

Bacteria Shape Classification using Small-Scale Depthwise Separable CNNs

Duc-Tho Mai¹ and Koichiro Ishibashi²

Abstract—Fast detection and classification of bacteria species play a crucial role in modern clinical microbiology systems. These processes are often performed manually by medical biologists using different shapes and morphological characteristics of bacteria species. However, it is clear that the manual taxonomy of bacteria types from microscopy images takes time and effort and is a great challenge for even experienced experts. A new revolution has been inaugurating with the development of machine learning methods to identify bacteria automatically from digital electron microscopy. In this paper, we introduce an automated model of bacteria shape classification based on Depthwise Separable Convolution Neural Networks (DS-CNNs). This architecture has excellent advantages with lower computational costs and reliable recognition accuracy. The experiment results indicate that after training with 1669 images, the proposed architecture can reach 97% validation accuracy and work well to classify three main shapes of bacteria.

I. INTRODUCTION

Convolutional Neural Networks (CNNs) have recently made significant progress in several fields and reached remarkable achievements, such as image processing [1], [2], object detection [3], and semantic segmentation [4]. The high recognition accuracy and good performance of CNNs even better than humans in some cases have been promoting the development of medical image analysis. Nowadays, bacterial infections are the leading cause of mortal in both developed and developing countries in all over the world, killing millions of lives each year. Therefore, it is absolutely essential to apply the rapid diagnosis and classification of the infections caused by bacteria to allow earlier preclusion and reduce their impacts.

A. Related Works

The classification of bacteria is a significantly important study for different medical purposes. In this context, it is necessary to classify the kind of strain accurately; however, in other contexts, we only need to distinguish between the possible options of bacteria, such as color or shape. Many recommendations are currently suggested to identify and recognize bacteria from the medical images based on Neural Networks (NNs).

The most fundamental technique applied for classifying bacteria is based on the bacterium's shape and cell arrangement. Rod, coccus, and spiral are three common shapes

of bacteria. Liu et al. [5] suggested the first work for identifying the shape of bacteria named CMEIAS (Center for Microbial Ecology Image Analysis System) with two object classifiers to extract size and shape measurements of segmented, digital images of microorganisms and classify them into their appropriate morphotype. The experimental accuracy that CMEIAS shaped classifier reached, was approximately 97% when performing on a dataset of 11 bacterial morphotype classes. The authors in the study [6] utilized the shape detector to recognize bacterias. The image processing consists of two stages, initially enhancing the image and extracting only essential bacteria shapes, then detecting rod shape and spherical (coccus) shape bacteria. A model of bacteria shape recognition based on CNNs and region variance was introduced by Polap et al. [7]. The experiments of that study were performed on 63 images of bacteria shapes, and the results after training could precisely classify rod-shaped and spherical(coccus) one over 91% and 78%, respectively.

Different traditional machine learning methods and modern deep learning techniques have been widely used in several works to recognize and identify bacteria types. In [8], CNNs were used for counting bacterial species from microscopy colony images with an overall accuracy of 92.8% on a large training dataset. Lei Huang et al. [9] presented automated applications for bacterial colony classification tasks using Deep CNNs on a dataset of 18 classes derived from Peking University First Hospital. The authors announced that the classification accuracy was from 73% to 90%. With the same idea using transfer learning techniques [10], Wahid et al. [11] employed Deep CNNs, called Xception architecture, to identify bacteria from microscopic images with 97.5% accuracy. At the same time, M.Talo [12] implemented the pre-trained CNN architecture based on ResNet-50 to classify digital bacteria images in the Digital Image of Bacteria Species (DIBaS) dataset into 33 categories. That work archived an average classification accuracy of 99.12% higher than that of Wahid's.

However, the traditional Deep CNNs request more computational complexity and memory-intensive when increasing accuracy requirements and the large scale size of the architecture model, making it challenging to apply on portable and limited hardware resource devices. The DS-CNNs can exceedingly reduce model parameters and improve the computational speed with negligible accuracy reduction, so it is an appropriate option for low resources demand and low power consumption. One of the first works on this research was executed by Wang et al. [13] with the model of a

¹Duc-Tho Mai is with Graduate School of Informatics and Engineering, The University of Electro-Communications (UEC), Tokyo, Japan thomaiduc@uec.ac.jp

²Koichiro Ishibashi with the Graduate School of Informatics and Engineering, The University of Electro-Communications (UEC), Tokyo, Japan ishibashi@uec.ac.jp

large number of DS-Conv blocks to identify human tissue types with the highest classification accuracy of 97% from 3D OCT images. In this article, we introduce a small-scale architecture that utilizes Depthwise Separable Convolution (DS-Conv) to classify the bacteria shape as an initial fast bacteria recognition tool from microscopy images.

