Pedestrian detection system based on deep learning

Mohammed Razzok[1], Abdelmajid Badri[1], Ilham EL Mourabit[1], Yassine Ruichek[2], Aïcha Sahel[1]
[1]Laboratory of Electronics, Energy, Automation, and Information Processing, Faculty of Sciences and Techniques Mohammedia, University Hassan II Casablanca, Casablanca, Morocco
[2]Laboratory CIAD, University Burgundy Franche-Comté, Belfort, France

Article Info

Article history:

Received Jun 20, 2021
Revised Dec 25, 2021
Accepted Apr 15, 2022

Keywords:

Deep learning
Image classification
Neural network
Pedestrian detection

ABSTRACT

Pedestrian detection is a rapidly growing field of computer vision with applications in smart cars, surveillance, automotive safety, and advanced robotics. Most of the success of the last few years has been driven by the rapid growth of deep learning, more efficient tools capable of learning semantic, high-level, deeper features of images are proposed. In this article, we investigated the task of pedestrian detection on roads using models based on convolutional neural networks. We compared the performance of standard state-of-the-art object detectors like Faster region-based convolutional network (R-CNN), single shot detector (SSD), and you only look once, version 3 (YOLOv3). Results show that YOLOv3 is the best object detection model than others for pedestrians in terms of detection and time prediction.

Corresponding Author:

Mohammed Razzok
Laboratory of Electronics, Energy, Automation, and Information Processing
Faculty of Sciences and Techniques Mohammedia, University Hassan II Casablanca
N°51, Hay Ifriquia Rue Echahid Eloualid Essaghir, Casablanca 20000, Morocco
Email: mohammedrazzok@gmail.com

1. INTRODUCTION

The increasing number of vehicles during this century has made road accidents a major cause of death. Traffic accidents in Morocco cause more than 4000 deaths each year, 25% are pedestrians. Both the scientific community and the automobile industry have contributed to the development of various types of protection systems to improve the vehicle's safety and environmental performance. At the moment, the main goal in this field is to provide drivers with information about their environment and any potential dangers. Two of all useful information are the detection and location of pedestrians in front of a vehicle. Traditional object detection techniques in the past were based on are based on handcrafted features such as integral channel features (ICF) [1], [2], scale-invariant feature transform (SIFT) [3], histogram of oriented gradients (HOG) [4], local binary patterns (LBP) [5], general forward-backward (GFB) [6] ,and their variations [7]-[9] and combinations [10], [11], followed by a trainable classifier such as support vector machines (SVM) [7], [12], boosted classifiers [13], or random forests [14]. Their performance can be easily degraded by constructing complex ensembles that combine numerous low-level features with high-level context from object detectors and scene classifiers. With the rapid progress of deep learning technology, more effective tools capable of learning semantic, high-level, and deeper features are being introduced to address the issues in traditional architectures.

Deep convolutional neural networks (DCNN) [15]-[19] provide us such ability with high performance for various computer vision applications. Our study focuses on detecting pedestrians in individual monocular images using state-of-the-art object detection approaches based on neural networks. The rest of this research study includes: section 2 describes object detection models we will address. After presenting and discussing our testing results in section 3, we conclude in section 4.

2. OBJECT DETECTION MODELS

2.1. Faster R-CNN

Faster region-based convolutional network (R-CNN) proposed by Ren *et al.* [20] , runs at 7 FPS using Nvidia TiTan X graphic card. It employs a separate network known as the region proposal network to identify region proposals. The predicted regions are then reshaped with the help of a region of interest (ROI) pooling layer. After that Faster R-CNN classifies the image within the proposed region and predicts the bounding box offset values. Its structure is illustrated in Figure 1.

