Optimization of Loss Function for Pedestrian Detection

Shuo Zhang[1], Kailiang Zhang[1(✉)], Yuan An[1], Shuo Li[1], Yong Sun[1], Weiwei Liu[2], and Likai Wang[2]

[1] Jiangsu Province Key Laboratory of Intelligent Industry Control Technology, Xuzhou University of Technology, Xuzhou 221018, China
zhangkailiang@xzit.edu.cn

[2] Traffic Police Detachment, Xuzhou Police Bureau, Xuzhou 221002, China

Abstract. The advanced intelligent driving assistance system has improved the current traffic congestion to a great extent and effectively reduced frequent traffic safety accidents. Pedestrian detection technology is the core of autonomous driving technology, and its accuracy, real-time and complexity will directly determine the safe operation of autonomous driving. In the case of heavy traffic, detecting a single pedestrian in a crowd is still a challenging problem. Considering the problem of mutual occlusion between pedestrians in dense crowds, an improved function algorithm based on YOLOv3 is proposed to optimize the loss function and increase the accuracy of detection by replacing the anchor frame. Experimental results show that this method can effectively reduce the missed detection rate, increase the average accuracy, and help improve the effectiveness of pedestrian occlusion detection, ensure accurate pedestrian detection in traffic congestion scenarios, and ensure driving safety.

Keywords: Computer vision · Deep learning · Pedestrian detection · Loss function

1 Introduction

With the development of urbanization, the density of urban population and vehicles continues to increase, and the frequency of traffic accidents is gradually increasing. Therefore, how to effectively reduce traffic accidents and ensure traffic safety has become an urgent social problem to be solved. Some new network technologies and communication methods have been proposed [1, 2], network quality and experience have been continuously improved [3, 4], and intelligent driving systems have also been continuously improved. In road traffic, pedestrians are a disadvantaged group because they are more casual and active than car driving. Reducing the driving pressure of drivers and ensuring the safety of pedestrians are important goals in the design of urban traffic systems and intelligent driving systems to ensure pedestrian safety and reduce the occurrence of traffic accidents. The information around the car is collected through multiple sensors. The smart routing technology can transmit this information in real time [5, 6]. After intelligent

D. Jiang and H. Song (Eds.): SIMUtools 2021, LNICST 424, pp. 523–531, 2022.
https://doi.org/10.1007/978-3-030-97124-3_39

analysis and big data processing in cloud platform [7–10], the key information of pedestrian dynamics will be detected and fed back to the system. Then the system accurately provides the driver with road condition information, so that the driver can accurately and timely grasp the pedestrian movement information, and avoid pedestrian collision accidents caused by negligence. Of course, in an emergency, the driver does not need to operate the intelligent driving system and take mandatory safety measures to ensure that dangerous accidents can be stopped in time. The scene of pedestrian detection is very complicated. First, compared with vehicles, the postures of pedestrians are diverse and uncertain. Current pedestrian detection methods cannot fully adapt to these changes. However, based on the energy efficient transmission approaches, the complex algorithm can be deployed in mobile devices [11–14]. In a busy traffic environment, interference factors such as the mutual influence between characters, occlusion and changes in ambient light perception need to improve the accuracy of detection and recognition, and the algorithm structure is usually more complicated. Pedestrian detection technology is an indispensable key part in the field of self driving. It is a solid foundation to ensure driving safety and promote the development of intelligent transportation, and has important practical significance.