The remaining of this paper can be divided into parts as follows: In section II, a brief DS-Conv operation and the proposed architecture are introduced. Section III describes the material and experimental setups. The results are provided in Section IV, and we finally conclude this study in Section V.

II. THE PROPOSED METHODS

In this section, we describe the methods that have been used in our study. These are assembling of the fundamental DS-Conv knowledge and proposed architecture. All these subsections are shown below:

A. Depthwise Separable Convolution

The depthwise separable convolution [14] appeared the first time on Laurent Sifre's thesis and then was applied for image classification. DS-Conv can be divided into depthwise convolution for space filter operation and 1×1 convolution (also called pointwise convolution) for combining the input channels linearly. Therefore, the redundant computation and model size can be effectively decreased. Figure. 1 depicts the mechanism operations of standard convolution and DS-Conv. A ratio of the parameters between two convolution blocks is constructed to demonstrate the reduction of computational cost and model size.

As shown in Figure 1(a), in which the size of the input image is $D_k \times D_k \times M$, where D_k is the height and width of the input image, M is the number of input channels. Each convolution layer uses filters of size $D_f \times D_f$ (one channel per filter) with N filters. When N filters are taken to slide through the input image, one intermediate feature map is produced by convolving each input feature map with a 2D filter kernel in depthwise convolution block, then during point-wise convolution, the set of intermediate feature maps are convolved with a 3D filter kernel to create final output feature maps, $D_g \times D_g \times N$.

We assume that $P_{standard}$ is denoted the number of parameters of the traditional convolution, can be calculated:

$$P_{standard} = D_f \times D_f \times M \times D_g \times D_g \times N = D_f^2 \times M \times D_g^2 \times N \quad (1)$$

Following Figure 1(b), assuming P_{dsc} represents the quantity of parameters in the depthwise separable convolution, can be written as Eq. (2):

$$P_{dsc} = D_f^2 \times 1 \times D_g^2 \times M + 1^2 \times M \times D_g^2 \times N \quad (2)$$

Therefore, from Eq. (1) and (2), the ratio of parameters between separable convolution and the conventional convolution can be written as:

$$\frac{P_{dsc}}{P_{standard}} = \frac{1}{N} + \frac{1}{D_f^2} \quad (3)$$

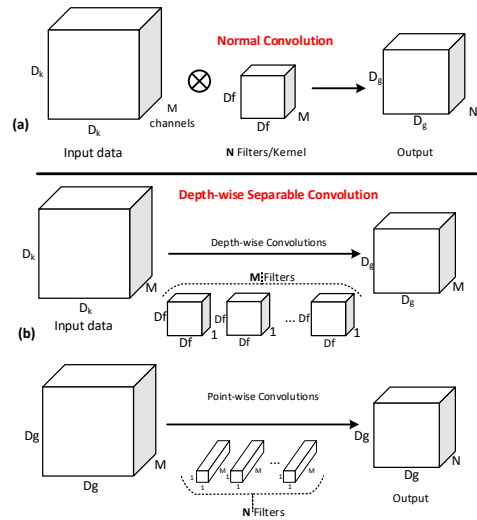


Fig. 1. Convolution mechanism in: (a) Traditional Convolution; (b) Depthwise Separable Convolution.

Eq. (3) shows that the computation cost can be reduced to $\frac{1}{N} + \frac{1}{D_f^2}$ compared to the conventional convolution process. Our study utilizes $D_f \times D_f = 3 \times 3$ and $N = 64$, so the computation complexity and the number of parameters of corresponding convolution layers are 10 ~ 13 times less than that standard convolution at only a small accuracy loss.

B. Proposed Model Architecture

Our bacteria shape classification process will be described in Fig. 2 from Input microscopy images block to Shape classification one in which 2^{nd} and 3^{rd} stages are the main contributions. In the data preprocessing step, the input images are resized to fit with the model, augmented the quantity of images for training, labeled with random weights, then the datasets are divided into a training set and test set with a ratio of 80/20, respectively. After the 2^{nd} step,



Fig. 2. Suggested stages of shape classification, starting from initial microscopy image to shape classification.

the input images with the size of $128 \times 128 \times 3$ fetch to Depthwise Separable CNNs for training, and the desired outputs are a weight file and the probability of bacteria shape (%). We employ the resize function's scale feature and an interpolation parameter to reduce the input image dimensions from the originality to expected results.

