

Bridging AI and Forensics: A Detectron2-Powered System for Crime Scene Image Analysis

G Sreeya

Dept. Artificial Intelligence &
Machine Learning
Global Academy of Technology
Bangalore, Karnataka

Rithesh Kunder

Dept. Artificial Intelligence &
Machine Learning
Global Academy of Technology
Bangalore, Karnataka

Sanjana Nagaraj

Dept. Artificial Intelligence &
Machine Learning
Global Academy of Technology
Bangalore, Karnataka

Prof. Vasugi I

Dept. Artificial Intelligence &
Machine Learning
Global Academy of
Technology Bangalore,
Karnataka

Dr. Roopa B S

Dept. Artificial Intelligence &
Machine Learning
Global Academy of
Technology Bangalore,
Karnataka

Abstract—Panoptic segmentation is a comprehensive computer vision task that combines the strengths of both semantic and instance segmentation, offering a complete understanding of visual scenes. It involves classifying every pixel in an image while also identifying and separating individual object instances. Recent developments in deep learning have greatly enhanced the performance and efficiency of panoptic segmentation models. This study presents a panoptic segmentation pipeline using Detectron2, specifically implementing the pre-trained model "panoptic_fpn_R_101_dconv_cascade_gn_3x" for forensic image analysis. The methodology includes steps such as image preprocessing, model integration, segmentation processing, and result visualization to achieve detailed scene interpretation. The key objectives of this project are: (1) to develop a real-time system for panoptic segmentation of forensic imagery, (2) to display confidence scores for each detected object, and (3) to provide a simple, web-based interface with remote access to the outputs. The segmentation results are assessed using both qualitative and quantitative metrics to ensure reliability and accuracy. By integrating deep learning-powered segmentation with an accessible web interface, this project offers a practical tool for forensic professionals, with potential extensions in smart surveillance and autonomous systems.

Keywords— Panoptic Segmentation, Forensic Image Analysis, Detectron2, Deep Learning, Instance Segmentation, Semantic Segmentation, Confidence Score, Gradio Interface, Image preprocessing, Segmentation processing, Visualization

1. INTRODUCTION

Deep learning has transformed numerous domains, including computer vision, natural language processing, and robotics, achieving performance levels comparable to humans in complex tasks. In the field of forensic science—where accurate visual interpretation is essential for crime scene investigation and evidence identification—deep learning-based segmentation techniques have emerged as highly effective tools. Traditional image processing methods often struggle in forensic settings due to challenges like overlapping objects, varying object scales, and diverse scene compositions.

Panoptic segmentation, a cutting-edge deep learning technique that merges semantic segmentation (categorizing each pixel) with instance segmentation (identifying individual object instances), has become a prominent solution for these challenges. By offering a complete understanding of visual scenes, panoptic segmentation enhances tasks such as evidence classification, crime scene reconstruction, and object detection. However, applying these models in real-world forensic scenarios presents ongoing challenges related to interpretability, robustness, and scalability.

This project introduces a real-time panoptic segmentation system tailored for forensic image analysis. Built on the Detectron2 framework, the system accurately segments and labels both foreground objects and background elements. Users can upload images or capture them directly using a live camera, after which the images are processed through the segmentation model. The outputs include visualized segmentations with confidence scores reflecting the model's certainty for each detected object. A key component of the system is a user-friendly web interface developed with Gradio, allowing for easy access and remote sharing of results through generated links. This ensures that the system is not only technically capable

but also practical and accessible for forensic professionals. By streamlining forensic image analysis, this project aims to improve the speed, accuracy, and clarity of investigations, reducing reliance on manual inspection.

2. LITERATURE SURVEY

Image segmentation has traditionally been divided into semantic and instance segmentation. Semantic segmentation labels each pixel with a class, focusing on background elements like sky or road, using models like FCNs and Deep Lab. Instance segmentation, on the other hand, identifies and segments individual object instances, with approaches like Mask R-CNN. These tasks were handled separately until Kirillov et al., [1], introduced panoptic segmentation, a unified framework that assigns both semantic labels and instance IDs to each pixel. To evaluate this new task, they proposed the Panoptic Quality (PQ) metric, combining recognition and segmentation accuracy. This work bridged the gap between semantic and instance segmentation, inspiring a wave of unified, end-to-end models for more comprehensive scene understanding.

