

# Abschlussbericht – videobasierte Vandalismuserkennung an Bahnhöfen

Forschungsprojekt TAHAI - TrustAdHocAI

Sandro Hartenstein | 28.2.2025 | Berlin

# Agenda



- Forschungsprojekt & Anwendungsfall
- Wissenschaftliche Ansätze
- Prototypische Evaluation
- Ergebnisse
- Erkenntnisse
- Fazit & Empfehlung



# Forschungsprojekt & Anwendungsfall

# Forschungsprojekt TAHAI



# Anwendungsfall

## Vandalismuserkennung

- Ziel:
  - Erkennen von Vandalismus (insbesondere Grafitysprayer)
- Herausforderungen:
  - Nutzung vorhandener Videoanlagen (100+ Anzahl Kameras)
  - Echtzeiterkennung
  - Geringe False Positives notwendig
- Status:
  - Experimentelle Daten (Videos) sind am Bahnhof Südkreuz erstellt wurden
  - Ansätze im Vergleich



Bild von [Gerd Altmann](#) auf [Pixabay](#)



# Wissenschaftliche Ansätze

# Möglichkeiten der Erkennung



<b>Methode</b>	<b>Prinzip</b>	<b>Vorteile</b>	<b>Nachteile</b>
<b>OpenCV</b>	Klassische Algorithmen	Einfach, geringe Rechenanforderungen	Ungenau, anfällig für Fehlalarme
<b>YOLO</b>	Echtzeit-Objekterkennung	Schnell, genau	Hohe Hardware-Anforderungen
<b>CRNN</b>	CNNs + RNNs	Lernt räumliche und zeitliche Informationen	Rechenintensiv, große Datensätze erforderlich
<b>Conv-LSTM-SVM</b>	CNNs + LSTMs + SVMs	Erkennt komplexe Verhaltensmuster	Rechenintensiv, erfordert Parametereinstellung

- Prinzip: Diese Methode basiert auf klassischen Algorithmen der Computer Vision, wie z.B. Kantenerkennung, Objekterkennung und Mustererkennung.
- Vorteile: Einfach zu implementieren, geringe Rechenanforderungen.
- Nachteile: Kann bei komplexen Szenarien oder schlechten Lichtverhältnissen ungenau sein.  
Anfällig für Fehlalarme.



<https://cvl.tuwien.ac.at/teaching/student-topics/change-detection-in-graffiti-images/>

Quellen:

Ghazal, Mohammed; Vazquez, Carlos; Amer, Aishy (2007): **Real-time automatic detection of vandalism behavior in video sequences**. In: 2007 IEEE International Conference on Systems, Man and Cybernetics. SMC 2007] ; Montreal, QC, Canada, 7 - 10 October 2007. 2007 IEEE International Conference on Systems, Man and Cybernetics. Montreal, QC, Canada, 10/7/2007 - 10/10/2007. Institute of Electrical and Electronics Engineers. Piscataway, NJ: IEEE Service Center, S. 1056–1060.



# YOLO (You Only Look Once)

- Prinzip: YOLO ist ein Echtzeit-Objekterkennungsalgorithmus, der das gesamte Bild auf einmal analysiert und Objekte mit hoher Geschwindigkeit und Genauigkeit lokalisieren und klassifizieren kann.
- Vorteile: Sehr schnell und genau, ideal für Echtzeit-Anwendungen.
- Nachteile: Benötigt eine leistungsstarke GPU für optimale Leistung.

Quellen:

Poirier, Fabien (2024): **Real-Time Anomaly Detection in Video Streams**. Online verfügbar unter [https://www.researchgate.net/publication/386335109\\_Real-Time\\_Anomaly\\_Detection\\_in\\_Video\\_Streams](https://www.researchgate.net/publication/386335109_Real-Time_Anomaly_Detection_in_Video_Streams), zuletzt aktualisiert am 29.11.2024.



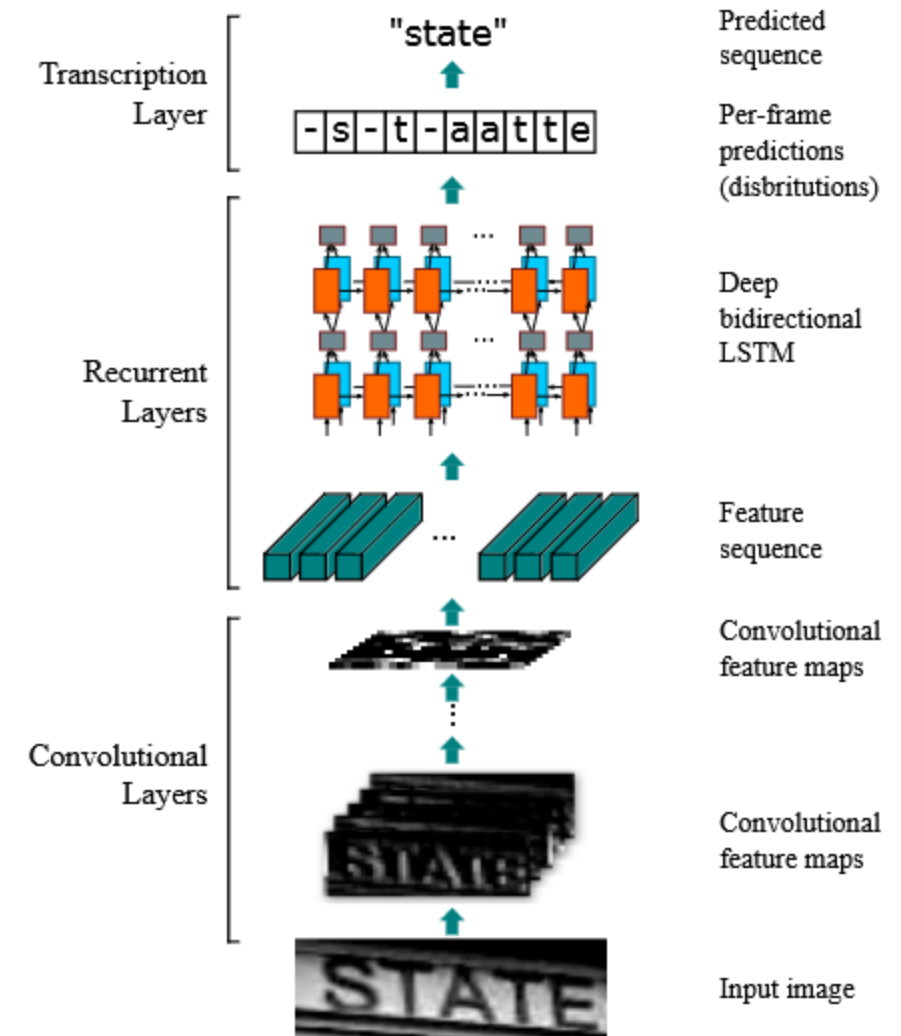
<https://universe.roboflow.com/danger-graffiti-project/dangergraffiti/images/HweSwQLmnz4KRDgVRpUo>

# Convolutional Recurrent Neural Network (CRNN)

- Prinzip: CRNN kombiniert Convolutional Neural Networks (CNNs) zur Extraktion räumlicher Merkmale mit Recurrent Neural Networks (RNNs) zur Verarbeitung sequenzieller Daten (z.B. Videoframes).
- Vorteile: Kann sowohl räumliche als auch zeitliche Informationen aus Videosequenzen lernen und analysieren.
- Nachteile: Kann rechenintensiv sein und erfordert große Datensätze zum Trainieren.

Quelle:

Poirier, Fabien (2024): **Real-Time Anomaly Detection in Video Streams**. Online verfügbar unter [https://www.researchgate.net/publication/386335109\\_Real-Time\\_Anomaly\\_Detection\\_in\\_Video\\_Streams](https://www.researchgate.net/publication/386335109_Real-Time_Anomaly_Detection_in_Video_Streams), zuletzt aktualisiert am 29.11.2024.



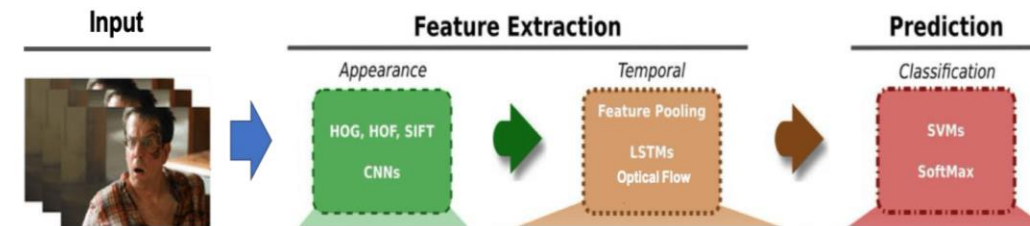
<https://arxiv.org/pdf/1507.05717>

# Conv-LSTM-SVM

**Prinzip:** Diese Methode kombiniert Convolutional Neural Networks (CNNs) zur Merkmalsextraktion mit Long Short-Term Memory (LSTM) Netzwerken zur Erfassung zeitlicher Abhängigkeiten und Support Vector Machines (SVMs) zur Klassifizierung.

