diverse microscopy conditions: bright-field and fluorescence imaging modalities
- Variation in staining protocols, magnification levels, and cell types
- Image sizes ranging from 256×256 up to 1080×1080 pixels
- Pixel-level instance masks, with each nucleus annotated separately
- All images resized to 512×512 with padding, to match YOLO input requirements

Method: Two Competing Architectures

Both models are pretrained on COCO and evaluated under identical conditions

YOLOv11-seg

- CSPNet-based backbone
- Path Aggregation Network (PAN) neck
- Anchor-free detection head, extended for mask prediction
- Mature, extensively validated design
- Small variant (YOLOv11s-seg) selected for its speed/accuracy trade-off
- Stable transfer learning from COCO pretraining

YOLOv12-seg

- Lightweight backbone with FlashAttention
- Redesigned segmentation head for mask refinement
- Focused on enhanced feature alignment, reduced overhead
- Newer, attention-centric architecture (2025)
- Small variant (YOLOv12s-seg) evaluated under identical conditions
- Scratch + COCO-initialized YAML configuration

Experimental Setup

Identical training conditions ensure a fair comparison

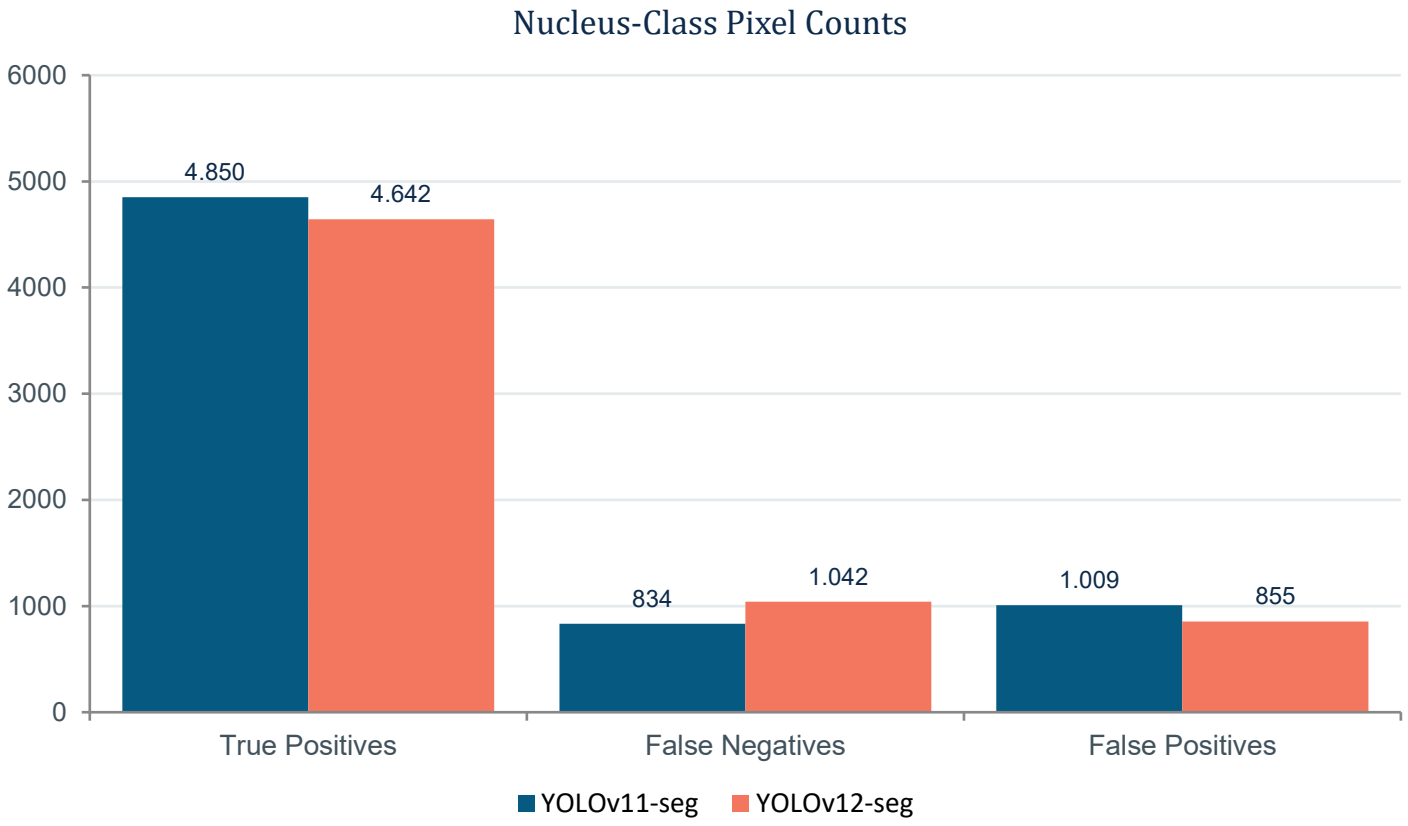
Parameter	Value
Optimizer	AdamW (learning rate = 0.002)
Batch size	16
Input resolution	512 × 512 pixels
Epochs	100 (cosine learning-rate schedule)
Data augmentation	Mosaic, HSV adjustment, random flips
Hardware	NVIDIA RTX 3050 Laptop GPU (6 GB VRAM), CUDA 11.8
Framework	PyTorch with automatic mixed precision (AMP)