The proposed architecture is intelligibly described in Fig. 3 with overall five layers, in which the first three convolution layers extracting image features with a depth of 64 and filter size 3×3 . After that, one Fully Connected (FC) layer and one Softmax function are employed for data flattening and

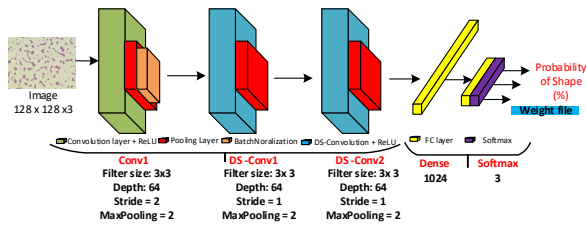


Fig. 3. The proposed architecture of DS-CNNs.

bacteria shape classification. In addition, we also insert Batch Normalization and Dropout layers to normalize data and avoid over-fitting during training. Table I also defines the model architecture based on some characteristics, including Layers, Filter Shape, Parameters, and Multiply-Accumulate (MACs). The total parameters and MACs are 0.27M and 13.2M, respectively.

TABLE I
THE PROPOSED ARCHITECTURE SPECIFICATION

Layer	Type	Filter Shape	Parameters	MACs
Conv1	Conv/stride 2	3 x 3 x 3 x 64	1792	7077888
	Batch Norm	-	256	0
	Max Pooling	Pool 2 x 2	0	0
DS-Conv2	DW-Conv	3 x 3 x 1 x 64	640	589824
	PW-Conv	1 x 1 x 64 x 64	640	4194304
	Max Pooling	Pool 2 x 2	0	0
DS-Conv3	DW-Conv	3 x 3 x 1 x 64	640	147456
	PW-Conv	1 x 1 x 64 x 64	640	1048576
	Max Pooling	Pool 2 x 2	0	0
FC4	Fully Connected	1024	66560	1332096
Classifier	Softmax	3	19661	0
Total			266499	13.2M

III. MATERIAL & EXPERIMENTS SET UP

A. Datasets Description

In our experiments, the used datasets are derived from DIBaS [15] which is publicly available and consists of 689 images divided into 33 bacterial species. Each bacteria type has about 20 photos with a high resolution of 2048×1532 pixels. Several bacteria-shaped images in our dataset are shown in Figure 4. In order to process the data in accordance with the training model, we deploy some computer vision methods to resize and augment (flip, rotate, shift, zoom...) image quantity to increase accuracy. As a result, the dataset has 1669 images with dimensions of $128 \times 128 \times 3$.

B. Experimental Setup

The model is trained and tested under the computational specification of 64-bit Windows 10, with Intel Core i9 processor (3.6 GHz), 32GB Random Access Memory (RAM), and NVIDIA GeForce RTX 2080 Graphic card. The Python programming language based on Tensorflow framework [16] and Keras libraries [17] are utilized. The dataset after preprocessing has a result of 1669 bacteria shape images, then 1335 images (80%) are allocated for the training set and the remaining 334 ones (20%) for the test set. The learning rate hyperparameter, which controls the speed of

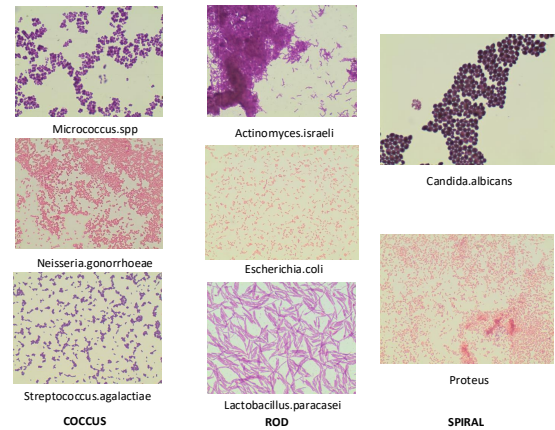


Fig. 4. Sample microscopy images with coccus-shaped, rod-shaped and spiral shaped of bacteria derived from DIBaS dataset.

weight update, is randomly set to 10^{-3} , and we randomly initialize the weight of filters and allow for updating weights automatically.

IV. RESULTS

The proposed model is trained with several scenarios using four popular optimizers used with 100 and 200 epochs.

TABLE II
TESTING ACCURACY (%) ON DIBAS DATASET FOR VARIOUS OPTIMIZERS

Models	Epochs	Adam	Adamax	Nadam	RMSprop
This work	100	94.01	91.43	94.31	94.61
This work	200	97.01	91.62	96.7	95.81

Table II shows the test accuracies (%) on our custom datasets for various optimizers as well as the number of epochs for training. As the results of the description table, the accuracies of all the optimizers are more than 91%. RMSprop optimizers training with 100 epochs and Adam optimizers with 200 epochs attained the highest accuracy of 94.61% and 97.01%, respectively.

