

Paper Jurnal/Prosiding

by Rani Purbaningtyas

Submission date: 10-May-2023 02:34PM (UTC+0700)

Submission ID: 2089305977

File name: 1-s2.0-S2214317322000026-main_1.pdf (1.76M)

Word count: 8003

Character count: 41888



Combining MobileNetV1 and Depthwise Separable convolution bottleneck with Expansion for classifying the freshness of fish eyes

Eko Prasetyo^{a,b,*}, Rani Purbaningtyas^a, Raden Dimas Adityo^a, Nanik Suciati^b, Chastine Fatichah^b

^a Department of Informatics, Engineering Faculty, Universitas Bhayangkara Surabaya, Jl. Ahmad Yani 114, Surabaya 60234, Indonesia

^b Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Jl. Raya ITS, Surabaya 60111, Indonesia

ARTICLE INFO

Article history:

Received 2 March 2021

Received in revised form

1 January 2022

Accepted 18 January 2022

Available online 22 January 2022

Keywords:

Depthwise separable convolution

Bottleneck

Classification

Freshness

Fish eye

Residual transition

ABSTRACT

Image classification using Convolutional Neural Network (CNN) achieves optimal performance with a particular strategy. MobileNet reduces the parameter number for learning features by switching from the standard convolution paradigm to the depthwise separable convolution (DSC) paradigm. However, there are not enough features to learn for identifying the freshness of fish eyes. Furthermore, minor variances in features should not require complicated CNN architecture. In this paper, our first contribution proposed DSC Bottleneck with Expansion for learning features of the freshness of fish eyes with a Bottleneck Multiplier. The second contribution proposed Residual Transition to bridge current feature maps and skip connection feature maps to the next convolution block. The third contribution proposed MobileNetV1 Bottleneck with Expansion (MB-BE) for classifying the freshness of fish eyes. The result obtained from the Freshness of the Fish Eyes dataset shows that MB-BE outperformed other models such as original MobileNet, VGG16, Densenet, Nasnet Mobile with 63.21% accuracy.

© 2022 China Agricultural University. Production and hosting by Elsevier B.V. on behalf of KeAi. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Fish is a product that significantly contributes to the global economy and trade [1] due to its delightful flavor and high nutrient contents [2], such as protein, unsaturated fatty acids, minerals, and vitamins [1]. This product is part of the daily

staple food of Indonesians consumed with rice and vegetables. According to a survey, fish is ranked fourth among the highly consumed side dishes after processed food, beverages, cigarettes, and grains [3]. Presently, some vendors sell both fresh and not fresh fish stored with ice at varying temperatures. The freshness of fish is inspected using sensory cues such as visual appearance, texture, sound, taste, and smell [4]. However, it is almost impossible for people to recognize the freshness of fish using a detector due to its cumbersome size, cost, and timeframe. Therefore, a system capable of automatically recognizing the freshness of fish quickly and easily, without destroying it, is needed. Non-destructive

* Corresponding author at: Department of Informatics, Engineering Faculty, Universitas Bhayangkara Surabaya, Jl. Ahmad Yani 114, Surabaya 60234, Indonesia.

E-mail address: eko@ubhara.ac.id (E. Prasetyo).

Peer review under responsibility of China Agricultural University.

<https://doi.org/10.1016/j.inpa.2022.01.002>

2214-3173 © 2022 China Agricultural University. Production and hosting by Elsevier B.V. on behalf of KeAi.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

freshness classification is conducted using an image-based automatic system by processing the visual features in an image. Due to the limitation of visual appearance, this system uses the eyes, skin, and tail of fish to determine its freshness. This study focused on the fish freshness classification based on eyes appearance.

Preliminary studies carried out to determine the eye differences of fish freshness showed that the visual appearance of both fresh and not fresh fish are different [4]. Several studies have been carried out to determine the fish freshness classification, such as developing a system based on eyes and gills during a 10-day ice-storage cycle using color features and an artificial neural network [2], location detection and counting fish numbers in the sea with extreme background variation [5]. The recent study in fish freshness classification used the Cosine Nearest Neighbor and eyes color features comprising 71 images showing an accuracy of about 60.89% [6,7]. In addition, a system for classifying common carp fish using a deep convolutional neural network (CNN), the accuracy reaches up to 98.21% [8]. The CNN technique is one of the common processes used to classify fish because it is cost-efficient, precise, non-destructive, automated, and produces real-time answers. This technique has also been combined in numerous studies, for example, CNN for age range classification from unconstrained face images [9], insect classification [10], and fruit grading [11]. Several CNNs were previously developed, such as the residual network (ResNet) for maintaining lower-level features using a skip connection [12] and VGGNet for reducing the convolutional kernel size [13]. Furthermore, Densenet was developed for skip connection with bottleneck [14], MobileNet for a small classifier suitable for mobile devices [15], and Nasnet with a search space transferable from low to a high dataset [16]. The MobileNetV1 was significantly enhanced using an inverted residual and linear bottleneck called MobileNetV2 to simplify the architecture further and improve the performance [17]. Although this system succeeded in simplifying the MobileNetV1 architecture, the performance achieved in the fish freshness classification was inadequate. Therefore, this research proposes a new architecture to resolve fish's freshness classification based on its eyes. In addition, PFS confirms that the eyes of fresh and not fresh fish can be distinguished with varying notions. Visual appearance is dominant in the analysis where the naked eye distinguishes fresh and not-fresh fish with difficulty. Consequently, the automatic system further does not have useful features to distinguish them.

As discussed earlier, MobileNet [5] delivers a vast decreasing number of parameters by moving from the standard to the depthwise separable convolution (DSC) paradigm. The standard convolution involves kernel size and input-output channels, sequentially broken down into DSC and pointwise with 1×1 kernel. DSC performance is slightly below the standard convolution from the empirical studies results, with relatively lighter parameters. Therefore, this smart concept presents a trade-off where the number of parameters is massively reduced with a slight drop in performance than the big model [17]. Slightly lower performance is still acceptable with a significantly smaller size model than the big model with slightly better performance. MobileNetV1 uses model parameters of

more than three million parameters at a width multiplier of 1.0, classic architectural paradigm, and plain convolution flow (PCF). This process is obtained from the initial to the final layer, using stratified, and straight, no skip connection [12], without bottleneck [17], inception [18], cross-stage partial [19], hierarchical spatial features [20], and additional image quality analysis [21]. Although PCF consists of architectural simplicity, its model is ineffective for learning dataset features. For instance, it increases accuracy during training, gets saturated, and drops fast [12] and increases the number of channels with binary-fold incremental inflict in higher memory traffic [19]. The channel number of MobileNet from initial to final layer is a binary-fold consisting of 32-64-128-256-512-1024 incremental parameters. Its size reduces models with performance similar to standard convolution, such as ResNet and VGGNet. In this research, MobileNetV1 uses fewer parameters to solve general classification problems, such as ImageNet, which becomes less optimal [8] when classifying the freshness of fish eyes. Therefore, this research proposed a CNN architecture that inherits the smart DSC idea from MobileNetV1, which is more effective in learning the fish eye features for freshness classification.

