2021-11

# Social and Safety Monitoring for Pandemic with YOLO

Chung, CL

http://hdl.handle.net/10026.1/18306

Journal of Industrial Mechatronics

# Social and Safety Monitoring for Pandemic with YOLO

Chia-Ling Chung, Hooman Samani, Li-Yu Yu and Chan-Yun Yang

*Abstract*—**During the pandemic of COVID-19, people have been suggested to keep social distance from the others. It is also beneficial to pay attention to the individuals with motion irregularities. In this research, we propose a visual anomaly analysis system based on deep learning with the aim to identify individuals with various types of anomaly which are specifically important during the pandemic. Based on the proposed monitoring system it would be easier to keep tracking the environment changes, and it would also be beneficial to the safety guard to reallocate resources accordingly to relieve the threat of anomaly. Types of the anomaly are very sensitive during the coronavirus pandemic. In the study, two types of anomaly detections are concerned. The first is monitoring the abnormally in the case of falling down in an open public area, and the second is measuring the social distance of people in the area to keep warning the individuals under an insufficient distance. By the implementation of YOLO, the related anomaly can be identified accurately in a wide range of open area. The reliable results make promisingly the use of a vision sensor as a ranger to detect anomaly in time in the open area. Through the implemented system to monitor the environment, the safety monitoring would be easier to manage the anomaly around a neighborhood which may help to avoid the spread of the virus.**

## I. Introductions

Coronavirus is an important pathogen causing human and animal diseases. In the past, several coronaviruses have been known to cause respiratory tract infections, such as Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS). The most recent suffering is COVID-19. As elapsed, there were already more than 150 million infected and 3.2 million death in early May 2021[1]. As a scientific brief released from Centers for Disease Control and Prevention (CDC) USA along May, 2021 [2], the infectious disease can even be transmitted through the respiratory fluids including the air carrying very small fine droplets and aerosol particles that contain the virus within three to six feet of an infectious source. As shown in Fig. 1, the epidemic outbreak of COVID-19, spreading almost to every regions in the world, has globally endangered the human lives severely.

As the epidemic is more difficult to control, the health risk of medical personnel who come into a contact with the infected people are also increasing. Various artificial intelligence (AI) based technologies are hence introduced to manage and reduce the risk, for example, UAVs with camera that support facial recognition from air [3], fusing face recognition and temperature measuring in an infrared thermal camera [4], and numerous regionally official apps in tracing out the possible contact with an infected individual [5]. However, the AI technologies provide variety of advanced functions in helping to identify the suspicious hot points [6-7] with the disease, and to find a new treatment method to track the spread of the disease.

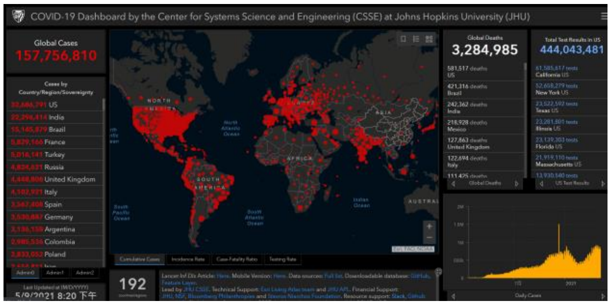

Figure 1. Update distribution numbers of infections and deaths worldwide of COVID-19 pandemic (as of 2021/05/09) [1]

## II. Research Materials and Methods

### A. The object detection YOLO

Nowadays, varieties of one-stage models have been developed and used in many mobile devices in which YOLO with its extensions is the most famous one due to its superior applicability. As it can generally reach 45 frames per second on GPU, and one simplified version can even reach 150 FPS, the identification rate of YOLO is very quick. With the excellent identification rate, YOLO with its potential are able to be applicable in many real-time video applications. Also as mentioned in [8], the deep neural networks are the most powerful machine learning scheme for an excellent object detection missions in the field of computer vision, and among them, YOLO is the one of the latest technology based on the deep neural network. The speed and accuracy confirms also the applicability to a multi-task detection, such as the pedestrian detection [9] and the underwater robot perceptron and control [10].

