information

MDPI

Article

A Novel Low Processing Time System for Criminal Activities Detection Applied to Command and Control Citizen Security Centers

Julio Suarez-Paez [1,*], Mayra Salcedo-Gonzalez [1], Alfonso Climente [1], Manuel Esteve [1], Jon Ander Gómez [2], Carlos Enrique Palau [1] and Israel Pérez-Llopis [1]

[1] Distributed Real-time Systems Laboratory (SATRD), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; maysalgo@doctor.upv.es (M.S.-G.); alclial@upvnet.upv.es (A.C.), mesteve@dcom.upv.es (M.E.), cpalau@dcom.upv.es (C.E.P.), ispello0@upvnet.upv.es (I.P.-L.)

[2] Pattern Recognition and Human Language Technologies, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; jon@dsic.upv.es

* Correspondence: juliosuarez@ieee.org or jusuapae@doctor.upv.es

Received: 16 October 2019; Accepted: 20 November 2019; Published: 24 November 2019

Abstract: This paper shows a Novel Low Processing Time System focused on criminal activities detection based on real-time video analysis applied to Command and Control Citizen Security Centers. This system was applied to the detection and classification of criminal events in a real-time video surveillance subsystem in the Command and Control Citizen Security Center of the Colombian National Police. It was developed using a novel application of Deep Learning, specifically a Faster Region-Based Convolutional Network (R-CNN) for the detection of criminal activities treated as "objects" to be detected in real-time video. In order to maximize the system efficiency and reduce the processing time of each video frame, the pretrained CNN (Convolutional Neural Network) model AlexNet was used and the fine training was carried out with a dataset built for this project, formed by objects commonly used in criminal activities such as short firearms and bladed weapons. In addition, the system was trained for street theft detection. The system can generate alarms when detecting street theft, short firearms and bladed weapons, improving situational awareness and facilitating strategic decision making in the Command and Control Citizen Security Center of the Colombian National Police.

Keywords: Command and Control Citizen Security Center; Command and Control Information System (C2IS); crime detection; homeland security

1. Introduction

Colombia is a country with approximately 49 million inhabitants, 77% of which live in cities [1], and as in many Latin American countries, some Colombian cities suffer from insecurity. To face this situation and guarantee the country's sovereignty, the Colombian government has public security forces formed by the National Army, the National Navy and the Air Force, which have the responsibility to secure the borders of the country as well as ensure its sovereignty. Additionally, the Colombian National Police has the responsibility of security in the cities and of fighting against crime.

To ensure citizen security, the Colombian National Police has a force of 180,000 police officers, deployed across the national territory and several technological tools, such as Command and Control Information Systems (C2IS) [2,3] that centralize all the strategic information in real time, improving *situational awareness* [2,3] for making strategic decisions [3,4], such as the location of police officers and mobility of motorized units.

The C2IS centralizes the information in a physical place called the Command and Control Citizen Security Center (in Spanish: Centro de Comando y Control de Seguridad Ciudadana), where under a strict command line, the information is received by the C2IS operators and transmitted to the commanders of the National Police to make the most relevant operative decisions in the shortest time possible (Figure 1).

Figure 1. Command and Control Citizen Security Center, Colombian National Police.

The C2IS shows georeferenced information using a Geographic Information System (GIS) of several subsystems [5], such as crime cases reported by emergency calls, the position of the police officers in the streets and real-time video from the video surveillance system [6].

However, this technological system has a weakness in the Video Surveillance Subsystem because of the discrepancy between the number of security cameras in the Colombian cities and the system operators, which hinders the detection of criminal events. In other words, there are many more cameras than system operators can handle, meaning that the video information arrives at the Command and Control Citizen Security Center but it cannot be processed fast enough by the police commanders, and as such, they cannot take the necessary tactical decisions.

Bearing this in mind, this paper shows a Low Processing Time System focused on criminal activities detection based on real-time video analysis applied to a Command and Control Citizen Security Center. This system uses a novel method for detecting criminal actions, which applies an object detector based on Faster Region-Based Convolutional Network (R-CNN) as a detector of criminal actions. This innovative application of Faster R-CNN as a criminal action detector was achieved by training and adjusting the system for criminal activities detection using data extracted from the Command and Control Center of the Colombian National Police.

This novel method automates the detection of criminal events captured by the video surveillance subsystem, generating alarms that will be analyzed by the C2IS operators, improving situational awareness of the police commanders present at the Command and Control Citizen Security Center.

2. Related Work in Crime Events Video Detection

In computer vision, there are many techniques and applications which could be relevant for the operators of the C2IS of the National Police, for instance, the detection of pedestrians, the detection of trajectories, background and shadow removing [7], and facial biometrics.

There are already several approaches to detect crimes and violence in video analysis, as shown by [8–11]. However, the Colombian National Police does not implement any method for the specific case of the detection of criminal events. The available solutions are not applicable because most of the cameras of the video surveillance system installed in Colombian cities are mobile (*Pan–Tilt–Zoom Dome*), which makes it difficult to use conventional video analysis techniques focused on human action

recognition because most of these methods are based on trajectory [12–15] or movement analysis [16–18] and camera movements interfere with these kinds of studies.

Owing to this, we decided to explore innovative techniques independent of the abrupt movement of video cameras, which perform a frame-by-frame analysis without independence between video frames.

Bearing this in mind, we discarded all the techniques based on trajectory detection and used prediction filters or metadata included in the video files, focusing on techniques that could take advantage of hardware's capabilities for parallel processing. As such, the criminal events detection system was developed using Deep Learning techniques.

Taking into account the technological developments of recent years, Deep Learning has become the most relevant technology for video analysis and has an advantage over the other technologies analyzed for this project: each video frame is analyzed and processed independently of all the others without temporary interdependence, which makes Deep Learning perfect for video analysis from mobile cameras such as those used in this project.

To choose the Deep Learning Models, we studied factors such as the processing time of each video frame, accuracy and model robustness. Therefore, several detection techniques were studied, such as R-CNN (Region-Based Convolutional Network) [19], YOLO (You Only Look Once) [20], Fast R-CNN (Fast Region-Based Convolutional Network) [21,22] and Faster R-CNN (Faster Region-Based Convolutional Network) [23,24] (Table 1). After analyzing the advantages and disadvantages of each technique, Faster R-CNN was chosen to implement the system for criminal events detection in the system for the C2IS of the National Colombian Police due to the fact that it has an average timeout that was 250 times faster than *R-CNN* and 25 times faster than Fast R-CNN [22,25,26]. Furthermore, in recent work, models based on two stages like Faster R-CNN have had better accuracy and stability than models based on regression like YOLO [27,28] and SSD, which is of great importance because in this work, a novel application focused in action detection was given to an object detector model.

Table 1. Deep Learning object detection models relative comparison.

Object Detector Model	Average Accuracy	Average Processing Time	Model Deployment Level (Number of Works Related)
R-CNN	High	High	Medium
Fast R-CNN	High	Medium	Medium
Faster R-CNN	Very High	Very Low	Very High
SSD	Very High	Very Low	Very High
YOLO	Very High	Very Low	Very High

Analyzing real-time video frame-by-frame is a task with a very high computational cost. This is considerable taking into account the sheer amount of video cameras surveillance systems available in Colombian cities. Therefore, it is necessary that each video frame has a low computational cost and processing time to secure a future large-scale implementation.

With this in mind, several previous studies have been studied where real-time video is analyzed with security applications. Among these studies, one stands out [29], in which the authors performed video analysis from a video surveillance system using the Caffe Framework [30] and Nvidia cuDNN [31] without using a supercomputer. Another study that demonstrated the high performance of Faster R-CNN for video analysis in real time is [32], in which the video was processed at a rate of 110 frames per second. Another interesting study is [33], in which the authors made a system based on Faster R-CNN for the real-time detection of evidence in crime scenes. One last study to highlight is [34], in which the authors created an augmented reality based on Faster R-CNN implementation using a gaming laptop.

Other authors have carried out related relevant research, such as [35], in which fire smoke was detected from video sources; [36], which showed a fire detection system based on artificial intelligence; [37], which detected terrorist actions on videos; [38,39], that showed novel applications to

object detection; [40,41], that showed an excellent tracking applications; [42] in which a Real-Time video analysis was made from several sources with interesting results in object tracking; [43] which proposed a secure framework for IoT Systems Using Probabilistic Image Encryption; [44] which showed an Edge-Computing Video Analytics system deployed in Liverpool, Australia; [45] where GPUs and Deep Learning were used for traffic prediction; [46] where a video monitor and a radiation detector in nuclear accidents were shown; [47] where an Efficient IoT-based Sensor Big Data system was detailed.

