Ubhara Surabaya

Eko Prasetyo <eko@ubhara.ac.id>

Confirming submission to Information Processing in Agriculture

1 message

Information Processing in Agriculture <em@editorialmanager.com> 2 March 2021 at 13:32
Reply-To: Information Processing in Agriculture <ipa_cau@163.com>
To: Eko Prasetyo <eko@ubhara.ac.id>

This is an automated message.

Combining MobileNetV1 and Depthwise Separable Convolution Bottleneck with Expansion for Classifying the Freshness of Fish Eyes

Dear Mr. Prasetyo,

We have received the above referenced manuscript you submitted to Information Processing in Agriculture.

To track the status of your manuscript, please log in as an author at https://www.editorialmanager.com/ipa/, and navigate to the "Submissions Being Processed" folder.

Thank you for submitting your work to this journal.

Kind regards,
Information Processing in Agriculture

More information and support

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

---

Ubhara Surabaya

Eko Prasetyo <eko@ubhara.ac.id>

Confirming handling editor for submission to Information Processing in Agriculture

1 message

Information Processing in Agriculture <em@editorialmanager.com> 3 March 2021 at 09:05
Reply-To: Information Processing in Agriculture <ipa_cau@163.com>
To: Eko Prasetyo <eko@ubhara.ac.id>

This is an automated message.

Manuscript Number: IPA-D-21-00070

Combining MobileNetV1 and Depthwise Separable Convolution Bottleneck with Expansion for Classifying the Freshness of Fish Eyes

Dear Mr. Prasetyo,

The above referenced manuscript will be handled by Editorial Office Dr Xiangyun Guo .

To track the status of your manuscript, please log into Editorial Manager at https://www.editorialmanager.com/ipa/.

Thank you for submitting your work to this journal.

Kind regards,

Information Processing in Agriculture

More information and support

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

---

Decision on submission to Information Processing in Agriculture

1 message

Information Processing in Agriculture <em@editorialmanager.com> 1 June 2021 at 08:58
Reply-To: Information Processing in Agriculture <ipa_cau@163.com>
To: Eko Prasetyo <eko@ubhara.ac.id>

Manuscript Number: IPA-D-21-00070

Combining MobileNetV1 and Depthwise Separable Convolution Bottleneck with Expansion for Classifying the Freshness of Fish Eyes

Dear Mr. Prasetyo,

Thank you for submitting your manuscript to Information Processing in Agriculture.

I have completed my evaluation of your manuscript. The reviewers recommend reconsideration of your manuscript following minor revision and modification. I invite you to resubmit your manuscript after addressing the comments below. Please resubmit your revised manuscript by Jun 30, 2021.

When revising your manuscript, please consider all issues mentioned in the reviewers' comments carefully: please outline every change made in response to their comments and provide suitable rebuttals for any comments not addressed. Please note that your revised submission may need to be re-reviewed.

To submit your revised manuscript, please log in as an author at https://www.editorialmanager.com/ipa/, and navigate to the "Submissions Needing Revision" folder under the Author Main Menu.

Information Processing in Agriculture values your contribution and I look forward to receiving your revised manuscript.

Kind regards,

Xiangyun Guo

Editorial Office

Information Processing in Agriculture

Editor and Reviewer comments:

Editor:
1. Please revise it based on the attached template.
2. Please read it THREE times after revision to make sure its correctness.

Reviewer #1: In general, the submitted manuscript is scientifically sound, interesting, and reflect some ingenuity and originality in its own field. In other words, it contains substantially new and interesting information that is of sufficient importance to justify publication. The only critical flaw is very very poor writing of the script. In needs very strong editing regarding grammar, punctuation, sentences and so on. I have specified only a few of them. It must be certainly revised by a native English language editor. The sentences must be rephrased. The Abstract and all parts of the paper have been written badly. The Figures and Tables have not been provided in a manner that deserves to be brought in a scientific Journal.

Reviewer #2: In this paper, deep learning is applied to identify the freshness of fish eyes. The mb-be model is compared with other models such as mobilenet, vgg16 and densenet, the results show that the mobile net V1 (mb-be) model based on bottleneck and extension has the highest classification accuracy.

In general, the research topic of this paper is novel and interesting. When revising this paper, the following advices should be considered:

1. In the validation of the model, only the accuracy is used to evaluate the multi classification problem, which is not convincing enough.

2. The author should further discuss the motivation of the proposed algorithm. For example, why not simply use deep learning technologies such as mobile netv2 and mobile netv3?

3. The super parameters in the model are not enough in detail.

4. The recognition accuracy is just 63.21%, it is not high.

5. There are some typos and grammar mistakes. As such, the English requires further improvement to further improve the quality of this paper.

