ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 9, September 2019

Pedestrian Detection based on Improved Faster RCNN Algorithm

Nadigottu Sai Dinesh

Dept of Computer Science Engineering, Vageswari College of Engineering, Karimnagar, Telangana, India

ABSTRACT: The majority of the pedestrian detection techniques depend close by made features that produce low precision on complex scenes. With the improvement of profound learning strategy, pedestrian detection has made extraordinary progress. This paper exploits a convolutional neural network, which depends on the Fast R-CNN structure to separate powerful pedestrian features for proficient and compelling pedestrian detection in convoluted conditions. In this paper, the proposed model is trained on the Daimler pedestrian dataset and, at that point, tried on open Caltech and INRIA pedestrian datasets, which accomplishes the miss rate of 24.27% and 10.31%. The results demonstrate that the improved Faster RCNN network outperforms the cutting edge techniques.

KEYWORDS: Pedestrian detection, Faster RCNN, feature concatenation, hard negative mining

I. INTRODUCTION

Pedestrian detection innovation has been a very dynamic field in the investigation of insightful transportation and astute monitoring frameworks. Therefore, researchers pay more and more consideration for pedestrian detection, and a series of pedestrian detection techniques are emerging. Traditional strategies nearly utilize single or numerous underlying features for pedestrian feature extraction, including Viola-Jones (VJ) and Histogram of Oriented Gradient (HOG) [1], which are human-made models and afterward characterized by sliding window furthermore, linear support vector machine (SVM). This sort of strategy accomplished high accuracy on the MIT pedestrian dataset. However, these techniques do not work well on some new testing Caltech pedestrian datasets. The detection results require improvement. In recent years, a profound convolutional neural network (CNN) has made significant progress in computer vision undertakings, object detection with profound learning techniques consistently has better results. Researchers along these lines endeavor to handle pedestrian detection assignments utilizing some effective profound learning strategy.

In contrast to traditional computer vision approaches, profound learning techniques accomplish better performance on the Caltech pedestrian dataset due to dodging the plan inconveniences of hand-craft feature, and object detection with profound learning has some notable benchmark assessments. One of the profoundly generative frameworks for object detection is the region-based Convolutional Neural Network (R-CNN) technique, which augments CNN for understanding the generic item detection errands. Notably, training a deep neural network is convoluted because of how internal covariate move[2]can degrade the effectiveness of training. The internal covariate move is a difference in distributing the information sources, which occurs while training the feed-forward neural networks by changing a layer's parameters. Cluster normalization is a strategy to control the distributions of feed-forward neural network enactments, thereby reducing internal covariate move. So deep neural networks trained with clump normalization can converge faster and generalize better. This paper utilizes Fast R-CNN architecture dependent on group normalization(BN) to extract robust pedestrian features. In the pedestrian detection task, another vital stage is finding the potential windows which may contain pedestrians. As a thorough and traditional strategy, the sliding window technique has two primary shortcomings. First, it needs searching for every conceivable situation in a picture. Second, it might produce numerous redundancy windows which influence the nature of detection. To improve the detection effectiveness, the approaches of region proposals have been proposed to produce excellent regions, in which the most popular strategy is Selective Search [3]. It can get a high caliber of regions by utilizing picture division. However, its

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 9, September 2019

speed is very moderate, and the detection performance is delicate to the picture quality. The Selective Search technique for extracting competitor windows is infeasible but cannot provide precise confinement of pedestrians. Following the emerging trend of exploring deep learning for pedestrian detection on testing datasets, in this paper, our algorithm broadens the current Faster RCNN network architecture by joining feature concatenation, what is more, challenging negative mining strategies. We utilize a particular pedestrian softmax classifier rather than a K-object classifier. We direct a broad arrangement of experiments to assess the proposed plot on the public pedestrian dataset and accomplished the best in class performance[4].

II. PROPOSED METHOD

The Architecture For Pedestrian Detection

The architecture of our proposed Fast R-CNN dependent on group normalization strategy is illustrated in Fig1. It has seven layers. As the group normalization layer can normalize every scalar feature freely layer by layer, it maintains a strategic distance from the gradient disappear, and gradient detonates problem. Cluster normalization can prompt a faster convergence during the training and improve generalization performance as a regularization procedure[5].