In addition to these, recently, interesting applications of Faster R-CNN have also been published, for example in [48], a novel application of visual questions answering by parameter prediction using Faster R-CNN was presented, [49] showed a modification of Faster R-CNN for vehicles detection which improves detection performance, in [50], a face detection application was presented in low light conditions using two-step Faster R-CNN processing, first detecting bodies and then detecting faces, [51] showed an application to detect illicit objects such as fire weapons and knives, analyzing terahertz imaging using Faster R-CNN as an object detector and [52] showed a Faster R-CNN application for the detection of insulators in high-power electrical transmission networks.

As shown previously, Deep Learning includes a variety of techniques in computer vision, which are suitable for the development of this work.

3. Novel Low Computational Cost Method for Criminal Activities Detection Using One-Frame Processing Object Detector

In many cases, the detection and recognition of human actions (like criminal actions) is done by analysis of movement [16–18,53,54] or trajectories [12–15], which implies the processing of several video frames. Nevertheless, when the video camera is mobile, it is very difficult to carry out the trajectory or movement analysis because camera movements may introduce noise to the trajectories or movements to be analyzed. In addition, in a Smart City application, the number of cameras could be hundreds or thousands, so motion or trajectories analysis involves processing several video frames for each detection, which would multiply the computational cost of a possible solution. It is necessary to analyze mobile cameras with the minimum computational cost possible because, in the Command and Control Citizen Security Center, thousands of cameras are pan–tilt–zoom domes and this makes it very difficult to perform a motion or trajectory analysis to detect criminal activities. On the other hand, since there are thousands of cameras, the computational cost becomes an extreme relevant factor.

For this reason, hours of video of criminal activities were studied and it was noted that all criminal activities have a characteristic gesture, such as threatening someone; therefore, we set out to analyze this characteristic gesture as an "object" so that it could be detected using techniques that are independent of camera movements and process only one video frame.

With this in mind, we propose a novel system called "Video Detection and Classification System (VD&CS)" in which Faster R-CNN is used in a hybrid way to detect objects used in criminal actions and criminal characteristic gestures treated as "objects". Considering that criminal actions always have fixed gestures such as threatening the victim, it is possible to consider that this criminal action can be understood by the system as an "object". This novel application has the potential to reduce the computational cost because only one video frame will be processed, compared to other action detection methods that must analyze several video frames [12–18,53,54]. With this novel method in mind, we proceeded with the system design and training.

3.1. Video Detection and Classification System (VD&CS)

The system proposed is based on a Faster Region-Based Convolutional Network (Faster R-CNN), involves two main parts: a region proposal network (RPN) and a Fast R-CNN [23] and it was developed using Matlab.

3.1.1. Region Proposal Network

The RPN is composed of a classifier and a regressor, and its aim is to predict whether, in a certain image region, a detectable object will exist or will be part of the background, as is shown in [23].

Regions of interest comprise short firearms, bladed weapons and street thefts, which are criminal actions but will be treated as objects in the training process.

In this case, the pre-trained CNN model AlexNet [55] was used as the core of the RPN. This CNN model is made up of Convolution layers, ReLU, Cross Channel Normalization layers, Max Pool layers, Fully Connected layers and Softmax layers, as shown in Figure 2.

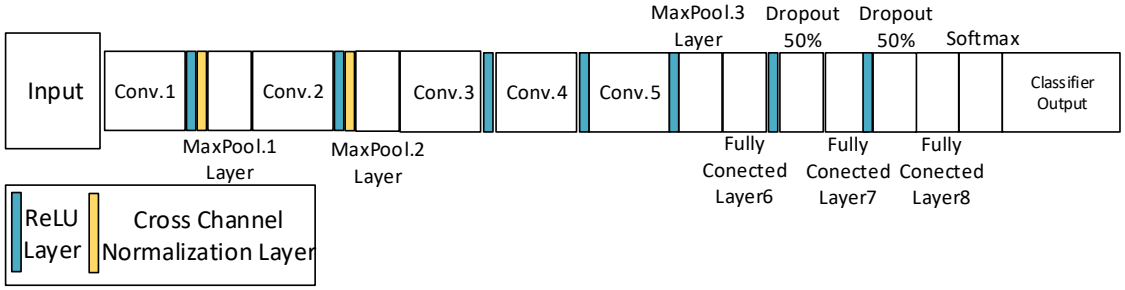

Figure 2. AlexNet Convolutional Neural Network Layers [55].

Figure 3 shows AlexNet used as RPN core. It has less layers than models like VGG16 [56], VGG19 [56], GoogleNet [57] or ResNet [58]. Hence, AlexNet has a lower computational cost and requires less processing time per video frame [22] (further implementation details are provided in Section 5).

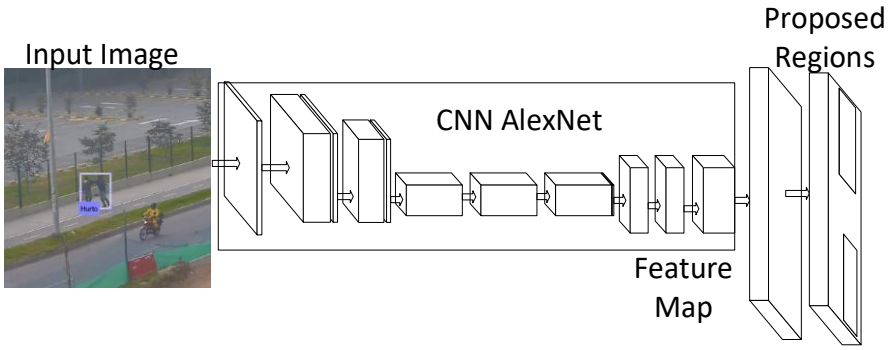

Figure 3. Video Detection and Classification System (VD&CS): Region Proposal Network (RPN).

3.1.2. Fast Region-Based Convolutional Network

Fast R-CNN acts as a detector that uses the region proposals made by the RPN and also uses AlexNet (Figure 2) as the CNN of the core model to detect regions of interest for the system, which are short firearms, bladed weapons and street thefts (Figure 4).

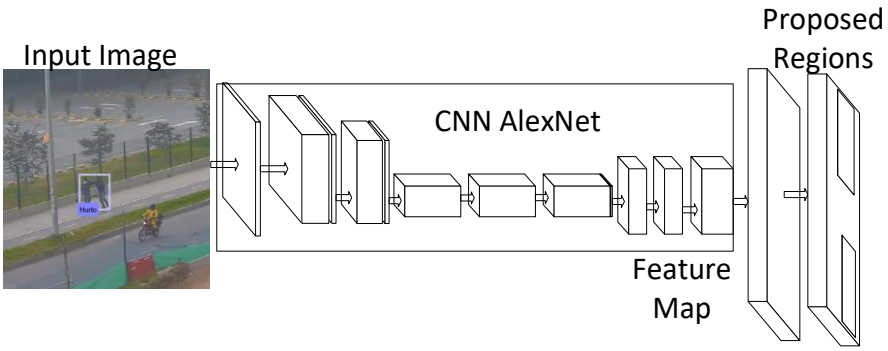

Figure 4. VD&CS: fast R-CNN.

3.1.3. Faster Region-Based Convolutional Network

Finally, RPN and Fast R-CNN are joined, forming a Faster R-CNN system (Figure 5) with the capacity of real-time video processing using AlexNet (Figure 2) as core.

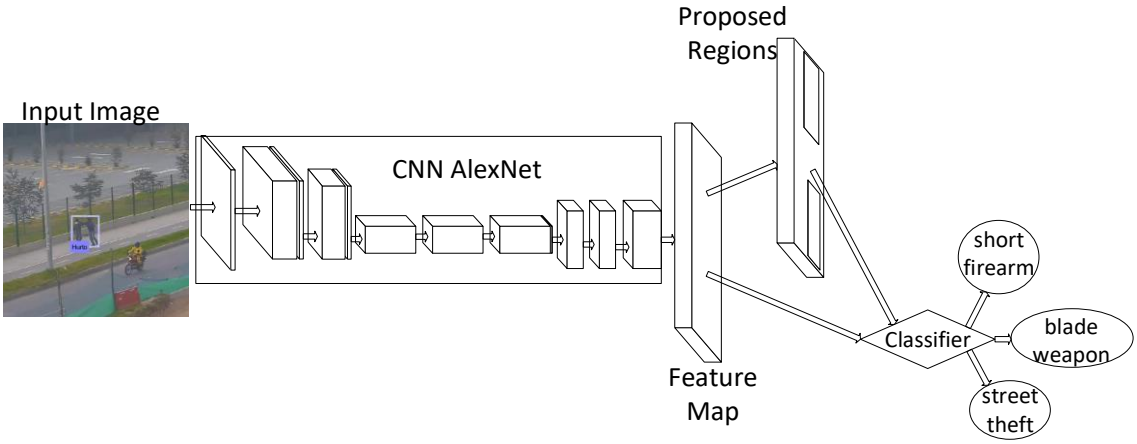

Figure 5. VD&CS: faster R-CNN.

