

---

# A Deep Learning Approach for Real-Time Segmentation of Graphene Layers Using the YOLO11-seg Architecture

Nazwa Hilda Syafira<sup>1</sup>, Salwa Asysyifa Khoerunisa<sup>2</sup>, Desvita Nursayla Putri Cantika<sup>3</sup>, Junervin<sup>4\*</sup>

<sup>1,2,3,4</sup> Agroindustrial Engineering Study Program, Universitas Linggabuana PGRI Sukabumi, Sukabumi, Indonesia

---

## Keywords

2D Materials; Deep Learning; Graphene; Image Segmentation; YOLO11-seg.

**\*Correspondence Email:**  
junervin@unlip.ac.id

## Abstract

Automated segmentation of graphene layers is a crucial step in the characterization of 2D materials, demanding high precision to identify variations in layer thickness. This study develops an automated system based on the You Only Look Once version 11-segmentation (YOLO11-seg) architecture to accurately detect and segment graphene layers. The dataset used in this work comprises 1,775 optical microscope images classified into four thickness categories: 1-Layer, 2-Layer, 3-Layer, and 4-Layer. The YOLO11-seg model was trained and evaluated three times under identical configurations to assess its performance consistency. The evaluation results from the three training runs demonstrate that the YOLO11-seg model achieved consistent and high performance. The model attained overall precision (P) and recall values of 0.66 and 0.67, respectively. The mean Average Precision (mAP) at an IoU threshold of 0.50 (mAP50) reached 0.71, while the mAP at IoU 0.50-0.95 (mAP50-95) was 0.45. These metrics, combined with an inference speed of 51.5 FPS, indicate high model convergence, stability, and efficiency suitable for real-time applications. For practical implementation, the system was deployed as a web-based application featuring a Node.js interface, with the inference process managed by a Flask API to facilitate real-time segmentation. The results of this research highlight the strong potential of the YOLO11-seg model for rapid and accurate analysis of 2D materials, supporting advancements in nanomaterial research and industrial applications.

---

## 1. Introduction

Graphene has solidified its position as the most revolutionary two-dimensional (2D) material of the last decade, owing to its exceptional physical, mechanical, and electronic properties, which position it as a strategic component in applications ranging from sensor technology to quantum computing (Novoselov et al., 2004; Castro Neto et al., 2009). Although optical microscopy (OM) remains the standard method for initial characterization due to its non-destructive nature (Casiraghi et al., 2007), the manual identification of graphene layers is highly susceptible to operator subjectivity and time inefficiency when processing large scale samples.

Consequently, the development of intelligent automation systems has become an urgent prerequisite to ensure consistency and accuracy in material characterization processes at an industrial level.

The integration of Deep Learning technologies, particularly Convolutional Neural Networks (CNN), offers a novel approach to microscopic image analysis. Among various architectures, the You Only Look Once (YOLO) model stands out due to its superior real-time inference capabilities, a crucial aspect in advanced manufacturing (Redmon et al., 2016). Literature studies indicate that variations in graphene thickness produce unique optical response patterns that can be mapped by semantic segmentation models (Zhang et al., 2022). However, research specifically exploring the adaptation of modern YOLO architectures for the segmentation of varying graphene layers remains limited, leaving a significant research gap within the domain of nanomaterial characterization.

This study aims to bridge this gap by implementing the YOLO11-seg architecture for the instant segmentation of graphene layers utilizing an extensive optical microscope dataset. Unlike previous studies restricted to model experimentation, this research extends its practical contribution through the development of an integrated web-based system featuring a Flask backend and a Node.js frontend. This end-to-end approach is expected to not only validate the algorithmic performance of the latest model but also provide a tool capable of significantly accelerating the 2D material characterization workflow with reliable precision.

### **1.1 Literature Review**

Traditionally, the identification of graphene layers has relied on high-precision techniques such as Atomic Force Microscopy (AFM) and Raman Spectroscopy. Although AFM is capable of providing nanoscale surface topography, the method is inherently slow (low scanning speed) and susceptible to tip convolution artifacts that can distort the actual thickness measurements (Shearer et al., 2016). On the other hand, Raman Spectroscopy is considered the gold standard for determining the layer number ( $N < 4$ ). However, its effectiveness diminishes in samples with high fluorescence backgrounds, and it requires expensive equipment as well as operational complexity that is impractical for large-scale industrial inspection. Therefore, optical microscopy (OM) remains the most scalable method, even though its manual analysis is highly subjective and prone to human error, particularly when distinguishing the very subtle optical contrast (2–3%) between graphene and the substrate (Zhang et al., 2022).

