

Preventing Crime and Terrorist Activities with a New Anomaly Detection Approach Based on Outfit

Gizem ORTAC KOSUN^{1*} , Seckin YILMAZ² , Yusuf KAYIPMAZ² , Rüya SAMLI¹ 

^{1*} Computer Engineering Department, Istanbul University - Cerrahpasa, 34320, Bursa, TURKEY

² Computer Engineering Department, Bursa Technical University, 16310, Bursa, TURKEY

Abstract

Video surveillance systems play an important role in ensuring security indoors and outdoors and detecting suspicious persons due to the increasing violence and terrorist acts every year. In the proposed study, an artificial intelligence-based warning system has been developed, which enables the detection of potential suspects who may carry out criminal or terrorist activities by detecting anomalies in surveillance videos. In this developed system, an abnormality is detected by using the outfits of the people. The YOLOv7 object detection model is trained on our customized data sets, and suspicious person detection is made through outfit information. Especially in cases where biometric data is hidden, dress information makes it easier to obtain information about people. For this reason, the knowledge of outfits is the main point of this study in the detection of suspicious persons. Thanks to this study, security guards will be able to focus on this suspicious person before they pre-empt any crime or terrorist activity. If there are other data confirming the suspicious situation as a result of this follow-up; security personnel will have time to eliminate the crime or attack. The experimental results obtained have been promising in terms of the usability of a person's outfit anomalies to ensure public confidence or avoid risk to human life. Although there are various studies in the literature for the prevention of terrorist or criminal activities; there is no study in which people's outfit is used to identify suspects.

Keywords: Forensic science, Anomaly detection, Soft biometrics, Surveillance video.

Cite this paper as:

Ortac Kosun, G., Yilmaz S., Kayıpmaz Y. and Samli, R. (2023). *Preventing Crime and Terrorist Activities with a New Anomaly Detection Approach Based on Outfit*. Journal of Innovative Science and Engineering, 7(2):167-182

*Corresponding author: Gizem Ortac Kosun

E-mail: gizemortac1@hotmail.com

Received Date:22/08/2023

Accepted Date:08/10/2023

© Copyright 2023 by

Bursa Technical University. Available online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

1. Introduction

With the increase in terrorism, violence, and crime rates in recent years, the importance of security and safety issues is increasing day by day. Government and private organizations are concerned about security in public and crowded areas such as airports and shopping malls. For this reason, video surveillance systems have been used in both private and public places. In forensic investigations, video and images from these systems are widely used in crime evidence investigations, which can provide important elements of forensic evidence, bring together existing elements of evidence, or make connections between evidence in a particular case [1]. However, most existing surveillance systems rely on the human factor. The efficiency of these systems, which rely on human control to detect any abnormality, decreases over time. This problem can be solved by the automation of video surveillance. These artificial intelligence-based studies have gained great importance in facilitating control operations with video surveillance systems and reducing the human factor error rate. The function of the automated system is to give an indication, in the form of an alarm or in any other form, when pre-defined abnormal activity occurs [2]. The intended use of the recorded images is to identify potential suspects or take appropriate action if they have no knowledge of where and when the incident occurred or even whether it occurred [3]. However, many of the systems, especially those originating from crime or terrorism, are designed to facilitate the investigation process after the event has occurred. Although the number of studies in this area is limited; focused on detecting offensive tools such as guns and knives, and developing preventive and warning systems. However, these tools, which constitute an element of attack, are usually hidden by people in outfits such as coats and revealed when the person approaches the crime environment. Studies show that the use of dress codes is very effective for detecting suspects, criminals, or missing persons [4,5]. Therefore, different approaches are needed to identify potential suspects.

In this study, an artificial intelligence-based warning system is proposed to help security guards identify potential suspects. It is focused on outfits to prevent crime or terrorist activities in video surveillance systems. In this developed system, if there is an anomaly in outfits worn by the people, early warning is given about the suspicious situation. Thus, security personnel will focus on the relevant person and follow the person closely with advance warning. As a result of this close follow-up monitoring, if there is any other data confirming the suspicious individual situation, the security personnel will have time to eliminate the crime or terrorist attack. Before the attack on the United States Capitol Hill building on January 6, 2021, video footage was obtained showing the route the bomber walked on January 5, 2021, wearing the bomber's face mask, glasses, hoodie, and gloves. Owing to these images, it was determined that the bomber placed the devices in the alley behind the Republican National Committee Headquarters between 19.30 - 20.30 in the evening [6]. This study is the first proposed study in this direction in the literature.

Based on the real events experienced, the study was advanced through various scenarios. For example, it is abnormal for a person to wear a coat on a summer day. If this person approaches the entrance of a public or private establishment, there may be a potential risk of attack. In such a case, this proposed system will provide a warning via the security camera while these subject is moving toward the entrance. Thus, alerting the security guard and being able to take precautions before the person approaches the building. To give another example, if a person's face is hidden among the crowd in public places such as shopping malls, a warning will be given that there is a potential risk of posing a threat to that person. In this way, the security guard will be able to follow this person closely. As a result of the close monitoring

of the security personnel, if there is a possibility to prevent a crime before it is committed, it can be prevented. If a crime has taken place (such as theft), it will be possible to take action much more quickly to catch this person. Otherwise, there will be situations where security personnel cannot detect people who may raise such suspicion during long-term monitoring of the security camera due to eye strain.

This proposed system, makes significant contributions to forensic prevention, catching the criminal, or clarifying the cases. The rest of the work is organized as follows: Part 2 is related studies, Part 3 is the methodology, Part 4 is experimental results, and Part 5 is the conclusion and acknowledgment.

