Volume 9, Issue 7, July – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1266

Counting Individuals in an Image using Machine Learning Technique

Harsha D P[1]; Hemanth Kumar[2]
[1*] MCA Student, JNNCE [2] Associate Professor,
Department of MCA, JNNCE, Shivamogga

Abstract:- In video surveillance system, the most complex thing is to detect individuals. In recent years, the research is done using deep learning technique, which gives powerful individuals detection results. A model in Deep Learning i.e. YOLO (You Only Look Once) has explored in individual detection in all the angles of a individuals. The model is tested and trained on viewing person dataset. Further, counting individuals has done through information of classified bounding box. The trained model is going to verify by giving several testable data set and takes two datasets for training and testing. Trained model is tested rigorously to find out the accuracy of a model. This methodology gives efficient results for counting individual in real world.

Keywords:- *Deep Learning; Bounding Box; Machine Learning.*

I. INTRODUCTION

Counting individuals in an image is important step in video surveillance. Locating individuals positions in frame is performed in detection phase. Other so many researchers has completed this work by applying various algorithms and tried to do this, can obtain good results in detection and counting while implementing YOLO. It detects multiple object detection in real-time video surveillance.

As of now the researchers has done the research on full view of the individual body and quite side angles that gives best results at the end. This is having a problem, that is in crowded scenes full human bodies cannot be detected in that situations the accuracy decreases. To overcome this, we have trained our model to detect only individual faces only, our proposed work detect individual in crowded scenes. This is a inspirational task for every researcher. So, to overcome this problem, above mentioned problem statement need to apply best methods to train and to detect individuals image. Required to determine which feature extraction is better to get good accurate classification of individuals in real time video surveillance.

For this proposed work there are many algorithms to solve this individual count. This application is used in several fields like in traffic, airports and in many IOT projects. We Can implement this in very less cost for the clients. In this how many individuals can see in a present screen it will give the count, if a person exits the screen then the count is decrement automatically. This projects main focus on if there are two humans in a single frame then it is challenging task to identify two individual so take the count for individuals head in bounding boxes.

II. LITERATURE SURVEY

In general, for individuals detection use two techniques: Detection based method and Map based method. Detection based method works as it counts each and every human individually and then gives the count. But in Map based method it creates a relationship if multiple humans detected in image and gives a count of individuals.

The work by kin-Yi Yam et al. in [1] used initial objects detector method IOD and existing objects detector method EOD has used for detecting individuals in their work. And they have tried for bi directional people flow counting. They have placed cameras in shopping malls and colleges in 40degree angle to check accuracy and they got the results as 88 percent for their work. The author Ya-li Hou et al. in [2] proposed a people counting and human detection using convolutional neural network to estimate number of humans in an image. For detection they used EM (expectation Maximization) based algorithm for individual detection. In [3] authors explained how they detected humans by using HOG Method and Network based tracking to detect the humans in circular manner.

The authors Xuan Zhou et al. in [4] used three techniques to detect individuals that static detection, dynamic detection and compressive judgement of detecting individuals in video surveillance. author trained images in 704X756 pixel size. The author Swapnil H Tathe et al. in [5] proposes three algorithms to detect people, one is Pixel Count algorithm, Kalman Filter algorithm and then Mean Shift algorithm is used for tracking. They successfully implemented on roads to track people. In [6] the author Huajie Wu et al. designs a counting system for humans by face detection method and skin detection method, authors made expermients in a lab entrance for counting. To detect faces they used AdaBoost algorithm to detect face closely. Author Mohammad Javad Abbaspour et al. in [7] the author gives plenty of objects to detect but gave count for humans only. They made experiment to detect objects by using KDE (Kernal Dimensity Estimation) and this work is limited to 3 humans in an image if it gives high then it doesn't work properly because it has to detect several kinds of objects too.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1266

In [8] author Mohana et al. gives humans can detected in Novel Matching Process and Novel Template Matching is used for counting purposes and they used iBall C20.0 web cam to try it and it has frame rate of 30 frames per second. The author Cao Jianzhao et al. in [9] Proposes Single Gaussian model (SGM) and Original Kernel Density Estimation (KDE) method to detect humans in image. They are often used in image processing and DSP (Digital Signal Processing) applications. In [10], the author Diping Song et al. says dataset 1 contains 2834 depth images with 4541 heads, and the dataset 2 contains 1500 depth images with 1553 heads. and they used Faster RCNN algorithm to detect humans with large dataset and got good accuracy results. The author Tsaipei Wang et. al. in [11] explains how Fisheye camera model, scene model and detection masks are used for human detection and support vector machine. They collected dataset in real retail shop.

