Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

A Novel Deep Learning Model Based on Yolo-V2 and Resnet for Pedestrian Detection

Kumar P[1*], Swaminathan B[2], Karthikeyan U[3]
[1,2,3]Rajalakshmi Engineering College, Chennai, India
*kumar@rajalakshmi.edu.in

ABSTRACT

Person on foot discovery is one of the significant errands in item location innovation. The person on foot identification calculation has been utilized in applications like shrewd video reconnaissance, traffic examination, and self-governing driving. Lately, numerous passer by discovery calculations have been proposed however the key downside is the precision and speed, which can be improved my coordinating productive calculations. The proposed model improves the person on foot identification calculation by incorporating two proficient calculations together. The model is created utilizing the joint rendition of ResNet and YOLO v2, which preforms highlight extraction and grouping separately. By utilizing this model, the effectiveness of the framework is expanded by improving the precision rate so it tends to be utilized with constant applications. The model has been contrasted and existing models like SSD, Faster R-CNN and Mask R-CNN. Contrasting and these models, the proposed model gives mAP esteem higher than these current models with less misfortune work when tried on the INRIA dataset.

Keywords
mAP - Mean Average Precision, R-CNN – Region-based Convolutional Neural Network, ResNet – Residual Neural Network, SSD - Single Shot Detector, YOLO - You Only Look Once.

Introduction

Article location is the PC vision and picture handling innovation that recognizes and characterizes the items. Item recognitions join picture arrangement and article restriction undertakings for recognizing the items from a picture which can likewise identify different items from a picture. Picture order is utilized for anticipating the class of the item in a picture. Article restriction is to find at least one items present in the picture and to find them utilizing a jumping box. This interaction can be utilized for recognition objects for self-ruling vehicles, which contain classes like vehicles, trucks, walkers, and so forth. By utilizing this in self-governing vehicles it assists with limiting the specific sensors in the vehicle. Person on foot recognition isn't just utilized for independent vehicles yet in addition it is a significant article in man-made consciousness since it is the critical hotspot for the machine to collaborate with. Walker location is utilized in some continuous applications for communication, security reason and significantly more. A portion of the person on foot identification applications are self-governing driving, traffic examination, astute observation, modern computerization, and so forth

The development of item location is so far improved from the conventional strategies, which limits the quantity of assessment steps in the current procedures and furthermore more proficient. In the customary technique, a fixed sliding window is utilized to slide from left to right and through and through in the picture to confine the item in the picture at various areas. After this cycle, a picture pyramid is utilized to recognize objects at different scales. After the consummation of these two cycles, the Region of Interest (ROI) is extricated and taken care of

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

into the convolutional neural organization. For every assessment of these means, in the event that the characterization likelihood is higher than the edge esteem, at that point a jumping box is utilized to find and mark the article in the picture. At long last, a non-max concealment is applied to the bouncing boxes to conclude the last anticipated or identified article in the picture.

The second strategy for object recognition is by utilizing a pre-prepared organization. The pre-prepared organization is an organization that is now prepared utilizing a bunch of informational collections for some other issue explanation. By utilizing this pre-prepared organization can utilize it for preparing the model, by utilizing the equivalent dataset or by utilizing own dataset. This strategy is so viable on the grounds that it isn't important to begin without any preparation for building a model. The lone thing need to do is to modify the model as indicated by the difficult assertion. Additionally, the arrangement part in the organization can be changed by eliminating the last completely associated layer and change it to the grouping calculations that required. A portion of the pre-prepared organizations are ResNet, VGG16, MobileNet and BaseNet. The order part can be adjusted to R-CNN, YOLO, and so on One of the benefits of utilizing profound learning calculation is neural organizations are equipped for extricating highlights itself with no extra cycle for it.