As explained earlier, the non-destructive fish freshness examination is conducted using an image-based automatic system. Although CNN promises optimal classification performance, it requires a vast number of image data and a wide variety of images. Several datasets, such as Caltech-101 [22] or Coil-100 [23], have many variations in scaling, rotation, shearing, color, lighting, viewpoint, and image quality. However, the eyes of fresh and not fresh fish do not have an adequate variety of images. Hence, the visual features between the internal and external classes are insufficient. Conversely, the slight differences in features do not require complicated CNN architecture. Therefore, to address this problem, a new CNN architectural system was proposed with the following contributions

1. Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE)

This technique is a bottleneck and expansion convolution for improving feature quality and generating more detailed features with non-linear functions. It is organized using the Depthwise Separable Convolution (DSC) by convoluting feature maps with a Bottleneck Multiplier (BM) ratio to obtain fewer feature maps. Furthermore, it is re-convoluted to expand the feature maps to the original size. This technique also introduced BM as a constant to determine the bottleneck level with the performance trade-off and model size. The concept effectively improves features quality for both classification [24] and object detection [25].

2. Residual Transition

CNN architecture generally uses pooling for bridging the feature maps from one layer to the next with varying sizes. This technique is only used when one feature map acts as an input. On the other hand, a transition block involves such features map and skip-connection from the previous layer.

This process of maintaining low-level features is insufficient in utilizing pooling. This research aims to determine the Residual Transitions (RT) for bridging some feature maps from one convolutional block to another using different sizes. The feature maps from the previous block are combined with the current using depthwise convolution and concatenation, and then the size is changed from one current block size to the next using pointwise convolution.

3. MobileNetV1 with Bottleneck and Expansion (MB-BE)

The MobileNetV1 with Bottleneck and Expansion (MB-BE) was proposed in this research using some parts as the backbone and DSC-BE as an additional layer; this model was utilized to generate the fish's features eyes and reducing the number of parameters. It is difficult to distinguish the freshness of a fish's eyes due to the insufficient visual features, both internal within classes and external between classes. DSC-BE is utilized to generate proper features for classifying the freshness of fish eyes in the MB-BE architecture. Furthermore, it investigated the depth of DSC-BE to obtain the configuration of a convolution depth of the model according to the freshness classification of fish eyes.

2. Materials and method

2.1. Dataset

The Freshness of the Fish Eyes (FFE) dataset [26] were used to evaluate the proposed model's performance. This dataset consists of 4 392 images of fish eyes, with 8 fish species, and each comprises highly fresh fish (day 1 and 2), fresh fish (day 3 and 4), and not fresh fish (day 5 and 6), it can be seen here. The eight fish species are as follows Chanos Chanos (500 images), Johnius Trachycephalus (240 images), Nibea Albiflora (421 images), Rastrelliger Faughni (769 images), Upeneus Moluccensis (792 images), Eleutheronema Tetractylum (240 images), Oreochromis Mossambicus (625 images), and Oreochromis Niloticus (805 images). These fish were classified into 24 classes with eight species of fish and three levels of freshness. However, there were difficulties in differentiating the freshness of fish eyes due to no adequate features internal within classes and external between classes. Image acquisition was carried out by treating the fish with storage in a styrofoam box for six days to adjust the ice storage process in a ratio of 1:1. A mobile phone is used to photograph the fish daily with various backgrounds and lighting; each image contains the same species with a varying number of fish. After that, a deep learning model is utilized to locate the eye used as input images. The data collection process is shown in Fig. 1.

An ablation study was also conducted on the model to evaluate its effectiveness using the FFE dataset in classifying the freshness of fish eyes. The experiments also compare the performance between MB-BE and several CNNs, such as the original MobileNet V1, MobileNet V2, ResNet50, Densenet, VGG16, and Nasnet Mobile on the Freshness of the Fish Eyes (FFE) dataset. Other datasets were also used, such as Caltech 101 (9 000 images, 101 classes) [22] and Coil (7 200 images,

100 classes) [23], to prove its ability to classify fish's freshness and other problems.

2.2. Depthwise Separable convolution (DSC)

Convolution of f_{k-1} features map uses M number kernel kernels $W = \{w_1, w_2, \dots, w_M\}$ with filter size of $D_k \times D_k$ that is operated to an image $I \{W(i, j) \times I(x - i, x - j)\}$ to produce N number feature map or channel output f_k as follows:

$$f_k = \{w_1 \times I, w_2 \times I, \dots, w_N \times I\} \quad (1)$$

Standard convolution in Eq. (1) uses computation costs of $D_k \times D_k \times M \times N \times D_f \times D_f$, where D_f denotes the size of the result feature map, M and N denote the input and output channel, respectively. Meanwhile, the major change to the convolution method with minimal costs is Depthwise Convolution (DC), which uses a single kernel for each feature map [15]. For example, 664 single kernels are needed for 64 features by a 3×3 size, where each kernel is used to convolve one feature map. Convolution results used to determine the 64 features of one DC are as follows:

$$f_k = \{\bar{w}_1 \times I, \bar{w}_2 \times I, \dots, \bar{w}_N \times I\} \quad (2)$$

Where, \bar{w}_N is a single kernel with size of $D_k \times D_k$, and f_k is a features map.

Furthermore, the DC output is convoluted using a 1×1 kernel pointwise convolution known as Depthwise Separable Convolution (DSC), as shown in Fig. 2. The cost of DSC computation is much smaller than the standard convolution, namely $D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f$.

This research used the DSC as the basis for light convolution and to reduce the number of parameters in the DSC-BE and the RT of fish freshness classification. Convolution uses DSC-BE in each block to achieve high-level features, while RT was transitionally used between blocks.

2.3. Depthwise Separable convolution bottleneck with Expansion (DSC-BE)

ResNet [12] and ResNext [24] solve saturated performance problems during training using the skip connection in the bottleneck scheme. ResNext also added cardinality to determine the amount of decomposition per block. This method shows the bottleneck with the residual concept where the feature map is convoluted into fewer features followed by the next convolution into more features. In this research, bottleneck with expansion convolution is proposed using DSC (Depthwise Separable Convolution), instead of standard convolution, to be less feature map with the number of BM (Bottleneck Multiplier) \times feature map, where $0 < BM \leq 1$. The DSC used this process to expand the number of feature maps to the original size. This bottleneck concept effectively improves features quality for both classification [24] and object detection [25], while expansion is powerful for generating more detailed features with non-linear functions.

The Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE) is the mechanism used to improve feature quality and generate more detailed features with low computational costs in this research. DSC offers feature map results with classification performance slightly below

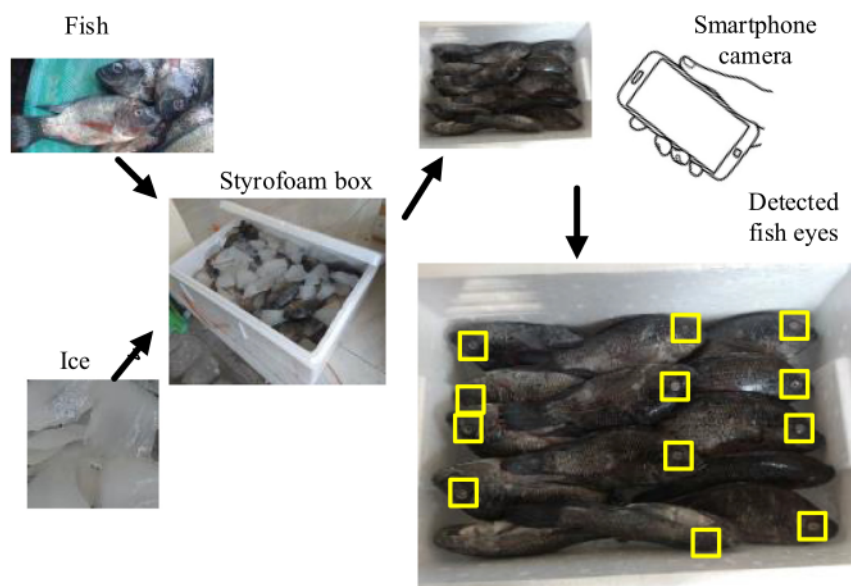


Fig. 1 – Fish eyes image acquisition.

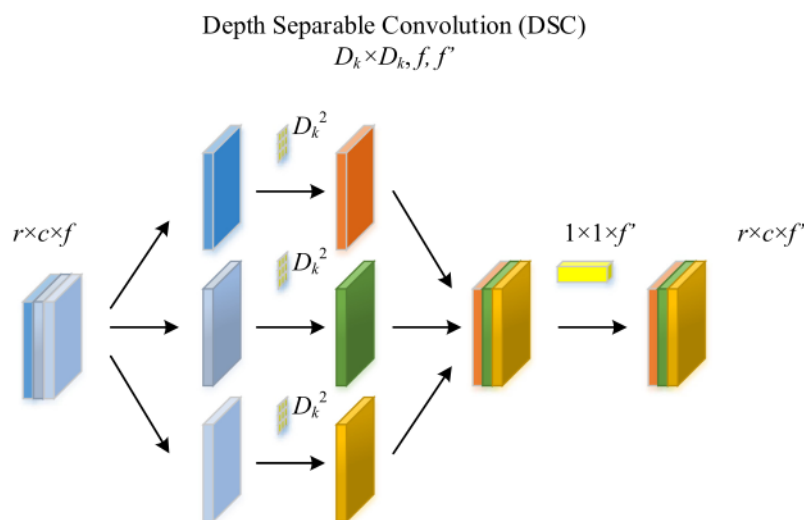


Fig. 2 – Depthwise Separable Convolution (DSC).

the standard convolution and low computational costs. Therefore, DSC-BM convolutes the feature map into a BM ratio measure because BM is a non-negative number below one, which is effective for scraping-out unimportant features. Furthermore, the DSC convolutes again with the same feature map size to generate more detailed features through non-linear transformation. In terms of usage, it requires kernel size parameters of $D_k \times D_k$ as depthwise convolution. However, this study uses the depthwise and pointwise convolution with kernel sizes of 3×3 and 1×1 , respectively.

This research also introduced BM as a constant for determining bottleneck levels with the trade-off of performance and model sizes. BM determines the convolution bottleneck-level; for example, using $BM = 0.5$ applies a bottleneck of half the number of feature maps. The smaller the BM, the higher the bottleneck occurs. If $BM = 1$, then there is no bottleneck, the convolution will become a plain convolution. Fig. 3 shows the conversion of the feature map size of $r \times c \times M$ to $r \times c \times N$. DSC-BE convolutes a bottleneck of BM, followed by a convolution of expansion to its original size.

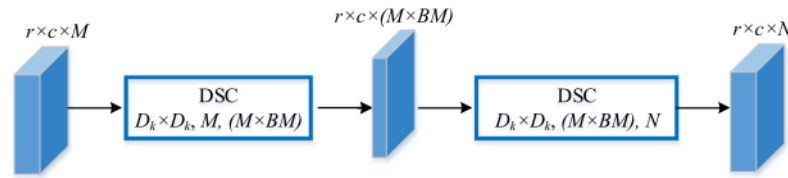


Fig. 3 – Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE).

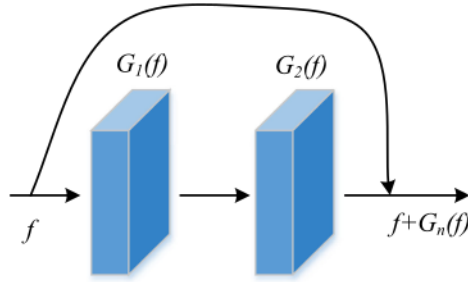


Fig. 4 – Identity mapping by ResNet [12].

2.4. Residual network

Residual Network (ResNet) is an architecture that introduces a concept of skip connection [12]. Saturated performance problems during training are resolved by adding residues from the previous convolution layer. ResNet's original concept allows a feature map to jump over several layers to join the certain layers, as shown in Fig. 4. This procedure is performed using identity mapping a feature map to the next layer using the adding operation. Supposing $G(f)$ is convolutional operation on a feature map f , then the residual network is calculated by adding identity mapping as follows:

$f_{k+1} = f + G_n(f)$ (3) Where $G_n(f)$ is n times convolutional operation conducted on a feature map f , f_{k+1} is the next feature map resulted from convolution. This method is proven to be able to avoid saturated performance problems when the training is conducted with a high number of epochs.

2.5. Residual transition

In most CNN architectures, such as Densenet and MobileNet, the model is divided into several blocks where each contains several layers of convolution with similar sizes and smaller resolutions. This changes the size of the feature map from one block to the next using transitions, as in Densenet, which uses 1×1 convolution on $2 \times 2/2$ max-pooling to obtain half of the resolution [14]. Therefore, this transition skips the connection from the previous block to a transition block to avoid saturated performance problems during training. Furthermore, additional padding or projection size [24] was used to maintain the feature map size.

The Residual Transition (RT) for bridging the skip connection in the transition block using 3×3 depthwise convolutions (DC) on the two feature maps was proposed in this research. The first feature map is the current convolution layer, and the second is the skip connection from the previous block. Furthermore, this research combined both maps through the concatenation and convolution process using 1×1 kernel, followed by batch normalization, relu activation, and max-pooling processes. The result is used in the next convolution block or passed to the fully connected layer, as shown in Fig. 5. The feature map f' with size $r \times c \times M$ is convoluted by $G_n(f)$ to be f . Both f and f' feature maps are involved together in the transition block. In addition, DC and PC are the mechanisms used to convolute them at low computational costs. Therefore, RT uses two DC and one PC to convolute them while keeping low computational costs. The next convolutional block would use the RT output as an input feature map.

