Khazar Journal of Science and Technology
© Khazar University Press 2022

Volume 7 №2 2023, 67-79
DOI: 10.5782/2520-6133.2023.7.2.67

Face Recognition in Smart Cameras by Yolo8

Farid Pashayev, Leyla Babayeva, Zuleykha Isgandarova, BehnamKiani Kalejahi

Department of Computer Science, School of Science and Engineering, Khazar University, Baku, Azerbaijan

Abstract

Smart AI Cameras have become a vital tool for enhancing security in several industries, such as industrial, transportation, and retail. This study investigates the methods that might be used to recognize moving objects in both daytime and nighttime settings. In this paper, convolutional neural networks, and recurrent neural networks—two deep learning techniques for object recognition—are investigated. We look at datasets containing a range of objects, lighting conFigureurations, and camera angles to determine how well these algorithms perform. In our research, we compared results from two separate datasets using YOLOv8. After all, we compared our methods and results with other scientists' research. We discussed the importance of camera placement, lighting issues, and algorithm choice for effective object detection. We evaluate the cameras' ability to recognize and follow moving things, as well as how well they can communicate with other security systems like alarms and access control. Our research demonstrates that smart AI cameras may significantly improve security in a variety of situations and that choosing the right algorithm and placing the camera is crucial for maximizing their effectiveness. For enterprises considering the usage of smart AI cameras for security, our research offers helpful information.

Keywords: *Smart cameras, YOLO, YOLOv8, object recognition, person recognition,* AI, IoT

Introduction

The explosion of artificial intelligence (AI) has had an impact on every part of community life, and deep learning (DL) and AI have made significant strides. Artificial intelligence (AI) encompasses a wide range of disciplines, including speech recognition, object identification, general algorithms, and brain modeling. As the need for safety and security has increased, more research has been done on intelligent surveillance. Intelligent surveillance can be used for a variety of purposes, including moving object recognition, object tracking, motion segmentation, object classification and identification, event detection, and behavior interpretation and description. Current study focuses on the processing and interpretation of camera-captured video sequences. Any moving items in a scene must first be detected by several apps in this field. To find out what is happening around monitored items, for instance, video event detection must first identify and track the tracked objects. In order to implement intelligent surveillance, monitoring and object detection are crucial components. Video object segmentation, which separates the moving element from each frame of a video clip, can be used to identify moving things. For segmenting video objects, a variety of techniques have been suggested. The two subcategories of these methods are foreground extraction-based and background construction-based video object segmentation. In background construction-based video object segmentation, the background information is first obtained, and then a video object is created by deleting the backdrop from the current frame in following frames. Foreground extraction-based video object segmentation first obtains an initial video object using temporal, spatial, or temporal-spatial information. The subsequent frames' motion, change, or other feature information is used to pinpoint the video object. Fast-moving objects can be tracked using background construction-based video object segmentation. It also costs little to implement and has simple computing requirements. For background construction-based video object segmentation, background modeling is required. In many applications (like video surveillance), it is primarily utilized in background subtraction foreground identification to model the background and identify the moving objects in an image. To show a background, the simplest method is to get a still photograph with no moving parts. The camera must be fixed, the illumination must be steady, and the background must not move in order to achieve decent results while utilizing background subtraction. The background can alter in specific circumstances, such as when the lighting changes. As a result, the backdrop representation model needs to be reliable and flexible. Background modeling has been used to detect moving objects in difficult circumstances.

Techniques for steadily updating models in the presence of dynamic backgrounds and variable lighting. Background subtraction approaches, both traditional and contemporary, have been studied. Our project intends to build a device that can recognize people in a picture or video. To address this issue, we went with the YOLOv8 approach.

Segmentation, object detection, and image categorization are among the applications for the newest and most complex YOLO model, YOLOv8. YOLOv8 was introduced by Ultralytics, the company that also created the well-known YOLOv5 concept. YOLOv8 offers a number of architectural changes and improvements over YOLOv5. In computer vision, the YOLO (You Only Look Once) model set has gained popularity. YOLO has grown highly popular due to its high level of realism and compact model size. A wide spectrum of developers can use YOLO models because they can be trained on a single GPU. On edge hardware or in the cloud, machine learning professionals can deploy it affordably. The YOLO concept has received significant backing from the computer vision community ever since Joseph Redmond initially proposed it in 2015. Versions 1-4 of YOLO were first kept up to date in C code inside of Redmond's own Darknet deep learning platform (Kiani et al., 2022). YOLOv8 is mostly used for the following reasons:

- The results of the Roboflow 100 and COCO tests indicate that YOLOv8 has a respectable accuracy rate.

