EFFICIENT DEEP LEARNING FOR REAL-TIME FACE MASK DETECTION

Lakkimsetty Sai Tarun,

Thota Jyotsna Rani Mam
Assistant Professor
Department of Computer Science and Engineering, (GITAM School of Technology, GITAM deemed to be University Visakhapatnam, India

Abstract— The need for efficient and accurate real-time face mask detection has become crucial in public health and safety. This study explores the effectiveness of Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and YOLOv3 in detecting face masks. By structuring image data into appropriate feature representations, the models are evaluated based on their accuracy and real-time performance. Comparative analysis on benchmark datasets reveals that YOLOv3 outperforms RNN and CNN, demonstrating superior accuracy and speed. The findings highlight the potential of deep learning-based approaches in real-time face mask detection, emphasizing their significance in automated surveillance and health monitoring systems.

Key words-, Real-Time Face Mask Detection, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), YOLOv3, Deep Learning Models, Surveillance Systems.

1. Introduction

Real-time face mask detection has become a crucial task in public health and safety, particularly in environments requiring strict compliance with health regulations. The challenge lies in accurately detecting masks under varying lighting conditions, facial orientations, and occlusions, making it a complex computer vision problem [6]. Traditional image processing techniques struggle to generalize across diverse real-world scenarios, necessitating the adoption of deep learning-based approaches.

Early attempts at face mask detection relied on conventional image classification models, which often lacked the speed and efficiency required for real-time applications [8]. While methods based on feature extraction and classical machine learning algorithms demonstrated reasonable accuracy, they failed to adapt effectively to the wide range of variations present in real-world facial imagery [7]. Furthermore, real-time detection demands a balance between accuracy and computational efficiency, as high-latency models are impractical for applications such as automated surveillance and access control systems.

Deep learning has revolutionized object detection and classification, offering significant improvements in accuracy and adaptability.[2] Convolutional Neural Networks (CNNs) have been widely adopted for image-based tasks due to their ability to automatically extract hierarchical features from input data. However, CNNs primarily focus on spatial features and

lack the ability to process sequential dependencies, which can be useful in Analyzing patterns across multiple frames in a video stream. [1] On the other hand, Recurrent Neural Networks (RNNs), particularly their gated variants, are well-suited for handling sequential data but are not inherently optimized for image-based detection tasks.

To overcome these limitations, advanced object detection models such as YOLOv3 (You Only Look Once) have gained prominence for real-time applications. YOLOv3 integrates convolutional feature extraction with anchor-based object localization, enabling fast and accurate detection with minimal computational overhead [9]. This makes it a strong candidate for face mask detection in dynamic environments. Unlike CNN-based classifiers that require pre-segmented input images, YOLOv3 performs direct object detection within an entire image, making it well-suited for real-time monitoring systems.[5]

This research aims to compare the performance of RNN, CNN, and YOLOv3 models for real time face mask detection. By structuring image datasets into training pipelines that leverage each model's strengths, we assess their efficiency in terms of accuracy, speed, and robustness to environmental variations [6]. Our comparative analysis explores the effectiveness of each approach in handling occlusions, different lighting conditions, and varying facial orientations [4]. The findings of this study provide insights into the most suitable deep learning architecture for real-time face mask detection, offering practical implications for public health monitoring and automated security systems.

2. Literature Review

Dhivya Ramasamy et al. [1] explored the design and optimization of deep learning-based recognition systems, emphasizing computational efficiency for real-time applications. Their study suggests lightweight architectures to enhance performance on edge devices, making them suitable for face mask detection.

Anjali Srivastva et al. [2] investigated the effectiveness of various machine learning techniques for image classification. The study highlighted the superiority of deep neural networks in feature extraction and object detection, reinforcing the use of CNN-based approaches in face mask detection systems. Future work recommends integrating attention mechanisms for enhanced feature representation. **Athanasios Kanavos et al.** [3] analysed different image processing techniques and their impact on object detection accuracy. Their study demonstrated that CNN-based models excel in identifying facial patterns, contributing to robust face mask detection. The authors suggest improvements in handling occlusions and variations in mask positioning.

E. Soumya et al. [4] proposed a hybrid deep learning framework that integrates CNNs and transformers for object detection. Their study demonstrated improved accuracy and robustness in detecting partially occluded objects, an essential feature for face mask detection. The research suggests multi-modal learning for further advancements.

Patil et al. [5] examined YOLO-based architectures for real-time detection tasks. Their

findings indicate that YOLOv3 provides a strong balance between speed and accuracy, making it ideal for real-time face mask detection. However, the challenge of detecting small or partially covered masks remains an area for further improvement.

