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

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

[a] *Faculty of Science and Defence Technology, National Defence University Malaysia, Kem Sungai Besi, Kuala Lumpur,57000, Malaysia*

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

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

Corresponding author: nurdiyana@upnm.edu.my, noormiza@upnm.edu.my

Abstract — **Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. According to the Ministry of Health (MOH), by August 21, Malaysia has recorded 1.59 million COVID-19 cases and about 14,000 deaths. COVID-19 is an infectious disease that causes respiratory infections similar to the flu, and those infected exhibited symptoms such as a cough, fever, and, in extreme cases, breathing difficulties. The World Health Organization (WHO) declared COVID-19 an epidemic. Unfortunately, the virus mutates and continues spreading in the surroundings. According to recent research, the Omicron variant is highly transmissible and spreads more easily than the other variants, even among vaccinated individuals. As a precaution, individuals are advised to maintain a safe distance of five feet from one another during a social meeting. This study develops an automated social distancing detector system using the Convolutional Neural Network (CNN) algorithm using still images and recorded Closed-Circuit Television (CCTV) videos as the inputs. The system automatically measures and monitors the social distance between people in a crowded environment. The system detects human social distancing accurately and categorizes the distance between people as dangerous or safe using red and green bounding box indicators. The results show that the system has a 90% detection accuracy. The proposed automated social distancing detector system has promising potential for implementation in large premises such as shopping malls or recreational parks because it offers instant notification to the security department or other enforcement agencies whether the public adheres to the safe social distance requirement.**

Keywords — **Convolutional Neural Network; Aggregate Channel Feature; Image and Video Processing, Automatic Social Distancing Detector.**

Manuscript received XXXX; revised XXXX; accepted XXXX. Date of publication XXXX.

I. INTRODUCTION

COVID-19 is a highly infectious disease caused by the coronavirus and was declared an epidemic by the World Health Organization (WHO). COVID-19 infection was first identified in late 2019 in Wuhan, China. As of August 24, 2021, about 212 million COVID-19 cases and 4 million deaths have been reported to the World Health Organization (WHO) [1]. According to the Ministry of Health (MOH), by August 21, Malaysia has recorded 1.59 million COVID-19 cases and about 14,000 deaths. COVID-19 is an infectious disease that causes respiratory infections similar to the flu, and those infected exhibited symptoms such as a cough, fever, and, in extreme cases, breathing difficulties. According to the World Health Organization, a person can contract COVID-19 from direct contact with an infected individual. Even though several vaccines are currently available, they are not suitable for all age groups and are not widely distributed to fight this dangerous and deadly virus [2]. Because of this, it is essential to implement an alternative precautionary measure to prevent thespread of this lethal virus.

COVID-19 variants continue to spread in the environment. According to recent research, the Omicron variant is highly transmissible and spreads more easily than the other variants, even among vaccinated individuals. Therefore, people must keep their distance from one another. Throughout the COVID-19 endemic, individuals must maintain an appropriate five feet distance from one another since this safe distance could help reduce virus transmission. However, it is not easy to maintain a safe social distance in a large crowd, and people often forget to maintain a safe distance.

The researchers develop the COVID-19 social distancing tracking expert system to deal with these problems. The system employs artificial intelligence to assess the social spacing between individuals to determine whether the situation is safe or risky.

II. PROBLEM STATEMENT

In identifying the easiest way to maintain social distancing, we realize that it is challenging to practice manual social distancing in everyday life. It is essential to develop a system to monitor and detect human social distancing to ensure that everyone maintains a minimum of five feet of social distancing to minimize the number of deaths or long-term effects from the infection.

The sector under the Critical National Information Infrastructure (CNII) must function as usual during the COVID-19 pandemic, and workers must maintain their social distance at the workplace. CNII is a set of critical systems and functions essential to the nation'ssurvival. Two sectors making up CNII arenational defense and security and health services. The World Health Organization (WHO) proposed social distancing to reducethe risk of contracting COVID-19 infections [28].

Presently, there is no accurate automatic system to monitor and control social distancing in the public space, which is a beneficial assistive surveillance system for the Malaysian government to curb the transmission of COVID-19, particularly in crowded places. We may not be able to eradicate COVID-19 any time soon, and it may become endemic. Therefore, everyone must adapt to the new normal, including maintaining safe social distancing to break the COVID-19 transmission chain [6].