There are mainly four versions of the YOLO series. Developed based on GoogleNet, YOLOv1 has 24 convolution layers and 2 fully connected layers. The activation function uses Leaky ReLu, and the last layer uses linear activation. YOLOv2 adopts a basic Darknet-19 model to design its prototype to be consistent with VGG16. Together with improved speed, accuracy and recognition types, the mean Average Precision (mAP) increased from 63.4 to 78.6 in the

Chia-Ling Chung, is with Department of Electrical Engineering, National Taipei University, New Taipei City, 23741 Taiwan (e-mail: permanent2001@gmail.com).

Hooman Samani, is with School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth PL4 8AA, UK (e-mail: hooman.samani@plymouth.ac.uk).

Li-Yu Yu, is with Department of Electrical Engineering, National Taipei University, New Taipei City, 23741 Taiwan (e-mail: 94liyu@gmail.com).

Chan-Yun Yang, is with Department of Electrical Engineering, National Taipei University, New Taipei City, 23741 Taiwan (e-mail: cyyang@mail.ntpu.edu.tw).

benchmark of VOC2007 dataset. YOLOv3, comparing with the previous two versions, has no big revolutionary innovation. It mainly optimizes the previous two versions of the model, amends multi-scale prediction, and improves the detection of small objects. In the case of YOLOv4 [11], the model can be trained faster on a single GPU. YOLOv4 improves the input of training, so that training can also have good results on a single GPU. For example, Mosaic data augmentation, cross mini-Batch Normalization (cmBN) and self-confrontation training (SAT). Traditionally, the neural network based YOLO consists of an input layer and an output layer, and in which there is at least one intermediate hidden layer. The advantage of YOLO is that it greatly reduces the requirements of hardware, and can greatly improve the detection accuracy of the model under a certain level of speed. As can be seen from Fig. 2, on the MS COCO dataset, YOLO obtained 43.5% Average Precision (AP) value. At the same efficiency, YOLO is twice as fast as EfficientNet.

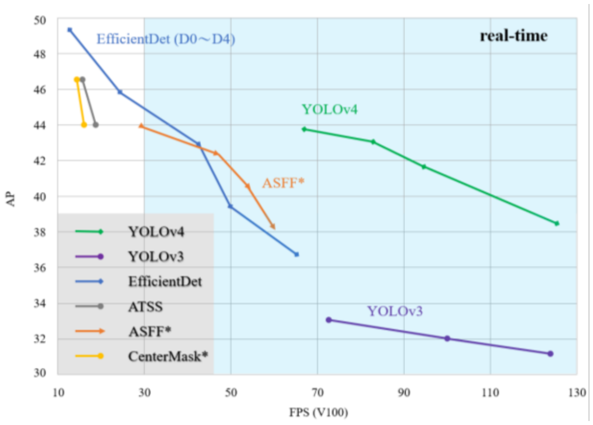

Figure 2. Comparison of speed and accuracy of YOLO [5]

### B. *Performance evaluation*

The method of YOLO is to cut the input image into S×S grid. If the center of the detected object falls into a grid, the grid is responsible for detecting the object. Each grid is responsible for predicting bounding boxes and probability belonging to different categories. The prediction of each bound box will output 5 prediction values: $x$, $y$, $w$, $h$ and *confidence* where $x$ and $y$ represent the ratio of the central coordinates of the bounding box to the width and height of the image which are normalized central coordinates of the bounding box, $w$ and $h$ denote the ratios of the width and height of the bounding box to the width and height of the input image which are normalized width height coordinates of the bounding box, and *confidence* is the IOU value representing the ratio of area of bounding box with respect to the area of ground truth.

The major evaluation indices of the study are IOU (Intersection Over Union) and mAP. IOU, as its name, is the intersection of two bounding boxes divided by the union of two bounding boxes.

