Kỷ yếu Hội nghị KHCN Quốc gia lần thứ XII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR); Huế, ngày 07-08/6/2019
DOI: 10.15625/vap.2019.0005

AUTOMATED PNEUMONIA DETECTION IN X-RAY IMAGES VIA DEPTHWISE SEPARABLE CONVOLUTION BASED LEARNING

Nghia Duong Trung [1], Tuyen Tran Ngoc[2], Hiep Xuan Huynh[3]

[1] Can Tho University of Technology, Can Tho city, Vietnam
[2] Can Tho University of Technology, Can Tho city, Vietnam
[3]Can Tho University, Can Tho city, Vietnam

dtnghia@ctuet.edu.vn,duong-trung@ismll.de; tntuyen.khmt0115@student.ctuet.edu.vn; hxhiep@ctu.edu.vn

ABSTRACT: *In the medical study of lungs' infection, pneumonia may be caused by bacteria, or viruses. When lungs are infected by pneumonia, the air sacs become inflamed and fill up with fluid or pus. Highly-trained radiography expert interpreters are responsible for interpreting possible pneumonia in industry radiography. Consequently, it intensively relies on the aptitude and experience of the interpreter, plus the inadequate X-ray image's quality. Thus, the detection of pneumonia using different medical imaging techniques becomes very challenging. Any wrong detection could lead to serious consequences in medical treatment. Accurate identification is preliminary to any kind of intervention. Therefore, leveraging technology in automatic detection of these radiography has become essential. Unfortunately, constructing and training a complex deep learning model from scratch is mostly infeasible due to the lack of hardware infrastructure. Therefore, this paper exploits the idea of transfer learning which is the improvement of learning in a new prediction task through the transfer of knowledge from a related prediction task that has already been learned. This will improve the current computer vision methods based on the use of deep learning to more effectively diagnose the presence of pneumonia in X-rays images. By utilizing convolutional neural networks re-trained with our obtained data, our experiment shows that the proposed idea performs perfectly and achieves the classification accuracy of* ***98.0% ± 0.17*** *with the acceptable deployment time on a normal laptop.*

Keywords: *Pneumonia Detection, Transfer Learning, Deep Learning, X-Ray Images.*

I. INTRODUCTION

According to the World Health Organization (WHO), pneumonia approximately kills 2 million children under 5 years old every year and is consistently estimated as the single leading cause of childhood mortality [1,2]. In fact, it kills more children than HIV/AIDS, malaria, and measles combined. A statistics conducted by [3, 4] showed that 1.57 million children under 5 years old died from pneumonia in 2008, accounting for 18% of 8.8 million global childhood deaths. It is noticeable that nearly all cases, approximately 95%, of clinical pneumonia occur in developing countries, particularly in Southeast Asia and Africa. Therefore, accurate and timely detection is mandatory. One key element of detecting pneumonia is radiographic data since chest X-rays are routinely obtained as standard of care and can help differentiate between normal and/or pneumonia. To solve this problem, we investigated the effectiveness of a transfer learning framework in classifying chest X-rays images to detect pneumonia facilitating rapid referrals for children needing urgent intervention.

From the machine learning point of view, the mentioned problem could be unquestionably addressable by the adoption of a new rapid solution that can bring X-Ray experts and computer scientists into one choir. A major assumption in applying many machine learning algorithms is that the training and future data must be in a similar feature space. Any differences may be eliminated before learning or they have no equivalent covariances during training a model. However, in many real-world applications, this assumption may be a solid drawback. The isolation insists on an entire learning procedure from dataset acquirement, model learning, model evaluation and hyperparameter tuning. Thus, a demand for computing infrastructure and financial support is obviously seen. Transfer learning, however, attempts to change it by developing methods to transfer knowledge learned in one or more source tasks and use it to improve learning in a related target task [5, 6]. Transfer learning improves learning in the target task by leveraging knowledge from the source task.

Transfer learning methods tend to be greatly dependent on the machine learning algorithms being used to learn the prediction tasks, and can often merely be considered extensions of those algorithms. Tremendous progress has been made in image classification and recognition, primarily thanks to the availability of large-scale annotated datasets, i.e. ImageNet [7]. Since Krizhevsky *et al.* [8] won the ImageNet 2012 competition, there have been much interest and work toward the revival of deep convolutional neural networks [9], especially in the task of image classification [10, 11, 12]. However, in this paper, we aim neither to maximize absolute performance nor to build a complete model from scratch, but rather to study transfer results of a well-known convolutional architecture. We use the reference implementation provided by Tensorflow [13, 14] so that our experiment results will be comparable, extensible and usable for a large number of upcoming research.