Artificial intelligence is set to transform all aspects of our lives – including our workplaces, homes and vehicles. AI tools are already widely familiar in Internet-searching, computers with speech recognition and games such as chess, but the next few years will see AI become ever more widespread, in everything from cars to robots to medicine. This will have significant repercussions for society, as AI performs many tasks that, until now, have been done by humans. As harder the applications become their need an approach better than the previous. In that case, Machine learning came into picture. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. Though machine learning is fast and advanced than AI, to deal with large datasets it needs large number of neural networks to work on it efficiently. In this case their booms a subset to machine learning called Deep Learning. Artificial intelligence uses deep learning models than machine learning for better decision making. Deep Learning attempt to model high-level abstractions in data by employing deep architectures composed of multiple non-linear transformations. The use of CNNs along with deep learning can bridge the semantic gap between detection of objects by the human brain and an Artificial Intelligence.

By knowing the viability and of profound learning, we endeavor to construct the framework utilizing profound learning methods. The choice of profound learning for this is, to deal with huge dataset as it is a person on foot identification application it includes huge volume of information and more profound the organization the precision and the exhibition of the application increments.

In this proposed framework, the pre-prepared ResNet model is utilized as an element extractor and the model is prepared utilizing the INRIA dataset. The characterization part in the organization is taken out and supplanted with the YOLO v2 network. The explanation behind YOLO v2 is it is the solitary organization in the YOLO family to recognize even the more modest articles from any side of the picture. Thus, by utilizing these two organizations together the precision of discovery can be expanded and the model can be utilized for ongoing applications.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

Literature Review

Various models in profound realizing where proposed and created for object recognition lately. Existing models have different calculations for both element extraction and characterization. Some model does both the measure utilizing single calculation which is the upside of utilizing Deep Learning. The current calculations for object discovery are examined underneath.

The different techniques that are utilized for extricating highlights are Haar-like [16], HOG [7], [10], [16], Gaussian models which are tangle lab highlights for extraction of highlights. The CNN strategies are LDCF [15], ZFnet [12]. Furthermore, after the development of profound catching on Quickly R-CNN [11] and YOLO [8], [11] were broadly utilized. Hoard is the generally utilized strategy for include extraction. Hoard gives an exact outcome to person on foot identification. Furthermore, alongside HOG, CSS [10] and Haar-like [16] are utilized for removing highlights. Different techniques for removing highlights are, Selective Search [13] which incorporates the component of comprehensive hunt and division. It bunches comparable districts utilizing shape, shading, and surface. Move learning [9] is utilized for the framework for highlight extraction and grouping. In this, the model prepared for another issue is utilized for tackling this difficult assertion by utilizing the information from it. A portion of the pre-prepared models for preparing the organization are ImageNet, ResNet, and so on When constructing the framework with profound learning similar strategies are generally utilized for both component extraction and grouping. Quick R-CNN [11] and YOLO [2], [11] are of those sort, utilized for both element extraction and order.
A novel ensemble deep network called collaborative deep networks where multiple deep networks are combined in a fully-connected network was used by Hongmeng Song, Wenmin Wang (2017), for maximizing the abilities of the model, a resembling process was incorporated to prepare diverse datasets and pre-train them to gain the ability of feature representation. The strategies for Classifications are SVM [7], [10], [16], AdaBoost [16], and K-implies bunching. These are the most well-known strategies utilized in a specific period. The possibility of SVM classifier calculation is straightforward; it makes a hyper plane what isolated two classifiers. AdaBoost or Adaptive boosting utilizes the troupe learning strategy to characterization. This joins numerous week classifiers to make it as a solid classifier. Zahid Ahmed, R. Iniyavan[1] proposed a method for achieving both accuracy and fast enough for real-time deployment. These requirements are addressed through the usage of depth wise separable convolution and Single Shot Detector framework employing different activation maps using OpenCV to achieve a reliable, robust and competent deep learning based pedestrian detection for real time operations. The Convolutional network architecture proposed for this is MobileNet. Thus, by the combination of MobileNet and SSD the system provides high accuracy and speed for real-time applications.
Finally, a collaborative learning method is presented to train the entire model. The performance of UDN model can be improved by 2.05% on largest Caltech dataset and 1.36% on ETH dataset which both outperform relevant approaches.