3.2. VD&CS: Training Process

The system proposed based on Faster R-CNN, was trained using Matlab in a four-stage process, as outlined below.

3.2.1. Train RPN Initialized with AlexNet Using a New Dataset

At this stage, AlexNet, shown in Figure 2, is retrained inside the RPN, using transfer learning with a new dataset of 1124 images specially created to train the VD&CS (Figure 6). This dataset was created by manually analyzing several hours of video taken from the Command and Control Citizen Security Center and finding criminal actions to extract. The dataset has three classes of interest: short firearm, bladed weapons and street theft (action as object), and its bound boxes were manually marked for each image.

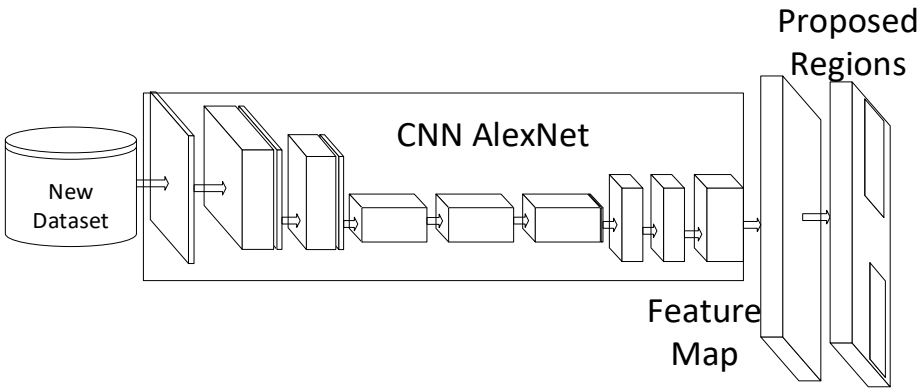

Figure 6. First stage: RPN training.

To improve the system performance, the training process data argumentation methods were used and as a result of the training procedure, in this stage, we obtained a feature map of the three classes mentioned above, from which the RPN is able to make proposals of possible regions of interest.

3.2.2. Train Fast R-CNN as a Detector Initialized with AlexNet Using the Region Proposal Extracted from the First Stage

In the second stage, a Fast R-CNN detector was trained using the initialized AlexNet as a starting point (Figure 7). The region proposals obtained by the RPN in the first stage were used as input to the Fast R-CNN to detect the three classes of interest.

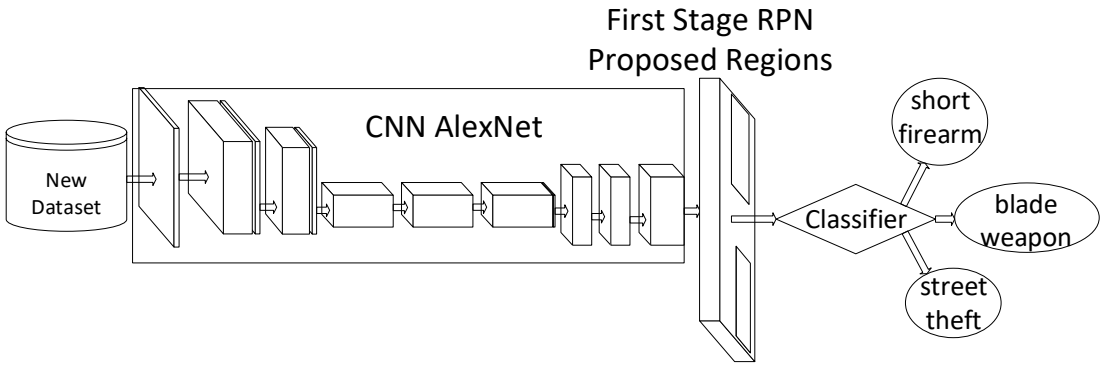

Figure 7. Second stage: Fast R-CNN training.

3.2.3. RPN Fine Training Using Weights Obtained with Fast R-CNN Trained in the Second Stage

With the objective of increasing the RPN success rate, in the third stage, fine training of the RPN that was trained in the first stage is carried out (Figure 8). In this case, weights obtained from the training procedure of the Fast R-CNN during the second stage were used as initial values.

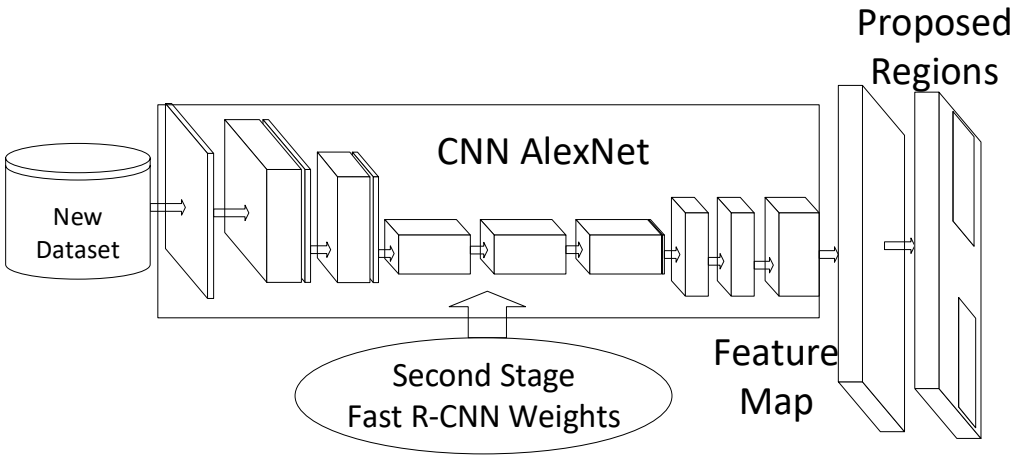

Figure 8. Third stage: RPN fine training.

3.2.4. Fast R-CNN Fine Training Using Updated RPN

To improve the accuracy of the Fast R-CNN trained in the second stage, in this last stage, fine training was carried out using the results of the third stage, as shown in Figure 9.

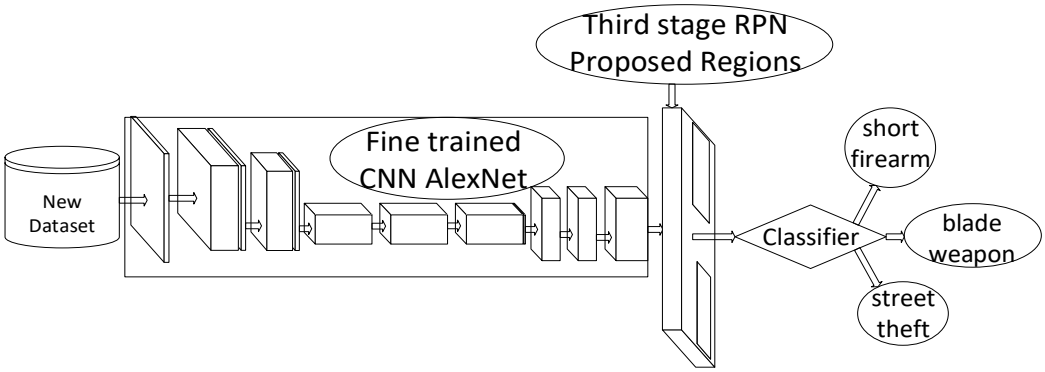

Figure 9. Fourth stage: Fast R-CNN fine training.

Finally, Figure 10 shows a system capable at generating alarms detecting short weapons, blade weapons and street theft by analyzing just one video frame, which would reduce the computational cost compared to models based on analysis of movement or trajectories.

Figure 10. VD&CS: criminal activities detection.

3.3. VD&CS: Testing

Once VD&CS was trained, its image processing time and accuracy were measured in order to evaluate its applicability to real scenarios of real-time video analysis. Two series of 500 images that were not used for training were used for testing using the same Hardware: MSI GT62VR-7RE with an Intel Core I7 7700HQ, 16 GB of DDR4 RAM, with a GPU NVIDIA GeForce GTX 1070 with 8 GB DDR5 VRAM). Table 2 shows the obtained results. The used performance indicators were the average processing time per frame, accuracy, undetected event rate, false positive rate and frame rate per second (FPS).

