ML Based Crowd Detection System – A review

Priyanka Ubale[1], Apurva Surve[2], Anamika Srivastava[3], Rakhi Vishwakarma[4], Shikha Malik[5]

[1-4]B.E. Student, Electronics and Telecommunication Engineering, Atharva College of Engineering, Mumbai, India
[5]Professor, Electronics and Telecommunication Engineering, Atharva College of Engineering, Mumbai, India

---

Abstract - *-Machine learning is a subpart of artificial intelligence (AI). In spite of the fact that machine learning is a field of computer science, it differs from traditional computational approaches. Object detection has a significant role in computer vision theory and practical applications , this detection involves tracking, detection and counting of objects. Object detection is one of the computer vision tasks which benefits from Deep learning techniques. These techniques involve Convolutional Neural Networks (CNN), Fast Convolutional Neural Networks (FCNN), You Only Look once (YOLO) and more. This paper conducts a survey on few of the important and recent developments on object detection with the use of deep learning.*

Key Words: **Machine learning, Deep learning, YOLOv4, object detection, computer vision**

1.INTRODUCTION

Nowadays the world is moving towards automation and machine learning is one of the trending technologies used in automation. Machine learning is a subset of artificial intelligence which uses data and algorithms to behave as humans do and improve the accuracy. Machine learning includes deep learning techniques. Deep is a subset of machine learning which uses neural networks with different layers to imitate the human brain behaviour. Deep learning has various applications in real time. In this paper we focus on one of its applications that is object detection.

Object detection is a branch of image processing and computer vision used to identify and locate objects from still images, videos or real time videos. The object detection methods fall under two categories- non-neural approach and neural approach. Non-neural approaches perform object detection by first extracting the features from an image and then feeding those features to a regression model for predicting the location and label of the object in an image. Non-neural based approaches include Scale-invariant feature transform (SIFT), Viola-Jones object detection framework using Haar features, Histogram of oriented gradients (HOG) features, etc. The neural networks approach performs object detection by recognizing patterns in an image using several layers like input layer, hidden layer and output layer. There should be atleast one hidden layer and the number of hidden layers can be increased for more accurate results of detection. The neural network approaches are RCNN, fast RCNN, faster RCNN, YOLO, SSD, etc.

2. ALGORITHMS

Some of the deep learning algorithms based on object detection are explained below:-

A) R-CNN (Region based convolutional neural networks): RCNN was introduced by Ross Girshick et al. in 2014. It uses a selective search algorithm for object localization. The selective search algorithm extracts 2000 regions which is comparatively smaller than the regions generated by brute force sliding window approach. These 2000 regions are also called region proposals which simply means smaller parts of the original image where an object of interest is likely to be found. The output of the selective search algorithm is fed to CNN. The feature vector of 4096 dimension is obtained as an output of CNN which is then fed into the SVM classifier to classify the objects. Then we use a boundary box regressor to put the object detected in a rectangle.

B) Fast R-CNN: It is an improved version of R-CNN. In fast R-CNN the whole image is fed to convolutional neural networks which produces a feature map. The region proposal obtained from a selective search algorithm is then combined with a feature map and sent to the RoI (Region of Interest) pooling layer to reshape into a fixed-length feature vector. Each of these feature vectors is fed to fully connected layers. The classes of objects are classified by using softmax probabilities and the location of the object is returned by bounding box regressor.

C) Faster R-CNN: In R-CNN and fast R-CNN, the region proposals are found using selective search algorithms. The selective search algorithm is slow and time consuming. This limitation is overcome in the faster R-CNN method. Faster R-CNN uses a separate network to predict the region proposals called RPN (Region proposal network). The rest of the process is similar to its previous version.

D) YOLO (You Only Look Once): Redmon et al. proposed YOLO algorithm in the year 2015. Unlike other object detection algorithms, YOLO contains a single convolution network that predicts the bounding boxes and their class probabilities. The input image is first divided into an SXS grid cell. Then each of these grid cells is

responsible for predicting the object present in that cell and their corresponding confidence scores.

3. LITERATURE SURVEY

In this section we present the review of various algorithms for detection, counting and tracking.

Feng Yizhou and et al. (2019) [1] proposed a person's flow monitoring system which uses Single shot Multibox Detector (SSD) for detecting the people's flow. The authors used the pedestrian's head as the parameter to detect the people due to the shooting angle. Since the traditional dataset considered face as a parameter to detect people, the authors prepared their own dataset that included people with umbrellas, hats and other objects which can block human characteristics. They used MobileNet to improve the performance of VGG-SSD and Non-maximum Suppression (NMS) algorithm to get rid of multiple bounding boxes around a single object. The accuracy achieved by the model was 93.345%.

Misbah Ahmad and et al. (2019) [2] explored the SSD algorithm for top view people detection and counting. The model is pretrained on frontal view dataset i.e., COCO dataset and tested on top view person images. Their testing dataset consists of 500 images of people taken from top view. Out of 500 images, 250 images were of indoor people sample images and other 250 images were of outdoor people sample images. The images contained people from 1 up to 7. The authors got an average TPR (True Positive Rate) of 95% and FPR (False Positive Rate) of 0.16% for indoor people sample images and for outdoor people sample images they got TPR of 94.42% and FPR of 0.17%.