Reviewer #3: Although, like I said, the paper is well structured, there are some problems in the article that should be addressed before acceptance and publication.
1) Correct some grammar and language issues, which make some expressions hard to understand at first sight. Maybe, even the help of a professional proficient in English would ease this process.

2) The selection of the baseline methods and the experimental results can benefit from revision. The models compared with this paper should be as new as possible. In this way, the proposed method can be the current state-of-the-art for freshness of fish eye.

3)I would like to ask the authors, to make explicit their contribution for the reader to understand it more easily.

4)Several papers have been surveyed by the authors which are base papers of the proposed work. Authors should include a conclusive summary about their own understanding that motivate them to propose this work should be added.

5) Conclusion is weak. Must be strengthened.

6) Some of the future aspects must be added.

IPA-D-21-00070_reviewer (1).pdf
1202K

(IPA-D-21-00070) Response Letter:

Title: Combining MobileNetV1 and Depthwise Separable Convolution Bottleneck with Expansion for Classifying the Freshness of Fish Eyes.

Editor comment to the authors:

[1] Please revise it based on the attached template.

Answer:
Thank you for your valuable comment.

1. Based on the attached template, we have revised as follow:
    a. non-destructive is replaced by non-destructiveness ………………………..(line 39)
    b. managed is replaced by arranged ………………………………………..(line 67)
    c. performance is replaced by the performance ………………………………(line 79)
    d. number channel is replaced by channel number …………………………...(line 104)
    e. compared is replaced by are comparing the performance between ………..(line 147)
    f. determine how MB-BE's performance is replaced by evaluate the performance of MB-BE …………………………………………………………………………(line 150)

Editor comment to the authors:

[2] Please read it THREE times after revision to make sure its correctness.

Answer:
Thank you for your valuable comment.

1. We have read this manuscript three times to ensure that the revision has been completed, grammar, and all language errors have been revised.

Reviewer #1 comment to the authors:

[1] In general, the submitted manuscript is scientifically sound, interesting, and reflect some ingenuity and originality in its own field. In other words, it contains substantially new and interesting information that is of sufficient importance to justify publication. The only critical flaw is very very poor writing of the script. In needs very strong editing regarding grammar, punctuation, sentences and so on. I have specified only a few of them. It must be certainly revised by a native English language editor. The sentences must be rephrased. The Abstract and all parts of the paper have been written badly. The Figures and Tables have not been provided in a manner that deserves to be brought in a scientific Journal.

Answer:
Thank you for your valuable comment.

1. We have made revisions to improve grammar, punctuation, sentences, typos. And also re-proofreading by a native English language editor.
2. Some sentences have been rephrased, for example:
    a. The survey results also indicate the high consumption of fish … was replaced with The survey results indicate the high consumption of side dishes …………(red font, line 33)
    b. Currently, apart from fresh fish, some sellers sell not fresh fish (ice-storage durations) … was replaced with Currently, some sellers sell both fresh fish and not fresh fish with ice-storage durations …………………………………………..........(red font, (line 35-36)
    c. the fish's freshness … was replaced with the freshness of fish ………(red font, line 38)
    d. no image quality analysis additional … was replaced with no additional image quality analysis …………………………………………………………(red font, line 100)
    e. CNN advises optimal classification performance, but CNN requires huge numbers of image data and a wide variety of images … was replaced with CNN promises optimal classification performance; on the other hand, CNN requires a vast number of image data and a wide variety of images ……………………………………..(red font, line 113-114)
    f. together with a new CNN architectural proposal, the contributions we made are as follows … was replaced with we propose a new CNN architectural proposal with contributions as follows ……………………………………………(red font, line 120)
    g. determine how … was replaced with evaluate the performance …...(red font, line 150)
    h. exciting because we can recognize … was replaced with impressive since it can recognize ……………………………………………………………(red font, line 155)
3. The abstract have been re-written ……………….. as follows:
    a. Second, third and fouth sentences
    MobileNet delivers a vast decreasing number of parameters for learning features by moving from the standard convolution paradigm to the depthwise separable convolution (DSC). However, classifying the freshness of fish eyes does not have adequate features to be learned; furthermore, slight differences in features do not require complicated CNN architecture.
    Was replaced with:
    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.
    b. The fifth sentences
    our contributions are: First, proposed DSC Bottleneck with Expansion for learning the differentiating fresh and not fresh fish eyes with a Bottleneck Multiplier
    was replaced with:
    our first contribution proposed DSC Bottleneck with Expansion for learning features of the freshness of fish eyes with a Bottleneck Multiplier.
    c. The eighth sentences

Our experimental results using the Freshness of Fish Eyes dataset show that MB-BE achieved 63.21% accuracy and outperformed other models such as original MobileNet, VGG16, and Densenet.
Was replaced with
Our experimental results using the Freshness of Fish Eyes dataset show that MB-BE outperformed other models such as original MobileNet, VGG16, Densenet, Nasnet Mobile with 63.21% accuracy

