Breakthrough Conventional Based Approach for Dog Breed Classification Using CNN with Transfer Learning

Punyanuch Borwarnginn, Kittikhun Thongkanchorn*,
Sarattha Kanchanapreechakorn, Worapan Kusakunniran
Faculty of Information and Communication Technology, Mahidol University, Nakhon Pathom, Thailand
Email: punyanuch.bor@student.mahidol.ac.th, kittikhun.tho@mahidol.ac.th,
sarattha.kar@student.mahidol.ac.th, worapan.kun@mahidol.ac.th

***Abstract*—Dogs are one of the most common domestic animals. Due to a large number of dogs, there are several issues such as population control, decrease outbreak such as Rabies, vaccination control, and legal ownership. At present, there are over 180 dog breeds. Each dog breed has specific characteristics and health conditions. In order to provide appropriate treatments and training, it is essential to identify individuals and their breeds. The paper presents the classification methods for dog breed classification using two image processing approaches 1) conventional based approaches by Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG) 2) the deep learning based approach by using convolutional neural networks (CNN) with transfer learning. The result shows that our retrained CNN model performs better in classifying a dog breeds. It achieves 96.75% accuracy compared with 79.25% using the HOG descriptor.**

Keywords— dog breed classification, LBP, HOG, transfer learning, deep learning

I. Introduction

Dogs are the most common domestic animals. Currently, there are approximately 900 million dogs in the world, which 75 % to 85% of them are not domestic dogs such as stray dogs and wild dogs [1]. In 2016, the Thai Bureau of Disease Control and Veterinary Medicine reported that there are 7.5 million dogs, of which 750,000 are non-owned dogs. Due to a large number of dogs, there are several issues such as population control, decreases outbreak such as Rabies, vaccination control, and legal ownership.

At present, there are over 180 dog breeds, according to the American Kennel Club [2]. Each dog breed has specific characteristics and health conditions. In order to provide appropriate treatments and training, it is essential to identify a dog breed. Dog breed identification has traditionally been carried out by experts. However, it could take time to evaluate each dog.

Various image processing techniques have been studied for dog breed classification. Chanvichitkul et al. [3] proposed a method to classify dog breed using dog face images. They presented a contour based classifier together with a Principle Component Analysis (PCA) based classifier. Parkhi et al. [4] introduced a model to classify 37 different breeds of cats and dogs using a combination of shape and appearance features. Liu et al. [5] proposed a method for dog face detection using SVM regressor with grayscale SIFT and build geometric and appearance model of their face parts (eyes and noses) to detect dog breed. The model achieved a 67% recognition rate across 133 breeds.

This paper presents the evaluation performances of dog breed classification using two image processing approaches 1) conventional based approaches by Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG) 2) the deep learning based approach by using pre-trained convolutional neural networks (CNN) with transfer learning. The result shows that our retrained CNN model performs better in classifying dog breeds. It achieves 96.75% accuracy compared with 79.25% using the HOG descriptor.

The remainder of this paper is organized as follows. Section II describes the conventional based approaches and the deep learning based approach is presented in Section III. Section IV explains our experiment settings. The results and discussion are given in Section V. Finally, the conclusion is drawn in Section VI.

II. Conventional Based Approach

In this section, we describe an overview framework of dog breed classification using conventional based approach. Given the input dog face image, and it is pre-processed such as grayscale conversion and histogram equalization to enhance the image. Then each feature extraction technique is applied to capture primitive texture patterns. Finally, extracted features are used to build a classifier using a support vector machine (SVM) to recognize dog breed.

A. Local Binary Pattern

Local Binary Pattern (LBP) is a texture descriptor which is widely used various application including face recognition [6, 7]. The LBP algorithm assigns labels to every pixel of an image using the threshold of neighbor pixels around the center pixel value and forms a binary number. If the grayscale value of the neighboring pixel is higher or equal to the value of the center pixel, the value is set to be one. Otherwise, it is set to be one as in (2). It is based on the setting parameters (P, R) where P is a number of neighbors around radius R (see Fig. 1). The binary pattern is transformed into to decimal value for each pixel as in (1) where is a grayscale value of the center pixel and is a grayscale value of circularly symmetric neighbor pixels from $p = 0, \dots, P - 1$.