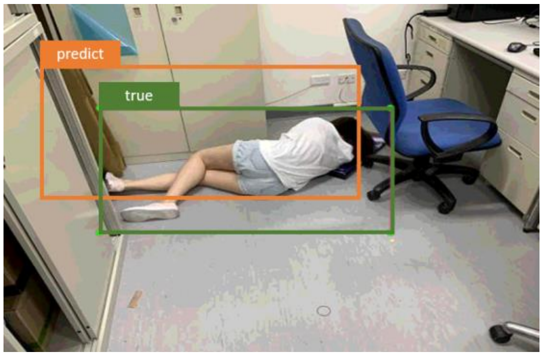

Figure 3. Intersection over Union

That is, the intersection of predicted bounding box and ground truth bounding box is divided by the join set:

$$\mathrm{IOU} = \frac{\text{Intersection of two bounding boxes}}{\text{Union of two boiunding boxes}}$$

The mAP is the mean value of AP in each category, while AP is the area under curve (AUC) of PR curve (precision recall curve). The PR curve is a curve drawn with recall as X axis and precision as Y axis. The higher precision and recall are, the better the model performance is.

## III. Implementations

### A. *The proposed system*

We will describe in detail the abnormalities that may occur during a coronavirus pandemic. Figure 4 is our process architecture. Anomalies are mainly divided into action and distance, and then subdivided into the following steps.

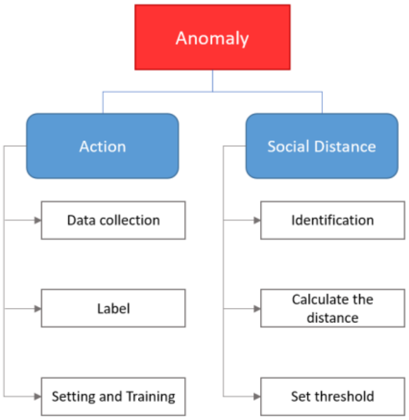

Figure 4. The overall structure of the proposed system

#### 1) *Anomaly action detection*

Coronavirus pneumonia is mainly attacked by lung cells, so it may cause sudden shock in infected people. In addition, patients with certain degree of discomfort may also lead to collapse, dizziness and fainting. We regard this behavior as abnormal. It is mainly aimed at the behavior of fainting caused by possible physical discomfort during the epidemic period.

Although ImageNet [12] and COCO data [13] are the first choice for most human training models, there are few human motion images. Finally, we searched the Shutterstock website for about 500 abnormal action pictures. In addition, we prepared our own materials as a training set, with a total of about 700 photos. In order to improve the accuracy, we adjust the image size to 1,024×1,024.

Then, Label Image [14] tag tool is used to manually label the feature category. Through this interface, we can frame and classify the exceptions in the screen, and specify the storage folder. Finally, write the data path to the model to start training.The maximum number of weight updates will be multiplied by 2,000 according to the category, but the minimum number should not be less than 4,000, so it is set to 4,000. The width and height of image output will be set to 1,024 as a multiple of 32. Steps is to set the change of learning rate. Steps will be 80% and 90% of the maximum number of weight updates. Then, the training set path and test set path are specified by using Darknet53.

When a person is in an abnormal situation, it may be covered by obstacles such as chairs, so that people nearby cannot detect it in time. Therefore, we try to detect the abnormal movement according to the different degree of obstacle cover, which are body cover 70%, body cover 50%, body cover 30% and body cover 0%. Figure 5 shows that the trained model can clearly identify these abnormal conditions.

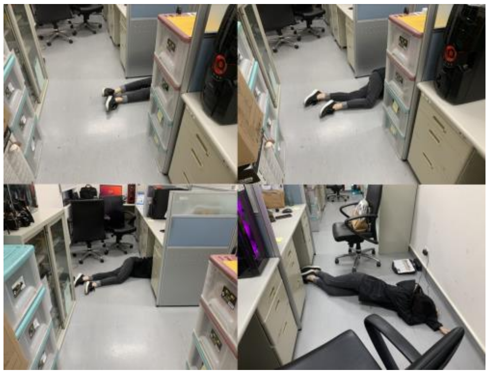

Figure 5. Abnormal actions with different degrees of concealment

Figure 6 shows that the trained model can clearly identify these abnormal conditions. It can be seen that our model can identify the anomalies of various occlusion effects.

