Multimedia Tools and Applications (2023) 82:44259–44275
https://doi.org/10.1007/s11042-023-15358-1

Concealed pistol detection from thermal images with deep neural networks

Ozan Veranyurt[1] · **C. Okan Sakar**[1]

Received: 31 January 2022 / Revised: 23 February 2023 / Accepted: 15 April 2023 /
Published online: 3 May 2023

Abstract
Violence involving firearms is a rising threat that requires precise and competent surveillance systems. Current surveillance technologies involve continuous human observation and are prone to human errors. To handle such errors and monitor with minimal human effort, new solutions using artificial intelligence approaches that can detect and pinpoint the threat are required. In this study, our aim is to develop a deep learning-based solution capable of detecting and locating concealed pistols on thermal images for real-time surveillance. For this purpose, we generate a dataset consisting of thermal video recordings of multiple human models and combine this dataset with thermal images from public sources. Then, we build up a deep learning-based framework by combining two deep learning models that detects and localizes the concealed pistol in the given thermal image. We evaluate multiple deep learning architectures for the classification and segmentation of the images. The best test set results in detecting the concealed pistol was achieved by a fine-tuned VGG19-based convolutional neural network model with an F1 score of 0.84 on the test set. In the second module of the system, a fine-tuned Yolo-V3 model trained as a multi-tasking model for both classification and location detection gave the highest mean average precision value of 0.95 in labeling and locating the pistol in a bounding box in approximately 10 milliseconds. The findings exhibit the potential of using deep learning techniques with thermal imaging for the real time concealed pistol detection.

Keywords Deep learning · Convolutional neural networks · Transfer learning · Thermal imaging · Concealed weapons

1 Introduction

Gun violence is a serious crime observed in either modern or emerging countries. Gun-aided assaults and burglary are different categories of crime in which hand-guns are utilized as primary aid. Using modern security cameras require human surveillance regardless

✉ Ozan Veranyurt
ozan.veranyurt@bahcesehir.edu.tr

C. Okan Sakar
okan.sakar@eng.bau.edu.tr

[1] Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey

of having smart features in order to detect any violent act or a concealed dangerous object. Moreover, current technologies for detecting concealed objects work in a high frequency spectrum which poses risk against human health [11]. Detection technologies such as backscatter scanner behave similarly as x-ray devices, and it is claimed that they pose ionizing radiation threat over acceptable levels [15]. Besides, in suspicious cases the security personnel may be required to search the person physically which may cause privacy and law issues against the searched party. A potential solution for these problems may be to develop an intelligent surveillance system that is capable of detecting concealed threats minimizing the human intervention that may lead to unwanted occasions.

There are some research efforts in building weapon detection systems that can be used to detect security threats in public areas [7, 12, 22]. These studies vary with respect to the aim of the study, electromagnetic method used to detect the weapon and classification/segmentation method used to process the obtained input. In this study, the aim is to develop an efficient thermal imagery-based deep learning solution for real time surveillance that is capable of detecting concealed pistols on images and video. To achieve this, our proposed solution works with thermal images captured with Infra-Red (IR) spectrum cameras that can read human body heat signature without emitting any high frequency signal that may impact persons health. Besides, the proposed framework does not require any additional sensors for detection as it is built solely with thermal cameras and a processing capable computer. This simplistic structure of the solution makes it more cost efficient compared to the systems based on multiple sensors and devices.

The proposed solution can work independent from continuous human monitoring, and it can detect pistols without the contact of rays or radio waves to the human body such as x-ray or other types of magnetic waves. Two deep learning models were combined two provide a flexible solution that fits different computation needs for real time security surveillance solutions. For the classification and pistol location detection tasks, VGG-19 and Yolo-V3 architectures achieved the best results, respectively. The system is mainly suitable to be used in places where constant manual surveillance is obligatory like hospitals, government offices, airports.

There are some recent works that utilize deep learning techniques for concealed weapon detection [3, 5, 9, 16, 20]. The main contributions of this study are as follows:

- Building a concealed pistol detection solution that does not pose risk to human health, offers both classification/location detection capabilities, and adaptable to different environments,
- Combining two deep learning models to provide a flexible solution that fits different computation needs for real time security surveillance,
- Proposal of a threshold function with the aim of reducing the false positive rates and eliminating duplicate detection of pistols in the solution,
- Achieving a deep learning model capable of detecting concealed pistols in less than 10 milliseconds with a single average GPU setup,
- Building a concealed pistol dataset consisting of thermal images belonging to multiple human models,
- Applying transfer learning and fine-tuning techniques with various deep neural network (DNN) architectures for concealed pistol detection and segmentation,
- Evaluation of the models in terms of both accuracy and frames per second (FPS) to consider the requirements of a real-time concealed pistol detection system.

The remainder of this paper is organized as follows: Section 2 presents the existing works related to the detection of concealed weapons with machine learning and deep

learning techniques. In Section 3, the description of the dataset consisting of thermal images used to train and test the system is given. In Section 4, we give the details of the proposed thermal image-based pistol detection framework along with the overview of the methods used. Section 5 presents the experiments performed to evaluate the pistol detection system. The conclusions are given in Section 6.

2 Related works

The existing works in building weapon detection systems to detect security threats in public areas vary with respect to the aim of the study. Ther are electromagnetic methods used to detect the weapon and classification/segmentation methods used to process the obtained input. The studies that use camcorder imaging with deep learning or computer vision techniques only aim to detect unconcealed weapons and hence are not capable of detecting concealed weapons. Although the solutions that are based on the use of devices working in the high frequency spectrum like millimeter wave scanner (MMW), x-ray imager, or microwave radar imagers are capable of detecting concealed weapons, they pose a health risk once exposed continuously or repeatedly [15]. Another fact regarding MMW based solutions is that if the sensors are used close to the human body, the compliance of radiation exposure may not be properly measured [18]. Purely sensor-based solutions such as metal detectors or x-ray devices can detect concealed metal objects only at a fixed location and do not support continuous monitoring. Besides, the solutions using MMW imaging require multiple frames of a livestream to be processed or further techniques applied on the images for accurate detection [24]. Since we focus on concealed gun detection in our study, which is a more realistic and useful scenario in video surveillance, we give the overview of studies that aim to detect concealed guns in this section.