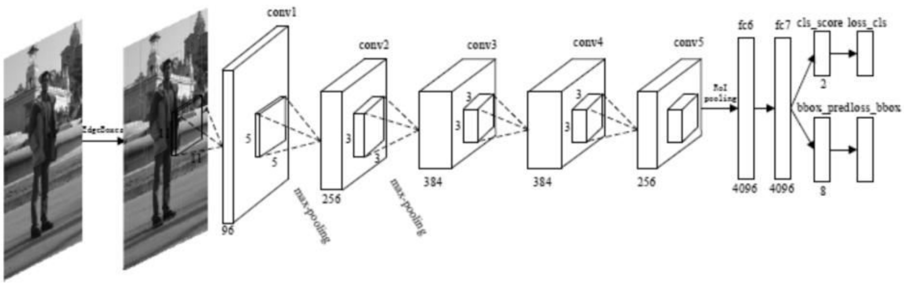

Fig 1: The architecture of our proposed pedestrian detection model

In our proposed architecture, the group normalization layer is set between each convolutional layer and the succedent ReLU layer, as illustrated. There are 96 kernels of size 11*11 with a stride of 4 pixels in the first convolutional layer, and in the following max-pooling layer, the kernel size is 3*3. The second convolutional layer takes as information the yield of the first convolutional layer, and there are 256 kernels of size 5*5 with a stride of 2 pixels. In the accompanying max-pooling layer, the kernel size is equivalent to the first layer. For the following two (the third and the fourth) convolutional layers, the corresponding pooling layers are precluded, and both contain 384 kernels.In the fifth convolutional layer, there are 256 kernels, and the accompanying pooling layer is a region of interest (RoI) pooling layer. The RoI pooling layer is used to pool each information object proposal's feature guides into a fixed-length feature vector, which is then taken care of into the completely associated layers. The following two layers are entirely associated layers, the two of which contain 4096 hubs[6]. Toward the architecture's finish, two yield layers produce two yield vectors per object proposal. In particular, one is a softmax layer, which yields grouping scores over K object classes in addition to a "background" class. The other is a bounding box regressor layer, which yields four real-esteemed numbers for every one of the K object classes.

Overview of the Proposed Model

The proposed pedestrian detection model follows the similar deep learning framework of Faster RCNN, which has been demonstrated to be a best in class in-depth learning plan for object detection. This framework primarily comprises of the Region Proposal Network (RPN) and the Fast RCNN network.

The region proposal network is mostly used to generate a series of interest (ROIs) containing objects. The Fast RCNN network is mostly used to arrange items (and background) and refine those regions' boundaries. The two networks of Faster RCNN share the parameters of the convolution layer got by the process of feature extraction, permitting this

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 9, September 2019

model to accomplish object detection assignments at a relatively quick speed. In our work, we propose to broaden the architecture of Faster RCNN for pedestrian detection and introduce feature concatenation and hard negative mining towards higher recall and accuracy[7].The process of training our pedestrian model. First of all, we train the CNN model utilizing the Daimler pedestrian detection benchmark dataset. We acquire hard negatives utilizing the equivalent dataset by testing the pre-trained model. To generate fewer bogus positives, these got hard negatives are sent into the network as the second step of the training process. In the last process, the resulting model will be tweaked on the Caltech pedestrian dataset.

Hard negative mining

The core thought of hard harmful mining is group tests with classifiers and utilizing the misclassified tests as negative examples to keep training the classifier [8]. We consider these negative examples and send them into the network, so the classifier will have the option to generate fewer bogus positives and accomplish better characterization performance. Hard negative mining is received as reinforcement for improving the viability of our training process. In our work, we train the model on the Daimler pedestrian detection benchmark dataset. During the first step of the training procedure, the region will be treated as hard negatives if itsintersection over association (IoU) over the ground truth region is under 0.5.

Feature concatenation

In the original Faster RCNN network structure, RoI pooling is performed on the feature guide to extract feature vector of fixed size about each item. Yields are further taken care of into the arrangement part of networks. This astute arrangement accomplishes the purpose of reducing plenty of count costs. However, such an approach may overlook several essential features of the picture. We found that the external network extracts texture and detail features through the technique for feature graph representation, which contains more features and the capacity to extract critical features[9]. The deep network extracts contour shapes and striking features, so the extracted features are more representative. According to the feature map characteristics, we join low-level and elevated level features in numerous convolution layers. To be more explicit, we extract the feature guides of different convolution layers separately. These feature maps are pooled in the RoI pooling layer, and the pooling result is L2-normalized.

