A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3

Md. Rafiuzzaman Bhuiyan
Dept. name of CSE
Daffodil International University
Dhaka, Bangladesh
rafiuzzaman15-9655@diu.edu.bd

Sharun Akter Khushbu
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
khusbhu15-6083@diu.edu.bd

Md. Sanzidul Islam
Dept. of SWE
Daffodil International University
Dhaka, Bangladesh
sanzid.swe@diu.edu.bd

***Abstract*—Computer vision learning pay a high attention due to global pandemic COVID-19 to enhance public health service. During the fatality, tiny object detection is a more challenging task of computer vision, as it recruits the pair of classification and detection beneath of video illustration. Compared to other object detection deep neural networks demonstrated a helpful object detection with a superior achievement that is Face mask detection. However, accession with YOLOv3 covered by an exclusive topic which through certainly happening natural disease people get advantage. Added with face mask detection performed well by the YOLOv3 where it measures real time performance regarding a powerful GPU. whereas computation power with low memory YOLO darknet command sufficient for real time manner. Regarding the paper section below we have attained that people who wear face masks or not, its trained by the face mask image and non face mask image. Under the experimental conditions, real time video data that finalized over detection, localization and recognition. Experimental results that show average loss is 0.0730 after training 4000 epochs. After training 4000 epochs mAP score is 0.96. This unique approach of face mask visualization system attained noticeable output which has 96% classification and detection accuracy.**

***Index Terms*—YOLOv3, DNN, face mask detection, mean average precision, GPU, Computer Vision.**

I. Introduction

Global pandemic COVID-19 circumstances emerged in an epidemic of dangerous disease in all over the world. The situation now has been under attack and also growing badly in all over the countries proclaimed by the WHO [1-3]. According to this epidemic beyond 114 countries being affected by this flu-like indications in the body regarding 6.4 days(2-14days). Million numbers of people getting sick in one day. During this disaster time period everyone should raise awareness and naturally should do some oneself activities. By this issue the country's government, social authority and working place should strictly follow necessary rules through continuous measurement and protest people's health. Hereby this germs reaches and viral in any region by one to many and many to million from hand shaking, germs of mouth and exchange accessories with others. On behalf of, nowadays people wearing masks for their own safety are concerned with reducing the flow of germs and the infected number of people reduction. Due to this violent topic we illustrate our work by detecting the mask position of who wears the masks and who are not both in outside and crowded places. The computer vision learning is the actual field to identify the image, convert descriptive image, output analyze and machine acquiring [4]. According to image detection it is identified by the numeric number for ability to understand humans [5]. Vision learning is deeply related to YOLO for any type of image detection [6-7]. Furthermore, YOLO detection is flexible in any place whether group into family members, colleagues and friends in a walk. Declared by the WHO that a potential speech by maintaining distance and wearing a mask is necessary [8]. Wearing a mask captured by the image detection where the machine can cover and translate only the mouth portion of the face part. Computer vision is a following section of Deep learning particularly an area of convolution neural network (CNN) [9]. Added with one main thing is CNN supports very high configuration Graphic Processing Units (GPU) thus as real time image or video extraction of visualisation is a bitter task. As we require people mask having or not which call a surveillance system there is a need for powerful validation such as video stream analysis that is fulfilled by advanced CNN [10]. Now cloud system real time video or image illustration among R-CNN getting complex within create unsatisfied occurrence [11]. Accordingly, new phase included which have consisted 27 CNN layer across 24 convolutional layers that fully connected algorithm is YOLO (You Only Look Once) as it affordable to identified tiny objects [12]. YOLO darknet algorithm while introduced to researchers who covered pretty much different segments of image detection easily specially face detection entirely grabbed YOLO [13].

To the best of our knowledge, throughout this paper we try to fill up inconvenient situations in the present world by the detection of masks in specific regions that people covered their mouth or not. In a sense the community remains safe while increasing flu stops are scattered in the air and make barriers to

enter in the human body. Therefore, modified YOLO detects the human face fully along with identified small regions of mouth from nose to chin area.