Table 2. VD&CS 500 image tests.

Item Tested	Results Test 1	Results Test 2
Crime Event Detections	355	367
Failures	145	133
Undetected	87	80
False positive	58	53
Average processing time	0.03 s	0.03 s
FPS (Frames per second)	33 FPS	33 FPS
Undetected event rate	17.4%	16%
False positive rate	11.6%	10.6%
Accuracy	71%	73.4%

According to previous results, the Confusion Matrix (Table 3) shows that VD&CS is useful for detecting criminal events in real-time video; its accuracy is within the parameters expected of a Faster R-CNN [23], taking into account that criminal actions were handled as objects within VD&CS, it confirms that VD&CS can be used for Criminal Activities Detection Applied to Command and Control Citizen Security Centers.

Table 3. VD&CS confusion matrix.

	Predictions	
Observations	49.6% (True Positive) 17.4% (False Negative)	11.6% (False Positive) 21.4% (True Negative)

3.3.1. Real-Time Video Testing

Training and tests with real-time video were performed on a laptop MSI GT62VR-7RE with an Intel Core I7 7700HQ, 16 GB of DDR4 RAM, with a GPU NVIDIA GeForce GTX 1070 with 8 GB DDR5 VRAM.

Real-time video testing consisted of two main video sources; the first source contained pre-recorded videos obtained from the Colombian National Police video surveillance system and the second source was a set of videos captured in real time by a laptop camera.

In these two scenarios, excellent results were obtained with respect to the processing time of each image, ranging from 0.03 to 0.05 s. This allows real-time video processing at a rate of 20 to 33 FPS, which is adequate considering the video sources of the C2IS of Colombian National Police.

Regarding the system accuracy, we checked that it is free of overtraining as the tests done on the system were performed with images not used in the training process and their results were confirmed in the confusion matrix and the system accuracy it is within the range expected for a Faster R-CNN; however, the system is designed to be used in public safety applications, so it always requires human monitoring because the detections depend on the lighting conditions and the distance of the cameras to the object, in addition to the success rate of the Faster R-CNN; additionally, in previous studies [22], authors evaluated other CNN models of a greater depth by choosing AlexNet for its performance and simplicity.

However, it achieves excellent results in terms of triggering alarms when it detects criminal events, improving situational awareness in the Command and Control Citizen Security Center of Colombian National Police.

3.3.2. Computational Cost Comparation

As previously stated, several detection and recognition of human actions techniques consist of movement or trajectories analysis. These techniques must analyze several video frames to be able to recognize actions, for example, in [59–61], sets of six to eight images are analyzed to identify actions.

In order to have computational cost low enough to be deployed in thousands of cameras, VD&CS just processes one video frame to detect criminal actions, which achieves a low computational cost that could be deployed in embedded systems or in cloud architecture, reducing high deployment costs.

To analyze the computational cost, different CNN models in the VD&CS core were compared with another action detection technique proposed in [59]. The results are shown in Table 4.

Table 4. VD&CS computational cost comparation.

Model	Average Processing Time	GPU	GPU Performance (Float 32)	Resolution (Pixels)
VD&CS (AlexNET)	0.03 s	Nvidia GTX 1070 MXM	6.738 TFLOPS	704 × 544
VD&CS (VGG-16)	0.23 s	Nvidia GTX 1070 MXM	6.738 TFLOPS	704 × 544
VD&CS (VGG-19)	0.28 s	Nvidia GTX 1070 MXM	6.738 TFLOPS	704 × 544
T-CNN	0.9 s	Nvidia GTX Titan X	6.691 TFLOPS	300 × 400

Therefore, assuming that GPUs have an equivalent performance and scaling the resolution of the video frames used in the tests, we consider deployments in cities like Bogotá where there are about 2880 Pan–Tilt–Zoom cameras (as of June 2019).

First, we analyzed computational cost measured TeraFlops and depict the variation of computational costs for processing of 2880 cameras (Figure 11).

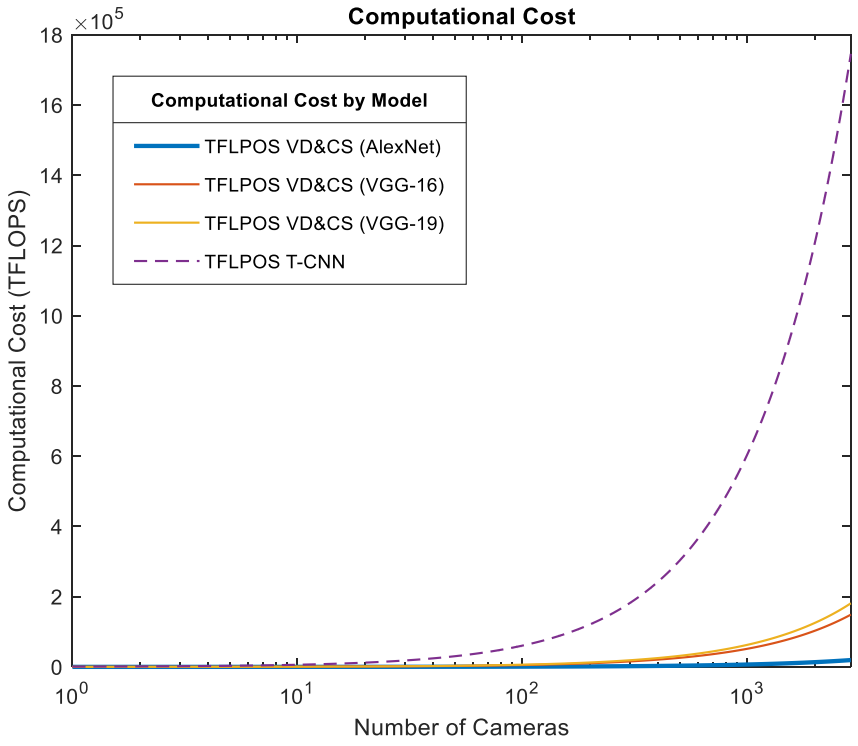

Figure 11. VD&CS low processing time system: computational cost comparation.

We also analyzed the Hardware cost and power consumption, assuming a deployment using Nvidia embedded systems [62] (Figures 12 and 13).

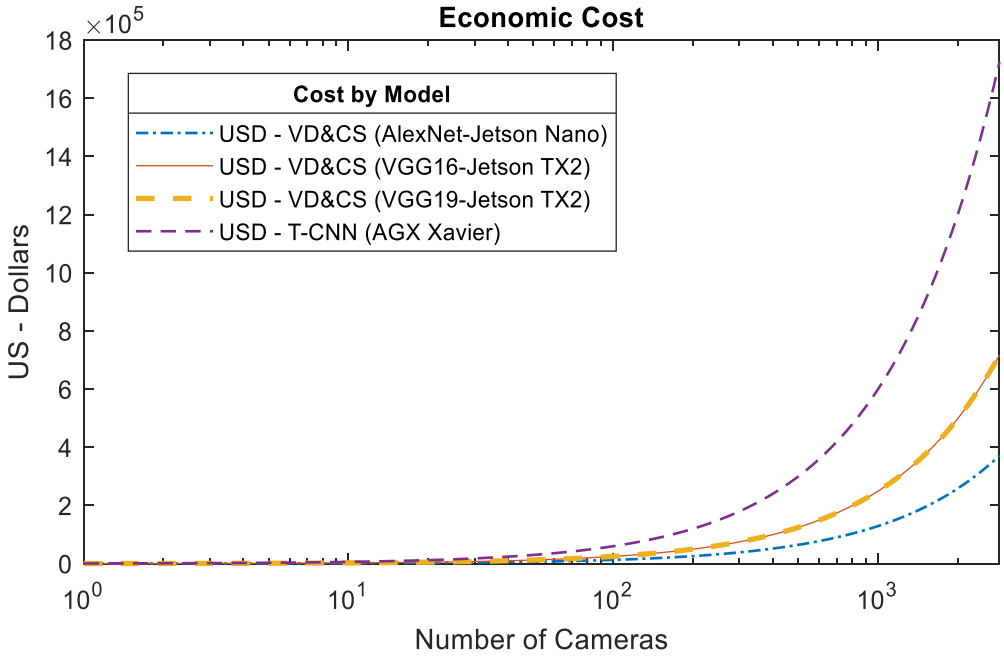

Figure 12. VD&CS Low Processing Time System: Economical Cost Comparation.

