INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

COVID-19 Social Distancing Tracking and Monitoring System (SDMOS-19)

Nurafrina Arrysya Binti Abdullah [a], Nur Diyana Kamarudin [a,*], Siti Noormiza Makhtar [b], Ruzanna Mat Jusoh [a], Alde Alanda [c]

[a] Faculty of Science and Defense Technology, Cyber Security and Digital Industrial Revolution Centre, National Defence University Malaysia, Sungai Besi, Kuala Lumpur,57000, Malaysia

[b] Faculty of Engineering, Cyber Security and Digital Industrial Revolution Centre, National Defence University Malaysia, Sungai Besi, Kuala Lumpur, 57000, Malaysia

[c] Department of Information Technology, Politeknik Negeri Padang, Limau Manis, Padang, 25163, Indonesia

*Corresponding author: *nurdiyana@upnm.edu.my*

Abstract*—The Coronavirus disease (COVID-19), stemming from the SARS-CoV-2 virus, has garnered global concern as a virulent infectious ailment. Recognized as an epidemic by the World Health Organization (WHO), the persistently mutating virus sustains its transmission within communities. Individuals have been advised to uphold a safe interpersonal distance, notably around five feet, to mitigate its spread during social interactions. Addressing this imperative, an innovative automated social distancing detection system is conceived, leveraging the Convolutional Neural Network (CNN) algorithm. This system operates on two distinct input modes: static images and recorded videos recorded on closed-circuit television (CCTV). Remarkably, the proposed automated system adeptly quantifies and surveils the extent of social distancing among individuals in densely populated settings. A sophisticated framework accurately discerns social distancing compliance, delineating between hazardous and secure intervals via distinct red and green bounding box indicators. The culmination of this endeavor reveals an impressive 90% detection accuracy for both input modes. Notably, this proposed system holds substantial promise for implementation within sprawling premises such as expansive shopping malls or recreational parks. Seamlessly enforcing automated safety distance assessment expedites real-time insights to security departments and other relevant authorities. Consequently, the efficacy of citizens in upholding safe interpersonal distances can be promptly evaluated and, if necessary, corrective measures can be expeditiously instituted. This automated system ensures public health and safety maintenance, particularly in difficult circumstances.

***Keywords*— Convolutional neural network; aggregate channel feature; image and video processing; automatic social distancing.**

Manuscript received 11 Sep. 2022; revised 30 Dec. 2022; accepted 20 Apr. 2023. Date of publication 31 Mar. 2024.

I. INTRODUCTION

COVID-19 is a highly infectious disease caused by the coronavirus, which was declared an epidemic by the World Health Organization (WHO). The COVID-19 infection was first identified in late 2019 in Wuhan, China. Since August 24, 2021, there have been roughly 212 million recorded cases of COVID-19, along with around 4 million deaths, reported by World Health Organization (WHO) [1]. According to the Ministry of Health (MOH), there were 1.59 million reported cases of COVID-19 in Malaysia until August 2021, with about fourteen thousand lives lost. Coronavirus Disease 2019 (COVID-19) seems to be an infectious disease that causes respiratory infection comparable to the flu with symptoms such as a cough, fever, and, in extreme symptoms, breath difficulties. According to the World Health Organization, an individual might become infected with COVID-19 if he contacts other virus-infected individuals. Various vaccines are being created, but they are not entirely suitable for all age groups and are not widely distributed to fight this dangerous and deadly virus [2].

As a result, an alternative solution precaution is implemented to avoid the spread of this lethal virus. COVID-19 variants continue to spread in our environment. According to recent research, the Omicron variant is easily contagious, and its spread contributed to its high transmissibility compared to other variants, even in vaccinated individuals. As a result, people must keep their distance from one another. Throughout a COVID-19 endemic, individuals must maintain an appropriate distance of five feet between themselves. Safe distance could help reduce virus transmission. Keeping a safe distance in a large crowd, on the other hand, is challenging. People frequently forget to maintain a safe distance.

This enables the advancement of the COVID-19 social distancing tracking expert system. The system employs artificial intelligence to assess individuals' social spacing to determine whether the situation is secure or risky. In identifying the easiest way to keep social distancing, we realize that manual social distancing is quite challenging and complex to adopt in everyone's regular lifestyle, so we need a system to monitor and might even help detect human social distancing, which is to make sure that everyone maintains a social distancing of at least five feet from each other to minimize the number of deaths or long- term effects from the infection.

Aside from that, the sector under Critical National Information Infrastructure (CNII) must function regularly, even during the COVID-19 pandemic, and maintain social distancing at the workplace. CNII is a set of critical systems and functions essential to the nation's survival. For instance, the sectors included in the CNII involve National Defense and Security, Health Services, and others. As a result, social distancing has been one of the methods set by the World Health Organization (WHO) to combat the risk of COVID-19 virus infections [3].

Until now, there has been a lack of an accurate automatic monitoring and control system for human social distancing when going outside, which could be helpful as an assistive surveillance system for the Malaysian government to curb the spread of coronavirus, particularly in crowded places. The virus outbreak may not be eradicated as quickly as we would like; it will become endemic, and everyone must adapt to changes in norms, such as maintaining safe social distance between people to break the COVID-19 chain [4].

The research project uses the Real-Time Human Detection in Thermal Infrared Images task and the K-means clustering method to demonstrate and verify real-time human detection in thermal infrared images. Besides helping to improve a ting-yolov3 network, a new network architecture for sensing and tracking pedestrians from Thermal Infrared images was invented. Moreover, all pictures were clustered using the K-means clustering approach. The results show that these methods can achieve the same optimum performance and precision as Thermal Infrared (TIR) images. Meanwhile, the K-means clustering method equipment is increasingly popular in bounding boxes. Nevertheless, the results show a 4.88ms delay in recognition rates, and the YOLO algorithm cannot recognize objects precisely because each matrix can only recommend two bounding boxes without K-means clustering [5].