*Corresponding author

978-1-7281-4055-1/19/$31.00 ©2019 IEEE

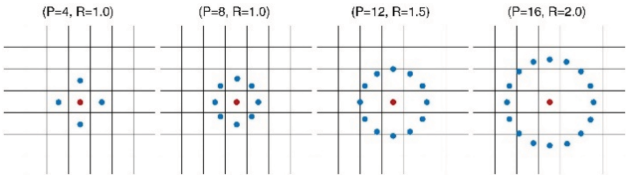


Fig. 1. Circular positons of neighbor pixels with different settings $(\boldsymbol{P}, \boldsymbol{R})$ [8]

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{1}$$

Where

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

After all pixels are computed with the LBP algorithm, the histogram of the LBP image is generated and used as a texture descriptor.

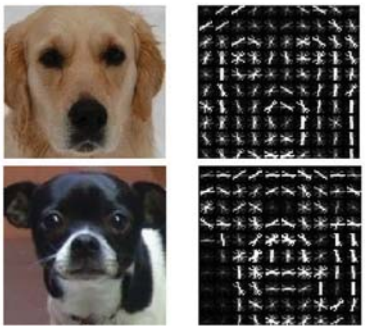


Fig. 2. The original images with their HOG descriptor

B. Histogram of Oriented Gradient

Histogram of Oriented Gradient (HOG) is also one of the widely used descriptors for object detection [9]. The HOG descriptor is based on the distribution of directions of gradients (see Fig. 2). A HOG feature is calculated as the following steps:

1. Calculate the gradient images of the grayscale image
2. Calculate the histogram of gradients in $n \times n$ cells. The image is divided into $n \times n$ pixels patches. For each pixel, there are the magnitude and direction values. Using 9-bin histogram corresponding to gradient direction values (0, 20, 40 … 160), the magnitude values of the pixels are accumulated and store in each bin based on their gradient direction.
3. Calculate a block normalization. For example, 16×16 block normalization of 8×8 cells means for each block includes 4 patches. Normalization divides each bin's value with the size of the block vector. Therefore, each block is represented by the normalized vector of 36 elements (9 bins × 4 patches)
4. Calculate the HOG feature vector by concatenating the vectors from all blocks.

III. Convolutional Neural Network

At present, the convolutional neural networks have been widely used in image classification. Several CNN architectures achieve high performance on Large Scale Visual Recognition Challenge [10]. However, it requires a lot of computational resources to train the network from scratch. Therefore, there is a widely used method called transfer learning that uses the feature vectors on pre-trained networks on a large dataset such as ImageNet [11] and COCO [12]. Then we can retrain a new classification on top using our dataset. In this paper, we choose the Inception V3 network to demonstrate the performance of deep learning based approach. Inception V3 [13] was the first runner up for image classification in ILSVRC2015 [10] by Google. The overview of transfer learning using Inception V3 is shown in Fig. 4. We reuse the parameters of the feature extraction part, remove the fully connected layer (Bottleneck layer) and then retrain the layer with a new dataset to produce feature vectors. Finally, the softmax layer is trained using these vectors and result as a new classifier. In this case, we apply dog breeds as the new categories and input dog face images as the input image.

IV. Experiments

A. Dataset

The Columbia dog with parts dataset is used for evaluating our models [5]. The dataset contains 8,351 dog images of 133 breeds by the American Kennel Club with 8 part locations annotated for each image. The sample images are shown in Fig. 5.