4. The Figures and Tables have been presented using the writing and connected link as in a scientific Journal. For example:
   a. Figure 1 and the citation has connected .... (red font, line 189)
   b. Figure 2 and the citation has connected .... (red font, line 223)
   c. Figure 3 and the citation has connected .... (red font, line 232)
   d. Figure 4 and the citation has connected .... (red font, line 254)
   e. Table 1 and the citation has connected .... (red font, line 370)
   f. Table 2 and the citation has connected .... (red font, line 418)
   g. Table 3 and the citation has connected .... (red font, line 455)

Reviewer #2 comment to the authors:

[1] In the validation of the model, only the accuracy is used to evaluate the multi classification problem, which is not convincing enough.

Answer:
Thank you for your valuable comment.

1. We classify multi-class (24 classes), so we have added other validation metrics such as Precision, Recall, and F1-score, aside from Accuracy. Performance comparisons using Precision, Recall, and F1-scores also show that our proposal outperforms other models. We also added a theoretical explanation of performance metrics in section 3.7.
   (For this answer, I give explaining in section 3.7 line 314-333, Table 1, section 4.3 line 392-401, and Table 2, line 418-431 …….. red font)

Reviewer #2 comment to the authors:

[2] The author should further discuss the motivation of the proposed algorithm. For example, why not simply use deep learning technologies such as mobile netv2 and mobile netv3?.

Answer:
Thank you for your valuable comment.

1. MobileNetV2 has a more simplistic architecture, it can be observed that it utilizes 2.25 million parameters, while MobileNetV1 utilizes 3.22 million parameters; nevertheless, based on our experimental results in fish freshness classification using eyes, the performance of MobileV1 outperformed MobileNetV2, so we chose MobileNetV1 as a baseline to propose a new architecture for fish freshness classification using fewer parameters.
   (We explain our reasons in section 1 line 79-82 and 104-111 …………………………. red font)

2. We also do not compare it with MobileNetV3, because all the models we compare in this research use the Pre-trained version, which can be downloaded from the hard application (https://keras.io/api/applications/). Unfortunately, at the time of our experiment, Pre-trained MobileNetV3 was not yet available. We apologize for this.

Reviewer #2 comment to the authors:

[3] The super parameters in the model are not enough in detail.

Answer:
Thank you for your valuable comment.

1. We think the super parameter of the model is Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE) and Residual Transition (RT). DSC-BE is used to improve feature quality and generate more detailed features with low computational costs, while RT is for bridging the skip connection. in the transition block using 3×3 depthwise convolutions (DC) on the two feature maps.
2. DSC-BM convolutes the feature map into a BM ratio measure because BM is a non-negative number below one; the number of feature maps decreases according to BM. This is effective for scraping-out unimportant features and leaving quality features. Next, the DSC convolutes again with the same feature map size as before. The expansion of the number of feature maps is the generation of more detailed features through a non-linear transformation.
3. In RT, the first feature map is the current convolution layer, and the second one is the skip connection from the previous block. Then, we combine the results using concatenation and convolute them using 1×1 convolution followed by batch normalization, relu activation, and max-pooling. The result is used by the next convolution block or passed to the fully connected layer.
   We have added some detail explanation of both in section 3.2 line 209-218, and section 3.4 line 246-254 …….. red font.

Reviewer #2 comment to the authors:

[4] The recognition accuracy is just 63.21%, it is not high.

Answer:
Thank you for your valuable comment.

1. This performance is not high due to two constraints, a lightweight model requirement and the lack of distinguishing features of fish eye freshness. The features of fish's freshness 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. On the other hand, Resnet50 achieves better performance with a higher computational cost. Therefore, our proposal is more appropriate to use for recognizing fish's freshness.
(For this answer, I give explaining in section 4.5 line 495-500 …….. red font).

Reviewer #2 comment to the authors:

[5] There are some typos and grammar mistakes. As such, the English requires further improvement to further improve the quality of this paper.

Answer:
Thank you for your valuable comment.

1. We've made a revision for some typos and grammar mistakes, and also proofreading from native reader to improve the grammatical of our manuscript.
For example:
    a. non-destructive is replaced by non-destructiveness ………………………..(line 39)
    b. managed is replaced by arranged …………………………………………..(line 67)
    c. performance is replaced by the performance ………………………………(line 79)
    d. number channel is replaced by channel number …………………………...(line 104)
    e. compared is replaced by are comparing the performance between ……….(line 147)

Reviewer #3 comment to the authors:

[1] Correct some grammar and language issues, which make some expressions hard to understand at first sight. Maybe, even the help of a professional proficient in English would ease this process.

Answer:
Thank you for your valuable comment.