Proposed Method

The model contains a neural organization with ResNet as a component extractor and YOLO v2 for order. The decision of these two calculations is to improve the proficiency of the article

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

identification framework by expanding the exactness through recognizing significantly more modest items precisely. A neural organization is made utilizing the ResNet network which does the component extraction measure. ResNet is utilized for its better learning and precision with more profound organizations. The debasement issue happens in more profound organizations which soaks the exactness of the model by rehashing a similar cycle and again at more significant levels. To defeat this issue the ResNet model is utilized, which stays away from a stage which is handled more than twice, so precision may not immerses even with more profound organizations. What's more, for characterization, the completely associated network is changed into YOLO v2 network. YOLO v2 is a superior organization among neural organizations to identify more modest articles precisely. It is superior to YOLO v1 in exactness and better than YOLOv3 in speed. The model is prepared utilizing the INRIA dataset with both positive and negative pictures. The picture surrendered to the model will be separated into N x N matrix cells and will be alloted for removing highlights in it. At that point the removed element will be utilized in the characterization interaction for recognizing the item in the given picture. When the article is recognized it will be stamped utilizing a bouncing box. The precision of discovery of an article is discovered utilizing the distinction between the genuine jumping boxes versus anticipated bouncing box which is known as Intersection over Union. Mean Average Precision or Average Precision is a measurement to figure the exactness of the article identifiers. Exactness quantifies how precise the forecast was. The model was prepared and tried utilizing the INRIA dataset with both positive and negative pictures. The forecast with IoU more prominent than or equivalent to 0.75 is considered as the best expectation and it is distinguished utilizing the bouncing boxes. The further clarification of the organization has been portrayed in Section IV.

A. SYSTEM DESIGN

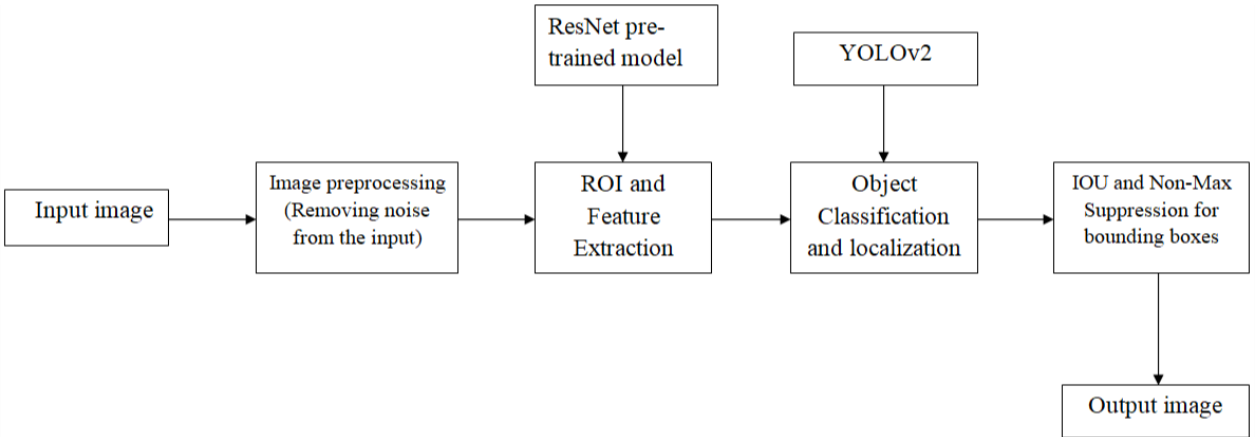

Fig 1: Pedestrian detection using deep learning architecture

As shown in fig 1, the input for the system is provided as an image or the video frames. In real-time applications, the input can be from the webcam of self-driving cars and camera footages in case of intelligence surveillances. The pedestrian detection system was built using deep learning techniques, and various datasets of the pedestrian are provided for the system for training and testing. As the inputs are assigned to the system, the first process of the system is to preprocess the image. That is, the noise and unwanted portion from the image are removed. The noise-free images are assigned for the next step, where the images are assigned for the ROI module (Region

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

of Interest). In this, the location of the objects from the image is marked or bounded for making the feature extraction process easy. Instead of processing the whole image, by using ROI the time and the area for processing are minimized, which will help find the objects at exact locations. The next step is feature extraction process.