II. Literature Review

A succession of study that comes when the justification is unique. YOLO darknet, this is the first dealing with face mask detection by image visualization. YOLOv3 made connections with CNN by hidden layers which through research easily fetch the algorithm and can detect and localize any type image. According to this motivation we demand mask detection as a unique and public health service system during the global pandemic COVID-19 epidemic. As we mentioned below, a comparative study regarding what other researchers have done with CNN based on YOLOv3 and its old version by detecting the several types of image. Added with discussion proceeded by the configuration of CNN, output and trained image details.

Bin Wang et al. [14] approached vision learning concept about tracking accuracies of small and dense objects any movement from video using YOLO applying their own formed dataset. Dongqing Shen et al. [15] originated by Deep learning models are different types of models presented for object detection such as Colored-based model , fuzzy-based model, motion and shape based model and also has another proposal method YOLO(You Only Look Once). It is used for Plume detection with optimizing the Plume datasets and enlarged with the efficiency of YOLO. First A. Jiangyun Li et al. [16] emerged in six cold steel strip surfaces can detect easily with the collaboration of YOLO network with giving a high efficiency rate.During proceduration completed by the 26 CNN (Convolution Neural Network) layers for surface detection with the found of accurate size and specific position. Jiajun Lu et al. [17] rendered by the YOLO object detection based on highlighting their detection on Stop Sign. Which has been insights with the detection of original Stop Sign while the duplicate Stop Sign comes and YOLO misclassified by the duplicate with a frame.This marked as a contradiction of two frames of Stop Sign which YOLO can understand sophisticatedly. Joseph Redmon et al. [18] inferred thus proposed better performance of detection in YOLO on Picasso dataset(5). In her recent paper continually identified and detected a Common Objects in Context dataset mirror) detection dataset' and 'Image net detection dataset'. At first detection of COCO dataset by YOLO rather than detection result was poor on the other hand Image detection dataset gives more efficient output behind on detecting 44 images over 200 images. YOLOv2 technology has been used for these find out objects which can discover 9000 objects. This method is very good performer and faster with the different speed 76.8mAP. At the end of the working procedure joining optimization results based on COCO and Image net dataset are stunning for the image size gap between detection and classification. Depending on this work in future it could be made more elaborate for image segmentation, matching strategies for weak labels and also improve detect result. Joseph Redmon et al. [19] Distinguish between YOLO and R-CNN which are both convenient for object detection none but identify a problem one is more efficient for background image and the other one is more localized objects. Connecting a webcam with YOLO can identify real time object detection. YOLO is faster than R-CNN because it generalizes well and creates new domains for better detection performance. They have used 'Picasso dataset' and 'People-Art dataset' for training on VOC2010(Visual Object Classes). However, comparison with YOLO and R-CNN one can give best performance in all way prediction of bounding boxes and detection also give high efficiency. Mennatullah Siam et al. [20] are detected motion segmentation on static moving vehicles which also a CNN based object detection .For detect object YOLO is one of the systematic models which gives more efficiency. They learned their own dataset which is (KITTI MOD). Rayson Laroca et al. [21] In this work , A new robust real time ALPR (Automatic License Plate Recognition) system based on the YOLO object detection had proposed three segments with a good explanation , elaborating , better recognition rates and also make their own datasets with 4500 images. Through the process lots of practical applications such as road monitoring any kind of detect image it depends on their image resolution , background , camera size so keeping the all the way in their mind they have collected a dataset which has 800 training images. ALPR is coming from DL (Deep Learning) concept and this ALPR approach dataset UFPR-ALPR which is a Brazilian dataset SSIG (Segplate Database) with 800 training image data with a simple background processed by YOLOv2. After that these 800 training data reprocessed by complex background and made a large dataset around 4500 images for filter face positive and processed video over two datasets 800 and 4500 image and detected the several approaches to increase recognition rates around . Achieve redundancy 93.53% for temporal redundancy and Recognition rate below 78.33% with also detect 30000 LP characters. They claimed there could be high possibilities on rate reach and also improve the ALPR pipeline. Shubham Shinde et al. [22] added a typical view of YOLO detection, localize and recognize action with the assistance of video scenes. Each object of video fixing within each single frame for data visualization. Perhaps, it proved the method YOLO is more effective and better fast from others methods by using their own made dataset. The execution process starts with taking 30 single frames of a dataset put into a model after combining results to fetch the prediction of action level. On behalf of runned the query of single CNN archived nearby 88% performance capability.