- YOLOv8 comes with a variety of features that are beneficial to developers, including an approachable CLI and a well-thought-out Python package.

- Because there is a sizable community surrounding YOLO and one that is growing around the YOLOv8 model, you can get help from numerous computer vision experts when you need it.

YOLOv8 obtains great accuracy on COCO. For instance, the medium YOLOv8m model yields a 50.2% mAP when assessed on COCO. Additionally, YOLOv8 outperformed YOLOv5 much better than Roboflow 100, a dataset that accurately measures model performance across a variety of task-specific domains.

YOLOv8's developer-friendly features are also crucial. For instance, YOLOv8 offers a CLI that simplifies model training compared to other models that require many executable Python files. Additionally, compared to earlier models, developing with a Python package is more comfortable. Two datasets were employed, one of which includes 13660 photos of people and another has 11324 pictures of both humans and other things (Yazdi & Bouwmans, 2018). However, both of these datasets were trained on photos of people. We obtained 75% accuracy after training the first dataset and 96% accuracy after training the second dataset. We will look at the issue at hand as well as the solutions put forth by both us and other researchers that have looked into this issue.

1. Related Works

An earlier article by Mehran et al. (2022) presents a summary of the most recent methods for moving object recognition in video sequences. The authors divide these methods into three groups: modeling-based background subtraction, low-rank and sparse matrix decomposition, and object tracking. In addition to performance measurements and multiple benchmark datasets available to evaluate the performance of different moving object identification systems, they look at difficulties and important concerns in this sector (Riechmann, et al., 2022). Additionally, we cited a work by Sivachandiran et al. (2021), who developed a novel DL-driven automated person detection and tracking (DLD-APDT) model for surveillance videos to identify and track persons in the movies. Their model's input video is initially turned into a collection of frames using a frame conversion technique. Furthermore, they employ EfficientDet, a model for object detection based on BiFPN. Again, their training is based on the PascalVOC and PenFudan datasets. Finally, to choose the hyperparameters as accurately as feasible, they employ the RMSProp optimizer (Imerit.net. 2021). In their study, three cutting-edge OD systems-YOLOv3 (SPP), YOLOv4 and PANet (SPP), and CSPResNeXt50—were examined, used, and assessed., Lukas Malburg et al. (2021) propose a general framework for multi-modal monitoring of industrial processes as a first step toward video-based monitoring and tracking. Their experimental assessment shows that, in an industrial research scenario, detection accuracies for customized objects and failure locations of these objects may be increased to beyond 90%. The authors intend to use the Internet of Things to deploy OD systems along with other sensors to track the production process (Malburg et al., 2021) Wu-Chih Hu et al. (2023) show yet another practical technique for recognizing and tracking several moving objects in a video sequence captured by a moving camera without the use of additional sensors. Their method is incredibly effective; it can be applied in real-time applications, for monitoring overlapped objects and objects with different scales, as well as for identifying objects that are moving swiftly or that are being filmed by a camera that is moving quickly (Wu-Chih Hu et al. 2023) Amin Al-Habaibeh et al. (2023) consider the use of low and high-resolution infrared sensors as a component of the Internet of Things (loT) to estimate crowds in cities in order to improve and maximize the efficacy of the transportation system and other public services for low-density situations. Various image processing methods are used to compare the findings of an experiment after numerous subject counts are photographed. According to the findings, both approaches can be used to assess crowd density. A high-resolution camera is more expensive and precise to place in infrastructures, as has been made public. This comparison uses the FLIR E25, a high-resolution alternative, and the IRI 1002, a reasonably priced infrared camera with 16x16 pixels. Radiation from a healthy body temperature is infrared. The average body temperature is influenced by a variety of things, including the type of food consumed, the clothing worn, the amount of physical activity, sleeping habits, and the time of day. Between the wavelengths of 7.5 and 13, infrared energy is detectable by infrared cameras (Al-Habaibeh, et al., 2023). To prevent camera trap systems from taking pictures of nothing, the authors of the article "Motion vectors and deep neural networks for video camera traps" by Miklas Riechmann et al. (2022) provide a solution. Utilizing less disk storage space, less bandwidth, and less time for examination later in the research process are all benefits of this method. Additionally, their approach includes a pipeline for video frame filtering based on images and motion. The use of machine learning (ML) on a live video feed powered by an RPi is suggested in light of the information above (Riechmann, et al., 2022). Several different objects in TIR pictures and videos were successfully recognized by the YOLO model-based UAV object recognition framework created for this study. The YOLO models (YOLOv3, YOLOv4, and YOLOv5) with 15 different architectures based on pre-trained checkpoints using multi-situation ground TIR video frames were used in this study to explore the vehicle and human multi-object recognition. The assignment with the highest overall mAP had a 14 MB validating job model, whereas the assignment with the highest total mAP had an 88.69%. The quickest detecting speed was 50 FPS. They used the YOLO models produced using the ground TIR dataset in the UAV object-based detection system they suggested. They also talked about using the right evaluation techniques to obtain appropriate detection results. It is predicted that the identification results and cross-application demonstrated in this work will be useful for item detection in diverse TIR datasets (DAVIDNYARKO123, 2023) The remaining comparisons can be seen here: Table1.