III. BACKGROUND OF THE STUDY

This research used 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 [5]. Besides improving the ting-yolov3 network, the researchers developed a new network architecture for sensing and tracking pedestrians from thermal infrared images. We clustered the pictures using the K-means clustering method. The results showed that these methods could achieve the same optimum performance and precision as thermal infrared (TIR) images [5]. The popularity of the K-means clustering method equipment is increasing in bounding boxes. Nevertheless, the results show a 4.88ms delay in the recognition rate, and the YOLO algorithm could not recognize objects precisely because each matrix can recommend only two bounding boxes without the K-means clustering [5].

Hussein *et al.* created a real-time computer vision sensor to track human movement in uncontrolled moving camera systems. The system was designed with two specific goals, responsiveness, and effectiveness [6]. They increased the validity and reliability of the structure by combining the techniques for activity recognition, monitoring, and motion tracking into a single framework, and the results were based on the integration of various algorithms. They increased the efficiency by using a multi-threading design and library. This work utilized people detection techniques, a machine learning algorithm, and a motion tracking method.

Previously, the tracking algorithm used frequency signals to detect the recognized object over time. The motion tracking method may use the movement duration indication to evaluate the extent of the identified and supervised object movements like a human. The human detection algorithm can use the form as a prompt to determine if a particular component of an image appears to contain a human [6]. Although these disadvantages are due to the high edge intensity, there are several erroneous warnings.

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 focused on the contrast of silhouette contours against the background. The classifier's participation was the second phase in human detection. The two crucial considerations in selecting classifiers are better generalization potential and minimal classification specificity. The two most frequently used methods which fulfill the requirements are the linear Support Vector Machine (SVM) and AdaBoost. The HOG feature and the SVM classifier are highlighted, and their own efforts are in successfully computing HOG features for human detection from video. The benefits are better detection, a low rate of false positives, reduced overall time, and a high accuracy-to-time implementation ratio. Moreover, enhancing the HOG algorithm by using this method to adjust the image resulted in a two-fold increase in detecting humans in a 768*576 image while decreasing detection performance, especially when the individuals were at the edges [7]. The disadvantage of this method is that it slows down when using large pictures.

The research teams in [8] presented a visual identification system for human activities using the data gathered using a single camera. The scenario was human activities in an open area not packed with people. The first step is object detection from video, followed by object categorization and analysis. There was increased participation in defining simple human activity, such as someone jogging or moving, as implied by a combined application of several motions. The top phase involves realizing the actions of multiple interacting individuals. The results showed a quick image processing time, with an average cycle duration of 61 milliseconds [8].

Fernando *et al.* focused on infrared images and proposed a novel human detection system utilizing a deep convolutional neural network (CNN) and a motion-based model [9]. The initial phase processed the raw images to make up the difference for the noise effect, followed by employing noise cancellation filters to overcome the outlier noise level resulting from the shifts in lighting. They also used histogram equalization to improve frame intensity to regulate uneven light distribution. The second phase considered human detection in the pre-processed images. The advantages of this method are noise reduction with noise cancellation filters, better overall system detection performance, and 5.1 percent better detection performance than the non-pre-processed step [9]. Its disadvantage is the relatively high false detectionrate due to the different body angles.

Bégard *et al.* investigated the performance of the AdaBoost algorithm to solve the problem of real-time pedestrian detection in images [10]. They used gradient-based local features with a cascaded sensor to create a powerful classification model. The researchers compared the original AdaBoost algorithm to two other enhancing algorithms under development. The study proposed a quick, simple, efficient visual identification system for human activities [10]. The disadvantage of the system is that insufficient classification methods could result in low margins and generalization errors.

One research team [11] proposed a two-step real-time pedestrian detection method. The first process is identifying the HOG and the cascade frame classification algorithm. Even though boosting is the weakest classifier in cascade, it correlates directly to the HOG block features. The method could significantly decrease the rate of false positives to varying degrees. The disadvantage of the HOG feature is the absence of the color and texture features, which resulted in several false warnings [11].

Thorat used MATLAB and Normalized Cross-Correlation to monitor and detect motion [12]. The researchers used the normalized cross-correlational research design for movement detection and element-linked assessment for shifting detection and tracking. The proposed method can perform real-time live video compression efficiently. This technique delivers a precise, reliable, sturdy, and immediately noticeable surveyance system. According to the users, the real-time consistency of the video stream with good precision provided the advantage of recognizing and tracking the object in frame sequence [12]. The main drawback is its poor performance in the presence of noise, resulting in a high number of erroneous warnings in the process.