The rest of this paper is as follows. First of all, Section II summarizes several key research on the task of X-Ray images classification. In Section III, we briefly discuss the overview of technical background including transfer learning, deep convolutional neural networks that summarizes a critical state-of-the-art review existing in the literature that is essential to solving the problems. In Section IV, we evaluate and perform the approach to our experiment dataset. Finally, Section V recapitulates the approaches and discuss achievements done in this research.

II. STATE OF THE ART

Computerized clinical specification and decision recommendation systems have been developed to help radiologists and clinical experts diagnose and detect disease. Success integration of these automated systems has empowered the work done in medical images and graphics research. An early attempt to classify pneumonia was conducted nearly two decades ago [15]. In that paper, they detected pneumonia by developing several natural language processing techniques to extract information from chest X-ray reports. They compared the performance of expert-crafted rules, a Bayesian network, and a decision tree to determine whether a chest X-ray report contained enough information to support a diagnosis of acute pneumonia. Later chest radiography has become pervasive in computer-aided diagnosis [16] and is one of the primary diagnostic imaging procedures for evaluating diseases in medical practices because it is economical and easy to use. Since then, X-ray images have been taken enough to provide data for multi-discipline research. Fortunately, deep neural networks have lately earned much attention due to its significant achievement in a wide range of computer science applications. Rapid advances in deep neural networks have resulted in the deployment of convolutional neural networks (CNNs) in the task of chest radiography classification. In a work done in [17] that applied CNNs to obtain deep learning features of X-rays images and executed random forest model on top to classify lung nodule. Image-based deep learning classifiers have gained potential generalization in biomedical image interpretation and medical decision making. The authors in [18] compared the performance of traditional deep neural networks and transfer learning in a vast application. Furthermore, they investigated a transfer learning model, called inception-V3, in the comparison of chest X-rays presenting as pneumonia versus normal. They achieved an accuracy of 92.8%, with a sensitivity of 93.2% and a specificity of 90.1%. We re-implement their methods and achieve the equivalent results. In this study, we evaluate a more effective transfer learning model in a similar classification task and a dataset of chest X-ray images.

III. MATERIAL AND METHODS

A. Transfer Learning

Given a training dataset X_train $\in \mathbb{R}^{n \times m}$ and the equivalent label Y_train $\in \mathbb{R}^{n}$ where *n*, and *m* are the number of observations, and the number of features respectively. To construct a prediction model *f*, traditional machine learning methods are trained by pairs of (X_train, Y_train) = {(x_i, y_i), …, (x_n, y_n)}. Traditionally, a major assumption in many machine learning algorithms is that the training and future data must be in the similar feature space, more specifically X_train and X_test must be under the same distribution *D*. It means that the training and the test dataset must share the similar marginal probability distribution P(X) over *D*. However, in many real-world applications, this assumption may be difficult to follow.

Transfer learning, however, attempts to change it by developing methods to transfer knowledge learned in one or more source tasks and use it to improve learning in a related target task [5, 19, 20]. It allows the model to be applied to datasets drawn from some distribution different from the one upon which it was trained. Modern object identification and classification models with millions of parameters can take weeks to fully train. Transfer learning is a technique that shortcuts a lot of this work by taking a fully-trained model for a set of predefined categories like ImageNet [7, 21], and retrains from the existing weights for new implemented classes. The goal of transfer learning is to improve learning in the target task by leveraging knowledge transferability from the source task. Thus, the transfer learning procedure can be defined as follows. Two identification datasets *D_source* and *D_test* are constructed by two different research groups. Our task is to assign labels *Y_target_test* to test data *X_target_test* drawn from distribution *D_target* given the training data (*X_target_train*, *Y_target_train*) drawn from distribution *D_target*. Among the different approaches to transfer learning [22, 19], we prefer instance-based transfer [23], which assumes that some instances in the source domain can be reused. By re-weighting weights of instances in the source domain, effects of dissimilar observations will be reduced, while similar observations will contribute more to the target domain and may thus lead to a more accurate model. The difference between traditional machine learning and transfer learning is presented in Figure (1).