Hussein et al [6] created a real-time computer vision sensor to track human movement in uncontrolled moving camera systems. The system has two specific goals: responsiveness and effectiveness [4]. The validity and reliability of the structure have been increased by combining techniques for activity recognition, monitoring, and motion tracking into a single framework, with the result based on the integration of various algorithms. The use of a multi-threading design and library improved efficiency. This work utilizes a people detection technique, a machine learning algorithm, and a motion tracking method.

Furthermore, using a frequency signal, the tracking algorithm had previously been employed to detect the recognized object over time. The motion tracking methodology may employ the movement duration indication to evaluate the extent to which the identified and supervised object moves like a human. Eventually, the human detection algorithm can use the form as a prompt to determine if a particular image component appears to contain a human [4]. While the disadvantages mentioned are due to the enormous edge intensity, several erroneous warnings exist.

Previous researchers devised a technique that merged a potentially beneficial and widely used vector-form feature: histograms of oriented gradients (HOG) [7]. The HOG features focus on the contrast of silhouette contours against the background, as shown—the classifier's participation is the second phase in human detection. Two essential criteria for selecting classifiers are better generalization potential and minimal classification specificity. The two most widely used approaches that fulfill the requirements are the linear support vector machine (SVM) and AdaBoost. The HOG feature and the SVM classifier are highlighted, and efforts are being made to compute HOG features for human detection from video successfully. The bolded benefits include superior detection, low false positives, and the capability to reduce overall time utilized and achieve a high accuracy-to-time implementation ratio. Moreover, enhancing the HOG algorithm by adjusting the image with this method results in a two-fold increase in detecting humans in a 768*576 image while decreasing detection performance, mainly when individuals are on the edges [7]. The disadvantage of this method is that it slows down when large pictures are used.

Research teams in [8] present a visual identification system for human activity based on data gathered from a single camera. In the scenario of human activity in an open area, let's assume it is not packed with people. Object detection from the video is delivered initially, followed by object categorization and analysis. There seems to be an increasing participation in defining simple human activity, such as when somebody is jogging or moving, as implied by a combined application of several motions. The top phase involves realizing the actions of multiple individuals who are interacting. The results show that the image processing time is quick, with an average cycle duration of 61 milliseconds.

T. Toprak et al. investigate different AdaBoost algorithm adoption and implementation to solve the problem of real-time pedestrian detection in images [9]. They used gradient-based local features and a cascaded sensor to create a powerful classification model. The researchers compared the original AdaBoost algorithm to two other enhancing algorithms currently under development. The study proposes a quick, simple, efficient visual identification system for human activity. The disadvantage is that insufficient classification methods resulted in low margins and generalization errors.

Shalini *et al.*, [10], research teams proposed a two-step real-time pedestrian detection method. The first process appears to be the identification of HOG and the cascade frame classification algorithm. Although boosting appears to be the weakest classifier in cascade, it correlates directly to HOG block features. The method could significantly decrease the rate of false positives to different degrees. The disadvantage of the HOG feature is that it does not include the color and texture features, which will result in several false warnings.

Thorat uses MATLAB and Normalizes Cross-Correlation to monitor and detect motion [11]. Besides, the research utilizes the normalized cross-correlational research design for movement detection and element-linked assessment for shifting detection and tracking, and the suggested methodology is capable of real-time live video compression with good efficiency. This technique delivers a precise, reliable, sturdy, and immediately noticeable surveyance system. Users also stated explicitly that real-time consistency of video stream with good precision provides advantages for recognizing and tracking the object in frame sequence[11]. Its main drawback is that it performs poorly in the presence of noise, resulting in more erroneous warnings.

Previous research proposed a Radio Frequency (RF) measurement for human detection [12]. An experimental design with two different measurement methods, anechoic chamber and outdoor range measurements, was used as the primary method. Pizzillo spoke about the benefits and drawbacks of the various techniques. They claimed their approach could genuinely occupy a new human detection methods data system with information that personifies humans' radio frequency (RF) biometric signers in an anechoic chamber measurement method and can be used to detect and classify mountable and enemy threats. The disadvantages of using both measurement methods include the RF radiation that can be extremely dangerous to volunteers by heating the tissue of the eyes and body, and it is reported that electronic devices like cardiac pacemakers can be sensitive to RF interference.

O. Potkin focused on an enhanced frame differencing method for tracking an object and used MATLAB to enforce an automated system[13] . Whenever the proposed approach was assessed on multiple video sequence data, it was revealed that the objects in motion were identified with a low rate of errors because once compared to the previous frame differencing technique. The benefits of the suggested method have included a simple and easy noise removal process, as well as increased processing speed and quickness. The downside is that object identification, which causes rapid irradiation alteration, would then interrupt the tracking procedure [13].

According to [14], MATLAB created a method that focuses on an improved version of the Codebook algorithm for backdrop modeling and the straightforward edition of the Skeletonization algorithm for human tracking in a connected platform for real-time human motion detection. The pluses of this algorithm are that it detects very few false alarms compared to existing approaches while being more proficient than the existing research method. They specify that the equipment used in this method should be improved, and that vastly larger spaces are required for this algorithm to be faster and potentially more suitable for deployment [14].

CNN is particularly adapted to the use of pictures as data, as per [15]. The relevance of these methodologies is that the identifier produces better classification outcomes than the previous technique. CNNs have their effectiveness, and this research indicates that adding a linear layer to term classification techniques may boost performance [15]. The restriction of the above algorithm is that it lacks the characteristics of some more excellent transformer design features.