**Vorteile:** Kann komplexe Verhaltensmuster und zeitliche Beziehungen in Videodaten erkennen.

**Nachteile:** Kann rechenintensiv sein und erfordert sorgfältige Parametereinstellung.



<https://www.mdpi.com/2076-3417/13/3/2007>

Quelle:

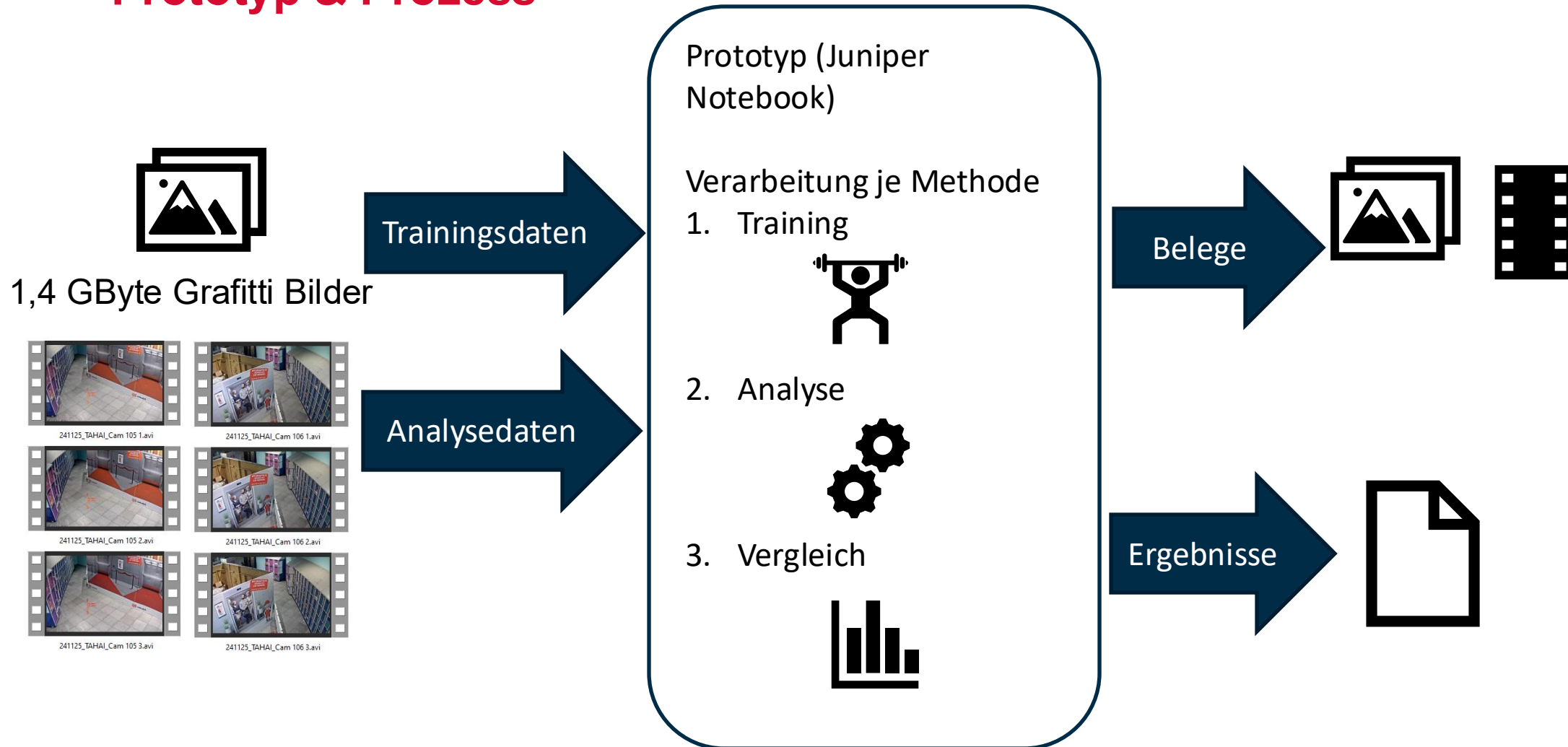
Poirier, Fabien (2024): **Real-Time Anomaly Detection in Video Streams**. Online verfügbar unter [https://www.researchgate.net/publication/386335109\\_Real-Time\\_Anomaly\\_Detection\\_in\\_Video\\_Streams](https://www.researchgate.net/publication/386335109_Real-Time_Anomaly_Detection_in_Video_Streams), zuletzt aktualisiert am 29.11.2024.



# Prototypische Evaluation



# Prototypische Evaluation Prototyp & Prozess





# Ergebnisse

## Begriffserklärungen:

- **Erkennungsrate:** Prozentsatz der als positiv klassifizierten Frames
- **Präzision:** Anteil korrekter Erkennungen an allen Erkennungen
- **Recall:** Anteil erkannter Vorfälle an tatsächlichen Vorfällen
- **F1-Score:** Harmonisches Mittel aus Präzision und Recall (Balance-Metrik)

# Ergebnisse Überblick



Methode	F1-Score	Erkennungsrate	Falsch-Positive	Verarbeitungszeit
YOLO v11	0,50	11,4%	11	90,7s
OpenCV	0,12	58,0%	63	106,1s
CRNN	0,20	49,7%	35	386,2s
CNN-LSTM-SVM	0,10	100,0%	73	420,2s

- **YOLO-basierter Ansatz:** Beste Balance zwischen Präzision und Geschwindigkeit
- **Hohe Falsch-Positiv-Raten:** Alle Methoden weisen zu viele Fehlalarme auf
- **Schwellenwerte zu niedrig:** Durchgehend übermäßige Sensitivität festgestellt
- **Effizienz vs. Komplexität:** Komplexere Modelle zeigen keine bessere Leistung

# Ergebnisse

## Erkennungsleistung: Recall vs. Präzision

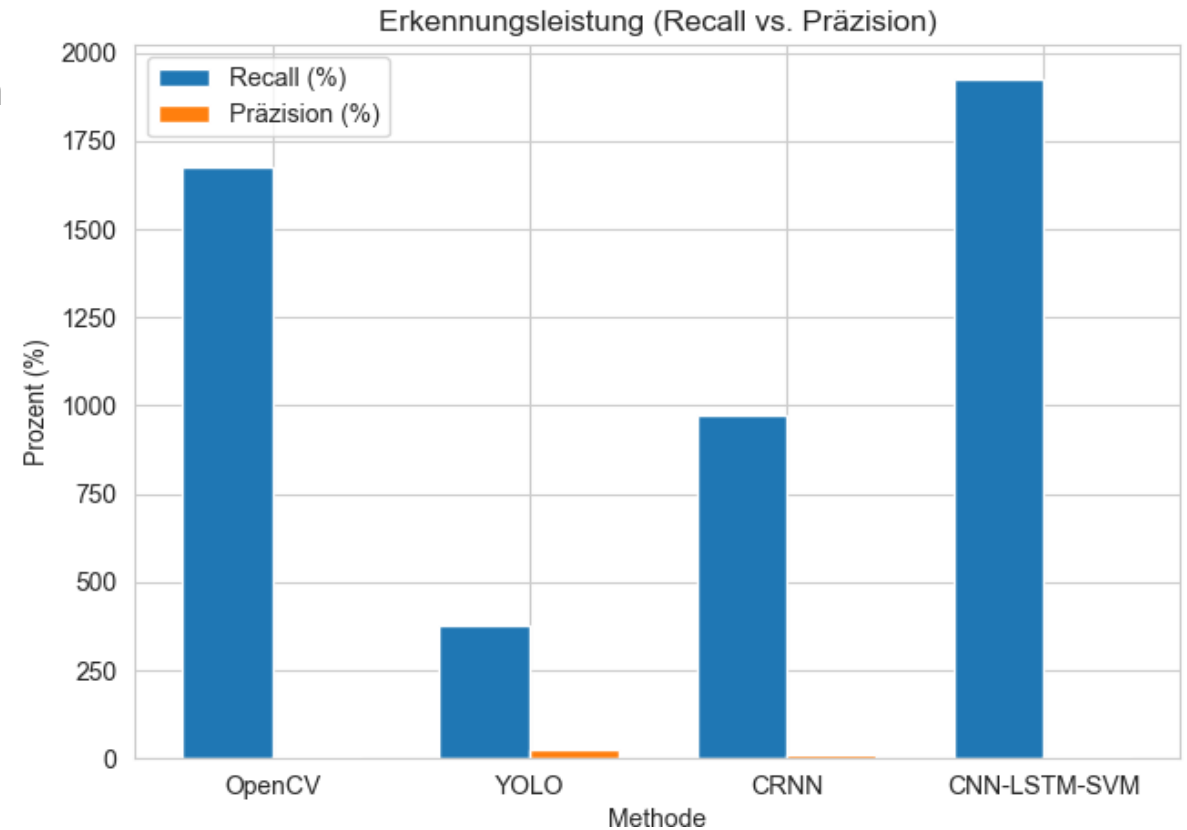
### Überempfindlichkeit aller Methoden

- CNN-LSTM-SVM: 1925% Recall, 5,19% Präzision
- OpenCV: 1675% Recall, 5,97% Präzision
- CRNN: 975% Recall, 10,26% Präzision
- YOLO: 375% Recall, 26,67% Präzision