In [12] author Shengke Wang et al, used multi head attention in Attention Mechanisms and trained over 16,000 images with 1920X1080 resolution and used Graph Match method to increase the success rate. The author Sanjukta Ghosh et al. in [13] uses convolutional neural networks to detect peoples in video surveillance and Multi Task Networking method for counting. They combined the counting model with instance separation for common feature extractor. In [14], author Misbah Ahmad et al. uses YOLO V3 version to detect and trained YOLO v3-M model to generate proper bounding boxes with high accuracy and made a overhead human detection and used COCO (Common in context) model to pretrained dataset, and got the accuracy rate of 90-95%.

The author J. Gronman et al. in [15] used RNN (recurrent neural network) is used to detect humans and used open cv (open-source computer vision) to count the humans and got a result of 91% accuracy. In [16] Shijie Sun et al. explains few methods to detect humans first removal of background, second re projection of image and third candidate head proposal method and last head proposal refinement method and got good results. The author Deepak Babu Sam et al. in [17] used LSC-CNN (Locally supervised convolutional neural network), used for automatic individual detection and gives counting in single time. the first ever to detect and count in single frame by Deepak Babu got 93% of accuracy.

In [18] the author Muhamad Izham Hadi Azhar et al. implemented YOLO V3 algorithm in Deep Sorting Method. They done their work into three phases and checks accuracy part by part and presents best of three. The author Imene Bakour et al. in [19] presents Soft-CSR (Soft Count Spatial Regularization) method is used for crowd counting and human detection. And got results of accuracy 97%. In [20] Ahmed Nidhal et al. presents HSV method to detect and they applied this method to control traffic congestion in roads they first tried and applied from accessing street cameras and then they applied their algorithm to it and they got highest accuracy rate of 97-100%.

III. PROPOSED METHODOLOGY

Machine Learning is a technique used to solve hidden patterns and structures from input data. Machine Learning algorithms and Deep Learning algorithms are trained on dataset to recognize patterns, detection and perform several tasks. In Machine Learning and Deep Learning there are multiple algorithms are there, few are like CNN (Convolutional neural network), R-CNN (Region Convolutional Neural Network), ANN (Artificial Neural Network), K-NN (K-Nearest Neighbors), PCA (principal component analysis) and etc.