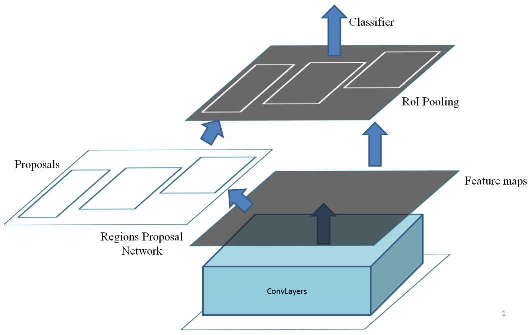

Figure 1. Faster R-CNN architecture

2.2. YOLOv3

You only look once (YOLO) invented by Redmon and Farhadi [21], is a convolutional network (CNN) based open-source object detection and classification algorithm. At first glance, it can tell which objects are present in an image and where they are located. The primary benefit of this technique is that a single neural network evaluates the entire image. Using an Nvidia TiTan X graphic card, the network can process images in real-time at 45 frames per second, and a simplified version called Fast YOLO can process images at 155 frames per second, outperforming other real-time detectors. Furthermore, in the background, YOLO generates fewer false positives in the background. The YOLO algorithm's structure consists of conventional neural networks, see Figure 2. YOLO begins by splitting the input image into SxS grids (S=13, S=26, and S=52), with B bounding locations predicted for each grid (B=3 for YOLOv3). Each boundary box includes many variables: x, y, w, h, box confidence score, and C class probabilities. The confidence score indicates the probability of an object being present in the box and the precision of the boundary box. x and y are cell offsets. The width w and height h of the bounding box are normalized by the image's width and height. The final output of each scale has a structure of (S, S, B×(5 + C)).

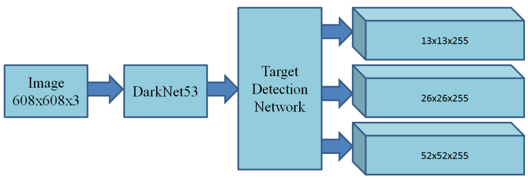

Figure 2. YOLOv3 architecture

2.3. SSD

The single shot detector (SSD) detector proposed by Liu *et al.* [22], runs at 19 FPS using Nvidia TiTan X graphic card. It's a feed-forward convolutional neural network consisting of a base net and an auxiliary architecture. The main aspect of SSD is that multiscale features are gathered to detect targets. SSD's main function is to predict category scores and box offsets for a predefined set of default bounding boxes by applying small convolutional filters to feature maps, followed by a non-maximum suppression step to produce the final detections. SSD like YOLO only needs one shot to identify different objects in an image, SSD allows more aspect ratios than YOLO. As a result, it can deal with objects of various sizes. The SSD network structure is shown in Figure 3.

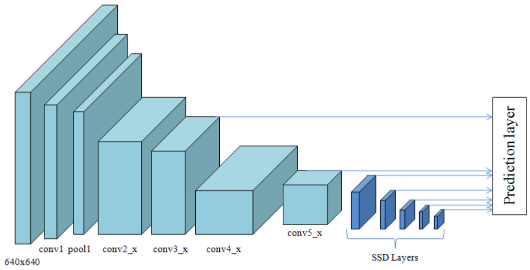

Figure 3. SSD architecture

3. RESULTS AND DISCUSSION

For our experiments, we used the common objects in context (COCO) database [23] for training and the daimler mono pedestrian dataset [24] for testing, experiments are executed on a computer with Intel Xeon CPU E3-1226 v3 Quard-core 3.3 GHz and 12 GB of RAM, Ubuntu 20 OS, Python 3.8.5, and Tensorflow 2.4.1 deep-learning framework. Our testing dataset consists of the following items: 4097 images captured from a vehicle at video graphics array (VGA) resolution (640x480) [25], and each image has a ground-truth [26] file indicating the real position of pedestrians existing in the image. For model evaluation we used: i) Time prediction represents the time that the model takes to predict the bounding boxes and the category class of objects; and ii) Average precision is a widely used metric for evaluating the accuracy of object detectors, it represents the surface under the precision-recall curve. In this paper, we calculated the average precision using the Cartucho code source [27]. Experimental findings presented in Table 1 reveal several interesting points: i) Faster R-CNN has better detection capabilities. however, is not suitable for real-time solutions; and ii) YOLOv3 is the best object detection model than SSD for Pedestrians in terms of detection and time prediction.

Table 1. Models performance

Model	AP(%)	Prediction time/image(s)
Faster R-CNN neural architecture search(NAS)	39.4	21.398288
SSD ResNet50 feature pyramid network(FPN)	22.68	0.822386
Yolo V3 DarkNet59	31.6	0.38899

4. CONCLUSION

In this paper, various types of pre-trained object detection models for pedestrians are implemented and tested. Results are compared using the performance parameters average precision, and time prediction. YOLOv3 is the best object detection model than others for pedestrians in terms of detection and time prediction. Results are very promising, but there are still some perspectives for our future research. Firstly, detect pedestrians in hard conditions (weather conditions and night vision). Secondly, investigating the impact of loss function on pedestrian detection models.