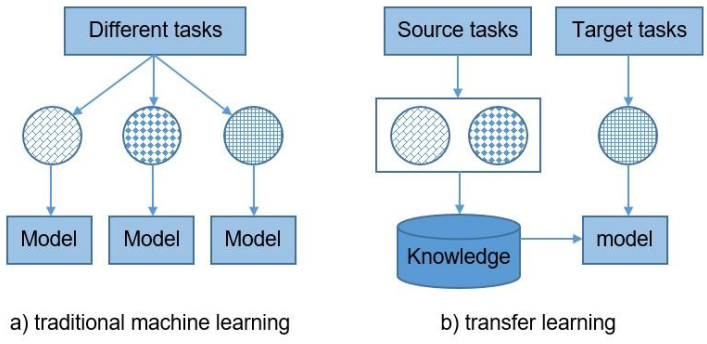

Figure 1. Difference between traditional machine learning and transfer learning

B. Convolutional Neural Networks

Convolutional neural networks (CNNs) have recently gained outstanding image classification performance in the large-scale challenges [24, 7, 21, 25]. The success of CNNs is achieved by their ability to learn rich level image representations as its hidden layers can be integrated theoretically unlimited. However, learning CNNs requires a very large number of annotated image samples and an estimation of millions of model parameters. This property obviously prevents the application of CNNs to problems with limited training data. There is a phenomenon in deep neural networks such that when trained the model on images, it tends to learn first-layer features that resemble either Gabor filters or color blobs [26]. Such first-layer features appear not to be specific to a particular dataset, but generally in the way that they are applicable to different tasks. The transition of knowledge is eventually transferred from the first layer to the last layer of the network. Expectedly, several large-scale datasets can be used to train the learning models, and then the learned models are applied to a particular target task where the parameters of the last layer are re-weighted based on its own dataset. The idea of transferring knowledge along deep neural networks have been explored by many previous researches [26, 27, 28, 29, 30]. Going to that research direction, we explore the performance of several state-of-the-art convolutional neural networks upon our collected data.

C. Depthwise Separable Convolution Based Model

MobileNets are a class of convolutional neural network designed by researches at Google [14]. It is designed to effectively maximize accuracy while being mindful of the restricted resources for an on-device or embedded application. The models are effectively small, low-latency, low-power parameterized to meet the resource constraints of a variety of use cases. They can be built upon for classification, detection, embeddings and segmentation similar to how other popular large-scale models. The main difference between the MobileNets architecture and a traditional CNNs is instead of a single 3×3 convolution layer followed by batch normalization [31] and ReLU [32], MobileNets split the convolution into a 3×3 depthwise convolution and a 1×1 pointwise convolution. The depthwise convolution applies a single filter to each input channel while the pointwise convolution combines the outputs of the depthwise ones. These two convolutions are shown in Figure (2). The depthwise separable convolution based model is described in Table (2). The comparison between a standard full convolution and 16 versions of MobileNets is presented in Table (3). Readers should refer to the original paper of the model [14] for greater details.

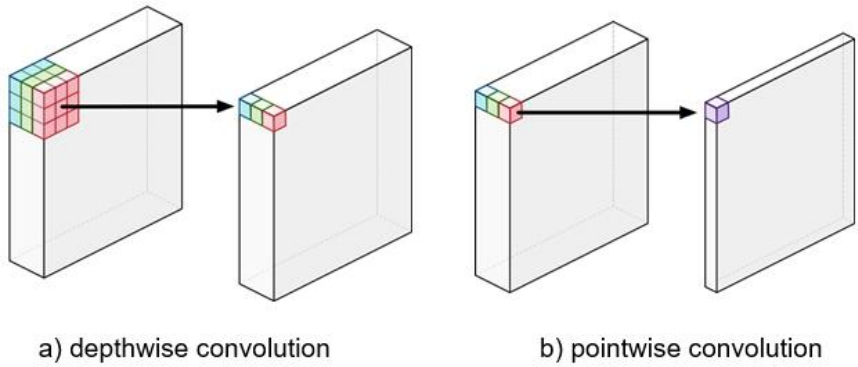

Figure 2. A depthwise and a pointwise convolutions

Thanks to the idea of depthwise separable convolution, the network architecture is lighter, and consequently, the computation expense is significantly reduced. In order to build such less computationally expensive architecture, the model surfaces two hyper-parameters, e.g. width multiplier and resolution multiplier, that we can tune to fit the resource and/or accuracy trade-off of our implemented model. The width multiplier allows us to thin the network, while the resolution multiplier changes the input dimensions of the image, reducing the internal representation at every layer. Given α and ρ be the width multiplier and resolution multiplier respectively. Given the value of $\alpha \in \{1, 0.75, 0.5, 0.25\}$ and the value of $\rho \in \{224, 192, 160, 128\}$, Table (3) shows the comparison between a full convolution and 16 combinations of α and ρ in terms of the number of fused multiplication and addition operations, and the number of learned parameters. The reduction of computational cost and the number of parameters is quadratically by roughly α^2 and ρ^2.