2 Related Work

So far, pedestrian detection technology has been developed for many years, and many research institutions and industries have achieved many outstanding research results. Some scholars [15] proposed a depth separable convolution model from the perspective of large background changes, overlaps or serious obstacles, which are optimized using different single-shot detector frameworks and have good reliability. Some scholars [16] proposed an improved R-CNN pedestrian retrieval framework, optimized the distance function, constructed the network using regions, and improved the hybrid similarity distance function. The experimental results show that compared with traditional methods, the learning ability has been enhanced, and has a certain driving performance. Some scholars [17] studied a pedestrian detection and recognition method based on deep learning from the perspective of active safety, which improved the speed and efficiency. Literature [18] proposed an integrated learning method and used structured learning optimization. At the same time, a visual feature extraction method based on spatial pooling is proposed. In the case of continuous translation, some improvements have been made in robustness. However, this method is still in the laboratory verification stage, and its application in actual scenarios has not yet been developed. Literature [19, 20] based on the detection model of YOLO, the network structure is light and the computing power is low, the detection efficiency is guaranteed, and there is a certain anti-collision function display, but the interference effect in complex scenes needs to be improved. With the development of deep learning, more and more scholars have gradually improved pedestrian detection [21, 22]. Whether it is the detection of small and medium-sized SSDs or the optimized PVANet to generate function maps, it will further improve accuracy and optimize performance. Literature [23] proposed a multi-class pedestrian detection network based on the fast R-CNN neural network, which detects different types of targets. In the training phase, multiple classification layers are defined. However, laboratory conditions have certain limitations. The methods and models proposed in [24–26]

are aimed at crowded pedestrian scenes. Pedestrian detection methods based on deep convolutional neural networks, Gaussian mixture models, HOG feature extraction and SVM classification can make full use of them. These methods continuously improve the performance of the model and can bring some optimization ideas, but they require constant running-in. The methods and models proposed by scholars [27, 28] have good performance in detection accuracy and processing speed. However, the impact of missed detection rate needs to be further optimized. In smart driving scenarios, pedestrian detection is an important task in applications such as monitoring driver assistance systems and autonomous driving. Detection and recognition models based on deep neural networks have been continuously proposed [24, 29]. In intelligent driving, an early warning mechanism needs to be established to ensure driving safety. In addition to the safety of the driver, pedestrians are also detected to ensure their safety. The efficiency of image processing and the accuracy of target detection are the first problems to be solved [30, 31]. In summary, a real-time and accurate pedestrian detection method is proposed, which is consistent with the development trend of intelligent driving. In particular, the security early warning mechanism in dense scenarios requires the higher recognition accuracy and lower time cost.

3 Model

The YOLO algorithm avoids the use of sliding windows for target detection. The feature is extracted through the Darknet-53 backbone network to obtain the image feature map with a size of N × N. YOLOv3 retrieves targets in three different scale feature maps. The range is determined to be 13 × 13 to detect large targets, the range is 26 × 26 to detect medium targets, and the range of 52 × 52 feature maps are responsible for searching and detecting other types. Through multiple detections of the target, the detection rate is improved. When predicting the bounding box, the dimensional cluster is used to anchor the box.

3.1 Darknet-53

The backbone network has been modified to Darknet-53, and its important feature is the use of Residual Network Residual. Its network structure is shown in Fig. 1.

It can be seen from Fig. 1 that there is only a convolutional layer in the model, and the size of the output feature map is controlled by adjusting the volume base layer step. There is no special restriction on the size of the input image. Using the idea of pyramid feature maps, small-size feature maps are used to detect large-size objects, while on the contrary, large-size feature maps are used to detect small-size objects. The output dimension of the feature map is N * N * [3 * (4 + 1 + 80)], where N * N is the number of grid points of the output feature map, there are three Anchor boxes, and each box has a 4-dimensional prediction box value, a 1-dimensional prediction box confidence, and the number of 80-dimensional object categories.

