Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230213

Deploying human body detection technologies in security systems: An in-depth study of the FASTER-GCNN algorithm

Chao Jiang

UWEC computer science college, University of Wisconsin Eau-Claire, Eau-Claire, 54701-6698, United States

jiangc0802@uwec.edu

Abstract. The field of human body detection, a pivotal area in computer vision, merits comprehensive discussion. Remarkable advancements have been achieved in the techniques for human body detection over the past few decades, with significant applications spanning various sectors. This discussion delves into the potential of human detection technology within the realm of security - a field that necessitates efficient and accurate human detection technology to promptly identify potential threats, suspicious behaviors, or unusual activities. Deep learning-based human detection algorithms have substantially improved capabilities in this domain, facilitating real-time tracking and identification of the human form, thereby enabling security personnel to respond swiftly. This paper employs the Faster-RCNN algorithm for model training, utilizing the Information and Automation Research (INRIA) database. The deep learning-trained model proves highly accurate in human body detection, effectively recognizing human movements and behaviors. Such capabilities hold immense potential for implementation within the security sphere, including video surveillance systems and other similar applications where effectiveness is crucial.

Keywords: Faster-RCNN, INRIA, deep-learning, human-detection.

1. Introduction

The function of the security sector primarily lies in safeguarding and preserving public safety, and mitigating security risks. In the contemporary era, security technologies and measures exert significant influences and implications on society, corporations, and individuals alike. The key roles in the realm of security encompass crime prevention through mechanisms such as video surveillance and intrusion detection. These measures help deter criminal activities, decrease the crime rate, and uphold societal security. They also protect property, guarantee public safety, facilitate monitoring and management, and enable timely response to emergencies. Incorporating human detection can sufficiently meet the demands of the security sector. This technology can identify potential threats, suspicious conduct, or anomalous activities, subsequently issuing a preliminary warning. In doing so, it aids in preventing criminal occurrences. Human detection technology can facilitate real-time monitoring and automatic alarms, swiftly initiating responsive and remedial actions. This significantly shortens the response duration to security incidents, enhances handling efficiency, and minimizes potential losses. Moreover, it amplifies the safety levels of public areas, transportation hubs, and crucial facilities, thereby ensuring the security of citizens' and tourists' lives and properties. Additionally, this technology aids in maintaining societal order, thwarting terrorist assaults, and responding to emergencies.

Surveillance videos, images, and data obtained can serve as vital clues and evidence, assisting investigators in solving cases. This technology aids law enforcement in tracking suspects, identifying vehicles implicated in crimes, or observing other significant details. The application of human body detection technology in the security sector elevates public awareness and concern about security matters. By raising consciousness about security risks, this can stimulate individuals and organizations to take anticipatory security actions and contribute to collective efforts to maintain societal safety.

2. Background

Through research and surveys, the field of security is often overlooked for the vast majority of people, with technology and applications investing relatively little in security, and the deployment and maintenance of security systems usually requiring considerable investment [1]. For organizations or individuals with limited budgets, other urgent needs may take precedence and security may be put on the back burner [2]. Some regions or organizations may have a low perception of security risks and perceive the likelihood of a security incident as low, so that security measures are not seen as an urgent need. Some specific scenarios where the security field may not be able to provide a solution to meet a specific need due to technological limitations or immaturity, resulting in it not being a preferred consideration [3]. However, the importance of the security field is unquestionable and plays a major role in preventing accidents and disasters, maintaining social order, and protecting lives and property, and with the development of society, security is increasingly emphasized. Various security threats and criminal activities pose a threat to public safety and personal property, so the study of security technology and measures is an important way to meet the needs of society, so the field of security is worth being studied, the human body detection technology in the field of security will have a very good progress and the future of this paper, through the training of the model for the model and the deep learning, the use of Faster-RCNN algorithm is applied to detectron2 carrier, so that the human body detection is more accurate and has good results in the field of security [4].