1. We have corrected some grammar and language issues and fixed them including the following:
    a. non-destructive is replaced by non-destructiveness (line 39)
    b. managed is replaced by arranged (line 67)
    c. performance is replaced by the performance (line 79)
    d. number channel is replaced by channel number (line 104)

e. compared is replaced by are comparing the performance between (line 147)
f. determine how MB-BE's performance is replaced by evaluate the performance of MB-BE (line 150)
g. right is replaced by good (line 31)
h. fish is replaced by side dish (line 34)
i. CNN advises optimal classification performance, but CNN requires huge number of image data and a wide variety of images is replaced by CNN promises optimal classification performance; on the other hand, CNN requires a vast number of image data and a wide variety of images (line 113-114)
j. To address this problem, together with a new CNN architectural proposal, the contributions we made are as follows is replaced by To address this problem, we propose a new CNN architectural proposal with contributions as follows (line 119-120)
k. Non-destructive fish freshness detection is exciting because we can recognize the freshness of fish without touching it is replaced by Non-destructive fish freshness detection is impressive since it can recognize the freshness of fish without touching it. (line 155-156)
l. cannot vary is replaced by are not diversified (line 486)
m. The experimental results on the Caltech-101 dataset presented in Table 3 show that MB-BE was not robust in completing the imbalanced dataset is replaced by Table 3 shows the experimental findings on the Caltech-101 dataset, which reveal that MB-BE could not complete the unbalanced dataset, (line 455-456)
n. It stands to reason that ResNet also uses a vast number of parameters is replaced by ResNet, needing to be noted, employs a vast number of parameters. (line 459-460)

2. We have also proofread to a professional proficient in English.

<u>Reviewer #3 comment to the authors:</u>

[2] The selection of the baseline methods and the experimental results can benefit from revision. The models compared with this paper should be as new as possible. In this way, the proposed method can be the current state-of-the-art for freshness of fish eye.

<u>Answer:</u>
Thank you for your valuable comment.

1. During the study, we conducted a pre-experiment by comparing several CNNs such as MobileNet V1, MobileNet V2, ResNet50, DenseNet 121, VGG16. The results show that ResNet50 achieves <u>the highest performance but uses many parameters</u>, while MobileNetV1 is the <u>next highest with fewer parameters</u>. Therefore, we used MobileNetV1 as a baseline to propose a new state-of-the-art named MB-BE for fish freshness classification based on eyes. We also present the performance of all these models in Table 2.
2. For comparison MB-BE with newest model, we have add comparison with Nasnet Mobile (proposed in 2018), and add it in our revision version

   a. Mention the concept of Nasnet (section 1 line 73-74)
   b. Comparison model (section 4.3 Table 2 line 418-424)
3. The recent model in our submission version is MobileNetV2 (proposed in 2018) that the pretrained version is available on keras application. (For this answer, there is an explanation in the section 1, line 78-82 …….. red font)

Reviewer #3 comment to the authors:

[3] would like to ask the authors, to make explicit their contribution for the reader to understand it more easily.

Answer:

Thank you for your valuable comment.

1. We have revised the explanation of contribution in the background as follows:
   a. We proposed a bottleneck with 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; then convoluting again to expand the feature maps as the original size. We also introduced BM as a constant to determine the bottleneck level with the performance trade-off and model size. (section 1 line 122-127).

   b. CNN architecture generally uses pooling only for bridging the feature maps from one layer to the next layer with different size; nevertheless, it can be used only if one feature map is input. In other hand, a transition block that also involves skip-connection from the previous layer (for maintaining low-level features) is not enough to utilize pooling. In this paper, we propose Residual Transitions (RT) for bridging some feature maps from convolutional block to the next convolutional block with different size. Feature maps from the previous block are combined with the current block using depthwise convolution and concatenation, then change the size of the feature map from current block size to the next block size using pointwise convolution. (section 1 line 130-137)

   c. We proposed MobileNetV1 with Bottleneck and Expansion (MB-BE) using some parts of the MobileNetV1 as the backbone and DSC-BE as an additional layer for recognizing the features of the fish's eye and reducing the number of parameters. As previously mentioned, it is hard to distinguish the freshness of a fish's eye since there are insufficient visual features, both internal within classes and externally between classes. DSC-BE is utilized to generate proper features for classifying the freshness of fish eyes in the MB-BE architecture. (section 1 line 139-144)
2. We think using these sentences, our contribution can be read easily.

Reviewer #3 comment to the authors:

[4] Several papers have been surveyed by the authors which are base papers of the proposed work. Authors should include a conclusive summary about their own understanding that motivate them to propose this work should be added.

Answer:
Thank you for your valuable comment.