III. Methodology

In this section we divided our work into two parts. First part we'll be discussing about data acquisition and annotation part. In here we briefly cover about our dataset also pre-processing stuff. Next data annotation we used. There many ways to annotate the data but for our purpose we only care about 3 steps which will be discussed briefly. Later basic introduction about YOLOv3 and configuration discussed with

setup environment and further training procedure introduced. Fig-1 shows the workflow of our architecture.

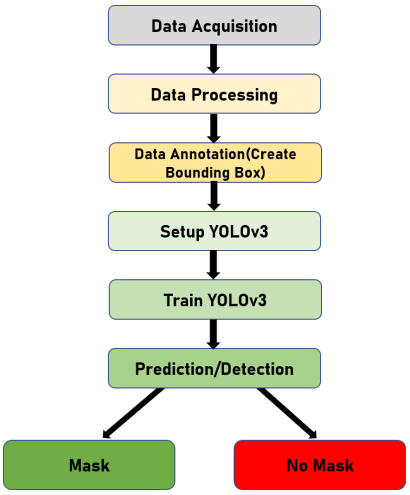

Fig. 1. Workflow Structure.

A. Data Acquisition

Data is really important in data driven techniques like machine learning, deep learning. The more the data the more the better result. For our purpose of working with YOLO we also need more data and with proper annotate. But for our work we don't find any type of annotate data. Using web-scrapping tool from website we have collected 650 images of both mask and no-mask. Further annotation part discussed in data annotation section. Next our data is not suitable for fed into the model. Before feeding we do some pre-processing. There are some irrelevant images inside the dataset. We remove them & finally got our dataset with 600 images where 300 for mask and rest for no-mask. Fig-2 shows the sample of our dataset.

Fig. 2. Dataset.

B. Data Annotation

Data annotation also refers as image annotation in our task. Data is label yet but for our problem we need to annotate them very well for object detection model. Detection is very different than classification tasks. So, data should be well annotated. Different types of annotations found. For our study we need bounding boxes. It's used for create a rectangle area over images that present in our dataset. We've used a tool called LabelIMG [23] to annotate our data. This process is labor-intensive and time consuming. Mention key characteristics that follow here :

- If only single image present then there is no issue, we just draw box around them (see upper left, right)
- If multiple objects present in a image like mask or no mask, then create bounding boxes both of them. (see down)
- If there are any blurry content present then we have to identify the objects presents here. If we can't identify ignore them. (see down)

The Fig-3 below shows some images that we have annotated.

Fig. 3. Bounding Boxes.

C. System overview

Here we divided the whole things into two parts. First will be covering setup of YOLOv3 for our problem next will be our applied yolov3 .

1) YOLOv3 Setup: Joseph Redmon et al. introduced You look only once also known as YOLO in 2015 [19]. Later some improvements came into them and YOLOv2 and YOLOv3 introduced respectively in 2016[18] , 2018[21] . Now, YOLOv3 is the state of art object detection model

followed by other versions of YOLO and YOLOv2 . It's been given amazing results regarding object classification and detection.
In previous version of Yolov2 Darknet-19 is used as a feature extractor. In yolov3 it changes with some improvements and they called it as darknet-53. Darknet is a framework for training neural network that written in c language which performs better in these tasks.
Before working with this architecture some steps we need to mention –

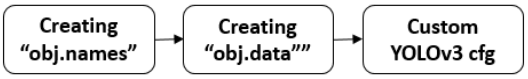

Fig. 4. Configuration steps.

From above Fig-4 of YOLOv3 configuration at first create a "obj.names" files which contains the name of the classes which model wanted to detect. Then a obj.data file which contains number of classes in here it is 2, train data directory, validation data, "obj.names" and weights path which gonna save on backup folder. Last a cfg file contains 2 classes. Next We change batch size as 64 and subdivisions as 16. For three yolo block set class as 2 and the previous convolution block set filter size at 21. Max batches for our case is 4000 which is calculated as "number of class x 2000".