Previous research proposed using radio frequency (RF) measurement for human detection [13]. The primary method used an experimental design with two measurement methods, anechoic chamber, and outdoor range measurements. Pizzillo discussed the benefits and drawbacks of the measurement methods. They contended that the approach could genuinely occupy a new human detection methods data system with information that personifies the radio frequency (RF) biometric signers of humans in an anechoic chamber measurement method and can be used to detect and classify mountable and enemy threats. The disadvantages of using these measurement methods are the damage to the eye and body tissues by RF radiation, and electronic devices such as cardiac pacemakersare sensitive to RF interference [13].

Yadav [14] employed an enhanced frame differencing method to track objects and used MATLAB to implement an automated system. Assessment of the multiple video sequence data revealed a low error rate in identifying objects in motion because once compared to the previous frame differencing technique. The benefits of the proposed method are a simple and easy noise removal process and higher processing speed, while the drawback is the object identification, which causes rapid irradiation alteration, interrupts the tracking procedure [14].

According to [15], MATLAB creates 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 advantages of this algorithm are very few false alarms than the existing approaches and higher efficiency than the currently available research method. They stated that the equipment for this method should be gotten better and require significantly larger space to function much quicker and be suitable for deployment [15].

CNN is particularly appropriate for using pictures as data [16]. The relevance of these methodologies is that the identifier produces better classification outcomes than the earlier technique. The researchers have demonstrated the effectiveness of CNNs and contended that adding a linear layer to term classification techniques could enhance their performance [16]. The limitation of this algorithm is the lack of characteristics of some greater transformer design features.

The approach adopted by [17] comprises four techniques, feature pyramid network, multi-scalefeature fusion, double-branch structure, and network architecture. The pyramid network gives broader interpretations at all stages and is formed rapidly from a single input scale without representational power, performance, or capacity. The multi-scale feature fusion is suitable for overcoming and over-enriching high-frequency information. The double-branch structure could increase and optimize algorithm effectiveness. The alternative approach, network architecture, captured image features andbuilt a depth map [17]. The limitation of multi-scale feature fusion is a high risk of data loss throughout the implementation, which necessitates regular maintenance. This approach is extremely slow inda processing when mixed with pictures.

[18] employed several methods and techniques, including sequential creation by vector quantization, improved hidden Markov model (M-HMM), robust method, and depth image processing using the least square method. The two popular methods for extracting features are feature extraction using depth shape features and feature extraction using joint information elements. The advantages of the methodologies are the ability to distinguish various activities across the system and high-quality photos and outputs. They identify, trace, and classify behavior patterns using detailed shapes, outlines, and body joint information. The disadvantage of this approach is a lower detection percentage, particularly in the absence of combined information and when viewing human silhouettes and subjects from a distant [18].

[19] implemented an improved YOLOV3 model with a Deep CNN, which can detect practically anything and has a high-exactness mechanism. The model is considerably faster than the original networks while maintaining consistency.This technology cannot detect something barely distinguishable even to the naked eye [19].

In summary, the findings of previous works indicate that scientific sensor development requires a comprehensive understanding of the underlying phenomenon, relative strengths, weaknesses, and limitations of the currently available sensing approach. Development of innovative models and algorithms requires approval with well-characterized ground truth data. Table 1 presents a summary of the literature review.

IV. RESEARCH METHODOLOGY

A. Data Collection

The researchers used an electronic camera to record still photos and videos of 100 pieces of data. These are the original footage from the researcher's single click and serve as the raw research relevant to the study goal. For the control photos, the researchers used primary images containing information about the measured distance among people to justify the automated system.

We compiled the raw images and videos using appropriate browser extensions, such as Microsoft Bing and Google Chrome, and the data was renamed and known as secondary. Microsoft Bing and Google Chrome are excellent Internet sites for accessing sources of still images and videos, including data from organizations' database systems, webpages, newspaper articles, or other source materials. The researchers saved the data in JPEG (.jpg,.jpeg), TIFF (.tif,.tiff), AVI (Audio Video Interleave), and MOV (QuickTime Movie) formats.