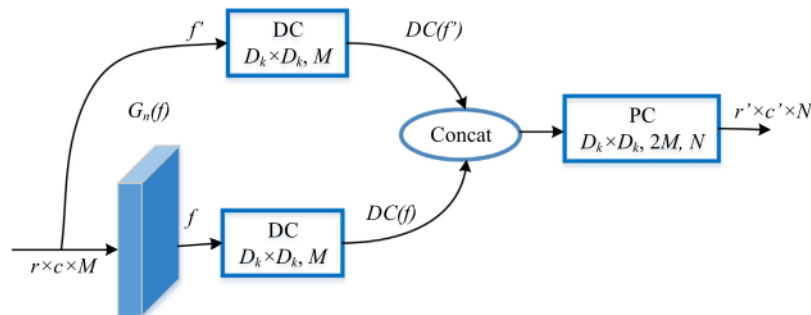


Fig. 5 – Residual Transition.

The RT formula is shown as follows:

$f_k = PC[DC(f') \cup DC(f)](4)$ Where f_k is the feature map result of RT operation, \cup denotes the union operation, f is the current feature map, and f' is the residual features from the previous block, DC and PC denote depthwise and pointwise convolution, respectively.

2.6. MobileNetV1 with bottleneck and Expansion (MB-BE)

MobileNet V1 convolution layer is classified into ten blocks, the first uses standard convolution, which produces 32 features, while the next block uses DSC and down-sampling with

max-pooling. The feature map increases by binary multiplication up to 1 024 features in last block. MobileNet V1 with Bottleneck and Expansion (MB-BE) is proposed as a new CNN architecture that partially inherits the architecture. Fig. 5 shows that among the ten blocks of MobileNet V1 architecture, the first six blocks (delimited by dotted lines) are inherited on MB-BE as the main backbone. The advantage of using some of the MobileNet V1 architecture is that it can be pre-trained as initial weights, using millions of images and thousands of classes. The pre-trained weights recognize various image features, which consists of significant advantages when trained using the initial weight of pre-trained MobileNet V1.

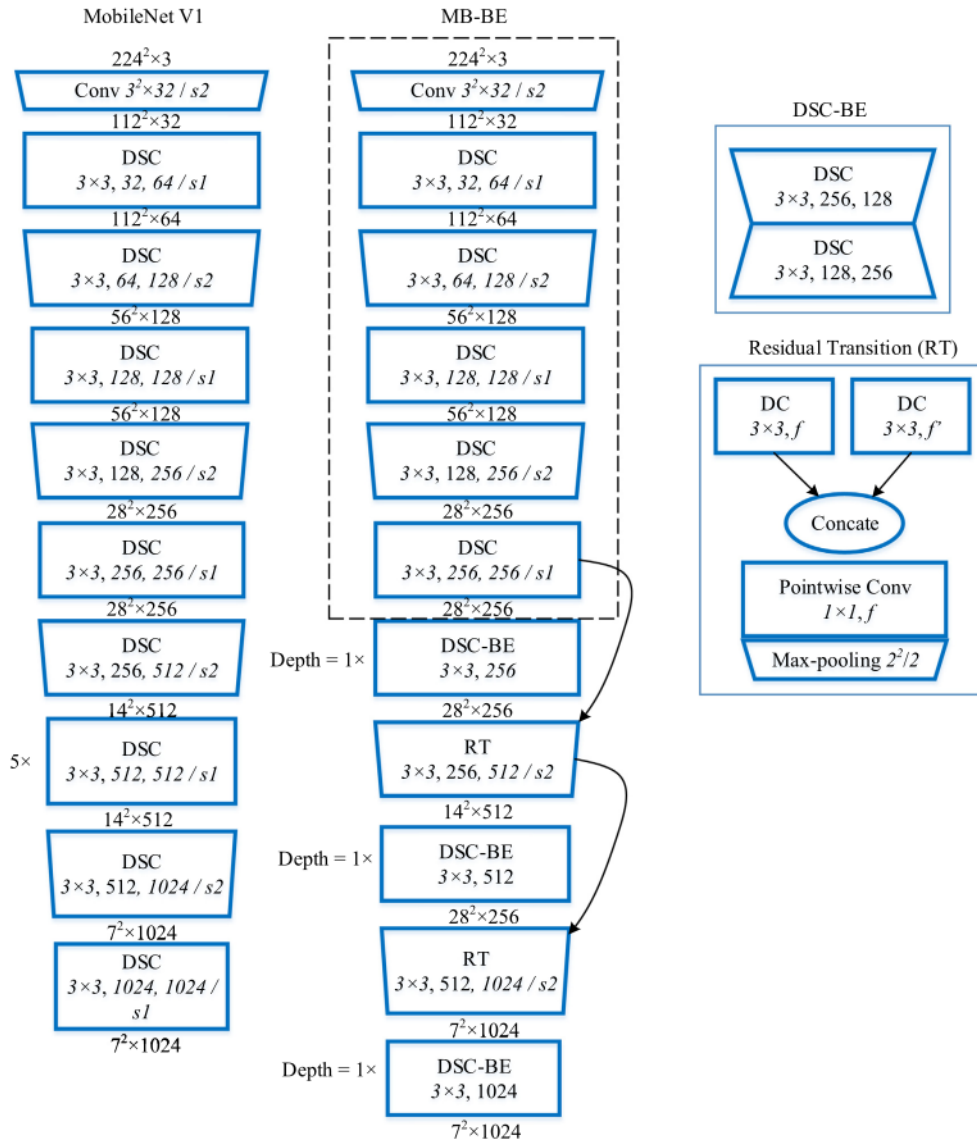


Fig. 6 – MobileNetV1 with Bottleneck and Expansion (MB-BE).

DSC is used as a convolutional basis for both Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE) and Residual Transition (RT). Three DSC-BE convolution blocks are added for the feature counts of 256, 512, and 1 024, respectively. As a feature map, the process of resizing transition, RT is used to bridge the changes by adding the residue from the previous block's end.

The four blocks of the MobileNet V1 convolution were removed and replaced with five new blocks to reduce the number of parameters and recognize the features of fish eyes. As explained before, the freshness of fish eyes is difficult to be distinguished because there are not many visual features, both internal within class and external between classes; then, DSC-BE in the MB-BE architecture is proposed to complete the correct feature for classifying the freshness of the fish eyes.

In terms of the number of parameters, the size of the MB-BE architecture depends on Bottleneck Multiplier (BM), as shown in Fig. 6. It is assumed when BM equals 0.5, a bottleneck convolution of 50% is carried out, followed by the expansion to the original size. Furthermore, in Fig. 5, DSC-BE 256 is convolved, which led to the evolution of 128 bottlenecks with 256 features expansion. In addition, an ablation study is con-

ducted to investigate the effect of BM on architectural measures and performance. The Bottleneck Multiplier with a value of 1.0 is also specified in the MobileNet V1 because it is the backbone of MB-BE.