1. We extend a conclusive summary about our understanding that motivate to propose this work as follows:

   We know that the channel number of MobileNet from initial to final layer is a binary-fold incremental parameter, i.e. 32-64-128-256-512-1024 parameters, and DSC managed to reduce the size model with performance similar to standard convolution, for example, compared to ResNet and VGGNet. In this case, MobileNetV1 uses fewer parameters to solve general classification problems, such as Imagenet. However, MobileNetV1 becomes less than optimal when classifying freshness of fish eyes. Therefore, we proposed a CNN architecture that inherits the smart DSC idea from MobileNetV1 nevertheless is more effective in learning the fish eye features for freshness classification.

   (For this answer, I give explaining in section 1, line 106-111 …….. red font).

Reviewer #3 comment to the authors:

[5] Conclusion is weak. Must be strengthened.

Answer:
Thank you for your valuable comment.

1. We strengthen the conclusion by conveying the concept and the results achieved by MB-BE, DSC-BE, and RT. We have also added performance comparisons to other models, as follows:

   This study experimented with 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 also addressed using Residual Transition (RT) instead of pooling and skip connection. Our experimental results show that MobileNetV1 with Bottleneck and Expansion (MB-BE) relatively classifies the freshness of the fish's eye with accuracy up to 63.21%, outperforming other models such as MobileNetV1 original, MobileNetV2, VGG16, Nasnet Mobile, and Densenet. Comparing with Resnet50, our proposal is outperformed by ResNet50 with accuracy up to 84.86%; however, the parameters utilized by ResNet50 are higher than MB-BE, approximately seven times higher, where ResNet50 uses 23.59 million parameters while MB-BE is only 3.16 million parameters. Therefore, MB-BE is a new state-of-the-art in fish's freshness classification based on eyes.

   (For this answer, I give explanation in section conclusion, line 502-512 …….. red font).

Reviewer #3 comment to the authors:

[6] Some of the future aspects must be added.

Answer:
Thank you for your valuable comment.

1. We have added future aspects that should be researched, as follows:

    During experiments, data augmentation is also employed to obtain more data variation though the performance of MB-BE still below 70%. We need to improve performance using other data variations, for example, Variational Auto Encoder or others. We used a one-model approach to predict eight fish species and three fish freshness, so our model classified 24 classes. This approach caused the trained model to classify many classes with few visual features. For future research, it is recommended to use a different approach where type and freshness should be separated so that the model's performance is more optimal.

    (For this answer, I give explanation in section conclusion, line 513-518 …….. red font).

Information Processing in Agriculture

Combining MobileNetV1 and Depthwise Separable Convolution Bottleneck with Expansion for Classifying the Freshness of Fish Eyes

--Manuscript Draft--

Manuscript Number:	
Article Type:	Full length article
Keywords:	depthwise separable convolution; bottleneck; classification; freshness; fish eye; residual transition
Manuscript Region of Origin:	Asia Pacific

Abstract:	Image classification using Convolutional Neural Network (CNN) achieve optimal performance with a particular strategy; MobileNet delivers a vast decreasing number of parameters for learning features by moving from the standard convolution paradigm to the depthwise separable convolution (DSC). However, classifying the freshness of fish eyes does not have adequate features to be learned; furthermore, slight differences in features do not require complicated CNN architecture. In this paper, our contributions are: First, proposed DSC Bottleneck with Expansion for learning the differentiating fresh and not fresh fish eyes with a Bottleneck Multiplier; Second, proposed Residual Transition to bridge current feature maps and skip connection feature maps to the next convolution block; Third, proposed MobileNetV1 Bottleneck with Expansion (MB-BE) for classifying the freshness of fish eyes. Our experimental results using the Freshness of Fish Eyes dataset show that MB-BE achieved 63.21% accuracy and outperformed other models such as original MobileNet, VGG16, and Densenet.

Highlights

- The problem in classifying the freshness of fish eyes is not enough appearance features.
- A slight differences of images do not require a complicated CNN architecture
- We proposed Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE) to learn the differentiating fresh and not fresh fish eyes
- We proposed Residual Transition (RT) to bridge current feature maps and skip connection features
- We proposed MobileNetV1 with Bottleneck and Expansion (MB-BE) inherited from MobileNetV1, and combined with DSC-BE, and RT

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

*Corresponding authors: nanik@if.its.ac.id, eko@ubhara.ac.id

Abstract

Image classification using Convolutional Neural Network (CNN) achieve optimal performance with a particular strategy; MobileNet delivers a vast decreasing number of parameters for learning features by moving from the standard convolution paradigm to the depthwise separable convolution (DSC). However, classifying the freshness of fish eyes does not have adequate features to be learned; furthermore, slight differences in features do not require complicated CNN architecture. In this paper, our contributions are: First, proposed DSC Bottleneck with Expansion for learning the differentiating fresh and not fresh fish eyes with a Bottleneck Multiplier; Second, proposed Residual Transition to bridge current feature maps and skip connection feature maps to the next convolution block; Third, proposed MobileNetV1 Bottleneck with Expansion (MB-BE) for classifying the freshness of fish eyes. Our experimental results using the Freshness of Fish Eyes dataset show that MB-BE achieved 63.21% accuracy and outperformed other models such as original MobileNet, VGG16, and Densenet.

