Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/45/20241691

A pneumonia detection system based on MobileNetv2 network and model callback

Fengwei Liu

Glasgow College University of Electronic Science and Technology
Chengdu, China

2021190903021@std.uestc.edu.cn

Abstract. This study focuses on automatically classifying radiographic images of the chest region into standard classification, COVID-19 classification, and viral pneumonia classification by utilizing complex neural networks. Using a chest radiograph dataset from the academic bastion Université de Montréal, the study introduces an innovative paradigm rooted in MobileNetV2. We conducted a comparative analysis to evaluate the efficacy of this avant-garde model by juxtaposing it with the typical DenseNet121 and RESNET50 popular in the field of medical image classification. This exploration revealed MobileNetV2 as an ingenious model distinguished by its tiny scale and commendable accuracy. Using DW convolution design greatly reduces the computational complexity and parameter count. Regarding the composition of the architecture, transfer learning is used to attach global mean pooling and fully connected layers on top of MobileNetV2, and is customized for nuanced tripartite classification work. Comparative evaluation shows that the MobileNetV2 model has only 2,643,187 parameters, completes training in just 37 seconds, and has an accuracy of 98.48%. In contrast, although DenseNet121 and RESNET50 demonstrate commendable proficiency, their large model dimensions and lengthy training intervals limit their usefulness in resource-limited environments. The findings highlight the superior performance of the MobileNetV2 model in the field of chest X-ray classification, providing a simplified and efficient alternative for deployment on mobile and embedded devices.

Keywords: MobileNetV2, DenseNet121, RESNET50, chest imaging radiological classification, viral pneumonia.

1. Introduction

The field of medical imaging has been pursuing more effective and accurate image classification methods. Established models such as DenseNet121 and RESNET50 have won praise for their deep network architecture and exemplary performance [1]. However, given the rapidly evolving healthcare data landscape and the need for real-time processing, there is an urgent need for solutions with enhanced flexibility and efficiency. This research focuses on improving the accuracy of chest X-ray image classification to address contemporary challenges in medical imaging. This includes traditional models such as DenseNet121 and RESNET50, which have produced noteworthy research results through the integration of dense connections and residual blocks, mitigating challenges such as vanishing and exploding gradients. However, these models face obstacles in terms of model dimensionality and training duration, especially when resources are limited.

In order to overcome certain limitations inherent in traditional models, this study introduces an avant-garde deep learning model based on MobileNetV2. MobileNetV2 stands out for its petite model size and commendable accuracy, leveraging a DW convolutional design to significantly reduce computational complexity and parameter quantization [2]. This enables MobileNetV2 to provide a lighter weight and more efficient alternative for chest X-ray image classification tasks while maintaining higher efficiency. In addition, in order to improve model efficiency, strategies such as data enhancement and model callback are also combined. Data augmentation enhances the model's generalization ability by deforming and expanding the training data set. In contrast, model callback enhances stability and accelerates convergence by dynamically adjusting the learning rate and retaining the best-fitting model.

From a review of the test set, we derived metrics encompassing the loss and accuracy of each model and analyzed model performance for various disease categories through the medium of confusion matrices and classification reports. Empirical results highlight that the pioneering approach proposed in this study outperforms traditional models in terms of classification accuracy, has fewer parameters, and requires less training time investment compared to traditional models such as DenseNet121 and RESNET50 [3]. This means that the seminal contribution of this study leaves behind a more flexible and efficient image classification model, bringing a new perspective to the automatic classification of chest X-ray images.

The structure of the research is organized as follows: After the 'Introduction', the 'Related Work' section reviews existing literature. 'Methods and Experiments' section details our MobileNetV2-based model and data augmentation strategies. 'Results' analyzes the performance of the models, and the 'Conclusion' discusses our findings and future perspectives in automatic chest X-ray image classification.

2. Related Work

In the vast field of medical imaging, DenseNet121 and RESNET50 have become typical deep learning prototypes. DenseNet121 was born due to the joint efforts of [4], who proposed the DenseNet model in their seminal 2017 paper "Densely Connected Convolutional Networks". In response to the dilemma of gradient dissipation in deep neural networks, DenseNet121 adopts dense connection mode to strengthen gradient propagation and improve training efficiency. In contrast, RESNET50 (proposed by Kaiming He, Xiangyu Zhuang, Shaoqing Ren, and Jian Sun [5]) left a deep impression in their 2015 book "Deep Residual Learning for Image Recognition". RESNET50 introduces the concept of residual blocks, which cleverly alleviates the dilemma of rapid gradient growth through skip connections. The depth of its architecture helps extract higher-level and more abstract features, thereby enhancing the model's ability to discern disease nuances.