### 100% Frame-Erkennungsrate bei CNN-LSTM-SVM

- Klassifiziert jeden Frame als "Graffiti"
- Paradoxerweise niedrigste Präzision trotz höchster Erkennungsrate

Hoher Recall, niedrige Präzision = Viele Fehlalarme

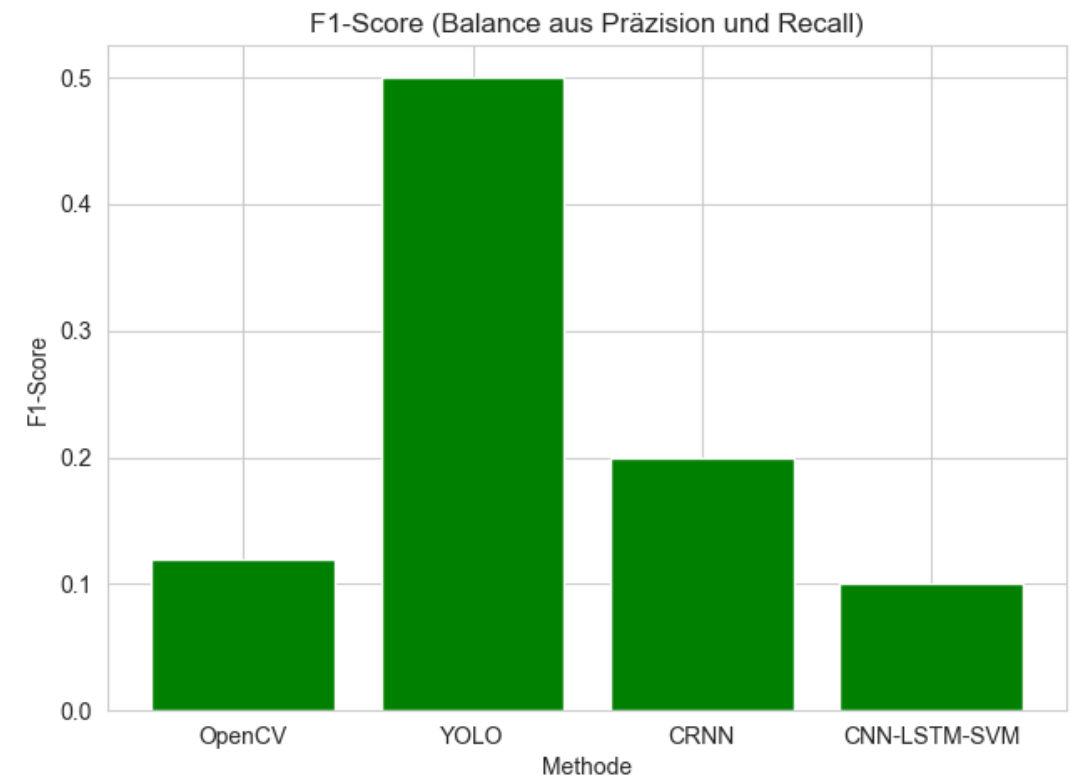




# Ergebnisse

## F1 Score

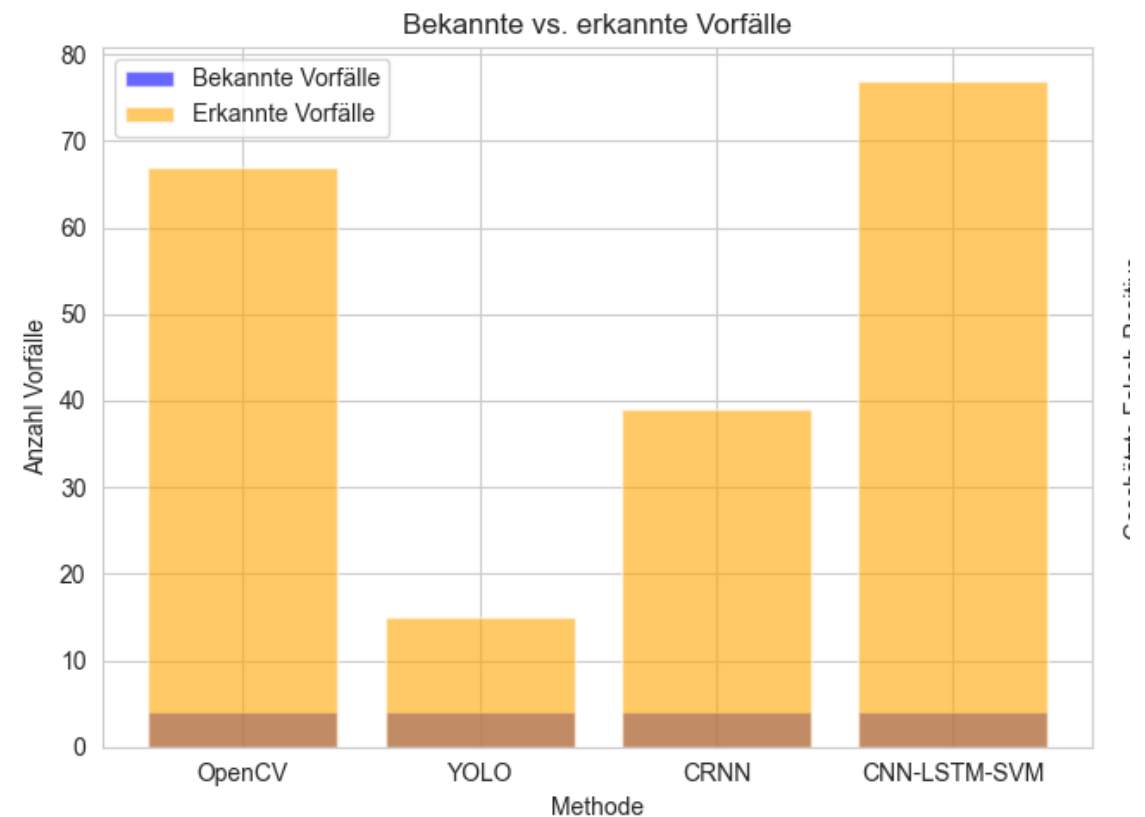
- YOLO zeigt mit  $\sim 0,50$  den deutlich höchsten F1-Score aller Methoden
- CRNN liegt mit  $\sim 0,20$  auf dem zweiten Platz
- OpenCV und CNN-LSTM-SVM haben beide niedrige F1-Scores ( $\sim 0,10$ )
- **Überraschend:** Das komplexeste Modell (CNN-LSTM-SVM) erzielt einen der schlechtesten Werte
- Der große Unterschied (5x höher) zwischen YOLO und den schwächsten Methoden ist bemerkenswert
- **Bestätigt:** Komplexität führt nicht automatisch zu besseren Ergebnissen
- **YOLO bietet die beste Balance zwischen Präzision und Recall**



