Technical Journal

Leveraging Convolutional Neural Networks for Face Mask Detection

Smita Adhikari[1,*]
[1]Lecturer, Department of Electronics and Computer Engineering, IoE, Pashchimanchal Campus
* Corresponding email : smita.adhikari@pasc.tu.edu.np
Received: January 20, 2024; Revised: March 20, 2024; Accepted: June 25, 2024

Abstract

In response to the widespread use of face masks for safety, security, and health reasons, the development of automated systems for face mask detection has garnered significant attention. This research addresses binary face mask detection by leveraging Convolutional Neural Networks (CNNs) to create a specialized model. Using a Kaggle dataset, preprocessing steps including image resizing and color space conversion are applied for standardized data. The method entails a purpose-built CNN architecture comprising multiple convolutional layers with ReLU (Rectified Linear Unit) activations and max pooling for efficient feature extraction and spatial information capture. Further bolstering the architecture, fully connected layers coupled with dropout layers mitigate overfitting risks, enhancing generalization. The final sigmoid-activated output layer facilitates precise binary classification, distinguishing individuals with or without masks. Training guided by the Adam optimizer ensures parameter optimization based on accuracy metrics. This work contributes a meticulously designed CNN architecture optimized for face mask detection, showcasing robust feature extraction, spatial complexity handling, and overfitting mitigation, thereby presenting a potent solution with broad implications across industries for safety and public health measures. By aligning technology with societal needs, the research aids diverse industries in integrating automation and Artificial Intelligence for mask-wearing compliance. The findings underscore mask detection's pivotal role in overcoming challenges for safer environments.

Keywords: *Automated system, Convolutional Neural Networks (CNNs), Face mask detection, Safe environments*

1. Introduction

The pervasive use of face masks, whether for safety, security, or anonymity, has spurred significant interest in automated systems capable of detecting mask presence on individuals. The task of face mask detection holds immense practical value across industries, from retail environments to surveillance systems, where adherence to mask-wearing protocols is crucial. In recent times, the COVID-19 outbreak has led to unprecedented changes in everyday life, with the widespread adoption of face masks as a critical measure to mitigate the virus's movement around the world. This necessity has spurred the exploration of novel technologies to automate the process of identifying individuals adhering to mask-wearing guidelines.

Machine learning has significantly advanced the field of automatic face mask detection by leveraging data-driven learning and sophisticated algorithms. Utilizing Convolutional Neural Networks (CNNs) and various deep learning models, machines can extract intricate visual features from images and videos, enabling them to discern between individuals wearing and not wearing masks. This data-driven approach eliminates the need for explicit programming, allowing the technology to adapt to various scenarios, lighting conditions, and mask types. Additionally, machine learning's ability to generalize patterns to unseen data ensures accurate detection across diverse environments. These systems not only reduce the need for human intervention in monitoring mask compliance but also facilitate real-time analysis, contributing to public health efforts by ensuring adherence to mask-wearing guidelines and creating safer spaces.

Face mask detection tackles public health compliance, ensuring adherence to mandates in crowded spaces and workplaces. It aids contactless access control, bolstering safety measures in entry systems. Automated surveillance benefits from this technology, monitoring adherence in public areas. Additionally, it supports educational efforts by providing real-time feedback, fostering responsible mask-wearing behavior.

Machine learning systems offer swift and adaptable solutions for crowded areas, enabling instant identification and intervention in places like airports and commercial spaces. Their scalability enhances safety protocols efficiently, seamlessly integrating with existing infrastructure and surveillance networks. Overall, their role in automating face mask detection demonstrates their potential to address modern challenges, fostering safer and more compliant societies. Convolutional Neural Networks (CNNs), a particularly effective technological development in the field of computer vision, present a convincing answer to this problem. CNNs are skilled at recognizing complex visual features and patterns, making them well-suited to distinguish between faces with and without masks. Convolutional Neural Networks (CNNs) offer a promising approach to successfully address this issue using deep learning innovations.

2. Related Work

This study presents a streamlined method for achieving facemask detection and specific notifications for non-compliance. Using datasets from Kaggle, the model is trained and evaluated. The system personally alerts users via text messages when it recognizes a facemask is missing from someone's face in real-time operation. Real-time faces are processed by the convolutional neural network (CNN) to extract mask presence, which is subsequently input into the system. An automated system is used to monitor face mask compliance and notify the appropriate authorities when non-compliance is discovered. The system assists the public in assuring correct mask usage to stop the COVID-19 infection from spreading by utilizing Computer Vision and MobileNet. Research also aids law enforcement or authorities by making face mask identification simple and offering visual proof for possible actions (Kumar et al., 2021).