Tavishee Chauhan, Hemant Palivela, Sarveshmani Tiwari and team performed image classification on chest X-ray images, orchestrating a meticulous comparative analysis of multiple optimizers, learning rate schedulers and loss functions to achieve the highest accuracy, which demonstrated A commendable success in the field of medicine. their proposed system [6]. In a parallel effort, Qasem Abu Al-Haija and Adeola Adebanjo achieved remarkable results using ResNet-50 convolutional neural networks to diagnose breast cancer in histopathology images [7]. These researchers not only enhanced and used DenseNet121 and RESNET50 in the context of medical image classification, but also made valuable contributions that promoted exploration and innovation to improve classification accuracy and address real-time operational requirements. Their combined efforts have synergistically advanced the development of DenseNet121 and RESNET50, advancing the field of medical image classification.

Although these works have shown commendable capabilities in medical image classification, traditional models still face challenges in model scaling and training duration, especially as medical data scale rapidly grows and real-time operation requirements intensify. Therefore, contemporary research focuses on improving classification accuracy while reducing model complexity and training duration. Researchers are increasingly recognizing the need to use lightweight models in large-scale data sets and applications that require improved real-time performance. Therefore, innovative initiatives strive to introduce deep learning paradigms that are thriving in different fields with the aim of achieving superior performance in medical image classification tasks.

Against this background, this study advocates the proposition of utilizing MobileNetV2 as a base model. MobileNetV2 stands out for its smaller model dimensions and commendable accuracy, which is achieved by leveraging depth wise separable convolutions, significantly reducing computational complexity and parameter count. This architectural design enables MobileNetV2 to provide more lightweight and efficient resources for chest X-ray image classification tasks while maintaining higher operational efficiency.

3. Methods and Experiments

3.1. X-ray Chest Image Dataset and Data Augmentation

X-ray images are the key medical basis for identifying the pathological status of a patient's lungs. The X-ray chest images used in this study came from a research consortium affiliated with the University of Montreal [8]. The dataset contains a total of 317 X-ray chest images, divided into a training set for calibrating model parameters and a test set for evaluating model effectiveness. In the training set, 251 images are carefully labeled and enumerated, while the corresponding test set encapsulates 66 images, following the same numerical nomenclature. Initially, the dataset classifies images into three different label types (1 "normal" and 2 disease): Normal, COVID-19, and Viral Pneumonia, with each image assigned a unique label. Chest X-ray image data is provided in jpg and jpeg formats, with image sizes up to 4280*3480 pixels.

The images were classified according to the type of pneumonia in the training set, with 70 images designated as "normal," 111 images designated as "Covid," and 70 images designated as viral pneumonia. In the test set, the distribution included 20 “normal” images, 26 “Covid” images, and 20 images designated as viral pneumonia. Images in the training set are randomly ordered and numerically named within each category. As an illustration, Figure 1 shows randomly selected X-ray chest images used for CNN model training, along with their respective labels and numerical names.

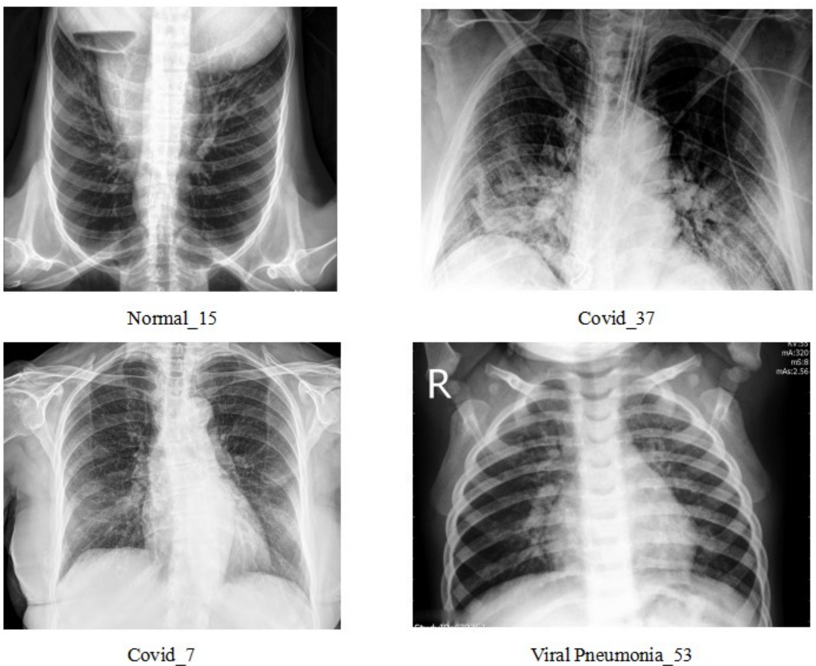

Figure 1. A Haphazard Selection of X-ray Chest Images.