IV. EXPERIMENTS

A. Dataset

The authors summarize the experiment dataset as follows. Chest X-ray images were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care. For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable

scans. This is a huge dataset on the X-Ray images and several data packages have been released to the research community. In this paper's experiments, the authors obtained a total of 5216 chest X-ray images from children, including 3874 characterized as depicting pneumonia and 1341 normal. The dimensions of images are varied ranging from 384 x 127 to 2772 x 2098. More details of the data can be found in [33]. Several examples of Chest X-Rays can be found in Figure (3). The authors have not applied any preprocessing techniques and keep these images as they are in designing our workflow.

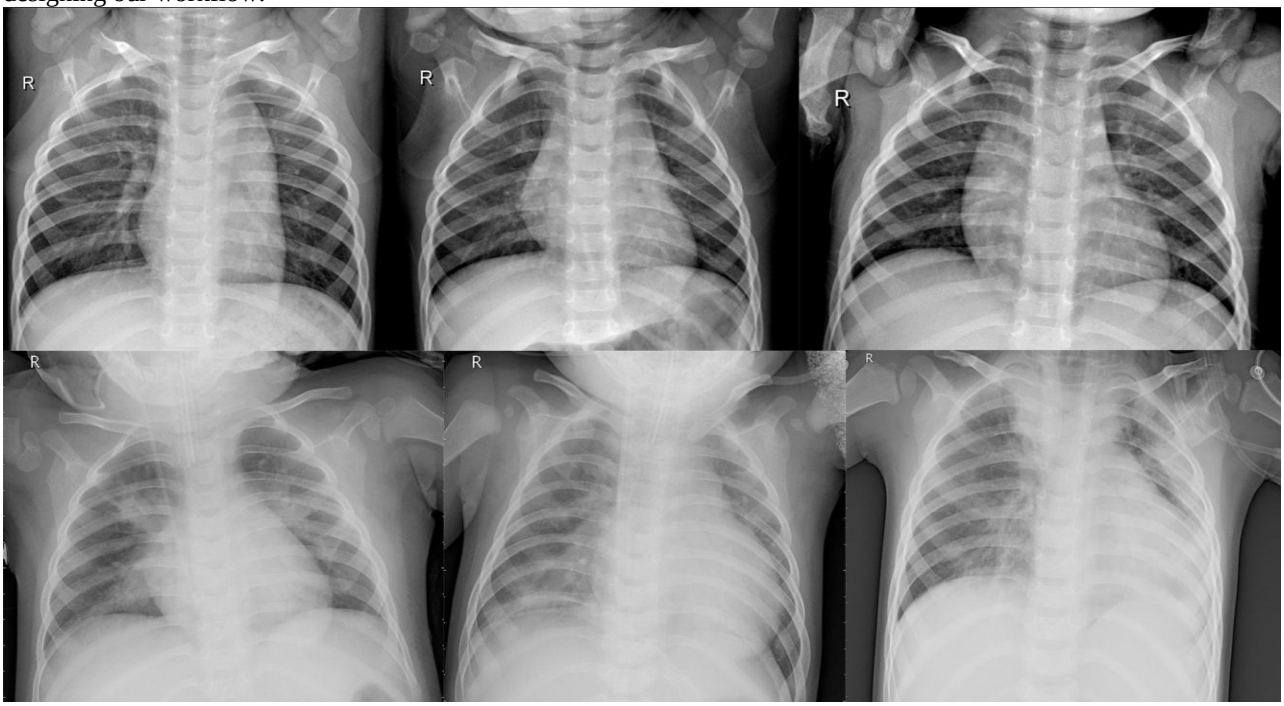

Figure 3. Illustrative examples of chest X-Rays in patients with pneumonia. The normal chest X-rays are the upper three images while pneumonia are the lower three images. Note that these images are selected for the ease of illustration

B. Data Splitting Schemes

Machine learning models have the fundamental goal of making accurate predictions on unseen instances beyond those appeared in the training set. To estimate the quality of models' predictions with data it has not seen, we can split a portion of the data for which we already know the answer as a proxy for the unseen data. Then we evaluate how well the model predicts for that data. Typically, training dataset contains observations used to fit a learning model. Validation dataset comprises instances used to provide an unprejudiced evaluation of the learning model by tuning hyperparameters. Test dataset includes samples of data used to provide an unbiased evaluation of the final learning model fit on the training dataset. The authors randomly shuffle the data into training, test, and validation sets without replacement in every experiment. In our experiment, we set up three different dataset splitting schemes by tuning various split ratios. We divide the obtaining data into three different splitting schemes which the readers can see a summation in Table (1).

