**1.Weapon Detection in Real-Time CCTV Videos Using Deep Learning**

URL : <https://ieeexplore.ieee.org/document/9353483>

Abstract :

Security and safety is a big concern for today’s modern world. For a country to be economically strong, it must ensure a safe and secure environment for investors and tourists. Having said that, Closed Circuit Television (CCTV) cameras are being used for surveillance and to monitor activities i.e. robberies but these cameras still require human supervision and intervention. We need a system that can automatically detect these illegal activities. Despite state-of-the-art deep learning algorithms, fast processing hardware, and advanced CCTV cameras, weapon detection in real-time is still a serious challenge. Observing angle differences, occlusions by the carrier of the firearm and persons around it further enhances the difficulty of the challenge. This work focuses on providing a secure place using CCTV footage as a source to detect harmful weapons by applying the state of the art open-source deep learning algorithms. We have implemented binary classification assuming pistol class as the reference class and relevant confusion objects inclusion concept is introduced to reduce false positives and false negatives. No standard dataset was available for real-time scenario so we made our own dataset by making weapon photos from our own camera, manually collected images from internet, extracted data from YouTube CCTV videos, through GitHub repositories, data by university of Granada and Internet Movies Firearms Database (IMFDB) imfdb.org. Two approaches are used i.e. sliding window/classification and region proposal/object detection. Some of the algorithms used are VGG16, Inception-V3, Inception-ResnetV2, SSDMobileNetV1, Faster-RCNN Inception-ResnetV2 (FRIRv2), YOLOv3, and YOLOv4. Precision and recall count the most rather than accuracy when object detection is performed so these entire algorithms were tested in terms of them. Yolov4 stands out best amongst all other algorithms and gave a F1-score of 91% along with a mean average precision of 91.73% higher than previously achieved.

**2. Systematic review on weapon detection in surveillance footage through deep learning**

URL : <https://www.sciencedirect.com/science/article/abs/pii/S1574013723000795>

Abstarct:

In recent years, the number of crimes with weapons has grown on a large scale worldwide, mainly in locations where enforcement is lacking or possessing weapons is legal. It is necessary to combat this type of criminal activity to identify criminal behavior early and allow police and [law enforcement agencies](https://www.sciencedirect.com/topics/computer-science/law-enforcement-agency) immediate action. Despite the human visual structure being highly evolved and able to process images quickly and accurately if an individual watches something very similar for a long time, there is a possibility of slowness and lack of attention. In addition, large surveillance systems with numerous equipment require a surveillance team, which increases the cost of operation. There are several solutions for automatic weapon detection based on computer vision; however, these have limited performance in challenging contexts. A [systematic review](https://www.sciencedirect.com/topics/psychology/systematic-review) of the current literature on deep learning-based weapon detection was conducted to identify the methods used, the main characteristics of the existing datasets, and the main problems in the area of automatic weapon detection. The most used models were the Faster R-CNN and the YOLO architecture. The use of realistic images and [synthetic data](https://www.sciencedirect.com/topics/computer-science/synthetic-data) showed improved performance. Several challenges were identified in weapon detection, such as poor [lighting conditions](https://www.sciencedirect.com/topics/computer-science/lighting-condition) and the difficulty of small weapon detection, the last being the most prominent. Finally, some future directions are outlined with a special focus on small weapon detection.

**3. Artificial intelligence and super-resolution for enhanced weapon detection in video surveillance**

**URL :** <https://www.sciencedirect.com/science/article/pii/S0952197624018426>

**Absract :**

The prevalence of crimes involving handguns and knives underscores the importance of early weapon detection. This, along with the spread of [video surveillance systems](https://www.sciencedirect.com/topics/computer-science/video-surveillance-system), boosted the development of automatic approaches for weapon detection from surveillance cameras. Despite the advancements from classical [computer vision](https://www.sciencedirect.com/topics/engineering/computervision) to [Deep Learning](https://www.sciencedirect.com/topics/engineering/deep-learning) (DL) techniques, accurately detecting weapons in real-time remains challenging due to their small size. Current DL methods, which attempt to mitigate this issue using complex detection architectures, are resource-intensive, resulting in high costs and energy usage, and hindering their deployment on efficient edge devices. This creates challenges in resource-limited environments, making these methods impractical for edge and real-time applications. To address these shortcomings, our work proposes YOLOSR, which integrates a You Only Look Once (YOLO) v8-small model with an Enhanced Deep [Super Resolution](https://www.sciencedirect.com/topics/computer-science/super-resolution) (EDSR)-based network using a shared [backbone](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/spine). During training, the [auxiliary](https://www.sciencedirect.com/topics/chemical-engineering/auxiliaries) Super Resolution (SR) helps in learning better features, which could benefit the weapon detection task. During inference, the SR branch is removed, keeping the detector’s [computational complexity](https://www.sciencedirect.com/topics/computer-science/computational-complexity) unchanged. The YOLOSR’s accuracy and efficiency were validated on our WeaponSense dataset and on a NVIDIA Jetson Nano, against other weapon detectors. The results exhibited that YOLOSR, compared to the state-of-the-art YOLOv8-small model, maintained the same computational complexity with 28.8 billion [floating point operations](https://www.sciencedirect.com/topics/computer-science/floating-point-operation) and on-device latency of 101 ms per image, while increasing the Average Precision by 10.2 percentage points. Thus, the YOLOSR emerges as an effective solution for real-time weapon detection in resource-constrained environments, achieving an optimal trade-off between efficiency and accuracy.

