Intelligent System for Social Distance Monitoring with Human Detection and Tracking using YOLOv3

Dhanya G Nair [1] , Dr.K.P.Sanal Kumar [2] , Dr.S.Anu H Nair [3]
[1]Research Scholar, Department of Computer and Information Sciences, Annamalai University, Chidambaram, India.
[2]Assistant Professor, PG Department of Computer Science, R.V Government Arts College, Chengalpattu, India.
[3]Department of CSE, Annamalai University, Chidambaram, India [Deputed to WPT, Chennai].
E-Mail:[1]dgn123.dhanya@gmail.com,[2]sanalprabha@yahoo.co.in,[3] anu_jul@yahoo.co.in

Abstract:

Everyday life and the global economy have been negatively impacted by COVID-19 (Coronavirus). Slowing the spread of coronaviruses through social distance is proven to be an effective strategy in the war against COVID-19. The social distancing is the best way to stop the spread of COVID-19, as it prevents people from coming into intimate touch with each other. Recently, due to the fast spreading outbreak of the COVID-19, one of the obligatory preventive measures to avoid physical contact has become social distance. Surveillance methods that use Deep Learning, Open-CV and Computer vision to follow pedestrians and prevent congestion are the focus of this article. Closed-circuit television (CCTV) and drones can be used for implementation, where the camera will use object detection to identify the crowd and compute the distance between the humans. Local law enforcement will be notified if the Euclidean distance between two persons is less than the standard distance, which is determined by converting it to pixels and comparing it to that value.

Keywords: Deep learning ,CCTV, Open CV, Social distancing, , Computer vision.
DOI Number:10.14704/nq.2022.20.8.NQ44399 **NeuroQuantology 2022; 20(8): 3688-3698**

1 Overview

Social isolation is one of the best ways to stop the spread of COVID-19 because there is no vaccine. Because of the pandemic, there is no vaccination available. People are encouraged to keep a certain distance from one another in order to practise social distancing, which is a concept whose name gives it away. The number of reported instances has been rising at a very rapid rate all across the world, and as a result, maintaining social distance is essential. This research study offers a locating approach that may be used to detect social distance in public spaces. During this epidemic era, we are able to maintain a check on human activities at public locations by employing CCTV and drones. As a result, we are able to compute and summarise the distances between humans andsocial distancingmonitoring over the city.

Furthermore, a recent study [3] found that with the current critical care facilities in the United States, one-time assessments would be insufficient to reduce the occurrence of COVID-19. As per the research, these measures may be necessary all the way up to 2023. "Physical distancing" or "Social distancing" measures support straighten the transmitting curve, reducing the burden on overburdened critical care units, especially in countries with insufficient public health and testing resources. A single-time social distancing measure will not be sufficient to prevent the COVID-19 epidemic from overwhelming intensive care resources because it will keep enough people alive to be influenced by the gradual recovery in transmission following the end of this pandemic cycle [4].

1.1 Need of social distancing

After a restart of economic activity, it was discovered that social isolation and self-isolation were the most effective strategies of reducing the impact of this coronavirus pandemic. Self-isolation was also proven to be effective. In point of fact, it has been found that a significant number of people are ignoring public health measures, notably those that attempt to decrease social isolation. People who are enthusiastic about returning to work are likely to forget or ignore the necessity of maintaining social distance between themselves and others.

1.2 Motivation

There are a number of existing approaches to combating COVID-19 situations that are more focused on predicting the spread of infection, classifying standard and COVID infected people, monitoring the performance of COVID-19 detection models, and density detection to measure social distance. Nevertheless, the sharing of real-time video feeds and images with the relevant authorities is critical in the fight against the COVID-19 pandemic by penalising those who violate social distance norms. As a result, we feel compelled to come up with a system that streamlines and streamlines this process.

1.3 Contribution

Following are the contributions of this research paper.