To improve the model's resilience and generalization capabilities in medical image classification tasks, a series of image enhancement strategies are employed. These strategies aim to introduce variability in the training data, allowing the model to learn from a more diverse set of images. Here's an explanation of each strategy:

1. Random Horizontal Flip: This strategy involves randomly flipping images along the horizontal axis. By doing so, the model is exposed to diverse perspectives, which helps in acquiring a more heterogeneous set of features from the data. Enhances the dataset's diversity, allowing the model to learn from images that are mirrored, thus simulating different viewing conditions.

2. Rotation: The strategy introduces random rotations to images. This simulates the scenario where X-ray images are captured from varying angles, thus challenging the model to recognize features in different orientations. Increases the resilience of the model by enabling it to assimilate and recognize features from various angles, enriching its generalization capabilities.

3. Zoom: This strategy simulates different image sizes and scaling ratios through random scaling operations. It effectively changes the apparent size and scale of features within the images. Enhances the model's adaptability to different sizes and scales of features, which is crucial in medical imaging where the size of anomalies can vary.

4. Random Changes in Height and Width: This approach involves introducing random variations in the images' height and width. The strategy aims to train the model with images that vary in size but maintain the same aspect ratio. Increases the model's proficiency in handling images of different sizes, ensuring it can accurately interpret features regardless of their dimensional variations.

By integrating these image enhancement operations in the enhancement layer of the sequential model, each training iteration incorporates random transformations. This orchestration of augmented paradigms significantly enhances the diversity of the dataset. It leads to improvements in the robustness and generalization capabilities of the model, which are critical for achieving excellent results in medical image classification tasks.

3.2. MobileNetV2 Model

MobileNetV2 is Google's convolutional neural network carefully designed for mobile devices. Compared with traditional convolutional neural networks, it is a lightweight solution that improves computational efficiency and reduces model footprint. This design concept makes it possible to seamlessly perform high-precision image recognition tasks on mobile devices.

At its core, MobileNetV2 uses depth-separable convolution to reduce the number of parameters and computational requirements in the model architecture. This innovative approach decomposes the traditional convolution operation into two different autonomous operations: depth convolution and point convolution [9]. In this paradigm, depth wise convolutions specialize in convolution within the channel dimension, while pointwise convolutions limit their convolution operations to the spatial dimension. This decomposition promises a significant reduction in computational complexity while maintaining a higher standard of classification accuracy.

MobileNetV2 further enhances its capabilities by integrating a linear bottleneck function to accelerate network training and adopting an inverted residual structure to cleverly exploit low-dimensional feature information to fully realize its potential. The model strategically adopts a lightweight feature network design space approach to optimize performance through careful calibration of the convolution kernel size and adjustment of network width and depth, ultimately resulting in a more streamlined and efficient network [10].

3.3. Deep Convolutional Neural Network Model for Pneumonia Detection

The detailed process of the network model established in this paper is shown in Figure 2. Initially, the input layer of the model is designed to accommodate image data of size IMAGE_SIZE + (3), where IMAGE_SIZE represents the predetermined size of the image and 3 represents the number of image channels (RGB). The input data then undergoes complex processing through previously defined image

DOI: 10.54254/2755-2721/45/20241691

enhancement layers, thereby enhancing the model's ability to absorb insights from large amounts of data variation.

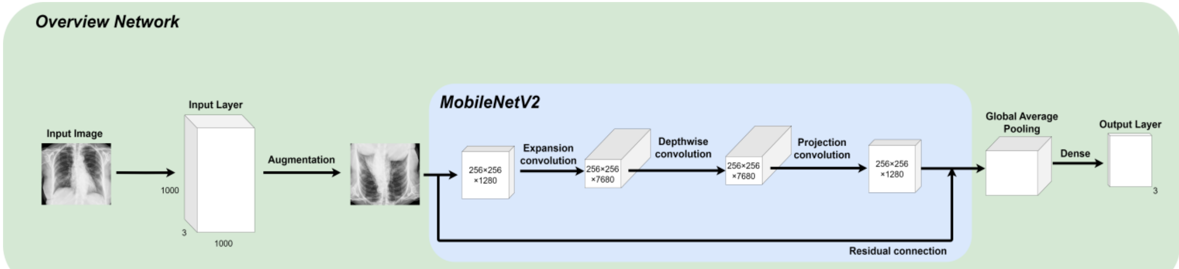

Figure 2. The structure of a network model for pneumonia detection.