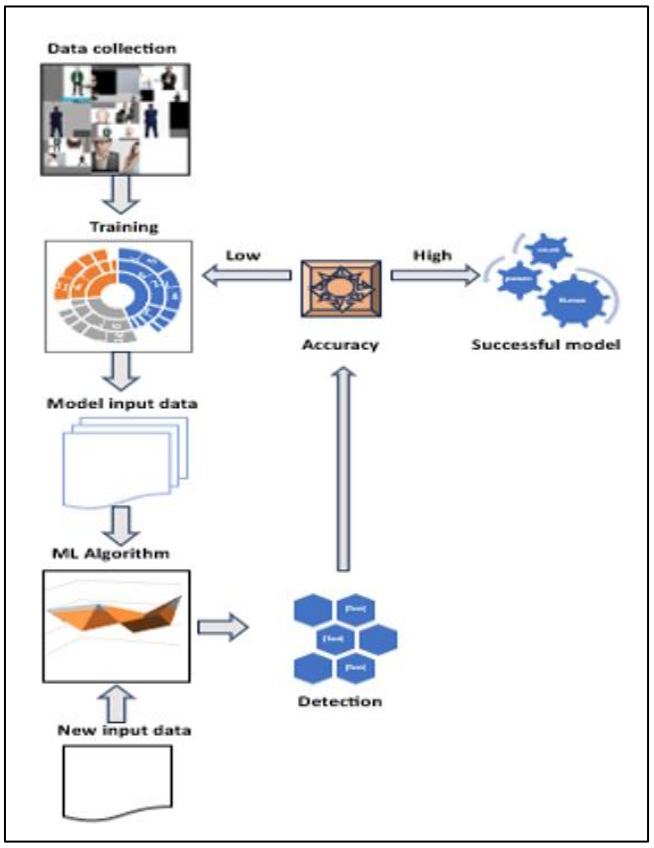

Fig 1: Machine Learning Model

Fig. 1 shows Machine Learning model. YOLO algorithm chosen as it detects multiple objects on single image in real time, the YOLO model has open-source libraries and it helps to reduce the lines of code to train and test the dataset.

YOLO model is first built by Joseph Redmon and Ali Farhadi in University of Washington, is used for image segmentation and real time object recognition. Fig. 2 shows development of YOLO versions. YOLO V8 is used for current work, a bounding box object recognition model that is created by Ultralytics. In recent versions of YOLO, they tried to include accuracy level and simultaneously its speed of training data.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1266

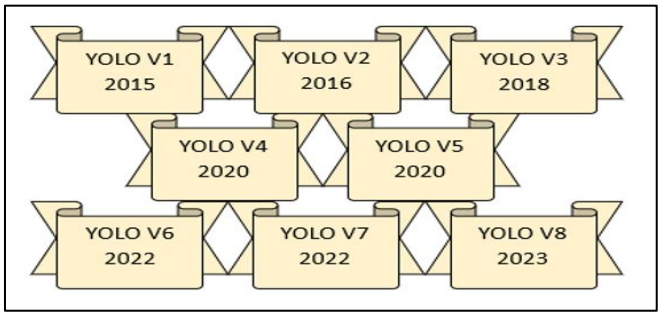

Fig 2: YOLO Versions

YOLO first divides image into grid cells as 13X13, each grid cell detects five bounding boxes (bounding box means it is in rectangle shape which covers our object to detect and train). YOLO gives a percentage of predicted image is correct or not which is present in bounding boxes. Now the image takes 13X13 images and each cell trains 5 bounding boxes means 13X13X5 there are 845 bounding boxes. So trained 845 images once predicted that explains us how speed YOLO is.

A. Comparison of YOLO V8 and R-CNN

Table 1: Comparison of YOLO V8 and R-CNN

Features	**YOLO V8**	**R-CNN**
Architecture	1-stage detector	2- stage Detector
Activation	Leaky ReLU, Mish	ReLU
Prediction	Direct prediction	2- stage Detector
Speed	Faster due to one stage detector	Slow due to two stage detector
Accuracy	High accuracy in real-time	Good accuracy
Application Suitability	Real time Applications	Higher precision Applications

R-CNN (Region-Convolutional Neural Network) takes lots of time when starts to train the model, but YOLO takes less time of training compared to R-CNN model. R-CNN is not used in real time implementation for object detection. Table 1 describes R-CNN is having lesser accuracy rate as compared to YOLO algorithm.

B. Architecture of YOLO V8

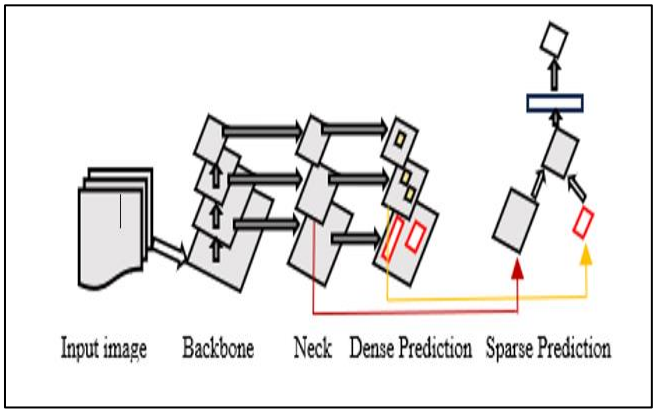

Fig 3: Architecture of YOLO V8

YOLO is having 5 stages. First, input image is given to model, next comes Backbone which takes responsible to extract the features from given input image. It is having pyramid in shape, this step having series of convolutional layers which helps to train model. Then comes Neck, it collects and processes all the features from back bone. This step includes FPN (Feature Pyramid Network) and Path Aggression Network (PANet) structure. Then comes Dense prediction, it is responsible to make predictions based on extracted features in every possible location. Fig. 3 explains Sparse prediction focuses on final detection results with making input bounding boxes on particular object to identify.

