Comparative study between face mask detection and social detection with YOLO and Hara Cascade

Name: Maurya Goyal

mauryagoyal@gmail.com

Name: Krishna Yanmantram

krishnayanmantram@gmail.com

Name: Jayakumar K

drjayakumar1979@gmail.com

***Abstract*—Even if the outbreak was locally contained today, we must abide by the laws. This process must be automated in order to ensure that residents are following the law because authorities cannot be everywhere at once.[1] One of the most important penalties for preserving social space between people is requiring a gap of at least 6 meters between any two persons. This must be enforced, especially in public settings, to assure safety and lower the chance of the virus spreading.**

Keywords—Facemask,social distancing, covid, constrain

I. Introduction

These public places are generally large places, so instead of wasting human resources hiring security guards to ensure this, deep learning can be used to easily automate this process. Along with this data, the program can determine whether protocols are being followed and take appropriate action, as there are many CCTVs in public places and access to every corner of the place. Since the input data is real-time data, we use deep learning to ensure that the output is received optimally and quickly, and to further optimize performance, we use YOLO, which stands for You Only Look Once. use. This is an optimization algorithm that speeds up the detection process. YOLO was invented in 2015 before sliding window detection, followed by R CNN and even faster R CNNS. The difference between object localization and image classification is that in object localization a bounding box is formed along the detected object and is only responsible for identifying the application while the classification reads the data from this box recognised and using several different Machine Learning models we use the data captured for various applications. Hence, using YOLO we are doing object detection i.e it is used to identify the people in our case while for face mask detection we use haar cascade detection technique.

II. Research gap and contribution

It has only been tested with test data and should be tested with other data to increase accuracy above 90%. More than 6 d.H. d.H. missing people are not recognized and should be made more reliable. Although the proposed method is much better than his FasterRCNN and SSD, it still fails to identify some people and can improve object detection performance. It is very sensitive to the spatial position of the camera. More and more new scheduling systems are being equipped to provide "item detection as a service" in complex applications such as errands. Deep CNN based object detectors: - Faster R-CNN and YOLOv4. It shows the qualitative results of pedestrian detection in images using Faster R-CNN and the corresponding social distance in world coordinates.

III. RELATED WORKS

The IRIS PX4 drone is used in the Robot Operating System and Gazebo simulation to create road segmentation. Build the method using the IRIS PX4 drone model and replicate it in the JDERobot environment using the Gazebo environment. The IRIS PX4 is a four-rotor drone powered by the Pixhawk all-in-one autonomous autopilot system. Darknet-53 with 53 layers of convolution followed by batch normalization layers and Leaky ReLU activation function is used for feature extraction [1]. Regarding the need for social distancing in populated areas, a clever vision-based PC system with a 10,000-foot viewpoint, running deep learning and using observational videos, has been proposed. This proposed technique leverages YOLO v3's object detection model and uses key problem regressors to identify key component centroids. Also, object bouncing boxes are preserved when large groups are detected, and red boxes are additionally displayed where social distancing may be abused. [2]. At the heart of this vast array of applications, visual affirmative frames involving image ordering, confinement, and identification have achieved extraordinary search power. With important advances in brain networks, especially deep learning, these visual recognition frameworks have achieved extraordinary performance. Object identification is one area where PC vision has made significant progress. [3] .Detections in the image domain are converted to real bird's eye view coordinates for social distancing detection. When social distancing is detected, information is passed to him to two branches for further processing. The proposed system first uses a pretrained deep convolutional neural network (CNN) [4]. In this work, using common CCTV surveillance cameras for people detection, tracking, and distance estimation, hybrid computer vision and his YOLOv4-based deep neural network for automatic crowd detection indoors and outdoors (DNN) model. [Five]. Improved network design. The following improvements to the YOLO network model are made while keeping the original model-driven idea. Improvement based on loss function. The original YOLOv1 network loss function assumes the same error for large and small objects. Using the SPP layer provides more feature-rich image information and also significantly improves network time efficiency. [6]. Similar to training, only one network evaluation is required to predict detection of test images. In PASCAL VOC, the network predicts 98 bounding boxes for each image and class probabilities for each box. YOLO is very fast when testing because it only requires a single network evaluation, unlike classifier-based methods. Grid design applies spatial diversity to bounding box predictions. In many cases it is obvious which grid cell an object will fit in, and the network predicts only one box per object. [7]. The goal is to distinguish objects using a You Just Look Once (YOLO) approach. This strategy has some advantages compared to other articles

IV. MATERIALS AND METHODS / PROPOSED SYSTEM ARCHITECTURE

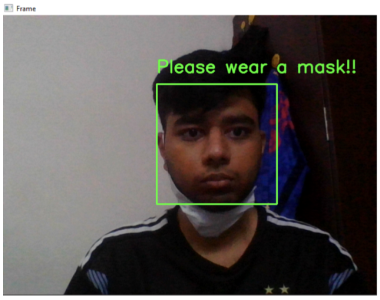
The main idea behind YOLO is to create a CNN network that can predict a (7, 7, 30) tensor. A CNN network is used to reduce the spatial dimensionality to 77 with 1024 output channels per site. To predict 772 border boxes, YOLO uses linear regression with two fully connected layers (middle image below). Keep the ones with the highest box confidence values (greater than 0.25) as final predictions and use them to build the model (right image). The class confidence value for each predictor box is computed as follows: Measures the certainty in identifying and classifying (where an object is). This score and probability terminology can easily be combined. Below is a math definition for your future.