In a recent study focusing on concealed weapon detection, Pang et al. [18] proposed a solution which processes passive MMW images with a DNN architecture. They built their own dataset called the Passive Millimeter Wave (PMMW) dataset which consists of 1624 MMW images. The experiments were performed with Yolo-V3, Single Shot MultiBox Detector (SSD) and SSD-VGG16 and the best performance was achieved with Yolo-V3 which demonstrated a processing speed of 36 FPS and a mean average precision (MAP) of 95% on a GPU-1080Ti computer. 5The studies focusing on the detection of concealed weapons use different types of data for detection. MMW is one of the recent and fast adapted technologies in detecting concealed objects. Via a dielectric lens and a receiver functional at 94 Gigahertz, images are analyzed with segmentation. Each segmented part is considered as a vector quantization and processed whether it hides a concealed object. In a related study, Yeom et al. [25] proposed an efficient method based on MMW for the detection and segmentation of a moving human carrying a concealed weapon. In this study, they used Bayesian decision making to make clusters from pixels and improve the segmentation performance.

Yuenyong et al. [26] used thermal/infrared images as in our study for knife detection which is an object that can used as a weapon in public areas. They used their own dataset of 8527 images including knives with different shapes and sizes. A pre-trained deep learning model was fine-tuned to classify thermal images as persons with or without knives. The researchers used the GoogleNet architecture as the baseline convolutional neural network (CNN) model and fine-tuned it with the thermal dataset. The hidden knives were detected with an accuracy of 97.91% using the fine-tuned GoogleNet.

Hussein et al. [8] proposed a new approach to detect concealed weapons using wavelet transforms merged with dimension reduced meta-heuristic algorithms. Shape matching for concealed weapon detection was developed with k-means clustering and SVM classifier. In the hybrid solution, infrared and RGB images were processed using different fusion techniques. The experiments were performed on a custom dataset built by the researchers consisting of a pair of visual and infrared images. The images were taken with two cameras which were a standard light camera and an infrared camera. Then, the study was conducted on 20 image pairs and false alarm generation rates were presented for different fusion techniques. Best results were achieved by PixelAvg technique which obtained lowest mean square error (MSE) (<100).

In another study, inception-v3 was used to detect humans with weapons using infrared images [14]. The purpose of the study was to prevent smuggling of Sandalwood trees by using a deep learning-based thermal detection solution. The researchers used their custom dataset for the experiments. Transfer learning over inception-v3 was applied and 99% accuracy was obtained in the tests.

Kowalski [10] proposed another approach for hidden object detection using passive imagers functional in terahertz (THz) range and a mid-wavelength (MWIR) at 50-100 THz. Researchers collected their own dataset of images with different items and various types of clothing for the experiments. The research proposed the comparison of two detection methods which were Yolo-V3 and R-FCN (Region Based Fully Convolutional Networks) in both spectrums. R-FCN outperformed Yolo-V3 in terms of accuracy (92.5% vs 80.5%); however, while YOLO-V3 was processing at 11 FPS, R-FCN was processing slower with 7 FPS.

Wei and Liu [23] suggested a deep learning-based method for detecting dangerous items in x-ray detectors. In their experiments, they used the GDXray dataset which contains 8150 images and four dangerous object categories; knife, handgun, shuriken and razor blade. Since it was a challenge to compile a dataset of x-ray images with items classified as dangerous, the researchers focused on using multitask learning on SSD300 for both object classification and location detection (localization). The final layers of the model were adjusted with multiple outputs to improve the generalization ability of the system. The study achieved 0.915 MAP on the location detection task.

In a recent study, Cheng et al. [4] proposed a concealed object detection solution by using terahertz images. They took the SSD deep learning object detection network and replaced the core of the network with a residual network in order to decrease training time. Additionally, they proposed an algorithm based on feature fusion on terahertz images in order to catch small target objects. In their experiments, SSD networks performance improved from 95.04% to 99.92% accuracy. The proposed model also achieved 99.92% MAP.

Lamas et al. [13] focused on the estimation of human pose in order to reduce false negatives in weapon detection while using CCTV video surveillance. CNN based deep learning algorithms may still yield false positive results in real time detection of weapons and in order to mitigate that the researchers focused on the human pose rather than focusing on estimating the exact location of the weapon. They defined a new factor called adaptive pose factor which considers the distance of human body from the camera and came up with a new methodology called WeDePE (Weapon Detection over Pose Estimation). Their experiments were held on the Sohas weapon detection dataset consisting of 3250 images which include knives and pistols. Their suggested methodology improved finetuned object detectors performances such as Faster RCNN, SSD and Centernet. In another recent study Goenka and Sitara [6], focused on usage of deep learning specifically Mask RCNN model in order to detect weapons in CCTV surveillance images. In their proposed solution they applied gaussian deblur technique to boost the

main features of pistol so that detection could be more effective especially in images where the pistol is blurrier. Their experiment results yield improved results if their pre-processing was applied.

3 Dataset description

In this study, the images obtained from the thermal camera were classified in two classes as positive and negative. The frames including a concealed pistol are labeled as positive. Below the datasets and their details are summarized:

- Self-Created Dataset: This dataset is created in the context of this study. It consists of 600 images belonging to 11 people taken in different locations with or without concealed pistols. Since it is not allowed to carry a pistol outside without an appropriate carry permit, 380 of the thermal images were taken from six police officers with the permission of the Police Department and Governorship of Istanbul, Turkey and annotations of the images were performed later on manually. The other part of the dataset was taken from five civilians that have carry permit.
- Trimodal Dataset: The trimodal dataset is a publicly available dataset [17]. It consists of 5724 annotated thermal images taken in three different scenes. Images had gone through a selection and adjustment process before they were used in our study.
- LTIR dataset: This dataset contains 8 and 16-bit thermal images of 20 different categories which are later used in the without a concealed gun category [1].