- Artificial Intelligence (AI) techniques could be used to alleviate social isolation during the COVID-19 pandemic.
- For COVID-19, we propose an intelligent social distance monitoring system that uses the YOLOv3 Model to detect and track people.
- Lastly, we examine the proposed scheme's performance by comparing the various datasets.

1.4 Organization of the research paper

Here are the rest of the sections of the paper's structure: Section 2 is devoted to works in this prestigious field of study. Section 3 introduces the suggested method's system design, which is followed by Section 4 on system implementation. Section 5 covers the findings and analysis of the suggested method's system testing. Section 6 concludes with thoughts about the future.

2 Related works

The importance of maintaining social distance among individuals in order to prevent the COVID-19 virus from continuously spreading has led to the introduction of a number of programmes that aim to execute and enforce this practise. In addition to the use of authority to ensure people are following the rules about social distance, it has also been considered using the most recent technology, such as the internet of things (IOT), in order to enhance the amount of people who comply with the laws regarding social distance.

2.1 Survey using machine learning methods

Using raspberry pi and Open-CV, Dr. S Syed Ameer Abbas and his co-authors suggested a system for tracking people and crowd control in 2017. A cascade classifier was trained to detect the head in the scene using Haar features in OpenCV. A camera and a Raspberry pi3 with a quad core ARMv8 central processing unit were used to capture video of the busy scene and process it frame by frame. By comparing the number of heads to the threshold, the crowd may be regulated and prevented if it exceeds the threshold [1]. Using image processing, Joel Joseph Joy and his colleagues devised a technique for detecting traffic intensity in 2018. The length of the wait and the density of the traffic were recorded using photos obtained by the camera. Fuzzy logic was applied to the video input to deal with the concept of incomplete truth. [2] The result of the partial truth concept could be either totally true or completely untrue.

Toward the end of 2020, Adrian Rosebrock[3] published a paper on social distancing detector that is based on OpenCV, Computer Vision, and Deep Learning. The essay focuses on social distance monitoring with street-installed CCTV cameras and sheds light on social distancing throughout the epidemic period. When the camera takes a picture of a person and compares it to a standard measurement, it acts as a distance detector. The file.py script contains the logic for this social distance detector programme, which ensures that people keep a safe distance from one another by looping over frames in the video stream. Video files and webcam broadcasts are also supported [6].

2.2 Survey using Deep learning methods

Human detection using deep learning is the focus of this section. [8] Deep learning is also mentioned in a large number of current object categorization and detection studies. There is a lot of focus on the latest research in object detection using machine learning as part of this review. [9] As a computer vision task, human detection can be considered an object detection for classifying and locating the shape of video images in real time. Multi-class object recognition and detection has become a hot topic in artificial intelligence research thanks to advances in deep learning. [10] An in-depth review of the current state of human detection technology was presented in this study by Nguyen et al.. Human descriptors, machine learning techniques, occlusions, and real-time detection are the primary emphasis of the survey. Techniques utilising deep convolutional neural networks (CNN) for visual recognition have been demonstrated to perform better on numerous picture recognition benchmarks [11].

Multiple convolutional layers, sub-sampling layers, and fully linked layers make up the multilayer perceptron neural networks in Deep CNN. [12] After then, the weights in all of the layers of the networks are trained to recognize different objects based on their respective datasets. In deep learning, the CNN model was one of the types of supervised feature learning approaches that are robust in detecting the object in varied situations. [13] Due to the new high-performance computer technology and massive datasets like Image Net, CNN has gained remarkable success in large-scale image classification jobs. In terms of network architecture, methods, and new ideas, a variety of CNN models for object detection with object localization have been developed. Alex Net, VGG16, InceptionV3, and ResNet-50 are some of the CNN models that have recently been trained to generate exceptional results in object recognition [15].