B. Data Pre-Processing.

Data pre-processing is a method for determining the data objects and features to be used in the next phase of object recognition and data classification. Image pre-processing is modifying images to enhance them through data cleaning and image enhancement. Data cleaning is a data pre-processing to classify insignificant and missing items. Data cleaning will be conducted to address lacking and imprecise in still images and movies. Noisy data is inconsistent data caused by ineffective data collection methods, thus making data evaluation difficult.

Remini is one of the most popular image enhancers for improving still images and videos. Remini is an automated image processing with 26 tools, and users do have to be concerned about the preferred level. It gives pixelated, damaged, low- resolution photographs and movies a fresh new look. Remini astonishes users with photograph clarity and sharpness in high definition, as shown in Figure 1. It uses cutting-edge AI to unblur, restore and optimize pictures and videos

C. Data Analysis and Classification

This phase analyzed and classified the clean and enhanced images from thepre-processing. We examined and classified the data to detect social distance using a few approaches from theCNN algorithm in the MATLABplatform. There might be a small percentage of errors causing false social distance detection. CNN technique, also known as ConvNets, is a frequently used deep learning tool. This research used CNN because of its effectiveness in pattern recognition,especially in recognizing objects, human faces, and scene images. We combined CNN with the Remini image enhancer to increase the accuracy of the social distancing detector and observed a 90% improved accuracy. The software and tools for analysis and classification are MATLAB R2021a, an Image Processing Tool, Computer Vision Tool, and Video Processing Tool.

We used a pre-trained model based on the Aggregate Channel Feature (ACF) classification model in the analysis and classification stage. The recognition method used CNN and a pre-trained model that used the ACF classification model 'inria-100x41' to detect individuals. The INRIA Person dataset was used to coach the 'inria-100x41' conceptual framework. This process optimized the people in the images and saved the results as bounding boxes and a scoring system.

CNNs are artificial neural networks explicitly designed to process pixels for image and video recognition. CNNs are powered by the AI techniques implemented in image and video processing systems to operate the conceptual and insightful duties.

The learning environments routinely use computer vision algorithms to recognize images and videos and classification techniques like the ACF system, which uses a pre-trained model to identify humans in images and videos. CNN begins with the convolution layers, the first layer of a convolutional network [20]. The convolutional layers in this research are guided by an additional convolution layer, also known as max pooling, and the top level of the convolutional layer is a completely connected layer. The CNN will become more detailed as it identifies the high number of videos and pictures for this task. The initial layer develops simple features such as color schemes and sides. As the image data evolves through the CNN layers, it learns to recognize and define the various components or forms of items.

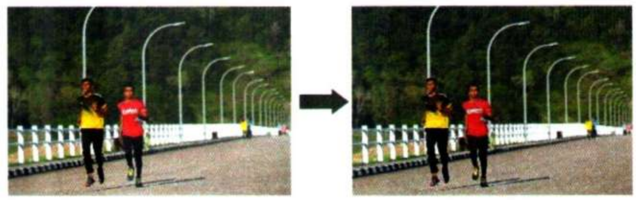

Fig 1. Sharper image after enhancement.

TABLE 1
COMPARATIVE STUDY OF THE CURRENTLY AVAILABE METHODS RELATED TO THIS STUDY.

No.	Title	Method	Advantages
1.	Real-time Human Detection in Thermal Infrared Images	- K-Mean clustering - YOLO algorithm	- Achieve the best high frequency - Excellent priority in bounding boxes
2.	Radio Frequency RF Measurements for Human Detection	- anechoic chamber measurement - Outdoor Range Measurement	- 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
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	- Achieve the best high frequency - Excellent priority in bounding boxes
7.	Real-time Human Activity Recognition	Moving object classification	- Good image processing time
8.	Detection and tracking of moving object	Normalized Cross-Correlation	- High accuracy - Rapid, visible surveyance system - High detection reliability
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 cancellation 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 joints 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 - 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 has an accurate classification - More accurate 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 has to be detected - High precision and accuracy. Much faster than other methods while maintaining accuracy

The fundamental structure of CNN is the convolutional layer, which makes most of the estimations. It consists of data entering and filtration. Like the previous venture, It uses a color photo composed of a 3D element array. The three aspects of the images are size, width, and intensity, which correlate with the RGB of the image. The sensor characteristic, also known as kernel or filter, moves continuously across the image area to verify if the required characteristic is relevant to the construction process. The process in this research is known as convolution. The detector has a two-dimensional or 2D weight range that reflects a segment of the image. Furthermore, the mass in the feature descriptor remained constant as it moved across the picture during the convolution of the layers, a process known as parameter sharing.