# Ergebnisse

## Erkannte Vorfälle

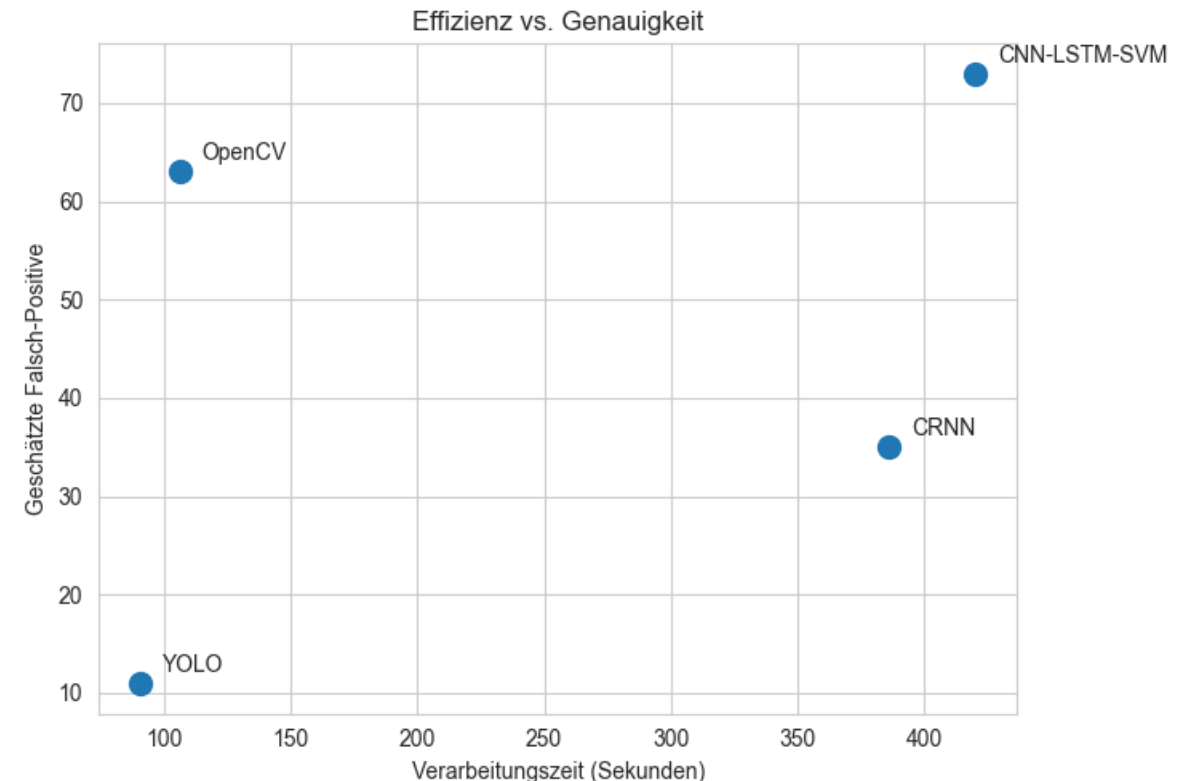
- alle Methoden erkennen deutlich mehr Vorfälle als tatsächlich bekannt sind (4 Vorfälle)
- **CNN-LSTM-SVM** erkennt mit ~77 die meisten Vorfälle
- **OpenCV** folgt mit ~67 erkannten Vorfällen
- **CRNN** erkennt ~39 Vorfälle
- **YOLO** erkennt mit ~15 die wenigsten Vorfälle, ist aber immer noch fast 4x höher als die tatsächliche Anzahl
- **Massive Überempfindlichkeit aller Systeme erkennbar**
- **YOLO** zeigt die geringste Überempfindlichkeit (niedrigster gelber Balken)
- Die große Diskrepanz deutet auf viele Falsch-Positive bei allen Methoden hin



# Ergebnisse

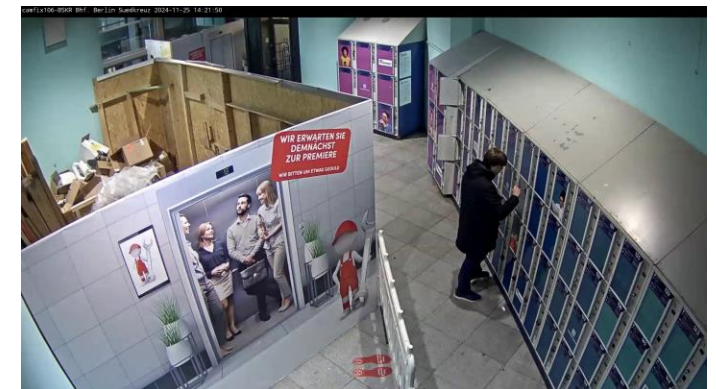
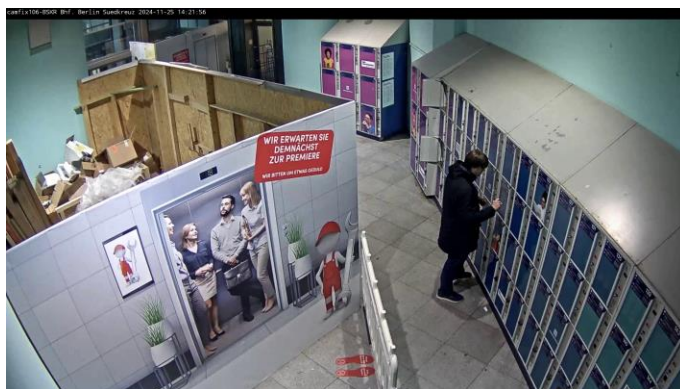
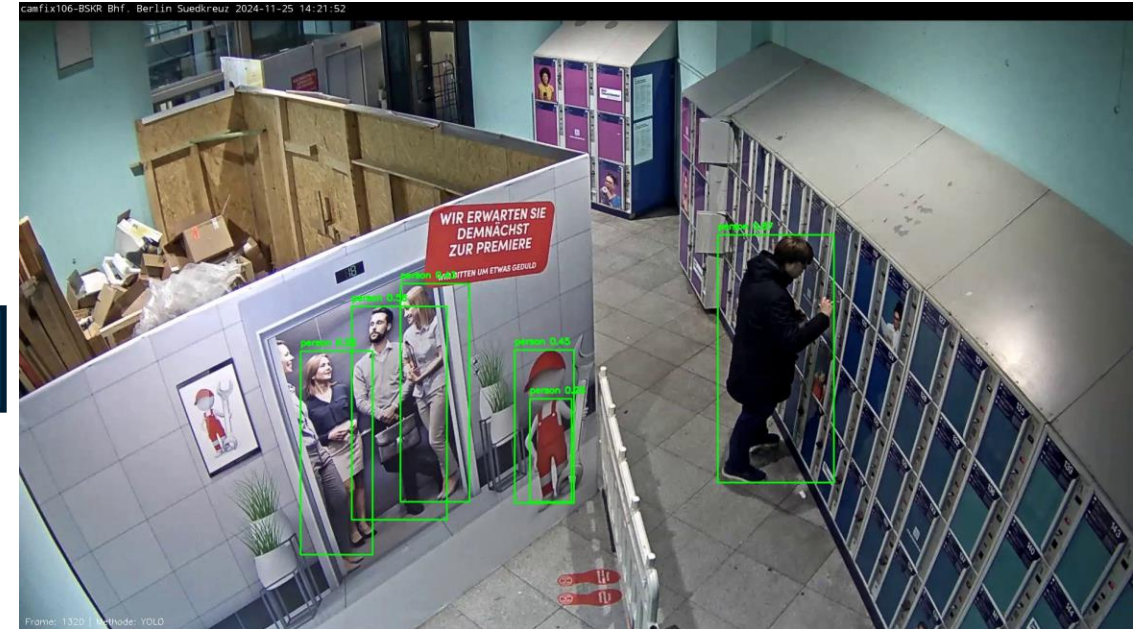
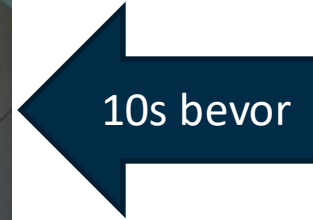
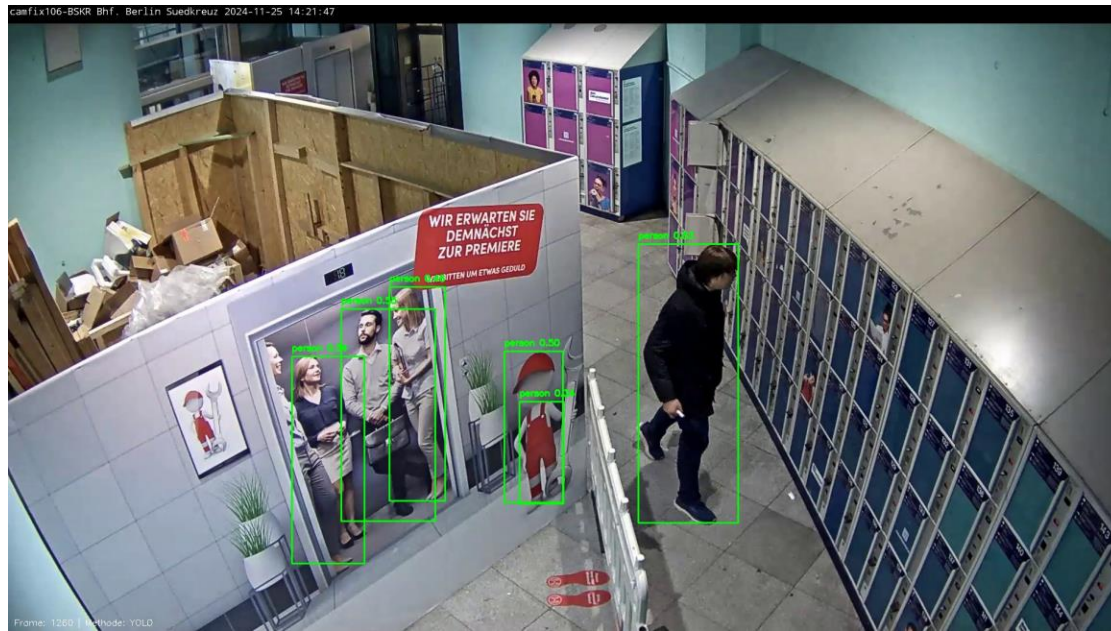
## Effizienz vs Genauigkeit