2. Related Works

Researchers have shown great interest in studies involving the detection of various anomalous objects and situations, such as people, anomalies, and masks, within the framework of forensic events in video systems. With these studies, it has become much easier to collect and analyze important evidence about cases in video systems. Detecting suspicious situations, people, or objects and developing warning systems has been an area of great interest in the past years. In the study of Narejo et al.[7], a computer-based fully automatic system was developed to detect various weapons, especially pistols, and rifles. Using the YOLOv3 algorithm on customized datasets, a model is proposed that provides a predictive machine or robot to identify weapons and can also alert the human manager when a gun or firearm is seen on the sidelines. The system has an accuracy value of 98.89%. By Grega et al. [8], a system was developed that detects knives and firearms in CCTV images and warns the security guard or operator. The specificity and sensitivity of the blade detection algorithm are 94.93% and 81.18%, respectively, when the edge histogram descriptor and the Decision Support Vector classifier are used. These results are significantly better than others published recently. A specificity of 96.69% and a sensitivity of 35.98% were obtained by using a three-layer neural network for the firearm detection algorithm. Mehta et al. [9] developed a real-time warning system that includes weapon and fire detection anomalies in areas monitored by cameras. Experimental results have shown that the proposed model is suitable for real-time monitoring and can be deployed in any GPU-based system. Marbach et al. [10] proposed a system for automatic fire detection based on the temporal variation of fire density. Dever et al. [11] designed an armed robbery detection algorithm according to the silhouette of the person and the position of the arms. The silhouette image was divided into separate parts and the position of the arms was determined and the determination was made. Anomaly detection in crowded areas is also one of the areas of increasing interest for public safety in video surveillance systems. However, although many studies have been carried out in this area in recent years such as Yin et al. [12], Ravanbakhsh et al. [13] and Mehran et al. [14] the subject is still an open field of study.

3. Material and Method

The flow chart of the study is presented in Figure 1. After the collected training images went through the preprocessing stage, the regions of the objects to be detected in the image were tagged and trained with the YOLOv7 network. In the test phase, images containing suspicious situations were applied to the trained YOLOv7 model, results were obtained and a warning was given in case of suspicious situation detection.

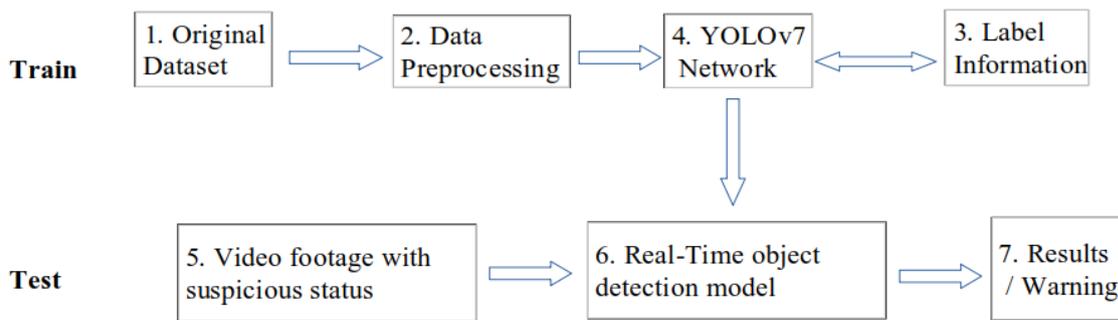


Figure 1. Flow chart of the proposed algorithm.

3.1. Datasets

In the proposed study, an algorithm is proposed to detect situations where people camouflage the face and trunk of their bodies in a way that prevents recognition and tracking. The human body is divided into two parts, the head and the trunk. It is aimed to identify people who hide some or all of their biometric data by wearing a surgical mask, prescription glasses, and sunglasses on the head area, and a coat, overcoat, or coat in a size that will cover their body regardless of the weather conditions, and to detect possible dangers in advance.

3.2. Data Collection

In the proposed study, the surgical mask, transparent glasses, and sunglasses are required for fixation in the head and trunk region; The images required for coats, coats, overcoats, and clothes other than those outfits were obtained from Google Images. There are 220 images in the first dataset to be used for coat detection, and 800 images in the second dataset that will be used in the detection of surgical masks, clear glasses, and sunglasses.

3.3. Data Preprocessing

After the necessary data has been obtained and collected, the next step is to annotate the data. This step has a direct impact on the efficiency and performance of the model. LabelImg [\cite{heartexlabs}](#), a graphic image annotation tool, was used to label the ground truth box of images. Label categories are divided into “mask”, “no mask”, “glass”, and “sunglass” for the head area, and “coat” and “other” for the body area, respectively. The dataset is annotated in the format $(\langle x_{min} \rangle \langle y_{min} \rangle \langle x_{max} \rangle \langle y_{max} \rangle)$ and converted to YOLO format $(\langle object-class \rangle \langle x_{center} \rangle \langle y_{center} \rangle \langle width \rangle \langle height \rangle)$.

For the head area of the people, 800 images are selected in categories: "mask", "no-mask", "glass" and "sunglass", considering whether people are masked at different angles, positions, and directions, whether they wear clear glasses and sunglasses. manually labeled. of the tagged photos are reserved for the training set and the validation set. During the labeling phase, the people in the images were labeled as wearing or not wearing a mask to cover their entire faces. If the person wears clear glasses, in addition to the mask information label, the glasses are labeled to cover the glasses. If the person wears sunglasses, a sunglasses label is given to cover the sunglasses in addition to the mask data. In the part of the study that included coats, and topcoats (t-shirts, sweaters, athletes, etc.), 220 images were labeled as "coat" and "other" according to whether the person wore them or not. 190 of these photographs were used as a training

set and as a validation set.