2. Data

We used two datasets in our code. The first was downloaded from the internet, and the second one was created using various datasets. Using YOLOv8n.pt, these two datasets were trained. The first dataset (human dataset, n.d.) has 13660 images at a resolution of 416 × 416 pixels. The only images in this dataset are private ones. People are labeled in YOLOv8 format as pedestrians. When we run it, the accuracy is 75%. We created a second dataset using various images we acquired from the internet. Then, we used robot flow to classify them into groups. To create the group, 11324 photos were employed. The collection contains 80 classifications in addition to personal photos. It has undergone multiple training with a batch size of 16 and an epoch of 30. Although this dataset can detect a large number of objects, our accuracy with simply human image training is 96%.

3. Experiments

Each image underwent the pre-processing steps of auto-alignment of pixel data using EXIF orientation stripping and scaling to 416 × 416 (Stretch). We have chosen a batch size of 16 and a training epoch of 30 for these datasets. These datasets were trained over the course of about 30 hours.

On a special dataset, a YOLOv8 model is trained using the training mode. In this mode, the model is trained using a set of hyperparameters and a few carefully chosen datasets. The model's training parameters are adjusted in order for it to correctly identify the classes and locations of objects in an image.

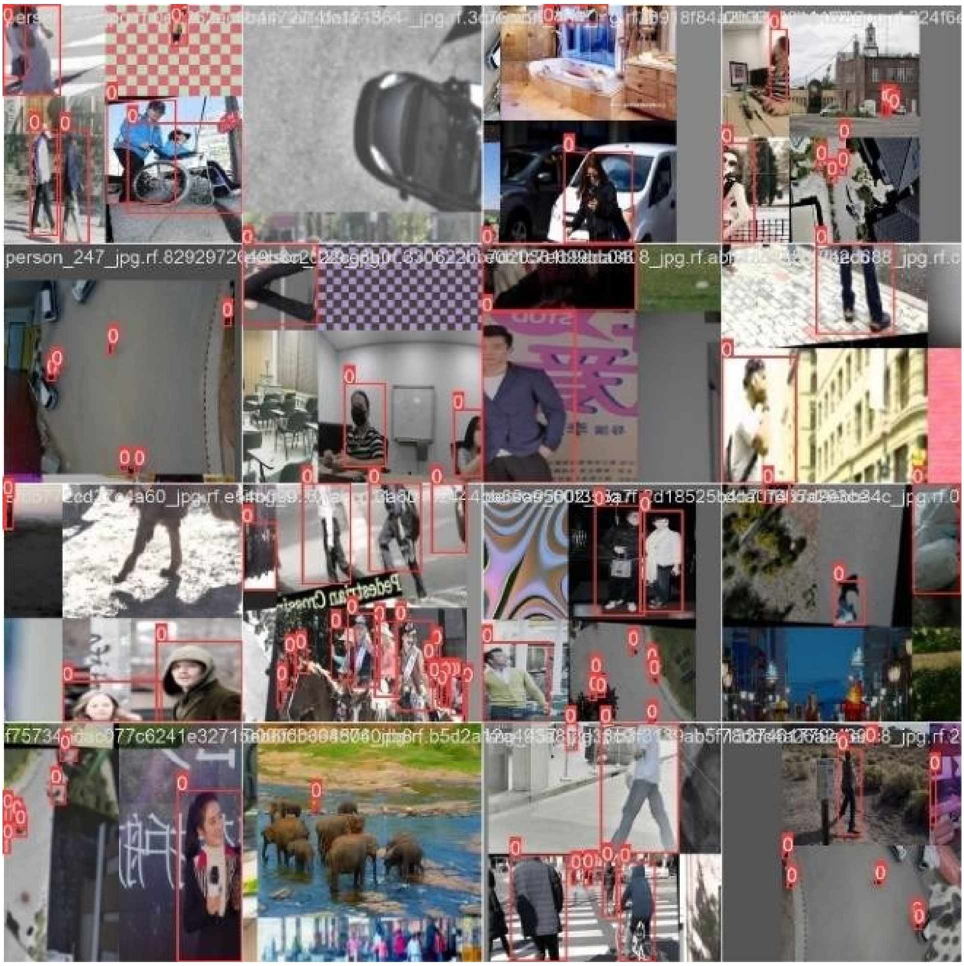

Figure 1. Train batch of first dataset

For YOLO models, the various conFigureurations and hyperparameters required to train the model on a dataset are referred to as "training settings". The performance, speed, and accuracy of the model can be affected by any of these options. The standard YOLO training parameters include the learning rate, momentum, weight decay, and batch size. Other factors that could affect the training procedure include the choice of an optimizer, a loss function, and the quantity and make-up of the training dataset. For optimum performance for a certain application, these parameters need to be carefully adjusted and experimented with. A way to determine how effectively your algorithm models your dataset is by using the loss function. It is a mathematical function of the machine learning algorithm's parameters. The loss is zero if the model's forecast is accurate; otherwise, the loss is higher. b0x_loss is a loss that evaluates how "tight" the projected bounding boxes are to the item in the real world. cls_loss is a loss that measures how accurately each predicted bounding box was classified; each box may contain either an object class or "background." Cross entropy loss is the common name for this loss. Summarized data and results of each dataset used can be seen in the following Table2:

Table2. Comparison table of two used datasets

Datasets	Number of data	Training model	Accuracy	box_loss	cls_loss
Dataset 1	13660	YOLOv8	75%	1,57	1,28
Dataset 2	11324	YOLOv8	96%	1,17	1,18

Data labeling in machine learning is the classification of unlabeled data (such as pictures, text files, videos, etc.) and the addition of one or more educational labels to give the data context in order to learn from it using a machine learning model. For instance, labels might say if a picture shows a bird or an automobile, whether someone spoke in an audio recording, or whether a tumor is visible on an X-ray. The use cases of speech recognition, computer vision, and natural language processing all call for the labeling of data.