- **YOLO** zeigt die beste Gesamtperformance: niedrigste Falsch-Positive (~11) und schnellste Verarbeitungszeit (~100s)
- **OpenCV** ist fast so schnell wie YOLO (~120s), erzeugt aber deutlich mehr Falsch-Positive (~63)
- **CRNN** benötigt mehr Zeit (~380s) bei mittlerer Anzahl Falsch-Positiver (~35)
- **CNN-LSTM-SVM** zeigt die schlechteste Kombination: längste Verarbeitungszeit (~420s) und meiste Falsch-Positive (~73)
- Ideale Position wäre links unten (schnell und genau) – YOLO kommt diesem Ideal am nächsten
- Klarer Trade-off zwischen Geschwindigkeit und Genauigkeit erkennbar
- **Die aufwändigeren neuronalen Netzwerke (rechte Seite) bieten keinen Genauigkeitsvorteil trotz höherem Rechenaufwand**



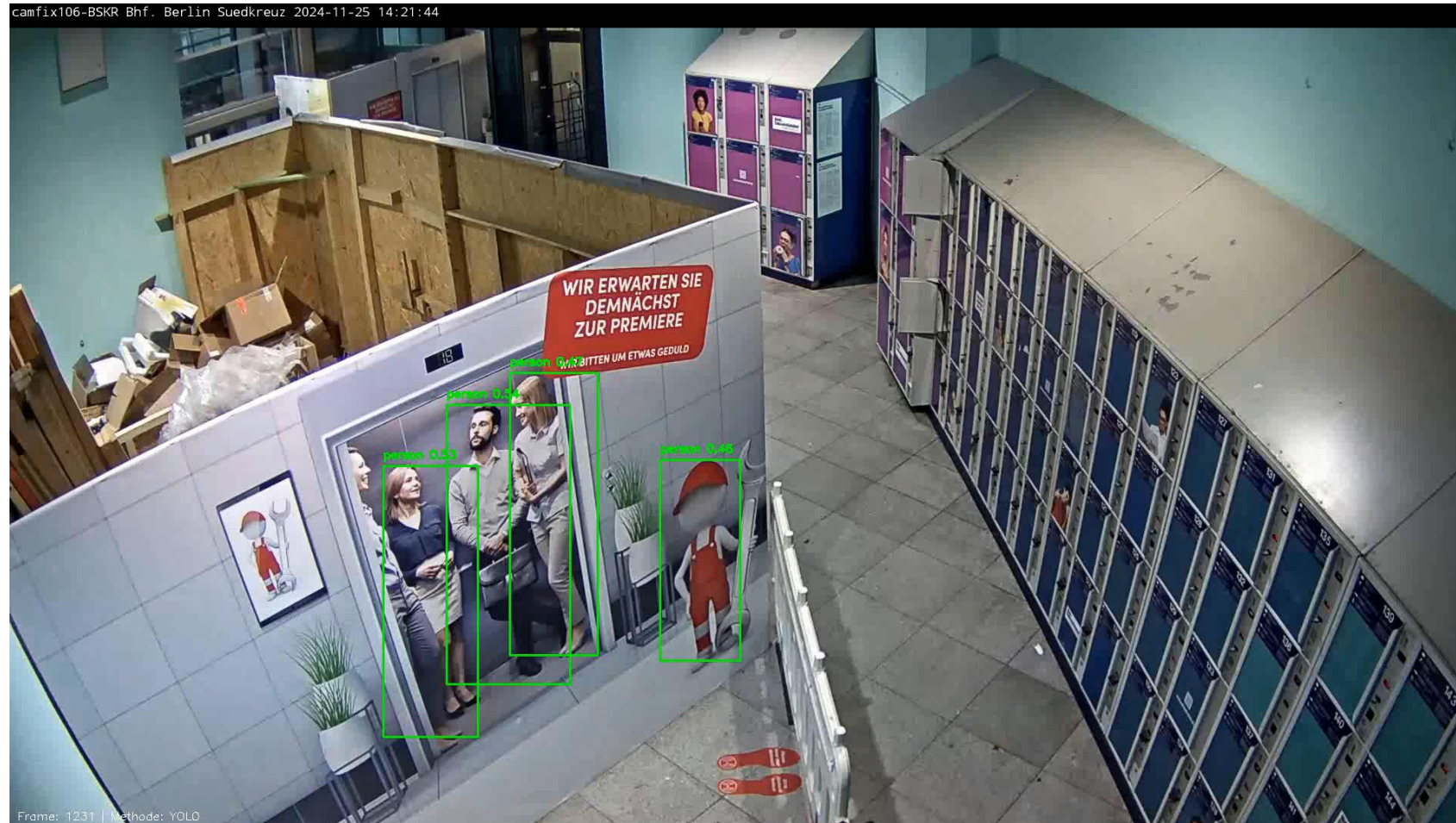