The proposed method for feature extraction is ResNet method where the model is trained by the pre-trained ResNet model for extracting the features from the image. The advantage of selecting ResNet for Feature extraction is it provides high accuracy rate. ResNet provides accuracy better than any other Convolutional architecture. After feature extraction the images are assigned for the classification process. The proposed method uses YOLOv2 network for classification. The advantage of selecting YOLOv2 for classification is it better than YOLO v1 by the accuracy rate and better than YOLO v3 by the speed. YOLO v2 can detect any size object better than other YOLO algorithms. If the pedestrian is found in the image at the end of classification process, the YOLO v2 algorithm detects the pedestrian using the bounding boxes. And, for mapping pedestrian using bounding boxes, it uses Non-max suppression method and Intersection over Union (IoU) for accurate marking. Thus, finally the pedestrian from the given image is detected and marked using the bounding boxes.

The organization until completely associated layer is adjusted utilizing the ResNet organization, so that highlight extraction is finished utilizing ResNet and YOLO v2 plays out the order interaction. Initial, a convolutional neural organization with ResNet as a base is made. The CNN contains numerous convolutional in addition to ReLU layer and pooling layers on the other hand. These layers are the structure squares of the convolutional neural organization. The convolutional layer applies a channel to the information pictures and results in a guide of initiation. After each aftereffect of the convolutional layer, the guide of actuation or the component map demonstrates the strength and areas of the distinguished element in the info picture and predicts the class to which it has a place. ReLU in the convolutional layer is utilized to change the non-straight properties over to direct. The vast majority of this present reality information is non-direct, so to manage certifiable information ReLU layer is utilized to change it over to straight for additional cycle. Another structure square of CNN is the pooling layer, which logically decreases the spatial size of the contribution to diminish the calculation and boundary of the organization, keeps up the main data. The last layer in the convolutional neural organization is completely associated layers, this layer is sub-separated into three layers in particular completely associated input layer, first completely associated layer and completely associated yield layer. In a completely associated input layer, the yield of the past layer is given as info, which straightens and changes over them into a solitary vector for making them as a contribution to the following layer. This single vector esteems are given as contribution to the main completely associated layer, which takes inputs and applies loads for the component investigation to anticipate the marks. At last, the completely associated yield layer gives the last likelihood for each name.

B. MODULE DESCRIPTION:

The entire methodology for building up the pedestrian detection framework is divided into five modules. The modules are:

1. Dataset collection and pre-processing
2. Building a neural network
3. ROI Module (Region of Interest).
4. Feature Extraction
5. Classifying and Detecting Pedestrian

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

The pedestrian datasets where gathered from INRIA passerby dataset and from Kaggle dataset. The gathered datasets where pre-handled for eliminating the clamor, undesirable locales from the picture where taken out and making the pictures widening for making the items to be find effectively for preparing. In spite of the fact that the pre-prepared ResNet model is being utilized for building up the neural organization, the dataset for preparing the model has been taken from INRIA person on foot dataset and prepared the pre-prepared model for redoing it to the proposed framework. The INRIA passerby dataset contains 15560 positive pictures and 6744 negative pictures for preparing and testing the model. All the pictures are pre-prepared and set for preparing the organization. The aftereffect of the framework totally dwells on how the organization has been prepared and with which sort of data sources. So the dataset ought to be adequate to prepare the organization and should prepare the organization with all potential ways for confronting another picture for location so issues like overfitting won't happen while testing the model.

The following module in building up the framework is building up the neural organization. For building a neural organization the loads and boundaries for the organization have been set. In this, the loads and boundaries of the ResNet model have been utilized for building the model. The pre-handled datasets are given has contributions to prepare the model. At that point the sigmoid is the initiation work utilized here for building the organization. After the finish of single forward proliferation, the organization begins learning and in the end it results with a misfortune work. To limit the misfortune work, the back spread strategy can be utilized to return to the past layers and can change the loads in like manner so the misfortune capacity can be limited. This can be rehashed until the organization is prepared up to the normal level.

