ATLANTIS PRESS
International Journal of Computational Intelligence Systems, Vol. 12 (2018) 123-130

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.

Julio Suarez-Paez[1], Mayra Salcedo-Gonzalez[1], M. Esteve[1], J.A. Gómez[2], C. Palau[1], I. Pérez-Llopis[1].

[1]Distributed Real-time Systems Laboratory (SATRD), Universitat Politècnica de València,

Camino de Vera, s/n
Valencia, 46022, Spain
E-mail: juliosuarez@ieee.org, m.l.salcedogonzalez@ieee.org, mesteve@dcom.upv.es, cpalau@dcom.upv.es, ispello0@upvnet.upv.es

[2] Pattern Recognition and Human Language Technology, Universitat Politècnica de València,

Camino de Vera, s/n
Valencia, 46022, Spain
E-mail: jon@dsic.upv.es

Received 7 April, 2018

Accepted 16 September, 2018

Abstract

This paper shows the implementation of a prototype of street theft detector using the deep learning technique R-CNN (Region-Based Convolutional Network), applied in the Command and Control Information System (C2IS) of National Police of Colombia, the prototype is implemented using three models of CNN (Convolutional Neural Network), AlexNet, VGG16 and VGG19 comparing their computational cost measuring the image processing time, according to the complexity (depth) of each model. Finally, we conclude which model has the lowest computational cost and is more useful for the case of the National Police of Colombia.

Keywords: Deep Learning, R-CNN, AlexNet, VGG16, VGG19, CNN (Convolutional Neural Network), Command and Control Information System (C2IS).

1. Introduction

The street theft to people is one of the biggest challenges for citizen security agencies, as it affects many cities in the world.
Colombia is not immune to this problem, since its cities have considerable indices of this crime.

To face this situation, the National Police of Colombia has around 178,000 agents and several technological tools, including command and control information systems (C2IS)[1]. This system provides strategic information in real time, improving the Situational Awareness[1], allowing to react in a timely way to situations such as street theft to people.

In many C2IS, the Video Surveillance System is a fundamental source of strategic information, however, it is common that some events (like theft) can't be detected by Video Surveillance operators, because commonly there are more cameras than one operator can manage.

A future aim is to automate the detection of street theft in the C2IS of the Colombian National Police using the Video Surveillance System. This article shows the

development and implementation of a prototype of street theft detector based on deep learning, specifically using the R-CNN (Region-Based Convolutional Network) technique[2], which is implemented using three different architectures of CNN (Convolutional Neural Network)[3,4], AlexNet[5], VGG16[6] and VGG19[6]. Finally, the prototypes were tested with information from the Video Surveillance system of the National Police of Colombia, with the support of the Telematics Office, in order to conclude which architecture may have the lowest computational cost for a future implementation in the National Police of Colombia.

2. Street theft detection using deep learning.

The National Police of Colombia has deployed thousands of security cameras across the country, however, the police do not have an efficient system to identify the street theft or other criminal activities. Additionally, most of these cameras are PTZ (pan-tilt-zoom) Dome type, i.e. mobile cameras with images always changing and which do not have a fixed background like in a fixed camera. Therefore, to design a street theft detector is a very relevant factor to define how the video analysis will be performed in these conditions.

For this development, we discard methods based on motion detection through prediction and other similar techniques[7,8,9,10], inasmuch as the PTZ (pan-tilt-zoom) Dome cameras of the National Police of Colombia changes abruptly the pan, tilt and zoom. Because of that the backgrounds of the images change drastically making useless several techniques of video analysis.
Although there are techniques to work with mobile background images[11,12,13], given its complexity and the specific case of the National Police of Colombia, alternative techniques are sought for detection of street theft, which are independent of camera movement and its relationship between each video frame.

For these reasons, the approach used for detection of street thefts was based on deep learning techniques which have shown in many publications[14,15,16,17] their efficiency in computer vision.

In development of the current prototype, it is proposed to use an R-CNN (Region-Based Convolutional Network)[2], which was trained with a dataset of street theft images created specifically for this project.

3. Prototype development and implementation

In development of the street theft detection prototype, it was used the deep learning technique (R-CNN) published by R. Girshick[2]. We have chosen R-CNN because it is the basis of techniques with a better performance respect to the processing time for each image like as Fast R-CNN[18] or Faster R-CNN[19] which are the state-of-the-art in object detection and have many applications, having hundreds of papers in different journals and conferences.