# Yolo11 Framebelege





# Yolo11 Videobelege



# Yolo11

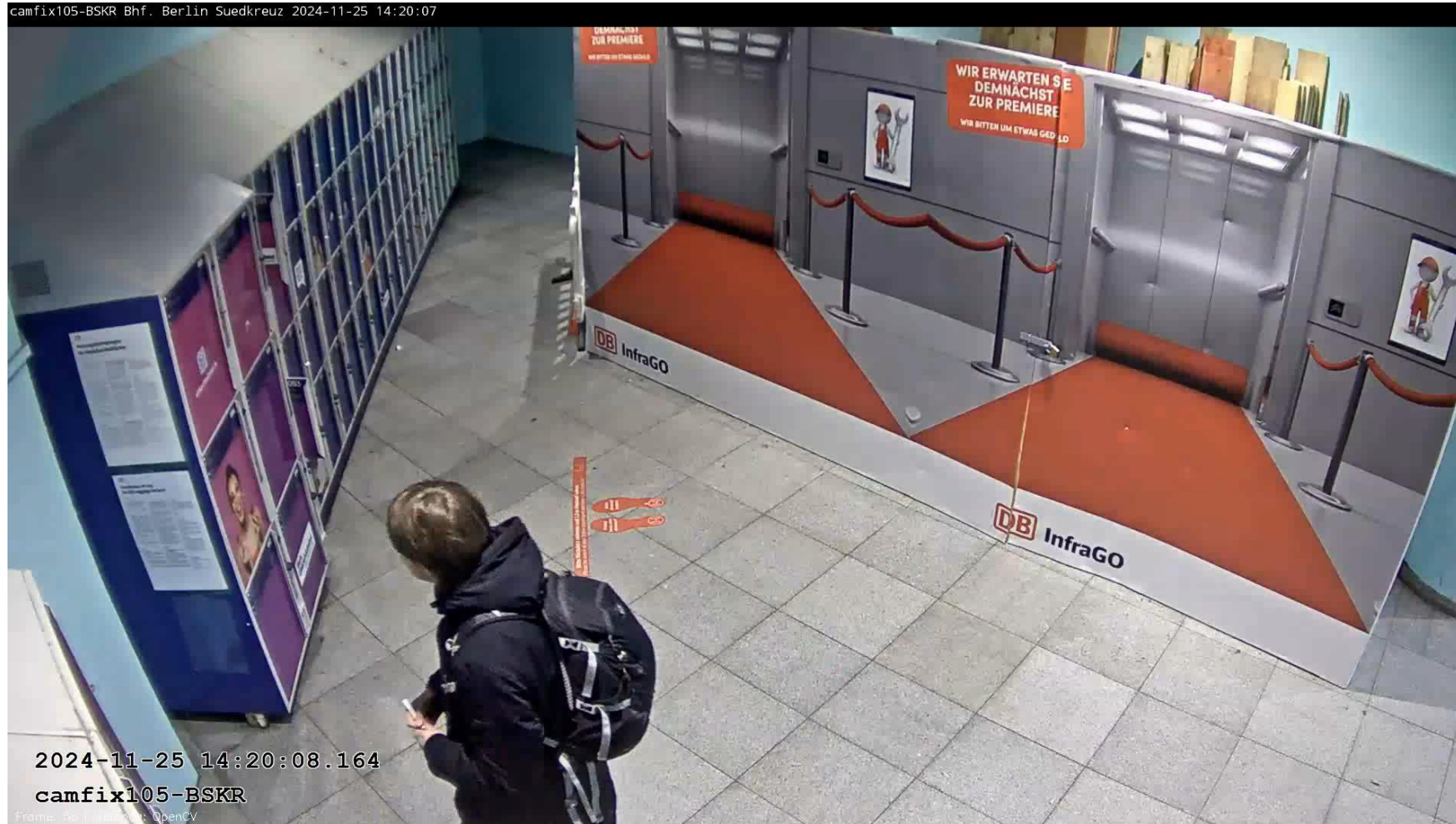
## Videobelege CAM 105 / clip 2





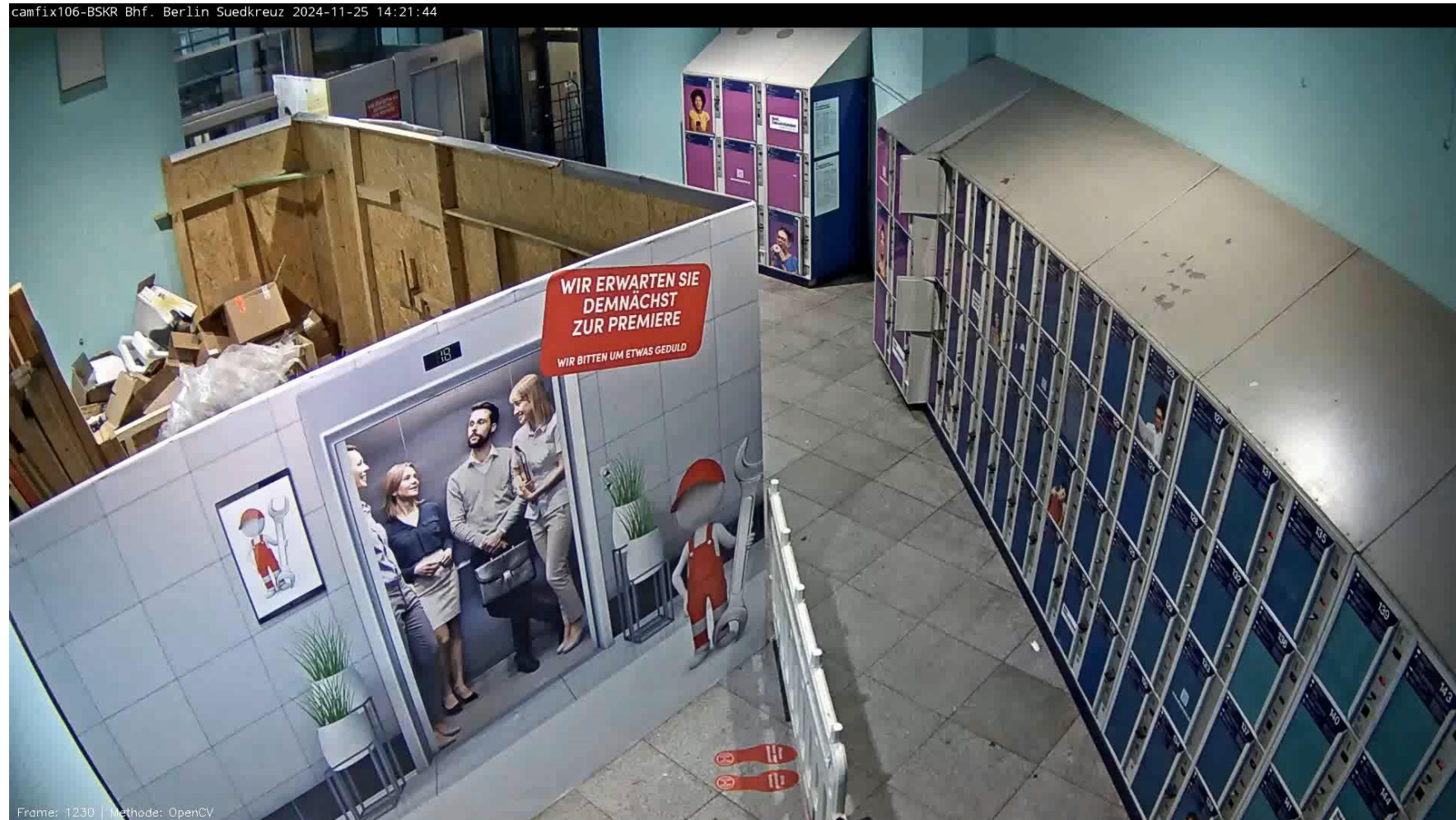
# OpenCV

## Videobelege 1/24 CAM 105 clip 2



# OpenCV

## Videobelege 2/24 CAM 106 clip 2







# Erkenntnisse

## Überragende Bedeutung der Präzision

- Zu viele Falschmeldungen führen zu "Alarmermüdung"
- YOLO mit nur 11 Falsch-Positiven klar überlegen

## Konservative Ansätze überlegen

- Trotz niedrigster Erkennungsrate bietet YOLO die beste Gesamtperformance

## Komplexität ≠ bessere Leistung

- Das komplexeste Modell zeigt die schlechteste Gesamtleistung
- Einfachere, gut optimierte Methoden übertreffen komplexere Ansätze

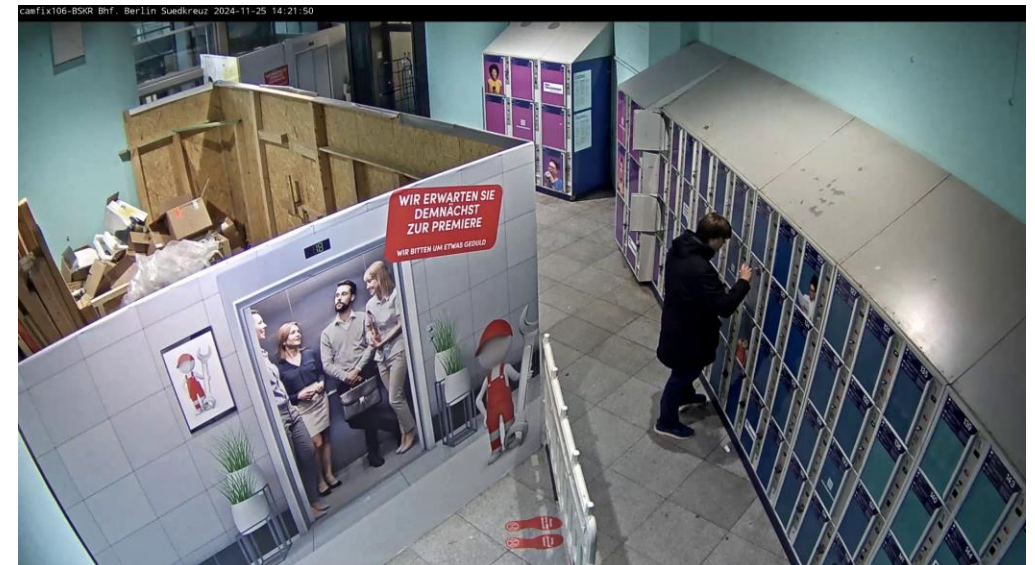
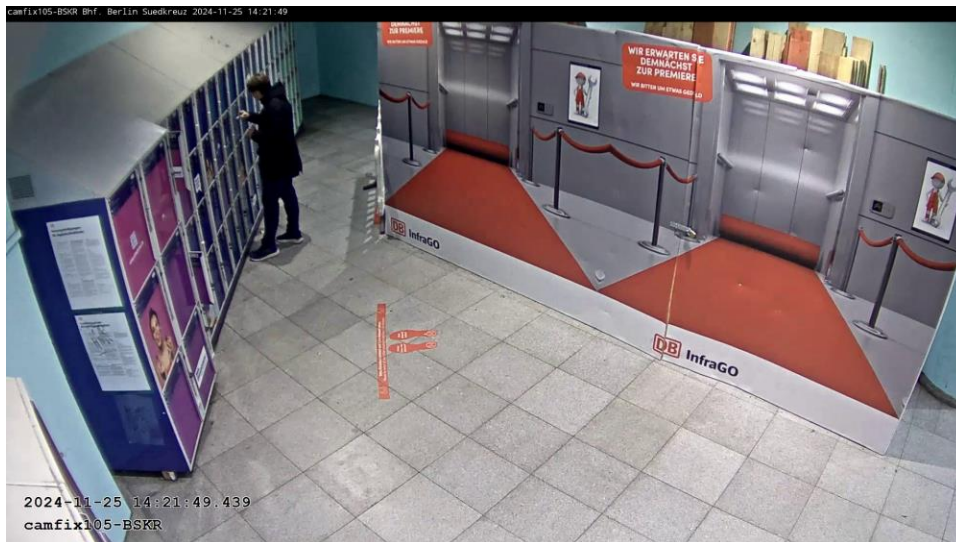
## Schwellenwertoptimierung entscheidend

