International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Real-Time Facial Cover Detection System based Deep Convolution Neural Network with Proposed Activation Function

Nay Kyi Tun[1*], Aye Min Myat[2]

[1]Associate Professor, Faculty of Computer System and Technology, Myanmar Institute of Information Technology
[2]Professor, Faculty of Information and Communication Technology, University of Technology (Yatanarpon Cyber City)
[*]Corresponding Author Profile at the end of the Paper

Abstract
The ongoing global contagion has highlighted the importance of effective preventive measures such as wearing face masks in public spaces. In this study, we suggest a deep learning-based approach for real-time facial covering detection to aid in enforcing mask-wearing protocols. Our system utilizes deep learning networks (CNNs) to automatically detect whether individuals in images or video streams are wearing mask or not. The suggested system includes of 3 main stages: face detection, facial cover detection, and real-time monitoring. Firstly, faces are localized in the input image or video frame using a proposed face detection model. Then, the detected faces are fed into a proposed CNN model for mask classification, which determines whether each face is covered with a mask or not. Finally, the system will provide real-time monitoring and alerts authorities or stakeholders about non-compliance with mask wearing guidelines. We appraise the execution of our system on publicly available datasets and demonstrate its effectiveness in accurately detecting face masks in various scenarios. Additionally, we discuss the challenges and limitations of deploying such as system in real-world settings, including issues related to privacy, bias, and scalability. Overall, our proposed facial covering detection system offers a viable solution for automated monitoring and enforcement of face mask policies, contributing to public health efforts in mitigating the spread of contagious diseases.

Keywords: CNN, Face Mask, Detection, Classification, YOLO.

1. Introduction

- The emergence of the COVID-19 outbreak has necessitated the adoption of stringent public well-being measures, with widespread face mask usage being among the foremost effective means of preventing the spread of infectious diseases in communal settings.
- However, guaranteeing adherence with mask-wearing protocols in community spaces presents a formidable challenge for, authorities and organizations worldwide.
- In response to this challenge, the progress in automated face mask detection technology has attracted considerable interest as a promising approach to effectively enforce mask-wearing guidelines.

- This article provides an in-depth exploration of latest developments in the detection of face mask technology, examining the underlying methodologies, applications, obstacles, and future prospects of systems like.
- The contagion persists to evolve, the need for reliable and scalable solutions for monitoring face covering compliance remains paramount.
- By leveraging cutting edge technologies such as computer vision, deep learning, edge computing, researchers and practitioners have made substantial progress in the development of robust and real-time face covering detection systems.
- Deep learning networks (CNNs), in particular, have displayed exceptional abilities in accurately detecting faces and distinguishing between masked and unmasked individuals.
- Moreover, the integration of complementary technologies such as thermal imaging, edge computing, and privacy-preserving techniques has further enhances the utility and efficacy of facial covering detection systems in diverse operational settings.
- Beyond the immediate imperative of pandemic management, face mask detection systems hold immense potential for addressing broader societal challenges, including security surveillance, access control, and public safety monitoring.
- By leveraging the insights gained from research and real-world deployments, policymakers, businesses, and public health authorities can develop evidence-based strategies to promote mask compliance and mitigate the risk of disease transmission.
- However, the ubiquitous acceptance of facial covering detection systems also raises important ethical, legal, and societal considerations.
- Issues related to privacy, bias, algorithmic fairness, and consent must be carefully addressed to ensure that these systems are deployed responsibly and ethically.
- Furthermore, the interoperability and standardization of facial covering identification technologies are crucial to facilitate smooth incorporation into current infrastructure and compatibility across various platforms.
- In light of these considerations, this article aims to present a comprehensive synopsis of the YOLO [5] based facial covering identification systems, offering insights into the technological advancements, practical applications, and ethical implications of this rapidly evolving field.
- By synthesizing existing research and identifying key challenges and opportunities, we seek to inform future research directions and contribute to the development of effective and equitable solutions for promoting mask compliance and safeguarding public well-being.

2. Related Work

- A. Velip and A. Dessai [1] proposed a multi-task learning framework combining face detection and mask classification, achieving high accuracy in diverse environmental conditions. This study proposes a real-time facial covering detection system using deep learning techniques. The authors employ a CNN architecture for face detection and mask classification.
- S. Sakshi et al. [2] present a facial covering detection system based on a CNN model trained on a large dataset of masked and unmasked faces. Their system exhibits robust performance across various environmental conditions and lighting scenarios. Additionally, they explore the impact of different CNN architectures on detection accuracy.

