Exploring AI: Detecting Objects in a Walk Through SRH Berlin with YOLO

Supervisor: Prof. Dr. Kristian Rother

SRH University of Applied Sciences Berlin

Berlin, Germany, March 18th, 2025

by

Amin Niaziardekani (3119828)
Swapnaneel Sarkhel (3119924)
Ajdin Buljko (3119754)

1. Introduction:

Artificial Intelligence (AI) has come a long way, especially in computer vision, where it can now detect objects in real time. This project explores how well the YOLO (You Only Look Once) model performs in a real-world setting by applying it to a recorded video while walking through the SRH Berlin building. Using the YOLOv8 model, the system processes each video frame to identify and classify objects, showcasing its practical use in everyday environments. The results show that while the model is quite effective at detecting common objects, it still faces challenges, particularly with smaller objects. This report walks through the methodology, results, and limitations of the project while discussing ethical considerations in AI-driven surveillance and privacy, offering insights into the current state of AI-driven object detection and areas for improvement.

2. YOLO Model:

YOLO is a widely used deep learning model for object detection, first introduced by Redmon et al. in the paper *You Only Look Once: Unified, Real-Time Object Detection*. Unlike traditional detection models that scan an image multiple times, YOLO processes the entire image in a single forward pass, making it exceptionally fast and efficient. This project utilizes YOLOv8, an improved version with enhanced accuracy and performance compared to previous iterations.

3. Methodology:

To conduct this experiment, the following steps were taken:

- A video was recorded while walking through the SRH Berlin building.
- The YOLOv8 model was loaded using the Ultralytics YOLO library.
- The recorded video was processed frame by frame.
- Object detection was performed on each frame.
- Bounding boxes and labels were drawn around detected objects.
- The processed video was saved with annotations.

4. Dependencies:

To begin implementing the Object Detection in Video project, the following Python packages need to be installed. This code with install all the necessary libraries.

```
pip install ultralytics opencv-python torch torchvision numpy
```

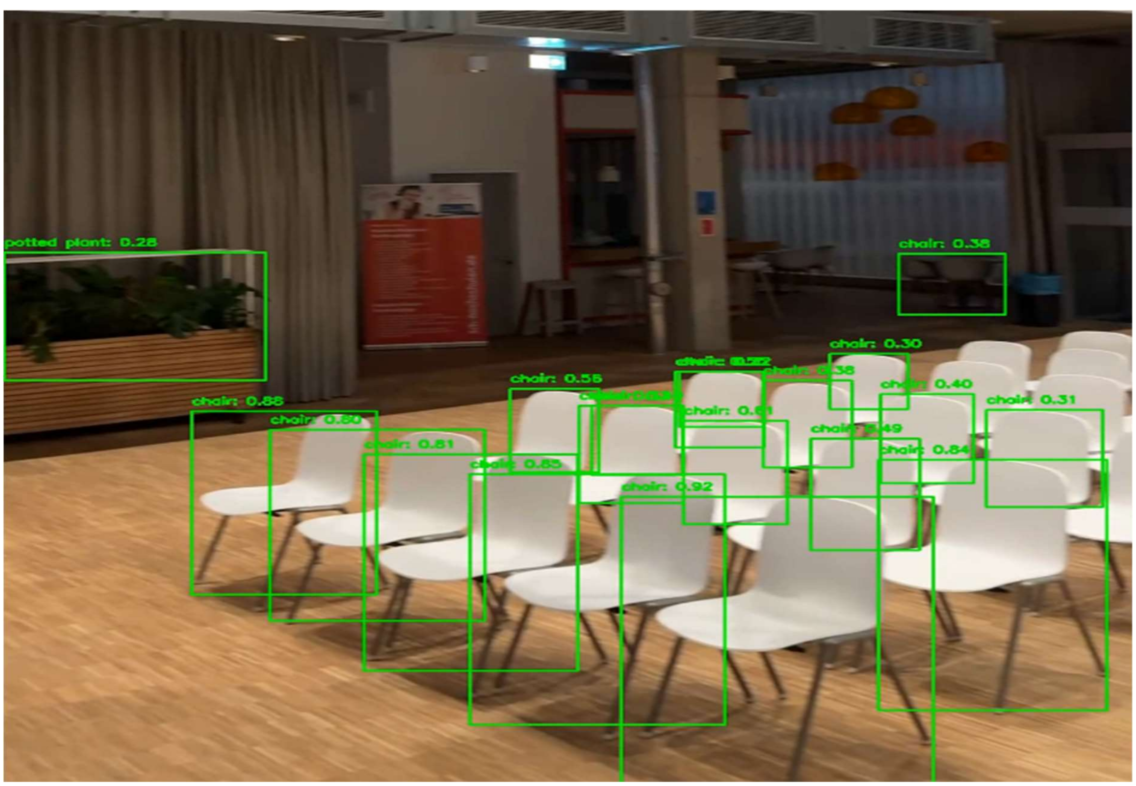
Code and Output:

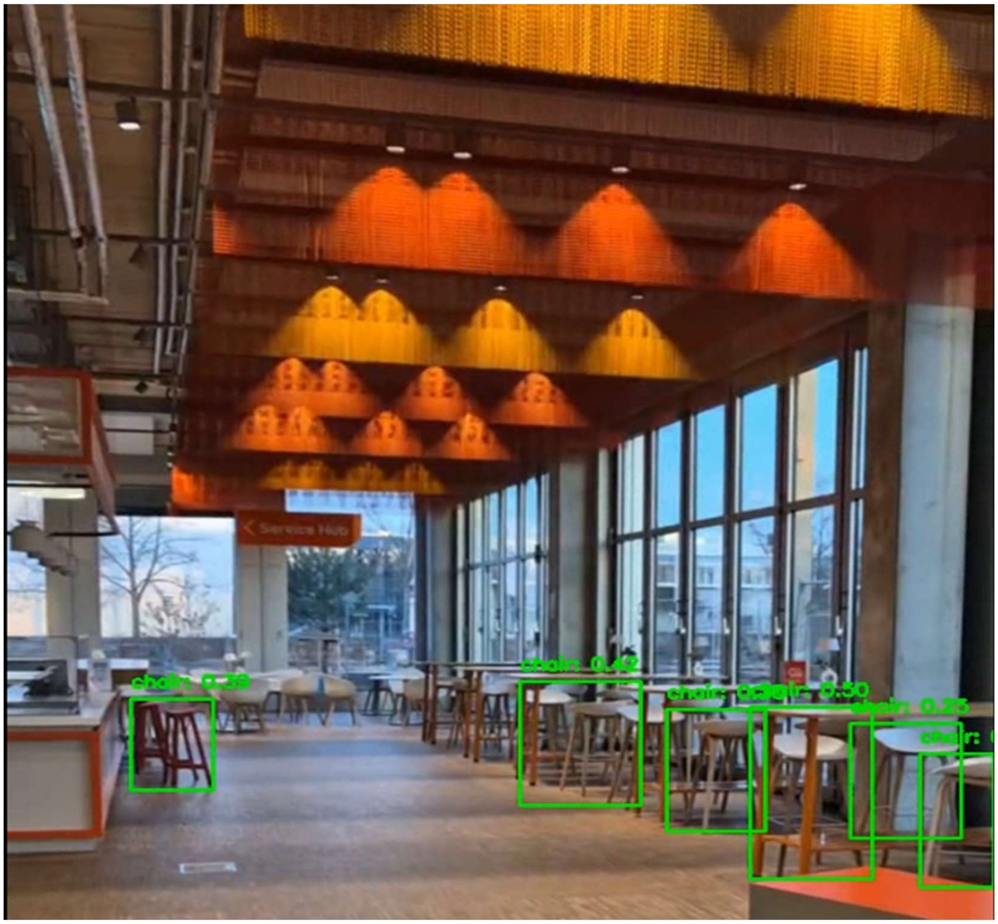
The full code, the original and the output videos can be found in this git repository: https://github.com/amin4554/YoloImageDetection

5. Results

The processed video successfully displayed real-time detection of various objects, such as furniture, and electronic devices. The model's performance was generally accurate, with clear bounding boxes and confidence scores. However, detection quality varied based on lighting conditions.