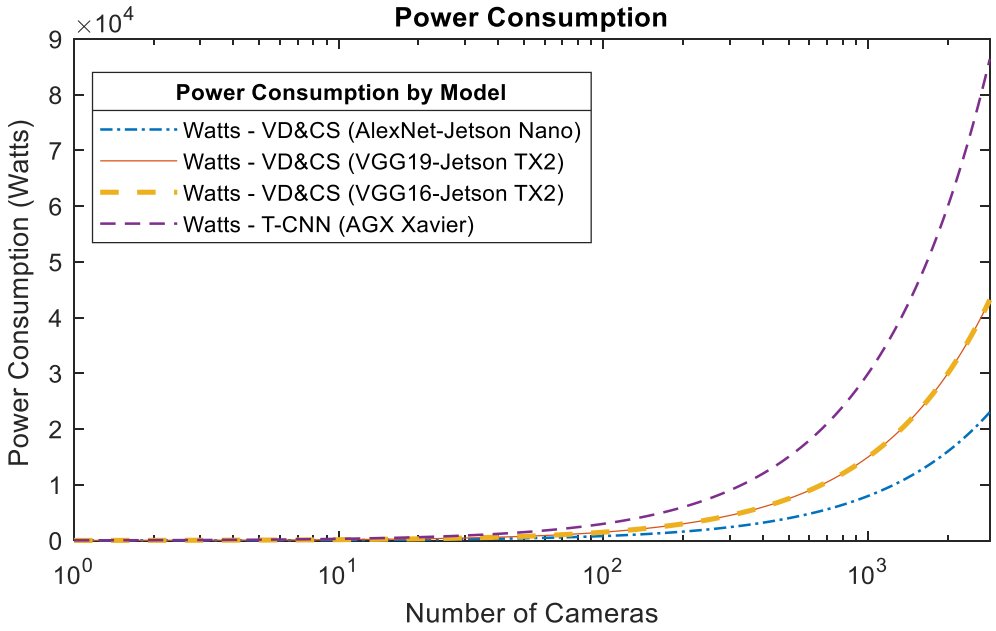

Figure 13. VD&CS Low Processing Time System: Power Consumption Comparation.

As Figures 11–13 show, having thousands of video sources in a Low Processing Time System, the computational cost is a factor of extreme relevance, since the economic and energy costs could make the implementation not feasible, and for this reason VD&CS proves be appropriate in a Low Cost System.

3.4. VD&CS: Final System

Once the process of training and testing are completed, we propose the system shown in Figure 14 to be applied in a larger city architecture.

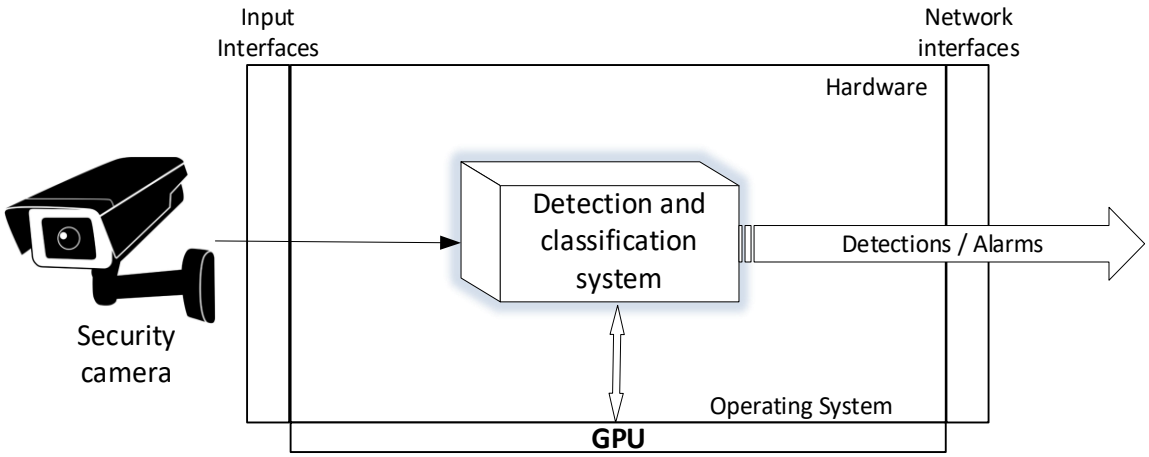

Figure 14. VD&CS: System.

In this approach, the VD&CS runs in an environment independent of the operating system because it can be implemented using any framework or library that supports Faster R-CNN, such as Caffe [30], cuDNN [31], TensorFlow [63], TensorRT [64], Nvidia DeepStream SDK [65], which uses real-time video coming from the security cameras and uses GPU computational power to run. Finally, the VD&CS uses network interfaces to send the generated alarms to the Command and Control Citizen Security Center.

This system is expected to be applied in different scenarios based on cloud architectures or embedded systems compatible with IoT (Internet of Things) solutions [66,67].

4. Low Processing Time System Applied to Colombian National Police Command and Control Citizen Security Center

To propose a Low Processing Time System to detect criminal activities based on a real-time video analysis applied to National Police of Command and Control Citizen Security Center, we must consider the Colombian Police Command and Control objectives, as detailed below:

Situational awareness: Police commanders must know in detail and real-time the situation of citizen security in the field, supported by technological tools to make the best tactical decisions and guarantee the success of police operations that ensure citizen security.

Situation understanding: Improving situational awareness by improving crime detection, allows police commanders to gain a better understanding of the situation, helping to detect more complex behaviors of criminal gangs.

Decision making-improvement: Decisions made in the Command and Control Citizen Security Center can be life or death because many criminal acts involve firearms and violent acts; therefore, the proposed system will improve decision making because it will provide real-time information to commanders, improving the effectiveness of police operations.

Agility and efficiency improvement: As mentioned above, decisions made by the police can mean life or death. Therefore, the improvement offered by the proposed prototype to the agility and efficiency of police operations relies on information that is unknown by commanders, impeding the deployment of police officers in critical situations.

4.1. Decentralized Low Processing Time System for Criminal Activities Detection based on Real-time Video Analysis Applied to the Colombian National Police Command and Control Citizen Security Center

The Command and Control Citizen Security Center is formed of subsystems such as the emergency call attention system (123), Police Cases Monitoring and Control Information System (SECAD), Video Surveillance Subsystem and the crisis and command room. Command and Control Citizen Security is supported by telecom networks that can be owned by the National Police or belong to the local ISP (Internet Service Provider).

These subsystems have different types of operators which are in charge of specific tasks such as monitoring the citizen security video (Operators Video Surveillance system), answering emergency calls (123 Operators) and assigning and monitoring field cop to police cases (Dispatchers).

Another important part of the Command and Control Citizen Security Center is the crisis and command room, in which the police commanders make strategic decisions according to their situational awareness and situation understanding [2].

In this decentralized system, the VD&CS will be implemented in embedded systems with GPU capability such as Nvidia Jetson [62] or AMD Embedded Radeon™ [68]. Then, it will be installed in each citizen video surveillance camera, detecting criminal activities locally (Figure 15).