3. The FASTER-GCNN algorithm

Detectron2 is renowned for its exceptional performance in the field of object detection and segmentation. It is a library built on top of the PyTorch deep learning framework, taking full advantage of the computational power of Graphics Processing Units (GPUs). Due to its high level of optimization and parallelization, Detectron2 enables efficient training and inference on large-scale datasets, offering fast and accurate results [5]. This library also provides numerous pre-trained models and model libraries, including traditional object detection models like Faster Region-Convolutional Neural Networks and RetinaNet, as well as instance segmentation models such as Mask Region-Convolutional Neural Networks. These pre-trained models can be employed as starting points, enabling users to rapidly construct and train high-performance models of their own.

The Faster Region-Convolutional Neural Networks is an advanced deep learning algorithm for object detection. Proposed in 2015 by Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun et al., this model improves upon its predecessors, Region-based Convolutional Neural Networks and Fast Region-Convolutional Neural Networks, by introducing a Region Proposal Network. This innovation facilitates end-to-end object detection, significantly enhancing detection speed [6]. The functioning of the Faster Region-Convolutional Neural Networks algorithm involves several stages, As shown in Figure 1:

• An input image is processed by a convolutional neural network to produce a feature map.

• A Region Proposal Network generates potential object regions using a sliding window approach. For each region, it predicts the probability of the presence of an object and the offset from the anchor box to the object's bounding box.

• These proposed object regions are then filtered using a technique called non-maximum suppression. This process eliminates overlapping regions, retaining only those with the highest confidence.

• The remaining candidate object regions are passed to a detection network that utilizes Region of Interest pooling to extract fixed-size feature maps.

• The detection network has two components: the classification branch predicts the likelihood of each object class within the candidate regions, while the regression branch estimates the offsets to adjust the bounding box coordinates.

• Lastly, non-maximum suppression is applied to the candidate regions after the classification and regression phases, yielding the final set of object detections.

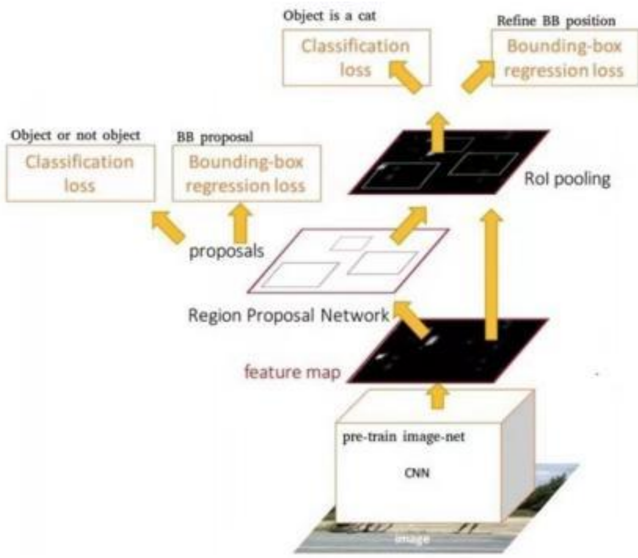

Figure 1. The architecture of Faster-RCNN (Photo/Picture credit: Original).

Fast R-CNN. RPN generates candidate boxes for objects, and Fast R-CNN extracts features from these boxes and classifies them [7]. The RPN uses a sliding window approach to generate these regions, and for each region, it predicts the probability of containing an object and the offset from the anchor box to the object's bounding box. After non-maximum suppression, the remaining candidate object regions are passed to the detection network. The detection network uses RoI (Region of Interest) pooling to extract fixed-size feature maps from the candidate regions. Detection Network: The classification branch predicts the probability of each candidate region containing each object class, and the regression branch predicts the offsets to adjust the bounding box coordinates of each candidate region [8]. The red box A represents the box before unregression, the blue box G' represents the box after regression. Each frame is represented as (x,y,h,w) and h,w denote the height and width of the frame [9].