- N. Kowalczyk et al. [3] propose a comprehensive face mask detection and recognition system integrated with thermal imaging technology. Their system combines deep learning-based mask detection with facial recognition to identify individuals and enforce mask-wearing protocols in public spaces. The study emphasizes the importance of multimodal approaches for enhanced accuracy and reliability.
- D. Singh and S. K. Joshi [4] introduce a lightweight CNN model tailored for real-time facial covering detection on edge devices with limited computational resources. Their system achieves competitive performance while maintaining low computational overhead, making it suitable for deployment in resource- constrained environments such as embedded systems and IoT devices.

3. Experimental Procedures

- These studies represent a diverse range of approaches and methodologies for facial cover detection, spanning from real-time systems using deep learning to privacy preserving solutions leveraging federated learning.
- By building upon and extending the findings of these works, our proposed facial covering detection system aims to contribute to the advancement of this critical area of research.
- In this strategy, we suggested us innovative YOLO-based detection algorithm for face covering detection model system.
- The strategy diagram with our suggested system includes both the learning and evaluation segment that were shown in Fig. 1.

Figure 1: Block diagram of proposed facial cover detection system

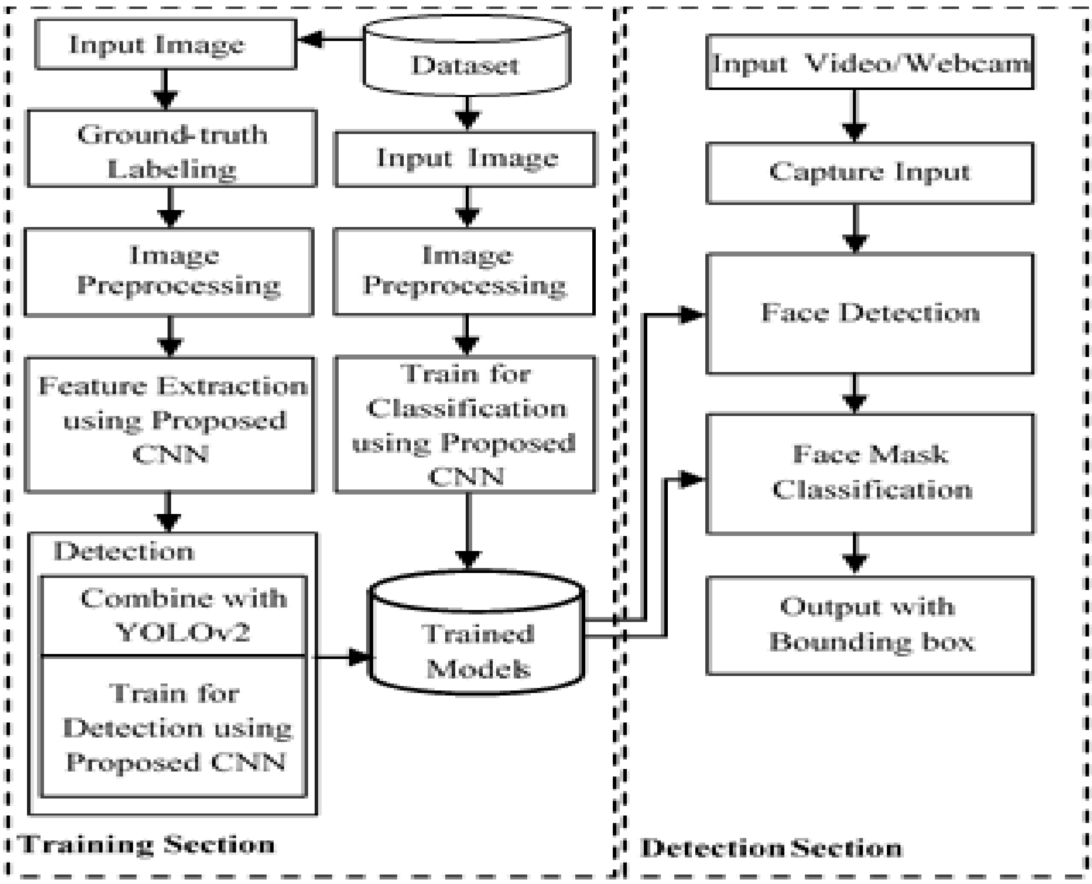
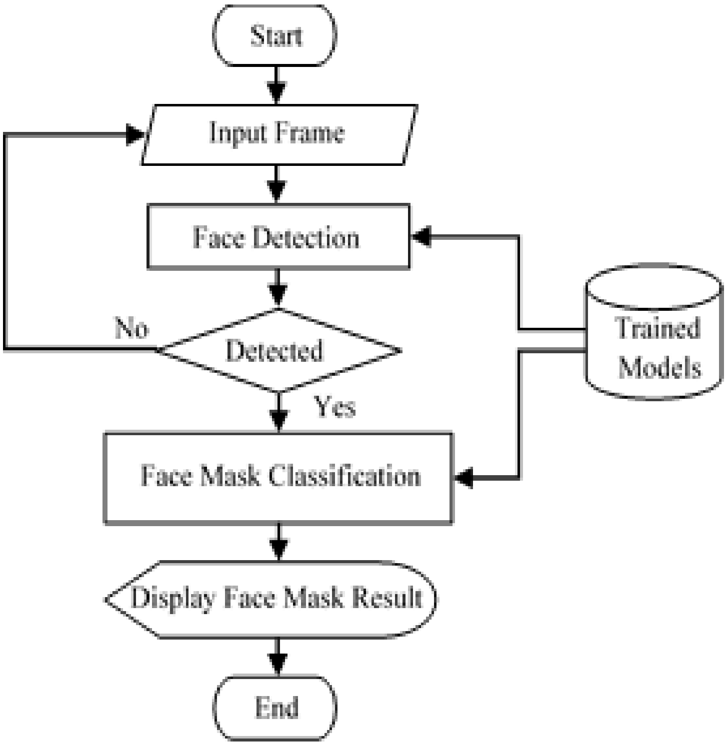