The CNN applied a Rectified Linear Unit (ReLU) transformation to the feature map after each convolution operation, which also introduced a nonlinear behavior into the model. The first convolution layer is followed by the second. Even as surface has access to the pixel value of the pictures within the visual field of previous layer upon layer, the CNN architecture will become more hierarchical. For the hierarchical sequence in the neural net, the section integration showed a relatively high structure that resulted in a functionality power structure within the CNN.

The following layer in the CNN for performing this task is the pooling layer, also known as bottom testing, which reduces the number of variables in the insight. Like the convolution layers, the pooling operation eliminated the filtration from the whole insight. Here on task, we used max pooling that further means also the filter would then start moving throughout the insight and choose the pixel with the highest production. Furthermore, the advantage of max pooling in CNN is reduced video and image complexities, higher productivity, and reduced video and image generalization risks.

The final layer used a fully connected layer, whereas, in slightly connected layers, the image pixels of the input were not directly connected to the output nodes. All entities in the fully connected layer as the output nodes automatically connect to the node in the previous layer, which applied characterization that classified the dataset using the attributes derived from the preceding layer and the various filtration used for their appropriate classification. Figure 2 illustrates the architectural style of the CNN layer or Convolutional Neural Networks surface.

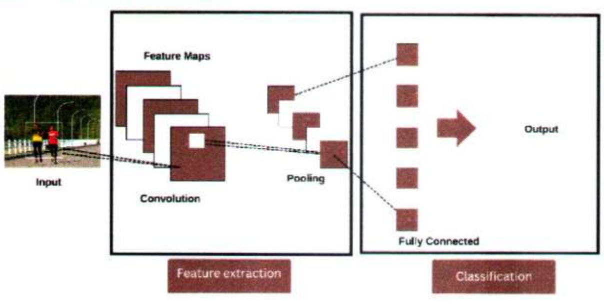

Fig 2. The architecture of the Convolutional Neural Network

D. Data Validation

We tested the accuracy and quality of the processed image throughout the data validation process to ensure the data sources before and after transportation and handling and analyzing the system output. The data validation determines if the proposed system for social distancing has the best accuracy and functions well in detecting human social distancing.

The still image in Figure 3 is an example of distance verification for the photos captured using a smartphone or digital camera. The picture shows the social distance between two individuals that follows the one-meter social distancing requirement. The bounding boxes for both individuals in the picture are green., which has proven stable for the social distancing between the two individuals.

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

V. PSEUDOCODE

Covid-19 Social Distancing Detector (the pseudocode for the image)

```
IF the detector using 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 human figure
    Show the bounding boxes and detection result for image
      Show red bounding boxes and declare as danger
IF ELSE
  IF safe social distancing
    Show green bounding boxes and declare as safe
  ELSE
    Stop the system to run the image
      Restart with 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 result and detect recorded videos in a box
  Show size of recorded videos
END IF
  IF the image chosen to detect human movement and figure
    Show the bounding boxes and detection result for
      recorded videos
    Show red bounding boxes and declare as danger
IF ELSE
  IF safe social distancing
    Show green bounding boxes and declare as safe
ELSE
  Stop the system to run the recorded videos
  Restart with new recorded videos
ENDIF
```

The first pseudocode describes the automatically generated scripting for still pictures to identify and continuously monitor regardless of whether people adhere to social distancing or flout the social distancing rule. An output image with red bounding boxes indicates that the individuals were not complying with the social distancing rule and disobeyed the policies or guidelines. Green bounding boxes indicate that the individuals keep their social distance. The people sensor ACF applied the Aggregate Channel Features 'inria-100x41' to diagnose humans in images. Additionally, peopleDetectorACF reverts an ACF-pretrained erect individual detector. A detector is an acfObjectDetector object which has been applied to the INRIA human set of data.