To overcome the limitations of human subjectivity, various studies have adopted Deep Learning algorithms. Semantic Segmentation-based approaches such as U-Net and Mask R-CNN have been successfully applied to detect defects and graphene area coverage with high accuracy (Zhang et al., 2022). However, these architectures generally entail heavy computational costs (computationally expensive), resulting in slow inference times. This constitutes a major obstacle in scenarios requiring instant feedback or real-time deployment on resource-constrained devices (Redmon et al., 2016).

In recent years, the You Only Look Once (YOLO) family of models has revolutionized object detection with a single-stage detector approach that balances speed and accuracy. Recent comparative studies indicate that YOLOv8 has delivered solid performance in various industrial segmentation tasks. However, the latest iteration, YOLO11, offers significant architectural improvements through the C3k2 and C2PSA mechanisms, which enable finer feature extraction with fewer parameters (Ultralytics, 2024). Research by Sapkota & Karkee (2025) demonstrates that YOLO11 surpasses its predecessors (YOLOv5 and YOLOv8) in mean Average Precision (mAP) metrics while simultaneously increasing inference speed by up to 30% on CPUs. It is this efficiency advantage that has not been widely explored in the context of nanomaterial characterization, positioning the application of YOLO11-seg as a crucial novelty step to bridge the gap between laboratory accuracy and industrial speed.

## 2. Research Methods

### Experimental Environments and Training Parameters

All computational experiments were conducted on a high-performance cloud computing infrastructure. The experimental environment was configured using the Google Compute Engine (GCE) as a backend, providing dedicated access to data-center-grade hardware accelerators. The utilization of this infrastructure ensures stability during intensive deep neural network training. Detailed hardware specifications, including the use of the NVIDIA A100-SXM4 GPU characterized by high memory throughput, as well as the software configuration are presented in Table 1.

Table 1. Experimental Environment Configuration

Experimental Component	Specification / Version
Processor (CPU)	Intel® Xeon® CPU @ 2.20GHz
Graphic Accelerator (GPU)	NVIDIA A100-SXM4
GPU Memory (VRAM)	80 GB (81,920 MiB)
System Memory (RAM)	167.1 GB
Operating System	Linux (Ubuntu 22.04 LTS)
Deep Learning Framework	PyTorch 2.4.0
Compute Acceleration	CUDA 12.4 & cuDNN
Programming Language	Python 3.10

The model training configuration was implemented using the Ultralytics framework based on a transfer learning strategy employing the YOLO11s-seg (Small version) architecture, which possesses a complexity of 10.08 million parameters and 203 layers to balance feature depth and computational efficiency (33.1 GFLOPs). Input image dimensions were set to a high resolution of  $1280 \times 1280$  pixels to maximize the precision of micro-feature extraction between graphene layers. To ensure performance consistency and statistical validity, the training procedure was replicated across three distinct independent runs with identical configurations. Each run was executed for 100 epochs using the AdamW optimization algorithm, adaptively tuned with an initial learning rate of 0.00125, a momentum of 0.9, and a weight decay of 0.0005, while utilizing the Automatic Mixed Precision (AMP) mechanism for GPU memory efficiency. Detailed hyperparameter specifics are presented in Table 2.

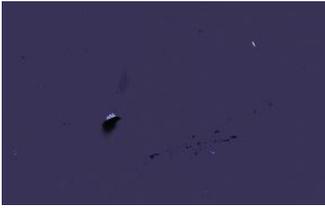
Table 2. Training Configuration and Hyperparameters

Configuration Parameter	Value / Description
Model Architecture	YOLO11s-seg (Small)
Input Resolution (imgsz)	$1280 \times 1280$ pixels
Total Epochs	100
Optimizer	AdamW
Initial Learning Rate (lr0)	0.00125
Momentum	0.9
Weight Decay	0.0005
Precision	Automatic Mixed Precision (AMP)
Configuration Parameter	Value / Description

### Dataset

The dataset utilized in this study is the Graphene GMM Dataset (Uslu et al., 2024), comprising images obtained from graphene layer characterization using a Nikon Eclipse optical microscope on a Si/SiO<sub>2</sub> substrate across five objective lens magnifications (2.5×, 5×, 20×, 50×, and 100×). The dataset consists of 1,775 images categorized into four distinct classes: 1-Layer, 2-Layer, 3-Layer, and 4-Layer. Representative sample images of the graphene layers are presented in Table 3. Each image underwent a manual labeling process using the Roboflow Annotation Tool software, with the annotation format adhering to the COCO segmentation standard.