3.4. Method

In this proposed system, when such a situation occurs, a warning will be given over the system while the person is moving toward the entrance. In this way, the security guard will be able to take precautions without getting too close to the building. The person's face recognition process is also very important for detecting the potential criminal. In this subject, the human body was divided into two parts the head and the torso, and experiments were carried out on the detection of abnormal situations. Masks, glasses, and sunglasses that prevent face recognition in the head area were detected. In the body part, the outfit will prevent recognition and explosives, weapons, knives, etc. It is aimed to determine whether a coat, overcoat, or overcoat is worn to hide objects. This study makes significant contributions to the prevention of forensic incidents, catching criminals, and illuminating cases. For this system, where real-time detection and speed are important, the most up-to-date version of You Only Look Once (YOLO) [16] algorithm, which is the fastest and most accurate among object recognition algorithms, is used.

YOLO stands for "You Only Look Once," an open-source object detection algorithm that relies on convolutional neural networks.

It's renowned as one of the most well-known deep learning algorithms, primarily due to its remarkable speed, attributed to its single-stage detection design. Traditional detection systems utilized classifiers or localizers for identifying objects, unlike YOLO, which takes a distinct approach. YOLO treats object detection as a regression task, employing a sole neural network for the entire image. In the YOLO framework, the input image is initially divided into a grid of size $S \times S$. Each grid cell is responsible for determining the presence of an object within its region, pinpointing its center, checking if it spans its center, measuring its dimensions (length and height), and classifying it. These processes culminate in the creation of bounding boxes [17].

Over the years, various iterations of the YOLO algorithm have been introduced by researchers. The lineage extends from YOLOv1 to the widely recognized YOLOv3, originating from the efforts of Joseph Redmon, a graduate student, and Ali Farhadi, a consultant. Following Redmon's withdrawal from computer vision research due to ethical concerns, Alexey Bochkovskiy introduced YOLOv4, carrying the torch forward. YOLOv7, the latest official version of the YOLO architecture, has been developed by its original creators.

The YOLOv7 algorithm has garnered significant popularity within the computer vision and machine learning communities as a potent object detection method. It eclipses all preceding object detection models and YOLO versions in both swiftness and precision. Notably, it operates efficiently on more budget-friendly hardware than many other neural networks and can be swiftly trained on modest datasets without relying on pre-trained weights. Taking into account the imperative of real-time detection, the YOLOv7 model, the most recent advancement in the YOLO series, has been selected, as illustrated in Figure 2.

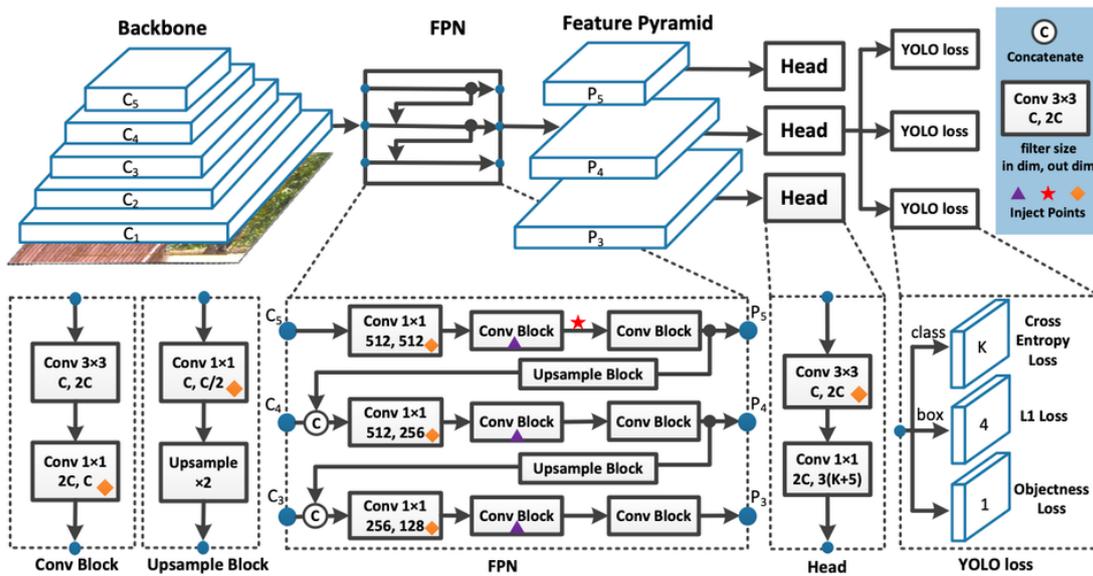


Figure 2. YOLOv7 architecture [18].

The YOLOv7 model has elevated both the pace and precision of object detection compared to its previous iterations. It introduces an enhanced integration approach, delivers more exact object detection performance, incorporates a more resilient loss function, and boasts an improved network architecture that features an upgraded tag assignment function. In a general sense, all YOLO architectures comprise three main components: the spine, head, and neck. The spine takes on responsibilities such as foundational work and essential feature extraction, channeling this information to the head through the intermediary neck component.