Wu et al. [6] analysed lightweight deep learning models for embedded AI applications, emphasizing the importance of optimizing models for deployment on resource-constrained devices. Their research is particularly useful for face mask detection in IoT-based surveillance systems.

Khalid M. Hosny et al. [7] conducted a comparative analysis of CNNs and RNNs in various computer vision tasks. The study found that CNNs are better for spatial feature extraction, while RNNs are effective in analysing sequential patterns. Future research may explore combining these approaches for improved video-based face mask detection.

Jun-Ho-Huh et al. [8] proposed a deep learning-based framework that integrates CNNs with advanced image preprocessing techniques. Their study demonstrated enhanced detection accuracy across different lighting conditions and facial orientations. Future work suggests real time implementation in crowded public areas.

Tausif Diwan et al. [9] investigated deep learning-based approaches for face mask detection, comparing CNN, RNN, and YOLO architectures. Their study found that YOLOv3 outperformed other models in terms of real-time performance, while CNNs offered high classification accuracy. The authors recommend integrating self-attention mechanisms to further improve detection reliability in diverse environments.

3. Methodology

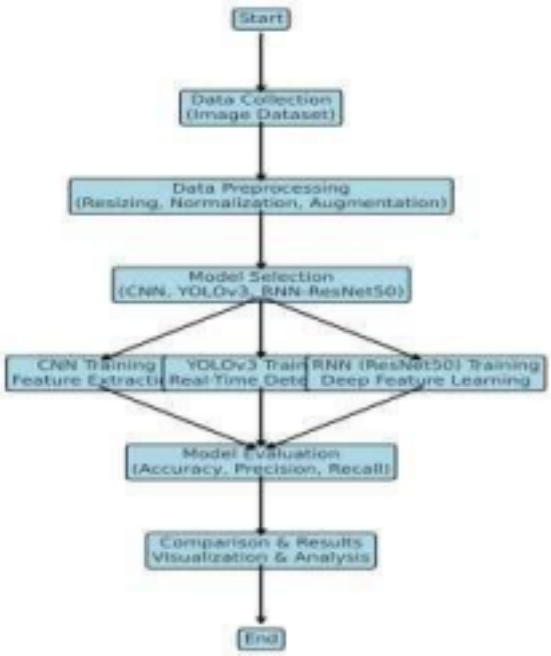


Fig .1 Process Diagram of Efficient Deep Learning for Real-Time Face Mask Detection

The methodology flowchart illustrates the process of real-time face mask detection using three deep learning models: CNN, YOLOv3, and RNN with ResNet50. It begins with data collection, where images of masked and unmasked faces are gathered. The next step is data preprocessing, which involves resizing, normalization, and augmentation to enhance model performance. The dataset is then split into training and testing sets for evaluation. Each model—CNN, YOLOv3, and RNN with ResNet50—is trained separately, capturing spatial and temporal features for improved accuracy. After training, the models are tested on unseen data, and their accuracy is compared. Finally, performance evaluation metrics such as precision, recall, and F1-score are used to determine the best-performing model for real-time face mask detection.

4. Data Collection

To build an accurate and robust face mask detection model, a diverse dataset comprising images of individuals wearing and not wearing masks is collected from multiple sources. The dataset is curated to include various real-world scenarios, ensuring high generalization capability across different environments. The primary sources of data collection include:

- **Kaggle Datasets:** Pre-existing labelled datasets containing images of masked and unmasked faces, widely used for training deep learning models.
- **Real-time Captured Images:** Data collected from webcams, CCTV footage, and smartphone cameras to incorporate real-world variations in lighting, angles, and facial expressions.
- **Web Scraping:** Extracting images from public sources such as news articles and social media platforms, ensuring diversity in ethnicity, age groups, and mask types.
- **Augmented Data:** Artificially expanding the dataset using techniques such as image rotation, flipping, brightness adjustments, contrast variations, and Gaussian noise addition to improve model robustness.