Figure 1 includes a set of exemplary thermal images from the self-created dataset used in the study. Examples of both classes, with and without concealed pistols, are shown. As it is seen, the images were taken from human models wearing different types of clothing such as t-shirt, thin jacket, and shirt with the aim of testing the accuracy of the proposed solution under different conditions.

4 Proposed framework

In this section, we provide the details of the proposed system. The pistol detection module of the framework has been explained in Section 4.1. In Section 4.2, the methodology used for pistol localization is given. Finally, we provide the details of the proposed combined framework in Section 4.3.

4.1 Pistol detection

The first step of the proposed framework is to detect the concealed pistol in the thermal video imagery. For this purpose, we use thermal images captured with IR spectrum cameras that can read human body heat signature without emitting any further signal. The heat signature of the person is recorded in a thermal matrix and pixel mapped to an image within the thermal camera sensor. Then, the image is pre-processed with sharpening and contrast operations so that concealed objects in the image gain clarity for

Fig. 1 Thermal images from the combined dataset with no pistols (above) and concealed pistols (below)

detection. Afterwards, the thermal image is processed by the first module of the proposed framework for pistol detection. For this task, we evaluated the performance of various pre-trained CNN models including VGG19, VGG16, ResNet50, InceptionV3, MobileNetV2 and DenseNet12 constructed for object detection.

The CNN models were trained with the same set of hyper-parameters for an unbiased hyper-parameter optimization. On top of each pre-trained model a flattening layer and 4 fully connected layers were integrated for the pistol detection task. We used 128×128 images for faster processing and SGD (Stochastic Gradient Descent) optimizer with Nesterov momentum. The thermal images labeled as positive are given to the second module of the proposed framework for location detection.

4.2 Pistol location detection

The aim of the second module is to classify the image and detect the location of the pistol. In this part of the study, the state-of-the-art real time object detection systems that have been used successfully for various object detections tasks including Yolo-V3, Yolo-V2, Tiny-Yolo, SSD, Masked-RCNN models were fine-tuned and evaluated. The YOLO (You Only Look Once) object detection system is based on an artificial neural network

architecture that can detect objects in the image with their positions at once. By dividing the image into cells, it produces output vectors that try to predict confidence values and bounding boxes for objects in each cell [19]. The Yolo models used in this study apply similar principles but vary in architecture. The region-based object detection algorithms such as R-CNN first determine the areas where objects are likely to be found and then execute separate CNN classifiers there.

Mask R-CNN is also based on a DNN architecture used in computer vision for object segmentation. It executes in two main steps. Firstly, it generates predictions about the regions that may include objects based on the input image, and then it predicts the object's class, refines the bounding box, and creates a pixel-level mask of the object based on the first stage estimation [27, 28]. SSD, on the other hand, acts like YOLO and processes only one shot of the given image to detect multiple objects present in an image using multi-boxing. In SSD, the bounding-box proposals are eliminated as in R-CNN variations resulting in a faster processing speed.

4.3 Proposed model

The framework of the proposed system including pistol detection and location detection modules is shown in Fig. 2. Based on the experiments which will be further discussed in Section 5, VGG19-based classifier and Yolo-V3 based pistol location detector were chosen for the related tasks in the proposed system. As shown Fig. 2, the images are taken via a modular or an external thermal camera and sent to the processing computer for the preprocessing step including sharpening and contrast boosting operations. Then, either the VGG19-based detector determines whether there is a concealed pistol in the image, or the classification and location detection tasks are both handled by the fine-tuned Yolo-V3 model. The location detector model can also work with the positive labeled image, that is fed from the VGG19 model, as a complementary function for demystifying the location of the concealed pistol.

Both concealed pistol detection and localization modules of the system are designed as configurable parameters considering that the hardware system on which the proposed framework is executed may not have sufficient computing capability for both tasks. If the target system has limited computational capability or no GPU, then it may use only the pistol detection module and bypass the location detector which will consume less resource and achieve faster processing speed. Additionally, this brings flexibility in the usage of deep learning modules. When the location detector module is bypassed, the system is designed to purely show a warning by constantly checking the configurable security counter threshold and increasing the security index counter. Once the classifier detects consecutive images with a probability estimate exceeding the threshold value, a system alarm, which informs the user that a pistol has been detected, is generated. In case the location detector module is enabled, detection of the concealed pistol and the estimated location are shown in a bounding-box along with a warning once the security index counter reaches its threshold. In this solution if the location detector module is executed solely, it is trained as a multitask model capable of classifying the concealed object and providing its location as a second output. The confidence threshold value in the range of [0, 1] is another numerical parameter used for location detector optimization. This setting triggers a function which impacts both the minimum confidence level for detection and the intersection over union (IoU) threshold for preventing duplicate detections of the same object. Additionally,