Unlike its predecessors, YOLOv7 adopts an extended and efficient layer aggregation network (E-ELAN) as the computational block for its backbone, refraining from deviating and continuing to employ DarkNet for the backbone [19]. To augment the network's capacity for continual learning enhancement without disrupting the original gradient pathway, YOLOv7 integrates a long-range attention network termed Extended-ELAN (abbreviated as E-ELAN). This innovation employs principles of expansion, mixture, and unification to bolster learning capabilities [18].

In terms of architecture, E-ELAN introduces alterations exclusively within the computational block, while the architecture of the transition layer remains largely unchanged. The strategic approach involves leveraging group convolution to broaden the channel and materiality of the computational blocks, as depicted in Figure 3 [20].

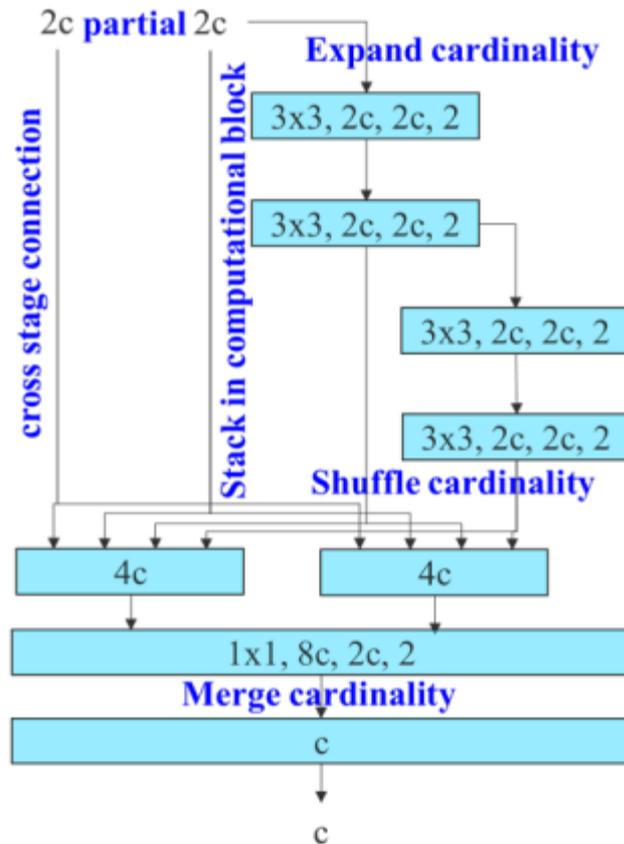


Figure 3. Extended ELAN (E-ELAN)

The YOLOv7 algorithm provides reparameterization planning (RP). RP is based on averaging several models to produce a performance-solid final model. Module-level reparameterization has been an active area of research where certain parts of the model have specific reparameterization strategies [20]. After training several models with the same parameters and different training sets, the weights are averaged to obtain the final model. The final model is formed by combining the outputs. The reparameterized convolution architecture in YOLOv7 uses RepConv [21] without the identity link (RepConvN). The goal is to prevent identity bindings when reparameterized convolution is used to replace a convolution layer with residue or concatenation.

Training Environment and Training Parameters

The experiments described in this paper were carried out using a 64-bit CPU operating at 2.20 GHz with twelve cores, along with 8 GB of memory. The computing setup also included an NVIDIA GeForce RTX 3050 Laptop GPU with 4 GB of video memory. The utilized version of the Compute Unified Device Architecture (CUDA) was 11.03. The deep learning framework of choice was PyTorch 1.11.0, and the code was compiled using Python 3.8.

Within the experimental model outlined in this article, the Adam optimization algorithm was deliberately employed to enhance the pace of training. The model processes images with a resolution of 640×640 pixels as its input. The initial learning rate assigned to the model is 0.01, coupled with a learning rate momentum of 0.937, and a weight decay value of 0.0005. These parameter choices consider both training speed and video memory capacity.

In terms of the training process, the batch size for each training set within the study was fixed at 4 samples. The model underwent 100 rounds of training iterations, each consisting of an equivalent number of iterations.

Evaluation Metrics

The evaluation of our model's performance in this article is carried out comprehensively and objectively using metrics such as the confusion matrix, precision, and sensitivity values.

TP represents the correct perception that the model predicts positively and the true value is also positive. FN detects detection errors that the model predicts negatively but have a positive true value. FP represents the detection errors that the model predicts positively but has a negative true value. TN represents accurate detection; the model estimate is negative and the true value is also negative.

Expressions of precision and sensitivity are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

As precision and recall alone may not suffice as exclusive performance indicators for the model, the F1 score is introduced as a compromise metric that combines both aspects, as defined in the following formula [22]:

$$F1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

The model's performance was assessed using a PR curve, which takes into account the precision and sensitivity ratio for each detected category. Additionally, the mean precision (mAP) is employed as a measure of the model's accuracy, depending on precision and recall. In this context, (AP) refers to the area under the precision-recall curve (PR curve), and (mAP) represents the average of (AP) values across various classes. Here, (N) represents the number of classes within the test sample [23].

$$AP = \int_0^1 P(R) dR \quad (4)$$

$$mAP = \frac{\sum_1^N \int_0^1 P(R) dR}{N} \quad (5)$$

Loss Function

The loss function of the YOLOv7 model is calculated by summing up three loss values: loss of localization (L_{box}), loss of confidence (L_{obj}), and loss of classification (L_{cls}). Among them, loss of confidence and loss of classification functions, binary cross-entropy loss, and localization loss use the CIoU loss function.

$$\text{Loss} = W1 \times L_{\text{box}} + W2 \times L_{\text{cls}} + W3 \times L_{\text{obj}} \quad (6)$$

In this context, the weight values for the three loss functions are denoted as W1, W2, and W3, respectively [22].