Another determining factor for choosing the deep learning technique R-CNN[2] to develop the prototype, is considering that Fast R-CNN[18] and Faster R-CNN[19] are based in R-CNN[2], because that, the results of the performed tests in R-CNN[2] could be applied to Fast R-CNN[18] and Faster R-CNN[19].

The hardware available to train and execute the Convolutional Neural Network, was a laptop MSI GT62VR-7RE whit a CPU Intel Core I7 7700HQ, 16 GB of DDR4 RAM, with an GPU NVIDIA GeForce GTX 1070 in MXM format which consists of 2048 CUDA cores at 1442 MHz base frequency and 8 GB DDR5 VRAM memory[20], even though it has sufficient computing power for prototype development using R-CNN, is not a comparison with the latest processors and GPUs for desktops and servers, specialized in the training and execution of this type of algorithms. However, for the Colombian National Police, it is very important to reduce computational cost of this prototype due to the large number of citizen security cameras.

Once the configuration was selected and adjusted to the needs of the National Police of Colombia, the implementation was carried out in MATLAB 2018a following the methodology described below:
(*it must be considered that the implementation and testing were done three times, first with AlexNet[5], another with VGG16[6] and the last one using VGG19[6]*).
As proposed in [2], the training of an R-CNN is divided into three stages which are:

- Training of a Convolutional Neural Network as an image classifier.

- Fine training of the Convolutional Neural Networks for the detection of Thefts.
- Support vector machine linear predictor for the detection of street theft.

3.1. Training of a Convolutional Neural Network as an image classifier:

In this stage, three well-known and proved models of Convolutional Neural Network were used. Fig. 2 shows AlexNet[5] and Fig. 4 shows VGG16[6] and VGG19[6], that are defined as very deep networks by the authors from the University of Oxford.

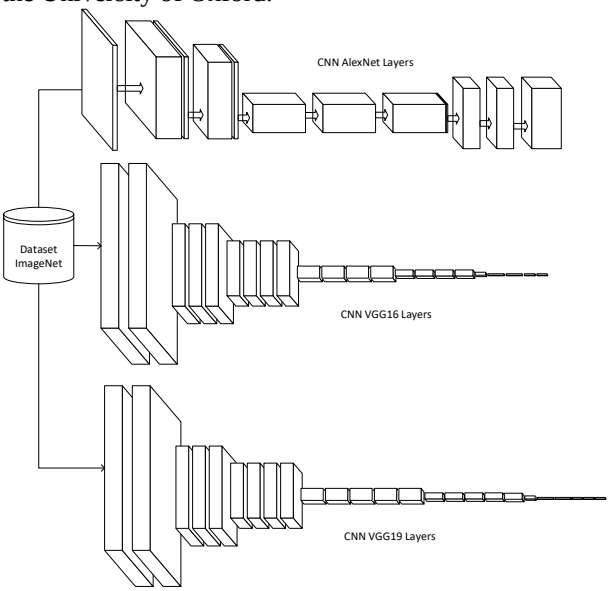

Fig. 1. Pre-trained Convolutional Neural Networks AlexNet[5], VGG16[6] and VGG19[6].

For this case, several topologies, were analyzed as well as pre-trained models, and afterwards, the implantation was carried out with the pre-trained models known as AlexNet, VGG16 and VGG19. These pretrained models can classify 1000 different kinds of images and were trained with approximately 1.2 million images using the well-known ImageNet dataset[21] as shown in Fig.1.

The topologies of CNN AlexNet, VGG16 and VGG19 used in this prototype are shown in Fig. 2. and Fig. 4 respectively.

3.2. Fine training of the Convolutional Neural Networks for the detection of Thefts.

In this stage of the training process, it was made a fine retraining of the chosen models. For the case of this prototype, the topologies used were AlexNet, VGG16 and VGG19. However, the ImageNET dataset was not used in this stage, instead, it was used a street theft dataset created for this specific case (Fig. 5). Fine-training was done in the laptop MSI GT62VR-7RE previously mentioned.

In this stage, models AlexNET, VGG16 and VGG19 were fine-trained independently using a new dataset made with street theft images, shown by in Fig 3.