After this preparation phase, the MobileNetV2 network will be seamlessly integrated as the model's infrastructure. Utilizing depth wise separable convolutions (DW convolutions), MobileNetV2 significantly alleviates computational complexity and parameter richness.

After adding the MobileNetV2 network layer, efforts are made to reduce parameter diffusion, enhance the generalization ability of the model, and extract global feature information. To achieve these goals, a global average pooling layer is introduced. This layer calculates the average of each channel in the feature map, thereby merging the entire feature map into a single singular value. This transformation process facilitates encapsulation of key information related to the integrated image.

On the basis of the global average convergence layer, a fully connected layer with 300 units is introduced, supplemented by the ReLU activation function. This fully connected layer is cleverly designed to further refine abstract features and enhance the learning efficiency of the model.

To summarize the architectural blueprint, the model’s output layer is carefully delineated with a SoftMax activation function tailored for multi-classification. The unit count in the output layer is consistent with the number of classes in the dataset, ensuring that the model is able to provide a probability distribution for each class. The comprehensive model structure is succinctly encapsulated in Table 1.

Table 1. Neural Network Model Structure for Pneumonia Detection.

Layer	Out Shape	Param
input_layer	(None, 1000, 1000, 3)	0
augmentation_layer	(None, None, 3)	0
mobilenetv2	(None, None, None, 1280)	2257984
global_average_pooling2d	(None, 1280)	0
dense	(None, 300)	384300
output_layer	(None, 3)	903

Total params: 2,643,187
Trainable params: 385,203
Non-trainable params: 2, 257,984

3.4. Model Compilation

The loss function chosen in this study is classification cross entropy, which is a widely recognized and effective function, especially suitable for multi-class classification tasks.

For optimization purposes, choose the Adam optimizer, a hybrid of momentum and stochastic gradient descent algorithms. Adam’s adaptive learning rate adjustment and enhanced convergence capabilities make it ideal for achieving more accurate convergence to the global optimum, thereby improving the overall training efficiency of the model.

DOI: 10.54254/2755-2721/45/20241691

The evaluation metric used in this study is accuracy, focusing on the classification accuracy of the model in the three categories of normal, COVID-19, and viral pneumonia. Accuracy is an intuitive measure that represents the proportion of correct classifications relative to all predictions made by the model.

During the model training process, the callback mechanism plays a key role, enabling specific operations at different intervals during each training stage. These callback functions help dynamically adjust the model's learning rate, retain optimal weights, stop training early, and other operations to optimize overall model performance. These callback functions can monitor indicators such as the accuracy of the validation set, thereby giving flexibility and being able to respond skillfully to different scenarios during the training process, thus improving the training effectiveness of the model.

4. Results

The exemplary performance of the MobileNetV2 model on the training set is indeed noteworthy, characterized by a parsimonious number of parameters of 2,643,187. The model achieved an impressive accuracy of 98.48% in just 37 seconds of training, demonstrating remarkable capabilities in an image classification task spanning three categories. Its commendable classification capabilities are further emphasized by the precision, recall, and F1 scores for each category, as shown in Table 2.

Table 2. MobileNetV2 model classification performance.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	26
1	0.95	1.00	0.98	20
2	1.00	0.95	0.97	20
accuracy			0.98	66
macro avg	0.98	0.98	0.98	66
weighted avg	0.99	0.98	0.98	66

In comparison, the DenseNet121 model with 7,169,091 parameters and a training duration of 404 seconds achieved a slightly lower accuracy of 90.91% on the training set compared to MobileNetV2. Table 3 describes category-specific performance metrics.

Table 3. DenseNet121 model classification performance.

	precision	recall	f1-score	support
0	0.84	1.00	0.91	21
1	1.00	0.77	0.87	13
2	1.00	0.94	0.97	17
accuracy			0.92	51
macro avg	0.95	0.90	0.92	51
weighted avg	0.93	0.92	0.92	51

The ResNet50 model has the highest number of parameters at 23,858,435 and takes only 46 seconds to train. This efficiency is accompanied by an accuracy of 96.97% on the training set, with class-specific performance metrics illustrated in Table 4.

Table 4. ResNet50 model classification performance.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	26
1	0.95	0.95	0.95	20
2	0.95	0.95	0.95	20
accuracy			0.97	66
macro avg	0.97	0.97	0.97	66
weighted avg	0.97	0.97	0.97	66

Through the comparative analysis of MobileNetV2, DenseNet121 and ResNet50, some obvious observations and conclusions can be drawn:

1. Overall accuracy: MobileNetV2 is the leader with an overall accuracy of 98.48%, surpassing DenseNet121 (90.91%) and ResNet50 (96.97%).