1. Dataset splitting scheme 1: The proportion of training, validation and test sets are 80%, 10%, and 10% respectively. We denote is as 80|10|10 hereafter.

2. Dataset splitting scheme 2: The proportion of training, validation and test sets are 70%, 15%, and 15% respectively. We denote is as 70|15|15 hereafter.

3. Dataset splitting scheme 3: The proportion of training, validation and test sets are 60%, 20%, and 20% respectively. We denote is as 60|20|20 hereafter.

Table 1. Summary of several splitting schemes.

Pneumonia Data	Splitting schemes		
	80 | 10 | 10	70 | 15 | 15	60 | 20 | 20
A total of 5216 images	4172 | 522 | 522	3652 | 782 | 782	3130 | 1043 | 1043

C. Implementation and Results

The implementation of the models is done using Python in the Anaconda Python Distribution environment in the Windows 10 platform. In our experiments, we set the required model hyper-parameters as follows. The learning rate is {0.1,0.01,0.001}. The number of epoch is {1500,2000,2500}. Our experiments were conducted on a normal

laptop Core i7-6500U with 2.5GHz clock speed, 8GB of RAM. The low-end GPU NVIDIA GeForce 940MX with 4GB of RAM is activated by default. The input size of the depthwise separable convolution based model is $n \times n \times 3$ for height, width and channel respectively. $n \in \{224, 192, 160, 128\}$. $\alpha \in \{1.0, 0.75, 0.50, 0.25\}$ is the hyperparameter that controls total amount of parameters used by the model. The classification decision is made at the softmax layer where its input is the probability distribution of investigated labels. The depthwise separable convolution based model is described in Table (2). We have executed 6912 experiments that each experiment requires from 5 minutes to 12 minutes to complete depending on the models' hyperparameters configuration. The identification accuracy is reported in Table (4).

1. Scenario 1: Experiment results on dataset splitting scheme 1

In this scenario, we evaluated our adapted deep architecture on dataset splitting scheme 1. We randomly shuffle dataset without replacement 3 times and execute 16 versions of the MobileNets model. Then we take an average in the end. The best percentage of accuracy score is 98.0% $\pm$ 0.17 in case of epoch 2000, α = 1.0, learning rate 0.01. The results of this scenario is presented in Table (4), column 80 | 10 | 10 specifically.

2. Scenario 2: Experiment results on dataset splitting scheme 2

Similar to scenario 1, we randomly shuffle dataset without replacement 3 times, execute 16 versions of the MobileNets model and take an average of classification accuracy in the end. The best accurate score is 97.2% in case of epoch 2000, α = 1.0, learning rate 0.01. The results of this scenario is presented in Table (4), column 70 | 15 | 15 specifically.

3. Scenario 3: Experiment results on dataset splitting scheme 3

The third scenario is investigated to examine the model's performance in the case of the lower proportion of training set, e.g. 60%. We apply a similar experiment configuration as the previous two scenarios. The best detection score is 97.5% which evidently confirms the effectiveness of transfer learning in this image classification task. The results of this scenario is presented in Table (4), column 60 | 20 | 20 specifically.

Table 2. The outline of depthwise separable convolution based architecture. At the softmax layer, c is the number of predicted labels

Type / Stride	Filter shape	Input size
Conv / s2	3×3×3×32	$n \times n \times 3$
Conv dw / s1	3×3×32 dw	112×112×32
Conv s1	1×1×32×64	112×112×32
Conv dw / s2	3×3×64 dw	112×112×64
Conv / s1	1×1×64×128	56×56×64
Conv dw / s1	3×3×128 dw	56×56×128
Conv / s1	1×1×128×128	56×56×128
Conv dw / s2	3×3×128 dw	56×56×128
Conv / s1	1×1×128×256	28×28×128
Conv dw / s1	3×3×256 dw	28×28×256
Conv / s1	1×1×256×256	28×28×256
Conv dw / s2	3×3×256 dw	28×28×256
Conv / s1	1×1×256×512	14×14×256
5× Conv dw / s1 Conv / s1	3×3×512 dw 1×1×512×512	14×14×512 14×14×512
Conv dw / s2	3×3×512 dw	14×14×512
Conv / s1	1×1×512×1024	7×7×512
Conv dw / s2	3×3×1024 dw	7×7×1024
Conv / s1	1×1×1024×1024	7×7×1024
Avg Pool / s1	Pool 7×7	7×7×1024
FC / s1	1024×1000	1×1×1024
Softmax / s1	Classifier	$1 \times 1 \times c$