III. EXPERIMENTS AND ANALYSIS

The pedestrian detection benchmark dataset at the first step of the training procedure and pedestrian detection is assessed utilizing the notable Caltech and INRIA pedestrian dataset. In this segment, we first introduce two datasets and describe our experimental subtleties. Lastly, we compare our model with several astounding techniques.

Caltech

The Caltech pedestrian dataset has gathered nearly 10 hours of 30Hz video (around a million frames), and a camera takes the video on the car in an urban scene. The dataset has chosen 137 minutes of video, nearly 250,000 frames of pictures, which contained around 2,300 pedestrians, and the author marked 350,000 jumping boxes. This dataset is rich in information and is currently the most extensive pedestrian data set[10].

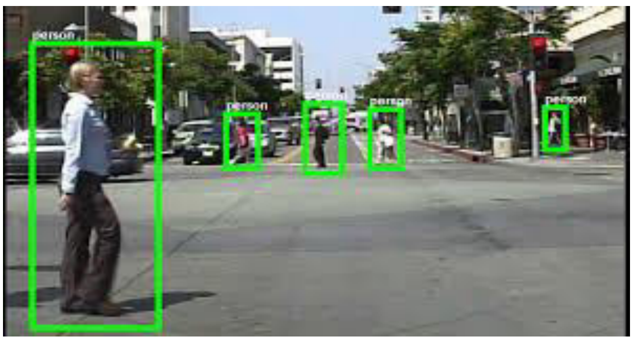

Fig 2:pedestrian detection for Caltech

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 9, September 2019

INRIA

The INRIA pedestrian dataset is currently the most utilized static pedestrian detection information base, part of the training, and a testing set. The training set comprises of 614 positive pictures and 1,218 negative pictures. The testing set comprises of 288 pictures.To grow a complimentary model set, we map every region proposal to the ground-truth, which has a high intersection over association (IoU). If the IoU is more extensive than 0.5, at that point, we will mark the chose region proposal as a positive example for the pedestrian class[11]. The rest region proposals, whose IoUs are smaller than 0.1, are regarded as negative examples.

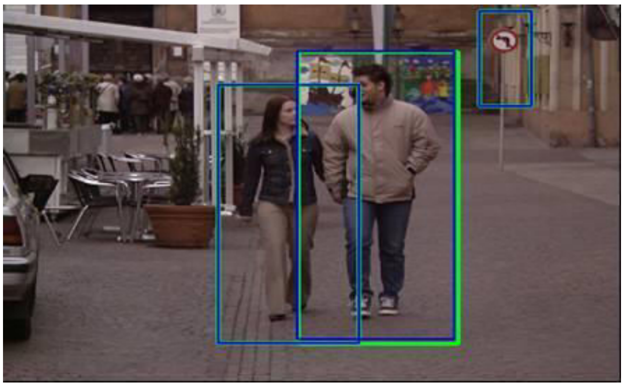

Fig 3: Pedestrian detection ForINRIA.

Implementation Details

We lead thorough assessments of the proposed approach and investigate its performance on generally utilized benchmarks, including Caltech,INRIA. Furthermore, we pick a log-average Miss Rate on False Positive Per Image(FPPI) as the assessment metric since this assessment strategy is closer to the classifier's genuine circumstance.The experimental results are presented in Fig.4. Instances of pedestrian detection utilizing the proposed technique on the Caltech testing set[12].We compare the proposed strategy with several strategies that accomplished the Caltech testing set's best performance, including VJ, HOG, ACF, LatSvm-V1, Shapelet, HikSvm, RandForest, and Multi Ftr+ CSS.

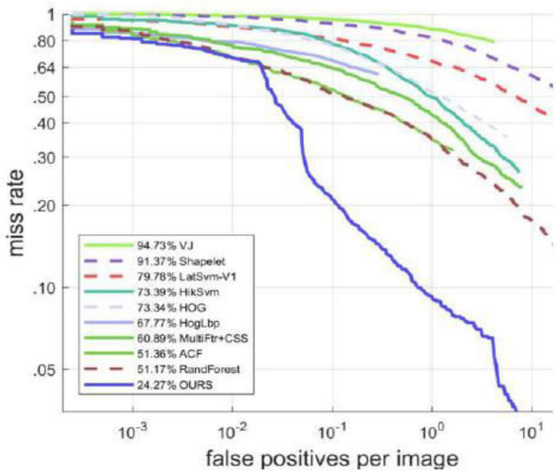

Fig 4: Performance comparison of several notable pedestrian detection strategies on the Caltech pedestrian dataset.