The depth of DSC-BE is also considered because it determined the number of DSC-BE convolutions conducted. Depths values of 1 and 2 indicate that DSC-BE 1 and 2 times, etc. Three blocks pass using depth, as shown in Fig. 5. Furthermore, the effect of the depth of DSC-BE is investigated to determine the adequate value through an ablation study.

The model described earlier is a feature extraction block, whereas a classification block needs to be added in CNN. Furthermore, a fully connected layer with 1 024 neurons is added to the hidden and one output layer in accordance with the number of neurons according to the experiment's dataset.

2.7. Framework research classifying the freshness of fish eyes

The system framework for classifying the freshness of the fish eyes starts with segmentation, as shown in Fig. 7. This section is conducted using the object detection method with

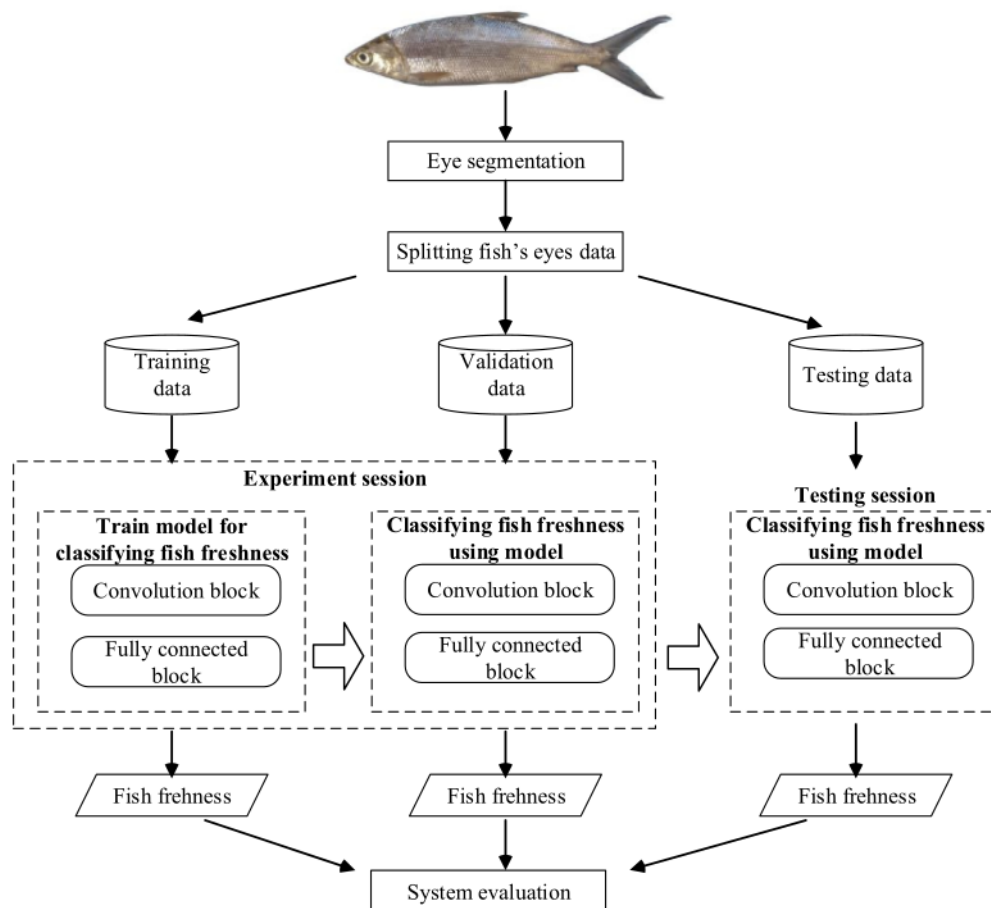


Fig. 7 – Framework Research Classifying the Freshness of Fish Eyes.

the images ¹⁰ put segmented into the fish eyes. A total of 4392 images in the dataset, divided into training, validation, and testing data were used to carry out this research. This data splitting process is explained in the next section. The proposed model ⁷ is trained and validated using the associated methods in the experimental session. Furthermore, the trained model ⁷ is used to classify the test data due to its importance in the reliability process in classifying images not seen during the experiment session. Furthermore, this study evaluated the classification results on all data with accuracy metrics and in-depth analysis.

2.8. Performance metric

This study classifies a multi-class problem consisting of 8 fish species and 3 levels of freshness, therefore a total of 24 classes were classified with Accuracy, Precision, Recall, and F1-score used to determine the performance metrics. Accuracy is used to measure the correct classification as follows.

Accuracy = $\frac{TP+TN}{TP+FN+FP+TN}$ (5) where TP, TN, FN, FP are true positive, true negative, false positive, and false negative of classification results, respectively. TP + TN denotes the number of images of fish species classified according to each species, TP + FN + FP + TN is the number of images classified through training, validation, and testing data. The Precision is utilized to evaluate the model's ability for avoiding misclassification based on positive predicted data in a class (TP) in accordance with the predicted results (TP + FP). The Precision Formula employed is as follows:

Precision = $\frac{TP}{TP+FP}$ (6) The FP is the class of species data that is failed to be detected as such species. The recall is used to evaluate the ability of the model to recognize a class; it is calculated based on positive predicted data in a class (TP) in accordance with all data that should be recognized as such class (TP + FN). The Recall formula is as follows:

Recall = $\frac{TP}{TP+FN}$ (7) The FN is the unrecognized data in a class, while F1-score is used to evaluate the overall model performance based on Precision and Recall. The formula is as follows.

$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ (8) In this study, Precision, Recall, and F1-score were calculated for each class and used to determine the average of all these metrics as the final performance.

3. Result and discussions

3.1. Experimental setting

This study was carried out using the pre-trained MobileNet V1 [15], which is publicly available, as a backbone architecture with trained weights from Imagenet to complete classifications of fish eye freshness. The model also used an input image of 224×224 pixels with a classification block consisting of a fully connected layer with 1024 neurons as the hidden layer. After adding an output layer to the number of neurons according to the classes in the dataset, 0.5 dropouts were further included. RMSProp with a learning rate of $1e-5$ and a loss function ¹⁰ categorical cross-entropy for the optimizer was utilized. The dataset was divided into training, validation, and testing with the respective percentages of 60, 20, and 20, respectively. For the FFE dataset, the authors used batch-size 24 and 22 for training and validation, respectively. Meanwhile, Caltech-101, and Coil-100 used batch sizes 30 and 31, as well as 44 and 14, respectively.