Keywords: depthwise separable convolution, bottleneck, classification, freshness, fish eye, residual transition

1. Introduction

Fish are a favorite food product for all ages; besides delicious taste, fish also contain right human health nutrients (Mohammadi Lalabadi, Sadeghi, & Mireei, 2020). For Indonesian people, fish are also a part of their daily staple food with rice and vegetables. The survey results also indicate the high consumption of fish, in which fish are in the fourth rank after processed food and beverages, cigarettes, and grains (Badan Pusat Statistik Indonesia, 2015). Currently, apart from fresh fish, some sellers sell not fresh fish (ice-storage durations). The freshness of fish can be inspected using sensory cues such as visual appearance, texture, sound, taste, and smell (Murakoshi, Masuda, Utsumi, Tsubota, & Wada, 2013). Ordinary people cannot bring a detector only for recognizing the fish's freshness; therefore, people need a system that automatically recognizes the freshness of fish quickly, easily, and non-destructive. Non-destructive freshness classification can only be conducted using an image-based automatic system by processing the visual features in the image. Due to the limitation of visual appearance, fish freshness classification can be carried out on the appearance of the eyes, skin, or tail. In this study, we focused on the fish freshness classification based on eyes appearance.

In their essential study of visual eye differences for fish freshness, (Murakoshi et al., 2013) showed that fish eyes' visual appearance, both fresh and not fresh, is different. The differences between individual fish in the perceived freshness scores (PFS) were higher for the shorter degradation time and lower for the longer degradation time. This information indicates that it is possible to develop a classification of fish's freshness. There have been some studies in fish freshness classification, such as (Mohammadi Lalabadi et al., 2020) developing a system for classifying fish's freshness based on eyes and gills during a 10-day ice-storage cycle, (Jalal, Salman, Mian, Shortis, & Shafait, 2020) developing a location detector and a count of fish numbers in the sea with extreme background variation. In addition, (Kunjulakshmi et al., 2020) developed portable and non-destructive systems as freshness sensors for Rastrelliger kanagurta stored under ice, and (Taheri-Garavand, Nasiri, Banan, & Zhang, 2020) developed a system for classifying common carp fish using deep convolutional neural network (CNN). CNN is also used by many other researchers, for example, (Zheng, Lei, & Zhang, 2020) combined Fully Convolutional Networks (FCN), Regionswith CNN feature (R-CNN), and Richer Fully Convolutional Networks (RFC) to amplify and extract the features of the data and previous studies in building surface data. (Abu Mallouh, Qawaqneh, & Barkana, 2019) used CNN for age range classification from unconstrained face images, and (Xiong, Zhang, & Zhang, 2020) used CNN as feature extraction for classifying water leakage diseases of shield tunnels. CNN was also used for insect classification (Kasinathan, Singaraju, & Uyyala, 2020) and fruit grading (Ismail & Malik, 2021). The system for freshness classification using CNN provides the advantage of integrating feature extraction and classification; consequently, system training is also carried out simultaneously between the two sections. Simultaneous training on CNN results in forming all parts of the system model to be interdependent and supportive. For example, changing the lower layer's parameter value also results in changing the parameter value of the next layer to the classifier. This mechanism distinguishes the CNN-based approach from the hand-crafted features-based approach. In the hand-crafted features-based approach, we managed the generation of features and classifiers independently or separately so that the train system is only on the classifier. Several CNNs were previously developed, for example, (He, Zhang, Ren, & Sun, 2016) developed a residual network (ResNet) to maintain lower-level features using a skip connection, and (Simonyan & Zisserman, 2015) developed VGGNet by reducing the convolutional kernel size. Furthermore, (G. Huang, Liu, Van Der Maaten, & Weinberger, 2017) developed Densenet as an alternative skip connection with bottleneck connections, (Howard et al., 2017) developed MobileNet as a small classifier suitable for implementation on mobile devices, (Mahmood, Bennamoun, An, Sohel, & Boussaid, 2020) collected features from multiple convolution layers to get more powerful features. (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018) further developed MobileNetV1 using an inverted residual and linear bottleneck called MobileNetV2 to simplify architecture further and improve performance. All these architectures attempt to achieve the CNN model for solving classifications with optimal performance. We reviewed MobileNet and proposed a new architecture to resolve fish's freshness classification based on eyes images. In addition, although PFS confirms that fresh and not fresh fish eyes can be distinguished, but the freshness classification based on eyes images has a different notion, visual appearance is more dominant to be analyzed; since it is difficult for the naked eye for distinguishing fresh and not fresh fish, the automatic system also does not have proper features to distinguish them.