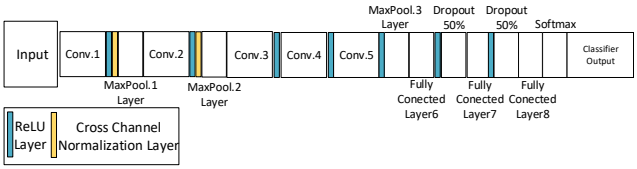

Fig. 2. CNN Layers: AlexNet[5].

It is important to note that in this training stage, VGG19 and VGG16 networks had a much longer training time than AlexNet, due to the depth (complexity) of the network topology, even though the same hardware and the same dataset were used for the fine training.

The fine training of the CNN was done with the dataset described below (Fig. 3), using parallel processing in the NVIDIA GeForce GTX 1070 GPU, and a total of 50 epochs, this fine training was done on the Convolutional Neural Network Topology proposed in[2] AlexNet, VGG16 and VGG19 shown by Figures 2 and 4.

Considering that we want to identify a relationship between the complexity of the model and the processing time of each image, when using CNN Region-Based Techniques, to ensure that three models are on equal terms, the hyperparameters used in fine training were equal for VGG19, VGG16 and AlexNet and had the following values:

- Size of the mini-batch: 50
- Initial Learning Rate: 0.001
- Factor for dropping the learning rate: 0.1
- Epochs for dropping the learning rate: 10
- Factor for L2 regularization: 0.0001
- Contribution of previous step: 0.9
- Decay rate of gradient moving average: 0.9
- Decay rate of squared gradient moving average: 0.999
- Denominator offset: 10^{-8}

The fine training described in this stage was performed on the equipment described above and took about 8 hours of processing. This time can be reduced using Multi-GPU schemes, or a cluster of several computers that have GPUs.

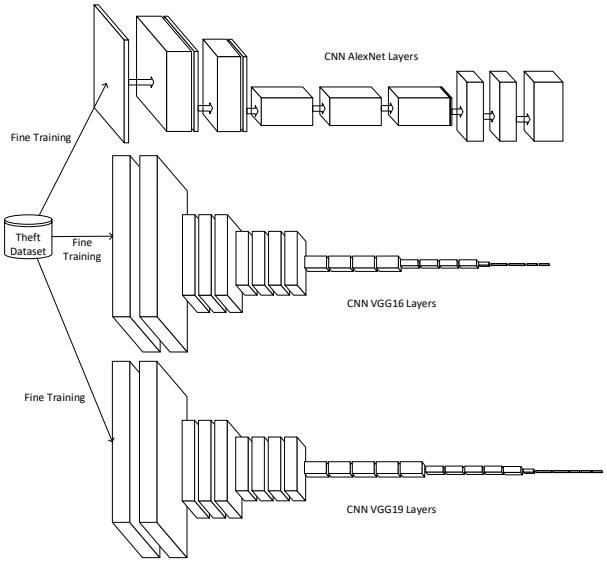

Fig. 3. Fine-training of Convolutional Neural Networks AlexNet[2], VGG16[4] and VGG19[6] with Street Theft Dataset.

3.3. Support vector machine linear predictor for the detection of street theft

In the third stage of the training process, it was trained a binary classifier using a Support Vector Machine with linear kernel [4,22], which aims to classify two classes: "Theft" and "No Theft".

Fig. 5. Sample images of the fine training dataset.

According to this, the SVM was trained whit two classes:

- Class “Theft”: this class includes the regions proposals obtained from the labeled images in the street thefts dataset, and characteristics extracted of the images obtained in the previous stage, from the fine-training of models VGG19, VGG16 and AlexNet individually, shown in Figures 6 and 7.
- Class “No Theft”: in this class where used pieces of images extracted from street thefts dataset, different from those labeled manually and those obtained in the previous stage of the R-CNN training process, as shown in Figures 6 and 7.

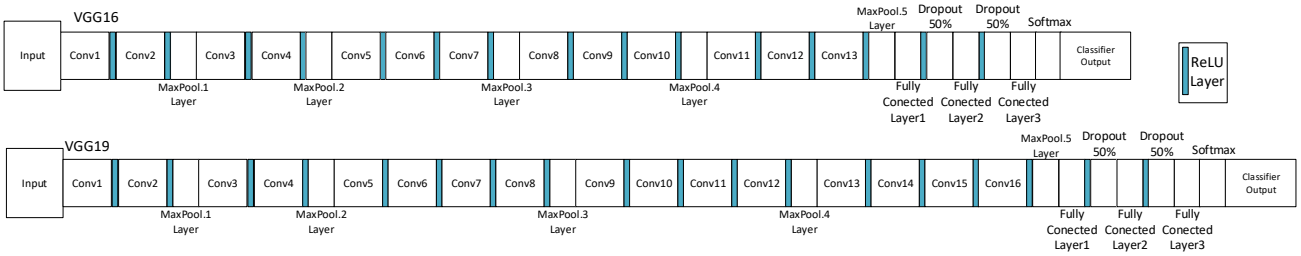

Fig. 4. CNN Layers: VGG16[6] and VGG19[6].