Chuang et al., [2] review's recent deep learning-based advances, from early two-branch models like Panoptic FPN to more unified, end-to-end approaches such as UPS Net and transformer-based methods like MaskFormer. They highlight key challenges, including class imbalance, overlapping objects, and context integration. The paper also outlines future directions like lightweight models and improved generalization, emphasizing the growing importance of panoptic segmentation in real-world vision tasks.

Elharrouss et al., [3] provide an in-depth review of the field, categorizing methods into top-down and bottom-up approaches, and analyzing their strengths and limitations. They also discuss key datasets like COCO and Cityscapes, along with the Panoptic Quality (PQ) metric used for evaluation. The paper highlights major challenges such as occlusion handling, real-time performance, and consistent segmentation. It concludes with future directions including end-to-end models, domain adaptation, and multi-modal integration to improve robustness and scalability.

Chen et al., [4] proposed a generalist framework that handles both images and videos using a single transformer-based architecture, improving temporal consistency and cross-domain generalization. Their method eliminates the need for task-specific models, offering efficient and robust performance across dynamic environments. The approach highlights the growing trend toward unified, multi-task models capable of handling complex real-world scenarios in panoptic segmentation.

Liu et al., [5] proposed a pioneering end-to-end framework for panoptic segmentation, streamlining the process by integrating both tasks into a single model with shared features, reducing the need for separate pipelines. Their approach improves efficiency and consistency by optimizing both segmentation tasks jointly, marking a significant advancement in unified segmentation architectures.

Zhao et al., [6] enhanced semantic segmentation with the Pyramid Scene Parsing Network (PSPNet), using multi-scale context to capture global scene structures, which influenced later panoptic segmentation models. Chen et al. (2023) extended this by developing a transformer-based framework for both images and videos, improving temporal consistency.

A dynamically instantiated network for pixel-wise instance segmentation was introduced by Arnab and Torr [7]. This approach improves the separation of object instances by dynamically generating instance-specific features during the segmentation process, allowing for better adaptability to varying object shapes, sizes, and complexities. By instantiating separate features for each object, the network is able to more accurately distinguish between overlapping or adjacent instances, a key challenge in instance segmentation.

A token-sparsity-based method for panoptic segmentation, designed for natural scenes, was proposed by Liu et al., [8]. This approach focuses on sparse token representations to reduce computational complexity while maintaining segmentation accuracy. It efficiently handles large-scale natural scene datasets, addressing challenges such as object occlusion and complex backgrounds. This work contributes to optimizing panoptic segmentation for both performance and efficiency, particularly in real-world, resource-constrained applications.

3. METHODOLOGY

The development of a real-time forensic image analysis system using panoptic segmentation involved several key stages, including system design, environment setup, image processing, segmentation, interface development, and performance evaluation.

3.1 System Design and Planning

The project started by defining its main goals: to create an easy-to-use application for forensic image analysis that could accurately detect and label both foreground objects and background elements using panoptic segmentation, while also showing confidence scores for each detected object. A system architecture was laid out, combining a backend powered by Detectron2 for processing and a frontend built with Gradio for user interaction. This setup ensured a smooth and efficient workflow for forensic image interpretation.

3.2 Environment Setup and Configuration

Python was chosen as the main programming language due to its compatibility with deep learning libraries and frontend tools. Key libraries like PyTorch, OpenCV, NumPy, and Gradio were installed to support both backend and frontend operations. Detectron2, a PyTorch-based library designed for object detection and segmentation, was used along with a pre-trained model—panoptic_fpn_R_101_dconv_cascade_gn_3x—sourced from the Detectron2 Model Zoo. The model was loaded using the DefaultPredictor function and set up to run on a CPU to ensure compatibility across systems without requiring high-end GPUs.