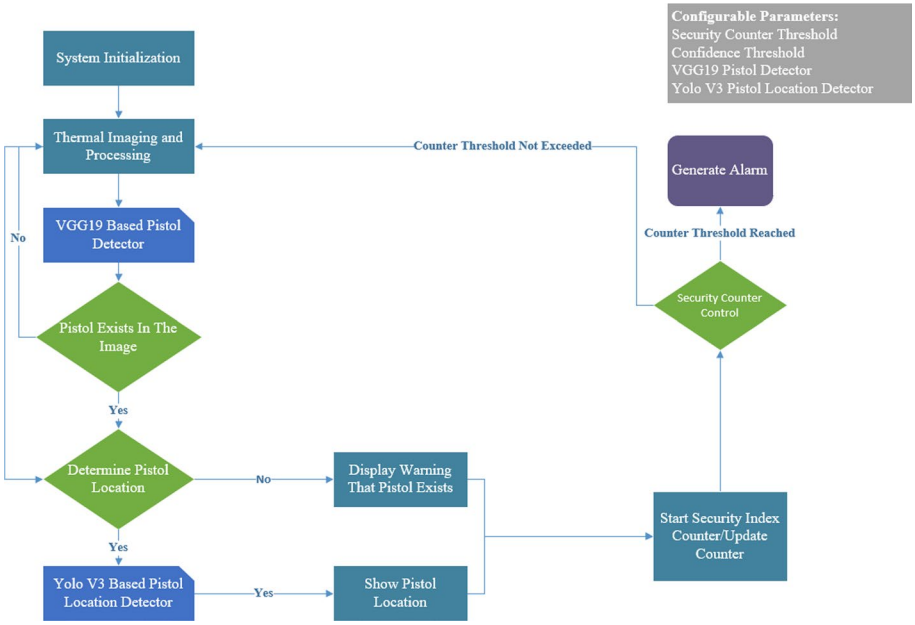

Fig. 2 Thermal image-based concealed pistol detection system proposed in this study

fine-tuning this method helps in reducing the false positives caught by the location detector module. When both modules work together, pistol detection module feeds the pistol detector module with filtered images of already classified as containing pistols in order to speed up the detection. In this case both modules will reduce the security counter and an alarm will be generated once the counter reaches the threshold. Although these CNN architectures were typically used for classification tasks consisting of RBG images, they have also been successfully used on thermal or MMW images with fine-tuning. Therefore, we used a similar approach in our methods. The proposed system can also be applied on a mobile phone with the same detection modules using an attachable thermal camera [21].

In the pistol detection module of the study, we used VGG19 architecture. The model was finetuned thoroughly and merged with 3 consecutive fully connected layers. Eventually the model had 8192 parameters and applied output with a softmax activation function for classification purpose.

As illustrated in Fig. 3, thermal images that are downsized to 128×128 after preprocessing are fed into the down-sampling layer which produces three feature vectors for producing small, medium and large-scale grids. This is one of the built-in and salient features of Yolo-V3 which makes detections in three different scales. As Yolo-V3 is a fully convolutional network, it generates the output by application of a 1×1 kernel on the feature map. In Yolo-V3, the detection is achieved by application of 1×1 kernels on the feature maps which are three different sizes at three different locations in the network. This makes the architecture more successful compared to its predecessors. The derived feature vectors are then given as inputs to the detection layer which produces the relative location offsets, object probability and probability for the two object classes relevant to the study. The detection layer produces three filters for each feature vector as Yolo-V3 architecture offers three anchor boxes to be used in each grid

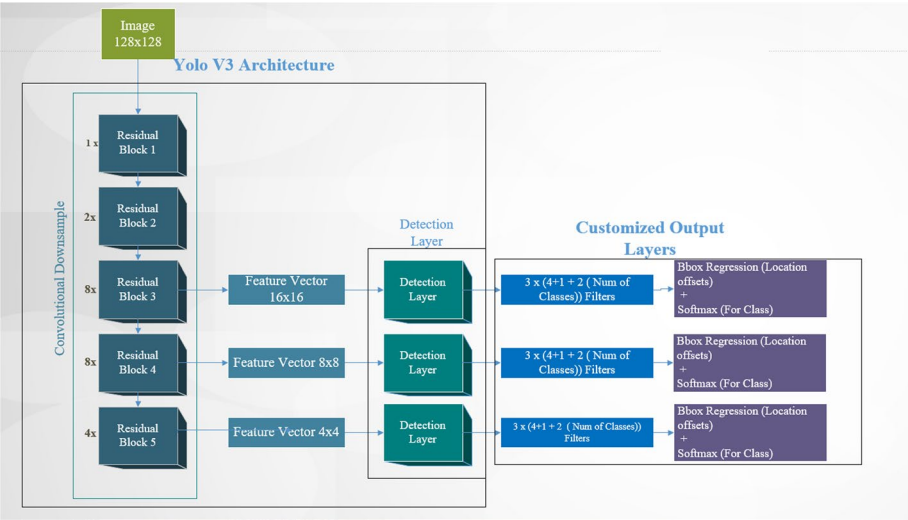

Fig. 3 Concealed pistol detector/locator model architecture with Yolo-V3

cell. The output vector contains the relative location offsets, object probability and class probability per anchor [23]. In the output section of the architecture, we divided the localization and classification tasks and fed into a composition of losses which are regression for bounding boxes and soft-max for the object classes. In this approach, the fine-tuned architecture of Yolo-V3 is used to generate one output vector that is optimized with a composite loss function instead of evaluating multiple outputs in separate loss functions. The output layer of Yolo v3 was also finetuned and reduced in terms of filter layers in order to classify only two objects which were the background and pistols.

5 Results

In this section, we present the experimental setup and results for the pistol detection and segmentation tasks. Specifically, in Section 5.1 we give the details of the experimental setup for the experiments. We also present the definition and formulas of the evaluation metrics used for both tasks in the same section. Then, the results obtained with various DNN architectures for the pistol detection and pistol segmentation tasks are given in Section 5.2 and 5.3, respectively.

5.1 Experimental setup

We evaluate the main steps of the proposed system in two main parts. In the first part, we evaluate the performance of different fine-tuned CNN classifiers for pistol detection from the thermal images. In the second part, different CNN models were fine-tuned and

evaluated with the purpose of concealed pistol detection and determining the location of the concealed pistol.

In the first part of our experiments, for choosing the right fine-tuned classifier for concealed pistol detection, we used accuracy, AUC (Area Under Curve), precision, recall and F1 score as evaluation metrics. To avoid an overlap between training and test sets, we used leave-subject-out cross-validation strategy during experiments. For this purpose, 572 images belonging to 15 human models were separated for training and the rest 100 images belonging to the remaining four human models for testing so that the images belonging to a specific human model are either in the training or test set. The hyper-parameter optimization procedure was performed on the training set by dividing it into further training and validation sets. For statistical significance, the train-test split procedure is repeated 15 times and the average values of the metrics are reported. We also used a non-parametric statistical test, Wilcoxon test, to evaluate the statistical significance of the results.