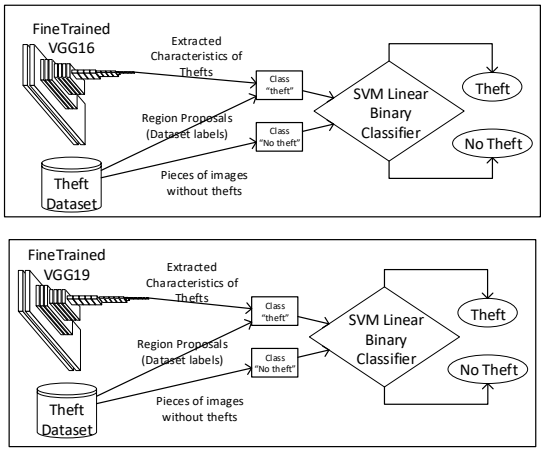

Fig. 6. Binary Classifier using a Support Vector Machine with linear kernel using fine trained VGG16 and VGG19.

This process had been done three times, fist using AlexNet model, another with VGG16, and last using VGG19.

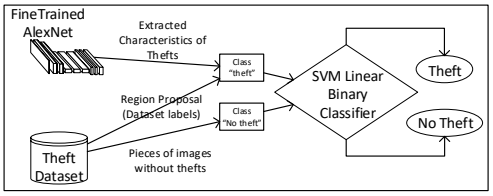

Fig. 7. Binary Classifier using a Support Vector Machine with linear kernel using fine trained AlexNet.

4. Results of the implementation

In this study, interesting results were obtained, by implementing the prototypes of detection of street theft using deep learning, thanks to the fine-tuned models VGG19, VGG16 and AlexNet, getting positive detections as Fig. 9 shows. However, the prototypes are not completely infallible, having false detections and situations in which there are street thefts that are not detected.

To verify the effectiveness of this prototype, a series of tests were carried out with images of street theft not used in the training of the system. Another important measured parameter was the processing time in each image to measure the computational cost. It should be noted that the tests were performed by using the same laptop with the technical characteristics previously described.

To analyze the reduction of the computational cost focused on the image processing time of the Street Theft Detector prototype, 300 static images were analyzed with street theft and we also analyzed Video from the C2IS security cameras of the National Police of Colombia; these results are shown in table 1.