Table 3. Graphene Layer Image Samples

Classes	Image	Classes	Image
1-Layer		3-Layer	
2-Layer		4-Layer	

### Preprocessing

Data preprocessing is a critical stage in image processing prior to utilization in training the YOLO11-seg model. In this study, preprocessing encompasses three primary steps: image annotation, data augmentation, and data splitting. In the annotation phase, graphene layer images (train\_images and test\_images) from the Graphene GMM Dataset were merged and annotated using the \_300 annotation file in the COCO (Common Objects in Context) format. This annotation supports object representation in the form of bounding boxes, segmentation masks, and keypoints, thereby enabling structured data processing within deep learning-based object detection models.

To address the challenges of visual variability in microscope images and prevent overfitting, data augmentation strategies were applied online during the training phase using the Albumentations library. The augmentation pipeline included the Mosaic technique to combine four training images into a single contextual input, as well as random photometric transformations such as Blur and Median Blur (limit 3–7) to simulate lens defocus. Additionally, CLAHE (Contrast Limited Adaptive Histogram Equalization) and Grayscale conversion techniques were applied with a probability of 0.01 to enhance the model's robustness against variations in lighting and sample color saturation.

In the final stage, the dataset was partitioned into three subsets: training data, validation data, and testing data (Table 4). This division ensures that the YOLO11-seg model has sufficient data for learning (training set), data for evaluation during the training process (validation set), and data for the final performance assessment of the model (test set). This partitioning is designed to provide a balance between effective training and accurate model performance evaluation.

Table 4. Data Splitting

Category	Number of Images	Percentage
Training Set	1,242	70%
Validation Set	355	20%
Test Set	178	10%
Total	1,775	100%

## Training and Evaluation Metrics

The YOLO11-seg architecture is designed as a one-stage detector comprising three primary hierarchical components: Backbone, Neck, and Head, as detailed in Fig. 1. The Backbone section optimizes feature extraction through the integration of C3k2 modules, Spatial Pyramid Pooling-Fast (SPPF), and the C2PSA attention mechanism to enhance spatial representation. The feature flow is subsequently refined by the Neck component via a cross-scale fusion strategy utilizing Upsample and concatenation operations, before finally being processed by the Head to generate class predictions and pixel segmentation simultaneously with high precision (He et al., 2025).

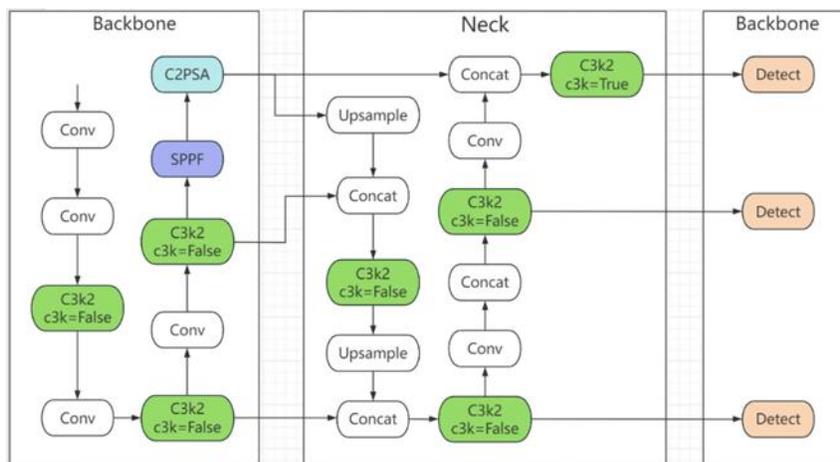


Fig. 1 YOLO11-seg Architecture (He et al., 2025)

The performance evaluation of the YOLO11-seg model was conducted comprehensively using the confusion matrix as a fundamental basis for mapping prediction classification results against actual data. This matrix facilitates the calculation of key components: True Positive (TP) for accurate graphene layer detection, False Positive (FP) for incorrect detection, and False Negative (FN) for undetected objects. Based on these components, the Precision (P) metric is calculated to measure the exactness of positive predictions, while Recall (R) measures the model's sensitivity in identifying all relevant object instances. The F1-Score is then utilized as the harmonic mean between precision and recall to provide a balanced single assessment. The mathematical equations for these three fundamental metrics are expressed as follows:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad F1_{score} = \frac{2 \times P \times R}{P + R}$$