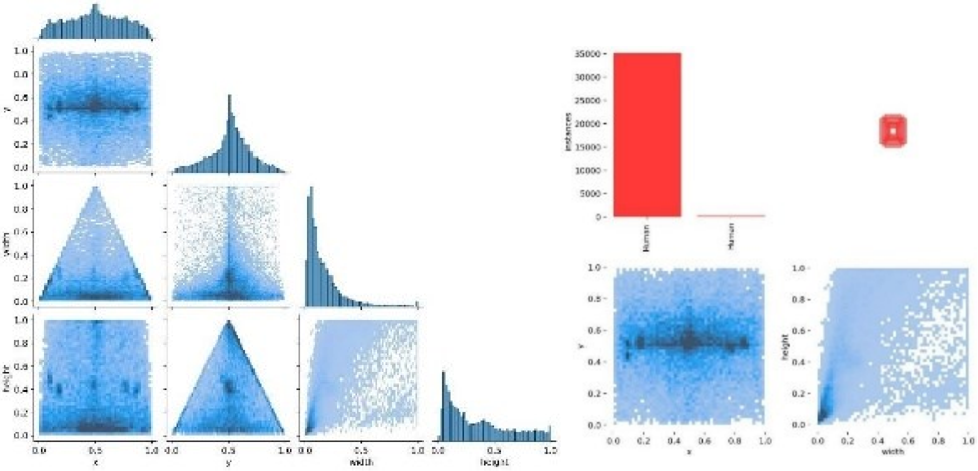

Figure 2. Labels and Label’s Correlogram for first dataset

A set of two-dimensional histograms called a correlogram is used to compare each axis of our data to every other axis. You can view labels and labels correlograms for both datasets in Figure2 and Figure4. Train batch for second dataset:

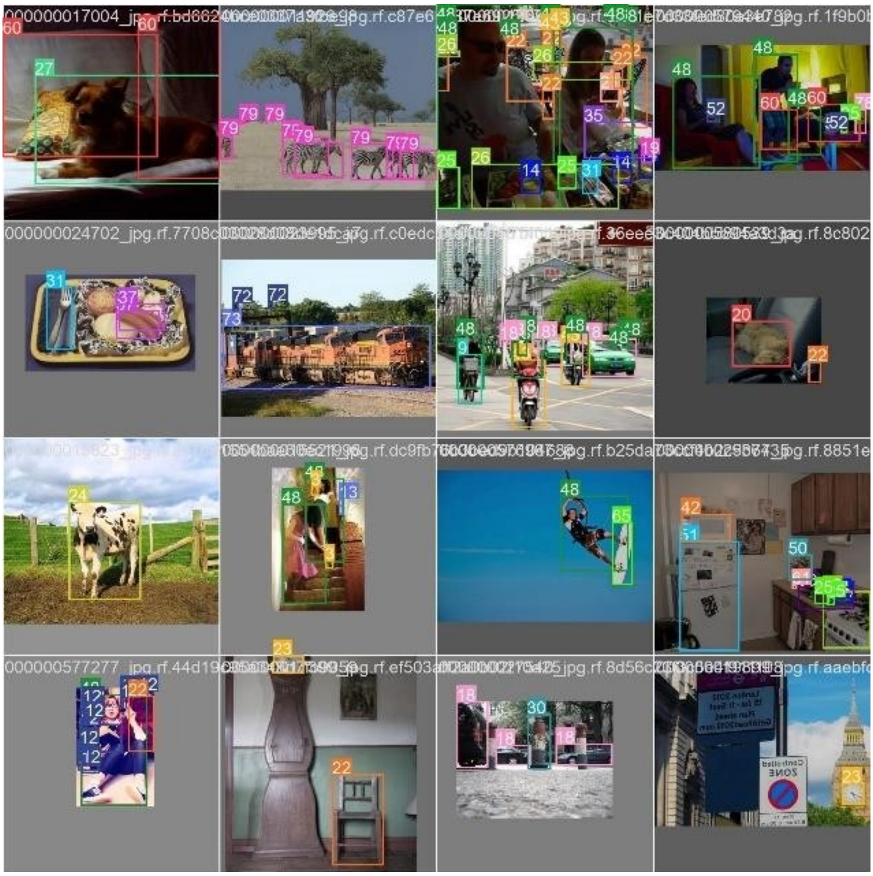

Figure 3. Train batch of second dataset

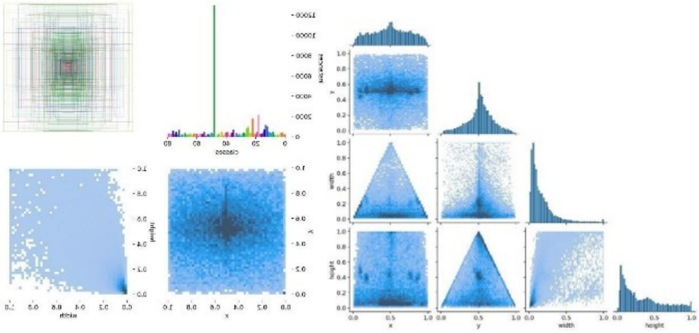

Figure 4. Label's Correlogram for second dataset

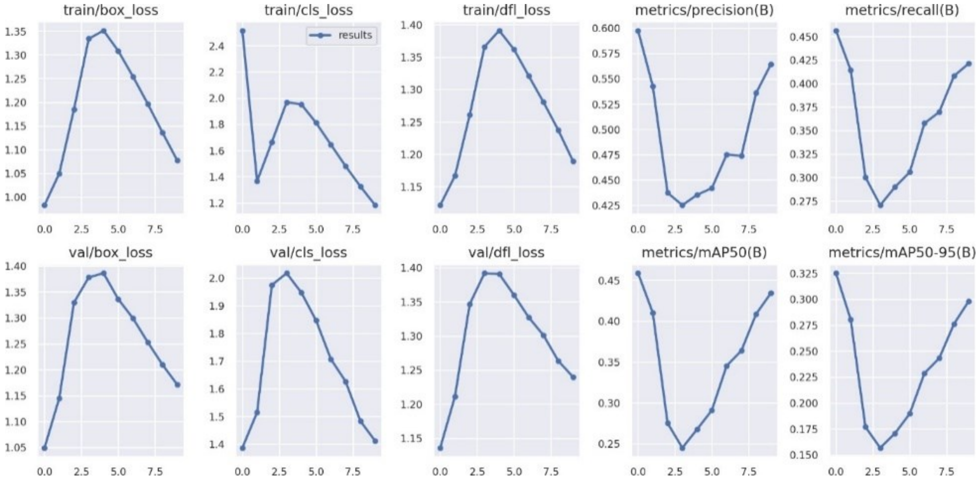

Figure 5. Metrics and results of second dataset

Conclusion

In our research, we offered a method for surveillance systems' image and video cameras to use when looking for humans. The YOLOv8 method was used to train two datasets. The batch size and epoch for training the first dataset were 16 and 30, respectively. After 30 hours of training, we were 75 percent accurate. The second data set had a batch size of 16 and an epoch of 30. Using solely training images of individuals, this dataset has a 96% accuracy rate for detecting a variety of objects. Despite significant progress, there are still many problems that need to be tackled. It might be challenging to get the greatest performance due of things like occlusion, various lighting conditions, and privacy concerns. Future research should focus on developing advanced methods and algorithms to handle these problems and improve the general reliability and accuracy of human recognition systems. Furthermore, it is crucial to consider how morally questionable person recognition technology might be. The potential for abuse and privacy concerns needs to be carefully taken into account. Researchers and practitioners should aim to create transparent frameworks that prioritize privacy protection and guarantee the proper usage of smart camera systems. Person identification for