Table 3. The comparison between a full convolution and 16 versions of MobileNets

Models	Million Mult-Adds	Million Parameters
Full convolution	4866	29.3
MobileNets 1.0-244	569	4.24
MobileNets 1.0-192	418	4.24
MobileNets 1.0-160	291	4.24
MobileNets 1.0-128	186	4.24
MobileNets 0.75-244	317	2.59
MobileNets 0.75-192	233	2.59
MobileNets 0.75-160	162	2.59
MobileNets 0.75-128	104	2.59
MobileNets 0.50-244	150	1.34
MobileNets 0.50-192	110	1.34
MobileNets 0.50-160	77	1.34
MobileNets 0.50-128	49	1.34
MobileNets 0.25-244	41	0.47
MobileNets 0.25-192	34	0.47
MobileNets 0.25-160	21	0.47
MobileNets 0.25-128	14	0.47

Table 4. The classification accuracy of evaluated models on the chest X-ray dataset. The best score is in **bold**

Models	Splitting schemes		
	80 | 10 | 10	70 | 15 | 15	60 | 20 | 20
MobileNets 1.0-244	**98.0% ± 0.17**	96.2% ± 0.06	97.1% ± 0.60
MobileNets 1.0-192	97.3% ± 0.31	97.1% ± 0.15	97.0% ± 0.15
MobileNets 1.0-160	97.2% ± 0.21	95.8% ± 0.87	97.5% ± 0.29
MobileNets 1.0-128	96.8% ± 0.15	96.4% ± 0.35	96.8% ± 0.35
MobileNets 0.75-244	97.6% ± 0.36	96.8% ± 0.75	96.0% ± 0.35
MobileNets 0.75-192	97.4% ± 0.25	96.7% ± 0.17	96.5% ± 0.61
MobileNets 0.75-160	97.8% ± 0.23	97.2% ± 0.44	96.2% ± 0.96
MobileNets 0.75-128	97.7% ± 0.10	96.8% ± 0.51	95.7% ± 1.50
MobileNets 0.50-244	97.3% ± 0.87	96.1% ± 0.30	96.7% ± 0.81
MobileNets 0.50-192	96.8% ± 0.45	96.2% ± 0.38	96.4% ± 0.40
MobileNets 0.50-160	97.0% ± 0.66	95.8% ± 0.26	94.8% ± 0.72
MobileNets 0.50-128	96.6% ± 0.06	96.4% ± 0.68	95.7% ± 0.06
MobileNets 0.25-244	96.5% ± 0.21	95.6% ± 0.35	95.5% ± 0.38
MobileNets 0.25-192	96.2% ± 0.21	95.0% ± 1.32	96.7% ± 0.61
MobileNets 0.25-160	97.4% ± 0.15	95.4% ± 0.70	95.5% ± 0.15
MobileNets 0.25-128	97.2% ± 0.21	96.3% ± 0.21	95.9% ± 1.22

D. Remarks

Overall, the experiment results signify the robustness and effectiveness of the combination of knowledge transferability and depthwise separable convolution in the task of pneumonia detection. A highly good accuracy score is achieved by 98.0% with a slight standard deviation of only ± 0.17. Interestingly, the lowest score accomplishing by the MobileNets 0.50-160 version under the data splitting scheme 60 | 20 | 20 is 94.8% ± 0.72 which also indicates a very good benchmark. Although MobileNets 0.25-128 has 14 million multiplication addition and 0.47 million parameters compared with 569 million multiplication addition and 4.24 million parameters of MobileNets 1.0-244, see Table (3), the performance of these models is not significantly different. Looking further into all experiment approaches, the performance of 16 versions of the architecture when the training epoch greater than 1500 is identical for all models. The accuracy score becomes stable beyond 1500^{th} epoch. One of the things to note is that the dataset is unbalanced as 3874 and 1341 images are pneumonia and normal respectively. The authors inspect the confusion matrix of the most powerful architecture, e.g. the MobileNets 1.0-224, showing that the sensitivity and specificity are similarly reliable results, e.g. 97.0% and 98.1% respectively. The best achievement of our adapted model's architecture on the training and test sets, e.g. MobileNets 1.0-244, is presented in Figure (4).