In the second experiment, we aimed to find the best CNN model for classifying and detecting the location of the concealed pistol. For this purpose, the performances of the models given in Section 4.2 used to detect the most approximate bounding boxes were evaluated. We used MAP for evaluation which takes into account the average precision and recall of a model given a certain confidence threshold. In this experiment, we used 0.4 as the confidence threshold value for comparing the intersection over union (IOU) between the ground-truth images and the predicted ones. In the experiments, this value has been found as the optimum confidence threshold based on the hyperparameter optimization procedure applied for training and validation. During the experiments we observed that using confidence threshold lower than 0.4 lead to too many false positives and objects being detected in the same space and location. In the evaluation of object localization we used MAP, as it is one of the most common metrics used in object detection and localization comparison. For MAP calculation, we used the method proposed by Cartucho et al. [2]. For a given range, the average precision is calculated and then the mean of all calculated AP scores for the number of queries is taken as one summary score. For the precision and recall calculations used to obtain MAP, the definition of a true positive is a bounding box higher than the IoU confidence threshold.

Since the proposed solution is aimed to be used for real-time concealed pistol detection, we also evaluated the detection performance with the FPS metric. For this part of the study, the images containing a pistol were divided into three sets as 220 for training, 40 images for validation and 95 images for testing. The trials were performed with a hardware setup consisting of an Intel Core I7 8th Generation CPU and a GTX 1060 GPU. In the final step, the classification capability of each model was assessed with accuracy, F1 score and AUC in the test set of 190 images which was equally balanced in terms of classes.

Table 1 summarizes the metrics used in both experiments and their formulas. Accuracy is a standard metric used in classification experiments which calculates the successful detection of true positives and true negatives against all detected samples. Recall calculates the ratio of true positives against the sum of true positives and false negatives. In a similar fashion precision observes the ratio between the true positives against the sum of true positives and false positives. Both metrics are important signs for success in unbalanced datasets and demonstrates the achievement of the network against false positive and false negative samples. F1 score takes into account both precision and recall by calculating their harmonic average and yields a score in the range of [0, 1]. ROC-AUC score looks at the area covered when a roc curve is plotted with true positive

Table 1 Evaluation metrics used for classification and location detection tasks

Measure	Formula
Accuracy	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
Recall	$Recall = \frac{TP}{TP+FN}$
Precision	$Precision = \frac{TP}{TP+FP}$
F1 score	$F1 - score = 2 * \frac{PrecisioxRecall}{Precision+Recall}$
ROC-AUC Score	$\int_0^1 TPR(FPR)dFPR$
MAP	$\frac{1}{n}\sum_1^n APk$

rate (TPR) against the false positive rate (FPR). The curved area yields results between 0 and 1. Mean average precision is the metric used in the second part of the experiment and it is generally used for object segmentation tasks. It takes into account the average precision (AP) of each class and calculates the mean based on n which is the number of classes being detected.

5.2 Pistol detection results

We evaluated the performance of VGG19, VGG16, Resnet50, InceptionV3, MobileNetV2 and DenseNet12 pre-trained CNN models for the pistol detection module. For each model, the tests were performed 15 times for statistical significance and average values of accuracy, F1 score, precision, recall and AUC were reported.

As seen in Table 2, the highest average test set accuracy with 85% was obtained with the VGG19 model. Similarly, the best performance in terms of F1 score and precision were also achieved by the same model. We also see that although the average accuracy for InceptionV3 was fairly low compared to VGG19 and VGG16, it demonstrated the highest recall showing that it is more successful in detecting the positive cases.

Considering that the proposed solution addresses a security problem, recall should have more precedence compared to precision since it is critical to detect the positive cases that carry a gun. However, the Wilcoxon test showed that the F1 score of VGG-19 is significantly higher than that of InceptionV3 (p value <0.01). The results also show that although MobileNetV2 also gave a higher recall than VGG19, its precision is even lower than Inception V3. The next highest recall of 0.88 was obtained with VGG16 which also performed well in terms of F1 score compared to Inception V3 and MobileNetV2. The Wilcoxon test showed that the difference between the F1 scores of VGG16 and VGG19 is not statistically significant. Therefore, VGG16 can be preferred over VGG19 in the systems in which false

Table 2 Average results obtained on test set for the concealed gun detection task

Model	Accuracy	F1 score	Precision	Recall	AUC
VGG19	85%	0.84	0.85	0.85	0.85
VGG16	84%	0.84	0.81	0.88	0.84
Resnet50	77%	0.76	0.75	0.78	0.76
InceptionV3	70%	0.76	0.64	0.94	0.69
MobileNetV2	56%	0.66	0.53	0.91	0.57
Densenet12	81%	0.81	0.81	0.83	0.81

negatives are more critical than false positives. We should also note that the simplicity of the architecture of VGG16 containing a smaller number of hidden layers than VGG19 can be considered as another advantage for such real time systems.

5.3 Pistol location detection results

The second module of the proposed system is designed to either collaborate with the first model and detect the location of the pistol or work independently for the detection and location tasks. For this part, the accuracies of multiple deep CNN object detection models including SSD, Mask R-CNN, Tiny Yolo, Yolo-v2, and Yolo-V3 were fine-tuned and evaluated in terms of the MAP metric. Besides, considering that speed is one of the main concerns in such real time detection systems, we present FPS as an additional evaluation metric. As the final step the classification capability of each model was assessed in a comparative manner.