Figure 2: Flow chart of proposed facial cover detection system

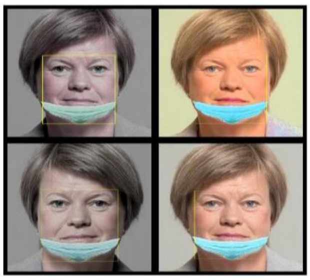


- During the learning phase we initially construct a ground truth collection by annotating the initial image collection for detection objective.
- And we apply preprocessing to the image. We extract features of the image by using our proposed Convolution Neural Network without consisting of the fully connected layer.
- The architecture is then combined with the YOLOv2 detection network model for this stage of detection. Following this, we proceed to train the pictures collection.
- Regarding the training of the classification model, the entry picture size standardized to 224 x 224 pixels.
- Decreasing the picture size results in a reduction of image characteristics.
- After resizing the face mask images, all these pictures are trained using the suggested CNN for face covering classification.
- In the detection phase, input is sourced from either a video feed or a webcam.
- The recommended system captured this entry picture frames and detected faces using our envisioned detection model.
- In the process flow diagram of the proposed system, the system is started by capturing the input frame from the webcam, Face is detected by using the proposed object detection method which is based on the YOLOv2 algorithm [6].
- If the input frame detects face, the proposed CNN architecture is applied for the classification of face mask.
- And then, the value of classification results is shown as an output.
- If the input frame devoid a face or does not detect face, the process of detection will return to the input stage, which is shown in Fig. 2.

3.1 Preprocessing

- In the stage of the preprocessing for detection, data augmentation is applied to the ground truth label image.
- We used data augmentation to improve the network accuracy and we also made a random transformation to original data to increase the labeled training data.
- We apply random horizontal flipping and random X/Y scaling.
- We also use color space transformation to convert RGB to HSV color space.
- And then, we jitter image color to make randomly augment the color of each pixel.
- As for the stage of training for face mask classification, the data augmentation techniques are also used that includes scale augmentation and position augmentation. Scale augmentation technique and position of training images.
- We crop and resize the detected face image into 224 x 224 x 3-pixel image.
- After the stages of preprocessing, we trained the detected model and classification model by using our proposed CNN architectures. The preprocessing result of detection are shown in Fig. 3.