The approach by [16] is classified into four techniques: feature pyramid network, multi-scale feature fusion, double-branch structure, and network architecture. The first approach is a pyramid network, which seems to have more decadent interpretations at all stages and is formed rapidly from a single input scale without representational power, performance, or capacity. The second is multi-scale feature fusion, suitable for overcoming and over-enriching high-frequency information. Additionally, they stated that a double-branch structure could increase and optimize the algorithm's effectiveness. The alternate approach includes network architecture, which could also capture the image's features and build a depth map [16]. The limitation caused by multi-scale feature fusion shows that there is still a high risk of data loss throughout this approach, indicating that this method requires regular maintenance. In addition, it is prolonged while processing data when pictures are mixed.

Several methods and techniques were used, according to [17], including sequential creation by vector quantization, improved hidden Markov model (M-HMM), robust method, and depth image processing using the least square method. Furthermore, there are two popular methods for extracting features: feature extraction using depth shape features and feature extraction using joint information elements. The advantages of the methodologies are that they can distinguish various activities across the system and provide high-quality photos and results. As a result, detailed shapes outlines, and body joint information can be used to identify, trace, and classify behavior patterns. On the contrary, I agree that flaws may result in a lower detection percentage, mainly when combined information is absent and human silhouettes and subjects are viewed from a distant place [17].

An improved YOLOV3 model with a Deep CNN was implemented in [18]. The model can detect practically anything that needs to be detected and has great exactness in its mechanism. Furthermore, it was demonstrated that, although being considerably faster than the original networks, this approach maintains its consistency. They point out that this technology cannot detect something that is barely distinguishable even to the naked eye[18].

II. Materials and Method

A. Data Collection

An electronic camera records still photos and recorded videos of 100 pieces of data. This is the original footage from the owner's single click, and it will serve as the raw research that is clearly relevant to the original study goal. As controlled photos, primary images containing information about the measured distance among people are utilized to justify the automated system.

Raw images and videos will also be compiled using appropriate browser extensions, such as Microsoft Bing and Google Chrome, and the data has been renamed secondary. Microsoft Bing and Google Chrome are excellent internet sites for gaining access to still images and video sources, including such data from organizations' database systems, webpages, newspaper articles, or other source materials. JPEG (.jpg,.jpeg), TIFF (.tif,.tiff), AVI (Audio Video Interleave), and MOV (QuickTime Movie) formats are utilized to save the data.

B. Data Pre-Processing

Data pre-processing determines which data objects and features will be employed in the next phase of object recognition and data classification[3]. Image pre-processing could be described as the method of modifying images to enhance images in two different ways: data cleaning and image enhancement, as in Fig 1. Data cleaning has been employed during pre-processing to classify unimportant and missing items. Data cleaning will be conducted to address lacking and imprecise in still images and movies[19]. Noisy data is inconsistent data that is complicated to evaluate because of ineffective data collection methods and unreliable data entry.

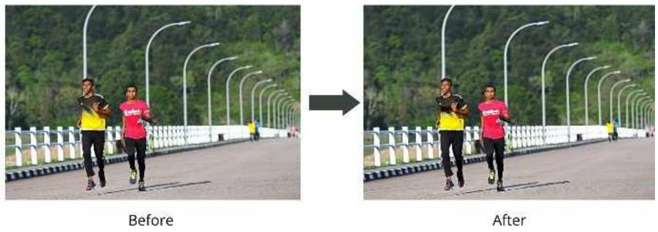

Fig. 1 Sharper image after enhancement.

Remini is one of the most popular image enhancers that can improve still images and videos. Because Remini appears to be an automated image processing tool with 26 tools, users do not need to be concerned about their preferred level while using it. Remini gives pixelated, damaged, low-resolution photographs and movies a fresh lease on life. Remini astonishes users with the clarity and sharpness of the pictures that appear in high definition, as shown in Figure 1. Because Remini uses cutting-edge AI, it could also unblur, restore, and optimize pictures and videos

C. Data Analysis and Classification

The clean and enhanced images obtained from the pre-processing phase will be analyzed and classified in this phase. Data will be examined, and classification will be implemented to detect social distancing using a few approaches from the CNN algorithm on the MATLAB platform. A small percentage of errors might still occur at this stage, which could cause false alarms in social distance detection. CNN technique, also known as ConvNets, uses a deep learning tool [20]. The usage of CNNs is due to their effectiveness in pattern recognition, especially in recognizing objects, people's faces, and scene images[21]. Moreover, we also combined CNN with Remini image enhancer to boost the accuracy of social distancing detectors, and it has been observed that the accuracy has improved to 90%.

A pre-trained model based on the Aggregate Channel Feature (ACF) classification model is employed for the analysis and classification stage. The recognition method relies on CNN and a pre-trained model that uses the ACF classification model 'inria-100x41' to detect individuals. The INRIA Person set of data is also utilized to coach the 'inria-100x41' conceptual framework. This process optimizes people in images and saves the results as bounding boxes and a scoring system.

A CNN is an artificial neural network explicitly designed to process pixels for image and video recognition [22][23]. CNNs are powered by AI techniques implemented in image and video processing systems to operate conceptual and insightful duties[24]. These learning environments routinely use computer vision algorithms to recognize images and videos, as well as classification techniques like the ACF system, which uses a pre-trained model to identify humans in photos and videos[25][26].