Ujwala Bhangale and et al. (2020) [3] presented a model for detecting and counting the crowd using Deep Convolutional Neural Network (DCNN). The DCNN architecture is based on CSRNet which includes 10 convolutional layers and 3 max pooling layers of VGG-16 in the front end and in the backend, they used 6 dilated convolution layers with rate of dilation 2. The 3X3 kernel size was maintained throughout the process. The dataset used is ShanghaiTech dataset. This dataset contains two parts. Part A consists of images having dense crowds and part B consists of images having sparse crowds.

Nurul Iman Hassan, Fadhlan Hafizhelmi Kamaru Zaman, Nooritawati Md. Tahir and Habibah Hashim (2020) proposed a people detection system in real time [4]. The model successfully detects most of the people from the images having complicated scenes and crowded areas. It also was able to detect people in armor and who were partially occluded efficiently. The model was trained on Google Colaboratory which has a built in Tesla K80 GPU. The authors prepared a custom dataset using Google's Open Images. The dataset consists of 500 high resolution images of person class. The achieved mean average precision(mAP) is 78.3% and final average loss of 0.6. To increase the mAP and decrease the final loss the authors recommended increasing the images used for training the model and for fast computing they suggested using a faster GPU.

Prashanth Kannadaguli (2020) [5] designed a human detection system using You Look Only Once (YOLO) v4. The YOLOv4 algorithm outshine in real time object detection from images and videos. The thermal data collected using thermal imaging is used for training and testing. In the preprocessing step the author annotated the images using Microsoft's Visual Object Tagging Tool (VoTT), removed the noise from the thermal images using median filter and then on the result of median filter histogram matching was performed. The model was experimented on four sets of data. In each set the number of training data was increased and it was observed that as the training data was increased the model runtime and accuracy was also increased. The author inferred that using YOLO we can successfully detect humans in thermal aerial images or videos.

Juan Byju and et al. (2021) [6] used the Improved YOLOv4 model to detect pedestrians in challenging conditions. In the Improved YOLOv4 algorithm transfer learning is used to detect and classify the objects from the image. Transfer learning is the process used to extract features from the custom dataset with the help of a pretrained model. The authors prepared a custom dataset by combining various datasets like COCO, Elektra, inRIA and videos of pedestrians in snowy, rainy and windy climate conditions. The model was trained with 3200 training and testing images of custom dataset on google colaboratory. To find the best performance metrics which consist of mean average precision (mAP), the number of iterations were increased till the model is overfitted. Overfitting happens when the model is over trained and after that the value of performance metrics goes on decreasing. The best mAP was achieved in the 2000 iteration. The result obtained from Improved YOLOv4 was compared with YOLOv4 and it was observed that there was 7% increase in the mAP of Improved YOLOv4 model.

Garima Mathur , Davendra Somwanshi and Dr. Mahesh M. Bandle (2018) [7] presented object tracking using Mean Shift Algorithm .The algorithm is used to track user defined objects by frequently updating object location. Region of Image(ROI) to be tracked is extracted to track the object and each frame is checked the presence of object. Their algorithm was implemented for video with 115 frames at rate of 30fps.Their experiment performed the object tracking without missing any frames and could successfully overlap bounding box.

Dushyant Kumar Singh, Sumit Paroothi ,Mayank Kumar Rusia and Mohd. Aquib Ansari (2020) [8] presented Human crowd detection with the use of different techniques. Firstly, background subtraction technique for motion detection, in which pixel position between two images shows the true difference between intensity with respect to displacement. After that Histogram of Gradient (HOG) well known technique of computer vision, it counts the existence of gradient orientation with image locally. And then Support Vector System (SVM) ,this classifier technique used to maximize the marginal difference between two distinct classes.[8]They also used the Vif descriptor for the detection of violence. They provided a good result in case of crowd violence detection in terms of accuracy and sensitivity.

Dr. Shailender Kumar, Vishal Pranav Sharma and Nitin Pal(2021) [9] did propose the model for object tracking and counting in the zone with the use of YOLOv4,Deep SORT and TensorFlow. The tracking and counting of objects involve three stages i,e. detecting, identifying and tracking the object in a particular zone. In their research they used a Kalman filter as it improved the accuracy of the model and yielded the best result. And the use of YOLO was for detection and recognition at the same time. Their model was able to produce 60% accuracy during detection.