For each grid cell, YOLO predicts a number of bounding boxes. I just want one of them to be responsible for the cause in order to calculate the real positive loss. For this reason, we choose the one with the highest her IoU (intersection on union) with the ground truth. As a result of this tactic, the bounding box prediction is specialized. Predicting a specific size and aspect ratio will make each prediction more accurate.

MTT (Moving Trajectory Tracking) & BEV: Direction following control in the light of waypoint behavior is a promising method for automated water vehicles (USVs) to achieve independent routes. It focuses on progress directions and employs sound learning computer-aided reasoning strategies to further develop UPS's post-degree directions. First, two deep brain tissue (DNN) models are built to assess root influences and step-by-step the limits of directional regulation. We then pre-train the DNN using Gaussian-Bernoulli-constrained Boltzmann machines to further improve the accuracy of predicting path effects. Finally, two of her DNNs are connected along his UPS control loop to provide predictive monitoring and decision support for traditional control technology. This type of equality adjusts the steering of the boat to pitch. Bird's Eye View The Higher Perspective is a dream-checking framework used in Auto ADAS innovations that provides a 360-degree hierarchical view. The main advantage of this framework is to allow the driver to exit the vehicle safely. Anyway, it's usually used for trajectory and chilling detection. This framework typically contains 4-6 fisheye cameras mounted around the vehicle, providing right, left, forward, and backward views of the elements around the vehicle. The terms flying perspective and aether perspective are sometimes used interchangeably with higher perspective. The term elevation may imply viewing a model from an exceptional height, even from a distance, while looking sideways from an airplane window or top. An upside view is inseparable from an upside view, but usually derives an advantage at a lower level than the last option period. For example, in PC and computer games, people.
The calibration phase of the pipeline begins by computing the transform to box view. The most basic calibration technique in this process is that from the perspective he selects four points and maps them to the corners of the box view rectangle. This step means everyone is on a flat surface.1.2. Object detection: The second stage of the pipeline, detection, uses a perspective view to apply a pedestrian detector and create a bounding box around each pedestrian. The company used his open-source DNN-based pedestrian detection network for this process.1.3 Distance measurement:- a. The third stage of the pipeline determines the estimated (x,y) position of each individual's bounding box within the frame view. After the boxview distances between subjects are calculated, the distances between each pair of subjects are scaled by a scaling factor derived from the calibration. To illustrate this metric, lines are drawn between people and red circles are placed around those who are closer than the minimum allowable distance.1.4. Feature extraction: His fourth step in the pipeline, which occurs after an object has been identified, is detailed below. This module separates people from all other objects in the photo. It uses dark web architecture to make the task easier. Edge detection is used to implement this work. The fifth step in the pipeline is tracing.1.5. This module assigns each person's relative her 2D spatial coordinates with respect to each person's closest or largest image. Think of them as the vertices of an entire polygon with n vertices. As a result, the polygon contains n*(n-1)/2 edges. Validation (1.6) The sixth and final step in the pipeline is validation. Since we found all possible edges in the previous module, we check all possible edges and then validate if(distance >= 6feet) . H. n*(n-1)/2 edges for n people. -

For the Haar Cascade classifier as we can see in the results the green box drawn is made by the Haar cascade object detection model. It uses the concept of edge or basically line detection. Hence what it detects is basically the change in the intensity of pixels. Hence it divides the image to certain boxes and in these boxes it tries to identify the instantaneous changes in the intensities of the pixels. Therefore, for this purpose it divides the pixels into 2 classifications: light or dark. After this it identifies the sum of all features lying in the dark and the sum of all features lying in the light area. Then we find out the difference between them. If this value is close to one it means that there is an edge separating the different variations which we have to identify. Then we draw the best block that separates the different intensities.

VI RESULTS AND DISCUSSION

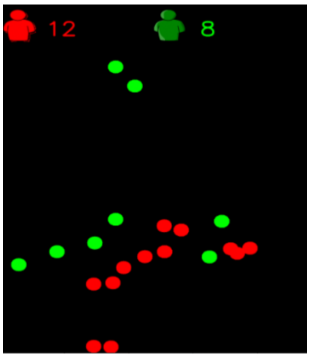


fig(i)

We can see that the sample person in means of the classification of face mask has not worn a mask and it has been detected that it has not been worn and then in fig2 where in we can now see that the same can be analyzed and then worked towards which will help us determine if the people have worn or not in the scenio of analyzing.

fig(ii)

When we can see that clusters are formed we can now see that the clusters of people near and far are seen and then we can see that the same was then seen to have people in good distance which is marked in red and green in fig iii respectively. Here we can now compare and study about which we want to consider and then take our reports out of the same.



fig(iii)

Moreover, fig v discusses the accuracy of the algorithms which helps us to study about why the algorithms of each are different and then we can see the time in which they perform the same. As we can see, the accuracy of haar cascade is usually less but really fast which really forces us to use the algorithm when we compare the same to YOLO when the situation arrives we can see that the Haar cascade algorithm has been an easy go with which we can accept these and then work on the same which helps us to analyze the found figures wherein they exist.