(1) Do the panning first:

$$Gx' = Aw * dx(A) + Ax \tag{1}$$

$$Gy' = Ah * dx(A) + Ay \tag{2}$$

(2) Do the scaling again:

$$Gh' = Aw * dh(A) + Ah \tag{3}$$

$$Gw' = Ah * dw(A) + Aw \tag{4}$$

The parameters to be learned at this point are dh(A), dw(A), dx(A), dy(A). The corresponding learning objectives can be defined as follows.

$$t_x = \frac{(G_x - A_x)}{A_x} \tag{5}$$

DOI: 10.54254/2755-2721/32/20230213

$$t_y = \frac{(G_y - A_y)}{A_y} \tag{6}$$

$$t_h = \log\frac{G_h}{A_h} \tag{7}$$

$$t_w = \log\frac{G_w}{A_w} \tag{8}$$

A linear regression model can be used to define the learning process for d*(A): d∗(A)=W∗T∗ϕ(A). The design loss function is: loss=i=1 ∑ 100 (t∗ i−W∗T∗ϕ(Ai)) The corresponding function optimization objective is: W∗ ‘ =argminw ∗(i = 1 ∑ 100(t∗ i−W∗T∗ϕ(Ai)) + ｜ ｜λ W∗T ｜ ｜) W* can be solved by a stochastic gradient descent algorithm.

4. The database

The INRIA database is a computer vision-related database resource maintained by the computer vision team at the French National Institute for Information and Automation Research for target detection and pedestrian detection tasks. The INRIA Person Dataset contains a number of positive and negative sample images. Positive sample images contain labeled pedestrian targets, while negative sample images have no pedestrian targets. The dataset has a total of about 2,500 images containing about 1,200 positive samples (pedestrian targets) and about 1,300 negative samples (non-pedestrian targets) [10]. The images in the dataset are of high resolution and good quality, covering a wide range of scenes and complexities. Each positive sample image is labeled with the location and bounding box of the pedestrian target and is used to train and evaluate the performance of the pedestrian detection algorithm.

5. Experimental analyses

Following an in-depth comparative analysis of the Faster R-CNN and HOG+SVM algorithms, it is discerned that the former proves more precise in human body detection. Faster R-CNN is an advanced deep learning model for target detection, which leverages a Region Proposal Network (RPN) to facilitate an end-to-end detection process.

Historically, HOG and SVM have been employed as a classic algorithm pairing for target detection. In this combination, Histogram of Oriented Gradients extracts the image's features, while Support Vector Machine categorizes the targets from non-targets. The procedure includes several stages: image pre-processing to grayscale to eliminate illumination effects; gradient calculation at each pixel point; division into cells for gradient direction assignment; normalization of each cell's gradient direction histogram for shadow and lighting robustness; and, if color is considered, the inclusion of a color histogram in the HOG feature. The culmination of these steps is a feature vector formed by concatenating all cell gradient histograms.

Support Vector Machine is a supervised learning algorithm for binary classification. In the HOG+SVM target detection model, SVM segregates the target from non-target regions. The process initiates with preparing labeled image samples and extracting their HOG features, followed by labeling the HOG feature vectors for SVM training. The detection phase utilizes the trained SVM classifier to detect the image's target region by extracting HOG features via a sliding window and classifying them through SVM. The sliding window's size and step can be adjusted to the target size and detection accuracy. Ultimately, while HOG-SVM relies on machine learning, Faster R-CNN harnesses deep learning. By classifying the database images with the trained model, it is apparent that Faster R-CNN's recognition is superior. HOG+SVM, on the other hand, does not efficiently recognize potential targets, with some pedestrians escaping detection. This inefficiency is untenable in a security system demanding high safety factors. The superiority of Faster R-CNN in human body detection is further illustrated in a subsequent comparison of pedestrian detection using both models on the same set of images. As shown in Figure 2.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230213

Figure 2. A comparison of pedestrian detection on images using these two models (Photo/Picture credit: Original).