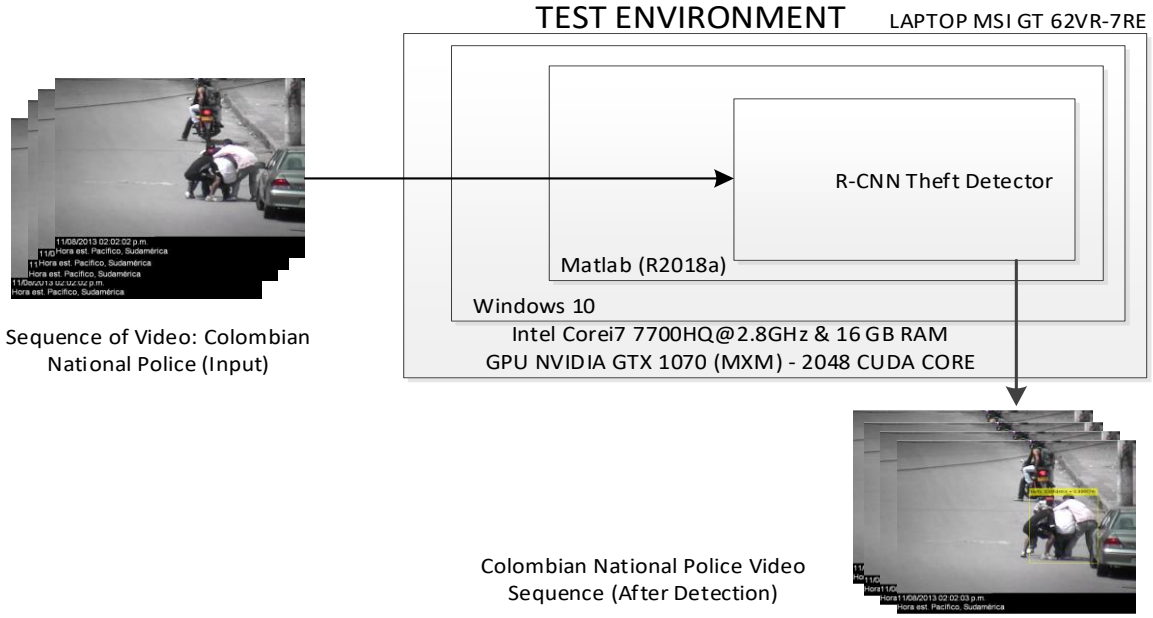

Fig. 8. Test Environment.

Table 1. Prototype test results

300 Image Test	AlexNet	VGG16	VGG19
Positive Detections	219	231	237
Fails	81	69	63
Average processing time	2.45 sec	22.75 sec	26.59 sec
Accuracy	73%	77%	79%
Frames per second in Video	0.4 FPS	0.04 FPS	0.03 FPS

After performing these tests, we conclude, AlexNet has shorter image processing time and lower computational cost than VGG19 and VGG16 if models are implemented in CNN Region-Based Techniques, under the same conditions.

However, this decrease in the computational cost affects prototype's accuracy, because the tests show that AlexNet has an accuracy 6% lower than VGG19 and 4% lower than VGG16. On the other hand, VGG16 has a computational cost 929 % higher than AlexNet and VGG19 has a computational cost 1085% higher than AlexNet. This reduction in the computational cost is very important for future applications in the National Police of Colombia and should compensate the loss of accuracy.

These results could be applied to Fast R-CNN[18] and Faster R-CNN[19] because they are based on R-CNN[2] and for future applications, these results could be considered to develop a theft detection system that analyzes real-time video.

5. Future Application into National Police of Colombia Command and Control Information System (C2IS)

Bearing in mind that the work shown in this paper is a prototype, it should be noted that the application to the Colombian National Police C2IS was not carried out with real-time videos, it was done on pre-recorded videos given by the Telematics Office of the Colombian National Police, which were analyzed frame by frame, as shown in Fig. 8

Fig. 9. Results of Street Theft Detector.

As a result, street theft detections were obtained while the conditions of the video allowed the detection. For example, in blurred videos or with very low resolution the detection was not detected. However, in videos where the street theft had a minimum pixel size, the detection is consistent with the results reported in Table 1.

When using the prototype in videos obtained from Video Surveillance Cameras (PTZ Dome) with average heights of 10 meters, there were only detections of Street theft when the camera made enough zoom, so that the street theft image had a minimum of pixels.

Fig. 10. Street Theft Detection in real Video of the Surveillance system of the Colombian National Police.

6. Prototype Limitations

Developing this prototype, important limitations were identified and must be considered for future implementations:

- Quality in video stream and size in pixels of the street theft to be detected: When applying the prototype in real situations its effectiveness is limited by these factors, restricting its applicability to the parts of the Video Surveillance system with good video quality.

- Lighting conditions: if images of the Video Surveillance System have dim lighting conditions and the people are not distinguished very well from the background (e.g., a person dressed in dark in a dark background), the effectiveness of the system is greatly reduced.

- Partial obstructions: the partial obstructions generated by the electrical infrastructure, traffic lights, trees and other common elements in an urban environment could cause the prototype to make erroneous detections or not to detect a street theft present in the environment.

- Erroneous detections: in real situations erroneous detections are presented which would trigger false alarms, so the system must always work under human supervision.

- Computational Cost: As shown in Table 1, the computational cost is directly related to deep learning model used in the prototype. In this case, AlexNet had a computational cost of around 950% lower than VGG16 and VGG19. In the case of the National Police of Colombia, VGG16, VGG19 models are discarded for short-term implementation, at least until the computing power of the GPU available in the market increases.

- Accuracy lost: As mentioned above, reducing the computational cost entails a loss of accuracy of 4% to 6%, which should be taken in future implementations.