smart cameras is becoming a reality as a potent technology with numerous applications in security, surveillance, and human-computer interaction. Future study in this area can be built upon the accomplishments discussed in this publication. With further effort on refining algorithms, resolving problems, and adding ethical considerations, person identification technology will progress and become widely used, ultimately making our societies safer and more efficient. In the future, we intend to go into more detail about this and recommend it to malls for use in identifying persons and tracking, as well as in observing people's entrances and exits for increased security.

References

Al-Habaibeh, A., Yaseen, S., & Nweke, B. (2023). A comparative study of low- and high-resolution infrared cameras for IoT smart city applications. Ain Shams Engineering Journal. 14(6): 102108.

Data labeling: why it is important to manage it efficiently. (2022). Labellerr. https://www.labellerr.com/blog/data-labeling-why-it-is-important-to-manage-it-efficiently/

Davidnyarko123. (2023). GitHub-DAVIDNYARKO123/yolov8-silva. GitHub. https://github.com/DAVIDNYARKO123/yolov8-silva

Imerit.net. (2021). This Is How You Know It's Time to Bring in Professional Data Labelers. iMerit. https://imerit.net/blog/when-to-outsource-your-data-labeling-all-sol/

Hu, W. C., Chen, C. H., Chen, T. Y., Huang, D. Y., & Wu, Z. C. (2015). Moving object detection and tracking from video captured by moving camera. Journal of Visual Communication and Image Representation. 30: 164–180.