- Alle Methoden benötigen signifikant höhere Erkennungsschwellen

## Kamerasynchronisationsherausforderung

- Mehrfachzählung von Vorfällen durch verschiedene Kameras
- Derselbe Vorfall wird von 2 Kameras als 2 separate Vorfälle gewertet
- Verfälscht die tatsächliche Anzahl an Vorfällen

- Synchronisation der Vorfälle (Mehrfacherkennung durch verschiedene Kameras)





# Fazit & Empfehlung



# Fazit und Empfehlung

## Systemarchitektur für Echtzeit-Skalierung

### **Verteiltes System zielführend:**

- Edge-Computing-Geräte für YOLO/OpenCV direkt an Kameras
- Parallele Verarbeitung mehrerer Videostreams
- Nur Erkennungen werden an zentrales System weitergeleitet

### **Hardware-Empfehlungen:**

- NVIDIA Jetson-Module für Edge-Processing
- 1 Edge-Gerät kann ~4-8 Kameras in Echtzeit verarbeiten (YOLO)
- Kosteneffizienter als zentrale Serverarchitektur



# Fazit und Empfehlung

## Empfehlungen für die nächsten Schritte

### **Zweistufiger Erkennungsprozess:**

- YOLO als Haupterkennung
- Manuelle Validierung oder sekundäre Überprüfung bei Verdacht

### **YOLO-Optimierung:**

- Weitere Feinabstimmung der Konfidenzschwellen
- Training mit bahnhofsspezifischen Daten (bestenfalls Supervised Learning)

### **Kameraübergreifende Ereigniskorrelation:**

- Implementierung eines Systems zur räumlichen und zeitlichen Zuordnung
- Vermeidung von Mehrfachzählung desselben Vorfalls durch verschiedene Kameras
- Abgleich von Erkennungszeitpunkten und -positionen

### **Kontinuierliche Verbesserung:**

- Feedback-Schleife für falsche und korrekte Erkennungen
- Regelmäßiges Nachtraining mit neuen Daten

# Abschlussbericht – videobasierte Vandalismuserkennung an Bahnhöfen

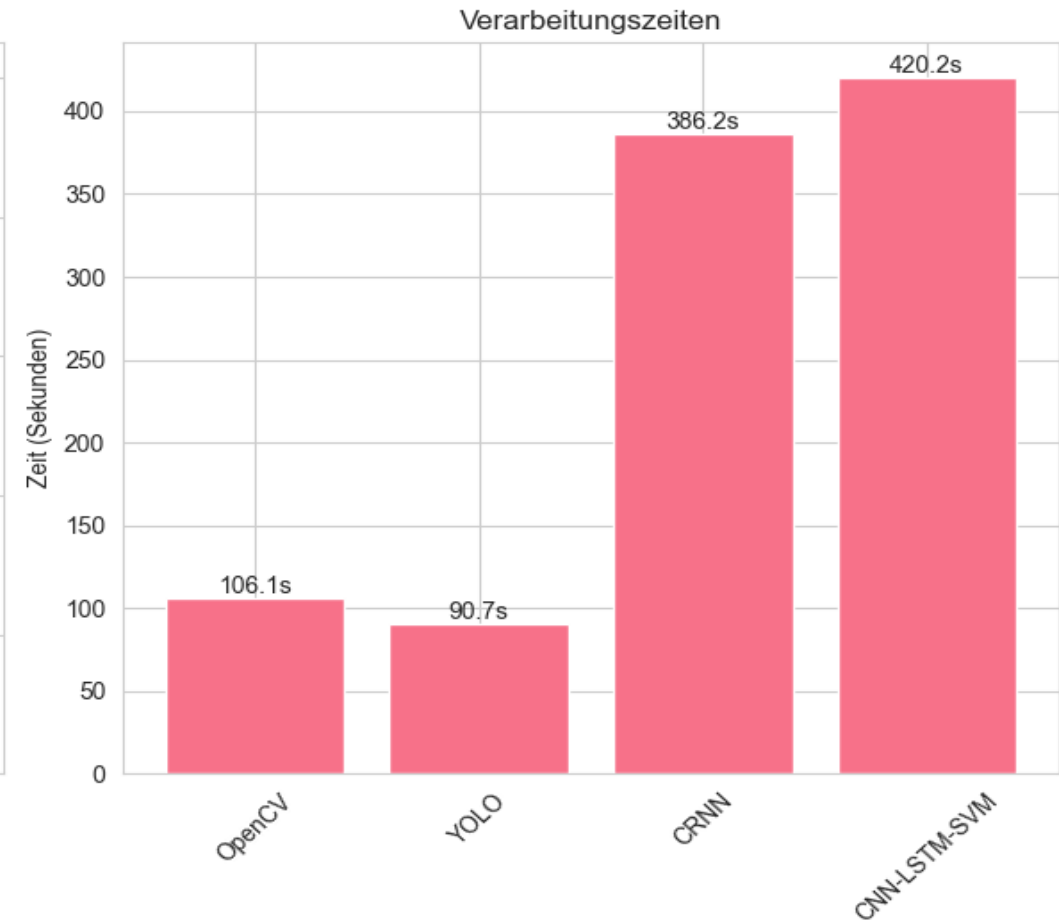
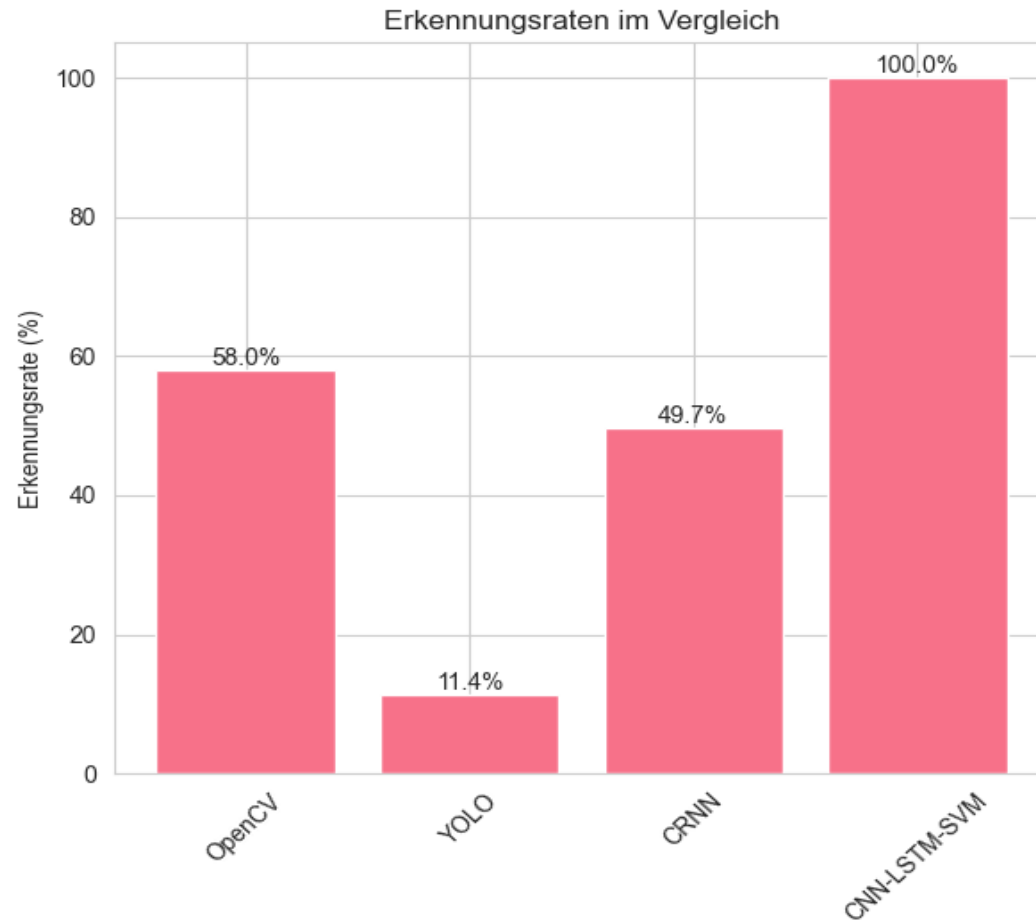
Forschungsprojekt TAHAI - TrustAdHocAI