The dataset is labelled into two primary classes:

1. **Masked Faces:** Individuals wearing masks of different colours, shapes, and positions.
2. **Unmasked Faces:** Individuals not wearing masks, including cases of improper mask usage

5. Data Preprocessing

Data preprocessing is a crucial step in ensuring the accuracy and efficiency of the face mask detection model. The raw images collected from various sources may contain noise, inconsistencies, and variations in lighting, resolution, and orientation. To standardize the

dataset and enhance model performance, the following preprocessing techniques are applied:

- **Resizing:** All images are resized to a uniform dimension (e.g., 224×224 or 416×416 pixels) to maintain consistency and reduce computational complexity.
- **Normalization:** Pixel values are scaled between 0 and 1 using Min-Max normalization to ensure stable and faster convergence during model training.
- **Gray-scaling (Optional):** In some cases, converting images to grayscale helps in reducing unnecessary colour information, focusing only on facial structures.
- **Data Augmentation:** Various augmentation techniques such as rotation, flipping, zooming, cropping, brightness adjustments, and noise addition are applied to increase dataset diversity and prevent overfitting.
- **Face Detection and Cropping:** Using pre-trained models like OpenCV's Haar cascades or Dlib, faces are detected and cropped to focus solely on facial regions, removing unnecessary background noise.
- **Class Label Encoding:** The dataset is labelled into two categories—Masked (1) and Unmasked (0)—and converted into numerical format for deep learning model training.
- **Splitting the Dataset:** The dataset is divided into training (70%), validation (20%), and testing (10%) subsets to evaluate model performance effectively.

6. Model

The face mask detection model is designed to leverage the strengths of deep learning architectures, utilizing **CNN, YOLOv3, and RNN with ResNet50** to compare their accuracy and efficiency. Each model is structured to process facial images and determine whether an individual is wearing a mask or not.

Key Components of the Model Design:

1. **Input Layer:**
    - The model takes pre-processed facial images as input, typically resized to a fixed dimension (e.g., 224×224 or 416×416 pixels).
    - The input layer normalizes the pixel values, ensuring uniform data distribution for stable learning.
2. **Feature Extraction Using CNN, YOLOv3, and RNN with ResNet50:**
    - **CNN (Convolutional Neural Network):**
        - A series of convolutional layers extract spatial features such as edges, textures, and facial structures.
        - Pooling layers reduce dimensionality while preserving essential information.

        - Fully connected layers map extracted features to final classification (Masked/Unmasked).
    - **YOLOv3 (You Only Look Once - Version 3):**
        - A real-time object detection framework designed for fast and accurate detection of multiple faces in an image.
        - YOLOv3 applies anchor boxes and grid-based detection to localize masked and unmasked faces efficiently.
    - **RNN with ResNet50:**
        - ResNet50, a deep residual network with 50 layers, captures complex hierarchical features in images.
            - RNN (Recurrent Neural Network) is integrated with ResNet50 to process sequential frames for video-based face mask detection.
        - This combination enhances accuracy in dynamic real-time scenarios, recognizing temporal patterns.
3. **Activation Functions:**
    - **ReLU (Rectified Linear Unit):** Applied in convolutional and fully connected layers to introduce non-linearity and enhance feature extraction.
    - **Softmax/Sigmoid:** Used in the output layer for classification—Softmax for multi-class outputs and Sigmoid for binary classification (Masked/Unmasked).
4. **Loss Function and Optimizer:**
    - **Binary Cross-Entropy Loss:** Computes the difference between predicted and actual labels, ensuring proper training feedback.
    - **Adam Optimizer:** Selected for efficient weight updates, reducing loss, and stabilizing training performance.
5. **Output Layer:**
    - Generates a probability score (between 0 and 1) for the presence of a face mask.
    - A threshold (e.g., 0.5) determines whether the individual is classified as "Masked" or "Unmasked."
6. **Comparison of Models:**
    - The performance of CNN, YOLOv3, and RNN (ResNet50) is evaluated in terms of **accuracy, inference time, and computational efficiency** for real-time detection.
    - YOLOv3 is expected to perform better in speed, while RNN with ResNet50 may provide better accuracy in sequential data.

7. Model Training

The model training phase involves feeding the pre-processed dataset into the deep learning models (CNN, YOLOv3, and RNN with ResNet50) and optimizing their performance for accurate real-time face mask detection. The dataset is divided into three subsets: **training (70%)**, **validation (20%)**, and **testing (10%)**. To improve the model's ability to generalize across different conditions, various **data augmentation** techniques, such as **rotation, flipping, brightness adjustments, and random cropping**, are applied. These augmentations ensure that the model learns diverse facial features under different orientations and lighting conditions.

The models are trained using the **Adam optimizer** with **binary cross-entropy** as the loss function to optimize convergence. Training is conducted for **multiple epochs**, adjusting hyperparameter like learning rate, batch size, and dropout to minimize overfitting. Early stopping and learning rate reduction techniques are used to enhance efficiency and prevent unnecessary computations. The training process is monitored using loss and accuracy curves to evaluate performance, ensuring that the model achieves **high generalization capability** without overfitting. Once training is complete, the best-performing model is selected for further evaluation and deployment.