The pre-handled pictures are doled out for Region of Interest. In this, all the potential areas of the articles are found from a picture utilizing pooling cycle and imprint the discovered item utilizing the bouncing boxes. This cycle is accomplished for making the further interaction to be finished utilizing a specific zone where the articles are situated, rather than preparing the entire picture. Also, for choose whether the picture has a place with class or to the foundation class. Highlight extraction is for extricating the highlights of the article for arranging them. Highlight extraction diminishes the quantity of assets expected to depict an enormous dataset. By extricating the element, the framework realizes a little about what the article is. So during arrangement it contrasts the article and less number of dataset that is just with the most extreme coordinating datasets as opposed to contrasting and the entire set. The benefit of profound learning is; it does the component extraction measure without help from anyone else. The thing to done is to prepare our datasets with the pre-prepared ResNet and to utilize the extricated highlight to arrangement measure.

The subsequent stage is separating the highlights from the pictures. One of the benefits of utilizing profound learning methods is, the base organization or neural organization itself does the component extraction measure and needn't bother with any extra interaction for it. The interaction that happens inside the organization for extricating highlights is the district of interest technique, which is done prior to separating highlights. By utilizing this technique, the area where the article dwells in the picture can be taken independently for the extraction interaction so the space should have been prepared can be limited when contrasted with handling the entire picture. So the time needed for the extraction cycle will be limited.

The last module in building up the framework is the arrangement and recognition measure. Subsequent to extricating the highlights of the pictures, they are allocated for grouping measure.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

YOLOv2 network is utilized for arrangement measure. The upside of choosing YOLOv2 for characterization is, it gives more exactness rate than some other CNN design. For continuous applications we need a calculation which recognizes even little articles from the info gave. All things considered YOLOv2 gives a decent identification rate. Likewise, YOLOv2 is quicker than YOLOv1 and YOLOv3. The certainty score for the pictures with the removed element is determined utilizing the accompanying eight qualities, which is additionally clarified in section 4.

Results and Discussions

The INRIA dataset contains two arrangements of passerby pictures: Static person on foot pictures. In both, it contains the first pictures with explanations and positive pictures in 64 x 128-pixel design alongside unique negative pictures. The positive preparing set contains 1208 pictures and the test set contains 566. The negative preparing set contains 1218 pictures and a negative test set 453 pictures. MIT dataset contains 709 persons on foot pictures taken in city roads. Each picture is 64 × 128 pixels and contains either a front or a back perspective on a focused, standing individual. The scope of postures is generally restricted, and individuals are standardized to have roughly a similar size in each picture. It contains 509 preparing and 200 test pictures. MIT informational index doesn't contain negative pictures, so we utilized the negative pictures from the new INRIA static informational collection. KAGGLE dataset contains the passerby dataset with 785 pictures and doesn't contain any negative pictures and that can be taken from the INRIA dataset. The pictures of the walker in this is of strolling pictures of people on foot and intersection the crossways. This dataset is compelling for the interaction since it contains the exceptionally edited pictures of the passerby so pre-preparing is simple.

The examination has been finished utilizing the INRIA individual dataset as the preparation tests. 80 % of the dataset is utilized for preparing the model and rest 20 % is doled out for testing the model. For the proposed model, a pre-prepared ResNet model is taken and prepared the model utilizing the INRIA dataset. The model has been adjusted utilizing the YOLO v2 network. The organization is prepared utilizing an enormous volume to walker information with the goal that the organization can recognize any common information later on for application purposes.

For the characterization part the YOLO calculation has been utilized. In that, the picture will be partitioned into N x N network cells and will be appointed for removing highlights in it. At that point the removed component will be utilized in the order cycle for distinguishing the item in the given picture. The arrangement interaction is continued by YOLO v2. The certainty score for the pictures with the extricated highlight is determined utilizing the accompanying eight qualities,

Y = pc, bx, by, bh, bw, c1, c2, c3.

pc – Represents whether an object is present in the frame or not. If present pc=1 else 0.

bx, by, bh, bw – are the bounding boxes of the objects (if present).