Jiang, C., Ren, H., Ye, X., Zhu, J., Zeng, H., Nan, Y., Sun, M., Ren, X., & Huo, H. (2022). Object detection from UAV thermal infrared images and videos using YOLO models. International Journal of Applied Earth Observation and Geoinformation. 112: 102912.

Kiani, Kalejahi, Behnam et al. (2022). 'Diagnosis of Liver Disease by Computer- Assisted Imaging Techniques: A Literature Review'. 1 Jan: 1097 – 1114.

Kiani Kalejahi, B., Meshgini, S. & Danishvar, S. (2023). Brain tumor segmentation by auxiliary classifier generative adversarial network. SIViP 17, 3339–3345. https://doi.org/10.1007/s11760-023-02555-6.

Malburg, L., Rieder, M. P., Seiger, R., Klein, P., & Bergmann, R. (2021). Object Detection for Smart Factory Processes by Machine Learning. Procedia Computer Science. 184: 581–588.

Mutepfe F, Kalejahi BK, Meshgini S, Danishvar S. (2021). Generative Adversarial Network Image Synthesis Method for Skin Lesion Generation and Classification. J Med Signals Sens. 11(4):237-252.

Paul, R., & Mukhopadhyay, S. (2015). Foreground Object Extraction Based on Independent Component Analysis. Advances in Vision Computing: An International Journal. 2(4): 1–11.

Riechmann, M., Gardiner, R., Waddington, K., Rueger, R., Leymarie, F. F., & Rueger, S. (2022). Motion vectors and deep neural networks for video camera traps. Ecological Informatics. 69: 101657.

Sivachandiran, S., Jagan Mohan, K., & Mohammed Nazer, G. (2022). Deep Learning driven automated person detection and tracking model on surveillance videos. Measurement: Sensors. 24: 100422.

Yazdi, M., & Bouwmans, T. (2018). New trends on moving object detection in video images captured by a moving camera: A survey. Computer Science Review 2018; 28: 157–177.

Table1. Comparison table of related works

	Models	**Datasets**	**Experiments**	**Results**
Our work	YOLOv8	Human dataset and image dataset	We trained our model, and then tested it on our datasets. It first resizes photos, removes background, detects all objects on them and then detects which object is person	75% and 96%
Mehran Yazdi et al.	They summarized different techniques in their work	Benchmark datasets	They divided multiple techniques into object tracking, low-rank and spare matrix decomposition, and modeling-based background subtraction	Summarized and compared multiple methods
S. Sivachandiran et al.	Used DLD-APDT	Surveillance videos	Their model first carries out a frame conversion operation, turning the input video into a collection of frames. Additionally, they use the BiFPN-based EfficientDet model as an object detector. Again, PascalVOC and PenFudan datasets are being used in their training. Finally, they use the RMSProp optimizer to select the hyperparameters as best as possible	Person detection using DL-driven automated person detection and tracking model
Lukas Malburg et al.	Used YOLOv3, YOLOv4 and PANet	Custom dataset	As a first step towards video-based monitoring and tracking, Lukas Malburg et al. propose a generic framework for multi-	Their experimental evaluation demonstrates that detection accuracies for custom objects and

			modal monitoring of manufacturing processes in their study and deploy and assess three cutting-edge OD systems (YOLOv3 (SPP), YOLOv4 and PANet (SPP) CSPResNeXt50)	failure places of these objects can be improved to above 90% in an industrial research setting
Wu-Chih Hu et al.	Method for detecting and tracking moving objects without using extra sensors	Videos of moving objects taken by moving cameras	Tested their model on moving objects from a video sequence taken by a moving camera without using extra sensors	It is effective for tracking items that overlap and objects whose scale changes, as well as for detecting fast-moving objects or things under fast-moving cameras
Amin Al-Habaibeh et al.	Used multiple image processing techniques	FLIR E25, a high-resolution substitute, and IRI 1002, a low-cost infrared camera with 16x16 pixels, are used in this comparison	In an experiment, multiple subject counts are photographed, and the results are then compared using different image processing techniques	The results suggest that both technologies help determine crowd density
Miklas Riechmann et al.	YOLO model-based UAV	TIR dataset	Their method introduces an image- and movement-based video frame filtering pipeline	The identification results and cross-application shown in this work are anticipated to be helpful for item detection in various TIR datasets