In the calculation of the MAP value for each model, we also analyzed the precision-recall curves. The precision-recall curves cover 0-1 range. The measurement starts from confidence threshold 0.4 and measures precision-recall changes as confidence threshold increases. After a certain confidence threshold point, precision and recall values were not measurable as the model was not capable of detecting any object and recall values dropped to 0 as seen in Fig. 4 which shows the precision-recall graph for SSD300 and Mask R-CNN models. We applied fine tuning on all hidden layers of these models. As seen from Fig. 4, SSD300 demonstrated 80.66% average precision and as the area under the curve implies both precision and recall values showed a positive trend over 80% for different confidence thresholds measured in scale from 0.4 to 1. In the next attempt, we observed the average precision for the Mask RCNN model. As illustrated in Fig. 4, the average precision obtained for this model was 29.87% which was significantly lower compared to SSD300. Besides, we also see that precision and recall values obtained with Mask R-CNN did not show a smooth transition.

The performance results for Tiny-Yolo and Yolo-V2 are given in Fig. 5. As it is seen, the Tiny Yolo model achieved an average precision of 27.92% and the precision-recall relation showed an imbalanced behavior with changing confidence threshold. Figure 5 shows that the Yolo-V2 model performs better than Tiny-Yolo with an average precision of 41.61%. Besides, the model's precision-recall relation was smoother compared to Tiny-Yolo. Therefore, in the last attempt, we evaluated the performance of Yolo-V3. As seen in Fig. 6, the Yolo V3 model showed the highest average precision with 95.02%.

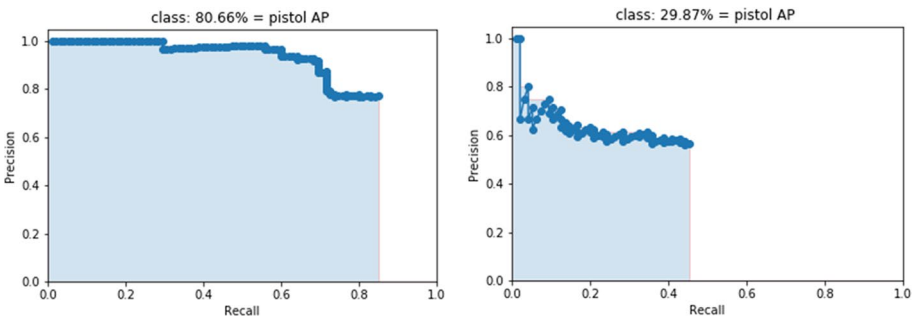

Fig. 4 Precision-recall curve for (**left**) SSD300 (**right**) mask R-CNN

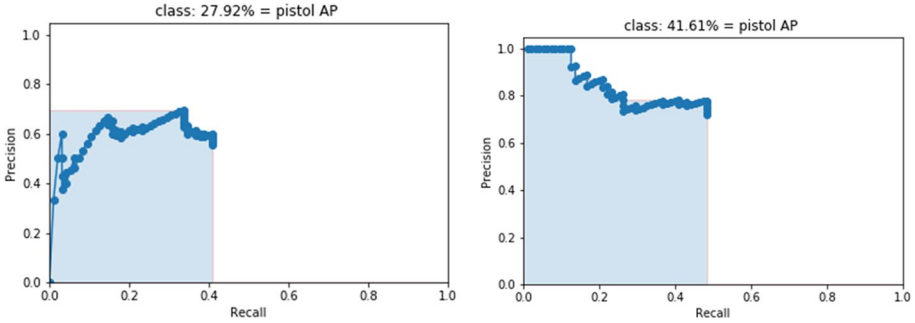

Fig. 5 Precision-recall curve for (**left**) Tiny Yolo (**right**) Yolo V2

The model also demonstrated the highest precision values against recall variations within the given confidence range.

Table 3 summarizes the average precision, FPS and classification results for all models used to detect the location of the pistols. As seen, the highest FPS was observed with Tiny-Yolo which is due to the reduced size of the model. Considering that precision and speed are both important factors in a real-time detection system, Yolo-V3 which gave the highest AP value with 10 FPS performance was chosen to be used in the final detection system. The model could detect the location of a concealed pistol in an image in approximately 10 milliseconds. Besides, based on the classification results, the highest accuracy, F1 score, and AUC values were achieved by the Yolo-V3 model. The model achieved 92% accuracy in the equally balanced test set and eventually similar AUC values were observed. Highest F1 score was also achieved by Yolo-V3.

In Fig. 7, sample images from the detection results of Yolo-V3 are shown. In the left-side images, the ground truth and predicted bounding box by Yolo-V3 had an IoU over 0.9 which can be regarded as a successful prediction. On the other hand, as seen in the right-side images, the model detected two concealed pistols shown with the red bounding boxes while there was only one concealed pistol. During the tests, we observed that in some cases the model made multiple predictions which were false positives accompanying the right prediction. To alleviate this problem, we applied a fine-tuning method for reducing the

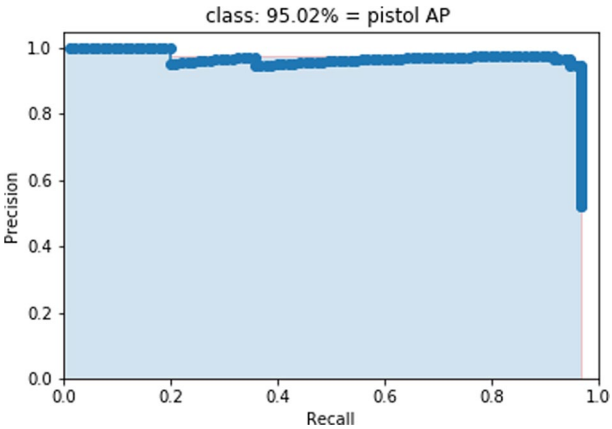

Fig. 6 Precision-recall curve for Yolo-V3

Table 3 Summarized test set results obtained for the pistol location detection task

	Location Detection		Classification		
	Mean AP	FPS	Mean AP	F1 Score	AUC
Yolo-V3	95%	10	92%	0.90	0.92
Yolo-V2	42%	8	40%	0.38	0.40
Tiny-Yolo	28%	34	30%	0.26	0.29
Mask R-CNN	30%	4	35%	0.31	0.35
SSD 300	81%	10	84%	0.81	0.84

false positives by building a function that adjusts both the confidence threshold for object detection and IoU thresholds for duplicate objects. By optimizing both values at once, false positives, derived from the same pistol or suspicious objects with a pistol look, were minimized.