6. Conclusion

In essence, the Faster-RCNN screening and detection functionality proves highly efficient in security systems. Comparative studies with other algorithms reveal the precision and effectiveness of Faster-RCNN, making it an ideal choice for high-stake security applications. However, it is essential to recognize some inherent limitations in the use of Faster-RCNN within the security paradigm. Faster-RCNN is a sophisticated deep learning model, composed of a Region Proposal Network and a Target Detection Network, both of which demand substantial computational resources and considerable storage space. This proves difficult for embedded devices with limited resource capabilities. In security situations, the potential subjects may range in size from small (like pedestrians) to significantly large (like vehicles). Faster-RCNN employs predefined anchor frames for target detection, but these frames may not cover all target scales, resulting in under-detection of small objects and misdetection of larger ones. Frequently, multiple targets may overlap significantly, as seen with crowded pedestrians. In such scenarios, Faster-RCNN could yield redundant detection results, causing inaccuracies in target counts or incomplete detection frames. Speed is crucial in security scenarios, but Faster-RCNN's detection rate on large-scale images can be slow, making it unsuitable for high real-time performance requirements. Complex security situations may present targets from varied viewpoints or obscured by other objects. These conditions may lead to the Faster-RCNN failing to detect targets accurately, resulting in false detections. Additionally, Faster-RCNN requires extensive, well-labeled training data for optimal detection results, a task that can be costly and time-consuming within the security sector. Tailoring the model to specific security situations is also necessary to ensure real-world application success, representing yet another limitation. Nevertheless, future improvements to complement and enhance the security system could include the deployment of lighter weight target detection algorithms, model optimization for specific scenarios, or the incorporation of other sensors and technologies to bolster detection accuracy and real-time performance. Emerging trends suggest that the integration of reinforcement learning into target detection and security will gradually become more commonplace. By learning through environmental interaction, security systems can adaptively optimize detection strategies, delivering superior performance in complex environments. As target detection technology expands, concerns around data privacy and security will also rise. Future advancements will prioritize creating intelligent, efficient, comprehensive, and secure target detection systems while preserving user privacy and data security. The evolution of target detection technologies signals a move towards more intelligent, efficient, comprehensive, and secure solutions. These advancements will profoundly influence the development of security systems and other fields, resulting in improved safety and convenience for society at large.

References

[1] Hung G L , Sahimi M S B , Samma H ,et al.Faster R-CNN Deep Learning Model for Pedestrian Detection from Drone Images[J].SN Computer Science, 2020, 1(2):116.
[2] Albinali H , Alzahrani F A .Faster R-CNN for Detecting Regions in Human-Annotated Micrograph Images[C]//2021 International Conference of Women in Data Science at Taif University (WiDSTaif).2021.
[3] Lukac Y N .Pedestrian Detection based on Faster R-CNN[J].International Journal of Performability Engineering, 2019, 15(7).
[4] Baussard A , D'Acremont A , Quin G ,et al.Faster-RCNN with a compact CNN backbone for target detection in infrared images[C]//Conference on Artificial Intelligence and Machine Learning in Defense Applications.2020.
[5] Panigrahi S , Raju U S N .Pedestrian Detection Based on Hand-crafted Features and Multi-layer Feature Fused-ResNet Model[J].International Journal on Artificial Intelligence Tools, 2021.
[6] Zhao R , Li C , Ye S ,et al.Butterfly Recognition Based on Faster R-CNN[C]//IOP Publishing.IOP Publishing, 2019:032048-.DOI:10.1088/1742-6596/1176/3/032048.
[7] Gao F , Fu L , Zhang X ,et al.Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN[J].Computers and Electronics in Agriculture, 2020(176-):176.
[8] Kihara H , Ikejiri K , Ishizaki S ,et al.A Study of the Suitable Electrode Position on the Head for High Accuracy R-peak Detection from Electrocardiogram[J].The Proceedings of the Symposium on sports and human dynamics, 2019:A-29.DOI:10.1299/jsmeshd.2019.A-29.
[9] Huang X , Li X , Hu Z .Cow tail detection method for body condition score using Faster R-CNN[C]//2019 International Conference on Unmanned Systems and Artificial Intelligence (ICUSAI).2019.
[10] Zhang W , Mi Z , Zou Y ,et al.Joint HFaster-RCNN and Bayesian Posterior Probabilities for Pedestrian Detection[C]//2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE).IEEE, 2019.