Figure 3: Preprocessing of input dataset image for detection



3.2 Proposed Detection Network based on YOLOv2 Algorithm

- In our proposed system, we make we make detection of face by using the proposed detection model which is transferred from the You Only Look Once version 2 model.
- Before the development of the YOLOv2 algorithm, we augmented the dataset which is not only to enlarge our utilized dataset artificially but also to lessen the likelihood of overfitting throughout the training stage [7].
- In the architecture of CNN, the input stage accepts the image with a size of 416 x 416 with RGB.
- The architecture and parameters of the proposed face detection model is shown in Fig. 4 and Table 1.
- The architecture consists of the attribute extraction part and detection part.
- In the feature extraction, we used our tiny CNN architecture which is shown in Table 1.
- In a typical CNN, fully connected layers are usually placed toward the conclusion of architecture [8].
- In our proposed architecture, we operate a series of convolution processes without consisting of a fully connected layer because we want to replace the YOLO output layer.
- Finally, the result of face detection displays as an output. There are 8 convolution layers and each convolution layer consists of convolution, batch normalization and proposed activation function layer.

Table 1: The Layer of Proposed CNN Architecture for Face Detection

Type	Filters	Size/Stride	Output
Convolute - 1	32	3 × 3	416 × 416
MaxPool-1		2 × 2/2	208 × 208
Convolute - 2	64	3 × 3	208 × 208
Convolute - 3	64	3 × 3	208 × 208
MaxPool-3		2 × 2/2	104 × 104
Convolute– 4	128	3 × 3	104 × 104
MaxPool-4		2 × 2/2	52 × 52
Convolute– 5	256	3 × 3	52 × 52
MaxPool-5		2 × 2/2	26 × 26
Convolute– 6	512	3 × 3	26 × 26
MaxPool-6		3 × 3/2	12 × 12
Convolute– 7	1024	3 × 3	12 × 12
MaxPool-7		3 × 3/2	10 × 10
Convolute– 8	1024	3 × 3	10 × 10

Table 2: The Variables of Proposed CNN Architecture for Facial Covering Classification

Layer	Filter/Weight	Number of Parameters
Convolute Layer (C1)	((3×3×3) +1) ×16	448
Convolute Layer (C2)	((3×3×16) +1) ×32	4,640
Convolute Layer (C3)	((3×3×32) +1) ×64	18,496
Convolute Layer (C4)	((3×3×64)+1) ×128	73,856
Convolute Layer (C5)	((3×3×128)+1) ×256	295,168
Convolute Layer (C6)	((3×3×256)+1) ×512	1,180,160
Convolute Layer (C7)	((3×3×512)+1) ×1024	4,719,616
Fully Connected (F1)	((1024) +1) ×100	102,500
Fully Connected (F2)	((100) +1) ×3	303
Total parameter		6,395,187

3.3 Proposed Classification Network Architecture

- After the detection of the face, the process of classification is applied to the detected face by using the proposed CNN model.
- There are several architectures in the field of Convolution Networks that have a name.
- The most common are AlexNet, VGGNet, and GooglrNet.
- The architecture of the proposed classification method using deep learning network (CNNs) is shown in Fig. 5.
- In the architecture of CNN for face mask classification, we resize the input image into 224 x 224 with RGB.
- There are seven convolution layers and six max-pooling layers. After passing the two Fully-Connected layers and Software layer, the final size will be reduced to 1 x 1 x 2.
- The classification output layer produces the output corresponding to each group.
- The output is related to the 3 classes for the classification of without facial covering, with facial covering and wrong position of facial covering.
- The variables of the proposed CNN for facial covering classification are shown in Table 2. The filters count and the size of output feature maps are shown in it.