Sandro Hartenstein | 28.2.2025 | Berlin



# Backup - Ergebnisse

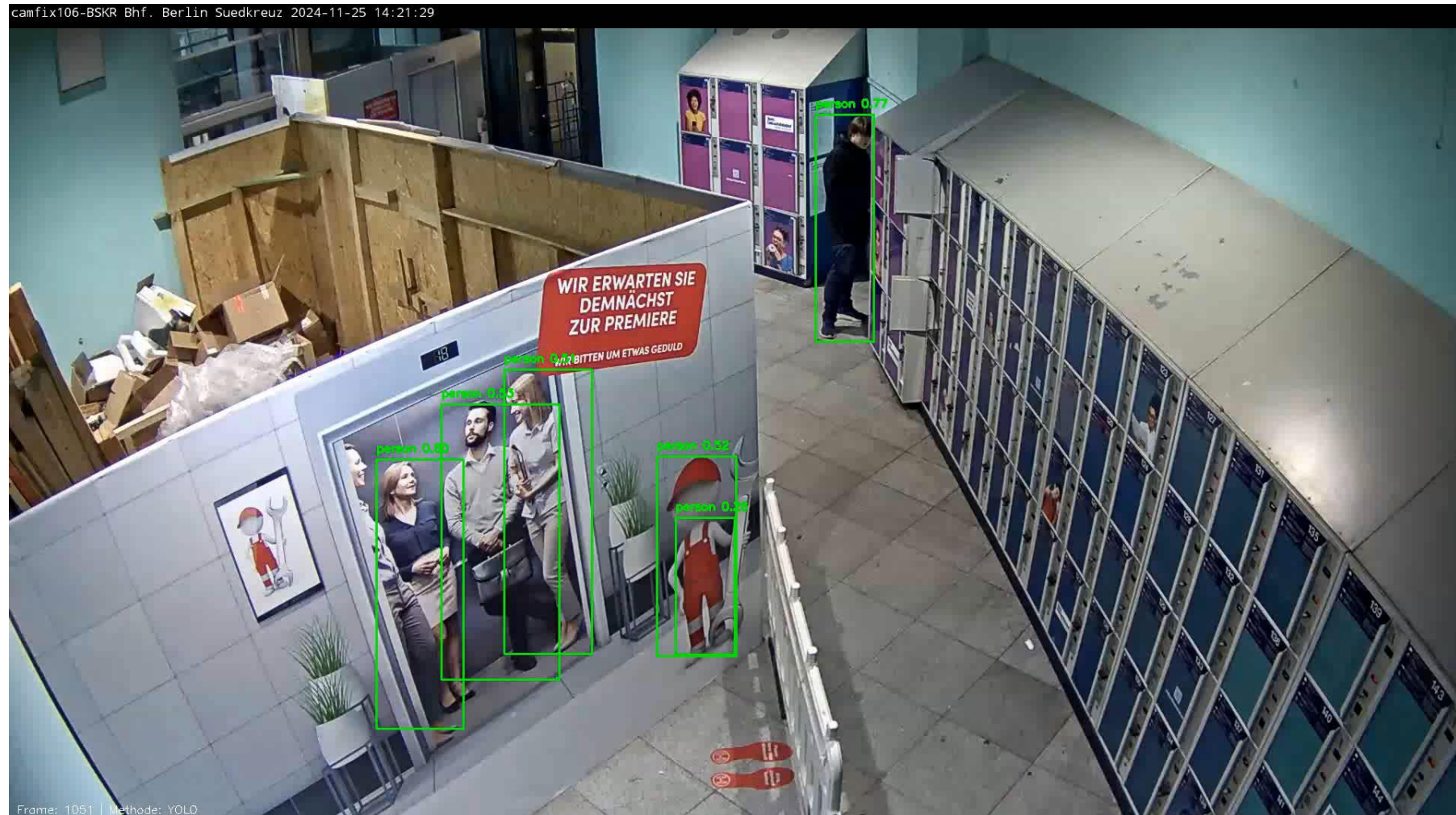
## Erkennungsraten und Verarbeitungszeiten





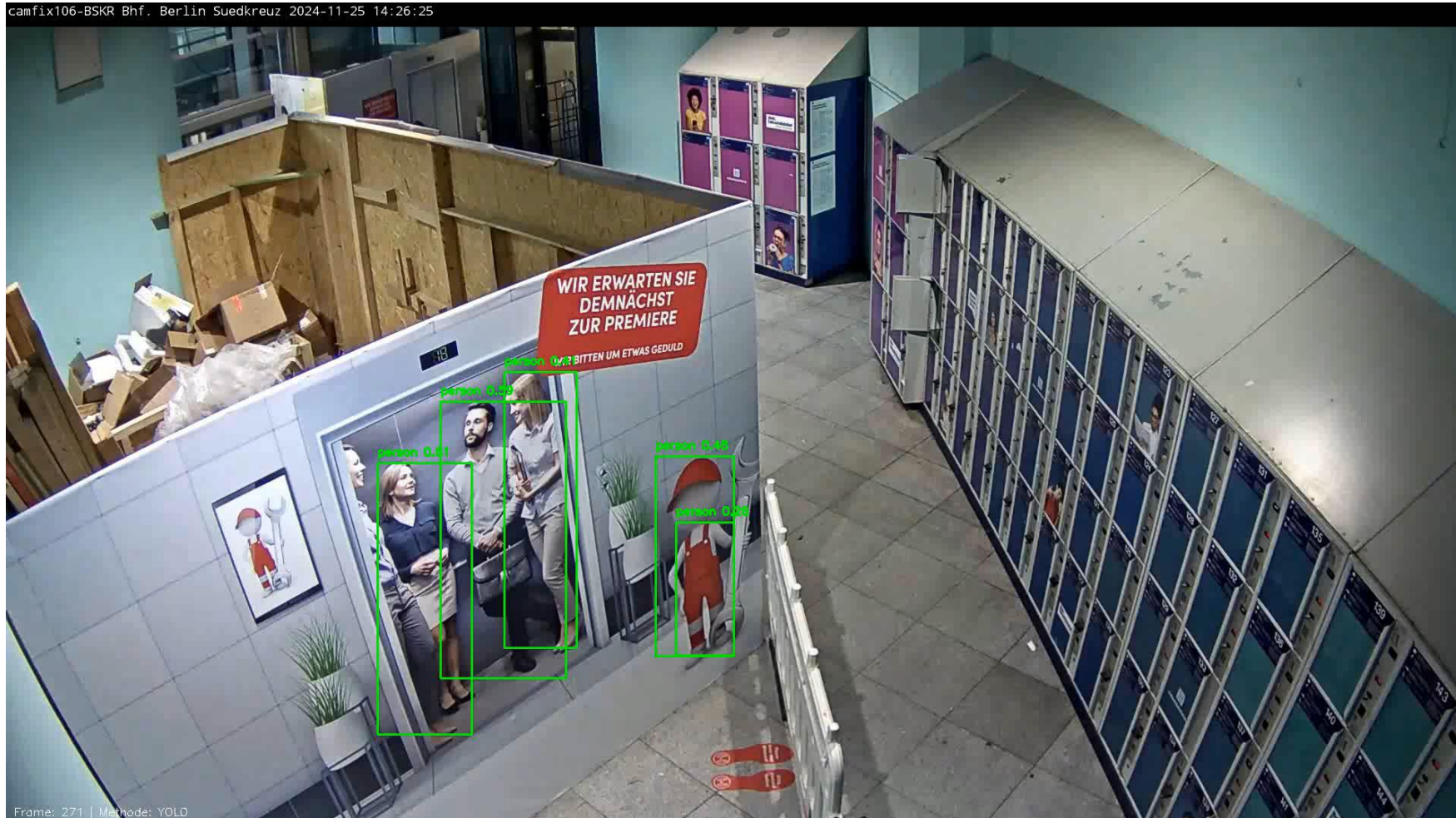
# Backup Videos

## Yolo Cam 106 Clip2



# Backup Videos

## Yolo Cam 106 Clip1 False Positive





# Backup Videos

## OpenCV Cam 106 Clip1 False Positive