CNN began with the convolution layers, the very first layer of a convolutional network [27]. The convolutional layers in this research project are guided by an added convolution layer, also renowned as max pooling, and the top level of this convolutional layer is a completely connected layer. CNN will become more detailed as it identifies more significant quantities of the videos and pictures used in this task[28]. On the initial layer, there is a growing interest in developing simple features such as color schemes and sides. Besides, as the image data evolves through the CNN layers, it learns to recognize and define the various components or forms of the item, which in this proposal will be to discover, until it eventually detects its planned object, which in the present proposal is to recognize.

TABLE I
COMPARATIVE STUDY OF EXISTING METHODS RELATED TO THIS STUDY.

No.	Title	Method	Advantages
1.	Real-Time Human Detection in Thermal Infrared Images	- K-Mean clustering - YOLO algorithm	- Obtain the best high frequency - Excellent priority in bounding boxes
2.	Radio Frequency RF Measurements for Human Detection	- anechoic chamber measurement - Outdoor Range Measurement	- To detect and classify mounted and dismounted threats
3.	Real-Time Human Detection, Tracking and Verification in Uncontrolled Camera Motion Environments	- Human detection algorithm - Object tracking algorithm - Motion analysis algorithm	- Robustness and efficiency
4.	Real-Time Human Detection in Video Streams	- HOG algorithm - SVM algorithm	- Superior detection - Low false positives - Reduce the total time used
5.	Real-Time Human Detection in Urban Scenes: Local Descriptors and Classifiers Selection with AdaBoost-like Algorithms	- AdaBoost-like algorithms	- Easy and simple to program - Flexible to combine with any machine learning algorithm
6.	Real-Time Human Detection based on Cascade Frame	- AdaBoost algorithm - HOG algorithm	- Obtain the best high frequency - Excellent priority in bounding boxes
7.	Real-Time Human Activity Recognition	Moving object classification	- Good processing time for image
8.	Detection and tracking of moving object	Normalized Cross-Correlation	- High accuracy - Rapid, visible surveyance system - Have high reliability to detect

No.	Title	Method	Advantages
9.	Video-based Detection, Tracking, and Classification of Vehicles	- Background Subtraction Algorithm	- Superior detection - Low false positives - Reduce the total time used
10.	Real-Time Human Detection and Tracking in Infrared Video Feed	- Deep Convolutional Neural Network	- Reduce noise with noise cancelation filters - Enhance the detection accuracy
11.	Depth Images based Human Detection, Tracking and Activity Recognition Using Spatiotemporal Features and Modified HMM	- Robust Method - Depth Image Analysis (apply least Square Method) - Feature Extraction using depth shape features - Sequence generation using vector quantization - Modified Hidden Markov Model(M-HMM)	- Can recognize different activities - Can detect, track and recognize activities using depth silhouettes and body joint information - Access high-quality images and overcome
12.	Embedded system for real-time human motion detection	- A modified version of the Codebook algorithm for background modeling - Simplified version of Skeletonization algorithm for human detection by MATLAB	- More accurate detection than the original one - Reduce the false alarm detection compared to the original algorithm - Clean picture from the noise that stuck in
13.	Hand Gestures Detection, Tracking, and Classification Using Convolutional Neural Network	- Convolutional Neural Network	- The classifier demonstrates the accuracy of classification - More accuracy than the previous method
14.	Ball and Player Detection and Tracking in Soccer Videos Using Improved YOLOV3 Model	Improved YOLOV3 Model with a deep convolutional neural network	- Able to detect almost everything that need to be detected - High precise and accuracy It is much faster than other methods and still maintains accuracy

The fundamental basic structure of CNN is the convolutional layer, in which most of the estimation in this journal takes place. It necessitates a few parts, including entering data and filtration. Like the previous venture, it utilized a color photo of a 3D array of elements. Its image does have three aspects: size, width, and intensity, which correlate to RGB in an image. The sensor characteristic, also known as kernel or filter, will keep moving across areas of the image and verify if the characteristic required is prevalent to be employed in the construction process; the procedures we have been through are known as convolution. Besides, the detector has a two-dimensional, or 2D, range of weights that reflect a segment of the image. Furthermore, during the convolution layers, its mass in the feature descriptor will stay constant as it moves across the picture, a process known as parameter sharing.

A CNN appears to apply a Rectified Linear Unit (ReLU) transformation towards the feature map after each convolution operation, which also introduces nonlinear behavior into the model[29][30]. The first convolution layer is followed by the second. The CNN architecture will become more hierarchical even as surface has access to the pixel value of the pictures within the visual field of previous layer upon layer. For the hierarchical sequence in the neural net, the integration of its sections will portray and show the relatively high structure, resulting in a functional power structure within the CNN.

Further to that, the following layer in the CNN utilized for this task is the pooling layer, also identified as bottom testing, which is employed to decrease the number of variables in the insight. Like the convolution layers, its pooling operation was applied to eliminate filtration from the whole insight. Here on task, we were using max pooling, which further means the filter would start moving throughout the insight and choose the pixel with the highest production. Furthermore, the max pooling on CNN provides advantages by lowering the complexity of videos and images utilized, increasing productivity, and limiting the danger of generalization on our videos and images.

The final layer employs a fully connected layer, whereas in slightly connected layers, the image pixels of the input are not directly connected to the output nodes. Throughout the fully connected layer, each entity as in output nodes would automatically connect to a node in the earlier one. This layer applies characterization, which classifies the dataset based on the attributes derived from the preceding layer and the various filtrations used to classify it appropriately. Figure 2 illustrates the architectural style of the CNN layer, or Convolutional Neural Networks surface.