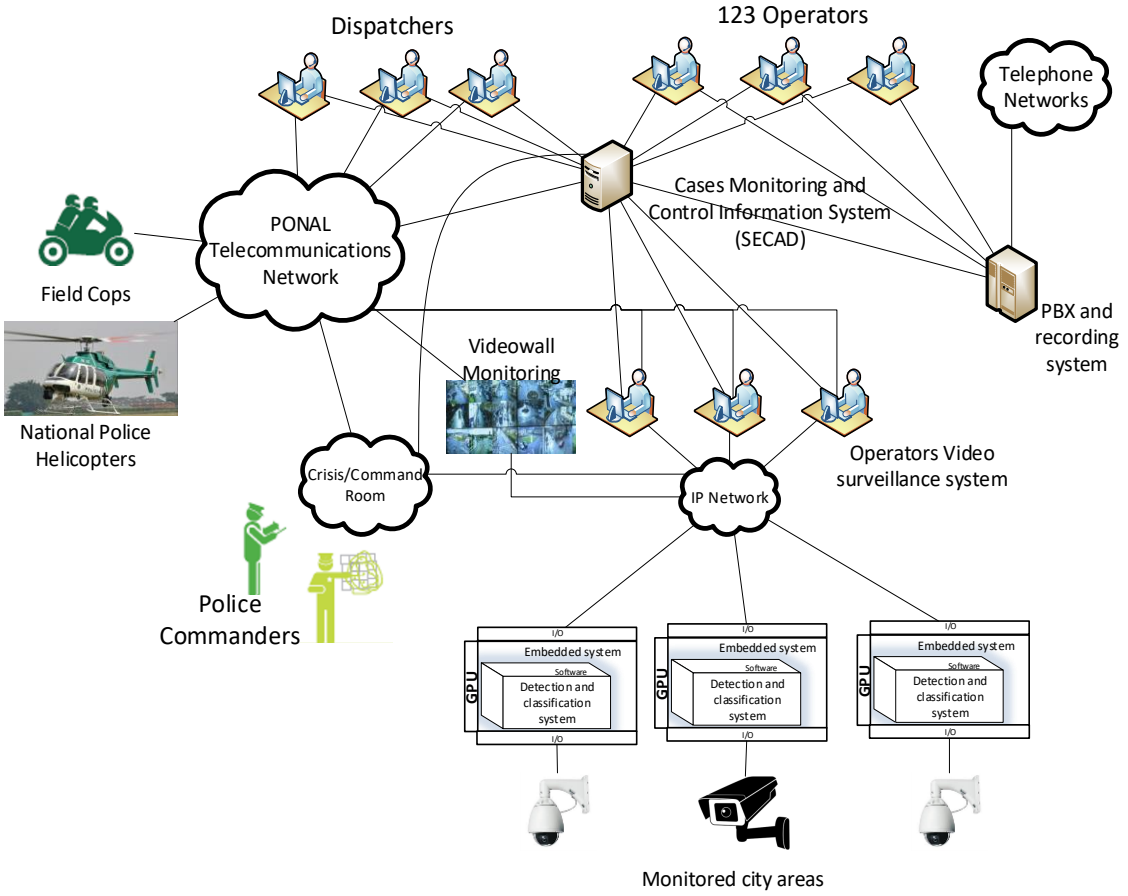

Figure 15. Decentralized Low Processing Time System for criminal activities detection.

After each detection, alarms will be generated and will be sent by a network to the Video Surveillance Subsystem where operators can take actions to prevent and respond to criminal actions.

4.2. Centralized Low Processing Time System to Criminal Activities Detection Based on Real-Time Video Analysis Applied to Colombian National Police Command and Control Citizen Security Center

In contrast to the previously decentralized system shown before, in this case, the video will be processed in a centralized infrastructure with high computational power and GPU capabilities (Figure 16).

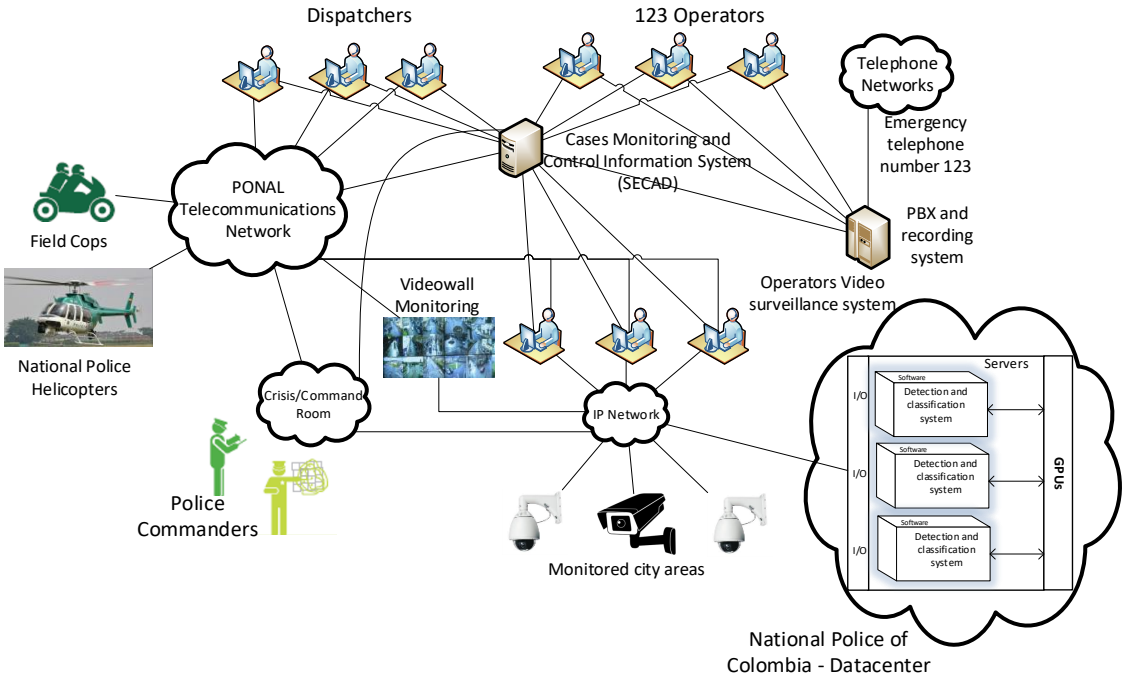

Figure 16. Centralized Low Processing Time System to criminal activities detection.

The datacenter runs the VD&CS individually for each video signal coming from each of the city video surveillance cameras, generating alarms when criminal activities are detected, and sends it back to the Video Surveillance Subsystem through the network, where operators can take actions to prevent and respond to criminal actions.

5. Possible Implementation and Limitations

Considering that the development of VD&CS was performed on a laptop using Matlab and Windows 10 and that an image processing rate of 20 to 30 frames per second was obtained, it is feasible to migrate the VD&CS to an environment with greater efficiency using the libraries optimized for Deep Learning, such as cuDNN [31], TensorFlow [63], TensorRT [64], Nvidia DeepStream SDK [65], further reducing the computational cost.

With this reduction in the computational cost, it would be possible to implement VD&CS in embedded systems, such as the Nvidia Jetson [62] and optimize the implementation using Nvidia DeepStream [65], to be installed directly in citizen security cameras and subsequently generate alerts upon the occurrence of criminal events which would be reported to the Command and Control Citizen Security Center of the Colombian National Police, like in the decentralized Low Processing Time System.

Currently, in June 2019 in Bogotá D.C., there are about 2880 Pan–Tilt–Zoom cameras that are monitored in the Citizen Security Control Center, and these domes generate around 22.4 Gbps of real-time video traffic. Given that currently, in Colombia, there is no cloud provider that has datacenters in the country, it would not be applicable to use cloud solutions with datacenters in the United States or Brazil because the international channel cost would be very high; therefore, in June 2019, the best

solution is to use embedded systems, at least until a cloud provider provides a datacenter with GPU capability in Colombia.

The VD&CS limitations must be considered in future implementations because, like all systems based on Deep Learning, it is not 100% reliable and its precision is linked to critical factors such as lighting and partial obstructions, meaning that human supervision is necessary.

However, the implementation of this Low Processing Time System in a large-scale environment depends on the budget availability of the Government of Colombia.

6. Discussion and Future Application

VD&CS have proven to be effective in a hybrid operation as an object detector and the treatment of criminal actions as objects. If the characteristic gestures are identified in certain actions, it should be possible to use object detectors based on Deep Learning in various applications such as the detection of suspicious activities, fights, riots and more.

As shown above, several recent applications of Faster R-CNN have shown great performance as object detector [48–52], however, this work demonstrated that applying object detection techniques based on Deep Learning like Faster R-CNN in actions detection could be an alternative to action recognition based on analysis of trajectories or movements and could be applied more easily in highly mobile video environments, such as military operations, transportation, citizen security, and national security to name only a few, nevertheless, human supervision is always required, because after a while, the quantity of False Negatives and False Positive could drastically reduce the system effectiveness, which is very serious in safety applications.

In future research, we could identify human actions that could be recognized using object detectors based on Deep Learning.

These actions should have characteristic gestures like in the case of criminal activities, which always have recognizable gestures such as threatening the victim.

Although the system's accuracy is around 70%, this percentage can be considered acceptable because the system is tolerant to the sudden movements of the Pan–Tilt–Zoom cameras of the Colombian National Police. It also shows that it is possible to use an object detector to detect criminal actions and in future applications, the system's accuracy could be improved.

Further future research work consists of maximizing the recognition of human actions using an objects classifier, minimizing system failures. This can be achieved by building more complete datasets and experimenting with diverse Deep Learning techniques such as YOLO, and several CNN models such as ResNet, GoogleNet.

7. Conclusions

By applying the secure city architectures in command and control systems, situational awareness and situation understanding of police commanders will improve, as well as their agility and efficiency in decision making, thus improving the effectiveness of police operations and directly increasing citizen security.