For the coordinate loss, the established CIoU loss [24] is utilized, which factors in considerations like overlapping area, center distance, and aspect ratio. This loss contributes to enhanced detection accuracy, particularly addressing the issue of non-overlapping detection boxes.

The computation formula for the BCE cross-entropy loss is defined as follows, where w_n denotes the averaging of outcomes, and y_n signifies the actual sample label:

$$L_n = -w_n \cdot \left[y_n \cdot \log(S(x_n)) + (1 - y_n) \cdot (1 - \log(S(x_n))) \right] \quad (7)$$

The calculation formula for the CIoU loss is specified as follows, where IoU denotes the intersecting area between the prediction box and the actual box:

$$\text{CIoU} = \text{IoU} - \left(\frac{\rho^2(b, b^{\square})}{b^2} + \alpha v \right) \quad (8)$$

The parameter v quantifies the alignment of the detection frame's aspect ratio, while the parameter α serves as a trade-off parameter, allowing for a greater emphasis on regressing the overlapping area factor [23].

4. Results and Discussion

The YOLOv7 model is trained on two datasets. Loss of localization, loss of confidence, loss of classification, precision, recall, and graphs are presented for both datasets during the training period of the training and validation sets. The loss types included in these graphs are object loss, classification loss, precision, sensitivity, and values. The localization loss indicates how well the algorithm can detect the center of an object and how well the estimated bounding box covers an object. Confidence is a measure of the probability that an object will exist in a suggested area of interest, and high objectivity means that the viewport likely contains an object. Classification loss gives information about how well the algorithm can predict the correct class of a given object.

Different performance measures for both the training and validation sets of the first data set consisting of the “coat” and “other” classes are presented in Figure 4. As can be seen from the graphs, the precision, sensitivity, and average precision values developed rapidly throughout the training. On the other hand, it is seen that localization loss, confidence loss, and classification loss decrease rapidly throughout the training.

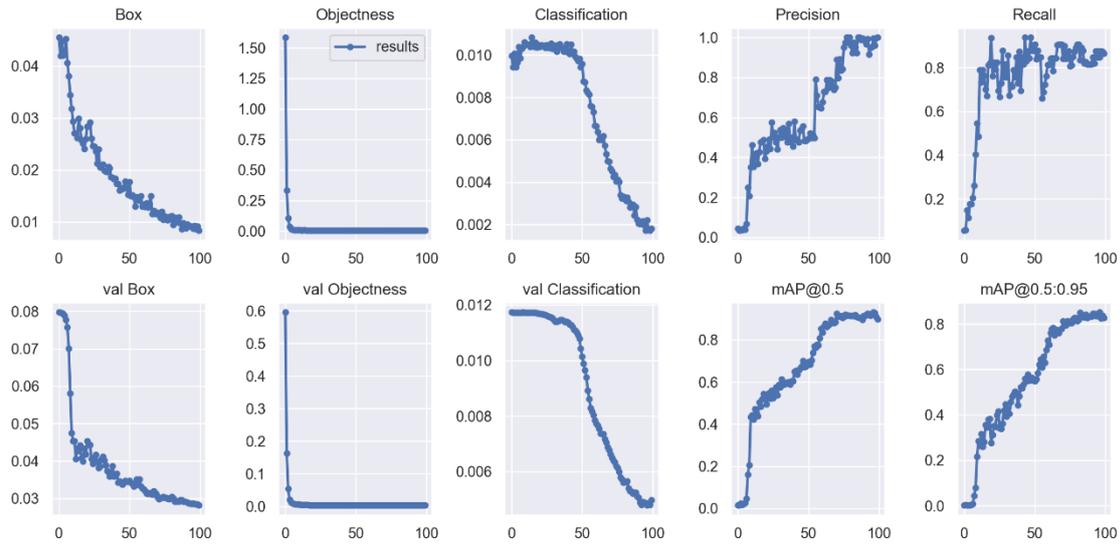


Figure 4. Loss of localization, loss of confidence, loss of classification, precision, sensitivity and mAP plots during training for the first dataset training and validation set.

For the first dataset, 100 rounds of training were completed in approximately 0.73 hours. As can be seen in Table 1, the precision of the dataset trained on YOLOv7 is 0.865, mAP@0.5 is 0.898, and mAP@0.5:0.95 is 0.827, indicating that the performance in the validation set is quite high.

Table 1. First data set validation performance values.

Class	Precision	Recall	mAP@.5	mAP@.5:.95
Coat	1	0.865	0.898	0.827
Other	1	1	0.996	0.938
All	1	0.729	0.779	0.717

As can be seen from the F1 curve presented in Figure 5, the confidence value optimizing precision and sensitivity is 0.702. A high confidence value indicates that a suitable design has been obtained.