B. Conventional Based Approach VS Deep Learning

In this experiment, we select 20 breeds of 133 breeds from the dataset. We use the constraint that each breed contains more than 60 images. We segment dog face images from the ground truth provided and normalize them to 100 × 100 pixels. For each breed, 40 images are used for training and the other 20 images are used for testing. With this dataset, several scenarios of the experiments are explored as follows:

1) Histogram of LBP

In this experiment, several LBP settings are selected to evaluate. Fig. 3 shows the sample LBP images of a dog face. For each image, histograms of LBP images are calculated and used as a feature vector. During the training process, we compare the result from 128-bin histogram and 256-bin histogram. The 128-bin histogram returns a higher result.

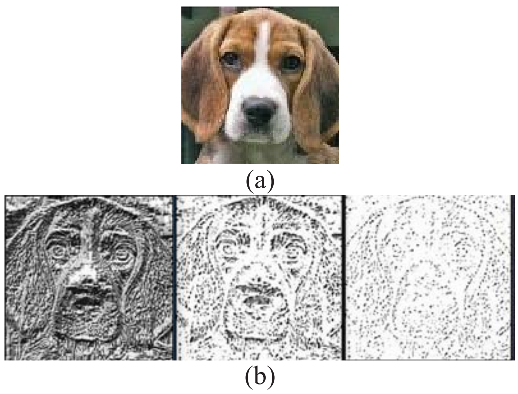


Fig. 3. (a) the Original image (b) Sample LBP images with different settings. From left to right $(\boldsymbol{P} = \mathbf{8}, \boldsymbol{R} = \mathbf{2})$, $(\boldsymbol{P} = \mathbf{12}, \boldsymbol{R} = \mathbf{3})$, $(\boldsymbol{P} = \mathbf{24}, \boldsymbol{R} = \mathbf{3})$

2) Histogram of LBP on sub-images

As the previous experiment, we select histogram of LBP, but the histogram is calculated on the sub-images instead of the whole image, as shown in Fig. 6. We calculate histogram for each sub-image and all histograms of sub-images are concatenated into a single feature vector. The histogram of 128 bins and 256 bins are evaluated with 4 and 9 sub-images. The 128-bin histogram remains a higher result.

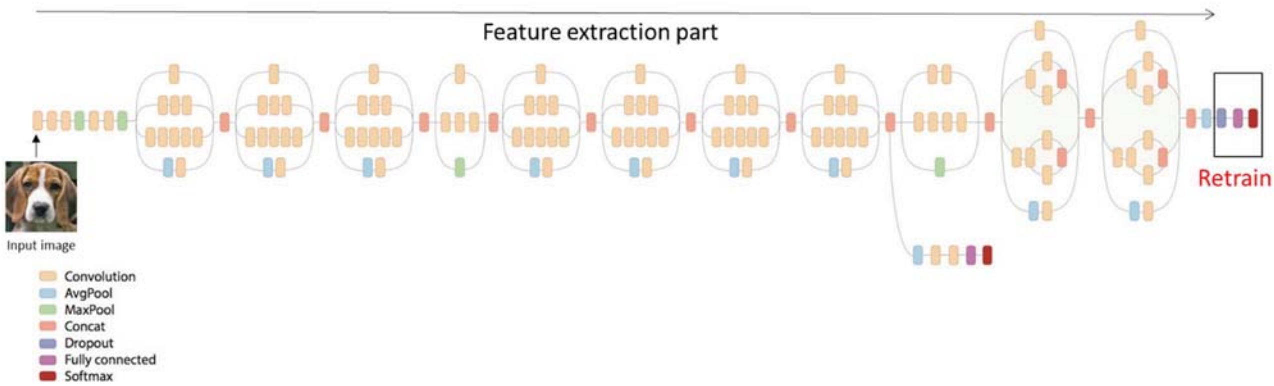


Fig. 4. An overview of transfer learning using Inception V3 model. Transfer learning uses the feature extraction part from a trained model and retrains a new classification on the top layers. Reproduced from [14]