It is because of the neural network structure of deep learning's success in object recognition that it is able to self-build object descriptors and learn high-level features that are not directly provided in the dataset. In terms of accuracy and speed, current state-of-the-art object detectors with deep learning have several advantages and disadvantages to consider. The object may be positioned in a variety of ways in the image. Since the development of the real-time object identification algorithms based on the CNN model like R-CNN [16] and YOLO [18], multi-class detection has been further explored. For deep CNN object identification, YOLO (You Only Look Once) is the most often used algorithm. Figure 1 shows the YOLO model in action. Based on [20], we describe a computer vision technique for people detection using a camera positioned on the roadside or at the workplace that adapts the original concept. In the camera's vision, the persons who are walking in a certain area are captured. The YOLO approach was used to detect the video stream acquired by the camera in order to use these existing deep CNN algorithms to determine the number of persons in an image or video using bounding boxes. The software will show whether or not there is enough social distance between the persons in the video by measuring the Euclidean distance between them.

3 Methodology

Yolo v3, a fully convolutional neural network technique for recognizing individuals in video frames, is employed in the proposed work. For the detection of persons in video frames, Yolo v3 uses a fully convolutional neural network technique. A real-time computer vision library, Open CV, is utilised to feed the yolo v3 neural network, which in turn feeds the output frames to the cameras, the input frames from the photos or videos collected by cameras. The system will do statistical analysis of the area collected by the cameras and report back to the user after the results for the selected region have been presented. The Yolo v3 recognition engine is used in the algorithm's object detection phase and can identify up to 80 different things from a given batch of input photos and videos. A 53-layer network trained on Imagenet is the first and most important feature of Yolo v3. It employs a darknet variation for its operation. On the other hand, 53 extra layers are added to the preceding set of layers for the purpose of detecting an item. As a result, the Yolo v3 network contains 106 layers.