c1, c2, c3 – are the classes. If the object is a car then c1 and c3 will be 0 and c2 will be 1. These qualities will be again rehashed if there happen numerous articles in a solitary casing. The pictures with the presence of items are taken and the articles are identified utilizing the jumping boxes. IOU and non-max concealments and used to distinguish the item precisely. That is, when the item is distinguished it will be stamped utilizing a jumping box. The exactness of identification of an article is discovered utilizing the distinction between the real jumping boxes versus anticipated bouncing box which is known as Intersection over Union.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

$$\text{Intersection over Union} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Mean Average Precision or Average Precision is a measurement to ascertain the exactness of the item identifiers.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Utilizing these conditions the precision or the mAP estimation of the model is found.In this methodology, the joint form of ResNet and YOLO v2 has been come out with an exactness of 87.7% which is the most noteworthy precision rate as contrasted and some current models. The current passerby's models taken for examination are Single Shot Detector, Faster R-CNN and Mask R-CNN. Among these three models the model Mask R-CNN is the better in both exactness and got less misfortune rate. While contrasting these models and the proposed framework, the proposed model is superior to the Mask R-CNN in both exactness and got less misfortune rate. It got the exactness rate higher than Mask R-CNN with the distinction of pretty much 3% higher than it, with the misfortune rate not as much as Mask R-CNN with the distinction 0.056. The precision and misfortune pace of each model is appeared in table 1.

Models	mAP	Loss
Single Shot Detector	62.5%	1.92
Faster R-CNN	78.8%	0.159
Mask RCNN	84.9%	0.099
Proposed Approach	87.7%	0.043