As discussed earlier, MobileNet delivers a vast decreasing number of parameters by moving from the standard convolution paradigm to the depthwise separable convolution (DSC). In standard convolution, convolution involves kernel size and input-output channels. DSC breaks it down into depthwise convolution (single kernel for every channel) and pointwise convolution (standard convolution with 1×1 kernel); both of them are used sequentially. From the empirical studies results, DSC achieves performance slightly below the standard convolution, but the number of parameters is relatively lighter. Thus, this smart concept presents a trade-off where the number of parameters is massively reduced with a slight drop in performance (Sandler et al., 2018). In our opinion, 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 width multiplier = 1.0; its architectural paradigm is also classic, which uses plain convolution flow (PCF). PCF is a convolutional flow from the initial layer to the final layer, stratified and straight, no skip connection (He et al., 2016), no bottleneck (Sandler et al., 2018), no inception (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016), no cross-stage partial (Wang et al., 2020), no spatial hierarchical features (Razzaghi, Abbasi, & Bayat, 2020) and no image quality analysis additional (J. C. Huang, Huang, & Liu, 2020); all the PCF stages are straight to the end. PCF presents architectural simplicity, but the model is ineffective for learning dataset features; for example, accuracy increases during training but then gets saturated and drops fast, (He et al., 2016), increasing in the number of channels with binary-fold incremental inflict in higher memory traffic (Wang et al., 2020). We know that the number channel of MobileNet from initial to final layer is a binary-fold incremental parameter, i.e. 32-64-128-256-512-1024 parameters. Therefore, we proposed a CNN architecture that inherits the smart DSC idea from MobileNetV1 nevertheless is more effective in learning the fish eye features for freshness classification.

As explained earlier, the non-destructive fish freshness examination can only be conducted using an image-based automatic system. CNN advises optimal classification performance, but CNN requires huge numbers of image data and a wide variety of images. Several datasets, such as Caltech-101 (Fei-Fei, Fergus, & Perona, 2004) or Coil-100 (Nene, Nayar, & Murase, 1996), have many variations in scaling, rotation, shearing, color, lighting, viewpoint, and even the quality of the image itself. However, the fish eyes, both fresh and not fresh, do not have adequate variety of images (tiny differences), so there are not sufficient visual features, both internal within classes and external between classes. On the other hand, slight differences in features do not require complicated CNN architecture. To address this problem, together with a new CNN architectural proposal, the contributions we made are as follows:

1. Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE)
   We proposed a bottleneck with expansion convolution where the feature map was convoluted using the Depthwise Separable Convolution (DSC) into fewer feature maps according to Bottleneck Multiplier (BM) × feature map. This bottleneck concept effectively improves features quality for both classification (Xie, Girshick, Dollár, Tu, & He, 2017) and object detection (Redmon & Farhadi, 2017), while expansion is to generate more detailed features with non-linear functions. We also introduced BM as a constant to determine the bottleneck level with the performance trade-off and model size.
2. Residual Transition
   Transitions are used for bridging a particular convolutional block to the next convolutional block. The size of the two blocks' feature maps is not the same, so a transition block is utilized to bridge the transfer of features from a particular block to the next block. CNN architecture generally uses pooling only to bridging; for feature maps from one layer to the next layer, pooling can solve it. Nevertheless, a transition block that also involves skip-connection from the previous layer (for maintaining low-level features) is not enough to utilize pooling. We proposed Residual Rransition (RT) to bridge the skip connection in the transition block with 3×3 depthwise convolutions on two

feature maps, then combine them using concatenation followed by 1×1 convolution, batch normalization, relu activation, and max-pooling.

3. MobileNetV1 with Bottleneck and Expansion (MB-BE)
   We proposed MobileNetV1 with Bottleneck and Expansion (MB-BE) using some parts of the MobileNetV1 as the backbone and DSC-BE as an additional layer for recognizing the features of the fish's eye and reducing the number of parameters. As explained before, the fish's eye freshness is difficult to be distinguished because there are not adequate visual features, both internal within classes and external between classes. We proposed DSC-BE in the MB-BE architecture to create correct features for classifying the freshness of fish eyes. We also 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.

The experiments in this study compared MB-BE's performance with several CNNs, such as the original MobileNet V1, MobileNet V2, ResNet, Densenet, and VGG16. The experiments were carried out on the Freshness of Fish Eyes (FFE) dataset (4392 images, 24 classes) to determine how MB-BE's performance in classifying the freshness of fish eyes. Other datasets were also used, such as Caltech 101 (9000 images, 101 classes) and Coil (7200 images, 100 classes), to prove that MB-BE can solve not only the classification of fish's freshness but also other classification problems.