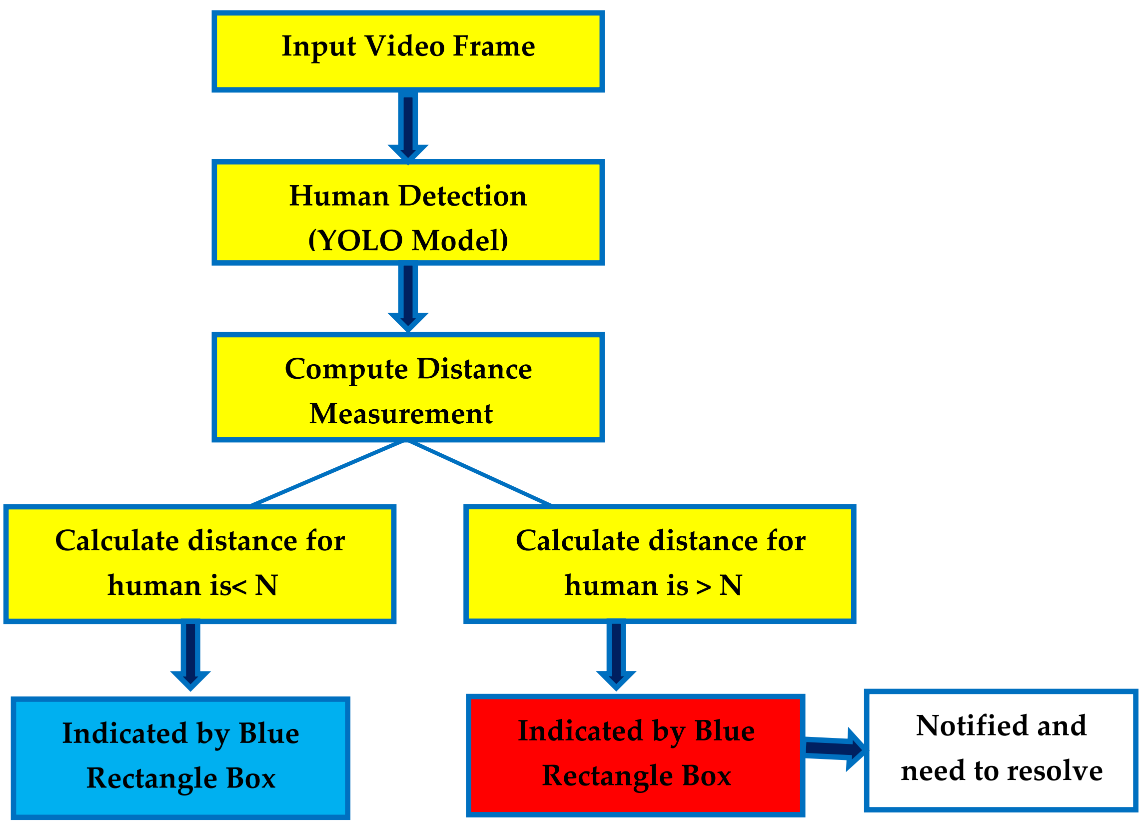

Figure 2: General architecture of proposed method

If the object's centre is found to be within a grid cell, then that grid cell is the one that will be held responsible for the detection and location of the object. The algorithm employs a threshold to predict several labels for an object, and it uses logistic regression to predict the scores for each class in order to forecast each label. Together, these two techniques are used to predict each label. Classes that have scores higher than fifty percent, which is considered to be a confidence parameter, are used to filter out the things in the image that belong to the "people" categorization. This is done in order to focus on the people-related objects. It is possible to calculate the interpersonal distance between people by comparing the distance between centroids, or more specifically, by comparing the distance between the bounding boxes contained inside the image. If it is detected that the distance is larger than a specified threshold, then it is regarded to be in breach of social norms regarding appropriate distances, and as a result, the image will display a red bounding box to indicate that it is hazardous. People who are safe to be around are indicated by the presence of a green bounding box around them as long as the social distance guidelines are followed.

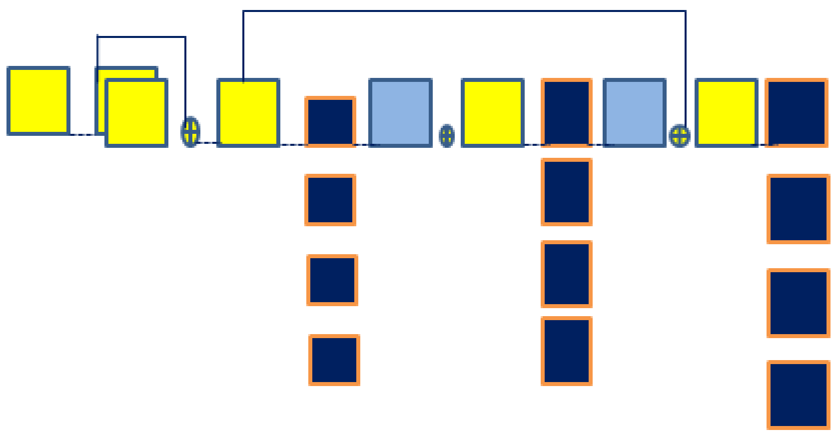

Figure 3: YOLO V3 architecture

The following list is comprised of the necessary mathematical equations for the system: Using a method based on probabilities, the YOLO v3 network makes four predictions regarding the coordinates of each bounding box.

$$z_x = \sigma\,(s_x) + m_x \quad \text{(i)}$$
$$z_y = \sigma \textstyle\sum(s_y) + m_y \quad \text{(ii)}$$
$$z_h = q_h e^{s_h} \quad \text{(iii)}$$
$$z_w = q_w e^{s_w} \quad \text{(iv)}$$

The zx, zy, zh, and zw represent the x, y, and height, as well as width, of our prediction's centre, while the sx, sy, and sw represent the grid's outputs. The variables mx and my indicate the network's upper left coordinates. Its dimensions are defined by the anchors qh and qw. We use the following formula to calculate the distance between the centres of each box that represents a person:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

With the help of the aforementioned equation, we are able to determine the distance that separates the two individuals who are symbolised by the centres of their respective bounding boxes.