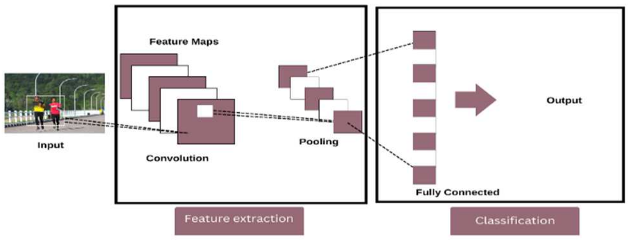

Fig. 2 Architecture of Convolutional Neural Network

D. Data Validation

The accuracy and quality of the processed image will be tested throughout data validation, which also involves ensuring a data source both before and after transporting and handling it, in addition to analyzing the system's result. Data validation will determine if the proposed system for social distancing has the best accuracy and appears to work well for detecting human social distancing [12].

The still image in Figure 3 is an instance of distance verification utilizing photos captured with a smartphone or digital camera that adhere to the social distancing guidance. The picture depicts a two-person social separation distance following the one-meter social distancing principle. As asserted, the bounding boxes for both individuals in the picture seem to have been green, proving stable for the social distancing between the two individuals.

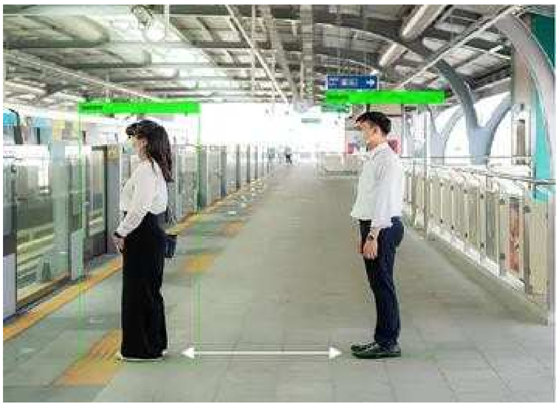

Fig. 3 The result of an image used for data validation

E. Pseudocodes

Covid-19 Social Distancing Detector (Pseudocode for image)

```
If the detector uses people detector ACF (Inria- 100x41)
 is being called
  It will detect image insert and run the
    code Show result and detect image in
    a box.
      Show size of
image END IF
  If the image chosen to detect a human figure
    Show the bounding boxes and detection result for the
      image Show red bounding boxes and declare them
      as danger
IF ELSE
  IF safe social distancing
    Show green bounding boxes and declare
them safe
   ELSE
    Stop the system to run the image
       Restart with a new image
ENDIF
```

Covid-19 Social Distancing Detector (Pseudocode for CCTV Video

```
If the detector using video file reader ACF (Inria-
100x41) is being called
  It will detect recorded videos, insert and run the
    code
   Show results and detect recorded videos in a
    box
  Show the size of recorded
videos
  END IF
    If the image was chosen to detect human movement and
      figure
      Show the bounding boxes and detection results for
        recorded videos
      Show red bounding boxes and declare them as
 danger
  IF ELSE
    IF safe social distancing
      Show green bounding boxes and declare them
safe
  ELSE
    Stop the system to run the recorded
      videos
    Restart with newly recorded videos
 ENDIF
```

The first pseudocode describes automatically generated scripting for still pictures to identify and continuously monitor regardless of whether folks are adhering to social separation or breaking the WHO rules. Suppose indeed the boundary boxes in the output image are red. In that case, it indicates that now the individual did not comply with the social separation and disobeyed policies or guidelines provided. If the rectangles are green, the person adheres to the social distancing guidance. The People sensor ACF applies the Aggregate Channel Features 'inria-100x41' to diagnose humans in images. Additionally, peopleDetectorACF reverts an ACF-pre trained erect individuals' detector. The detector is an acfObjectDetector object practiced upon that INRIA human set of data [22].

The next coding is for CCTV video recordings. This section requires perspective. VideoFileReader restores an ACF-trained upright people sensor. The analyzer is an acfObjectDetector entity that was learned by utilizing an INRIA large dataset. As shown in the image above, a pre-trained model is used to identify humans and determine regardless of whether they adhere to social distancing guidelines or not. If the boundary boxes in recorded footage turn red, it indicates that the participant does not comply with the regulations of separate channels but instead disobeys the instructions or guidelines granted. If the boundary box coloration is green, it indicates that the people follow the social distancing recommendation.

III. Results and Discussion

Secondary data is shown in Figures 4, 5, and 6. The images represent various locations, each in a recreational park at a different period. Figure 4 depicts two persons jogging in the park who are socially separated. As a result, the meters measured follow the World Health Organization's social distancing regulations, and the bounding boxes that detect the human are shown as green, indicating that the distance between two individuals is secure.