During the development of the VD&CS, it has been proven that it is possible to improve situational awareness in the Command and Control Citizen Security Center of the Colombian National Police, triggering alarms of criminal events captured by the video surveillance system.

Reducing the computational cost for using Deep Learning or any other technique in citizen security applications is fundamental for achieving real-time performance and feasible implementation costs, especially given the amount of information generated by surveillance systems. The processing time is vital to achieve a real improvement of situational awareness.

The Low Processing Time System to Criminal Activities Detection Applied to a Command and Control Citizen Security Center could be deployed in Colombia because the VD&CS showed that it is possible to detect criminal actions using a Deep Learning Object Detector as long as the system is trained to detect actions (these actions must have characteristic gestures such as threatening the

victim). Deep Learning can be a powerful tool in citizen security systems because it can automate the detection of situations of interest which can escape from system operator view in Command and Control Information System of a security agency such as the Colombian National Police.

Author Contributions: Conceptualization, J.S.-P., M.S.-G. and A.C.; methodology, M.E., J.A.G. and C.E.P.; software, J.S.-P. and, M.S.-G.; validation, A.C., and I.P.-L.; formal analysis, J.S.-P., M.S.-G. and A.C.; investigation, J.S.-P., M.S.-G. and A.C.; writing—original draft preparation, J.S.-P., M.S.-G. and A.C.; writing—review and editing, J.S.-P., M.S.-G. and A.C.; visualization, J.S.-P., M.S.-G. and A.C.; supervision, M.E., C.E.P. and J.A.G.; project administration, I.P.-L.; funding acquisition, M.E., C.E.P. and I.P.-L.

Funding: This work was co-funded by the European Commission as part of H2020 call SEC-12-FCT-2016-Subtopic3 under the project VICTORIA (No. 740754). This publication reflects the views only of the authors and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Acknowledgments: The authors thank Colombian National Police and its Office of Telematics for their support on the development of this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Bank United Nations. *Perspectives of Global Urbanization*; Command and Control and Cyber Research Portal (CCRP): Washington, DC, USA, 2019.
2. Alberts, D.S.; Hayes, R.E. *Understanding Command and Control the Future of Command and Control*; Command and Control and Cyber Research Portal (CCRP): Washington, DC, USA, 2006.
3. Esteve, M.; Perez-Llopis, I.; Hernandez-Blanco, L.E.; Palau, C.E.; Carvajal, F. SIMACOP: Small Units Management C4ISR System. In Proceedings of the IEEE International Conference Multimedia and Expo, Beijing, China, 2–5 July 2007; pp. 1163–1166.
4. Wang, L.; Rodriguez, R.M.; Wang, Y.-M. A dynamic multi-attribute group emergency decision making method considering experts' hesitation. *Int. J. Comput. Intell. Syst.* **2017**, *11*, 163. [CrossRef]
5. Esteve, M.; Perez-Llopis, I.; Palau, C.E. Friendly force tracking COTS solution. *IEEE Aerosp. Electron. Syst. Mag.* **2013**, *28*, 14–21. [CrossRef]
6. Esteve, M.; Pérez-Llopis, I.; Hernández-Blanco, L.; Martinez-Nohales, J.; Palau, C.E. Video sensors integration in a C2I system. In Proceedings of the IEEE Military Communications Conference MILCOM, Boston, MA, USA, 18–21 October 2009; pp. 1–7.
7. Spagnolo, P.; D'Orazio, T.; Leo, M.; Distante, A. Advances in background updating and shadow removing for motion detection algorithms. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the 11th International Conference on Computer Analysis of Images and Patterns, CAIP 2005, Versailles, France, 5–8 September 2005*; Springer: Versailles, France, 2005; Volume 3691, pp. 398–406.
8. Nieto, M.; Varona, L.; Senderos, O.; Leskovsky, P.; Garcia, J. Real-time video analytics for petty crime detection. In Proceedings of the 7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016), Madrid, Spain, 23–25 November 2017; pp. 23–26.
9. Senst, T.; Eiselein, V.; Kuhn, A.; Sikora, T. Crowd Violence Detection Using Global Motion-Compensated Lagrangian Features and Scale-Sensitive Video-Level Representation. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2945–2956. [CrossRef]
10. Machaca Arceda, V.; Gutierrez, J.C.; Fernandez Fabian, K. Real Time Violence Detection in Video. In Proceedings of the International Conference on Pattern Recognition Systems (ICPRS-16), Talca, Chile, 20–22 April 2016; pp. 6–7.
11. Bilinski, P.; Bremond, F. Human violence recognition and detection in surveillance videos. In Proceedings of the 13th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2016, Colorado Springs, CO, USA, 23–26 August 2016; pp. 30–36.
12. Xue, F.; Ji, H.; Zhang, W.; Cao, Y. Action Recognition Based on Dense Trajectories and Human Detection. In Proceedings of the IEEE International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, 16–18 November 2018; pp. 340–343.
13. Shi, Y.; Tian, Y.; Wang, Y.; Huang, T. Sequential Deep Trajectory Descriptor for Action Recognition with Three-Stream CNN. *IEEE Trans. Multimed.* **2017**, *19*, 1510–1520. [CrossRef]

14. Dasari, R.; Chen, C.W. MPEG CDVS Feature Trajectories for Action Recognition in Videos. In Proceedings of the IEEE 1th International Conference on Multimedia Information Processing and Retrieval, Miami, FL, USA, 10–12 April 2018; pp. 301–304.
15. Arunnehru, J.; Chamundeeswari, G.; Bharathi, S.P. Human Action Recognition using 3D Convolutional Neural Networks with 3D Motion Cuboids in Surveillance Videos. *Procedia Comput. Sci.* **2018**, *133*, 471–477. [CrossRef]
16. Kamel, A.; Sheng, B.; Yang, P.; Li, P.; Shen, R.; Feng, D.D. Deep Convolutional Neural Networks for Human Action Recognition Using Depth Maps and Postures. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *49*, 1–14. [CrossRef]
17. Ren, J.; Reyes, N.H.; Barczak, A.L.C.; Scogings, C.; Liu, M. Towards 3D human action recognition using a distilled CNN model. In Proceedings of the IEEE 3rd International Conference Signal and Image Processing (ICSIP), Shenzhen, China, 13–15 July 2018; pp. 7–12.
18. Zhang, B.; Wang, L.; Wang, Z.; Qiao, Y.; Wang, H. Real-Time Action Recognition with Deeply Transferred Motion Vector CNNs. *IEEE Trans. Image Process.* **2018**, *27*, 2326–2339. [CrossRef]
19. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 142–158. [CrossRef]
20. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
21. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Washington, DC, USA, 3–7 December 2015; pp. 1440–1448.
22. Suarez-Paez, J.; Salcedo-Gonzalez, M.; Esteve, M.; Gómez, J.A.; Palau, C.; Pérez-Llopis, I. Reduced computational cost prototype for street theft detection based on depth decrement in Convolutional Neural Network. Application to Command and Control Information Systems (C2IS) in the National Police of Colombia. *Int. J. Comput. Intell. Syst.* **2018**, *12*, 123. [CrossRef]
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
24. Hao, S.; Wang, P.; Hu, Y. Haze image recognition based on brightness optimization feedback and color correction. *Information* **2019**, *10*, 81. [CrossRef]
25. Jiang, H.; Learned-Miller, E. Face Detection with the Faster R-CNN. In Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017—1st International Workshop on Adaptive Shot Learning for Gesture Understanding and Production, ASL4GUP 2017, Washington, DC, USA, 3 June 2017; pp. 650–657.
26. Peng, M.; Wang, C.; Chen, T.; Liu, G. NIRFaceNet: A convolutional neural network for near-infrared face identification. *Information* **2016**, *7*, 61. [CrossRef]
27. Wu, S.; Zhang, L. Using Popular Object Detection Methods for Real Time Forest Fire Detection. In Proceedings of the 11th International Symposium on Computational Intelligence and Design, ISCID, Hangzhou, China, 8–9 December 2018; pp. 280–284.
28. Chen, J.; Miao, X.; Jiang, H.; Chen, J.; Liu, X. Identification of autonomous landing sign for unmanned aerial vehicle based on faster regions with convolutional neural network. In Proceedings of the Chinese Automation Congress, CAC, Jinan, China, 20–22 October 2017; pp. 2109–2114.
29. Xu, W.; He, J.; Zhang, H.L.; Mao, B.; Cao, J. Real-time target detection and recognition with deep convolutional networks for intelligent visual surveillance. In Proceedings of the 9th International Conference on Utility and Cloud Computing—UCC '16, New York, NY, USA, 23–26 February 2016; pp. 321–326.
30. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the ACM International Conference on Multimedia—MM '14, New York, NY, USA, 18–19 June 2014; pp. 675–678.
31. Nvidia Corporation. NVIDIA CUDA®Deep Neural Network library (cuDNN). Available online: https://developer.nvidia.com/cuda-downloads (accessed on 24 November 2019).
32. Song, D.; Qiao, Y.; Corbetta, A. Depth driven people counting using deep region proposal network. In Proceedings of the IEEE International Conference on Information and Automation, ICIA 2017, Macau SAR, China, 18–20 July 2017; pp. 416–421.