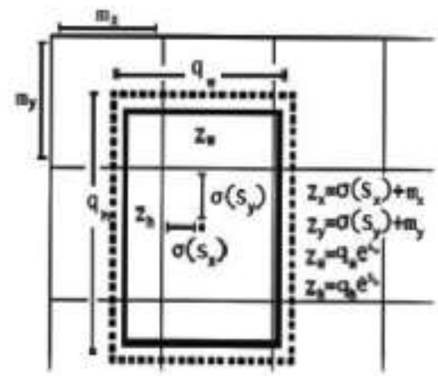

Figure 4: Location prediction with coordinators in bounding boxes

3.1 Dataset

There is a Penn-Fudan Database that contains photos utilised for pedestrian detection in the tests. The photos were taken on campus and in the city. Pedestrians are what we're looking for in these photographs. Pedestrians will be present in every image. The heights of tagged pedestrians in this database fall into [180,390] pixels. Pedestrians classified as such are all straight ahead. All 345 tagged pedestrians may be found in 170 photos taken around Fudan University and University of Pennsylvania, with 96 images recorded surrounding the latter.

Figure 5: Sample images from Penn-Fudan Database

4. Implementation

The flowchart below illustrates the processes required to process the frames and identify members of the "people" class from those processed frames. Following the flowchart below, you can see an overview of every stage in the system. OpenCV is used to take an image or a frame from a real-time video as input in the initial step of the system. To save time, the system doesn't run the process every frame. The video frame will be used to build a blob that will be used to process our frame. An picture can be processed using an Image blob to execute operations like mean subtraction, scaling, and swap. A neural network called Yolo is used to preprocess images before they are sent back into the system. Any object with a greater than 50% confidence level is excluded from our Yolo neural network's detection procedure.

After determining that individuals are an object, the system will generate a bird's-eye view image. Aside from that, four places on the final image would be selected by the algorithm as the corners of the image to be transformed. Points on opposite sides of a square must be aligned in order to produce a rectangle. In order to create a bounding box, the system returns two points for each person it detects. These two points represent the box's upper left and lower right corners. To begin, the midpoint of the line linking these two opposed sides will be determined. Once a threshold of 2 feet or around 120 pixels has been established, the distance between two centroids, also known as two people, can be calculated. If there are more violations than the number of violations that was specified beforehand, an email warning will be issued to the appropriate authorities, allowing the concerned authority to easily keep an eye on the situation in that particular area and this is the fourth phase in the process.

4.1 Performance Evaluation Metrics

In general, the work prediction accuracy and loss value are evaluated. While your classifier is being tested on a dataset, you can keep track of its performance. In this example, True Positive denotes the number of correctly predicted positive events, True Negative the number of correctly predicted negative events, and False Positive is the number of incorrectly predicted negative events. False Negative denotes the number of correctly predicted negative events that actually occurred.Figure 5 Matrices of general confusion

Actual Class

Predicted_ Class		
	True_Positive	False_Positive
	False_Negative	True_Negative

Figure 5: Table for Confusion Matrix

This page contains information about the proposed classification model's performance metrics. Using equation 1, the total number of true predictions made by the three machine learning models is called accuracy in classification problems.