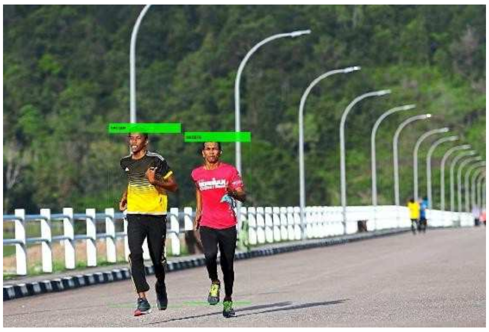

Fig. 4 Image detection result at recreational park

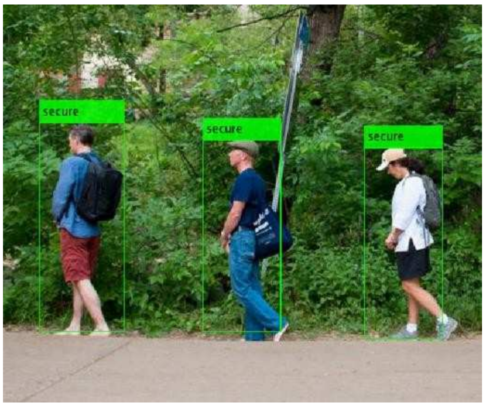

Fig. 5 Image detection result at recreational park

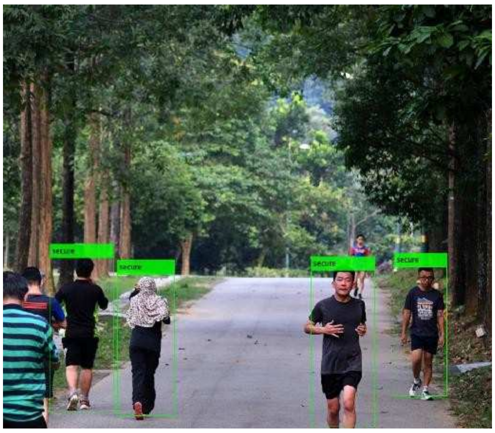

Fig. 6 Image detection result at recreational park

Figure 5 represents a one-meter social distance between three people in the image. The bounding boxes are green to indicate that there is appropriate social separation between the individuals. It can also be compared to the image used to validate whether individuals are adhering to the rules regarding social distancing, which indicates that guidelines for social distancing are one meter apart. Figure 6 illustrates a group of people jogging and walking in a recreational park. The system assesses safe social separation by displaying green boxes that indicate safety.

Figures 7 to 9 show images captured with a smartphone and a camera, referred to as primary data. The source data from the creator and personal data are examples of primary data. The data is collected from crowded places such as malls.

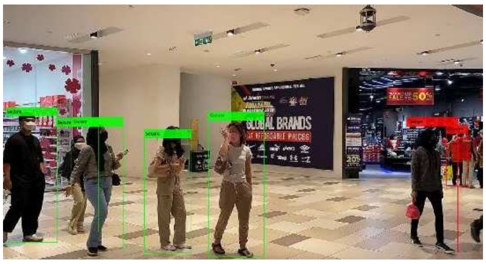

Fig. 7 Image detection result at a shopping mall.

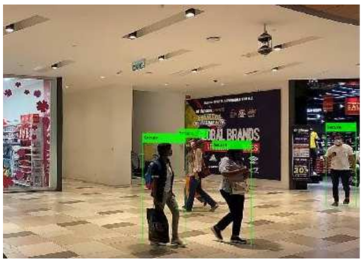

Fig. 8 Image detection result at a shopping mall.

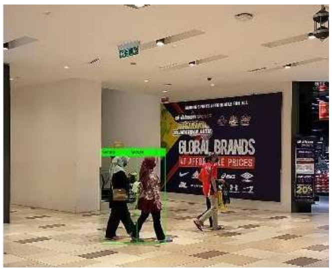

Fig. 9 Image detection result at a shopping mall.

Figures 10 to 13 were generated from recorded videos acquired with a camera and categorized as primary data. The video was recorded at the Kuala Lumpur LRT station to test the system's ability to detect and identify humans in video and human social separation in video. As a result, the percentage of accuracy in the video can increase from 80 percent to 87 percent because the system mistakenly detected several individuals in the video. Furthermore, compared to the still image used to validate, the proportion of detected is quite good for a video to identify human social distancing.

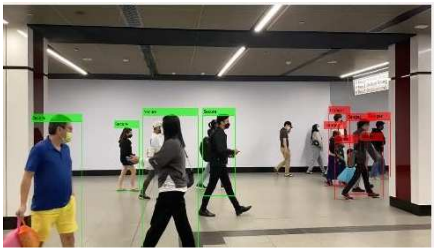

Fig. 10 Video detection result at an LRT station in Kuala Lumpur.

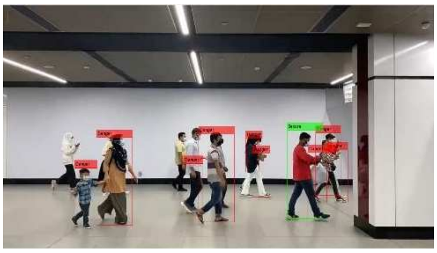

Fig. 11 Video detection result at an LRT station in Kuala Lumpur.

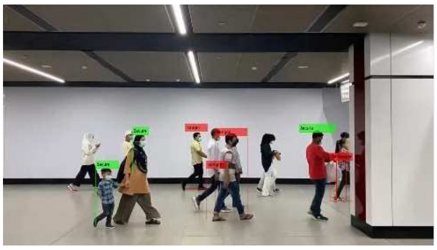

Fig. 12 Video detection result at an LRT station in Kuala Lumpur.

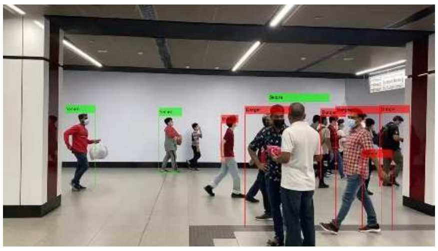

Fig. 13 Video detection result at an LRT station in Kuala Lumpur.

In a crowded space, the proposed system will indicate and identify human social distance in full compliance with WHO regulations and categorize it as risky or protected using a red box and green box clear indication, as shown in the results below. According to the results, the detection rate of this method has been estimated to be as high as 90%. The focus of this research, which uses a CNN with a pre-trained model employing the ACF classifier model stipulated as 'inria-100x41,' could be used in massive community arenas such as a recreation center or possibly even a business district because it offers a surveillance system all while providing real-time information to governmental regulatory agencies as to whether the public are complying by WHO social distancing rules.