7. Conclusion

With the tests carried out on these prototypes we demonstrated that the implementation of a theft detection system based on deep learning is possible. In this case it was applied to the National Police of Colombia C2IS with the aim of improving situational awareness, but taking into account the above mentioned limitations.

Given the computational cost (running time) for processing every video frame, in the case of video applications is highly recommended to use deep learning models with a good balance between complexity and accuracy. As it can be observed from the obtained results in this work, AlexNet has a computational cost around 950% lower than VGG16 and VGG19. This low image processing time is fundamental to be considered in future implementations on real environments.

Deep learning offers many possibilities for city security applications and if any of these is implemented in a Command and Control Information System of a security agency such as the National Police of Colombia, it will possibly provide vital information, which would improve the response to situations of criminality or public order.

Acknowledgements

The authors belonging to the Universitat Politècnica de València, Spain, thank National Police of Colombia and its Office of Telematics the support in the development of this project.

References

1. D. S. Alberts and R. E. Hayes, Understanding Command And Control, CCRP, 2011.
2. R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142 - 158, Jan 2016.
3. S. Lawrence, C. Giles and A. C. Tsoi, "Convolutional neural networks for face recognition," in *1996 IEEE Computer Society Conerence on Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96*, San Francisco, CA, USA, USA, 1996.

4. C. Silva, D. Welfer, F. P. Gioda and C. Dornelles, "Cattle Brand Recognition using Convolutional Neural Network and Support Vector Machines," *IEEE Latin America Transactions*, vol. 15, no. 2, pp. 310-316, 2017.
5. A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional," in *Advances in Neural Information Processing Systems 25*, C. B. L. B. K. W. F. Pereira, Ed., Curran Associates, Inc., 2012, pp. 1097-1105.
6. K. Simonyan and A. Zisserman, "VERY DEEP CONVOLUTIONAL FOR LARGE-SCALE IMAGE RECOGNITION," in *ICLR 2015*, San Diego, USA., 2015.
7. A. Schumann and E. Monari, "A soft-biometrics dataset for person tracking and re-identification," in *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*, Seoul, South Korea, 2014.
8. M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier and L. V. Gool, "Online Multiperson Tracking-by-Detection from a Single, Uncalibrated Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1820-1833, 2011.
9. A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan and M. Shah, "Visual Tracking: An Experimental Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442-1468, 2013.
10. Z. H. Khan, I. Y.-H. Gu and A. G. Backhouse, "Robust Visual Object Tracking Using Multi-Mode Anisotropic Mean Shift and Particle Filters," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 1, pp. 74-87, 2011.
11. L. Zhang, Z. Zhang and H. Xiong, "Visual Pedestrian Tracking from a UAV Platform," in *Multimedia and Image Processing (ICMIP), 2017 2nd International Conference on*, 2017.
12. X. Lu and D. Li, "Research on target detection and tracking system of rescue robot," in *Chinese Automation Congress (CAC), 2017*, Jinan, China, 2017.
13. T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," in *Applied Imagery Pattern Recognition Workshop, 2008. AIPR '08. 37th IEEE*, Washington DC, USA, 2008.
14. H. Venkateswara, S. Chakraborty and S. Panchanathan, "Deep-Learning Systems for Domain Adaptation in Computer Vision: Learning Transferable Feature Representations," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 117-129, 09 November 2017.
15. J. X. Z. L. T. L. W. Y. Peiqin Li, "Facial Peculiarity Retrieval via Deep Neural Networks Fusion," *International Journal of Computational Intelligence Systems*, vol. 11, no. 1, pp. 58-65, 2018.
16. N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," *IEEE Access*, 19 February 2018.
17. S. Nie, M. Zheng and Q. Ji, "The Deep Regression Bayesian Network and Its Applications: Probabilistic Deep Learning for Computer Vision," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 101-111, 2018.
18. R. B. Girshick, "Fast R-CNN," *CoRR*, vol. 1504.08083, 2015.
19. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.
20. B. T. Nugraha, S.-F. Su and Fahmizal, "Towards self-driving car using convolutional neural network and road lane detector," in *Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT), 2017 2nd International Conference on*, Jakarta, Indonesia, 2017.
21. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211-252, 2015.
22 W. Zeng, J. Jia, Z. Zheng, C. Xie and L. Guo, "A comparison study: Support vector machines for binary classification in machine learning," in *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*, Shanghai, China, 2011.