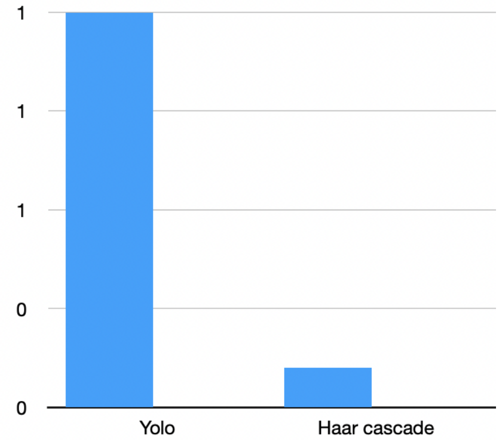
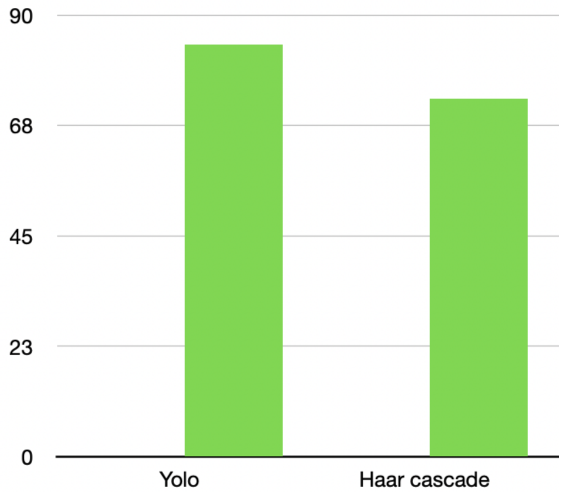


fig iv

Thereafter we can then see that the same has been in means of acceptance where we can see that they sync only over the fact that Yolo is much slower fig iv as it runs frame by frame and usually works over a video which is of greater size and workable area than the latter.



fig(v)

VII CONCLUSION AND FUTURE WORK

From this we can conclude that maintaining accuracy is a problem and should be done according to industry standard requirements. So adding a layer of noise will allow us to implement the same thing in terms of future acceptability, and thus help answer our demands. We have now noticed the difference between the two said models which were the

The major difference between haar cascade is that haar cascade classifier is clearly faster than YOLO and other faster CNNs wherein we might see some lags in the videos sometimes. But, the accuracy provided by the YOLO based model outweighed the accuracy by haar cascade classifier in which it sometimes wasn't even able to identify even the faces which were easily recognised by the YOLO model. Hence for surveillance footage to measure the distances YOLO fits the best whereas for a model like the face mask detection where time is of importance a model like haar cascade classifier combined with other classification techniques like logistic regression, SVMs, Naive Bayes etc.

VII References

[1]Somaldo, P., Ferdiansyah, F. A., Jati, G., & Jatmiko, W. (2020, December). Developing smart COVID-19 social distancing surveillance drone using YOLO implemented in robot operating system simulation environment. In *2020 IEEE 8th R10 humanitarian technology conference (R10-HTC)* (pp. 1-6). IEEE.

[2]Magoo, R., Singh, H., Jindal, N., Hooda, N., & Rana, P. S. (2021). Deep learning-based bird eye view social distancing monitoring using surveillance video for curbing the COVID-19 spread. *Neural Computing and Applications*, *33*(22), 15807-15814.

[3]Pathak, A. R., Pandey, M., & Rautaray, S. (2018). Application of deep learning for object detection. *Procedia computer science*, *132*, 1706-1717.

[4]Yang, D., Yurtsever, E., Renganathan, V., Redmill, K. A., & Özgüner, Ü. (2021). A vision-based social distancing and critical density detection system for COVID-19. *Sensors*, *21*(13), 4608.

[5]Rezaei, M., & Azarmi, M. (2020). Deepsocial: Social distancing monitoring and infection risk assessment in covid-19 pandemic. *Applied Sciences*, *10*(21), 7514.

[6]Ahmad, T., Ma, Y., Yahya, M., Ahmad, B., & Nazir, S. (2020). Object detection through modified YOLO neural network. *Scientific Programming*, *2020*.

[7]Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[8]Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[9]Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. (2021). Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, *8*(1), 1-27.

[10]Zhao-zhao, J. I. N., & Yu-fu, Z. H. E. N. G. (2020, October). Research on application of improved YOLO V3 algorithm in road target detection. In *Journal of Physics: Conference Series* (Vol. 1654, No. 1, p. 012060). IOP Publishing.