6 Conclusions

Considering the impact of MMW and x-ray technologies, detection of concealed weapons using thermal imaging is the least harmful solution to human health. Additionally, with the support of deep-learning techniques it provides detection capability without constant human observation. This work offered a new framework to detect concealed pistols using thermal images by combining two deep learning models. The proposed two-step solution can be purely used for only detection, or for both localization and detection based on the system capacity. For this purpose, we have compared various deep learning models using the suitable evaluation metrics for the related task. Our experiment results revealed that VGG19 and Yolo V3 models have demonstrated the best performance in terms of concealed pistol classification and localization. While the fine-tuned VGG19 model achieved the best classification accuracy results (85%) compared to other deep learning models, Yolo V3 outperformed other compared object detection algorithms with a significant improvement in MAP (95%). Moreover, Yolo V3 achieved to detect concealed pistols in 10 milliseconds which is an applicable time for real-time surveillance.

As the processing of thermal images highly depends on heat signature of the target object or human, one of the challenges was the detection of concealed objects under thick clothing. Besides, clothing with thermal features that do not reflect heat waves outside also make the thermal camera impossible to discern hidden objects. Another potential problem is that if the subject in the photo has multiple layers of clothes, the heat signature of the hidden metal object may be absorbed and not be emitted to the thermal camera. In order to overcome these problems, we evaluated different preprocessing approaches. In the experiments, we took RGB and thermal images of the same scene and we used these two images to produce improved images for detection. As a result RGB and thermal images of the same scene were fused with Discrete Wavelet Transform (DWT). In this trial, this approach did not result in suitable images for the further steps. Then, we came up with a preprocessing layer of sharpening and contrast to clarify the potential segments of the images with concealed pistols. This method achieved better results which can be also detected by the human eye. The methods for preprocessing the thermal images could be proliferated and different approaches could be applied to overcome and detect well-hidden objects. At this point of the study, the proposed model is suggested for indoor usage where layer of

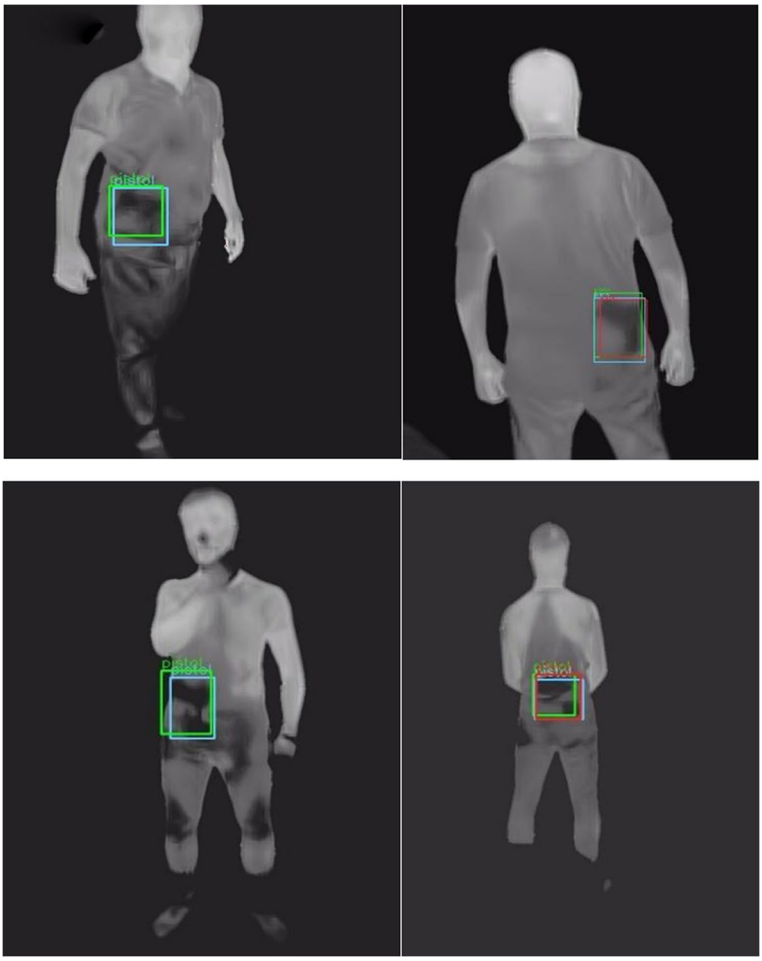

Fig. 7 Sample images processed with Yolo V3 by the pistol location detection module of the proposed system

clothing would be minimized and thermal cameras with higher wavelength spectrum can be utilized for better results. In our experiments, we took photos of the models wearing t-shirt, uniform or thin jackets as outer layer concealing the pistol. We did not evaluate the performance of the model under thick/insulated type of clothing. In the future, we plan to include insulated layers of clothing in our experiments and use a higher spectrum thermal camera to achieve better results under this type of clothing. Moreover, with the aim of improving the performance of the overall solution, we aim to build a feedback architecture between the detection and segmentation modules designed for the proposed solution.