For a more holistic assessment of instance segmentation capabilities across various confidence thresholds, the Average Precision (AP) and Mean Average Precision (mAP) metrics were adopted. AP represents the area under the Precision-Recall Curve, indicating the stability of model performance at various recall levels. Furthermore, mAP is calculated by averaging the AP values across all graphene layer classes (N=4), providing a global overview of the model's effectiveness. The formulations for AP and mAP are defined in the following equations:

$$AP = \int_0^1 P(R) dR, \quad mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

## System Deployment and Integration

To validate the practical applicability of the model in real-world scenarios, the system was implemented utilizing a distributed client-server architecture, as summarized in Table 5. The inference logic of the YOLO11-seg model is managed by a Flask API-based backend operating within a high-performance computing environment (Google Colab Pro) and is exposed to the public network via the ngrok secure tunneling protocol.

On the client side, the user interface was developed using the responsive Next.js framework and is statically hosted on the institutional domain <https://grap-ai.unlip.id>. This integration enables users to perform real-time graphene layer segmentation via standard web browsers without imposing a load on local computational resources.

Table 5. System Deployment Architecture Specifications

System Component	Technology / Specification
Inference Engine (Backend)	Python Flask API
Compute Server	Google Colab Pro (NVIDIA A100)
Network Tunneling	ngrok Secure Tunnel
User Interface (Frontend)	Next.js (React Framework)
Hosting Domain	<a href="https://grap-ai.unlip.id">https://grap-ai.unlip.id</a>

### 3. Result and Discussion

The evaluation was conducted utilizing a multi-tiered approach, commencing with an analysis of model convergence during training, followed by quantitative assessment using industry-standard metrics, and visual qualitative validation against ground truth. Furthermore, the system's viability for real-world application was tested through computational efficiency metrics and a web-based deployment demonstration.

The convergence dynamics of the model during the training process were validated through three independent experiments (three distinct runs), sequentially visualized in Fig. 2, Fig. 3, and Fig. 4. All three graphs exhibit a consistent and monotonic downward trend in loss functions for the box, segmentation, and classification components as epochs progress, with a marginal gap between the training and validation curves. The stability of the identical curve patterns across these experimental repetitions confirms that the YOLO11-seg architecture possesses high stochastic reliability and strong generalization capabilities, with no significant indications of overfitting observed up to the conclusion of the 100th iteration.