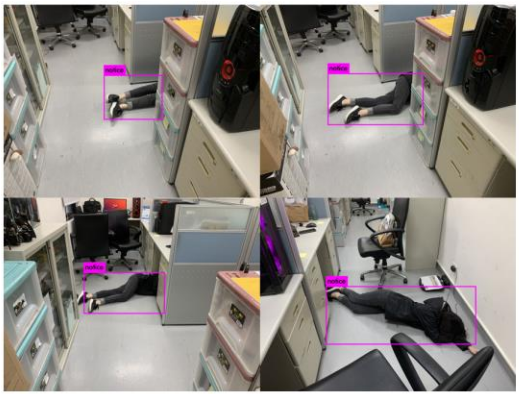

Figure 6. They were body cover 70%, body cover 50%, body cover 30% and body cover 0% of the abnormal behavior detection results

Figure 7 shows the training results of abnormal behavior. The blue line is the loss function, and the red line is the accuracy of judging the abnormality. The key point for a model to learn the characteristics is the loss function. When the loss function approaches the smaller value, the better the performance of the model. We can see that with the passage of time, The loss slowly drops to the position close to 0.5, and there is no overfitting, and the accuracy of the model is as high as 91%.

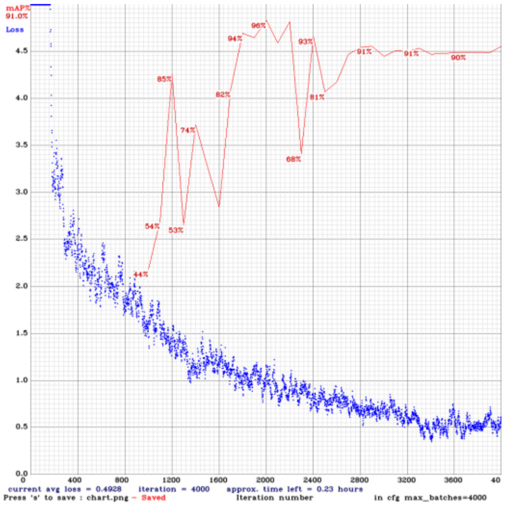

Figure 7. Covergence of the training procedure

## 2) *Social distance measurement*

It is necessary to keep social distance to protect the transmission of viruses. As for the definition of social distance, each country has different norms. In Taiwan, the recommendedsocial distance is 1.5 meters, which is equivalent to the distance between two hands of one person. First of all, we can use the labeled data in YOLO to detect whether there are people in the picture. If someone in the picture will box, make sure that at least two people are detected, and then extract the boundary box coordinates. Then we can convert the central coordinates into rectangular coordinates, and then obtain the framed centroid, namely x1 and y1 punctuation positions.

After obtaining the centroid x and y, we use Euclidean formula to calculate the linear distance between people. The pixel value is converted into the distance from the point to the nearest background point, and whether the distance between two people is less than N pixels is detected. If it is less than, it is at a safe distance, otherwise, it is not.

In the formula, d is the distance, x1 is the centroid x value of the first box, y1 is the centroid y value of the first box, x2 is the centroid x value of the second box, and x2 is the centroid y value of the second box. And so on to calculate the distance between the boxes.

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)}$$

It can be directly seen that two people on the left are too close, so a red box will be displayed. On the right, a green box will be displayed for safety status. Finally, the upper left corner will output a picture showing several people who do not meet the requirements of social distance.

Next, the social distance threshold is set to 1.5 and check that it meets the conditions. Finally, we set the color for the bounding box. Red indicates the person in danger, and green indicates the person under protection.

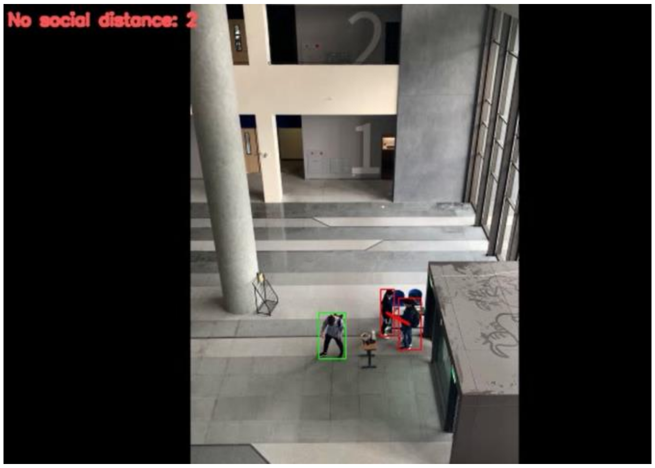

Figure 8. Social distance detection at the bulding eneterance

Finally, we combine the two functions as shown in the figure 9. It is shown that the two people on the right do not keep social distance, so they will be detected as red boxes. Although the person on the left of the picture keeps social distance, it is detected as abnormal because the person falls to the ground. In the upper left corner of the figure, the number of people who have not kept social distance will be calculated and displayed.