The second coding is for the CCTV video recordings. This section requires perspective. A VideoFileReader restores an ACF-trained upright people sensor. The analyzer is an acfObjectDetector entity that learned utilizing the INRIA individual large dataset. A pre-trained model was used to identify humans and decide regardless of whether they adhere to social distancing guidelines. Red bounding boxes in the footage indicate that the individuals do not adhere to the required social distancing. Green bounding boxes indicate that the individuals abide by the social distancing recommendation.

VI. RESULTS

Figures 4, 5, and 6 show the secondary data. The images are from various locations of arecreation park at different times. Figure 4 shows two individuals jogging in the recreation park while maintaining social distance. As a result, the meters measured follows the World Health Organization's social distancing regulations, and The boundingboxes which detected the individuals are green, indicating a safe distance between them.

Fig 4. The image detected at a recreation park

Fig 5. The image detected at a recreation park

Fig 6. The image detected at a recreation park

Figure 5 shows a one-meter social distance between three people in the image. The bounding boxes are green, indicating an appropriate social separation between theindividuals. It is comparable to the image used to validate whether individuals adhere to the one-meter social distancing rule. Figure 6 shows a group of people jogging and walking in a recreation park. The system assessed for safe social separation and displayed greenboxes that indicate safety.

Figures 7 to 9 show the images captured with a smartphone and a camera, which serve as the primary data. The source data from the creator and personal data are examples ofprimary data. The data was from crowded places such as shopping 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 are from the videos recorded with a camera and categorized as primary data. Thevideo was recorded at the Kuala Lumpur LRT station to test the system's ability to detect and identify humans in video and detect human social separation in the video. The accuracy of the video increased from 80 to 87 percent because several individuals were mistakenly detected by the system in the video. Furthermore, a comparison with the still image used for validation showed that the detection percentage is quite good for the video to identify human social distancing.

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. The results showed that the method has a detection rate of as high as 90%. This research used a CNN with a pre-trained model employing the ACF classifier model known as 'inria-100x41' to apply the research findings in massive community arenas such as a recreation center or even a business district because it provides a surveillance system and real-time information to government regulatory agencies whether the public complies with the WHO social distancing rules.

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

TABLE 2
ACCURACY COMPARISONS BETWEEN THE IMAGES AND VIDEOS.

Images	- The image detection accuracy is 90%. - The algorithm successfully identified most images. - Compared to the still image used for validation, the result showed an accuracy of approximately 90% since the method showed the correct bounding boxes if the individuals practiced the one-meter social distancing.
Videos	- The collected videos have an accuracy of between 80 to 87 percent. - The recorded video is realistic 80 to 85 percent of the time since some videos did not precisely and accurately locate the individuals after loading the data for recognition. - Some results were not precise and did not produce accurate bounding boxes because they had messy data.

CNN is a powerful algorithm for image and video processing. These algorithms are currently the most effective for automated image and video processing. Many have used these algorithms to detect people or features in images and videos. Images and videos have RGB values, and MATLAB can extract an image or video from a file. The color information is in three-dimensional arrays, where the first two dimensions are the image height and width or pixel count. The third is the image's red, green, and blue color schemes. After detecting a human figure in an image or video, the RGB coding specifies the bounding box as green or red.

The research findings showed that the average accuracy for the images was higher than the video, where the bounding boxes in the images were more accurate in detecting people than in the recorded video. Thecorrect detection is when the bounding boxes identify the required social distance. Green bounding boxes mean that the individuals who practice social distancing are safe, while red bounding boxes show that those individuals flout the social distancing rule and put themselves at risk of contracting COVID-19. The average accuracy for videos is lower because multiple individuals in the video were assigned the wrong-colored bounding boxes.

VII. CONCLUSION

This framework could help the government control COVID-19 transmission. The experiment has demonstrated that the system is functional and capable of detecting the distance between humans and classifying it as safe or unsafe. The use of CNN gives this scheme an approximate accuracy rate of 90%. It is a pre-trained human detection method employing a pre-trained model leveraging the ACF classifier model, or Aggregate Channel Feature (ACF) known as the 'inria-100x41', which is the best option for recognizing humans. Moreover, this system could encourage new social norms and help the authorities reduce the number of COVID-19 infections.

We urge future research to enhance and improve the system by adding a scoring number in the bounding box. The security teams controlling the system in their zones or locations, including shopping centers and recreation facilities, could obtain an exact and reliable count of the number of people passing through the area.