3.2. Ablation study

An ablation study was carried out to determine the appropriate configuration of the MB-BE parameters to complete the freshness classification of fish eyes. MB-BE used MobileNet V1 as the backbone model and connected to DSC-BE and RT thrice and twice. In each DSC-BE block, the depth of BSC-BE was investigated to determine the number of times the DSC-BE convolution was carried out in order to achieve high performance. The first, second, and third MB-BE were recorded to determine the notation of the depth of DSC-BE. The higher the depth of DSC-BE, the greater the parameters utilized with deeper convolution. Furthermore, this study investigated the appropriate BM values by evaluating BM = 0.3, 0.5, 0.7, and 0.9. These two parameters in the ablation study were combined to obtain an adequate configuration for achieving optimal performance.

Table 1 shows the evaluated BM from 0.3 to 0.9 at the same depth of DSC-BE, i.e. (111). The results further show that the performance of training, validation, and testing were similar, where accuracy is low for small BM at 0.3, with accuracy, training, validation, and testing scores of 50.30%, 51.03%,

Table 1 – Ablation study of MB-BE and performance on test data.

Model	Parameters (million)	BM	Train data accuracy /%	Val. data accuracy /%	Test data			
					Accuracy /%	Precision /%	Recall /%	F1-score /%
MB-BE (111)	2.33	0.3	50.30	51.03	52.96	54.69	52.16	51.15
MB-BE (111)	3.16	0.5	56.75	63.21	60.02	58.41	58.06	57.66
MB-BE (111)	3.44	0.7	55.69	59.00	59.23	61.83	57.99	57.13
MB-BE (111)	4.00	0.9	55.73	59.00	59.00	60.58	58.02	57.69
MB-BE (251)	4.05	0.5	44.99	49.09	48.41	49.48	46.44	45.77
MB-BE (151)	4.50	0.5	44.46	47.95	49.43	49.95	48.42	46.95
MB-BE (122)	4.52	0.5	51.63	56.72	55.81	55.39	54.00	53.70

and 52.96%, respectively. For BM between 0.3 and 0.9, the performance peaked at 0.5, has accuracy, training, validation, and testing scores of 56.75%, 63.21%, and 60.12%, respectively. An increase in the BM significantly decreased the MB-BE performance contracts by approximately 55%. BM determines the bottleneck-level, and the number of parameters, therefore, the higher the bottleneck (small BM), the smaller parameters, and vice versa. The ablation study results showed that too large or small a bottleneck does not ensure better performance. The best bottleneck level was determined at BM = 0.5.

Furthermore, an ablation study was carried out for the same BM with various Depths used to determine the most optimal performance. Ablation also played attention to the number of parameters regarded by MobileNet, which is identical with small parameters. This means that the greater the depth, the higher the number of convolution conducted, which is in addition to the simplicity of the architecture used to properly model the system on mobile devices. The results of the ablation study at the Depth level of (251), (151), and (122) show that the accuracy of all Depth alternatives does not exceed Depth (111), where the highest accuracy is 56.72%. We found an increase in depth, leading to a rise in the number of parameters (more than 4 million parameters) with no effect on performance. Therefore, the results of the ablation study also indicated that the best depth is (111).

Performance comparisons were also carried out on test data using Precision, Recall, and F_{1-score}. Table 1, shows that the ablation study has a Depth (111) similar to the model, which achieved the best Precision at BM = 0.7 (61.83%). Recall shows different results where BM of 0.5 achieved the best precision results of 58.06%. This performance is slightly different from BM = 0.9, where recall of 58.02 still uses more parameters than 0.5. The F_{1-score} as a combined performance of Precision and Recall also shows that MB-BE with Depth (111) achieves the best performance, where using BM 0.9 and 0.5 achieves F_{1-scores} 57.69% and 57.66%, respectively. However, BM of 0.5 is superior to 0.9 because it uses a fewer parameter. The results of the comparison with other Depth differences, namely (251), (151), and (122), also show that the depth (111) outperforms other

The performance of MB-BE in classifying the freshness of fish eyes using the FFE dataset was compared in this research, as shown in Table 2. The ResNet50 reaches the best accuracy of training data, validation, and testing with an accuracy of 84.86%, 78.47%, and 78.82%, respectively. Meanwhile, the

MB-BE achieved an accuracy of 56.75%, 63.21%, and 60.02% for training, validation, and testing. The performance of ResNet50 is higher than MB-BE, however, the parameters utilized are approximately seven times higher, parameters of 23.59 million and 3.16 million.

MB-BE also outperforms MobileNet, Densenet, VGG16, and Nasnet Mobile, with approximately 34% to 59% accuracy. In terms of the number of parameters, MB-BE is more excellent except for MobileNet2. MB-BE uses 3.16 million parameters while the other CNN use more than 3.2 million, and MobileNet V2 uses 2.25 million; nevertheless, the accuracy reached by MB-BE is higher than MobileNetV2. This research was further experimented with by classifying datasets, namely Caltech-101 and Coil-100.

Table 2 showed that the best Precision, Recall, and F1-score were achieved by ResNet50, despite receiving more parameters than all others. MB-BE had Precision, Recall, and F1-scores of 58.41%, 58.06%, and 58.24%, equivalent to the original MobileNetV1 at 59.74%, 57.98%, and 58.85%. However, our proposal used fewer parameters of 3.16 and 3.22 compared to original MobileNetV1 and other models, and superior on all performance during the experiment, such as MobileNetV2, DenseNet 121, VGG16, and Nasnet Mobile.



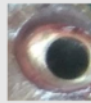

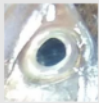
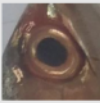
MobileNetV2 and MobileNetV1 have simplistic architecture, with 2.25 million and 3.22 million parameters. Nevertheless, MobileV1 was chosen as a baseline because it outperformed MobileNetV2 to propose a new architecture for fish freshness classification using fewer parameters. Table 2 shows that the proposed model outperforms the original MobileNetV1 and MobileNetV2 in terms of accuracy, while the Precision, Recall, and F1-score are equivalent to MobileNetV1 and superior MobileNetV2.

The samples results of fish eye freshness classification are shown in Table 3. The results in columns (a), (b), (d), (f) indicate that the proposed model was successfully predicted according to ground truth, as opposed to others. The original MobileNet V1 and ResNet50 also achieved four correct predictions and outperform others. The sample indicates that successful classification of the six image samples, wherein general success is still evaluated using the previously discussed metrics. However, there are images that failed to be classified by the proposed model, despite being successful by other models, such as column (e), which was successfully classified by MobileNet V1, ResNet50, and Nasnet Mobile. Some images also failed to be classified by all models, such as those in column (c).

Table 2 – Model performance using FFE dataset.