2. Training time: MobileNetV2 not only performs well in accuracy, but also shows obvious advantages, with the shortest training time of 37 seconds. In comparison, ResNet50 and DenseNet121 require 46 seconds and 404 seconds respectively. This makes MobileNetV2 an excellent choice for real-time applications on mobile devices and embedded systems.

3. Number of model parameters: MobileNetV2 stands out as a lightweight contender with 2,643,187 parameters, while ResNet50 and DenseNet121 bear a larger burden with 23,858,435 and 7,169,091 parameters respectively. The efficiency of MobileNetV2 in terms of model size and computational resources is evident.

4. Category performance: MobileNetV2 and ResNet50 show commendable precision, recall, and F1 scores in each category, highlighting their strong performance. In contrast, DenseNet121 exhibits slight deficiencies on some categories.

All in all, MobileNetV2 emerged as an excellent choice, demonstrating excellent accuracy, fast training, and efficient model footprint. Its capabilities in real-time applications make it a perfect for medical image classification tasks on mobile and embedded platforms.

5. Conclusion

This study delves into the training and evaluation of three different neural network models (MobileNetV2, DenseNet121, and ResNet50) in the field of chest X-ray image classification. The seminal findings of this study highlight the superior performance of MobileNetV2 in this specific domain. Notably, MobileNetV2 outperforms traditional models, namely DenseNet121 and ResNet50, with significantly improved accuracy while maintaining a lighter model footprint and shorter training duration.

The originality of MobileNetV2 lies in its depth wise separable convolutional architecture, enhanced by a combination of linear bottlenecks and inverse residual connections. This merging not only retains commendable accuracy, but also significantly reduces the number of parameters and computational requirements of the model. Therefore, MobileNetV2 becomes the best solution for medical image classification, especially on resource-constrained mobile devices and embedded systems.

The efficacy of MobileNetV2 is of particular importance in real-time applications and scenarios with limited computing resources. In fields such as mobile medicine, telemedicine, and smart medicine, MobileNetV2 is expected to provide fast and reliable medical image classification services.

Certain limitations of this study must be acknowledged, particularly related to the size and diversity of the training data. Future efforts could enhance model robustness by expanding the size and diversity of the dataset. Additionally, for specialized medical imaging tasks, MobileNetV2 may need to be fine-tuned to meet specific requirements.

This study proposes a practical solution for chest X-ray image classification, introducing a novel, flexible and efficient method. This contribution bodes well for simplifying chest X-ray image analysis, reducing the burden on physicians, and improving the accuracy of medical diagnosis. As we chart our future course, our expectation is to expand our reach to a broader range of medical imaging tasks through continued research and improvements to MobileNetV2.

References

[1] Wu, Bo, et al. "An ameliorated deep dense convolutional neural network for accurate recognition of casting defects in X-ray images." *Knowledge-Based Systems* **226** (2021): 107096. (LinkOut)

[2] Hosseini, Azamossadat, et al. "A mobile application based on efficient lightweight CNN model for classification of B-ALL cancer from non-cancerous cells: a design and implementation study." *Informatics in Medicine Unlocked* **39** (2023): 101244. (LinkOut)

[3] Hazarika, Ruhul Amin, Debdatta Kandar, and Arnab Kumar Maji. "An experimental analysis of different deep learning-based models for Alzheimer’s disease classification using brain magnetic resonance images." *Journal of King Saud University-Computer and Information Sciences* **34.10** (2022): 8576-8598. (LinkOut)

[4] Taneja, Soham, et al. "Histopathological Cancer Detection Using CNN." *Cancer Prediction for Industrial IoT 4.0: A Machine Learning Perspective* (2021): 9781003185604-2. (LinkOut)

[5] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. (LinkOut)

[6] Chauhan, Tavishee, Hemant Palivela, and Sarveshmani Tiwari. "Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging." *International Journal of Information Management Data Insights* **1.2** (2021): 100020. (LinkOut)

[7] Al-Haija, Qasem Abu, and Adeola Adebanjo. "Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network." 2*020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. 2020. (LinkOut)

[8] N. I. of Health Clinical Center, Nih chest x-ray dataset of 14 common thorax disease categories. (LinkOut)

[9] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. (LinkOut)

[10] Xiang, Qian, et al. "Fruit image classification based on Mobilenetv2 with transfer learning technique." *Proceedings of the 3rd international conference on computer science and application engineering*. 2019. (LinkOut)