Figure 4: Proposed detection network architecture

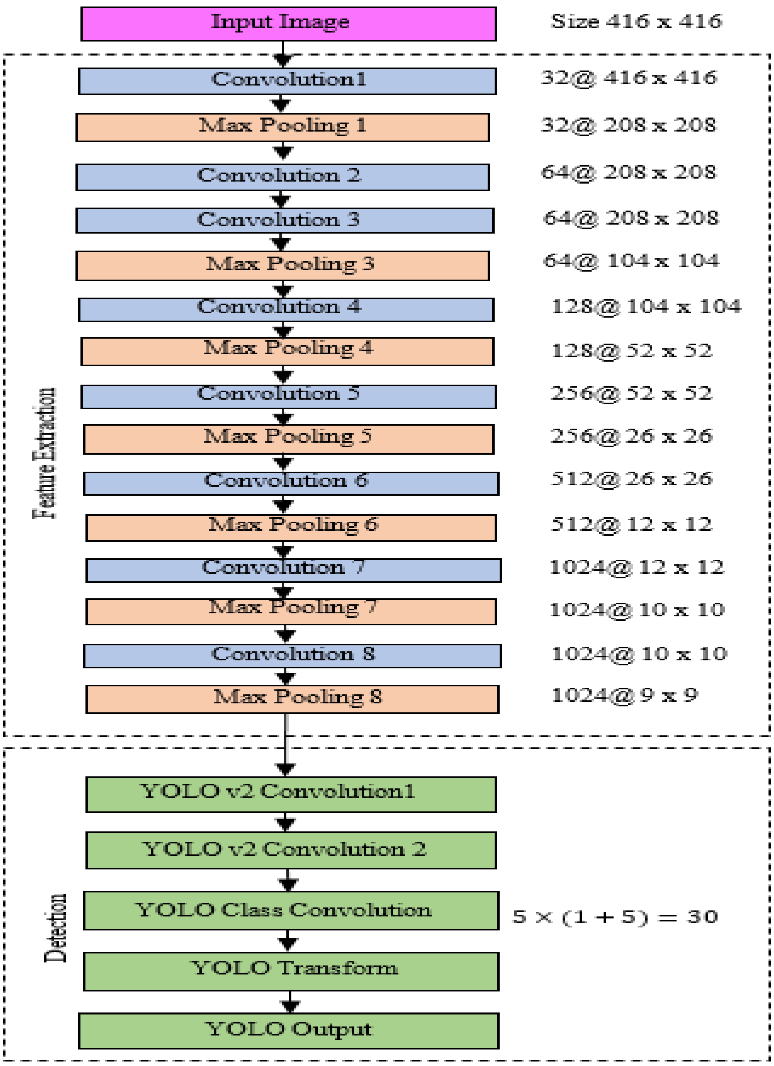


Figure 5: The architecture of the proposed CNN for facial cover classification

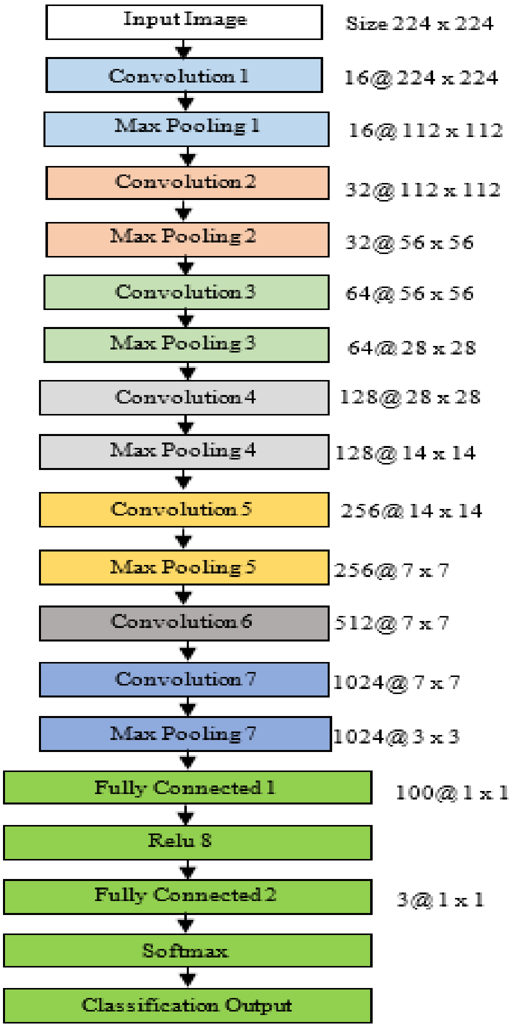


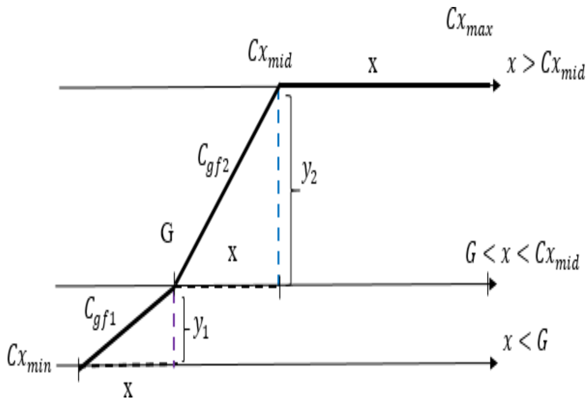
Table 2: The Variables of Proposed CNN Architecture for Facial Covering Classification