Model	Parameters (million)	Train data accuracy %	Val. data accuracy %	Test data			
				Accuracy %	Precision %	Recall %	F _{1-score} %
MobileNet V1	3.22	53.87	57.97	59.11	59.74	57.98	58.85
MobileNet V2	2.25	54.29	55.35	53.87	52.12	50.95	51.53
ResNet50	23.59	84.86	78.47	78.82	79.14	77.70	78.41
DenseNet 121	7.04	35.96	43.05	42.37	42.50	38.41	40.35
VGG16	14.71	34.26	41.00	43.85	45.81	41.38	43.48
Nasnet Mobile	4.27	35.77	40.89	37.24	33.66	30.61	33.37
MB-BE (111)	3.16	56.75	63.21	60.02	58.41	58.06	58.24

Table 3 – Sample detection results.

Model	Images, label of ground truth and predicted results					
						
	(a)	(b)	(c)	(d)	(e)	(f)
Ground truth	0	1	3	5	9	10
MobileNet V1	0	2	17	5	9	10
MobileNet V2	2	2	9	14	10	10
ResNet50	0	1	12	14	9	10
DenseNet 121	2	0	12	14	10	10
VGG16	2	2	9	12	10	11
Nasnet Mobile	2	9	9	14	9	9
MB-BE (111)	0	1	9	5	10	10

The value [0–23] is inner-product between {Chanos Chanos, Johnius Trachycephalus, Nibea Albiflora, Rastrelliger Faughni, Upeneus Moluccensis, *Eleutheronema Tetradactylum*, *Oreochromis Mossambicus*, and *Oreochromis Niloticus*} and {highly fresh, fresh, not fresh}, for example, 0, 1, 2 are highly fresh, fresh, and not fresh of Chanos Chanos, respectively.

In conclusion, the model performance in the testing session determined the system's robustness when implemented in real cases. This is because the higher the testing performance, the more robust the system completes the classification. The experimental results showed that the highest performance on the test data is ResNet50 and MB-BE because they both outperformed the state-of-the-art accuracy. Although MB-BE had less accuracy than ResNet50, it utilizes highly smaller parameters, which makes it a proper technique for classifying the freshness of fish eyes.

3.3. MB-BE For image classification

The Caltech-101 and Coil-100 datasets were used to determine the MB-BE's performance in solving other classification problems. Caltech-101 consists of 9 144 images, divided into 102 classes, which ranges from 31 to 800 pictures for each class. This means Caltech-101 is an imbalanced dataset, as shown in Table 4. The experimental findings on the Caltech-101 dataset revealed that MB-BE was unable to complete the unbalanced dataset, where the performance achieved in the training session was 78.56%. This performance was superior to Densenet, Nasnet Mobile, while MobileNet and ResNet50

outperformed VGG16 with an accuracy of 93.78%. The performance of MB-BE in validation and testing session was also similar with an accuracy of 70.11% and 68.56%, respectively. This result outperformed VGG16 as opposed to MobileNet, Densenet, Nasnet Mobile. The highest accuracy and validation scores of 92.73% and 92.67% were achieved by ResNet50.

The data obtained from the Columbia University Image Library (Coil-100) consists of 100 colored image objects arranged against a black backdrop on a motorized turntable. Regarding a fixed color camera, the turntable was rotated by 360 degrees to vary the object pose taken at 5-degree intervals, which correspond to 72 poses per object (7 200 images in total). The experimental results on the Coil-100 dataset shown in Table 5 indicate that all models achieve high accuracy. However, Densenet, Nasnet Mobile, and VGG16 accuracies were below 90% for training sessions, while MB-BE achieved the best performance at 99.36%. All models achieved accuracy above 96% in the validation and testing session, while ResNet and MB-BE were 100% and 99.93%. Although the MB-BE performance is not the best, its difference with ResNet is only 0.07 with an accuracy of 99.93%, outperforming other state-of-the-art, to achieve optimal performance. Other models such as MobileNet, Densenet, Nasnet Mobile, and

Table 4 – Performance classification using Caltech-101.

Architecture	Accuracy/%		
	Train	Val.	Test
MobileNet V1	80.51	89.40	89.67
MobileNet V2	81.99	88.63	89.50
Densenet 121	71.38	83.06	83.98
ResNet50	93.78	92.73	92.67
VGG16	42.71	59.67	51.94
Nasnet Mobile	70.34	81.58	82.89
MB-BE	78.56	70.11	68.56

Table 5 – Performance classification using Coil-100.

Architecture	Accuracy/%		
	Train	Val.	Test
MobileNetV1	95.09	99.71	99.57
MobileNetV2	96.59	99.86	99.79
Resnet50	98.25	100.0	100.0
Densenet121	88.39	97.86	97.43
VGG16	80.52	97.14	96.86
Nasnet Mobile	80.30	93.93	92.64
MB-BE	99.36	99.86	99.93

VGG16 have achieved optimal results as well, with all above 93%. The results also prove that MB-BE is appropriate for a dataset with many varieties such as scaling, rotation, and viewpoint but not suitable with color, lighting, shearing, and even the image's quality.

3.4. Analysis of MB-BE performance for image classification

There are inconsistencies associated with the process of classifying the freshness of fish eyes using fresh and not fresh fish due to non-sufficient adequate features that can be perceived. The images of fish eyes, both fresh and not fresh, are not diversified as in the general classification cases, so the proper CNN model for solving it may not be complicated as models in general, such as ResNet, Densenet, Nasnet, and VGG16. Furthermore, a lightweight model was implemented on mobile devices with MobileNet as a proper alternative because of its small parameter compared to other CNNs. However, MobileNet could not optimally classify the freshness of fish eyes because of the straightforward convolution flow architecture and not adequate features to distinguish fresh and not fresh fish eyes. As our proposed CNN architecture, MB-BE partially inherits the MobileNet V1 architecture combined with DSC-BE and RT to provide a precise representation of features in classifying the freshness of fish's eyes. MB-BE with Depth (111) has fewer parameters than MobileNet V1, but the accuracy reached the most optimal compared to other models except ResNet50, where the accuracy reached 63.21%. This performance is not high due to two constraints, a lightweight model requirement and the lack of distinguishing features of the freshness of fish eyes. They can be observed by naked eyes where it is hard to identify fresh and not fresh fish. Our proposed model attempts to get the right features and higher performance at a lower computational cost. Conversely, ResNet50 achieves better performance with a higher computational cost; meanwhile, although it is below 70%, MB-BE achieved the most optimal performance with a lighter architecture compared to the state-of-the-art.

4. Conclusions and future works

In conclusion, this study experimented the Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE) to improve feature quality using bottleneck convolution and generate more detailed features using expansion convolution. Feature maps transition among blocks with residual was also

addressed using Residual Transition (RT) instead of pooling and skip connection. The experimental result showed that MobileNetV1 with Bottleneck and Expansion (MB-BE) relatively classifies the freshness of the fish eyes with accuracy up to 63.21%, thereby outperforming other models such as MobileNetV1, MobileNetV2, VGG16, Nasnet Mobile, and Densenet. The proposed model is exceeded by Resnet50 with accuracy of 84.86%. However, the parameters utilized by ResNet50 are greater than MB-BE, by approximately seven times higher, where ResNet50 uses 23.59 million parameters, while MB-BE is only 3.16 million. Therefore, MB-BE is a new state-of-the-art in fish freshness classification based on eyes.