3.3 Image Acquisition and Preprocessing

The system supports two input modes: uploading image files and capturing real-time images via a webcam. Once an image is provided, it is resized to a fixed resolution of 640×480 pixels to match the model's input size. This standardization helps ensure uniform processing speed and consistency across different inputs. Since a pre-trained model is used, there was no need for custom dataset training or splitting.

3.4 Segmentation and Inference

The core functionality lies in panoptic segmentation, which combines semantic and instance segmentation to classify each pixel and identify distinct object instances. Once an image is submitted, the system processes it through the model, which outputs both a segmentation mask and confidence scores indicating the model's certainty for each detected region or object.

3.5 Visualization and Output

After segmentation, the results are visualized using the Visualizer module from Detectron2. This module overlays color-coded segmentation masks and labels onto the original image for clear understanding. The annotated image is then displayed in the frontend, allowing users to easily interpret the results.

3.6 Web Interface Development

The web interface was built using Gradio, offering a lightweight, browser-based experience. Users can upload images or capture them in real time for analysis. Once processed, the segmented output is shown directly in the browser. Gradio's sharing feature also allows the generation of public links, making it easy to demonstrate the tool or collaborate remotely. This makes the system practical for both individual use and group investigations.

3.7 Evaluation and Testing

The system was tested on various forensic- like images to assess performance. Evaluation criteria included:

- Visual accuracy: Ensuring correct object and background detection through manual inspection.
- Label clarity: Verifying that segmentation masks were well- defined and correctly labeled.
- Confidence score reliability: Checking if the confidence levels matched the actual accuracy of detection.
- Interface performance: Testing responsiveness and usability across different devices and browsers.

3.8 Deployment Strategy

Initially, the system was deployed locally, using Gradio to provide both local and remote access through a public URL. This approach allowed for quick testing, easy demonstrations, and user feedback without complex deployment setups.

Additionally, the option to generate shareable links for remote access proved useful for field deployment and collaborative investigations. Overall, the system demonstrated strong potential as a practical tool for real-time forensic image analysis.

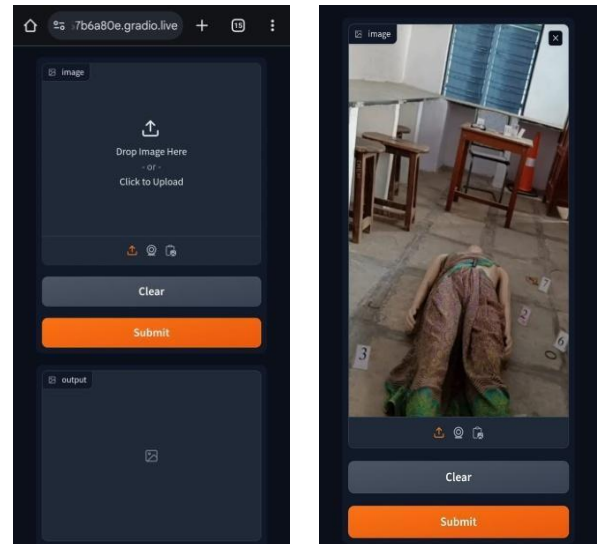
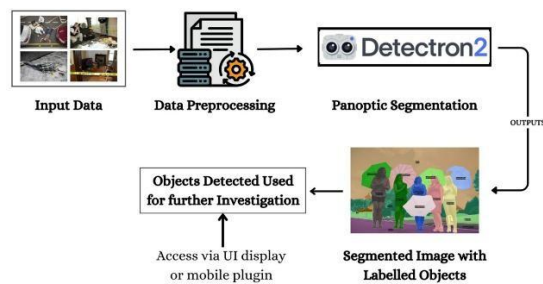


Fig 4.1: Frontend Web Interface

Fig 3.8.1 System Architecture



4. RESULTS

The panoptic segmentation system was tested on a variety of forensic images to evaluate its performance in segmentation accuracy, object detection, and confidence scoring. The model effectively segmented both foreground objects and background regions with high precision, clearly outlining object boundaries and accurately labelling individual instances. The confidence scores generated by the system reliably reflected the model's certainty, with higher scores corresponding to more accurate detections— an important factor for verifying segmentation quality in forensic applications. The visual output was clear and informative, with color-coded masks and labels overlaid on the original images, making it easy to identify and distinguish different objects. The Gradio-based interface performed smoothly, enabling users to upload or capture images and instantly view segmentation results.