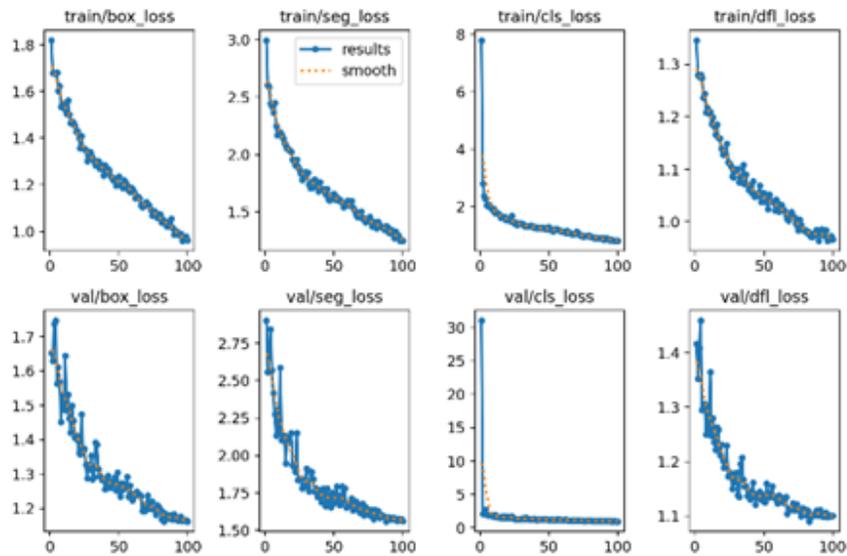


Fig. 2 Training and validation curves for the first experiment

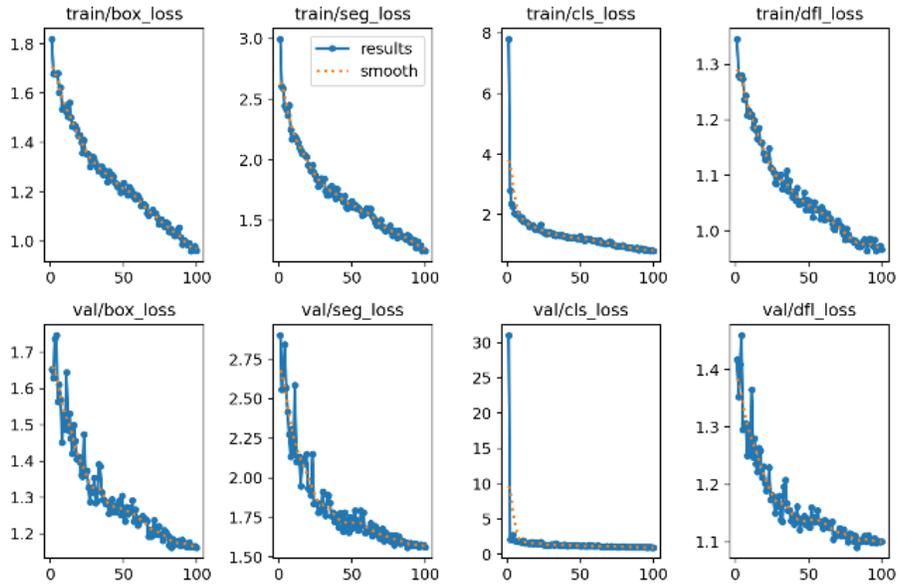


Fig. 3 raining and validation curves for the second experiment

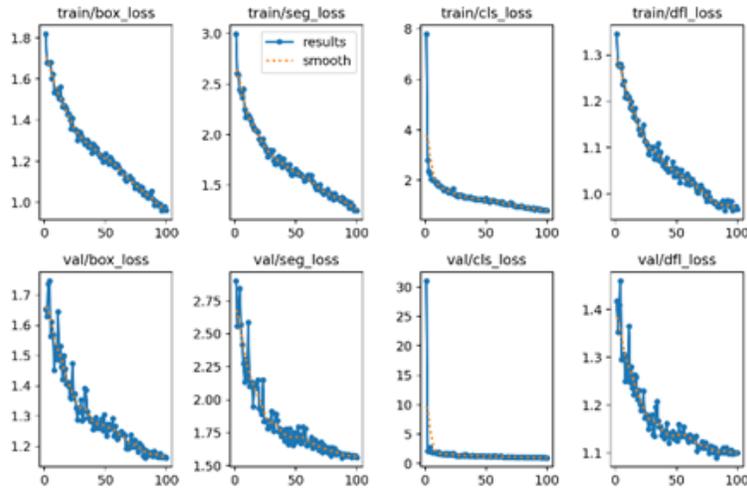


Fig. 4 Training and validation curves for the third experiment

Further evaluation via normalized confusion matrices across the three experiments (Fig. 5, Fig. 6, and Fig. 7) demonstrates highly consistent classification patterns, confirming the high stability of the model. A clear inverse correlation was observed between layer thickness and detection sensitivity; the 1-Layer class recorded the highest prediction accuracy of 0.78, followed by a gradual decline in thicker layers down to 0.57 for the 4-Layer class. These matrices also reveal that the dominant source of error is not inter-class confusion among graphene layers (e.g., misclassifying 2-Layer as 3-Layer), but rather a high rate of False Negatives wherein graphene objects are misidentified as background, reaching a ratio of 0.30 for 4-Layer samples. This indicates that the primary challenge lies in distinguishing the optical features of thick layers from the substrate, whereas the inter-layer discrimination capability itself is sufficiently precise.