2) Applied YOLOv3: An input here is an image is passed into the YOLOv3 model. This object detector is going through the image and find the coordinates that present in an image. It's basically divides the input into a grid and from that gird it'll analyzes the target objects features. From the neighboring cells that features were detected with high confidence rate are add at one place for produce model output. Fig-5 shows how YOLOv3 works in our purpose.

Fig. 5. Workflow of YOLOv3.

From above Fig-5 first image shows how actually model divided image as a grid. Next shows how it's find the features and last shows the detected object.

IV. Experimental Result & Discussion

After all the setup has been done our custom model is ready for training. Unlike other networks YOLOv3 used logistic regression as loss function. As far our resources is limited we used Google co-laboratory for our training purpose. 80% data used for training , rest used for validation. Over 4000 epochs of training we got a good accuracy of 96% and average loss is reduce to 0.0730 and our mean average precision score is 0.96.
After training with test data our model also detect the object more accurately. Fig-4 shows the average loss curve in our model.

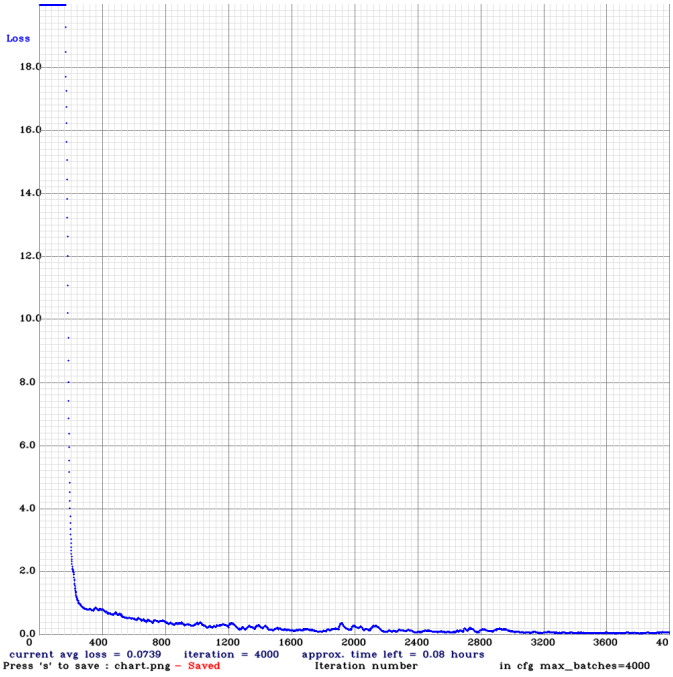

Fig. 6. Average loss curve.

Fig-5 shows the input image and the predicted output of that model. We see that the input images represented as no mask and our model also detect this very well as no mask.

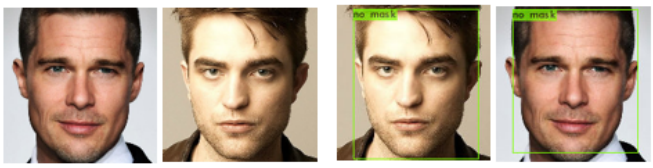

Fig. 7. No-mask Detection.

Fig-6 shows the input image and the predicted output of that model. We see that the input images represented as mask face and our model also detect this very well as mask face.

Fig. 8. Mask Detection.

Further we have also introduced our model into real-time video and got a impressive average 17 fps rate.

V. Conclusion

In our study, we have introduced an approach for detecting a person is wearing a mask or no mask using state of art YOLOv3 architecture. It's perfrom really well in images and our detection results was also quite good. Later we also applied this model into a real-time video to check whether our model's fps rate inside video and its detection performance with two classes mask/no mask. Inside video our model get impressive output where average fps is 17. In this piece of research mainly focus on building a custom object detection model using YOLOv3 not to make the whole architecture. Though the dataset that we collected is not that much variety after all it gives us promising accuracy in testing with some real world data.