TABLE III
COMPARE WITH SEVERAL COMMON CNNs ARCHITECTURES

Model	Layers	Parameters (M)	MACs (M)	Memory (MB)	Accuracy (%)
AlexNet	8	37.36	441	7.60	91.72
VGG - 16	16	65.07	5130	19.92	94.6
ResNet-50	50	23.69	1125	11.47	99.4
MobileNet	28	6.02	185	6.79	90.62
This study	5	0.27	13.2	1.93	97.01

Table III presents the model parameters, MACs, and memory usage of various popular CNN models when training on our custom dataset. It is clear that with the lower resources we use, the classification accuracy is higher than that of numerous complex and multilayer models. The number of parameters (0.27M) is far smaller than the conventional

TABLE IV
TEST RESULTS OF THE PROPOSED ARCHITECTURE

	Coccus	Rod	Spiral
No. Test Images	112	116	106
Incorrect	4	4	3
Sensitivity (%)	96.43	96.55	97.17

CNNs such as AlexNet (37.36M), ResNet-50 (23.69M), or even the compact ones like MobileNet v1 (6.02M).

The testing results of 334 images for three bacteria shapes: coccus, rod, and spiral are described in Table IV.

Fig. 5 and Fig. 6 depict the validation accuracy of the proposed model on the custom dataset derived from DIBaS when training model with 100 and 200 epochs to verify the data on Table II.



Fig. 5. Validation accuracy using Adam optimizer.



Fig. 6. Validation accuracy using RMSprop optimizer.

V. CONCLUSIONS

In this paper, a lightweight model for bacteria shape classification was constructed and discussed. The proposed method works well to classify three main shapes of bacteria, including rod, coccus, and spiral, with a classification accuracy of 97% and using lower parameters than other

traditional CNNs. This rapid classification is vital for the following stages to distinguish bacteria between different strains from microscopy images automatically. This method is enormously beneficial to all institutions working in microbiology, where results of medical experiments will be executed automatically and rapidly.

REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [2] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556 [cs], Apr. 2015.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.
- [4] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [5] J. Liu, F. Dazzo, O. Glagoleva, B. Yu, and A. Jain, "CMEIAS: A Computer-Aided System for the Image Analysis of Bacterial Morphotypes in Microbial Communities," *Microbial Ecology*, 2001.
- [6] M. Woźniak, D. Połap, L. Kośmider, and T. Clapa, "Automated fluorescence microscopy image analysis of *Pseudomonas aeruginosa* bacteria in alive and dead stadium," *Engineering Applications of Artificial Intelligence*, vol. 67, pp. 100–110, Jan. 2018.
- [7] D. Połap and M. Wozniak, "Bacteria shape classification by the use of region covariance and Convolutional Neural Network," *IJCNN*, 2019.
- [8] A. Ferrari, S. Lombardi, and A. Signoroni, "Bacterial colony counting by Convolutional Neural Networks," in *37th EMBC*, pp. 7458–7461. Milan, 2015.
- [9] L. Huang and T. Wu, "Novel neural network application for bacterial colony classification," *Theoretical Biology and Medical Modelling*, vol. 15, no. 1, p. 22, Dec. 2018.
- [10] "CS231n Convolutional Neural Networks for Visual Recognition." [Online]. Available: <https://cs231n.github.io/transfer-learning/>
- [11] M. F. Wahid, M. J. Hasan and M. S. Alom, "Deep Convolutional Neural Network for Microscopic Bacteria Image Classification," *5th ICAEE*, pp. 866-869, Bangladesh, 2019.
- [12] M. Talo, "An Automated Deep Learning Approach for Bacterial Image Classification," *ICATCES2019*, Turkey, 2019.
- [13] H. Wang, Q. Zhang, H. Lu, D. Won, and S. W. Yoon, "3D Medical Image Classification with Depthwise Separable Networks," *Procedia Manufacturing*, vol. 39, pp. 349–356, 2019.
- [14] L. Sifre, "Rigid-motion scattering for image classification," Ph.D. dissertation, Ecole Polytechnique, CMAP, France, 2014.
- [15] B. Zielinski, A. Plichta, K. Misztal, Spurek P, Brzywczy-Wóoch M, Ochońska D, "Deep learning approach to bacterial colony classification," *PLOS ONE*, Vol 12, No 9, p. 14, Aug. 2017.
- [16] "TensorFlow." [Online]. Available: <https://www.tensorflow.org/>
- [17] "Keras: the Python deep learning API." [Online]. Available: <https://keras.io/>