One other important point to consider in concealed firearm detection is the type of firearm frame. Metal frames have heat signatures which are easier to detect compared to the body carrying the firearm. On the contrary, polymer pistol frames are harder to detect on the body. In our study, we used both polymer and metal framed (aluminum and steel framed) pistols in both training and test sets so that the deep learning models can detect

both frame types. The temperature of the body or the body segment where the pistol is hidden is also an important factor in the capability of the detection system. If the gun has been fired recently, the barrel and possibly the slide will have a higher temperature closer to that of the body, causing the thermal camera to confuse the heatwaves of the body and pistol. Alloys and processing technologies used in gun barrels today allow them to cool quickly. Therefore, it is unlikely that a suspect will enter a security point with a hot-gun barrel. In future studies, our target is to work on new approaches to address these problems for improved detection of concealed pistols from thermal images.

Acknowledgements We would like to thank all institutions that have supported us in the data creation in this study.

Authors' contributions Ozan Veranyurt: Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Visualization C. Okan Sakar: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Validation, Project administration.

Data availability The dataset generated during this study is partially available at Figshare repository, https://figshare.com/articles/dataset/Concealed_Pistol_Detection_Dataset/20105600

Code availability Not applicable.

Declarations

Ethics approval (include appropriate approvals or waivers) This study was approved by the Scientific Research and Publication Ethics Committee at Bahcesehir University (ethics application number: E-20021704-604.02.01-7208).

Consent to participate (include appropriate statements) Not applicable.

Consent for publication (include appropriate statements) Not applicable.

Conflicts of interest/competing interests (include appropriate disclosures) The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Berg A, Ahlberg J, Felsberg M (2015) A thermal object tracking benchmark. In 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp 1–6). IEEE
2. Cartucho J, Ventura R, Veloso M (2018) Robust object recognition through symbiotic deep learning in mobile robots. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp 2336–2341). IEEE
3. Castillo A, Tabik S, Pérez F, Olmos R, Herrera F (2019) Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning. Neurocomputing 330:151–161
4. Cheng L, Ji Y, Li C, Liu X, Fang G (2022) Improved SSD network for fast concealed object detection and recognition in passive terahertz security images. Sci Rep 12(1):1–16
5. Fernández-Carrobles MM, Deniz O, Maroto F (2019) Gun and knife detection based on faster R-CNN for video surveillance. In Iberian Conference on Pattern Recognition and Image Analysis (pp 441–452). Springer, Cham
6. Goenka A, Sitara K (2022) Weapon detection from surveillance images using deep learning. In 2022 3rd International Conference for Emerging Technology (INCET) (pp 1–6). IEEE
7. González JLS, Zaccaro C, Álvarez-García JA, Morillo LMS, Caparrini FS (2020) Real-time gun detection in CCTV: an open problem. Neural Netw 132:297–308
8. Hussein NJ, Hu F, He F (2017) Multisensor of thermal and visual images to detect concealed weapon using harmony search image fusion approach. Pattern Recogn Lett 94:219–227

9. Jain H, Vikram A, Kashyap A, Jain A (2020) Weapon detection using artificial intelligence and deep learning for security applications. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp 193–198). IEEE
10. Kowalski M (2019) Hidden object detection and recognition in passive terahertz and mid-wavelength infrared. J Infrared Millim Terahertz Waves 40(11-12):1074–1091
11. Kowalski M, Kastek M, Piszczek M, Życzkowski M, Szustakowski M (2015) Harmless screening of humans for the detection of concealed objects. Saf Secur Eng VI 151:215–223
12. Lai J, Maples S (2017) Developing a real-time gun detection classifier. In Course: CS231n. Stanford University, Stanford, CA, USA
13. Lamas A, Tabik S, Montes AC, Pérez-Hernández F, García J, Olmos R, Herrera F (2022) Human pose estimation for mitigating false negatives in weapon detection in video-surveillance. Neurocomputing 489:488–503
14. Naresh K, RajKumar SS, Ganesh MS, Sai L (2018) An infrared image detecting system model to monitor human with weapon for controlling smuggling of sandalwood trees. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp 962–968). IEEE
15. National Research Council (1996) Airline passenger security screening: new technologies and implementation issues (Vol. 482, No. 1). National Academies Press
16. Olmos R, Tabik S, Herrera F (2018) Automatic handgun detection alarm in videos using deep learning. Neurocomputing 275:66–72
17. Palmero C, Clapés A, Bahnsen C, Møgelmose A, Moeslund TB, Escalera S (2016) Multi-modal RGB–depth–thermal human body segmentation. Int J Comput Vis 118(2):217–239
18. Pang L, Liu H, Chen Y, Miao J (2020) Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm. Sensors 20(6):1678
19. Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767
20. Vallez N, Velasco-Mata A, Deniz O (2021) Deep autoencoder for false positive reduction in handgun detection. Neural Comput Appl 33:5885–5895
21. Veranyurt O, Sakar CO (2020) An object detection method (Turkey Patent Application No. 2020, 14269). Turkey Patent and Trademark Agency
22. Verma GK, Dhillon A (2017) A handheld gun detection using faster r-cnn deep learning. In Proceedings of the 7th International Conference on Computer and Communication Technology (pp 84–88)
23. Wei Y, Liu X (2020) Dangerous goods detection based on transfer learning in X-ray images. Neural Comput Appl 32(12):8711–8724
24. Wu T, Rappaport TS, Collins CM (2015) The human body and millimeter-wave wireless communication systems: interactions and implications. In 2015 IEEE International Conference on Communications (ICC) (pp 2423–2429). IEEE
25. Yeom S, Lee DS, Son JY, Jung MK, Jang Y, Jung SW, Lee SJ (2011) Real-time outdoor concealed-object detection with passive millimeter wave imaging. Optics Exp 19(3):2530–2536
26. Yuenyong S, Hnoohom N, Wongpatikaseree K (2018) Automatic detection of knives in infrared images. In 2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON) (pp 65–68). IEEE
27. Zhang S, Wen L, Bian X, Lei Z, Li SZ (2018) Single-shot refinement neural network for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp 4203–4212)
28. Zhang D, Zhan J, Tan L, Gao Y, Župan R (2020) Comparison of two deep learning methods for ship target recognition with optical remotely sensed data. Neural Comput Appl 33:4639–4649