Layer	Filter/Weight	Number of Parameters
Convolutional Layer (C1)	((3×3×3) +1) ×16	448
Convolutional Layer (C2)	((3×3×16) +1) ×32	4,640
Convolutional Layer (C3)	((3×3×32) +1) ×64	18,496
Convolutional Layer (C4)	((3×3×64) +1) ×128	73,856
Convolutional Layer (C5)	((3×3×128) +1) ×256	295,168
Convolutional Layer (C6)	((3×3×256) +1) ×512	1,180,160
Convolutional Layer (C7)	((3×3×512) +1) ×1024	4,719,616
Fully Connected (F1)	((1024) +1) ×100	102,500
Fully Connected (F2)	((100) +1) ×3	303
Total parameter		6,395,187

3.4 Proposed Activation Operation

- For the feature extraction of CNN, the suggested activation operation was used to follow the convolution layer.
- The suggested activation operation extracts the feature without reducing the negative value of the convoluted feature map.

Figure 6: Proposed Activation Operation



The mathematical of the suggested activation operation is shown in equation (3.1)

$$\text{PAct} = \begin{cases} x & , \quad x > Cx_{mid} \\ x \times C_{gf2} & , \quad G < x < Cx_{mid} \\ x \times C_{gf1} & , \quad x < G \end{cases} \tag{3.1 a}$$

$$Cx_{mid} = \frac{Cx_{max} + Cx_{min}}{2} \tag{3.1 b}$$

$$G = \frac{Cx_{max} - Cx_{min}}{4} + Cx_{min} \tag{3.1 c}$$

$$C_{gf2} = \frac{x - G}{Cx_{mid} - G} \times y_2 \tag{3.1 d}$$

$$C_{gf1} = \frac{x - Cx_{min}}{G - Cx_{min}} \times y_1 \tag{3.1 d}$$

- Where x is a convolution layer value, CXmax is maximum value of convolution layer, CXmin is minimum value of convolution layer, CXmid is midpoint value of convolution layer, G is gradient start point of feature value, $Cgf1$, $Cgf2$ are gradient of feature values and y1 and y2 are the constant leak factors. In the suggested system, the value of y1 is 1 and the value of y2 is 0.01.

4. Results and Discussion

- The training of detection network used a Stochastic Gradient Descent with Momentum (SGDM) optimizer, with an initial learning rate of 0.001, mini-batch size of 5, and 32 maximum number of epochs.
- The training of detection model is done to 7200 images of 1 class.
- The detector model is done to 7200 images of 1 class.

- The detector model is done to 7200 images of 1 class.
- The detector model is then saved and used for the detection of 2400 testing images.
- This is done by randomly using 60% of all images for training, 20% for validation and 20% for testing.
- The average precision of detection is 0.97%. Our proposed system is work well in different positions and light conditions. It also works in blur conditions.
- As for the facial covering classification, the training used the SGDM optimizer, with an initial learning rate of 0.001, mini-batch size of 30 and 7 maximum number of epochs.
- Then training is done to 15000 images of 3 classes.
- The classifier model is then saved and used for the classification of 3000 validation images, and 3000 testing images.
- This is done by randomly using 60% of all images for training, 20% for validation and 20% for testing.
- The process of proposed feature extraction is compared with ReLU activation function which is shown in Fig. 7.
- The suggested activation operation base feature extraction method extracts the feature of image more accurately than ReLU activation function-based method.
- The average precision of proposed detection model is compared with YOLOv1, YOLOv2 that is shown in Fig. 8.
- The process is based on training with SGDM optimizer.
- The average precision of YOLOv1 based model is 0.919, YOLOv2 based model is 0.941 and proposed model is 0.979.
- According to the result, the SGDM optimizer based proposed detection model has a higher accuracy than YOLOv1 and YOLOv2.