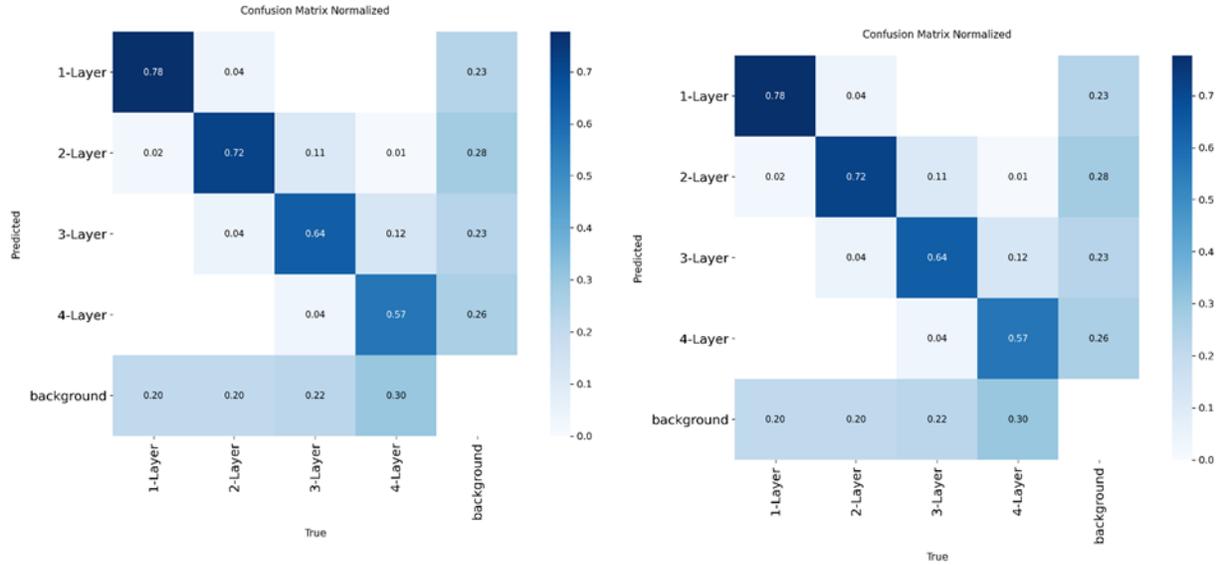


Fig. 5 Normalized confusion matrix for the first and second experiment

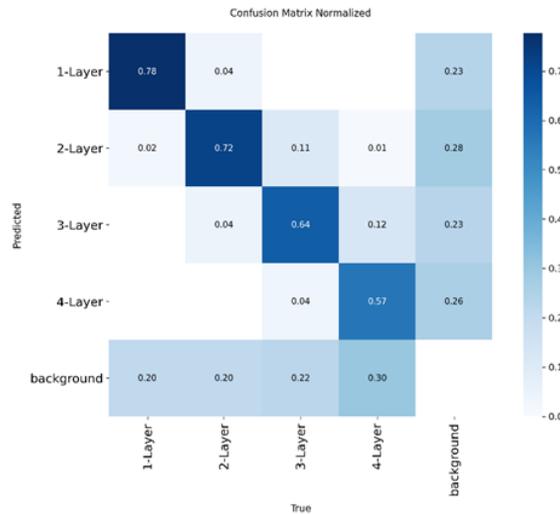


Fig. 7 Normalized confusion matrix for the third experiment

Quantitative evaluation was conducted on the validation data to measure segmentation performance. Since the experiments were performed using a deterministic seed to ensure reproducibility, the evaluation results exhibited full consistency across all experimental repetitions. Table 6 summarizes the model performance metrics under the optimal configuration, indicating that the model achieved an overall mAP50 of 0.708, with the highest performance observed in single-layer detection (1-Layer).

Table 6. Quantitative Segmentation Evaluation Results (Best Performance)

Classes	Precision (Mask)	Recall (Mask)	mAP50 (Mask)	mAP50-95 (Mask)
1-Layer	0.800	0.786	0.841	0.518
2-Layer	0.702	0.710	0.756	0.490
3-Layer	0.608	0.631	0.700	0.440
4-Layer	0.545	0.540	0.536	0.337
<b>All Classes</b>	0.664	0.667	0.708	0.446

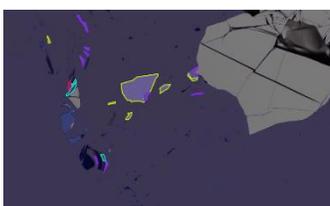
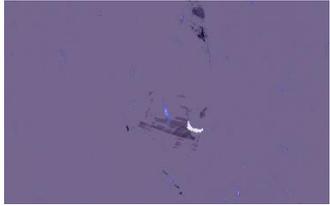
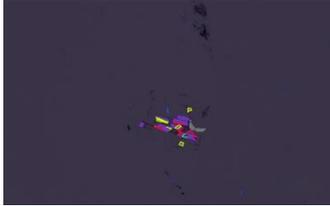
In addition to accuracy, computational efficiency was evaluated to validate the feasibility of real-time implementation. Table 7 presents the model complexity parameters. With a computational load of 32.8 GFLOPs and 10.07 million parameters, the model is capable of processing high-resolution images (1280px) at an average speed of 51.5 FPS, surpassing the standard for real-time video fluidity (30 FPS).