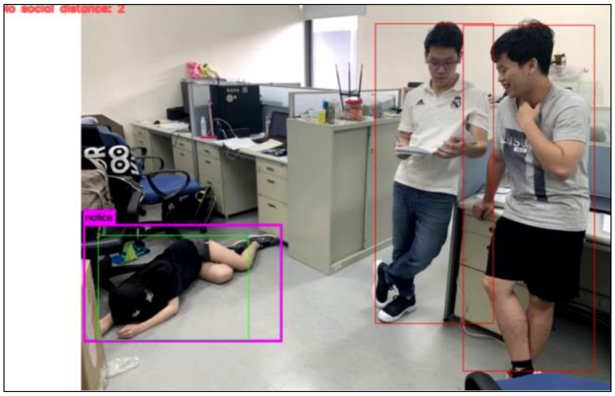

Figure 9. Result of the integrated final system in progress

## IV. Conclusion

In this research, we have explored the abnormal conditions and phenomena of people during the pandemic, and used deep learning vision system to solve social and security problems. This system not only could detect people's abnormal behavior in the public, but also could monitor the social distance between people and reduce the risk of infection and social resources.

In the future, it is expected that this vision system can be applied to robots or surveillance systems. In addition to enhancing the recognition accuracy of system detection, depth cameras are also considered to replace pixel based distance measurement.

## References

[1] E. Dong, H. Du and L. Gardner, “An interactive web-based dashboard to track COVID-19 in real time”, The Lancet infectious diseases, 2020, pp. 533-534.

[2] Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Scientific Brief: SARS-CoV-2 Transmission, https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html

[3] H. J. Hsu and K. T. Chen, “DroneFace: an open dataset for drone research”, In Proceedings of the 8th ACM on multimedia systems conference, 2017, pp. 187-192.

[4] Z. Xie, P. Jiang and S. Zhang, “Fusion of LBP and HOG using multiple kernel learning for infrared face recognition”, In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 2017, pp. 81-84.

[5] L. Ferretti et al., “Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing”, *Science*, 2020, 368.6491.

[6] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini and J. Wu, “How big data and artificial intelligence can help better manage the COVID-19 pandemic”, *International journal of environmental research and public health*, 2020, 17(9), 3176.

[7] C.-L. Chung, D.-B. Chen and H. Samani, “Action Detection and Anomaly Analysis Visual System using Deep Learning for Robots in Pandemic Situation”, 2020 International Automatic Control Conference (CACS), 2020, pp. 1-6, doi: 10.1109/CACS50047.2020.9289819.

[8] M. J. Shafiee, B. Chywl, F. Li and A. Wong, “Fast YOLO: A fast you only look once system for real-time embedded object detection in video”, arXiv preprint arXiv:1709.05943, 2017.

[9] W. Lan, J. Dang, Y. Wang and S. Wang, “Pedestrian detection based on YOLO network model”, In 2018 IEEE international conference on mechatronics and automation (ICMA), 2018, pp. 1547-1551.

[10] J. Yu, X. Chen and S. Kong, “Visual Perception and Control of Underwater Robots”, CRC Press, 2021.

[11] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection”, arXiv:2004.10934, 2020, [online] Available: http://arxiv.org/abs/2004.10934.

[12] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database”, In 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248-255.

[13] T. Y. Lin, et al., “Microsoft coco: Common objects in context”, In European conference on computer vision, Springer, Cham, 2014, p. 740-755.

[14] Labelimg tool. Tzutalin - Git code (2015): https://github.com/tzutalin/labelImg.