Data augmentation was also employed during the experiment to obtain more data variation though the performance of MB-BE still below 70%. Therefore, the performance variation needs to be improved using other data variations, such as Variational Auto Encoder. Furthermore, a one-model approach was used to predict eight fish species and 3 fish freshness, therefore, it is classified into 24 classes. Further research needs to be conducted by separating species and freshness to obtain the model's optimal performance.

9 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are grateful to the Deputy in Strengthening Research and Development, Ministry of Research and Technology/National Research and Innovation Agency, Indonesia, for supporting this research through the 2020 Basic Research theme at the Universitas Bhayangkara Surabaya under the contract number 008/SP2H/AMD/LT/MULTI/L7/2020 on 10th June 2020, and 048/VI/AMD/LPPM/2020/UBHARA on 11th June 2020.

REFERENCES

- [1] Wu L, Pu H, Sun DW. Novel techniques for evaluating freshness quality attributes of fish: A review of recent developments. Trends Food Sci Technol 2019;83:259–73. <https://doi.org/10.1016/j.tifs.2018.12.002>.

- [2] Mohammadi Lalabadi H, Sadeghi M, Mireei SA. Fish freshness categorization from eyes and gills color features using multi-class artificial neural network and support vector machines. *Aquacult Eng* 2020;90:102076. <https://doi.org/10.1016/j.aquaeng.2020.102076>.
- [3] Badan Pusat Statistik Indonesia. Ringkasan Eksekutif Pengeluaran dan Konsumsi Penduduk Indonesia. Badan Pusat Statistik; 2015.
- [4] Murakoshi T, Masuda T, Utsumi K, Tsubota K, Wada Y, Saygin AP. Glossiness and Perishable Food Quality: Visual Freshness Judgment of Fish Eyes Based on Luminance Distribution. *PLoS ONE* 2013;8(3):e58994. <https://doi.org/10.1371/journal.pone.0058994>.
- [5] Jalal A, Salman A, Mian A, Shortis M, Shafait F. Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecol Inf* 2020;57:101088. <https://doi.org/10.1016/j.ecoinf.2020.101088>.
- [6] Prasetyo E, Purbaningtyas R, Adityo R. Cosine K-Nearest Neighbor in Milkfish Eye Classification. *Int J Intell Eng Syst* 2020;13(3):11–25.
- [7] Prasetyo E, Adityo RD, Purbaningtyas R. Classification of segmented milkfish eyes using cosine K-nearest neighbor. In: Proceedings of ICAITI 2019–2nd International Conference on Applied Information Technology and Innovation: Exploring the Future Technology of Applied Information Technology and Innovation, Institute of Electrical and Electronics Engineers Inc. p. 93–8. <https://doi.org/10.1109/ICAITI48442.2019.8982124>.
- [8] Taheri-Garavand A, Nasiri A, Banan A, Zhang Y-D. Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish. *J Food Eng* 2020;278:109930. <https://doi.org/10.1016/j.jfoodeng.2020.109930>.
- [9] Abu Mallouh A, Qawaqneh Z, Barkana BD. Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images. *Image Vis Comput* 2019;88:41–51. <https://doi.org/10.1016/j.imavis.2019.05.001>.
- [10] Kasinathan T, Singaraju D, Uyyala SR. Insect classification and detection in field crops using modern machine learning techniques. *Information Process Agric* 2021;8(3):446–57. <https://doi.org/10.1016/j.inpa.2020.09.006>.
- [11] Ismail N, Malik OA. Real-time Visual Inspection System for Grading Fruits using Computer Vision and Deep Learning Techniques. *Information Processing Agric* 2021. <https://doi.org/10.1016/j.inpa.2021.01.005>.
- [12] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2016.90>.
- [13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.
- [14] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2017.243>.
- [15] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications 2017.
- [16] Zoph B, Vasudevan V, Shlens J, Le QV. Learning Transferable Architectures for Scalable Image Recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2018.00907>.
- [17] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation Mark. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2018.00474>.
- [18] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2016.308>.
- [19] Wang CY, Mark Liao HY, Wu YH, Chen PY, Hsieh JW, Yeh IH. CSPNet: A new backbone that can enhance learning capability of CNN. *IEEE Comput Soc Conf Computer Vis Pattern Recogn Workshops* 2020. <https://doi.org/10.1109/CVPRW50498.2020.00203>.
- [20] Razzaghi P, Abbasi K, Bayat P. Learning spatial hierarchies of high-level features in deep neural network. *J Vis Commun Image Represent* 2020;70:102817. <https://doi.org/10.1016/j.jvcir.2020.102817>.
- [21] Huang J-C, Huang H-C, Liu H-H. Research on the parallelization of image quality analysis algorithm based on deep learning. *J Vis Commun Image Represent* 2020;71:102709. <https://doi.org/10.1016/j.jvcir.2019.102709>.
- [22] Fei-Fei L, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *IEEE Comput Soc Conf Comput Vis Pattern Recogn Workshops* 2004. <https://doi.org/10.1109/CVPR.2004.383>.
- [23] Nene S, Nayar S, Murase H. Columbia Object Image Library (COIL-20). *Techn Rep* 1996.
- [24] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2017.634>.
- [25] Redmon J, Farhadi A. Yolov3. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recogn* 2017. <https://doi.org/10.1109/CVPR.2017.690>.
- [26] Prasetyo E, Adityo RD, Suciati N, Fatichah C. The freshness of fish eyes dataset 2020. 10.17632/xzyx7pbr3w.1.

Paper Jurnal/Prosiding

ORIGINALITY REPORT

11%

SIMILARITY INDEX

9%

INTERNET SOURCES

5%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

doaj.org

Internet Source

3%

2

researchrepository.murdoch.edu.au

Internet Source

1%

3

data.mendeley.com

Internet Source

1%

4

www.alice.cnptia.embrapa.br

Internet Source

1%

5

www.researchgate.net

Internet Source

1%

6

Eko Prasetyo, Rani Purbaningtyas, Raden Dimas Adityo. "Performance Evaluation of Pre-trained Convolutional Neural Network for Milkfish Freshness Classification", 2020 6th Information Technology International Seminar (ITIS), 2020

Publication

1%

7

Mirye Kim, Yongjang Kwon, Joouk Kim, Youngmin Kim. "Image Classification of Parcel Boxes under the Underground Logistics

1%

System Using CNN MobileNet", Applied Sciences, 2022

Publication

8

Eko Prasetyo, Nanik Suciati, Chastine Fatichah. "Multi-level residual network VGGNet for fish species classification", Journal of King Saud University - Computer and Information Sciences, 2021

Publication

1 %

9

repositorio.bc.ufg.br

Internet Source

1 %

10

link.springer.com

Internet Source

1 %

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On