ACKNOWLEDGMENT

This work was supported by the University Hassan II Casablanca's Technology of Information and Communication Center as part of the Big Data & Connected Objects research project.

REFERENCES

[1] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," *Proceedings of the British Machine Vision Conference*, 2009, pp. 91.1-91.11, doi: 10.5244/C.23.91.

[2] J. Hua, Y. Shi, C. Xie, H. Zhang, and J. Zhang, "Pedestrian-and vehicle-detection algorithm based on improved aggregated channel features," *IEEE Access*, vol. 9, pp. 25 885-25 897, 2021, doi: 10.1109/ACCESS.2021.3057616.

[3] T.-W. Chu, S.-F. Su, M.-C. Chen, S. S.-D. Xu, and K.-S. Hwang, "Edge enhanced sift for moving object detection," in *2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS)*, 2016, pp. 11-14, doi: 10.1109/ICCSS.2016.7586414.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[5] S. Divya, L. P. Suresh, and A. John, "Image feature generation using binary patterns—LBP, SLBP and GBP," in *ICT Analysis and Applications*, Singapore: Springer, 2022, pp. 233-239.

[6] S. Lee, J.-W. Jang, and K.-R. Baek, "Pedestrian detection algorithm using a gabor filter bank," *Journal of Institute of Control, Robotics and Systems*, vol. 20, no. 9, pp. 930-935, 2014, doi: 10.5302/J.ICROS.2014.13.0020.

[7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1627-1645, 2010, doi: 10.1109/TPAMI.2009.167.

[8] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *2009 IEEE 12th international conference on computer vision*, 2009, pp. 24-31, doi: 10.1109/ICCV.2009.5459205.

[9] M. Razzok, A. Badri, Y. Ruichek, and A. Sahel, "Street crossing pedestrian detection system a comparative study of descriptor and classification methods," in *Colloque sur les Objets et systèmes Connectés*, 2019.

[10] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *2010 IEEE Computer society conference on computer vision and pattern recognition*, 2010, pp. 1030-1037, doi: 10.1109/CVPR.2010.5540102.

[11] M. Razzok, A. Badri, I. E. Mourabit, Y. Ruichek, A. Sahel, "A new pedestrian recognition system based on edge detection and different census transform features under weather conditions," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 2, pp. 582-592, 2022, doi: 10.11591/ijai.v11.i2.pp582-592.

[12] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *2008 IEEE conference on computer vision and pattern recognition*, 2008, pp. 1-8, doi: 10.1109/CVPR.2008.4587630.

[13] G. Monteiro, P. Peixoto, and U. Nunes, "Vision-based pedestrian detection using haar-like features," *Robotica*, vol. 24, pp. 46-50, 2006.

[14] T. Xiang, T. Li, M. Ye, and Z. Liu, "Random forest with adaptive local template for pedestrian detection," *Mathematical Problems in Engineering*, vol. 2015, p. 767423 , 2015, doi: 10.1155/2015/767423.

[15] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, 2021, doi: 10.1109/TNNLS.2021.3084827.

[16] T. Patil, S. Pandey, and K. Visrani, "A review on basic deep learning technologies and applications," in *Data Science and Intelligent Applications*, Singapore: Springer, 2021, pp. 565-573.

[17] L. Chen *et al.*, "Deep neural network based vehicle and pedestrian detection for autonomous driving: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3234-3246, 2021, doi: 10.1109/TITS.2020.2993926.

[18] Y. Mao, "A pedestrian detection algorithm for low light and dense crowd based on improved yolo algorithm," in *MATEC Web of Conferences*, 2022, vol. 355, doi: 10.1051/matecconf/202235503020.

[19] M. A. Feroz, M. Sultana, M. R. Hasan, A. Sarker, P. Chakraborty, and T. Choudhury, "Object detection and classification from a real-time video using ssd and yolo models," in *Computational Intelligence in Pattern Recognition*, Singapore: Springer, 2022, pp. 37-47.

[20] 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, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[21] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018, doi: 10.48550/arXiv.1804.02767.

[22] W. Liu *et al.*, "SSD: Single shot multibox detector," in *European conference on computer vision*, 2016, pp. 21-37, doi: 10.1007/978-3-319-46448-0_2.