An introduction is made to a facemask detection system that incorporates transfer learning alongside convolutional neural networks (CNNs). The model undergoes comprehensive training, validation, and evaluation employing a dataset consisting of 993 images featuring masked individuals and 1918 images of unmasked individuals, gathered from diverse sources including Kaggle and RMFD datasets. The system's efficacy is demonstrated across both static images and real-time video streams. The model's performance is gauged based on criteria encompassing recall, accuracy, and precision. Notably, the computational efficiency inherent to this architecture enhances its suitability for embedded devices. Prepared for deployment in bustling locations such as malls, airports, and public spaces, this technology contributes significantly to infection control efforts by enforcing safety protocols and identifying instances of non-compliance (Shah, 2021).

Due to citizen resistance and agencies' limited capacity to monitor and enforce compliance, The guidelines for the mandatory use of face masks in public places made by the World Health Organization and other safety regulations encounter challenges. the development of facial mask recognition software for CCTV to make it easier to monitor and enforce this protocol. Such software models have the potential to be used in security applications, particularly in the monitoring of disease transmission. A pre-trained deep convolutional neural network (CNN) with a primary focus on the areas around the eyes and forehead was used as the research methodology. The most likely limit (MPL) was used during classification. The model was trained using two datasets, allowing for the recognition of important facial traits and the use of a decision-making method. The Real-World-Masked-Face Dataset's experimental findings show that recognition is highly successful. This study establishes a proof-of-concept and development basis to stop the spread of COVID-19 by enabling people to confirm their mask usage via a webcam. The study recommends making use of the created software and investigating further the improvement of reliable detectors by deep learning model training particular to specified face-feature or mask-wearing categories (Matthias, 2021).

This model built by using a convolutional neural network and more specifically the MobileNet architecture, shows promise for real-time door control automation and utility in places like temples, retail centers, subway stations, airports, and hospitals. The suggested technique makes face mask detection easier by utilizing a convolutional neural network model. It can adjust to a variety of sites, including malls, subway stations, airports, and public areas. The system uses a variety of facial photos, angles, and poses and is trained using convolutional neural networks based on MobileNet. The dataset consists of 3801 photos that have been classified as "with masks" and "without masks." While the latter contains unmasked faces, hand-covered faces, and instances of inappropriate mask wearing, the former group includes a variety of mask angles, poses, and colors (I. Journal, 2022).

A critical preventive measure against the spread of viruses, particularly in public spaces, involves wearing face masks. Detecting mask adherence presents a significant challenge necessitating effective solutions. To tackle this issue, an automated system employing deep learning (DL) algorithms was proposed to address the spread of infectious diseases. For precise face mask detection in this study, deep convolutional neural network (DCNN) and MobileNetV2-based transfer learning models are used. The evaluation of these models utilizes two distinct datasets. The research leverages a deep learning-based approach for automated face mask detection, assessing the performance of Deep Convolutional Neural Network (DCNN) and MobileNetV2 transfer learning models. The datasets comprise self-collected images (dataset-1) and external sources (dataset-2). Results demonstrate that MobileNetV2outperforms than Deep Convolutional Neural Network (Hussain et al., 2022).

This study focuses on using Convolutional Neural Networks (CNN) to detect household objects in video frames. It employs ResNet50 for object classification and Support Vector Machine (SVM) to train and store object data, aiding in object recognition. While showing promising accuracy in recognizing household items, there's room for improvement in certain categories. Future work aims to enhance recognition of multiple objects simultaneously, potentially expediting household item identification for users (Vatchala, 2022).

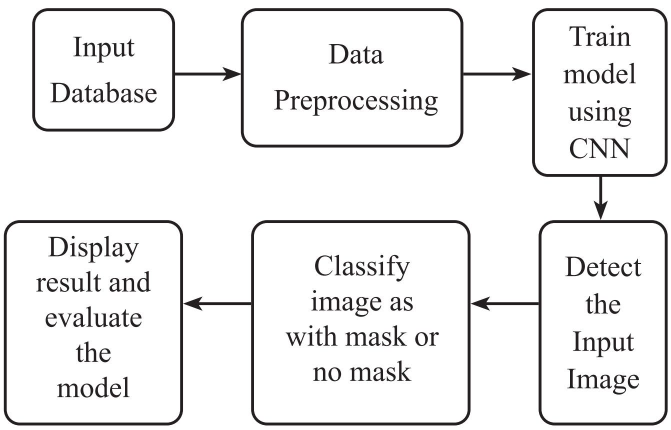
This study explores object detection in video frames, emphasizing household items, using a CNN-ResNet50-SVM workflow. SVM enhances dataset accuracy, with future work aimed at faster multi-object recognition.