TABLE II
ACCURACY COMPARISONS BETWEEN IMAGES AND VIDEOS.

Images	- The percentage of image detection accuracy is 90%. - The algorithm successfully identifies most of the images. - When compared to the still image that was used to validate, the result shows that the percentage of accuracy is approximately 90% since it does present the correct bounding boxes if the individual is following the social distancing meters, which is one meter.
Videos	- The accuracy of the collected video is between 80 and 87 percent. - It can indicate that the recorded video is realistic 80 percent to 85 percent of the time since some videos are not too precise and accurately located when the data for recognition is loaded. - Some of the results seem to be inaccurate and unpredictable, bounding boxes because they have their messy data.

CNN is a powerful image and video processing algorithm, as described in the methodology. These algorithms are currently the most effective in automated image and video processing. Many users use these algorithms to detect people or elements in pictures and videos [3]. Additionally, images and videos possess RGB values. MATLAB is required to extract an image or video from a file. Color information is kept in three-dimensional arrays. The first two dimensions are the image's height and width, also called the pixel count. The final element represents the red, green, and blue color schemes in every image. When a human figure is detected in pictures or videos, the RGB coding specifies green and red as the appearance of bounding boxes.

According to the findings, the image has a greater average accuracy than the video. It was demonstrated that the bounding box in the picture was more precise than the recorded video on detecting persons. As previously stated, the correct way to detect people is when the bounding boxes identify the proper social distancing. A bounding box that appears green and proves safe is for people who are following the social distancing regulations. In contrast, the bounding box that detects persons who refuse to follow the rules will appear red as well as display danger. As stated, photos are more accurate at detecting persons. However, videos have such a lower average accuracy because multiple people in the video are detected with faulty bounding boxes.

IV. Conclusion

This framework could be used to assist the government in controlling the virus from spreading further. The experiment illustrates that the system is functional and capable of detecting human separation and classifying it as secure or unsafe. The present scheme is approximated to achieve an accuracy rate of around 90%, thanks to the inclusion of CNN in this study. Since it was a pre-trained human detection method, it employed a pre-trained model leveraging the ACF classifier model or recognized as Aggregate Channel Feature (ACF) that was described as 'inria-100x41' and can be rated as the best alternative for recognizing humans. Moreover, this system is expected to encourage new social norms and aid authorities in reducing the number of COVID-19 infections caused.

The best recommendation for further research is to enhance and improve the system by adding a scoring number, as in the bounding box. The security teams that control the system in their zones or locations, including shopping centers and recreation facilities, could obtain an exact and trustworthy count of how many people have passed through the area.

Another future work is to provide a real-time camera embedded with CNN algorithm for the target market consumer, including a government organization, a shopping complex manager, or even a client who requires it on their area to make monitoring the social distance between each other much more accessible. Moreover, the system and sensor may additionally feature an alert or siren that rings if the individual disobeys the social distancing guidelines. This serves as a reminder to the person.

In summary, additional recommendations can improve the system's accuracy and precision in detecting human movement and figures. Based on the results acquired during the research, the existing arrangement developed effectiveness is 80-90 percent. More suggestions can help the algorithm perform at a higher level. It may be more beneficial to ensure that everyone follows the social distancing rules as a strategy to avoid and reduce the number of COVID-19 cases.

ACKNOWLEDGMENTS

The authors thank National Defense University of Malaysia for the publication fees. We are also grateful for the comments from individuals who were involved in this paper publication.

REFERENCES

[1] E. Breck, N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "Data Validation for Machine Learning," *SysML*, 2019.

[2] Y. Fan, Y. Luo, and X. Chen, "Research on Face Recognition Technology Based on Improved YOLO Deep Convolution Neural Network," Journal of Physics: Conference Series, vol. 1982, no. 1, p. 012010, Jul. 2021, doi: 10.1088/1742-6596/1982/1/012010.

[3] S. Kamal, A. Jalal, and D. Kim, "Depth Images-based Human Detection, Tracking and Activity Recognition Using Spatiotemporal Features and Modified HMM," Journal of Electrical Engineering and Technology, vol. 11, no. 6, pp. 1857–1862, Nov. 2016, doi:10.5370/jeet.2016.11.6.1857.

[4] B. A. Plummer, M. Brown, and S. Lazebnik, "Enhancing Video Summarization via Vision-Language Embedding," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, doi: 10.1109/cvpr.2017.118.

[5] X. He and Y. Peng, "Fine-Grained Image Classification via Combining Vision and Language," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, doi:10.1109/cvpr.2017.775.

[6] M. Hussein, W. Abd-Almageed, Yang Ran, and L. Davis, "Real-Time Human Detection, Tracking, and Verification in Uncontrolled Camera Motion Environments," Fourth IEEE International Conference on Computer Vision Systems (ICVS'06), 2006, doi:10.1109/icvs.2006.52.

[7] Y. Ioannou, D. Robertson, R. Cipolla, and A. Criminisi, "Deep Roots: Improving CNN Efficiency with Hierarchical Filter Groups," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, doi: 10.1109/cvpr.2017.633.

[8] J. Zhuang, J. Yang, L. Gu, and N. Dvornek, "ShelfNet for Fast Semantic Segmentation," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Oct. 2019, doi:10.1109/iccvw.2019.00113.

[9] T. Toprak, B. Belenlioglu, S. Dogan, B. Aydın, and M. A. Selver, "On Diversity and Complementarity of Pedestrian Detection Models," Journal of Physics: Conference Series, vol. 1141, p. 012152, Dec. 2018, doi: 10.1088/1742-6596/1141/1/012152.