**4. Fighting against terrorism: A real-time CCTV autonomous weapons detection based on improved YOLO v4**

URL: <https://www.sciencedirect.com/science/article/abs/pii/S1051200422004079>

Abstract

In recent years, [deep learning](https://www.sciencedirect.com/topics/computer-science/deep-learning) has demonstrated tremendous potential in the real world. Object detection is a critical real-world task for [deep learning](https://www.sciencedirect.com/topics/engineering/deep-learning). You Only Look Once (YOLO) object detection model recognizes interesting regions in images with impressive accuracy and real-time performance. The objective of this paper is to apply object detection to the field of security and counter-terrorism. Individuals are protected from violence by recognizing and locating the guns on closed-circuit television (CCTV). This paper presents a real-time detection approach for CCTV autonomous weapons based on YOLO v4. For the characteristics of CCTV scenarios, we propose the YOLO v4 backbone with Spatial Cross Stage Partial-ResNet (SCSP-ResNet). Meanwhile, the receptive field [enhancement module](https://www.sciencedirect.com/topics/computer-science/enhancement-module) is shown to capture fine [semantic features](https://www.sciencedirect.com/topics/computer-science/semantic-feature) of high-dimensional small objects. The Fusion-PANet (F-PaNet) module has been used to fuse multi-scale information to improve the model's perceptive power on the region of interest. Furthermore, we merge synthetic and real-world datasets to comprehensively investigate the effects of [synthetic datasets](https://www.sciencedirect.com/topics/computer-science/synthetic-datasets) on detectors. The experimental results reveal that our suggested detection model improves mAP (mean Accuracy Precision) and inference time by 7.37% and 4.2%, respectively. The model's parameter is reduced by 0.349 BFLOP/s(billion [floating point operations](https://www.sciencedirect.com/topics/computer-science/floating-point-operation) per second). The proposed detector outperforms the [baseline model](https://www.sciencedirect.com/topics/computer-science/baseline-model) in terms of accuracy, real-time, and robustness.

**5. Human pose estimation for mitigating false negatives in weapon detection in video-surveillance**

URL : <https://www.sciencedirect.com/science/article/abs/pii/S0925231221019159>

Abstract:

Applying CNN-based object detection models to the task of weapon detection in video-surveillance is still producing a high number of [false negatives](https://www.sciencedirect.com/topics/computer-science/false-negative). In this context, most existing works focus on one type of weapons, mainly firearms, and improve the detection using different pre- and post-processing strategies. One interesting approach that has not been explored in depth yet is the exploitation of the human pose information for improving weapon detection. This paper proposes a top-down methodology that first determines the hand regions guided by the human pose estimation then analyzes those regions using a weapon detection model. For an optimal localization of each hand region, we defined a new factor, called Adaptive pose factor, that takes into account the distance of the body from the camera. Our experiments show that this top-down Weapon Detection over Pose Estimation (WeDePE) methodology is more robust than the alternative bottom-up approach and state-of-the art detection models in both indoor and outdoor video-surveillance scenarios.

**6. Automatic Detection of Weapons in Surveillance Cameras Using Efficient-Net**

URL : <https://www.sciencedirect.com/org/science/article/pii/S1546221822009948>

Abstarct:

The conventional Close circuit television (CCTV) cameras-based surveillance and control systems require human resource supervision. Almost all the criminal activities take place using weapons mostly a handheld gun, revolver, pistol, swords etc. Therefore, automatic weapons detection is a vital requirement now a day. The current research is concerned about the real-time detection of weapons for the surveillance cameras with an implementation of weapon detection using Efficient–Net. Real time datasets, from local surveillance department's test sessions are used for model training and testing. Datasets consist of local environment images and videos from different type and resolution cameras that minimize the idealism. This research also contributes in the making of Efficient-Net that is experimented and results in a positive dimension. The results are also been represented in graphs and in calculations for the representation of results during training and results after training are also shown to represent our research contribution. Efficient-Net algorithm gives better results than existing algorithms. By using Efficient-Net algorithms the accuracy achieved 98.12% when epochs increase as compared to other algorithms.