$$Accuracy = \frac{tp+tn}{tn+fp+tp+fn} \quad (1)$$

$$Recall = \frac{tp}{tp+fn} \quad (2)$$

$$FScore = 2\frac{Precision\ X\ Recall}{Precision+Recall} \quad (3)$$

$$Specification = \frac{tn}{tn+fp} \quad (4)$$

$$Precission = \frac{tp}{tp+fp} \quad (5)$$

5 Discussions

It is clear from Figure 6 that this system's Yolo v3 algorithm finds the "people" class from a list of different classes in the frames. A frame rate of 1.39 frames per second is used by the system. With the use of video and still photographs, we can test the system's capabilities. The distance between the participants in a frame taken in the video is used to identify each frame as either safe or unsafe. The overall number of safe and unsafe individuals is shown on a dashboard. You can get a statistical summary of a particular area of interest by looking at the bar chart on the dashboard. People in a safe region are represented by the blue colour, while those in an unsafe location are represented by the redcolour. At this time, the threshold value has been set at 7.These people are highlighted with RED bounding boxes that suggest they pose a threat because of their proximity and lack of respect for the distance rules. The colour BLUE denotes the rest of the group which is represented in the figure 7. The Penn-Fudan dataset gives better accuracy than other dataset for the proposed method.

Table 1: Performance metrics

Dataset	Scenario	Precision in %	Recall in %	F1-score in %	Accuracy in %
One Meter Distance	Outdoor	99.5	95.6	97.4	94.4
Two Meter Distance	Outdoor	99.4	94.3	96.8	93.6
Penn-Fudan	Outdoor	100	95.7	97.9	95.8

Figure 6: Performance metrics of three dataset

Input image

Output image with violations

Input image

Output image without violations

Figure 7: Output for Penn-Fudan Dataset with violations and without violations

6 Conclusion

The proposed method provides a straightforward approach to tracking violations of social distance, making it simple for users to grasp. Using the dashboard, you can see a visual representation of the ROI statistics and focus in on regions in which social distance is not maintained. Whenever a certain number of violations is exceeded in a specific area, the framework is utilized to deliver alarm signal to the proper authorities to alert them to the violations.A social separation identification tool has been developed by us through the use of a deep learning model. The utilisation of computer vision will result in a red frame and a red line being drawn around any two people who are found to be in violation of the rules. The proposed method was evaluated with the use of a video showing pedestrians walking down a street. As the findings of the visualisation show, the proposed method can be utilised in a variety of settings, including the workplace, restaurants, and schools, to ascertain the social distances that exist between groups of individuals. Increase the hardware's computing power, optimise the pedestrian detection techniques, integrate supplementary detection techniques such as helmet and mask identification and human body thermal detection, calibrate the camera's perspective view, etc.

Reference

1. W. H. Organization, “WHO corona-viruses (COVID-19),” https://www. who.int/emergencies/diseases/ novel-corona-virus-2019, 2020, [Online; accessed May 02, 2020].

