흉부 엑스선 결핵 검출을 위한 EfficientNets 비교

아흐마드 이자즈, 신석주*
컴퓨터공학과, 조선대학교

ahmadijaz@chosun.kr, *sjshin@chosun.ac.kr(교신저자)

A Comparison of EfficientNets for Tuberculosis Detection in Chest Radiographs

Ijaz Ahmad, Seokjoo Shin
Dept. of Computer Engineering, Chosun University

Abstract

Deep learning (DL) models keep improving for natural image classification not only in terms of model performance but also in their number of parameters, floating-point operations (FLOPs), training and inference speed. Consequently, the applications of such models are extended to other computer vision (CV) related fields such as medical image analysis domain. Among the existing DL models, a new model of EfficientNet family namely EfficientNetV2 provides parameter and FLOPs efficiency along with training speed on natural images. The present study analyzes EfficientNetV2 suitability for Tuberculosis (TB) detection in Shenzhen chest X-ray (CXR) image dataset and compares it with EfficientNetV1 model. The analysis have shown that the EfficientNetV2 delivers the same sensitivity score as that of EfficientNetV1 model for the same number of parameters with better training speed.

Ⅰ. Introduction

Tuberculosis (TB) remains one of the leading cause of death from a single infectious agent according to the World Health Organization (WHO) [1]. The disease spread and mortality is preventable by its early detection [1], [2]. The WHO organization recommends chest X-ray (CXR) images for screening pulmonary abnormalities due to its wide availability and relative low cost [2]. With the recent success of Machine Learning (ML) in the field of computer vision, automatic computer aided diagnosis (CAD) systems have emerged to assist doctors and practitioners. Particularly, DL, a subfield of ML, has achieved state-of-the-art performance for image classification [3].

Several DL models have been proposed for diagnosis of TB [4]. Among them EfficientNet models provide parameter and FLOPs efficiency with better training and inference speed. For example, EfficientNetV1 [5] based model has achieved 89.92% accuracy when the images were enhanced with unsharped masking in [6]. Their accuracy can be improved to 94.25% with transfer learning (TL) as in [2]. For the most recent EfficientNetV2 [7], accuracy of 89.52% can be achieved without TL and pre-processing [3].

The present study compares EfficientNetV1 and EfficientNetV2 in terms of accuracy, sensitivity and specificity for TB screening in Shenzhen dataset[8]. The baseline B0 model was extended to B1, B2 and B3.

Ⅱ. Methods

This section provides a brief summary of EfficientNet models and highlights their differences. EfficientNet [5] (EfficientNetV1) is a family of lightweight convolutional models that are optimized for parameter and FLOPs efficiency. It takes advantage of neural architecture search (NAS) to design a baseline EfficientNet-B0 that has better trade-off on FLOPs and accuracy. The baseline network is then uniformly scaled up (depth, width, and resolution) with a simplified and effective compound scaling strategy to obtain a family of models B1-B7. The models have superiority over existing CNN models in terms of number parameters and FLOPs as they use depthwise convolutions. However, such convolutions often cannot fully utilize modern accelerators; therefore, EfficientNetV1 have main a limitation in terms of training or inference speed (for example, compare to ResNet-RS-420)[7]. Therefore, EfficientNetV2 [7] improves training speed of EfficientNetV1 models while maintaining the parameter efficiency. Specifically, EfficientNetV2 provides three solutions to EfficientNetV1 training bottleneck. First, for better training speed, EfficientNetV2 proposed to adjust the image size and regularization progressively during training. Second, EfficientNetV2 proposed a non-uniform scaling strategy to add more layers to later stages gradually as opposed to EfficientNetV1 that equally scales up all stages by using a simple compound scaling rule. Finally, to fully utilize modern accelerators, EfficientNetV2 proposed that Fused-

MBConv in early stage can improve training speed with a small overhead on parameters and FLOPs. The Fused-MBConv replaces the combination of depthwise Conv3×3 and expansion Conv1×1 in MBConv with a single regular Conv3×3 layer.

Ⅲ. Results

Dataset: In this study, a publicly available chest radiograph dataset called Shenzhen (SH) China dataset [8] was used. It consists of 326 CXR images of normal and 336 CXR images of TB cases. Only 326 images per class were used to balance the dataset. The dataset was split into training, validation and testing sets which account for 80%, 10 and 10%, respectively. In addition, the input images were pre-processed as: (1) all black borders and regions on the edges of images were cropped and (2) the images were resized from the center to meet the models input size requirements.