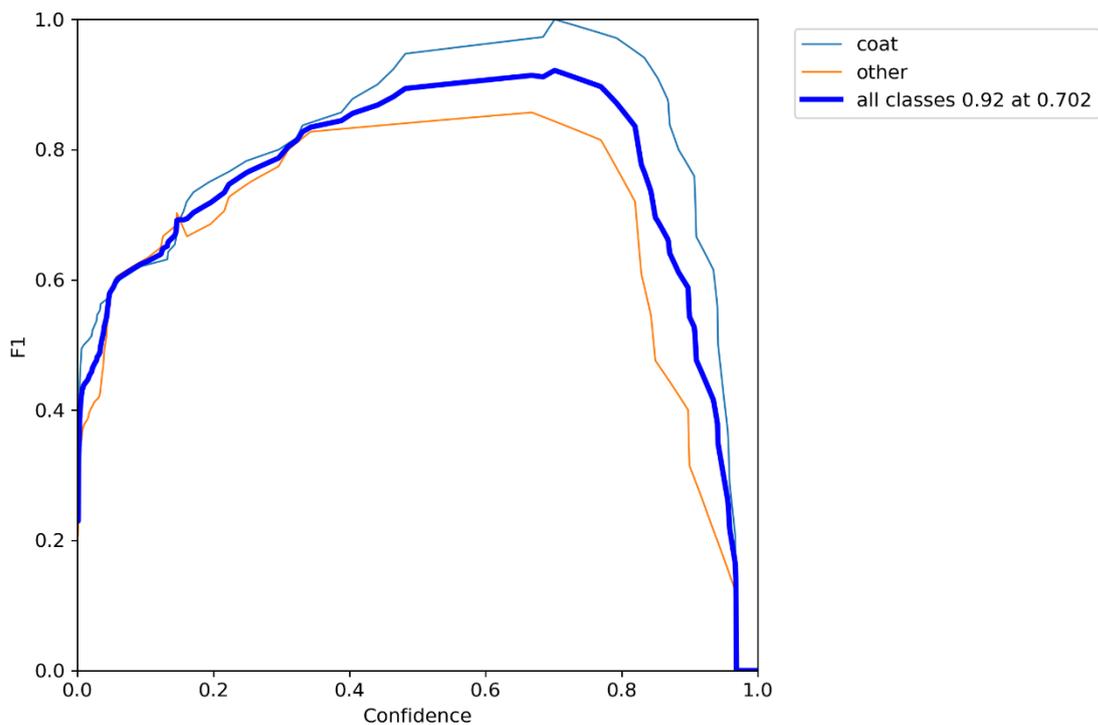


Figure 5. F1 curve for the first data set.

After the model was trained, predictions were made for new and invisible images in the test set. The examples in Figure 6 show that the algorithm can detect the person wearing the coat with greater precision.

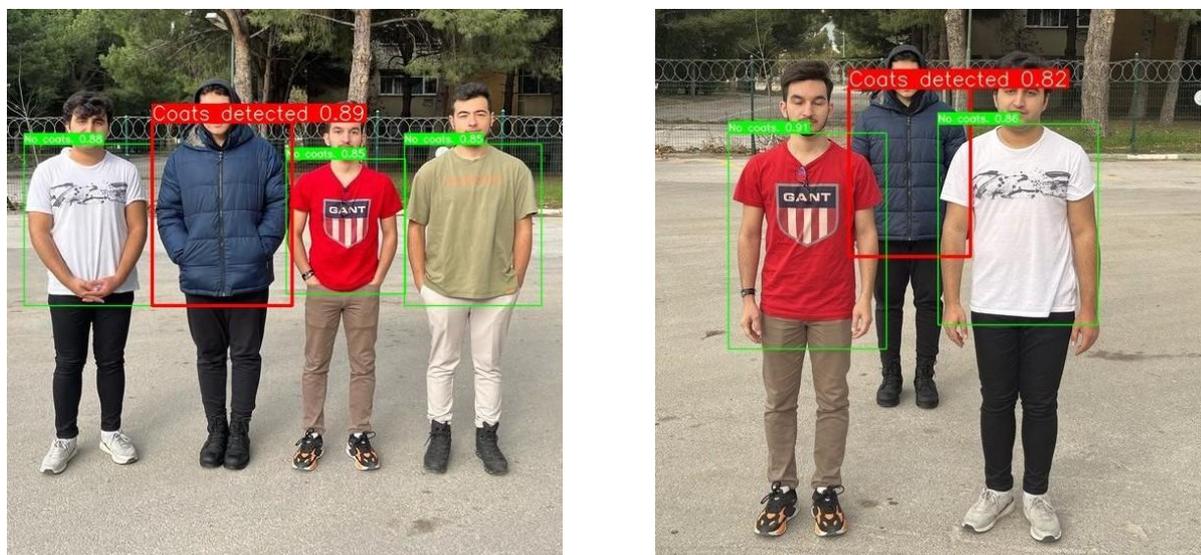


Figure 6. An example test set result for the first data set.

Different performance measurements for both training and validation sets of the second data set, consisting of “no-mask”, “mask”, “glass” and “sunglass” classes, are presented in Figure 7. As can be seen from the graphs, the precision, sensitivity, and average precision values developed rapidly throughout the training. On the other hand, localization loss, confidence loss, and classification loss showed a rapid decrease for the training set throughout the training. However, it is seen that the classification and confidence loss increase for the validation set.