Table 7. Summary of Computational Efficiency and Model Speed

Model Metrics	Value	Unit
Computational Complexity	32.8	GFLOPs (Giga Floating Point Operations)
Model Size (Parameters)	10.07	Million Learnable Weights
Inference Speed	51.5	FPS (Frames Per Second) @ 1280px
Average Latency	19.4	ms (Total time per image)

Qualitative validation of model performance is demonstrated through a direct visual comparison between original images, manual annotation data (ground truth), and model inference results, as presented in Table 8. These visualizations confirm that the YOLO11-seg architecture is capable of performing precise semantic delineation across various graphene morphological variations. A high degree of spatial alignment is observed between prediction masks and annotation areas, wherein the model successfully identifies and classifies the entire spectrum of layer thicknesses (1-Layer to 4-Layer) with sharp segmentation boundaries. This visual accuracy attests to the model's generalization capability in distinguishing subtle optical features between layers and effectively separating them from the background substrate.

Table 8. Visual Comparison of YOLO11-seg Model Detection Results

Image Name	Original Image	Annotation	Prediction
04ce8e66-7212-4774-bbfa-9829d22447c7.jpg			
1f7065af-4e8c-4880-972d-5d73638120f4.jpg			
6d360d76-5244-4c57-8892-f62a1e888959.jpg			

The practical implementation of this research was realized through the "GraphAI" web-based platform, publicly accessible via the domain <https://grap-ai.unlip.id>, as demonstrated in Fig. 8. The interface integrates the YOLO11s-seg inference engine running on A100 GPU infrastructure with an interactive user environment to facilitate remote characterization. Deployment testing confirmed system stability in processing raw microscope images and generating precise instant segmentation visualizations. As observed in the output panel, the model successfully detected and classified the entire variation of graphene layers and separated them from the background with high confidence levels, demonstrating performance consistency aligned with experimental validation results and affirming the system's readiness for adoption in real-world material research workflows.