C. Working of YOLO V8

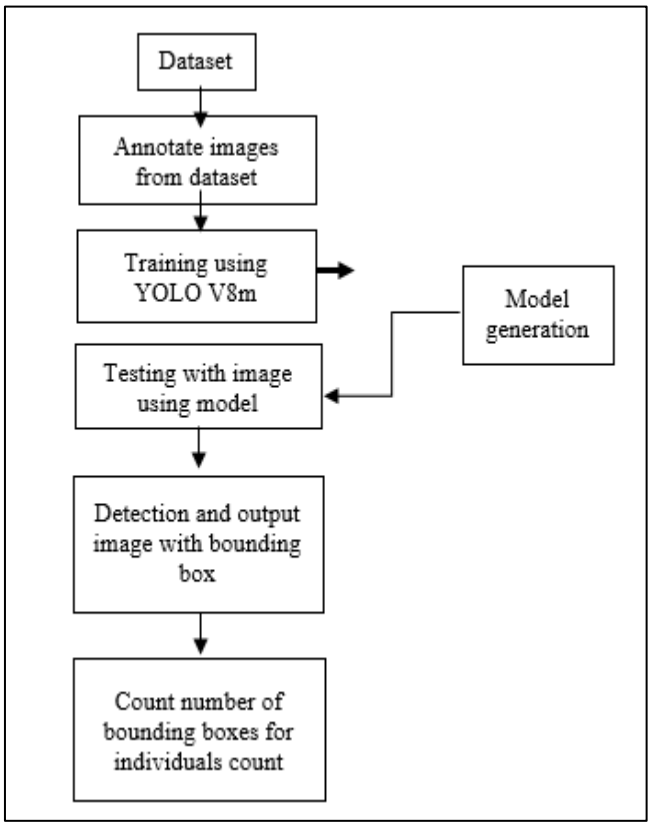

Fig 4: Work flow diagram

How YOLO V8 involved in individuals detection and counting? First step to collect dataset of individuals to train, test, split then annotate each image with particular individuals to train the model to get results. Training is done using YOLO V8-m model. This is a robust and open-source model is easy to implement. YOLO V8-m model is moderate version of YOLO. The Fig. 4 shows workflow of detection and counting

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1266

of individuals. After training a model it generates a pretrained model. Testing is done using this pretrained model and other input data. A method named NMS (Non-Maximum Suppression) is used to avoid overlapping of bounding boxes in output screen. Instance classes for each and every bounding box that processes iterations of counting individuals.

IV. RESULTS

Dataset involves large numbers of image collection. For head detection we try to gain the images like individuals head from all angles and wearing glasses, cap and masks and for women, specially having long hair is going different types of images taken the dataset images in 640X840 pixels of image size, trained 24 frames per second in jpg format. Trained different aged individuals to gain a good accuracy. Taken dataset into three parts that having training, testing and validation in the ratio 70:20:10. This helps us in giving better results. For expecting good results the first step is collecting the dataset. If proper collection of dataset is done then good results are achieved.

Fig 5: Predicting Humans in Group Image

Fig 6: Shows on only Individuals Detected not Animals

Fig. 5 gives individuals detection with bounding boxes with group of humans in several angles and poses. Fig. 6 shows the detection of humans rather than detecting animals like cat and dog in image and detecting individuals with bounding boxes with percentage of detection of humans and that to calculate the accuracy of individuals.

V. CONCLUSION

In this proposed work, a Machine Learning model is ready to detect individuals and give a count. The model (Pre trained on all angles of human images) is undergone testing by giving several individuals images. The trained model detects individuals with bounding boxes and is used for counting purpose. The results obtained by Deep Learning model can be further developed by providing the same training and testing overhead individual dataset. The model individuals detection abilities are shown by experimental results.