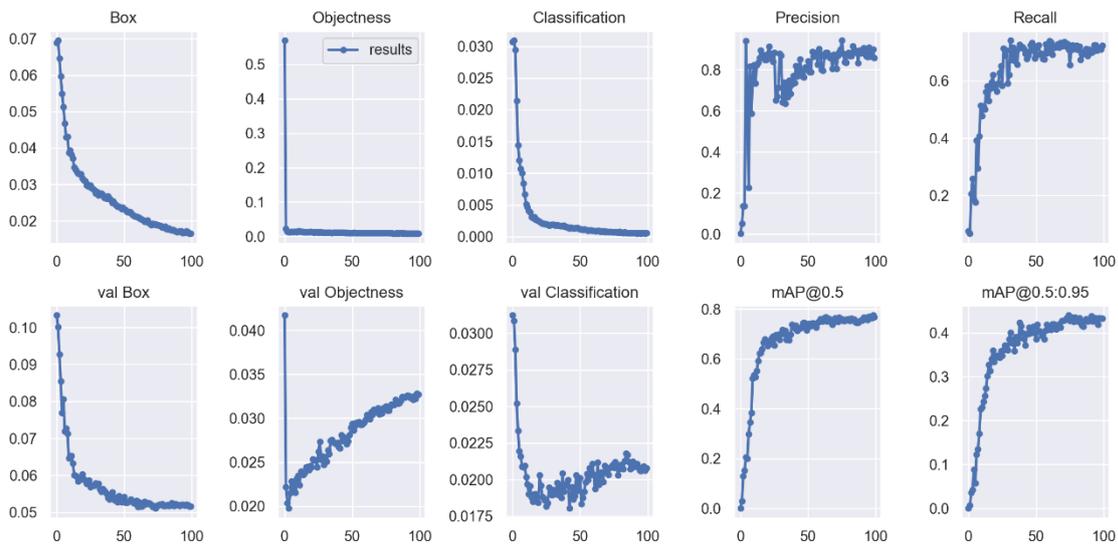


Figure 7. Graphs of loss of localization, loss of confidence, loss of classification, precision, sensitivity and mAP during training for the second dataset training and validation set.

For the second dataset, 100 rounds of training were completed in approximately 2.6 hours. As can be seen in Table 2, the precision of the dataset trained on YOLOv7 is 0.858, the sensitivity ratio is 0.724, mAP@0.5 is 0.768, mAP@0.5:0.95 is 0.433, and the performance in the validation set is quite high.

Table 2. Second data set validation performance values.

Class	Precision	Recall	mAP@.5	mAP@.5:.95
no-mask	0,764	0,824	0,846	0,451
mask	0,891	0,925	0,946	0,593
glasses	0,862	0,814	0,833	0,419

As can be seen from the F1 curve presented in Figure 8, the confidence value optimizing precision and sensitivity is 0.539. Confidence value shows that a suitable design is obtained.

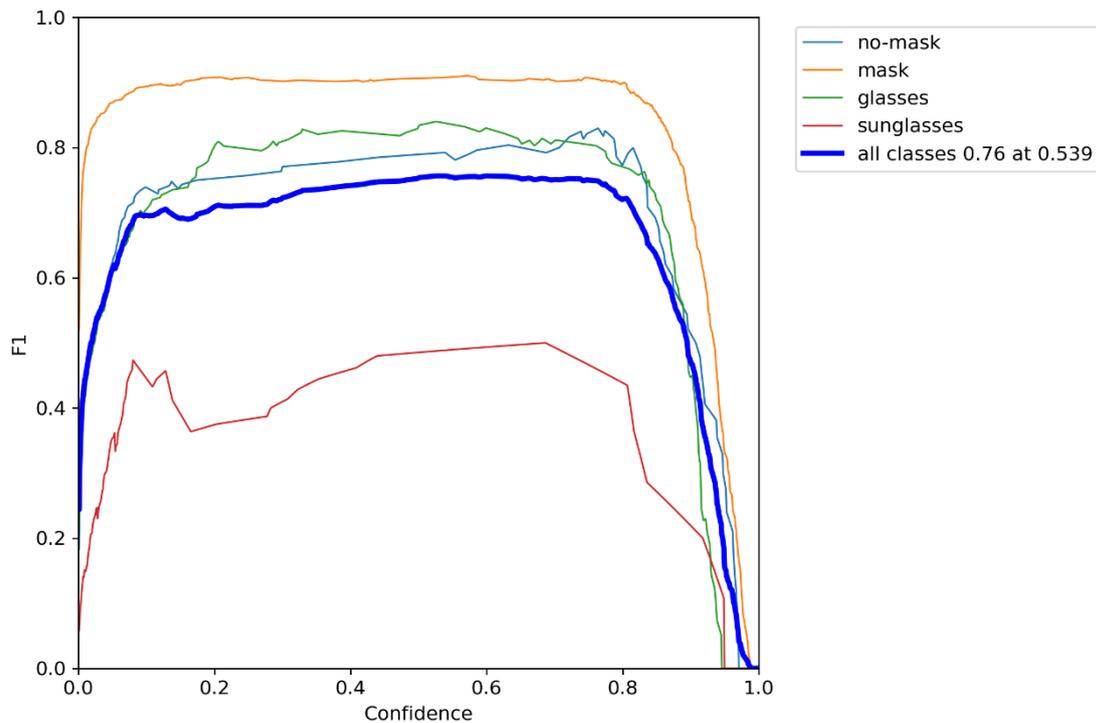


Figure 8. F1 curve for the second data set.

After the model was trained, predictions were made for new and unseen images in the test data set. The use of facial recognition and biometric data is an approach that increases performance in various studies such as person recognition and tracking, emotion and expression recognition, and social behavior analysis. Facial features such as eyes, nose, and mouth carry important information for face recognition. When the mask is worn, the nose and mouth are covered and face recognition cannot be performed. Wearing a mask disrupts the holistic face processing that supports face detection and recognition. As a result, face-matching performance deteriorates, making it difficult to track people. Besides this, people wearing sunglasses prevent the identification process if the eyes are closed. As can be seen from the results in Figure 9 and Figure 10, people who wear masks, sunglasses, or glasses to hide their biometric data can be detected at a high rate. Information alerts are created for cases where the mouth of people wearing masks and the eye area of people wearing sunglasses cannot be detected.