Screenshots:





6. YOLOv8n vs YOLOv8s

We tested two versions of YOLOv8—the nano (YOLOv8n) and small (YOLOv8s)—to see how well they could detect objects in our video. Here's what we found:

1. Accuracy & Detection Quality

- YOLOv8s was noticeably more accurate. It detected objects with confidence levels between 45% and 95%, while YOLOv8n often had lower confidence (as low as 30%).
- Smaller objects (like the potted plant) were easier to detect with YOLOv8s, whereas YOLOv8n sometimes missed them or misclassified them.
- YOLOv8n had more false positives, meaning it sometimes detected things that weren't actually there.
-

2. Speed & Performance

- YOLOv8n was faster, running at about 35 frames per second (FPS), while YOLOv8s ran at around 25 FPS.
- If you're running the model on a CPU, YOLOv8n performs much better since it's lighter. But if you have a GPU, YOLOv8s is still quite fast.

Detection Performance

The YOLO model does a solid job of detecting objects in the scene, especially the chairs, which are identified with confidence scores ranging from 0.30 to 0.92. This means the model is generally reliable, though it struggles a bit when objects overlap.

A potted plant in the background was also detected, but with a lower confidence score of 0.28. This suggests that smaller or partially hidden objects might be harder for the model to recognize with certainty.

Overall, the bounding boxes are well-placed, meaning the model is correctly identifying and locating objects within the scene.

Strengths

- The model performs consistently well in a busy environment, detecting multiple objects even when they are close together.
- It effectively differentiates between overlapping chairs, showing that it can handle cluttered scenes without major issues.
- Since YOLO processes frames in real time, it's well-suited for live video applications, making object detection quick and efficient.

Unexpected Model Behaviour

In some instances, the model incorrectly labelled objects or failed to detect them altogether, particularly when objects overlapped or were in motion. This limitation highlights the challenge of object detection in dynamic environments, where occlusion and perspective changes can impact model performance.

Limitations & Observations

- Some bounding boxes overlap, which may affect the clarity of object differentiation, particularly when objects are close together.
- The potted plant detection confidence is low (0.28), indicating that objects with non-standard shapes or textures may be harder for YOLO to classify with high certainty.
- A few chairs have low confidence values (around 0.30-0.40), suggesting that lighting conditions or partial occlusion might affect detection reliability.
- There are no misclassifications, meaning that the model did not incorrectly label any objects, but its ability to detect other types of furniture or decor is unknown.

Potential Improvements

- Using a higher-resolution YOLO model (e.g., YOLOv8s or YOLOv8m) could improve detection accuracy, especially for smaller objects.
- Fine-tuning the model on a dataset specific to indoor university environments might enhance recognition accuracy for objects commonly found in classrooms

or conference spaces.
- Post-processing techniques, such as Non-Maximum Suppression (NMS) tuning, could reduce overlapping bounding boxes, leading to clearer object differentiation.

7. Challenges

- Small objects were sometimes missed or misclassified.
- Performance was affected by changes in lighting.
- The lightweight "yolov8n.pt" model was used for speed; "yolov8s.pt" model could yield better accuracy.
- The model struggled with highly cluttered scenes where multiple objects overlapped.

8. Ethical Considerations

Object detection technology raises important ethical concerns, particularly regarding privacy and surveillance. In this project, no personal data was collected, and ethical guidelines were followed to ensure responsible use of AI. However, deploying such models in real-world scenarios requires compliance with data protection laws and ethical AI standards (Such as the GDPR and the EUs Trustworthy AI Guidelines).

9. Conclusion

This project successfully demonstrated the capabilities of YOLO in detecting objects in an indoor environment. While the model performed well in most cases, challenges such as lighting variations and partial obstruction highlighted areas for improvement. Future work could involve using a more advanced YOLO model and refining detection parameters to enhance accuracy and robustness.

10. References

- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. CVPR.
- Ultralytics YOLO documentation: https://github.com/ultralytics/ultralytics