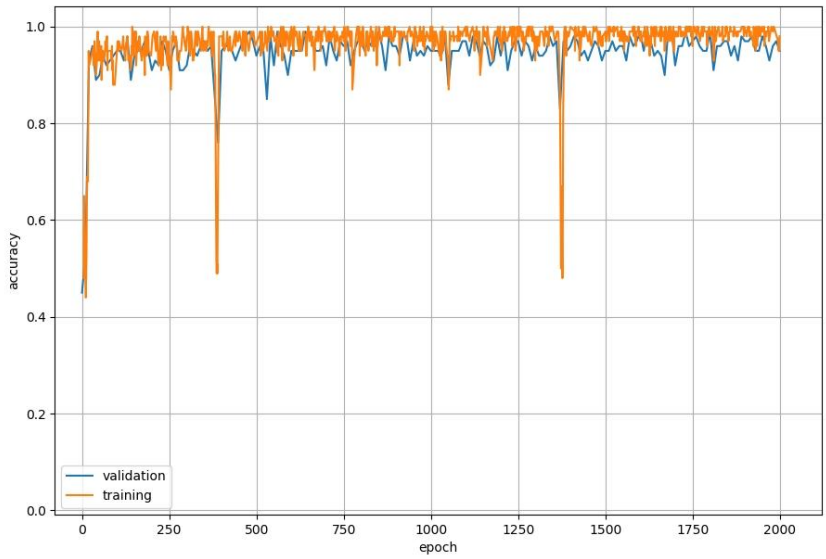

Figure 4. The classification accuracy on the training and validation sets in case of MobileNets 1.0-244 architecture

One of the interesting phenomena to note in Figure (4) is that several dramatical plunges in the training's line are observed. A possible reason is that the model encounters various complicated observations during training. However, these falls do not affect the model's performance on the validation set. We extend the number of training epoch to 4000 and 6000 and this phenomenon fades away.

V. CONCLUSION

By conducting this research paper, the authors have investigated a deep convolutional neural network combining with the idea of transfer learning and applied it on a critical task of radiography image classification. We aim our research at supporting image analysis in a hospital that paves an association between computer vision and medical duties. The potential generalization of deep learning and transfer in the context of image analysis has been highlighted. An application in pneumonia detection has been deployed by utilizing cutting-edge machine learning frameworks and depthwise separable convolution that would indicate blossom research in this direction. The development of the framework started with a specific focus on the binary of pneumonia versus normal, but soon expands toward many other image-categorization-based problems and quantification. We comprehend several vital results on both theory and practice such as convolution, neural network architecture, a configuration of CNN, implementation of CNN on an open source software, investigation of a crucial medical diagnosis. We obtain a benchmark classification accuracy of $\mathbf{98.0\% \pm 0.17}$ on the obtained observations.

REFERENCES

[1] I. Rudan, C. Boschi-Pinto, Z. Biloglav, K. Mulholland, and H. Campbell, "Epidemiology and etiology of childhood pneumonia," Bulletin of the world health organization, vol. 86, pp. 408-416B, 2008.

[2] I. Rudan, K. L. O'brien, H. Nair, L. Liu, E. Theodoratou, S. Qazi, I. Lukˇsi´c, C. L. F. Walker, R. E. Black, H. Campbell et al., "Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries," Journal of global health, vol. 3, no. 1, 2013.

[3] R. E. Black, S. Cousens, H. L. Johnson, J. E. Lawn, I. Rudan, D. G. Bassani, P. Jha, H. Campbell, C. F. Walker, R. Cibulskis et al., "Global, regional, and national causes of child mortality in 2008: a systematic analysis," The lancet, vol. 375, no. 9730, pp. 1969-1987, 2010.

[4] R. A. Adegbola, "Childhood pneumonia as a global health priority and the strategic interest of the bill & melinda gates foundation," Clinical infectious diseases, vol. 54, no. suppl 2, pp. S89-S92, 2012.

[5] L. Torrey and J. Shavlik, "Transfer learning," in Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, 2010, pp. 242-264.

[6] S. J. Pan, Q. Yang et al., "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," 2009.

[8] I. L.S.V.R. Competition, "Available online: http://www.image-net.org/challenges," LSVRC/(accessed on 27 December 2016), 2012.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[12] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 2, pp. 645-657, 2017.

[13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Man´e, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Vi´egas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: http://tensorflow.org/

[14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[15] W. W. Chapman, M. Fizman, B. E. Chapman, and P. J. Haug, "A comparison of classification algorithms to automatically identify chest x-ray reports that support pneumonia," Journal of biomedical informatics, vol. 34, no. 1, pp. 4-14, 2001.