[10] G. V. Shalini, M. K. Margret, M. J. S. Niraimathi, and S. Subashree, "Social Distancing Analyzer Using Computer Vision and Deep Learning," Journal of Physics: Conference Series, vol. 1916, no. 1, p. 012039, May 2021, doi: 10.1088/1742-6596/1916/1/012039.

[11] R. Keniya and N. Mehendale, "Real-Time Social Distancing Detector Using Socialdistancingnet-19 Deep Learning Network," SSRN Electronic Journal, 2020, doi: 10.2139/ssrn.3669311.

[12] T. Pizzillo, "Radio Frequency (RF) Measurements for Human Detection, Tracking, and Identification," 2007. [Online]. Available: https://www.researchgate.net/publication/228601002

[13] O. Potkin and A. Philippovich, "Hand Gestures Detection, Tracking and Classification Using Convolutional Neural Network," Analysis of Images, Social Networks and Texts, pp. 263–269, 2020, doi:10.1007/978-3-030-39575-9_27.

[14] J. Diers and C. Pigorsch, "Out-of-Distribution Detection Using Outlier Detection Methods," Lecture Notes in Computer Science, pp. 15–26, 2022, doi: 10.1007/978-3-031-06433-3_2.

[15] H. Fernando, I. Perera, and C. de Silva, "Real-time Human Detection and Tracking in Infrared Video Feed," 2019 Moratuwa Engineering Research Conference (MERCon), Jul. 2019, doi:10.1109/mercon.2019.8818862.

[16] B. T. Naik and M. F. Hashmi, "Ball and Player Detection &amp;amp; Tracking in Soccer Videos Using Improved YOLOV3 Model," Jun. 2021, doi: 10.21203/rs.3.rs-438886/v1.

[17] M. Sharma, "A Review : Image Fusion Techniques and Applications," *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, 2016.

[18] Y. Niu and Z. Meng, "Research on object detection technology for human detection," Journal of Physics: Conference Series, vol. 1544, no. 1, p. 012076, May 2020, doi: 10.1088/1742-6596/1544/1/012076.

[19] F. A. A. Naqiyuddin, W. Mansor, N. M. Sallehuddin, M. N. S. Mohd Johari, M. A. S. Shazlan, and A. N. Bakar, "Wearable Social Distancing Detection System," 2020 IEEE International RF and Microwave Conference (RFM), Dec. 2020, doi:10.1109/rfm50841.2020.9344786.

[20] J. Begard, N. Allezard, and P. Sayd, "Real-time human detection in urban scenes: Local descriptors and classifiers selection with AdaBoost-like algorithms," 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Jun. 2008, doi: 10.1109/cvprw.2008.4563061.

[21] H. Darwis, Z. Ali, Y. Salim, and P. L. L. Belluano, "Max Feature Map CNN with Support Vector Guided Softmax for Face Recognition," JOIV : International Journal on Informatics Visualization, vol. 7, no. 3, pp. 959–966, Sep. 2023, doi: 10.30630/joiv.7.3.1751.

[22] M. Thangaraj and S. Monikavasagom, "A Competent Frame Work for Efficient Object Detection, Tracking and Classification," Wireless Personal Communications, vol. 107, no. 2, pp. 939–957, Apr. 2019, doi: 10.1007/s11277-019-06310-4.

[23] J.-R. Lee, K.-W. Ng, and Y.-J. Yoong, "Face and Facial Expressions Recognition System for Blind People Using ResNet50 Architecture and CNN," Journal of Informatics and Web Engineering, vol. 2, no. 2, pp. 284–298, Sep. 2023, doi: 10.33093/jiwe.2023.2.2.20.

[24] S. K. Mohamed, N. A. Sakr, and N. A. Hikal, "A Review of Breast Cancer Classification and Detection Techniques," International Journal of Advanced Science Computing and Engineering, vol. 3, no. 3, pp. 128–139, Oct. 2021, doi: 10.30630/ijasce.3.3.55.

[25] A. K. Ali, A. M. Abdullah, and S. F. Raheem, "Impact the Classes' number on the convolutional neural networks performance for image classification", Int. J. of Adv. Sci. Comp. and Eng., vol. 5, no. 2, pp. 119–128, Aug. 2023.

[26] Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, "Emotion Recognition by Facial Expression and Voice: Review and Analysis," Journal of Informatics and Web Engineering, vol. 1, no. 2, pp. 45–54, Sep. 2022, doi: 10.33093/jiwe.2022.1.2.4.

[27] J. Bai, S. Li, L. Huang, and H. Chen, "Robust Detection and Tracking Method for Moving Object Based on Radar and Camera Data Fusion," IEEE Sensors Journal, vol. 21, no. 9, pp. 10761–10774, May 2021, doi: 10.1109/jsen.2021.3049449.

[28] C. C. Chai, W. H. Khoh, Y. H. Pang, and H. Y. Yap, "A Lung Cancer Detection with Pre-Trained CNN Models," Journal of Informatics and Web Engineering, vol. 3, no. 1, pp. 41–54, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.3.

[29] Md. M. Islam, Md. R. Islam, and Md. S. Islam, "An Efficient Human Computer Interaction through Hand Gesture Using Deep Convolutional Neural Network," SN Computer Science, vol. 1, no. 4, Jun. 2020, doi: 10.1007/s42979-020-00223-x.

[30] D. Meidelfi, .-. Hendrick, .-. Yulherniwati, .-. Novi, and A. F. Zulfitri, "Implementation of Convolutional Neural Network and Vincenty Formula on Face Attendance System Web-Based for Managing the Attendance", Int. J. of Adv. Sci. Comp. and Eng., vol. 5, no. 3, pp. 287–297, Dec. 2023.