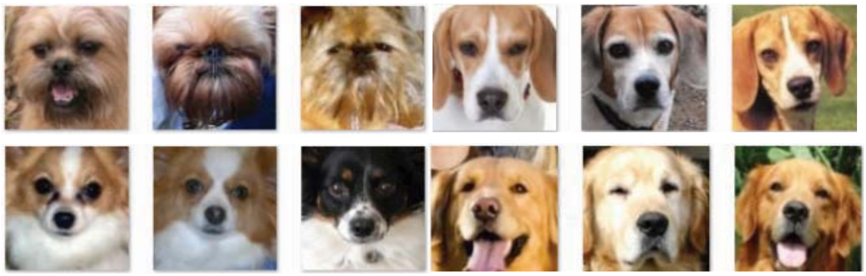


Fig. 5. Sample images of the dataset

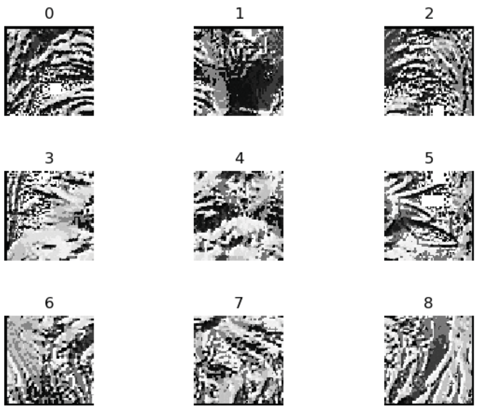


Fig. 6. An LBP image is divided into 9 sub-images.

3) Histogram of multi-channel LBP on sub-image

In this experiment, we combine 3 LBP settings $(P = 8, R = 2), (P = 12, R = 3), (P = 24, R = 3)$. This technique is similar to multi-channel histogram which an image has different colour information. On the other hand, each LBP setting captures different texture patterns. Each LBP settings are concatenated into a single feature vector.

4) Histogram of Oriented Gradient

In this experiment, HOG settings are 10×10 cells with 20×20 block size with 9-bin histogram.

5) Combination of features

In this experiment, we investigate the performance of combining different local features as image features. For example, a histogram of multi-channel LBP with HOG is selected as the feature vectors.

6) Inception V3 with transfer learning

In this experiment, the Inception V3 network is selected as the CNN architecture as described earlier. We retrain the network using Tensorflow library with 1,000 training steps. Each step chooses 100 random images from the training set and 88 images from the validation set. Fig. 7. shows a comparison of the training accuracy and the validation accuracy for each step and the loss function.

V. Results

For the traditional approach, we use a nonlinear histogram intersection SVM as a classifier and evaluate using 200 images. HOG achieves 79% accuracy. On the other hand, our retrained Inception V3 classifier achieves the high performance with 96.75% accuracy rates as shown in Table I. For each training step, 10 percent of training images are used as the validation set. We use cross-entropy as the loss function, and it decreases during each training step as a result shown in Fig. 7.

TABLE I. Comparison of different results

Technique	**Accuracy (%)**
LBP $(P = 8, R = 2)$	16.25
Multi-Channel LBP	23.00
LBP with 9 Sub Images $(P = 8, R = 2)$	44.00
Multi-Chanel LBP with 9 Sub Images	56.00
HOG	**79.25**
Multi LBP settings with HOG	62.25
CNN (Inception V3)	**96.75**