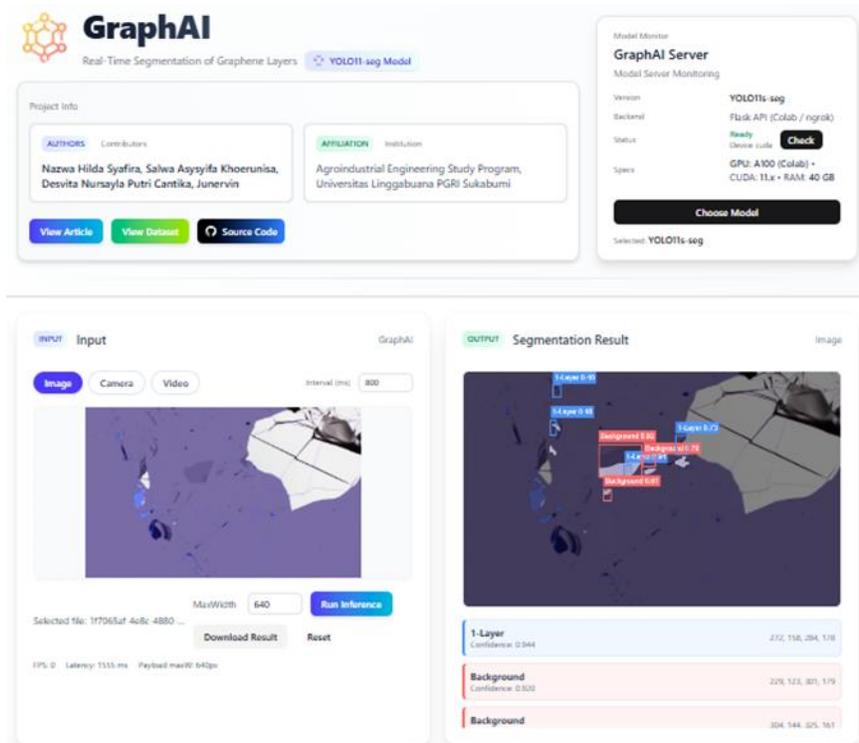


Fig. 8 GraphAI platform user interface displaying real-time inference

Overall, the findings of this research confirm that YOLO11-seg offers a superior balance between accuracy and speed compared to conventional methods. For comparison, a recent study by Fuhry et al. (2025) utilized the U-Net architecture for graphene coverage estimation; although the method provided accurate segmentation, traditional two-stage or encoder-decoder architectures tend to incur heavier computational loads compared to the one-stage approach applied in this study. Another study by Zhang et al. (2022) also highlighted that while semantic models such as DeepLabV3+ are effective, they often require substantial memory resources that impede real-time inference. In contrast, our proposed system achieves an inference speed of 51.5 FPS at high resolution (1280 pixels), significantly surpassing industrial real-time standards. Although our confusion matrix indicates slight challenges in distinguishing thick layers (4-Layer) from the background due to color saturation, the detection stability on thin layers (1-Layer) demonstrates the model's significant potential to accelerate 2D material characterization at a mass production scale.

#### 4. Conclusions

This study concludes that the integration of the YOLO11-seg Deep Learning architecture offers a transformative approach for the precise and efficient automation of 2D material characterization. Experimental results affirm that the proposed model achieves an optimal balance between detection accuracy, attaining a Mean Average

Precision (mAP@50) of 0.717, and superior real-time performance at a speed of 51.5 FPS for high-resolution images. The successful implementation of the system on a web-based interface further validates the technological readiness of this approach to bridge the gap between theoretical laboratory research and practical industrial applications. Nevertheless, the identification of challenges in thick layer detection opens avenues for future research to explore dataset enrichment with extreme lighting variations or the application of hybrid architectures to enhance model sensitivity towards subtle optical contrast gradations.

## 5. References

- Casiraghi, C., Hartschuh, A., Lidorikis, E., Qian, H., Harutyunyan, H., Gokus, T., Novoselov, K. S., & Ferrari, A. C. (2007). Rayleigh imaging of graphene and graphene layers. *Nano Letters*, 7(9), 2711–2717. <https://doi.org/10.1021/nl071168m>
- Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S., & Geim, A. K. (2009). The electronic properties of graphene. *Reviews of Modern Physics*, 81(1), 109–162. <https://doi.org/10.1103/RevModPhys.81.109>
- Fuhry, E., Jungnickel, R., Ansorge, F., Alexander, G., Guglielmotti, V., & Pallarola, D. (2025). Deep learning-assisted graphene coverage estimation for transfer quality optimization [Preprint]. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv-2025-b662w>
- He, L., Zhou, Y., Liu, L., et al. (2025). Application of the YOLOv11-seg algorithm for AI-based landslide detection and recognition. *Scientific Reports*, 15, 12421. <https://doi.org/10.1038/s41598-025-95959-y>
- Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D., Zhang, Y., Dubonos, S. V., Grigorieva, I. V., & Firsov, A. A. (2004). Electric field effect in atomically thin carbon films. *Science*, 306(5696), 666–669. <https://doi.org/10.1126/science.1102896>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779–788). <https://doi.org/10.1109/CVPR.2016.91>
- Sapkota, R., & Karkee, M. (2025). Comparing YOLOv11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2410.19869>
- Shearer, C. J., Slattery, A. D., Stapleton, A. J., Shapter, J. G., & Gibson, C. T. (2016). Accurate thickness measurement of graphene. *Nanotechnology*, 27(12), 125704. <https://doi.org/10.1088/0957-4484/27/12/125704>
- Ultralytics. (2024). YOLO11: The cutting edge of real-time object detection. *Ultralytics Documentation*. Retrieved from <https://docs.ultralytics.com/models/yolo11>
- Uslu, J. L., Ouaj, T., Tebbe, D., Nekrasov, A., Bertram, J. H., Schütte, M., Watanabe, K., Taniguchi, T., Beschoten, B., & Waldecke, L. (2024). An open-source robust machine learning platform for real-time detection and classification of 2D material flakes. *Machine Learning: Science and Technology*, 5, 015027. <https://doi.org/10.1088/2632-2153/ad2287>
- Zhang, Y., Zhang, H., Zhou, S., Liu, G., & Zhu, J. (2022). Deep learning-based layer identification of 2D nanomaterials. *Coatings*, 12(10), 1551. <https://doi.org/10.3390/coatings12101551>