Why this matters

Both YOLOv11s-seg and YOLOv12s-seg are trained with the exact same hyperparameters, hardware, and augmentation pipeline. This removes confounding factors so that any performance gap can be attributed to architectural design rather than training setup.

Results: Confusion Matrix Analysis

Pixel-level counts aggregated across 134 validation images

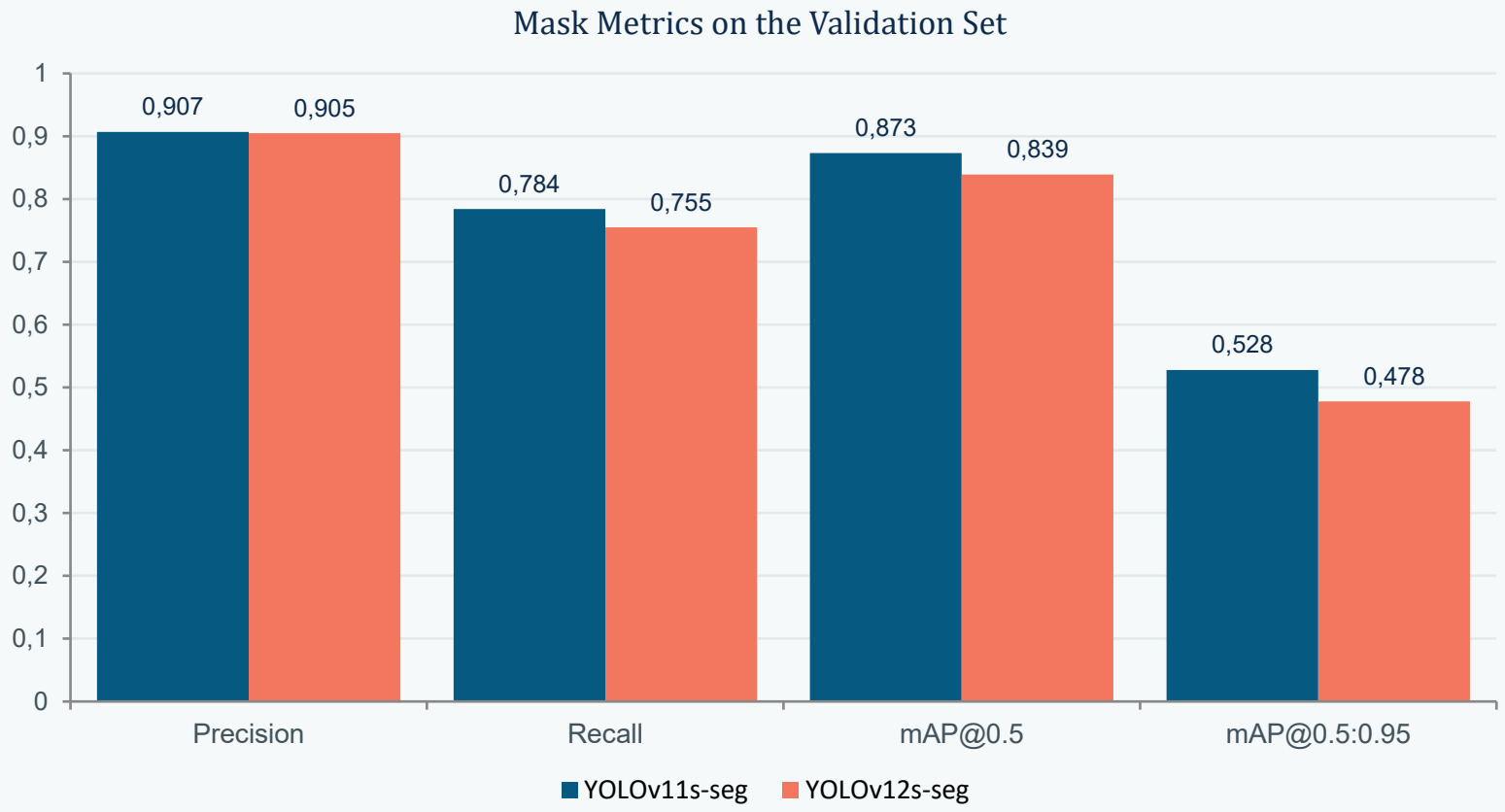


Reading the matrix

- YOLOv11-seg detects more true nuclei pixels (4,850) with fewer missed detections (834 false negatives).
- YOLOv12-seg produces fewer false positives (855) but misses more real nuclei (1,042 false negatives).
- Overall detection sensitivity is lower for YOLOv12-seg, despite its lighter design.

Results: Mask-Level Performance

YOLOv11s-seg outperforms YOLOv12s-seg across every evaluated metric



+10.7%

relative improvement in Mask mAP@0.5:0.95 (0.528 vs. 0.478)

6.3 ms

inference latency for YOLOv11s-seg, vs. 11.3 ms for YOLOv12s-seg

Training Dynamics & Computational Efficiency

Architectural maturity translates into a smoother, faster training process

Metric	YOLOv11s-seg	YOLOv12s-seg
Parameters	10.08 M	9.91 M
Network layers	203	294
Training duration	~35 min	~44 min
Training speed	2.5 it/s	1.7 it/s
Peak GPU memory	4.62 GB	5.52 GB
Best epoch (of 100)	88	98

Key takeaway

Although YOLOv12s-seg has slightly fewer parameters, its deeper 294-layer structure and FlashAttention mechanism introduce training instability, slower throughput, and higher memory demand. YOLOv11s-seg converges earlier and more reliably, thanks to its mature, well-optimized checkpoint.