Table 1: Accuracy Vs Loss

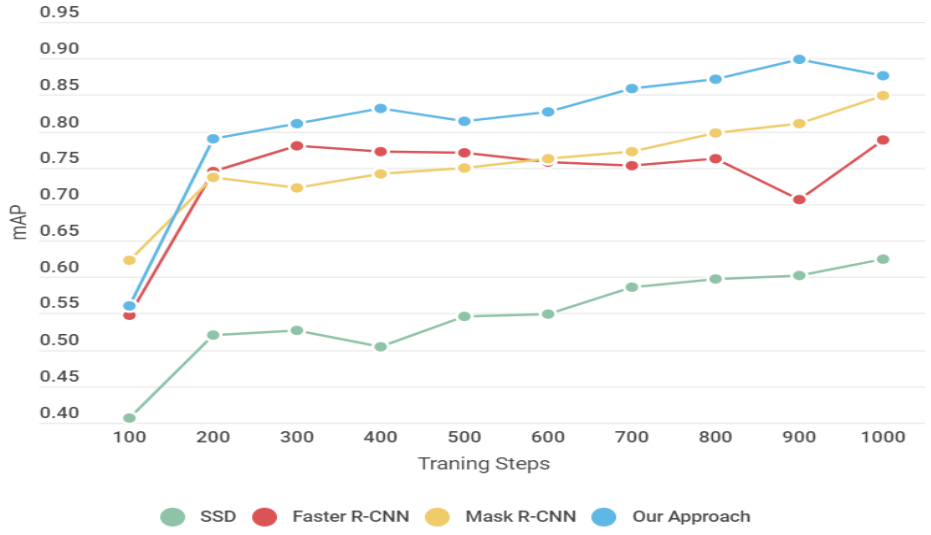

Fig2: Comparison of Accuracy with 4 models

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

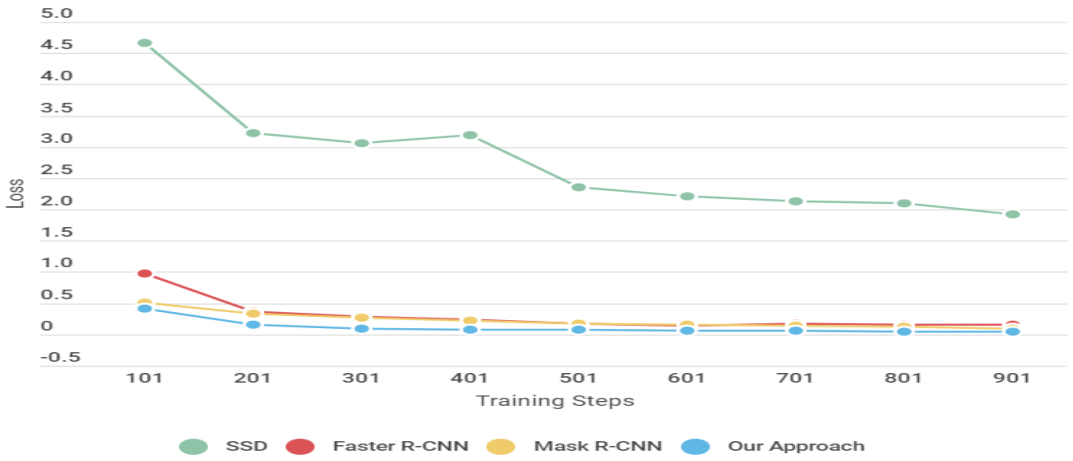

Fig3: Comparison of Loss Value with 4 models

The examinations of these models are plotted on a diagram for both mAP and Loss esteems, appeared in Figures 2 and 3 individually, which shows the mAP and misfortune estimation of the organizations from the first to the 1000th preparing steps. Distinction can be seen between those current models and the proposed approach, that this model has the most elevated exactness rate with less misfortune work. Likewise, the yield of the test is appeared in Figure 4A and 4B. The yield picture of the examination shows the location of people on foot utilizing the jumping confines with the certainty score it; from this the exactness of the model can be estimated. In figure 4.3.B, can see even a little piece of the picture is identified (set apart with red tone jumping box inside the vehicle) and set apart as common with the certainty score of 0.915, which is a precise identification. From this it is shown that by utilizing this calculation we can recognize more modest items from some random edges. So this calculation can be utilized continuously applications.

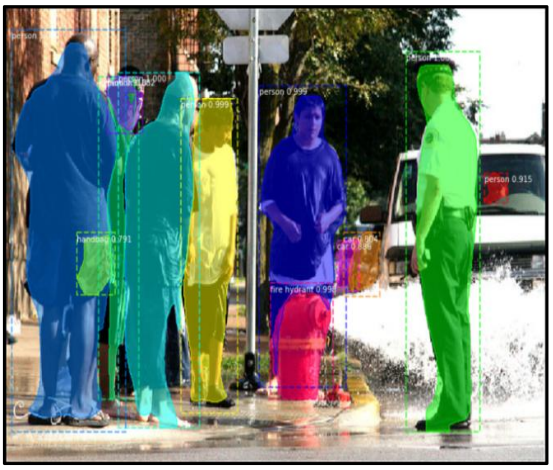

Fig 4A: Result 1Fig 4B: Result 2

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

Conclusion

Subsequently, the pedestrian detection calculation is created utilizing the coordination of ResNet and YOLO v2 networks. The decision of these two organizations is to build the precision rate and to limit the misfortune capacity of the framework to actualize it continuously applications. By exploring different avenues regarding these two incorporated organizations the calculation has come out with a precision of 87.7 percent with a deficiency of 0.043. The model has been contrasted and existing models like SSD, Faster RCNN and Mask RCNN which shows that this calculation has the most noteworthy precision rate with less misfortune work. This calculation can likewise be extemporized by coordinating with some other effective calculations. Additionally, it tends to be coordinated with equipment like CCTV cameras for clever video reconnaissance in real time applications.

References

[1] Ahmed, Z., Iniyavan, R., & P, M. M. (2019), "Enhanced Vulnerable Pedestrian Detection using Deep Learning"; International Conference on Communication and Signal Processing (ICCSP), pp: 0971-0974.

[2] Ash, R., Ofri, D., Brokman, J., Friedman, I., & Moshe, Y. (2018), "Real-time Pedestrian Traffic Light Detection"; IEEE International Conference on the Science of Electrical Engineering in Israel.

[3] Brunetti, A., Buongiorno, D., Trotta, G. F., &Bevilacqua, V. (2018), "Computer vision and deep learning techniques for pedestrian detection and tracking: A survey"; Neurocomputing, 300, pp: 17–33.