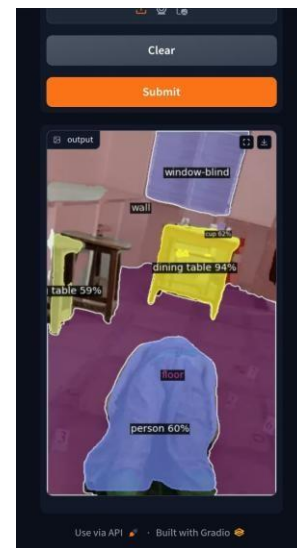


Fig 4.2: Segmented Forensic Image with Object Labels and Confidence Scores

5. CONCLUSION AND FUTURE WORKS

This research successfully demonstrates the potential of a deep learning-powered panoptic segmentation system using Detectron2 for enhanced forensic image analysis. By integrating semantic and instance segmentation into a unified framework, the model enables precise object detection, classification, and detailed scene interpretation—key factors in reconstructing crime scenes and identifying critical evidence. The use of a pre-trained model ensures a strong baseline performance, and the integration with a user-friendly web interface allows real-time interaction, making the tool practical for both field deployment and collaborative investigations. Despite these positive outcomes, certain limitations were observed when applying the model in real-world forensic settings. Specifically, the system may face difficulty when interpreting images with poor lighting, low resolution, or highly cluttered scenes. Moreover, the model's generalization ability may be restricted due to the lack of domain-specific forensic datasets, which can affect prediction accuracy and reliability.

To overcome these challenges, future work will focus on incorporating domain adaptation techniques to better tailor the model to forensic environments. Additionally, enhancing interpretability through explainable AI (XAI) features can help forensic experts understand the model's decisions more transparently. Another promising direction involves implementing automated report generation, which could streamline documentation by summarizing detected evidence, object types, and their confidence scores—saving time and reducing manual effort.

Moreover, expanding the current system to support video-based panoptic segmentation can lead to real-time analysis capabilities, enabling continuous monitoring in surveillance systems or live crime scene investigations. This progression could further benefit related domains such as smart security and autonomous navigation, thus expansion the scope and societal impact of the developed solution. In conclusion, the proposed system lays a solid foundation for intelligent forensic analysis. With continued improvements and targeted enhancements, it holds strong potential for deployment in real-world forensic operations and other vision-driven applications.

6. REFERENCES

- [1] Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp.9404-9413)
- [2] Chuang, Y., Zhang, S., & Zhao, X. (2023). Deep learning-based panoptic segmentation: Recent advances and perspectives. *IET Image Processing*, 17(10), 2807-2828.
- [3] Elharrouss, O., Al-Maadeed, S., Subramanian, N., Ottakath, N., Almaadeed, N., & Himeur, Y. (2021). Panoptic segmentation: A review. *arXiv*
- [4] Chen, T., Li, L., Saxena, S., Hinton, G., & Fleet, D. J. (2023). A generalist framework for panoptic segmentation of images and videos. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 909-919).
- [5] Liu, H., Peng, C., Yu, C., Wang, J., Liu, X., Yu, G., & Jiang, W. (2019). An end-to-end network for panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6172-6181).
- [6] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881-2890).
- [7] Arnab, A., & Torr, P. H. (2017). Pixelwise instance segmentation with a dynamically instantiated network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 441-450).
- [8] Liu, H., Zhang, P., Chen, D., & Fu, J. (2024, July). A Token-Sparsity-Based Image Panoptic Segmentation Method for Natural Scenes. In *2024 43rd Chinese Control Conference (CCC)* (pp. 8051-8056). IEEE.