Future research could provide a real-time camera embedded with the Deep Learning algorithm in large premises for the target market consumer, including government organizations, shopping complex managers, or even clients who require it in their area to facilitate the monitoring of social distance between individuals. The system and sensor could have additional alert mechanisms or sirens that sound if people fail to maintain social distancing.

VI. ACKNOWLEDGEMENT

The authors thank the National Defence University of Malaysia for financing the publication fees. We are also grateful for the comments from the individuals involved in this journal publication.

VIII. REFERENCES

[1] Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S., and Di Napoli, R. (2021). Features, evaluation, and treatment of coronavirus (covid-19). StatPearls.

[2] Naqiyuddin, F. A. A., Mansor, W., Sallehuddin, N. M., Mohd Johari, M. N. S., Shazlan, M. A. S., & Bakar, A. N. (2020). Wearable Social Distancing Detection System. *2020 IEEE International RF and Microwave Conference (RFM)*.

[3] Thompson, H. (2010). E-security and critical national infrastructures. In Handbook of Electronic Security and Digital Forensics, pages 295–320.

[4] Mohammadi, M., Shayestehpour, M., and Mirzaei, H. (2021). The impact of spike mutated variants of sars-cov2 alpha, beta, gamma, delta, and lambda on the efficacy of subunit recombinant vaccines. The Brazilian Journal of Infectious Diseases, page 101606.

[5] Manssor, S. A., Sun, S., Abdalmajed, M., and Ali, S. (2021). Real- time human detection in thermal infrared images.

[6] Hussein, M., Abd-Almageed, W., Ran, Y., and Davis, L. (2006). Realtime human detection, tracking, and verification in uncontrolled camera motion environments. In Fourth IEEE InternationalConference on Computer Vision Systems (ICVS'06), pages 41–41. IEEE.

[7] Sayadi, F., Said, Y., Atri, M., and Tourki, R. (2012). Real time human detection in video streams. Recent Research in Automatic Control, Systems Science and Communications, pages 154–158

[8] Albukhary, N. and Mustafah, Y. (2017). Real-time human activity recognition. In IOP Conference Series: Materials Science and Engineering, volume 260, page 012017. IOP Publishing.

[9] Fernando, H., Perera, I., and de Silva, C. (2019). Real-time human detection and tracking in infrared video feed. In 2019 Moratuwa Engineering Research Conference (MERCon), pages 111–116. IEEE.

[10] Bégard, J., Allezard, N., and Sayd, P. (2008). Real-time human detection in urban scenes: Local descriptors and classifiers selection with adaboost like algorithms. In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pages 1–8. IEEE.

[11] Zhihui, L., Chunyan, S., and Di, S. (2011). Real-time human detection based on cascade frame. In 2011 IEEE International Conference on Mechatronics and Automation, pages 514–518. IEEE.

[12] Kavita V. Wagh, Kulkarni R.K. (2013). *Human Tracking System*. International Journal of Engineering Sciences & Research Technology.

[13] Pizzillo, T. J. (2007). Radio frequency (rf) measurements for human detection, tracking, and identification. Technical report, Army Research Lab Adelphi Md Sensors and Electron Devices Directorate.

[14] Yadav, S. P. (2020). Vision-based detection, tracking, and classification of vehicles. IEIE Transactions on Smart Processing & Computing, 9(6):427– 434

[15] Mayya, M., Zarka, N., and Alkadi, M. S. (2010). Embedded system for real-time human motion detection. In 2010 2nd International Conference on Image Processing Theory, Tools and Applications,pages 523–528. IEEE.

[16] Potkin, O. and Philippovich, A. (2019). Hand gestures detection, tracking and classification using convolutional neural network. In International Conference on Analysis of Images, Social Networks and Texts, pages 263– 269. Springer.

[17] Niu, Y. and Meng, Z. (2020). Research on object detection technology for human detection. In Journal of Physics: Conference Series, volume 1544, page 012076. IOP Publishing.

[18] Kamal, S., Jalal, A., and Kim, D. (2016). Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified hmm. Journal of Electrical engineering and technology, 11(6):1857–1862.

[19] Naik, B. T. and Hashmi, M. F. (2021). Ball and player detection & tracking in soccer videos using improved yolov3 model.

[20] Ansari, M. A. and Singh, D. K. (2021). Monitoring social distancing through human detection for preventing/reducing covid spread. International Journal of Information Technology, 13(3):1255–1264.