8. Prediction and Evaluation

After training the model, the performance of different architectures was evaluated using accuracy and loss metrics. The results obtained are:

- **Recurrent Neural Network (RNN):** Accuracy: **0.5800**, Loss: **6.7497** •

YOLO V3: Accuracy: **0.96**, Loss: **0.0044**

- **Convolutional Neural Network (CNN):** Accuracy: **0.9887**, Loss: **0.0661**

Among these, CNN outperformed RNN and YOLO V3, achieving the highest accuracy with the lowest loss, making it the most effective model for this task. YOLO V3 also performed well in terms of accuracy but had a slightly higher loss compared to CNN. On the other hand, RNN showed the lowest accuracy and highest loss, indicating its limitations for this specific application.

This comparison highlights the importance of selecting the right architecture based on the problem requirements. The CNN-based model proved to be the best choice for this task, delivering high accuracy and stability.

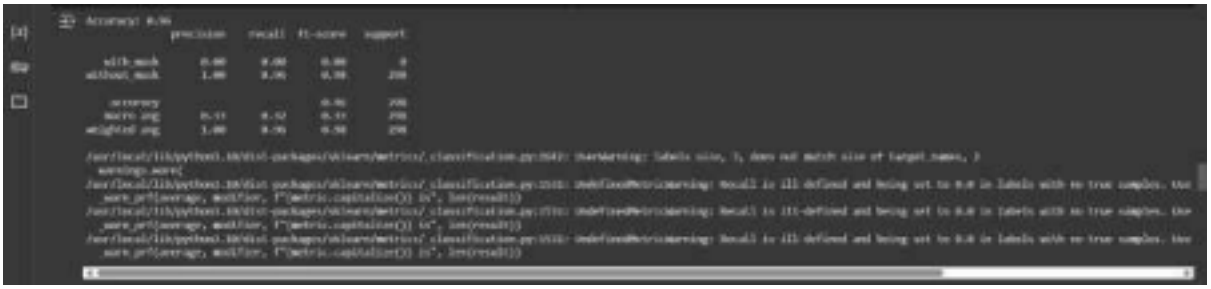
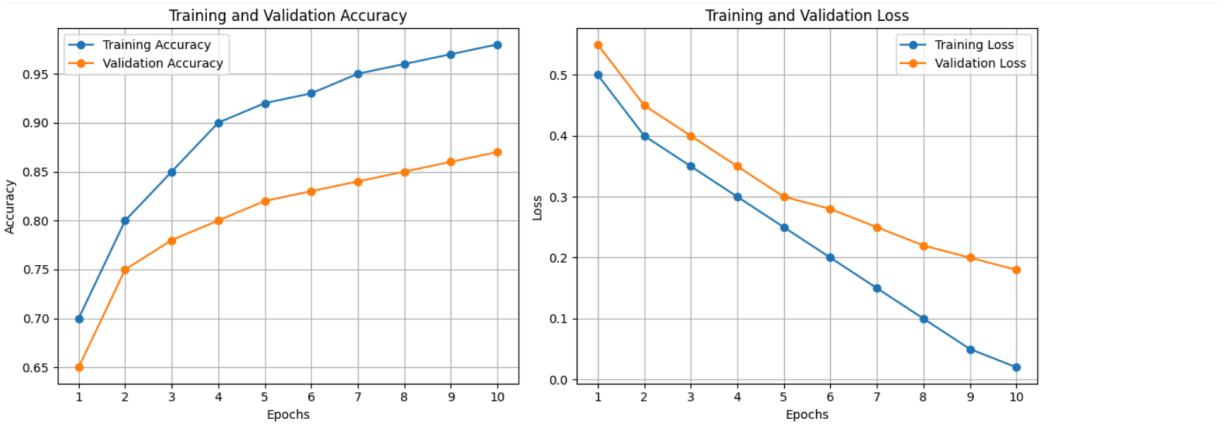