Figure 9. An example test set result for the second data set.



Figure 10. An example test set result for the second data set.

developed a catalytic system for the MBH reaction of methylphenyl glyoxylate with methyl vinyl ketone. The recently-proposed reaction mechanisms revealed the important role of a proton transfer mediator in this reaction. We designed our catalytic system according to the results of these mechanistic studies. We successfully performed the reaction with

DMAP and N-Boc-L-pipecolinic acid. We also proposed a proper reaction mechanism. Our ongoing studies are related to the expansion of the substrate scope and the asymmetric version of this reaction.

5. Conclusion

In the current study, an algorithm that detects outfit anomalies is proposed. The algorithm was developed with the YOLOv7 method. With this method, which provides high performance in real-time object detection, abnormal situations have been detected quite successfully by using some accessories and outfits. The results obtained from the experimental studies have been promising in establishing an early warning system for the detection of suspicious persons by testing them on scenarios applicable to daily life.

In future studies, the scope of outfits and accessories that people will use to hide their identities will be expanded. In addition, a more robust model will be developed for different camera positions, long distances, and more crowded environments.

References

- [1] Xiao, J., Li, S., & Xu, Q. (2019). Video-based evidence analysis and extraction in digital forensic investigation. *IEEE Access*, 7, 55432-55442.
- [2] Kamthe, U. M., & Patil, C. G. (2018, August). Suspicious activity recognition in video surveillance system. In 2018 Fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-6). IEEE.
- [3] Lavee, G., Khan, L., & Thuraisingham, B. (2005, August). A framework for a video analysis tool for suspicious event detection. In *Proceedings of the 6th international workshop on Multimedia data mining: mining integrated media and complex data* (pp. 79-84).
- [4] Bedeli, M., Geradts, Z., & van Eijk, E. (2018). Clothing identification via deep learning: forensic applications. *Forensic sciences research*, 3(3), 219-229.
- [5] Chen, Q., Huang, J., Feris, R., Brown, L. M., Dong, J., & Yan, S. (2015). Deep domain adaptation for describing people based on fine-grained clothing attributes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5315-5324).
- [6] Zapotosky, M. (2021). FBI releases new footage of possible RNC, DNC pipe-bomb suspect, says person probably 'not from the area'. *The Washington Post*, NA-NA.
- [7] Narejo, S., Pandey, B., Esenarro Vargas, D., Rodriguez, C., & Anjum, M. R. (2021). Weapon detection using YOLO V3 for smart surveillance system. *Mathematical Problems in Engineering*, 2021, 1-9.
- [8] Grega, M., Matiołański, A., Guzik, P., & Leszczuk, M. (2016). Automated detection of firearms and knives in a CCTV image. *Sensors*, 16(1), 47.
- [9] Mehta, P., Kumar, A., & Bhattacharjee, S. (2020, July). Fire and gun violence based anomaly detection system using deep neural networks. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 199-204). IEEE.

- [10] Marbach, G., Loepfe, M., & Brupbacher, T. (2006). An image processing technique for fire detection in video images. *Fire safety journal*, 41(4), 285-289.
- [11] Dever, J., da Vitoria Lobo, N., & Shah, M. (2002, August). Automatic visual recognition of armed robbery. In *2002 International Conference on Pattern Recognition* (Vol. 1, pp. 451-455). IEEE.
- [12] Yin, J. H., Velastin, S. A., & Davies, A. C. (1996). Image processing techniques for crowd density estimation using a reference image. In *Recent Developments in Computer Vision: Second Asian Conference on Computer Vision, ACCV'95 Singapore, December 5–8, 1995 Invited Session Papers 2* (pp. 489-498). Springer Berlin Heidelberg.
- [13] Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L., Regazzoni, C., & Sebe, N. (2017, September). Abnormal event detection in videos using generative adversarial nets. In *2017 IEEE international conference on image processing (ICIP)* (pp. 1577-1581). IEEE.
- [14] Mehran, R., Oyama, A., & Shah, M. (2009, June). Abnormal crowd behavior detection using social force model. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 935-942). IEEE.
- [15] Heartexlabs. Heartexlabs/labelimg: LabelImg is now part of the label Studio Community. the popular image annotation tool created by Tzutalin is no longer actively being developed, but you can check out label studio, the open-source data labeling tool for images, text, hypertext, audio, video and time-series data.
- [16] 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* (pp. 779-788).
- [17] Atik, M. E., Duran, Z., & ÖZGÜNLÜK, R. (2022). Comparison of YOLO versions for object detection from aerial images. *International Journal of Environment and Geoinformatics*, 9(2), 87-93.
- [18] Yang, F., Zhang, X., & Liu, B. (2022). Video object tracking based on YOLOv7 and DeepSORT. *arXiv preprint arXiv:2207.12202*.
- [19] Hussain, M., Al-Aqrabi, H., Munawar, M., Hill, R., & Alsoubi, T. (2022). Domain feature mapping with YOLOv7 for automated edge-based pallet racking inspections. *Sensors*, 22(18), 6927.
- [20] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7464-7475).
- [21] Soudy, M., Afify, Y., & Badr, N. (2022). RepConv: A novel architecture for image scene classification on Intel scenes dataset. *International Journal of Intelligent Computing and Information Sciences*, 22(2), 63-73.
- [23] Zheng, J., Wu, H., Zhang, H., Wang, Z., & Xu, W. (2022). Insulator-defect detection algorithm based on improved YOLOv7. *Sensors*, 22(22), 8801.
- [24] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921-2929).