33. Saikia, S.; Fidalgo, E.; Alegre, E.; Fernández-Robles, L. Object Detection for Crime Scene Evidence Analysis Using Deep Learning. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the International Conference on Mobile and Wireless Technology, ICMWT 2017, Kuala Lumpur, Malaysia, 26–29 June 2017*; Springer: Cham, Switzerland, 2017; pp. 14–24.
34. Sutanto, R.E.; Pribadi, L.; Lee, S. 3D integral imaging based augmented reality with deep learning implemented by faster R-CNN. In Proceedings of the Lecture Notes in Electrical Engineering, Proceedings of the International Conference on Mobile and Wireless Technology, ICMWT 2017, Kuala Lumpur, Malaysia, 26–29 June 2017; Springer: Singapore, 2018; pp. 241–247.
35. Wu, X.; Lu, X.; Leung, H. A video based fire smoke detection using robust AdaBoost. *Sensors* **2018**, *18*, 3780. [CrossRef] [PubMed]
36. Park, J.H.; Lee, S.; Yun, S.; Kim, H.; Kim, W.-T.; Park, J.H.; Lee, S.; Yun, S.; Kim, H.; Kim, W.-T. Dependable Fire Detection System with Multifunctional Artificial Intelligence Framework. *Sensors* **2019**, *19*, 2025. [CrossRef]
37. García-Retuerta, D.; Bartolomé, Á.; Chamoso, P.; Corchado, J.M. Counter-Terrorism Video Analysis Using Hash-Based Algorithms. *Algorithms* **2019**, *12*, 110. [CrossRef]
38. Zhao, B.; Zhao, B.; Tang, L.; Han, Y.; Wang, W. Deep spatial-temporal joint feature representation for video object detection. *Sensors* **2018**, *18*, 774. [CrossRef]
39. He, Z.; He, H. Unsupervised Multi-Object Detection for Video Surveillance Using Memory-Based Recurrent Attention Networks. *Symmetry* **2018**, *10*, 375. [CrossRef]
40. Zhang, H.; Zhang, Z.; Zhang, L.; Yang, Y.; Kang, Q.; Sun, D.; Zhang, H.; Zhang, Z.; Zhang, L.; Yang, Y.; et al. Object Tracking for a Smart City Using IoT and Edge Computing. *Sensors* **2019**, *19*, 1987. [CrossRef]
41. Mazzeo, P.L.; Giove, L.; Moramarco, G.M.; Spagnolo, P.; Leo, M. HSV and RGB color histograms comparing for objects tracking among non overlapping FOVs, using CBTF. In Proceedings of the 8th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2011, Washington, DC, USA, 30 August–2 September 2011; pp. 498–503.
42. Leo, M.; Mazzeo, P.L.; Mosca, N.; D'Orazio, T.; Spagnolo, P.; Distante, A. Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In Proceedings of the International Conference on Content-based Image and Video Retrieval, Niagara Falls, ON, Canada, 7–9 July 2008; pp. 525–534.
43. Muhammad, K.; Hamza, R.; Ahmad, J.; Lloret, J.; Wang, H.; Baik, S.W. Secure surveillance framework for IoT systems using probabilistic image encryption. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3679–3689. [CrossRef]
44. Barthélemy, J.; Verstaevel, N.; Forehead, H.; Perez, P. Edge-Computing Video Analytics for Real-Time Traffic Monitoring in a Smart City. *Sensors* **2019**, *19*, 2048. [CrossRef]
45. Aqib, M.; Mehmood, R.; Alzahrani, A.; Katib, I.; Albeshri, A.; Altowaijri, S.M. Smarter Traffic Prediction Using Big Data, In-Memory Computing, Deep Learning and GPUs. *Sensors* **2019**, *19*, 2206. [CrossRef] [PubMed]
46. Xu, S.; Zou, S.; Han, Y.; Qu, Y. Study on the availability of 4T-APS as a video monitor and radiation detector in nuclear accidents. *Sustainability* **2018**, *10*, 2172. [CrossRef]
47. Plageras, A.P.; Psannis, K.E.; Stergiou, C.; Wang, H.; Gupta, B.B. Efficient IoT-based sensor BIG Data collection—Processing and analysis in smart buildings. *Futur. Gener. Comput. Syst.* **2018**, *82*, 349–357. [CrossRef]
48. Jha, S.; Dey, A.; Kumar, R.; Kumar-Solanki, V. A Novel Approach on Visual Question Answering by Parameter Prediction using Faster Region Based Convolutional Neural Network. *Int. J. Interact. Multimed. Artif. Intell.* **2019**, *5*, 30. [CrossRef]
49. Zhang, Q.; Wan, C.; Han, W. A modified faster region-based convolutional neural network approach for improved vehicle detection performance. *Multimed. Tools Appl.* **2019**, *78*, 29431–29446. [CrossRef]
50. Cho, S.; Baek, N.; Kim, M.; Koo, J.; Kim, J.; Park, K. Face Detection in Nighttime Images Using Visible-Light Camera Sensors with Two-Step Faster Region-Based Convolutional Neural Network. *Sensors* **2018**, *18*, 2995. [CrossRef]
51. Zhang, J.; Xing, W.; Xing, M.; Sun, G. Terahertz Image Detection with the Improved Faster Region-Based Convolutional Neural Network. *Sensors* **2018**, *18*, 2327. [CrossRef]

52. Liu, X.; Jiang, H.; Chen, J.; Chen, J.; Zhuang, S.; Miao, X. Insulator Detection in Aerial Images Based on Faster Regions with Convolutional Neural Network. In Proceedings of the IEEE International Conference on Control and Automation, ICCA, Anchorage, AK, USA, 12–15 June 2018; pp. 1082–1086.
53. Bakheet, S.; Al-Hamadi, A. A discriminative framework for action recognition using f-HOL features. *Information* **2016**, *7*, 68. [CrossRef]
54. Al-Gawwam, S.; Benaissa, M. Robust eye blink detection based on eye landmarks and Savitzky-Golay filtering. *Information* **2018**, *9*, 93. [CrossRef]
55. Krizhevsky, A.; Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**. [CrossRef]
56. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
57. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2014; pp. 1–9.
58. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June –1 July 2016; pp. 770–778.
59. Hou, R.; Chen, C.; Shah, M. Tube Convolutional Neural Network (T-CNN) for Action Detection in Videos. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5823–5832.
60. Kalogeiton, V.; Weinzaepfel, P.; Ferrari, V.; Schmid, C. Action Tubelet Detector for Spatio-Temporal Action Localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4415–4423.
61. Zolfaghari, M.; Oliveira, G.L.; Sedaghat, N.; Brox, T. Chained Multi-stream Networks Exploiting Pose, Motion, and Appearance for Action Classification and Detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2923–2932.
62. Nvidia Corporation. Jetson Embedded Development Kit|NVIDIA. Available online: https://developer.nvidia.com/embedded-computing (accessed on 24 November 2019).
63. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv* **2016**, arXiv:1603.04467.
64. Nvidia Corporation. NVIDIA TensorRT|NVIDIA Developer. Available online: https://developer.nvidia.com/tensorrt (accessed on 24 November 2019).
65. Nvidia Corporation. NVIDIA DeepStream SDK|NVIDIA Developer. Available online: https://developer.nvidia.com/deepstream-sdk (accessed on 24 November 2019).
66. Fraga-Lamas, P.; Fernández-Caramés, T.M.; Suárez-Albela, M.; Castedo, L.; González-López, M. A Review on Internet of Things for Defense and Public Safety. *Sensors* **2016**, *16*, 1644. [CrossRef] [PubMed]
67. Gomez, C.A.; Shami, A.; Wang, X. Machine learning aided scheme for load balancing in dense IoT networks. *Sensors* **2018**, *18*, 3779. [CrossRef]
68. AMD Embedded Radeon^{TM}. Available online: https://www.amd.com/en/products/embedded-graphics (accessed on 24 November 2019).