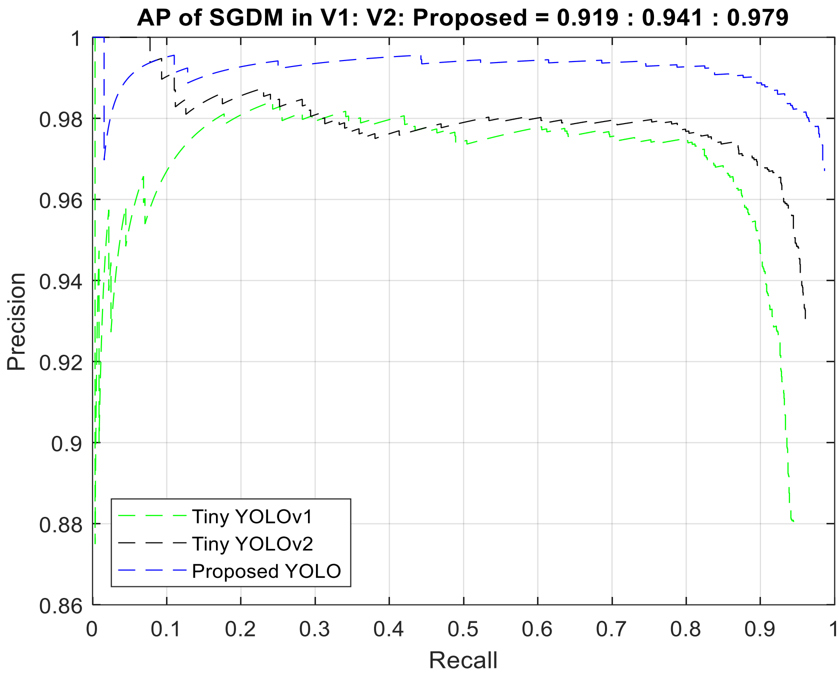
Figure 7: Visualization results of feature extraction (a) Original image, (b) Feature extraction with ReLU activation and (c) Feature extraction with Proposed activation

(a) (b) (c)

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Figure 8: Comparative analysis of the detection model



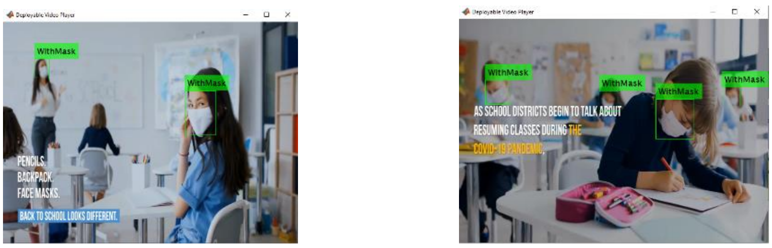
- Proposed detection model is tested facial covering detection in daylight, room light, different background color, horizontal position and vertical position.
- We expressed our experimental result of detection and classification that includes working in the real-time conditions in Fig. 9.
- The confusion matrix of the facial covering classification is shown in Fig. 10 which is based on the 3 classes of 3000 test dataset images.
- Table 3 shows the accuracy result of the proposed classification model that is obtained from the confusion matrix results.
- According to the average result, the precision is more than recall.
- The average result of the negative samples of all sentences is over 99.9% which shows the specificity of the classification model.
- The accuracy of each class is over 96% and the average accuracy of all classes is 96%.

Figure 9: Result of facial covering detection (a) Facial covering detection on image, (b) Facial covering on image, (c) Facial covering detection on video and (d) Facial covering detection on real-time.

(a)

(b)

(c)



(d)

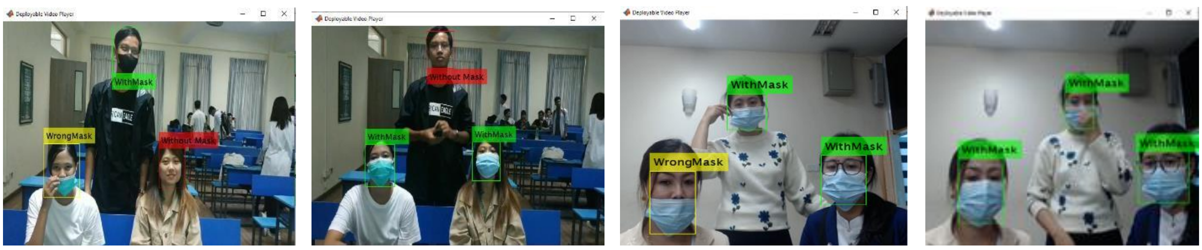


Figure 10: Confusion Matrix of proposed System

	FM	Without FM	Wrong FM
FM	942	12	46
Without FM	12	947	41
Wrong FM	50	15	935

Table 3: Accuracy Results of Facial Covering Classification

Class	Precision	Recall	Specificity	F1 Score	Accuracy
1	0.938	0.942	0.968	0.940	0.9592
2	0.972	0.947	0.985	0.959	0.9724
3	0.914	0.935	0.955	0.924	0.9489

5. Conclusion

- In conclusion, facial covering detection systems represent a crucial technological advancement in the ongoing efforts to combat infectious diseases and promote public health.
- Throughout this article. We have explored the evolution of facial covering detection technology, from traditional image processing methods to sophisticated deep learning algorithms.
- Our analysis has highlighted the significant strides made in improving the accuracy, reliability, and scalability of facial covering detection systems.
- Real-world deployments and case studies have demonstrated the practical utility of these systems in enforcing mask-wearing protocols and mitigating the spread of contagious diseases in various settings.
- Looking ahead, future research and development efforts should focus on addressing the remaining challenges and limitations of facial covering detection systems, including dataset bias, robustness to environmental conditions, and interoperability across different platforms.