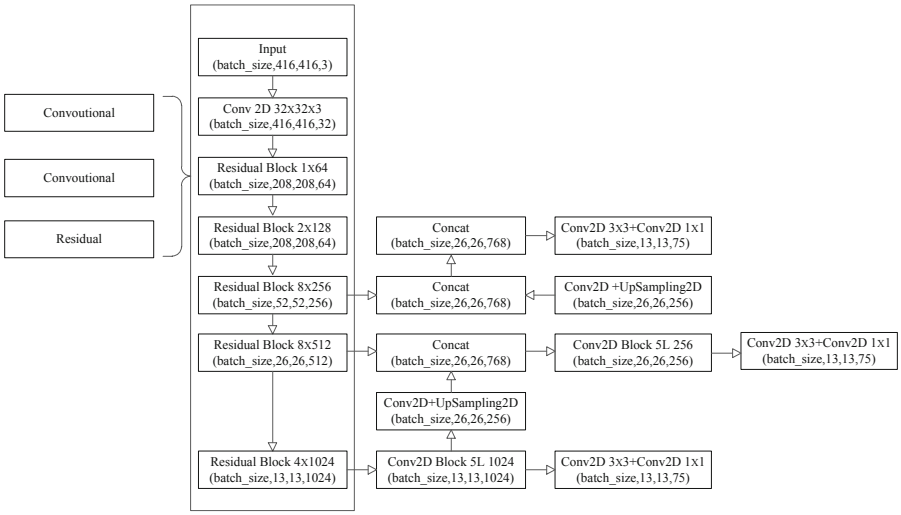

Fig. 1. Network Structure of Darknet-53.

3.2 Loss Function

In actual training, the loss function is lost by target positioning, position loss is used for translation operations, and size loss is used for retracting operations. For more pedestrians in dense scenes, Wang's team [37] can use the repulsive loss function to reduce the loss value, which refers to the characteristics of the magnet. $L = L_{Attr} + \gamma L_{\mathrm{Rep}GT} + \omega L_{\mathrm{Rep}Box}$ with L_{Attr} is to attract losses, said forecasting and loss between the real box, box with the real loss value is smaller that predict box the closer. In order to keep the prediction box as far away as possible from other target boxes and the prediction box of these target boxes which can affect it, the exclusion loss is proposed. The second part is the loss value produced by the prediction frame of the prediction frame and the adjacent real target, and the third part is the loss value produced by the prediction frame and the adjacent frame that are not predicting the same real target. Pass two correlation coefficients. Gamma and Omega are used to balance the loss of the second and third parts. γ and ω are the weight factors of repulsive force loss. The specific model is shown in Eq. (1).

$$L_{Attr} = \frac{\sum_{P \in P_+} Smooth_{L1}(B^P, G^P_{Attr})}{|P_+|} \tag{1}$$

In Formula (1), P represents candidate boxes, and represents all candidate boxes whose intersection with the real box is greater than the threshold value. G^P_{Attr} represents the real box with the maximum IOU with the candidate box, and B^P represents the prediction box regression from the candidate box. Finally, $Smooth_{L1}$ is used to establish the regression loss function, and the model is shown in Eq. (2).

$$L_{\mathrm{Rep}GT} = \frac{\sum_{P \in P_+} Smooth_{\ln}(IOG(B^P, G^P_{\mathrm{Rep}}))}{|P_+|} \tag{2}$$

In Eq. (2), G^P_{Rep} and B^P represent the real box and prediction box of other surrounding targets respectively. When the overlap is larger, the function value is larger. If IOG is used in the overlap calculation; there is only one way to reduce the loss function, that is, to reduce the overlap value between the prediction box and other surrounding target boxes, so that the two boxes are as far away as possible. In general, the fewer methods, the better the function optimization.

4 Function Optimization

In order to optimize the detection performance of YOLOV3 in dense crowds, we add two exclusions to the original YOLOV3 loss function. Specifically, in the increased rejection loss, the available loss function is included in these three parts, mainly to improve the target positioning loss. The repulsion function is aimed at the target detection algorithm based on the area candidate frame. The target detection algorithm based on the anchor frame is adopted here. The anchor frame is obtained by comparing the training data clustering. The positioning loss of the two algorithms in the loss function it is consistent with the attraction loss of Repulsion in the loss function, that is, the purpose is to make the prediction of the frame close to the true value of the match. In the original Repulsion loss function refers to the real match on G^P_{Attr} box, matching the real box is expressed as $G^P_{Attr} = \arg\max_{G \in g} IOU(G, P)$ which P said candidate box, G said real box, $g = \{G\}$ said all of the real target box in A picture. In Yolov3, anchor boxes are used instead of candidate boxes. We introduce the repulsive force loss $L_R\text{ep}GT$ into the loss function of Yolov3, with the purpose of moving the prediction box away from the real box of the surrounding targets. In order to be better applied in the algorithm, the following improvements are made to the exclusion loss function as $G^P_{Attr} = \arg\max_{G \in g} IOU(G, P)$, the loss of Repulsion, P for candidates, and only exist in the YOLOv3 anchor box, according to the repulsive force loss function design idea, the anchor box is similar to the original candidate box in the loss function. When the candidate box is replaced with the anchor box, the target box in the attraction loss is represented as the true box with the largest IOU in the anchor box. The function model is shown in Eq. (3).