**7. A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods**

URL : <https://www.sciencedirect.com/science/article/abs/pii/S0957417422017286>

Abstract

Surveillance systems do not give a rapid response to deal with suspicious activities such as armed robbery in public places. Consequently, there is a need for technology that can recognize criminal activities from Closed Circuit Televisions (CCTV) footage without the need of human help. Various high-performance computing algorithms have been developed but are limited to specific conditions. In this paper, we have identified gaps between existing technologies for weapon detection. The automatic detection of guns/weapons could help in the investigation of crime scenes. A new and difficult area of study is identifying the specific type of firearm used in an attack known as intra-class detection. The study examines and classifies the [strengths](https://www.sciencedirect.com/topics/materials-science/mechanical-strength) and shortcomings of several existing algorithms using classical [machine learning](https://www.sciencedirect.com/topics/computer-science/machine-learning) and [deep learning](https://www.sciencedirect.com/topics/computer-science/deep-learning) approaches, employed in the detection of different kinds of weapons. We have thoroughly compare and analyze the performance of several recent state-of-the-art methods on different datasets along with their future scope. We observed that [deep learning](https://www.sciencedirect.com/topics/engineering/deep-learning) techniques beat traditional [machine learning techniques](https://www.sciencedirect.com/topics/computer-science/machine-learning-technique) in terms of speed and accuracy.

**8. Improving handgun detection through a combination of visual features and body pose-based data**

URL: <https://www.sciencedirect.com/science/article/pii/S0031320322007312>

Abstract

Early detection of the presence of dangerous objects such as handguns in Closed-Circuit Television (CCTV) images is vital to reduce the potential damage. In this work, a novel method for automatic detection of handguns in CCTV-like images based on a combination architecture which leverages body pose estimation is proposed. Weapon appearance features along with body pose features are combined to perform robust detection in typical surveillance environments where appearance features alone are not sufficient (e.g., because the handgun may appear too small or dark). Both [CNN](https://www.sciencedirect.com/topics/computer-science/convolutional-neural-network) and recent transformer-based architectures are applied for visual feature extraction. Experiments on multiple datasets show that this approach improves state-of-the-art pose-based handgun detectors. An ablation study is also performed to verify the contribution of the pose processing branch and the false positive filter.

**9. A dual-stage deep learning framework for simultaneous fire and firearm detection in smart surveillance systems**

URL : <https://www.sciencedirect.com/science/article/pii/S2590123025024028>

Highlights

•A dual-stage deep learning framework enables unified, real-time detection of fire and firearms in smart surveillance footage.

•The Unified Threat Classification Network (UTCN) combines multiple CNNs for dynamic threat-specific routing to detection branches.

•Lightweight YOLOv5n-based detectors with specialized backbones deliver 97.1 % detection accuracy with reduced false positives.

•A Multi-Frame Confidence Evaluator (MFCE) enhances detection reliability by aggregating scores over consecutive frames.

•Real-world CCTV tests and a proof-of-concept domain adaptation loop show improved robustness under unconstrained conditions.

•Edge deployment on NVIDIA Jetson Nano confirms 17.6 FPS inference speed, low memory usage, and resilience under frame loss.

Abstract

Traditional video surveillance systems often treat fire detection and firearm recognition as separate tasks, missing the opportunity to address multiple security threats in an integrated manner. This paper presents a novel dual-stage deep learning framework for real-time, unified detection of fire and firearms in smart surveillance environments. At its core, a Unified Threat Classification Network (UTCN) dynamically routes frames to lightweight YOLOv5n-based fire and firearm detectors, with a Multi-Frame Confidence Evaluator (MFCE) verifying detection consistency to reduce false positives and false negatives. Beyond extensive benchmark testing, the system was validated on unconstrained real CCTV footage and enhanced with a proof-of-concept domain adaptation loop that fine-tunes detection modules using few-shot real-world samples. The proposed framework achieved up to 97.1 % detection accuracy and demonstrated improved robustness in real deployment scenarios, confirming its suitability for scalable, edge-deployable smart surveillance systems.