Qualitative Comparison

Visual inspection confirms the quantitative findings

YOLOv11-seg Predictions

- Tight boundary adherence around individual nuclei
- Accurate instance separation, even in dense clusters
- Minimal over-segmentation in crowded regions
- Consistent performance across staining types

YOLOv12-seg Predictions

- Occasional boundary artifacts around nuclei edges
- Small or faint nuclei sometimes missed entirely
- Adjacent instances occasionally merged together
- Slightly less stable across heterogeneous images

Discussion & Conclusion

YOLOv11-seg outperforms YOLOv12-seg on the heterogeneous DSB2018 dataset for nucleus instance segmentation.

Higher accuracy

Mask mAP@0.5:0.95 of 0.528 vs. 0.478

Better boundaries

Boundary-aware F1-score of 85.3% vs. 82.9%

Faster inference

6.3 ms vs. 11.3 ms per image

Why does the mature model win?

- Optimized feature aggregation via the Path Aggregation Network (PAN)
- Stable transfer learning from large-scale COCO pretraining
- Well-established design principles outweigh novel lightweight mechanisms

Future work: hybrid models with domain-specific attention, cross-dataset generalization across multi-organ histopathology, and interactive pathologist-in-the-loop refinement tools.

Thank You

Questions & Discussion

Sercan Külcü

sercan.kulcu@giresun.edu.tr

Computer Engineering Department, Giresun University, Türkiye

ORCID: 0000-0002-4871-709X