TABLE I-LITERATURE SURVEY ON OBJECT DETECTION

Sr No	Paper - Topic	Source - Year	Algorithm	Dataset	Accuracy
1.	Application of the SSD Algorithm in a People Flow Monitoring System	IEEE-2019	1.Single Shot Multibox Detector (SSD) 2.MobileNet 3. Non-maximum Suppression Algorithm	-	93.35%
2.	Intelligent Video Surveillance based on Object Tracking	IEEE-2018	Mean Shift algorithm	-	-
3.	A Deep Neural Network Approach for Top View People Detection and Counting	Researchagate-2019	Single Shot Multibox Detector (SSD)	COCO dataset	95%
4.	People Detection System Using YOLOv3 Algorithm	IEEE-2020	YOLOv3 Algorithm	customised dataset from Google's Open Images	mean average precision (mAP) of 78.3%
5.	Near Real-Time crowd counting using deep learning approach	Sciencedirect-2020	CNN-CSRnet	Shanghaitech dataset	68.2 - MAE
6.	Human crowd detection for city wide surveillance	Sciencedirect-2020	1.Violent Flows (ViF) descriptor 2. Support Vector Machine (SVM) classifier	-	-
7.	Object detection and count of objects in image in image using Tensor flow objects detection API	IEEE - 2019	1.RCNN 2.Faster RCNN inception V2 model 3.Tensor flow	-	81.81%
8.	An FPGA Based Approach For People Counting Using Image Processing Techniques	IEEE - 2019	1.Histogram Of Gradient (HOG) 2.Support Vector Machine(SVM) 3. FPGA	-	-
9.	Object tracking and counting in a zone using YOLOv4, DeepSORT and TensorFlow	IEEE - 2021	1.YOLOv4 2.DeepSORT 3.TensorFlow	-	60%
10.	Pedestrian Detection and	IEEE – 2021	YOLOv4 + Transfer	Custom dataset	-

	Tracking in Challenging Conditions		Learning		
11.	You Only Look Once: Unified, Real-Time Object Detection	IEEE-2016	YOLO	PASCAL VOC 2007	-
12.	YOLOv3 and YOLOv4: Multiple Object Detection for Surveillance Applications	IEEE - 2020	1.YOLOv3 2. YOLOv4	1. KITTI image dataset 2. KITTI video dataset	98% and 99% respectively
13.	Real-time Personal Protective Equipment (PPE) Detection Using YOLOv4 and TensorFlow	Researchgate - 2020	YOLOv4	Custom dataset	79%
14.	YOLO v4 Based Human Detection System Using Aerial Thermal Imaging for UAV Based Surveillance Applications	IEEE - 2020	YOLOv4	Thermal dataset	-

" – " – Not mentioned

4. CONCLUSION

This paper is a comprehensive survey for machine learning based crowd detection. The efficiency of deep learning in object detection, recently implemented and completed experiments and studies in the domain got reviewed and analyzed. To this end, we found that Convolutional Neural Networks, Fast Convolutional Neural Networks and You Only Look Once i.e. YOLOv3 and YOLOv4 have iteratively been used as baseline of robust object detection systems have obtained in many experiments, contemporary performance on various datasets.

REFERENCES

[1] Yizhou, Feng, et al. "Application of the SSD Algorithm in a People Flow Monitoring System." 2019 15th International Conference on Computational Intelligence and Security (CIS). IEEE, 2019.

[2] Ahmad, Misbah, et al. "A deep neural network approach for top view people detection and counting." 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON). IEEE, 2019.

[3] Bhangale, Ujwala, et al. "Near Real-time Crowd Counting using Deep Learning Approach." Procedia Computer Science 171 (2020): 770-779.

[4] Hassan, Nurul Iman, et al. "People detection system using YOLOv3 algorithm." 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE). IEEE, 2020.

[5] Kannadaguli, Prashanth. "YOLO v4 Based Human Detection System Using Aerial Thermal Imaging for UAV Based Surveillance Applications." 2020 International Conference on Decision Aid Sciences and Application (DASA). IEEE, 2020.

[6] Byju, Juan, et al. "Pedestrian Detection and Tracking in Challenging Conditions." 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2021.

[7] Mathur, Garima, Devendra Somwanshi, and Mahesh M. Bundele. "Intelligent Video Surveillance based on Object Tracking." 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE). IEEE, 2018.

[8] Singh, Dushyant Kumar, et al. "Human crowd detection for city wide surveillance." Procedia Computer Science 171 (2020): 350-359.

[9] Kumar, Shailender, Pranav Sharma, and Nitin Pal. "Object tracking and counting in a zone using YOLOv4, DeepSORT and TensorFlow." 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). IEEE, 2021.

[10] Kumar, Chethan, and R. Punitha. "Yolov3 and yolov4: Multiple object detection for surveillance applications." *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2020.

[11] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[12] Kumar, Chethan, and R. Punitha. "Yolov3 and yolov4: Multiple object detection for surveillance applications." 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2020.

[13] Protik, Adban Akib, Amzad Hossain Rafi, and Shahnewaz Siddique. "Real-time Personal Protective Equipment (PPE) Detection Using YOLOv4 and TensorFlow." 2021 IEEE Region 10 Symposium (TENSYMP). IEEE, 2021.

[14] Kannadaguli, Prashanth. "YOLO v4 Based Human Detection System Using Aerial Thermal Imaging for UAV Based Surveillance Applications." 2020