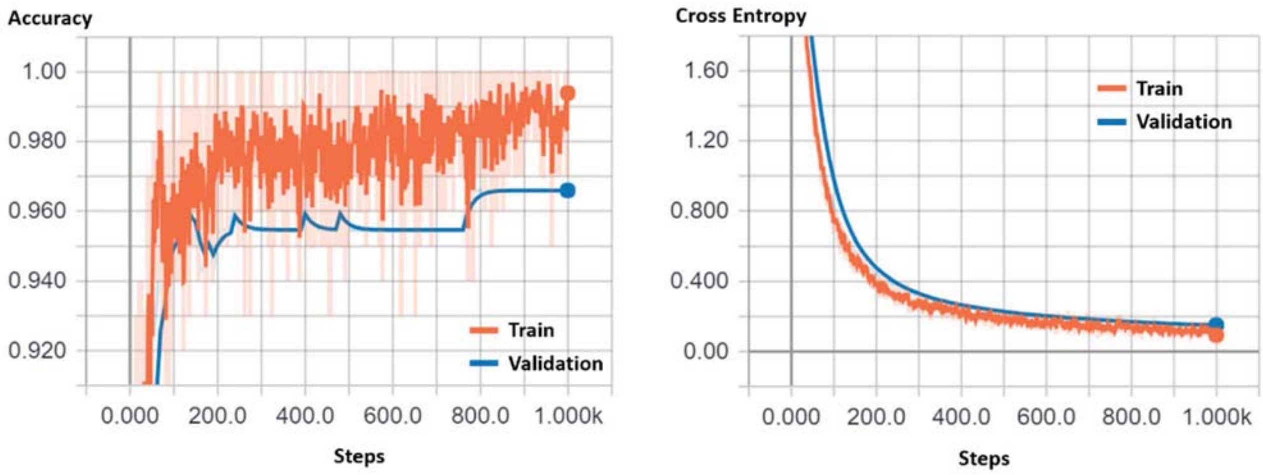


Fig. 7. Training and validation accuracy and Cross Entropy.

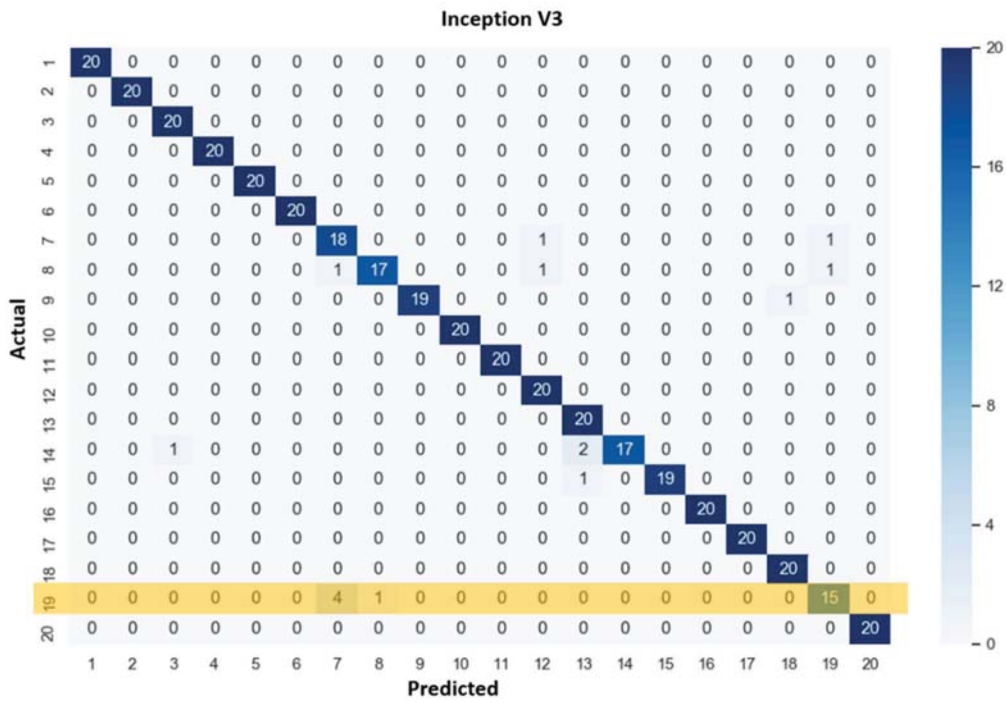
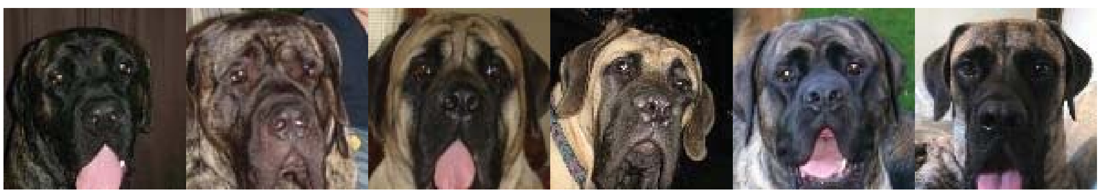