Addressing mask-wearing challenges in face recognition, our system combines preprocessing, CNNs, and Caffe models for accurate masked face identification via CCTV or webcam, ensuring efficiency and security (Saranya, 2021).

During the COVID-19 pandemic, face masks have become essential in public areas to curb virus transmission. To monitor mask compliance, Computer Vision through CCTV has been explored. Challenges arise in detecting masked faces from various angles. Past studies utilized Convolutional Neural Networks (CNNs) for high accuracy in front-facing mask detection. This research proposes a CNN approach using facial images from CCTV cameras in public spaces. Initial experiments achieved 97.33% accuracy in single-face mask detection using Retina Face segmentation. However, detecting masks on segmented face regions yielded 82.46% accuracy. The study combined these findings, optimizing for multiple-face mask detection, reaching 79.45% accuracy. Notably, the choice of face detection method and learning rate significantly influenced the mask detection system's accuracy, with the Retina Face model yielding the best results (Sidik and Djamal, 2021).

3. Methodology

This study only uses the Kaggle dataset, “Face Mask Detection Dataset” to demonstrate a Convolutional Neural Network (CNN) model for binary face mask identification. The dataset undergoes essential preprocessing, including resizing images to 128x128 pixels and converting them to the RGB color space, with class labels indicating "With mask" and "Without mask" instances. The designed CNN architecture encompasses two sets of convolutional layers employing 32 and 64 filters, respectively, each followed by rectified linear unit (ReLU) activations and subsequent max-pooling layers for effective down sampling of feature maps. A flattened layer is incorporated to transform the resulting 2D feature maps into a 1D vector, facilitating seamless integration with the subsequent fully connected layers. These dense layers comprise 128 and 64 units, respectively, leveraging ReLU activations to extract intricate features. To enhance model generalization and mitigate overfitting, dropout layers having dropout rate of 0.7 are introduced after each dense layer. The ultimate dense layer, employing a sigmoid activation function, generates probabilities for the binary classification task. For model compilation, the chosen Adam optimizer facilitates weight adjustment during training, while the sparse categorical cross-entropy loss function suits the binary nature of the classification. Model performance is assessed using the accuracy metric, encapsulating the overall effectiveness of the developed approach in face mask detection. This research underscores the pivotal role of meticulous data preprocessing and strategic model design in achieving accurate and reliable outcomes, demonstrating their significance in real-world applications.



3.1 Dataset

The dataset includes 7,553 RGB images grouped into two classes: "with_mask" and "without_mask" based on the presence or absence of masks in the images. It's designed for training a deep learning model to determine if a person wears a mask. Images follow the label_without_mask and the label_with_mask naming scheme, indicating their class which were pre-labeled. The dataset contains 3,725 with mask images and 3,828 without mask images, all in RGB color space with red, green, and blue channels.

3.2 The Proposed CNN Model

The image classification model is constructed using a sequential architecture. The model begins with convolutional layers, initially featuring 32 filters with a 3x3 kernel and ReLU activation, followed by max-pooling with a 2x2 pool size. Subsequently, a 64-filter convolutional layer with ReLU activation and additional max-pooling is applied. The feature maps are then flattened, leading to two dense layers. The first dense layer, containing 128 units with ReLU activation, incorporates dropout with a rate of 0.7 to enhance model robustness. The second dense layer, consisting of 64 units with ReLU activation, similarly employs dropout for generalization. The final dense layer is configured with an activation function tailored for binary classification tasks, which is the sigmoid activation in this case. The model's output corresponds to the desired number of classes, which is two in this instance.

The architecture's construction integrates convolutional and dense layers, culminating in a binary classification model. It effectively extracts features from input images through convolutional operations, followed by down sampling and flattening. Subsequent dense layers with dropout enhance the model's ability to generalize by mitigating overfitting. The final layer, utilizing sigmoid activation, yields class probabilities. With the help of this elaborate architecture, it is possible to accurately identify face masks by classifying photos into two groups: those with masks and those without.