$$G^A_{\text{Rep}} = \text{argmax}_{G \in g} IOU(G, A) \tag{3}$$

In Eq. (3), A is the anchor box, then the other surrounding real target boxes are expressed as $G^A_{\text{Rep}} = \arg\max_{G \in g}\{G^A_{Attr}\}$, so the new rejection loss function is shown in Eq. (4).

$$L_{\text{Rep}GT'} = \frac{\sum_{A \in A_+} Smooth_{\ln}(IOG(B^A, G^A_{\text{Rep}}))}{|A_+|} \tag{4}$$

In Eq. (4), the anchor frame representing all matched true frames. If the rejection loss is designed in this way, the following situations will occur: the direction of the adjustment of the prediction box has a great relationship with the position of the anchor box, and the selection of other real targets around it will have a large probability of vote difference. In the intelligent driving dense crowd scene, matching the targets that

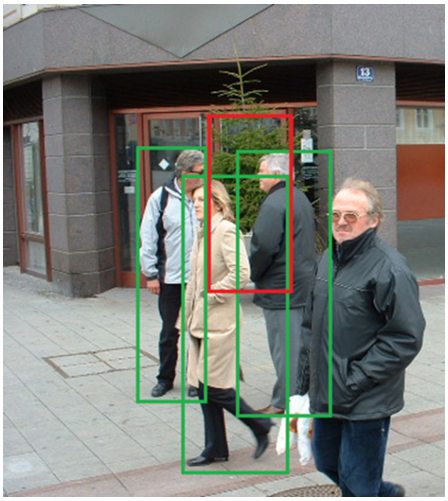

Fig. 2. Schematic diagrams of other real boxes around (Color figure online)

may exist around the anchor frame, this overlapping structure between them can be seen everywhere, and the result is shown in Fig. 2.

In Fig. 2, the green box is three real boxes, indicating that the red box matches the real target in the middle of the anchor box. If only the intersection of the anchor box and other real boxes is used to mark other real boxes, then the other real boxes on the right side of the pedestrian are the surrounding real boxes, so the exclusion loss can only be suppressed to intersect with the anchor frame on the right side of the box, but in the above case. The woman on the left side of the pedestrian frame will affect the reverse of the predicted frame, so the repulsive loss can only suppress the intersection with the anchor frame on the right side of the box.

5 Function Optimization

Caltech pedestrian data set is used in the training set, and the IOU threshold of the selected anchor frame is set to 0.5 during training. In which function $Smooth_{\ln}(x)$, parameter n = 1. The model trains 100 epochs. The trained models are tested in three subsets: Resonable, Occ = none and Occ = partical. In order to verify the effectiveness of the loss function, the original loss function is modified in different degrees. Firstly, the addition of the original loss function is the repulsive force loss of other real frames around the anchor frame that the prediction frame principle is close to the matched anchor frame, and the loss function is denoted as L_{RepGT}_tran01; Secondly, the prediction box is added into the loss function, which is far away from and close to other real boxes around the match on the real, and the loss function is denoted as L_{RepGT}_tran02; Finally, two repulsive force loss functions are added to the original loss function, denote as L_{RepGT}_tran. LAMR(log average miss-rate) values are compared as shown in Table 1.

Table 1. Comparison of LAMR values.

Subset	LOYO v3	L_{RepGT}_tran01	L_{RepGT}_tran02	L_{RepGT}_tran
Resonable	32.36%	31.20%	29.61%	27.28%
Occ = none	30.24%	29.12%	27.73%	26.27%
Occ = Partical	47.14%	45.89%	41.72%	40.73%

It can be seen from Table 1 that a single repulsive force loss can also effectively reduce the miss rate. In addition, L_{RepGT}_tran02 is more effective than L_{RepGT}_tran01 in reducing the rate of missed detection, while using the two loss functions together on the basis of the original loss function can greatly reduce the rate of missed detection. It is shown that in the training process of the model, when the prediction box is close to the matched anchor box and the prediction box is close to the real box, the performance of the detection can be slightly improved by adding the maximum, which indicates that the new loss function can effectively reduce the missed detection rate in the case of occlusion, which is in line with the original intention of the loss function.