[23] T.-Y. Lin *et al*, "Microsoft COCO: Common objects in context," in *European conference on computer vision*, 2014, pp. 740-755, doi: 10.1007/978-3-319-10602-1_48.

[24] S. Munder and D. M. Gavrila, "An experimental study on pedestrian classification," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 11, pp. 1863-1868, 2006, doi: 10.1109/TPAMI.2006.217.

[25] http://www.lookingatpeople.com/data/Daimler/pami09-enzweiler-gavrila/TestData_part1.tar.gz. (accesed May 5 2021).

[26] https://drive.google.com/file/d/1UTfXTZHlBa5GJYdjFAtljfRtTYfOoGZK/view?usp=sharing (accesed May 5 2021).

[27] J. Cartucho, R. Ventura, and M. Veloso, "Robust object recognition through symbiotic deep learning in mobile robots," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2336-2341, doi: 10.1109/IROS.2018.8594067.

BIOGRAPHIES OF AUTHORS

Mohammed Razzok received the Engineering degree in networks and telecommunications from the National School of Applied Sciences (ENSA), Oujda, Morocco, in 2017. He is currently pursuing his Ph.D. in the computer vision field at the Laboratory of Electronics, Energy, Automation and Information Processing, Faculty of Sciences and Techniques Mohammedia, University Hassan II Casablanca, Morocco. His main research is constructing and implementing on-board systems for intelligent vehicles to perform pedestrian recognition/detection, analyze pedestrian trajectories, warn the driver or initiate appropriate protective measures when detecting potentially dangerous situations. He can be contacted at email: mohammedrazzok@gmail.com.

Abdelmajid Badri holds a Ph.D. in Electronics and Image Processing from the University of Poitiers, France, in 1992. In 1996, he got the diploma of authorization to manage researches at the University of Poitiers France. He is a director at the Higher School of Technology (EST), Casablanca, and he is a University Professor (PES-C) at FST Mohammedia, University Hassan II of Casablanca, Morocco where he teaches electronics, signal processing, image processing, and telecommunication. He is a member of the electronic, energy, automatic and information processing (EEA and TI) laboratory which he managed since 1996. His research focuses on communication and information technology (electronics systems, signal/image processing, and telecommunication). He is qualified by the CNU-France in the 61st section: computer engineering, automatic, and signal processing. He managed multiple doctoral theses. He is a co-author of several national and international publications and is responsible for many research projects financed by the ministry or by the industrialists. He was a member of several committees and programs of international conferences and president of three international congresses in the same domain. He is a member and co-responsible for multiple scientific associations in touch with his domain of research. He can be contacted at email: abdelmajid_badri@yahooo.fr.

Ilham EL Mourabit holds a Ph.D. degree in Electronic Systems and Telecommunications in 2018. She is a University Professor at FST Mohammedia. She received the B.Sc. degree in Electronics and Computer Science, from the Faculty of Sciences and Technology of Mohammedia, Hassan II University Casablanca, Morocco, in 2009, and an M.Sc. degree in Electronic and Automatic Systems Engineering (telecommunication and information technologies specialty) from the same institution. Her main research areas are how to use satellite and cellular network's signals to determine the accurate position of a mobile target to perform real-time monitoring. She can be contacted at email: elmourabit.ilham@gmail.com.

Yassine Ruichek (Senior Member, IEEE) holds a Ph.D degree in control and computer engineering in 1997 and the authorization to manage researches (HDR) degree in physic science from the University of Lille, France, in 2005. In 2007 he became a full professor at the University of Technology of Belfort-Montbéliard (UTBM). His research interests include computer vision, image processing and analysis, pattern recognition, data fusion, and localization, with applications in intelligent transportation systems and video surveillance. He can be contacted at email: yassine.ruichek@utbm.fr.

Aïcha Sahel holds a Ph.D. in Electronics and Image Processing from the University of Poitiers, France, in 1996. She is a Professor at FST Mohammedia, University Hassan II of Casablanca, Morocco, and a member of the EEA and TI laboratory. Aicha Sahel's research focuses on electronic systems, signal/image processing, and telecommunications. She co-supervises doctoral theses and co-authors of several national and international publications and is a member of multiple financed research projects. She was a member of the steering committees of three international congresses in the same field of research. She can be contacted at email: sahel_ai@yahoo.fr.