[16] S. Katsuragawa and K. Doi, "Computer-aided diagnosis in chest radiography," Computerized Medical Imaging and Graphics, vol. 31, no. 4-5, pp. 212-223, 2007.

[17] C. Wang, A. Elazab, J. Wu, and Q. Hu, "Lung nodule classification using deep feature fusion in chest radiography," Computerized Medical Imaging and Graphics, vol. 57, pp. 10-18, 2017.

[18] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122-1131, 2018.

[19] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big Data, vol.3, no.1, p. 9, 2016.

[20] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill, "Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs," Scientific reports, vol. 8, no.1, p.16685, 2018.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, vol. 115, no. 3, pp. 211-252, 2015.

[22] R. Bekkerman, M. Bilenko, and J. Langford, Scaling up machine learning: Parallel and distributed approaches. Cambridge University Press, 2011.

[23] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010, pp. 1855-1862.

[24] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," International journal of computer vision, vol. 88, no. 2, pp. 303-338, 2010.

[25] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.

[26] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in Advances in neural information processing systems, 2014, pp. 3320-3328.

[27] O. Russakovsky, J. Deng, Z. Huang, A. C. Berg, and L. Fei-Fei, "Detecting avocados to zucchinis: what have we done, and where are we going?" in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2064-2071.

[28] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1717-1724.

[29] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1285-1298, 2016.

[30] N. Duong Trung, L. D. Quach, and C. N. Nguyen, "Learning deep transferability for several agricultural classification problems," International Journal of Advanced Computer Science and Applications, vol.10, no. 1, 2019. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2019.0100107

[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.

[32] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on cifar-10," Unpublished manuscript, vol. 40, no. 7, 2010.

[33] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (oct) and chest x-ray images for classification," Mendeley Data, vol. 2, 2018.

TỰ ĐỘNG NHẬN DẠNG BỆNH VIÊM PHỔI TỪ ẢNH X-QUANG SỬ DỤNG PHƯƠNG PHÁP HỌC KẾT CẤU TÍCH CHẬP THEO CHIỀU SÂU

Nghia Duong-Trung, Tuyen Tran Ngoc, Hiep Xuan Huynh

TÓM TẮT. *Trong nghiên cứu y học về nhiễm trùng phổi, viêm phổi có thể do vi khuẩn hoặc vi rút gây ra. Khi phổi bị nhiễm viêm phổi, túi khí bị viêm và chứa đầy chất lỏng hoặc mủ. Chuyên gia X quang được đào tạo chuyên sâu có trách nhiệm xác định đúng viêm phổi có thể trong chụp X quang công nghiệp. Công việc này phụ thuộc rất nhiều vào năng lực và kinh nghiệm của chuyên gia, cộng với chất lượng hình ảnh X quang không đầy đủ. Do đó, việc phát hiện viêm phổi bằng các kỹ thuật hình ảnh y tế khác nhau trở nên rất khó khăn. Bất kỳ phát hiện sai có thể dẫn đến hậu quả nghiêm trọng trong điều trị y tế. Xác định chính xác là sơ bộ cho bất kỳ loại can thiệp. Do đó, công nghệ tận dụng trong việc phát hiện tự động các hình ảnh X quang này đã trở nên cần thiết. Thật không may, việc xây dựng và đào tạo một mô hình học sâu phức tạp từ đầu hầu như không khả thi do thiếu cơ sở hạ tầng phần cứng. Do đó, bài viết này khai thác ý tưởng học chuyển giao, đó là cải tiến việc học trong một nhiệm vụ dự đoán mới thông qua việc chuyển giao kiến thức từ một nhiệm vụ dự đoán liên quan đã được học. Điều này sẽ cải thiện các phương pháp thị giác máy tính hiện tại dựa trên việc sử dụng học sâu để chẩn đoán hiệu quả hơn sự hiện diện của viêm phổi trong hình ảnh X quang. Bằng cách sử dụng các mạng thần kinh tích chập được đào tạo lại với dữ liệu thu được của chúng tôi, thử nghiệm của chúng tôi cho thấy ý tưởng được đề xuất thực hiện hoàn hảo và đạt được độ chính xác phân loại là 98,0% ± 0,17 với thời gian triển khai chấp nhận được trên một máy tính xách tay thông thường.*

Từ khóa. *Phát hiện viêm phổi, Học chuyển, Học sâu, Hình ảnh X-quang.*