Metrics: In analysis, we considered samples with TB as positive and healthy samples as negative classes. The number of observations belonging to the positive class and classified as such are true positives (TP) and misclassified as negative class are false negatives (FN). Similarly, the number of observations belonging to the negative class and correctly classified as such are true negatives (TN) and misclassified as positive class are false positives (FP). For the models performance evaluation, we have considered three measures namely, accuracy (Acc), sensitivity (Sen) and specificity (Spe). The performance metrics are defined as:

$$\begin{cases} \text{Acc} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \\ \text{Sen} = (\text{TP})/(\text{TP} + \text{TN}) \\ \text{Spe} = (\text{TN})/(\text{TN} + \text{FP}) \end{cases} \tag{1}$$

Accuracy measures the total number of correct predictions (TP + TN) made by the model, which is the ratio of correct predictions to total predictions. Accuracy is important when FN and FP have similar costs. However, for disease diagnosis, the occurrence of FN is intolerable and identifying the positives is crucial [3]. For this purpose, sensitivity is the metric that measures performance of a model in terms of how fewer number of FN are predicted. Another related metric is specificity, which gives the ratio of true negatives to total negative in the observations.

Training Setup: The models were trained for 500 epochs using Stochastic Gradient Descent (SGD) with batch size 16. The initial learning rate was set to 0.1, which was then reduced by a factor of 10 when validation accuracy stopped improving. In addition, we have used early stopping criteria once the model validation accuracy stopped improving for 60 epochs. During training, random flip, rotation, zoom, translation and contrast were used as augmentation methods.

Table 1. presents EfficientNet models performance in terms of accuracy, sensitivity and specificity for TB detection in X-ray images. For the EfficientNetV1 models, the accuracy improved with the model size. However, for the EfficientNetV2, the best accuracy is

TABLE I. Performance comparison of EfficientNet models for TB screening in Shenzhen dataset.

Metric	EfficientNetV1				EfficientNetV2			
	B0	B1	B2	B3	B0	B1	B2	B3
P (M)	5.3	7.9	9.2	12.3	7.2	8.2	10.2	14.5
Acc	86	87	87	89	87	84	85	86
Sen	81	89	79	86	89	84	82	84
Spe	91	85	94	91	85	84	87	88

P (M): # of parameters (million)

achieved for B0. Overall, there is a 2% difference in the best accuracies of both versions. For sensitivity, the most important metric in disease diagnosis, both versions have the same value of 89%, which is achieved by EfficientNet-B1 and EfficientNetV2-B0.

IV. Conclusion

This study provides a comparative analysis of EfficientNet family models versions 1 and 2 for TB screening in X-ray images. The EfficientNetV2 preserves parameter and FLOPs efficiency of EfficientNetV1 while improving their training speed. The analysis have shown that EfficientNet models are suitable for TB detection with a sensitive score of 89%.

In medical image analysis, transfer learning (TL) is widely used to improve a model performance on limited available data. Therefore, comparing both versions with TL is an interesting future research direction.

ACKNOWLEDGMENT

This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07048338).

References

[1] “Global tuberculosis report 2021.” https://bit.ly/3AQASEu (accessed Aug. 28, 2022).

[2] M. Oloko-Oba and S. Viriri, “Ensemble of EfficientNets for the Diagnosis of Tuberculosis,” *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, Dec. 2021, doi: 10.1155/2021/9790894.

[3] I. Ahmad and S. Shin, “A Perceptual Encryption-Based Image Communication System for Deep Learning-Based Tuberculosis Diagnosis Using Healthcare Cloud Services,” *Electronics*, vol. 11, no. 16, p. 2514, Aug. 2022, doi: 10.3390/electronics11162514.

[4] M. Oloko-Oba and S. Viriri, “A Systematic Review of Deep Learning Techniques for Tuberculosis Detection From Chest Radiograph,” *Front. Med.*, vol. 9, p. 830515, Mar. 2022, doi: 10.3389/fmed.2022.830515.

[5] H. Li *et al.*, “Rethinking the Hyperparameters for Fine-tuning,” *ArXiv*, vol. abs/2002.11770, 2020.

[6] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, “Image Enhancement for Tuberculosis Detection Using Deep Learning,” *IEEE Access*, vol. 8, pp. 217897–217907, 2020, doi: 10.1109/ACCESS.2020.3041867.

[7] M. Tan and Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” *arXiv:2104.00298 [cs]*, Jun. 2021, Accessed: May 10, 2022. Available: http://arxiv.org/abs/2104.00298

[8] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma,“Two public chest X-ray datasets for computer-aided screening of pulmonary diseases,” *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, 2014.