VI. Future work

In future we will add more data to get more accurate result in detection. As far our resources is limited we can't get higher fps rate in video. Future we will be train and evaluate our model into a better machine. Recently, more object detection architecture i.e. Mask RCNN, Faster RCNN etc are introduced. A new version of YOLOv4 also come into play recent couple of days. We will apply these models for compare the performances all of them.

References

[1] Hongzhou Lu, Charles W. Stratton, and Yi-Wei Tang. Outbreak of pneumonia of unknown etiology in Wuhan, china: The mystery and the miracle. Journal of Medical Virology, 92(4):401–402, 2020.

[2] Chih-Cheng Lai, Tzu-Ping Shih, Wen-Chien Ko, HungJen Tang, and Po-Ren Hsueh. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. International Journal of Antimicrobial Agents, 55(3):105924, March 2020.

[3] Hussin A. Rothan and Siddappa N. Byrareddy. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of Autoimmunity, page 102433, February 2020.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," In Conference on Computer Vision and Pattern Recognition, 2014.

[5] Reinhard Klette, "Concise Computer Vision". Springer, 2014.

[6] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglud, Ozal Yildirim, U. Rajendra Acharya", "Automated detection of COVID-19 cases using deep neural networks with X-ray images", 2020.

[7] Joao Carlos Virgolino Soares, Marcelo Gattass, Marco Antonio Meggiolaro, "Visual SLAM in Human Populated Environments: Exploring the Trade-off between Accuracy and Speed of YOLO and Mask R-CNN", 19th International Conference on Advanced Robotics (ICAR), 2019.

[8] L. Hensley, "Social distancing is out, physical distancing is inheres how to do it," Global News–Canada (27 March 2020), 2020.

[9] S. Wan, Y. Liang, Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification", Comput. Electr. Eng. 72 (2018) 274–282.

[10] A. Koubaa, B. Qureshi, M. Sriti, Y. Javed, and E. Tovar, "A service oriented cloud-based management system for the internet-of-drones", 2017 IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC 2017, April 26-28, 2017, pp. 329–335, 2017.

[11] S. Ren, K. He, R.B. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", CoRR abs/1506.01497, 2015.

[12] E. Dong, Y. Zhu, Y. Ji, S. Du, An improved convolution neural network for object detection using YOLOv2, 2018 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 1184–1188, 2018.

[13] M.H. Putra, Z.M. Yussof, K.C. Lim, S.I. Salim, "Convolutional neural network for person and car detection using YOLO framework", 67–713, 2018.

[14] Bin Wang, S. T.-B.-F.-D., "Detection and tracking based tubelet generation for video object detection. Journal of Visual Communication and Image Representation", 102-111, 2019.

[15] Dongqing Shen, X. C., "Flame detection Using deep learning", 2018 4th International Conference Control, Automation and Robotics, IEEE, 2018.

[16] First A. Jiangyun Li, S. B., "Real-time Detection of Steel Strip Surface Defects Based on Improved YOLO Detection Network", IFAC-PapersOnLine, (pp. 76-81), 2018.

[17] Jiajun Lu, H. S., "No Need to Worry about Adversarial Examples in Object Detection in Autonomous Vehicles", Computer Vision and Pattern Recognition (cs.CV), 2017.

[18] Joseph Redmon, A. F., "YOLO9000 :Better, Faster, Stronger.", arXiv:1612.0824v1 [cs.CV], Washington: IEEE, 2016.

[19] Joseph Redmon, S. D., "You Only Look Once: Unified, Real Time Object Detection" IEEE, 2016.

[20] Mennatullah Siam, H. M., "MODNet: Motion and appearance based Moving Object Detection Network for autonomous Driving", Computer Vision and Pattern Recognition (cs.CV), IEEE, 2017.

[21] Rayson Laroca, E. S., "A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector", 2018 International Joint Conference on Neural Networks (IJCNN), 2018.

[22] Shubham Shinde, A. K., "YOLO based Human Action Recognition and Localization", International Conference on Robotics and Smart Manufacturing, pp. 831-838, 2018.

[23] "tzutalin/labelImg", GitHub. https://github.com/tzutalin/labelImg. [Accessed: 09- Jun- 2020].