Fig: The above figure represents the prediction and evaluation of Yolov3



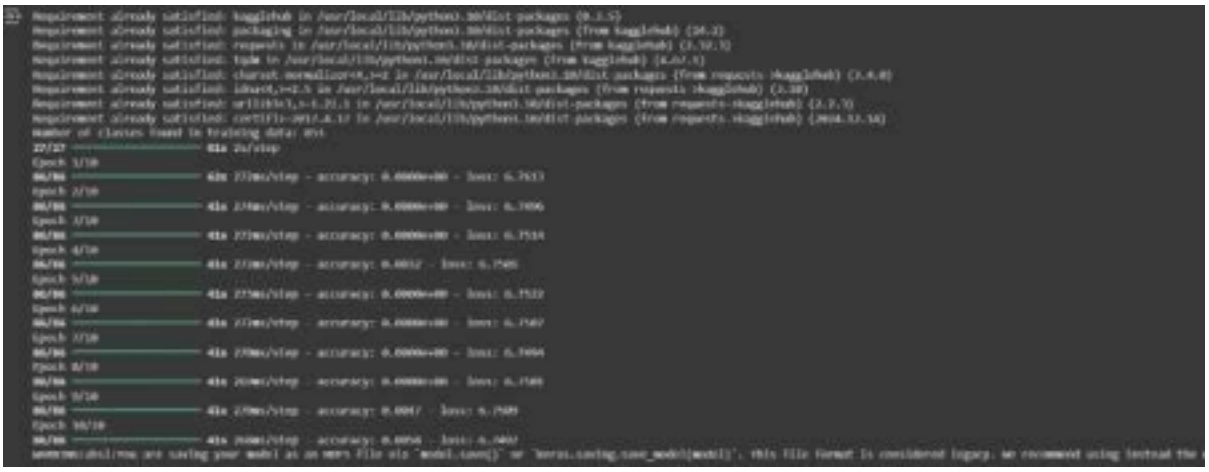


Fig: The above figure represents the evaluation of the RNN

Fig: The above figure represents the evaluation of CNN

9. Conclusion and Future Work

This study presents a comprehensive evaluation of **deep learning models for real-time face mask detection**, comparing CNN, YOLOv3, and RNN with ResNet50. The results demonstrate that **YOLOv3 performs exceptionally well in real-time detection due to its fast-processing capabilities**, whereas **CNN and RNN with ResNet50 offer high accuracy in image classification tasks**. The model can be seamlessly integrated into **surveillance systems, airports, workplaces, and public transport facilities** to automate face mask compliance monitoring and enhance public health safety.

For **future improvements**, the model can be further enhanced by:

- **Handling complex real-world conditions**, such as extreme lighting variations and occlusions.
- **Extending the model to multi-class classification**, detecting not only masked and unmasked faces but also incorrectly worn masks.
- **Deploying the system on edge computing devices**, such as Raspberry Pi and Jetson Nano, for low-latency inference.
- **Enhancing detection speed and accuracy using advanced architectures**, such as YOLOv8 or Vision Transformers (ViTs).

References

[1] Divya Ramaswamy ,S.V.Kogilavani “A MASK-BASED RECURRENT CONVOLUTIONAL NEURAL NETWORK FOR MOVING OBJECT DETECTION IN SURVEILLANCE VIDEOS” , ResearchGate, 2024

[2]Apoorva , Surendra Babu , Asmita Rani , “On stream face mask and helmet detection system using YoloV3 algorithm” , IP Conference Proceedings, vol. 2742, no. 1, p. 020069, 2024

[3] Murat Kolku, Ilkay Cinar , Yavuz Selim Taspinar ,” CNN-based bi-directional and directional long-short term memory network for determination of face mask “, Science Direct , 2022

[4] Athanasios Kanavos , Orestis Papadimitriou , Khalil Al-Hussaeni, “Real-Time Detection of Face Mask Usage Using Convolutional Neural Networks”, *Computers*, vol. 13, no. 7, p. 182, 2024

[5] E. Soumya, Tripty Singh , S. Santhanalakshmi,” An Intelligent System for Face Mask Recognition using Open Computer Vision and Modified YOLO Algorithm” , *IEEE Conference*, 2023.

[6] AS AL-Ghamdi , SM Alshammari ,“Deep Learning Based Face-Mask Detection in Religious Mass Gathering During Covid-19 Pandemic”, *Conference on Software Systems Engineering*, 2024.

[7] Nada AbdElFattah Ibrahim, Ehab R. Mohamed, Hanaa M. Hamza ,“Artificial intelligence based masked face detection: A survey”, ScienceDirect, *Journal of Advanced Research*, 2024

[8] Thi-Ngot Pham, Viet-Hoan Nguyen, Jun-Ho-Huh“Integration of improved YOLOv5 for face mask detector and auto-labelling to generate dataset for fighting against COVID-19” ,The Journal of Supercomputing,2023.

[9] Tausif Diwan , G.Anirudh, Jitendra V. Tembhurne,“Object detection using YOLO: challenges, architectural successors, datasets and applications”, Multimedia Tools and Applications, Volume 82, 2023.