To demonstrate our approach's adequacy, we train the proposed model on the INRIA training set and assess the model on the INRIA testing set. For the INRIA dataset, our strategy's miss rate is 10.31%, which outperforms the top-performing strategy[13]. Fig.5 shows the miss rate of our strategy assessed on the INRIA testing set.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 9, September 2019

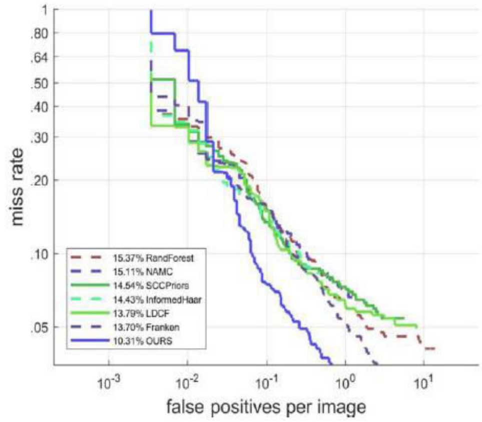

Fig 5: Performance comparison of several notable pedestrian detection techniques on INRIA pedestrian dataset

There is no uncertainty that the proposed strategy has self-evident preferences compared with other notable techniques. Experimental results show that our network architecture is ready to extract features and achieve pedestrian detection better than other detection algorithms.There are plenty of factors that could result in degrading performance. For model, if we train our model utilizing all provided comments, many of the marks will confound the feature extractor since the jumping confines these areas are not precise. Of course, the learning technique can additionally be improved from numerous perspectives.

III. CONCLUSION

In this work, we proposed another strategy for pedestrian detection utilizing a deep convolutional neural network. Specifically, we expand the Faster RCNN framework for generic item detection and introduce two successful strategies for improving pedestrian detection performance, including feature concatenation and hard negative mining. We directed a series of experiments on a well-known pedestrian dataset. Experimental results demonstrate that the proposed model outperforms the best in class detection algorithms.

REFERENCES

[1] S. Bell, C. Lawrence Zitnick, K. Bala, R. Girshick, Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, 2874 - 2883.
[2] C. Szegedy et al. Going deeper with convolutions. In Proceedings of Computer Vision and Pattern Recognition (CVPR), Boston, MA, June 2015, 1 - 9.
[3] VISHAL DINESH KUMAR SONI. (2019). ROLE OF ARTIFICIAL INTELLIGENCE IN COMBATING CYBER THREATS IN BANKING. *INTERNATIONAL ENGINEERING JOURNAL FOR RESEARCH & DEVELOPMENT*, *4*(1), 7. HTTPS://DOI.ORG/10.17605/OSF.IO/JYPGX
[4] Xiaogang Chen, Pengxu Wei, Wei Ke, Qixiang Ye, and Jianbin Jiao. Pedestrian detection with a deep convolutional neural network. In Computer Vision-ACCV 2014 Workshops, pp. 354–365, 2014.
[5] AnkitNarendrakumarSoni" Spam e-mail detection using advanced deep convolution neural network algorithms " Volume-2, Issue-5 (May-2019), Journal For Innovative Development in Pharmaceutical and Technical Science ISSN (O):- 2581-6934.
[6] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137 - 1149.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 9, September 2019

[7] Vishal DineshkumarSoni. (2019). FACE DETECTION BY IMAGE DISCRIMINATING. *International Engineering Journal For Research & Development*, *4*(6), 7. https://doi.org/10.17605/OSF.IO/D7KEW
[8] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (9): 1904 -1916.
[9] AnkitNarendrakumarSoni" Crack Detection in buildings using convolutional neural Network " Volume-2, Issue-6 (June-2019), Journal For Innovative Development in Pharmaceutical and Technical Science ISSN (O):- 2581-6934.
[10] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning, ”In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3626 – 3633.
[11] Vishal DineshkumarSoni. (, 2018). Artificial Cognition for Human-robot Interaction. *International Journal on Integrated Education*, *1*(1), 49-53. https://doi.org/10.31149/ijie.v1i1.482
[12] Jianan Li, Xiaodan Liang, ShengMeiShen, TingfaXu, and Shuicheng Yan. Scale-aware fast r-CNN for pedestrian detection. arXiv preprint arXiv:1510.08160, 2015
[13] AnkitNarendrakumarSoni " Application and Analysis of Transfer Learning-Survey" Volume 1 Issue 2, Nov-Dec2018, International Journal of Scientific Research and Engineering Development (IJSRED) :ISSN:2581-7175,