4. Experiment Result and Analysis

The dataset was split into three distinct subsets for comprehensive model assessment. 80% of the data was taken for training, while 20% was reserved for unbiased testing. Additionally, 10% of the training data was allocated for validation, aiding hyperparameter tuning and model performance monitoring. The training dataset consists of a set of facial photos showing people wearing and not wearing masks. The model's architecture, which is highlighted by its carefully thought-out layers, produced remarkable performance. The performance of the mode is tested using various parameters such as precision , recall , accuracy and F1 score. The achieved results highlight the model's robustness and efficiency in recognizing facial features with and without masks. The high precision and recall values (0.91 and 0.95, respectively) signify the model's ability to accurately classify both masked and unmasked instances. The F1 score of approximately 0.93 indicates a well-balanced trade-off between precision and recall, further affirming the model's efficacy. Moreover, the low training and validation losses (12.14% and around 13.7%, respectively) signify that the model effectively captured the essential features within the data while avoiding overfitting. The consistency in performance across training and validation datasets underscores the model's capability to generalize well to new instances.
The model can predict mask or without mask very quickly but can be made more faster by using simpler architectures like MobileNet, pair with hardware like GPUs.

The resolution of the images utilized in the dataset is essential for the performance and accuracy of the face mask detection model. In this research, the images are resized to 128x128 pixels during preprocessing. The impact of image resolution on the results can be significant. Higher-resolution images may contain more detailed information, allowing the model to capture finer features and nuances, potentially leading to improved performance. However, higher resolution also increases computational complexity and training time.

These findings suggest that the designed model can be relied upon for accurate detection of individuals wearing masks in facial images, showcasing its potential for real-world applications in scenarios that require mask detection for safety and security purposes. With a training accuracy of 95.53% and a validation accuracy of 93.58%. In accordance with this, the training procedure produced a remarkably low training loss of 12.14 %, while the validation loss remained noticeably low at roughly 13.7 %. These results demonstrate the model's accuracy in distinguishing between occurrences of masked and unmasked facial features. The graphs in Figures 2 and 3 show the following: the system will distinguish between images with and without face masks as shown below in figure 4.