Fig. 8. : Confusion matrix of retrained Inception V3.



(a)

(b)

Fig. 9. The example of misclassified classes (a) Class 19: Mastiff (b) Class 7: Bullmastiff.

Furthermore, the confusion matrix (Fig. 8) demonstrates that the retained network can mostly categorize dog breeds from the test set. However, 4 images of class 19 (Mastiff) were misclassified as class 7 (Bullmastiff). Bullmastiff is a cross breed between Mastiff and Bulldog. It shares similar appearances with Mastiff (see Fig. 9), which might confuse the network. In order to improve the performance, similar breeds can be grouped together and classify using the coarse-to-fine concept [3] to reduce the number of class in the first prediction. Later, the model is narrowed down to these similar breeds.

Since the current experiment focuses on 20 breeds from the total 133 breeds and achieves the significant result, the experiment is extended by using all 133 breeds. In the case, we select 3 pre-trained CNNs to implement the classifier which are InceptionV3[13], MobileNetV2[15], and NASNet[14]. Table 2 reports that pre-trained CNNs can achieve high accuracy rates. The best performance is 91% from NASNet.

TABLE II. ACCURACY RESULTS FOR 133 BREEDS

Network	Accuracy (%)
InceptionV3	89.50
MobilenetV2	89.60
NASNet	91.00

VI. CONCLUSION

This paper presents an improvement of the dog breed classification from the conventional based approach to the deep learning based approach. The experiment showed that using CNN with transfer learning achieves a significant accuracy of 96.75% while the conventional based approach is at 79.25%. Therefore, it is proven that CNN could be used on the dog breed classification.

ACKNOWLEDGMENT

This research has been supported by the Royal Golden Jubilee (RGJ) Ph.D. Programme under the Thailand Research Fund (Grant No. PHD/0053/2561) and Faculty of Information and Communication Technology, Mahidol University.

REFERENCES

[1] "How Many Dogs Are There In The World?" [Online]. Available: https://www.worldatlas.com/articles/how-many-dogs-are-there-in-the-world.html. [Accessed: 17-Mar-2019].

[2] "AKC Breeds by Year Recognized." [Online]. Available: https://www.akc.org/press-center/articles/breeds-by-year-recognized/. [Accessed: 17-Mar-2019].

[3] M. Chanvichitkul, P. Kumhom, and K. Chamnongthai, "Face recognition based dog breed classification using coarse-to-fine concept and PCA," in *2007 Asia-Pacific Conference on Communications*, pp. 25–29.

[4] O. M. Parkhi, A. Vedaldi, C V Jawahar, and A. Zisserman, "The truth about cats and dogs," in *2011 International Conference on Computer Vision*, 2011, pp. 1427–1434.

[5] J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur, "Dog Breed Classification Using Part Localization," in *European Conference on Computer Vision*, Springer, 2012, pp. 172–185.

[6] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[8] Y. Quan, Y. Xu, and Y. Sun, "A distinct and compact texture descriptor," *Image and Vision Computing*, vol. 32, no. 4, pp. 250–259, 2014.

[9] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[12] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, no. PART 5, Springer, Cham, 2014, pp. 740–755.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[14] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8697–8710.

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.