6. Acknowledgement

I am profoundly grateful of my colleagues, whose insightful opinions and life in general. Throughout this journey, my research mentor, Dr. Aye Min Myat, has been an invaluable source of guidance and nurturing, providing immense insight. The success of this study endeavor has been greatly attributed to the combined efforts and support of all of these collaborators, for which I am incredibly appreciative.

7. References

1. A. Velip and A. Dessai, "Face Mask Detection Using Machine Learning Techniques," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-5, doi: 10.1109/ASIANCON55314.2022.9908873.
2. S. Sakshi, A. K. Gupta, S. Singh Yadav and U. Kumar, "Face Mask Detection System using CNN," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 212-216, doi: 10.1109/ICACITE51222.2021.9404731.
3. N. Kowalczyk, M. Sobotka and J. Rumiński, "Mask Detection and Classification in Thermal Face Images," in IEEE Access, vol. 11, pp. 43349-43359, 2023, doi: 10.1109/ACCESS.2023.3272214.
4. D. Singh and S. K. Joshi, "Masked Face Detection using Lightweight Deep Learning based Model for IoT based Healthcare applications," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-7, doi: 10.1109/MysuruCon55714.2022.9972485.
5. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
6. J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 66518-66520.

7. C. R. Alimboyong, A. A. Hernandez and R. P. Medina, "Classification of Plant Seedling Images Using Deep Learning," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1839-1844.
8. Salman Khan, Hossein Rahmani, Syed Afaq Ali Shah, Mohammed Bennamoun, "A Guide to Convolutional Neural Networks for Computer Vision," 2018, pp. 56.

Corresponding Author Summary Profile

My name is Nay Kyi Tun. I'm an Associate Professor in the Faculty of Computer System and Technology of Myanmar Institute of Information Technology. I'm research interest including image processing, Deep Learning, Artificial Intelligent and Project Management. Now, I'm doing the Ph.D. research with Deep Learning.

Institution email address – naykyitun.utycc@gmail.com nay_kyi_tun@miit.edu.mm
Institution Website - https://www.miit.edu.mm/
Institution googles scholar page - https://www.miit.edu.mm/
Publish paper lists - https://www.miit.edu.mm/people/daw-nay-kyi-tun

Published Papers

No	Title	Conference/Journals	Local/ International
1	Template Matching Based Eye Detection in Facial Image	Parallel and Soft Computing, Yangon, Myanmar	PSC (2011), 6th Local Conference on parallel and Soft Computing
2	Arduino Based Voltage Measurement System	Annual University Journal on Innovative Research and Products (AUJIRP)	AUJIRP Vol-1, Issue-1, December 2018 (Local Journal)
3	Microcontroller Based Electronic Voting System	Annual University Journal on Innovative Research and Products (AUJIRP)	AUJIRP Vol-2, Issue-1, December 2019 (Local Journal)
4	Obstacle Detection System for Unmanned Ground Vehicle	Journal Of Research and Application (JRA)	JRA Volume-01, Issue-01, 2019 (1^{th} Local Journal)
5	Smart Home Security System Based on GSM network	Annual University Journal on Innovative Research and Products (AUJIRP)	AUJIRP (2020), Volume-03, **Issue-01(ISSN-2709-0418)** (Local Journal)

6	Real-Time Face Mask Detection Based on Convolutional Neural Network	Journal of Research and Innovation (JRI)	**2023, vol 6** **ISSN: 2709-6506** (Local Journal)
7	Deep Learning-Based Real-Time Mask Detection for Human Using Novel YOLOv2 with Higher Accuracy	Proceedings of Fifth International Conference on Computer and Communication Technologies	IC3T 2023, Volume 2 (Springer) International Conference Date: 22.10.2023(online)
9	Real-Time Face Mask Detection system based on Transfer learning for Public Safety	The 15^{th} International IEEE Conference on Computing, Communication and Networking Technologies (ICCCNT)	ICCCNT 2024, Conference Date: June 24-28(5day Hybrid Fest-participate online/offline)