[4] hen, E., Tang, X., & Fu, B. (2018), "A Modified Pedestrian Retrieval Method Based on Faster R-CNN with Integration of Pedestrian Detection and Re-Identification"; International Conference on Audio, Language and Image Processing, pp: 63-66.

[5] Guanqing Li, ZhiyongSong ,Qiang Fu (2018), "A New Method of Image Detection for Small Datasets under the Framework of YOLO Network"; IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, pp: 1031-1035.

[6] Lan, W., Dang, J., Wang, Y., & Wang, S. (2018), "Pedestrian Detection Based on YOLO Network Model"; IEEE International Conference on Mechatronics and Automation (ICMA), pp: 1547-1551.

[7] Rahul Pathak,P. Sivraj, (2018), "Selection of Algorithms for Pedestrian Detection During Day and Night"; Computational Vision and Bio Inspired Computing, pp 120-133.

[8] Zaatouri, K., &Ezzedine, T. (2018), "A Self-Adaptive Traffic Light Control System Based on YOLO"; International Conference on Internet of Things, Embedded Systems and Communications (IINTEC), pp: 16-19.

[9] Ghosh, S., Amon, P., Hutter, A., &Kaup, A. (2017), "Reliable pedestrian detection using a deep neural network trained on pedestrian counts"; IEEE International Conference on

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2258 – 2268
Received 05 March 2021; Accepted 01 April 2021.

Image Processing.

[10] Hongmeng Song, Wenmin Wang (2017), "Collaborative Deep Networks for Pedestrian Detection"; IEEE Third International Conference on Multimedia Big Data.

[11] Naghavi, S. H., Avaznia, C., &Talebi, H. (2017), "Integrated real-time object detection for self-driving vehicles"; 10th Iranian Conference on Machine Vision and Image Processing.

[12] Zhang, H., Du, Y., Ning, S., Zhang, Y., Yang, S., & Du, C. (2017), "Pedestrian Detection Method Based on Faster R-CNN. 2017 13th International Conference on Computational Intelligence and Security.

[13] Hailong Li, Zhendong Wu, &Jianwu Zhang. (2016), "Pedestrian detection based on deep learning model"; 9th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics.

[14] Peng, Q., Luo, W., Hong, G., Feng, M., Xia, Y., Yu, Li, M. (2016), "Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO"; 8th International Conference on Intelligent Human-Machine Systems and Cybernetics.

[15] Tomè, D., Monti, F., Baroffio, L., Bondi, L., Tagliasacchi, M., &Tubaro, S. (2016), "Deep Convolutional Neural Networks for pedestrian detection, Signal Processing: Image Communication, pp: 482–489.

[16] Wei, Y., Tian, Q., &Guo, T. (2013), "An Improved Pedestrian Detection Algorithm Integrating Haar-Like Features and HOG Descriptors"; Advances in Mechanical Engineering.

[17] Krizhevsky, I. Sutskever, G. E. Hinton (2012), "Imagenet classification with deep convolutional neural networks"; Advances in neural information processing systems.

[18] David Gero´ nimo, Antonio M. Lo´ pez, Angel D. Sappa (2010), "Survey of Pedestrian Detection for Advanced Driver Assistance Systems", IEEE Transaction on pattern analysis and machine intelligence, VOL. 32, NO. 7, pp: 1239-1258.

[19] Hong, Z., Zhang, L., & Wang, P. (2018). "Pedestrian Detection Based on YOLO-D Network"; IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), pp: 802-806.

[20] Xie, C., Li, P., & Sun, Y. (2019)," Pedestrian Detection and Location Algorithm Based on Deep Learning". International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), pp: 582-585.

[21] Nguyen, K., Fookes, C., &Sridharan, S. (2015), "Improving Deep Convolutional neural networks with unsupervised feature learning", IEEE International Conference on Image Processing (ICIP), pp: 2270-2274.

[22] Chen, X., Guo, R., Luo, W., & Fu, C. (2018). Visual Crowd Counting with Improved Inception-ResNet-A Module. 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp: 112-119.