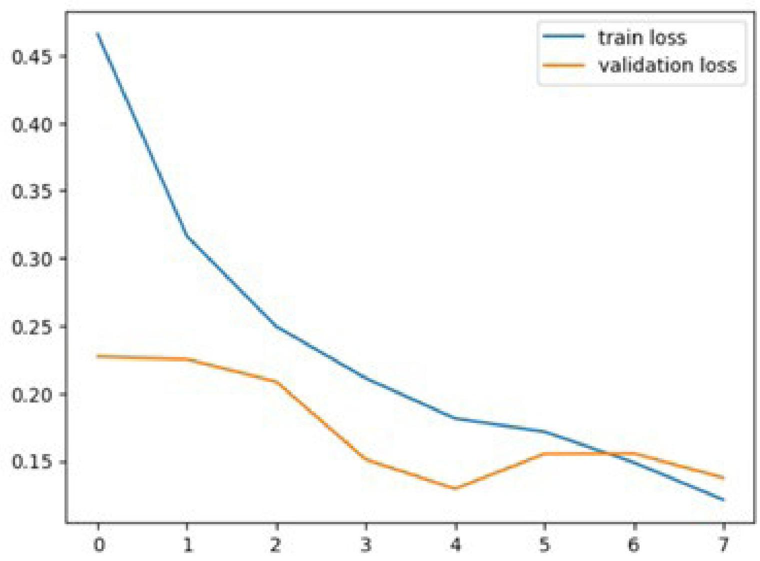


Figure 2: Loss plots

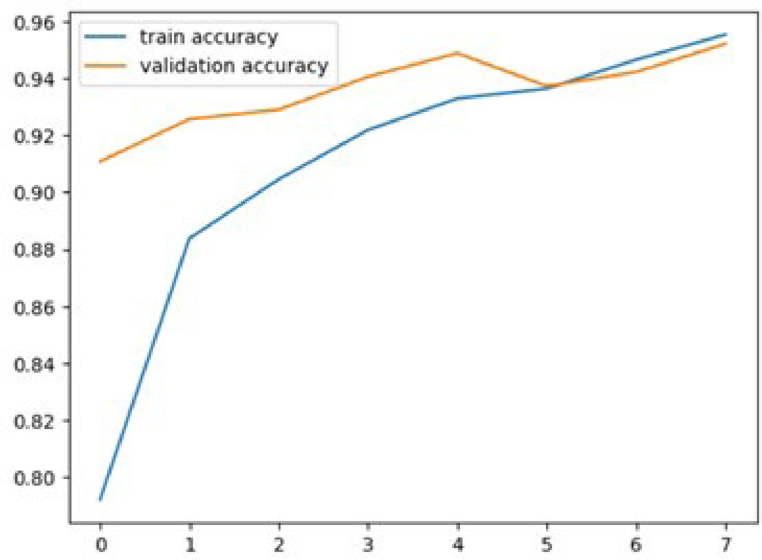


Figure 3: Accuracy plots

Figure 4: Test result of system

4.1 Confusion Matrix

During the model's testing for face mask detection, the confusion matrix revealed distinct performance metrics. The true negative count was 716, indicating accurate classification of instances without masks. The true positive count reached 698, demonstrating precise identification of masked faces. However, 63 instances were incorrectly classified as without masks, resulting in false negatives, while 34 instances were inaccurately categorized as masked, leading to false positive as shown in figure 5. Various performance metrices are calculated using formula given as in Eq. (1, 2, 3, 4)

Accuracy=(TP+TN)/(TP+TN+FP+FN) 1)

Precision=TP / (TP+FP) (2)

Recall=TP / (TP+FN) (3)

F1 score = 2 × ((precision × recall) / (precision + recall)) (4)

where TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

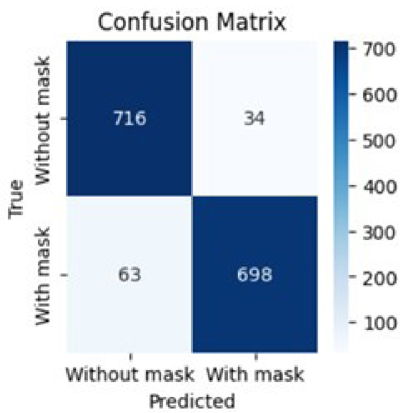


Figure 5: Confusion matrix

5. Conclusions

This research explored automated face mask detection through Convolutional Neural Networks (CNNs), addressing the critical need for monitoring mask compliance in public spaces. Leveraging a custom CNN architecture and a curated dataset, the research demonstrated the potential of deep learning methods in this application. Hyperparameter tuning increased the model's speed in detecting mask presence or absence. By aligning technology with safety and public health concerns, the research underscored the significance of accurate mask detection as industries adopted automation and AI to address evolving societal needs. The findings offered practical implications for safety protocols, security systems, and public health measures, showcasing the valuable role of CNNs in face mask detection across diverse industries.

6. Recommendations

Future work in automated face mask detection via CNNs involves augmenting datasets with diverse real-world scenarios, optimizing models through architecture exploration, and developing real-time applications for public spaces and transportation. Adapting models to detect various mask types, addressing privacy concerns, integrating with public health systems, and exploring multi-class classification were identified as key areas.

References

Kumar, S., P. M. B, and B. S, (Jan. 2021). Face Mask Detection by using Optimistic Convolutional Neural Network, *IEEE*, doi: 10.1109/icict50816.2021.9358653.

Hussain, D., Ismail, M. Hussain, I. Alroobaea, R. Hussain, S. and Ullah, S. S. (May 2022). Face mask detection using deep convolutional neural network and MobileNetV2-Based transfer learning, *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–10, doi: 10.1155/2022/1536318.

I. Journal, (Jan. 2022). Face Mask Detection using CNN and OpenCV, *Irjet*, [Online]. Available: https://www.academia.edu/77042791/Face_Mask_Detection_using_CNN_and_OpenCV.

Matthias, D., Managwu, C. and Olumide, O. (Mar. 2021). Face mask detection application and dataset, *Journal of Computer Science and Its Application*, vol. 27, no. 2, doi: 10.4314/jcsia.v27i2.5.

Saranya, G. Sarkar, D. Ghosh, S. Basu, L. Kumaran, K. and Ananthi, N. (Jun. 2021). Face Mask Detection using CNN, *IEEE*, doi: 10.1109/csnt51715.2021.9509556.

Shah, R. S. R. (Jun. 2021). Face mask detection using convolution neural network, *ResearchGate*, [Online]. Available: https://www.researchgate.net/publication/352308477_Face_mask_detection_using_convolution_neural_network.

Sidik, R. P. and Djamal, E. C. (Sep. 2021). Face Mask Detection using Convolutional Neural Network, *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, doi: 10.1109/ic2ie53219.2021.9649065.

Vatchala, S. Sasidevi, S. Dhanalakshmi, R. and SaiRamesh, L. (2022). Smart household object detection using CNN, in *Advances in parallel computing*, doi: 10.3233/apc220065.