6 Conclusion

In this paper, aiming at the problem of mutual occlusion between pedestrians in dense crowds in self driving scenarios, an improved function algorithm based on YOLOv3 is proposed. By replacing anchor frames, the loss function is optimized and the detection accuracy is improved. Experimental results show that this method can effectively reduce the missed detection rate, improve the average accuracy, and help improve the effectiveness of pedestrian occlusion detection, ensuring accurate pedestrian detection under traffic jams, and ensuring driving safety. The advanced intelligent driving assistance system has greatly improved the current traffic congestion and effectively reduced frequent traffic safety accidents. Pedestrian detection technology is the core of self driving technology, and its accuracy, real-time and computation complexity will directly influence the safety of self driving.

Acknowledgment. This work is partly supported by Jiangsu technology project of Housing and Urban-Rural Development (No. 2019ZD040) and Xu Zhou Science and Technology Plan Project (No. KC21309).

References

1. Zhang, K., Chen, L., An, Y., et al.: A QoE test system for vehicular voice cloud services. Mob. Netw. Appl. **26**, 700–715 (2019)
2. Chen, L., Jiang, D., Bao, R., Xiong, J., Liu, F., Bei, L.: MIMO scheduling effectiveness analysis for bursty data service from view of QoE. Chin J. Electron. **26**(5), 1079–1085 (2017)
3. Chen, L., et al.: A lightweight end-side user experience data collection system for quality evaluation of multimedia communications. IEEE Access **6**(1), 15408–15419 (2018)

4. Chen, L., Zhang, L.: Spectral efficiency analysis for massive MIMO system under QoS constraint: an effective capacity perspective. Mob. Netw. Appl. (2020)
5. Jiang, D., Wang, Z., Huo, L., et al.: A performance measurement and analysis method for software-defined networking of IoV. IEEE Trans. Intell. Transp. Syst. (2020)
6. Jiang, D., Wang, W., Shi, L., Song, H.: A compressive sensing-based approach to end-to-end network traffic reconstruction. IEEE Trans. Netw. Sci. Eng. **7**(1), 507–519 (2020)
7. Jiang, D., Huo, L., Song, H.: Rethinking behaviors and activities of base stations in mobile cellular networks based on big data analysis. IEEE Trans. Netw. Sci. Eng. **7**(1), 80–90 (2020)
8. Jiang, D., Wang, Y., Lv, Z., Qi, S., Singh, S.: Big data analysis based network behavior insight of cellular networks for industry 4.0 applications. IEEE Trans. Ind. Inform. **16**(2), 1310–1320 (2020)
9. Yang, B., Bao, W., Huang, D.-S.: Inference of large-scale time-delayed gene regulatory network with parallel MapReduce cloud platform. Sci. Rep. **8**(1) (2018). https://doi.org/10.1038/s41598-018-36180-y
10. Yang, B., Bao, W.: Complex-valued ordinary differential equation modeling for time series identification. IEEE Access **7**(1) (2019). https://doi.org/10.1109/ACCESS.2019.2902958
11. Jiang, D., Wang, Z., Wang, W., et al.: AI-assisted energy-efficient and intelligent routing for reconfigurable wireless networks. IEEE Trans. Netw. Sci. Eng. (2020)
12. Jiang, D., Huo, L., Zhang, P., et al.: Energy-efficient heterogeneous networking for electric vehicles networks in smart future cities. IEEE Trans. Intell. Transp. Syst. (2020). https://doi.org/10.1109/TITS.2020.3029015
13. Jiang, D., Wang, Y., Lv, Z., Wang, W., Wang, H.: An energy-efficient networking approach in cloud services for IIoT networks. IEEE J. Sel. Areas Commun. **38**(5), 928–941 (2020)
14. Jiang, D., Huo, L., Lv, Z., Song, H., Qin, W.: A joint multi-criteria utility-based network selection approach for vehicle-to-infrastructure networking. IEEE Trans. Intell. Transp. Syst. **19**(10), 3305–3319 (2018)
15. Ahmed, Z., Iniyavan, R., Madhan Mohan, P.: Enhanced vulnerable pedestrian detection using deep learning. In: 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 0971–0974 (2019). https://doi.org/10.1109/ICCSP.2019.8697978
16. Chen, E., Tang, X., Fu, B.: A modified pedestrian retrieval method based on faster R-CNN with integration of pedestrian detection and re-identification. In: 2018 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, pp. 63–66 (2018). https://doi.org/10.1109/ICALIP.2018.8455703
17. Song, H., Choi, I.K., Ko, M.S., Bae, J., Kwak, S., Yoo, J.: Vulnerable pedestrian detection and tracking using deep learning. In: 2018 International Conference on Electronics, Information, and Communication (ICEIC), Honolulu, HI, pp. 1–2 (2018). https://doi.org/10.23919/ELINFOCOM.2018.8330547
18. Paisitkriangkrai, S., Shen, C., van den Hengel, A.: Pedestrian detection with spatially pooled features and structured ensemble learning. IEEE Trans. Pattern Anal. Mach. Intell. **38**(6), 1243–1257 (2016). https://doi.org/10.1109/TPAMI.2015.2474388
19. Lin, S., Lin, M., Hwang, Y., Fan, C.: Deep-learning based pedestrian direction detection for anti-collision of intelligent self-propelled vehicles. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, pp. 387–388 (2019). https://doi.org/10.1109/GCCE46687.2019.9015528
20. Lan, W., Dang, J., Wang, Y., Wang, S.: Pedestrian detection based on YOLO network model. In: 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, pp. 1547–1551 (2018). https://doi.org/10.1109/ICMA.2018.8484698
21. Liu, S., Lv, S., Zhang, H., Gong, J.: Pedestrian detection algorithm based on the improved SSD. In: 2019 Chinese Control and Decision Conference (CCDC), Nanchang, China, pp. 3559–3563 (2019). https://doi.org/10.1109/CCDC.2019.8832518