2. WHO, "Who director-general's opening remarks at the media briefing on covid-19-11 march 2020. " https://www.who.int/dg/ speeches/detail/, 2020, [Online; accessed March 12, 2020].
3. L. Hensley, "Social distancing is out, physical distancing is in—here'show to do it," Global News–Canada (27 March 2020), 2020.
4. ECDPC, "Considerations relating to social distancing measures in response to COVID-19 – second update," https://www.ecdc.europa.eu/ en/publications-data/considerations, 2020, [Online; accessed March 23, 2020].
5. M. W. Fong, H. Gao, J. Y. Wong, J. Xiao, E. Y. Shiu, S. Ryu, and B. J. Cowling, "Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—social distancing measures," 2020.
6. F. Ahmed, N. Zviedrite, and A. Uzicanin, "Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review," BMC public health, vol. 18, no. 1, p. 518, 2018.
7. W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics–i. 1927." 1991.
8. C. Eksin, K. Paarporn, and J. S. Weitz, "Systematic biases in disease forecasting– the role of behavior change," Epidemics, vol. 27, pp. 96–105, 2019.
9. M. Zhao and H. Zhao, "Asymptotic behavior of global positive solution to a stochastic sir model incorporating media coverage," Advances in Difference Equations, vol. 2016, no. 1, pp. 1–17, 2016.
10. P. Alto, "Landing AI Named an April 2020 Cool Vendor in the Gartner Cool Vendors in AI Core Technologies," https://www.yahoo.com/ lifestyle/landing-ai-named-april-2020-152100532.html, 2020, [Online; accessed April 21, 2020].
11. L. AI, "Landing AI Named an April 2020 Cool Vendor in the Gartner Cool Vendors in AI Core Technologies," https://www.prnewswire.com/ news-releases/, 2020, [Online; accessed April 22, 2020].
12. B. News, "China coronavirus: Lockdown measures rise across Hubei province," https://www.bbc. co.uk/news/world-asia-china51217455, 2020, [Online; accessed January 23, 2020].
13. N. H. C. of the People's Republic of China, "Daily briefing on novel coronavirus cases in China," http:// en.n hc.gov.cn /2020-03/20/c 78006. htm, 2020, [Online; accessed March 20, 2020].
14. K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday et al., "The effect of control strategies to reduce social mixing on outcomes of the covid19 epidemic in wuhan, china: a modelling study," The Lancet Public Health, 2020.
15. C. Adolph, K. Amano, B. Bang-Jensen, N. Fullman, and J. Wilkerson, "Pandemic politics: Timing state-level social distancing responses to covid-19," medRxiv, 2020.
16. K. E. Ainslie, C. E. Walters, H. Fu, S. Bhatia, H. Wang, X. Xi, M. Baguelin, S. Bhatt, A. Boonyasiri, O. Boyd et al., "Evidence of initial success for china exiting covid-19 social distancing policy after achieving containment," Wellcome Open Research, vol. 5, no. 81, p. 81, 2020.
17. S. K. Sonbhadra, S. Agarwal, and P. Nagabhushan, "Target specific mining of covid-19 scholarly articles using one-class approach," 2020.
18. N. S. Punn and S. Agarwal, "Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks," 2020.
19. N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Covid-19 epidemic analysis using machine learning and deep learning algorithms," medRxiv, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/04/11/2020.04.08.20057679

20. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Thirty-first AAAI conference on artificial intelligence, 2017.
21. Google, “Open image dataset v6,” https://storage.googleapis.com/ openimages/web/index.htm l, 2020, [Online; accessed 25-February2020].
22. N. S. Punn and S. Agarwal, “Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images,” ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 16, no. 1, pp. 1–15, 2020.
23. A. Vaswani, S. Bengio, E. Brevdo, F. Chollet, A. N. Gomez, S. Gouws, L. Jones, Ł. Kaiser, N. Kalchbrenner, N. Parmar et al., “Tensor2tensor for neural machine translation,” arXiv preprint arXiv:1803.07416, 2018.
24. D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen et al., “Deep speech 2: End-to-end speech recognition in english and mandarin,” in International conference on machine learning, 2016, pp. 173–182.
25. A. Sonawane, “Medium : YOLOv3: A Huge Improvement,” https: //medium.com/@anandsonawane/yolo3, 2020, [Online; accessed Dec 6, 2019].
26. N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in 2017 IEEE international conference on image processing (ICIP). IEEE, 2017, pp. 3645–3649.
27. N. Wojke and A. Bewley, “Deep cosine metric learning for person re-identification,” in 2018 IEEE winter conference on applications of computer vision (WACV). IEEE, 2018, pp. 748–756.
28. S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. Iyengar, “A survey on deep learning: Algorithms, techniques, and applications,” ACM Computing Surveys (CSUR), vol. 51, no. 5, pp. 1–36, 2018.
29. A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey,” Neurocomputing, vol. 300, pp. 17–33, 2018.
30. N. S. Punn and S. Agarwal, “Crowd analysis for congestion control early warning system on foot over bridge,” in 2019 Twelfth International Conference on Contemporary Computing (IC3). IEEE, 2019, pp. 1–6.
31. Pias, “Object detection and distance measurement,” https://github.com/ paul-pias/Object-Detection-and-Distance-Measurement, 2020, [Online; accessed 01-March-2020].
32. J. Harvey, Adam. LaPlace. (2019) Megapixels: Origins, ethics, and privacy implications of publicly available face recognition image datasets. [Online]. Available: https://megapixels.cc