ACKNOWLEDGMENT

The author expresses gratitude to everyone who was supported to do work and research process.

REFERENCES

[1]. K. -Y. Yam, W. -C. Siu, N. -F. Law and C. K. Chan, "Effective bi-directional people flow counting for real time surveillance system," 2011 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2011, pp. 863-864

[2]. I. Bakour, H. N. Bouchali, S. Allali and H. Lacheheb, "Soft-CSRNet: Real-time Dilated Convolutional Neural Networks for Crowd Counting with Drones," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), Boumerdes, Algeria, 2021, pp. 28-33.

[3]. S. V. Tathe and S. P. Narote, "Real-time human detection and tracking," 2013 Annual IEEE India Conference (INDICON), Mumbai, India, 2013, pp. 1-5.

[4]. Mohana and H. V. R. Aradhya, "Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixed camera," 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, India, 2016, pp. 1-7.

[5]. M. Patzold and T. Sikora, "Real-time person counting by propagating networks flows," 2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Klagenfurt, Austria, 2011, pp. 66-70.

[6]. S. Sun, N. Akhtar, H. Song, C. Zhang, J. Li and A. Mian, "Benchmark Data and Method for Real-Time People Counting in Cluttered Scenes Using Depth Sensors," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 10, pp. 3599-3612, Oct. 2019.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1266

[7]. Xuan Zhou, Shan Xu and Yuanhao Wang, "A combination of new static detection and dynamic detection in people counting," Proceedings of 2012 International Conference on Measurement, Information and Control, Harbin, 2012, pp. 605-609.
[8]. S. Wang, R. Li, X. Lv, X. Zhang, J. Zhu and J. Dong, "People Counting Based on Head Detection and Reidentification in Overlapping Cameras System," 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Jinan, China, 2018, pp. 47-51.
[9]. A. Nidhal, U. K. Ngah and W. Ismail, "Real time traffic congestion detection system," 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, 2014, pp. 1-5.
[10]. M. J. Abbaspour, M. Yazdi and M. -a. M. Shirazi, "Robust approach for people detection and tracking by stereo vision," 7'th International Symposium on Telecommunications (IST'2014), Tehran, Iran, 2014, pp. 326-331.
[11]. H. Wu, L. Meng and J. Pan, "Research of people counting based on face detection and area-tracking," Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, Dalian, China, 2013, pp. 525-528.
[12]. M. I. H. Azhar, F. H. K. Zaman, N. M. Tahir and H. Hashim, "People Tracking System Using Deep SORT," 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2020, pp. 137-141.
[13]. A. S. A. Nasir, N. K. A. Gharib and H. Jaafar, "Automatic Passenger Counting System Using Image Processing Based on Skin Colour Detection Approach," 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA), Kuching, Malaysia, 2018.
[14]. J. Gronman, P. Sillberg, P. Rantanen and M. Saari, "People Counting in a Public Event—Use Case: Free-to-Ride Bus," 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2019, pp. 1055-1059.
[15]. S. Ghosh, P. Amon, A. Hutter and A. Kaup, "Deep Counting Model Extensions with Segmentation for Person Detection," ICASSP 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 2032-2036.
[16]. C. Jianzhao, O. C. Victor, O. M. Gilbert and W. Changtao, "A fast background subtraction method using kernel density estimation for people counting," 2017 9th International Conference on Modelling, Identification and Control (ICMIC), Kunming, China, 2017, pp. 133-138.
[17]. Y. -L. Hou and G. K. H. Pang, "People Counting and Human Detection in a Challenging Situation," in IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, vol. 41, no. 1, pp. 24-33, 2011.
[18]. T. Wang, C. -W. Chang and Y. -S. Wu, "Template-based people detection using a single downward-viewing fisheye camera," 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Xiamen, China, 2017, pp. 719-723.
[19]. D. Song, Y. Qiao and A. Corbetta, "Depth driven people counting using deep region proposal network," 2017 IEEE International Conference on Information and Automation (ICIA), Macao, China, 2017, pp. 416-421.
[20]. M. Ahmad, I. Ahmed and A. Adnan, "Overhead View Person Detection Using YOLO," 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2019, pp. 0627-0633.