22. Sun, W., Zhu, S., Ju, X., Wang, D.: Deep learning based pedestrian detection. In: 2018 Chinese Control and Decision Conference (CCDC), Shenyang, pp. 1007–1011 (2018). https://doi.org/10.1109/CCDC.2018.8407277
23. Zhang, J., Xiao, J., Zhou, C., Peng, C.: A multi-class pedestrian detection network for distorted pedestrians. In: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, pp. 1079–1083 (2018)
24. Ghosh, S., Amon, P., Hutter, A., Kaup, A.: Reliable pedestrian detection using a deep neural network trained on pedestrian counts. In: 2017 IEEE International Conference on Image Processing (ICIP), Beijing, pp. 685–689 (2017)
25. Luo, S., Qin, S.: Pedestrian detection of occlusion based on multi-marker method. In: 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, pp. 1033–1037 (2019)
26. Liu, T., Cheng, J., Yang, M., Du, X., Luo, X., Zhang, L.: Pedestrian detection method based on self-learning. In: 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, pp. 2161–2165 (2019)
27. Kim, D., Park, S., Kang, D., Paik, J.: Improved center and scale prediction-based pedestrian detection using convolutional block. In: 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, pp. 418–419 (2019)
28. Ayachi, R., Afif, M., Said, Y., Abdelaali, A.B.: Pedestrian detection for advanced driving assisting system: a transfer learning approach. In: 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, pp. 1–5 (2020)
29. Kulkarni, R., Dhavalikar, S., Bangar, S.: Traffic light detection and recognition for self driving cars using deep learning. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, pp. 1–4 (2018)
30. Kankaria, R.V., Jain, S.K., Bide, P., Kothari, A., Agarwal, H.: Alert system for drivers based on traffic signs, lights and pedestrian detection. In: 2020 International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1–5 (2020)
31. Hbaieb, A., Rezgui, J., Chaari, L.: Pedestrian detection for autonomous driving within cooperative communication system. In: 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, pp. 1–6 (2019)