2. Literature Review

Non-destructive fish freshness detection is exciting because we can recognize the freshness of fish without touching it. In contrast to other conventional ways, for example, using chemical substance can damage the fish's body; using a wave sensor still requires special tools; while using software with a camera to classify fish freshness is a more straightforward and non-destructive. Research by (Taheri-Garavand et al., 2020) developed a system for classifying common carp fish using VGG16; the accuracy reaches up to 98.21%. Using CNN, the freshness classification of fish is a fast, low-cost, precise, non-destructive, real-time, and automated technique. Research by (Kunjulakshmi et al., 2020) developed a portable, non-destructive freshness detection system for Rastrelliger kanagurta fish stored under the ice. The experiment uses multiple linear regression to assess the relationship between pixel count and quality indicators in the fisheyes images. The system classified three freshness categories, namely: extremely fresh, fresh, and spoiled with accuracy up to 0.989. The freshness classification of milkfish eyes was also developed by (Prasetyo, Adityo, & Purbaningtyas, 2019; Prasetyo, Purbaningtyas, & Dimas Adityo, 2020)using Cosine Nearest Neighbor and color features. The system uses 71 milkfish eye images showing an accuracy of up to 60.89%. Another research conducted by (Mohammadi Lalabadi et al., 2020) combined eye and gill color feature to classify fish freshness using an artificial neural network and a support vector machine. Experimental results show that gills present more robust features than the eye with an accuracy of up to 96%. Proper CNN and a small number of fish species were supported previous research's success as feature extraction and classification; therefore, they obtain high classification performance. However, we use eight species divided into three freshness levels; accordingly, there are 24 classes in our research, making the distinguishing features more difficult to be studied.

3. Method

3.1. Depthwise Separable Convolution (DSC)

Convolution of f_{k-1} features map uses M number kernel kernels $W = \{w_1, w_2, \dots, w_n\}$ 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 as follows:

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

Standard convolution in Eq. (1) will require computation costs of $D_k \times D_k \times \mathrm{M} \times \mathrm{N} \times D_f \times D_f$, where D_f is the size of the result feature map. The major change to the convolution method with minimal costs is Depthwise Convolution (DC), where DC uses a single kernel for each feature map (Howard et al., 2017). This means, for example, if we have 64 features, we also need 64 single kernels, generally 3×3 in size, where each kernel is used to convolve one feature map. Convolution results with one DC are also 64 features using equation as follow:

$$f_k = \{\overline{w_1} \times I, \overline{w_2} \times I, \dots, \overline{w_N} \times I\} \qquad (2)$$

Where, $\overline{w_N}$ is a single kernel with size of $D_k \times D_k$.

Then, the DC output is convoluted using standard convolution with 1×1 kernel (pointwise convolution), so this procedure is called Depthwise Separable Convolution (DSC) as presented in Fig. 1. The cost of DSC computation is much smaller than the standard convolution, namely $D_k \times D_k \times \mathrm{M} \times D_f \times D_f, + \mathrm{M} \times \mathrm{N} \times D_f \times D_f,$.

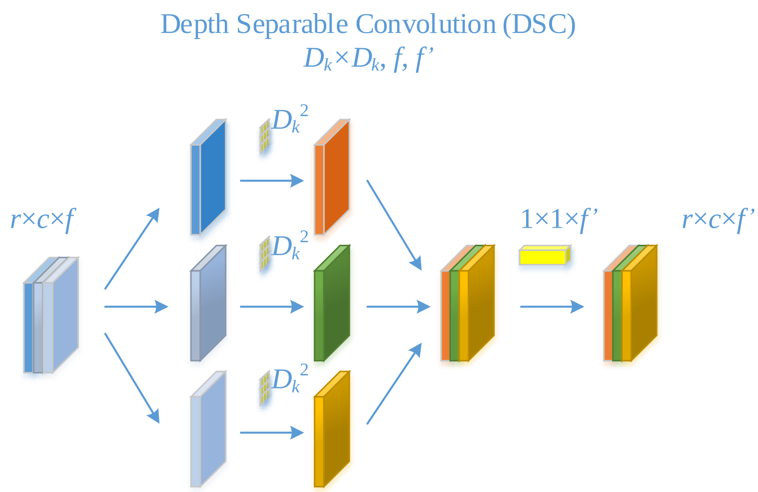

Fig. 1 Depthwise Separable Convolution (DSC)

We still use the DSC as the basis for light convolution in our proposed architecture, where the DSC is used in both the DSC-BE and the RT to reduce the number of parameters and recognize fish eye features. Convolution uses DSC-BE in each block to better achieve high-level features, while RT is the transitional convolution between blocks.

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

ResNet (He et al., 2016) dan ResNext (Xie et al., 2017) solve saturated performance problems during training using the skip connection in the bottleneck scheme. ResNext also adds 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. 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) × feature map, where $0 < BM \leq 1$. The DSC follows this to expand the number of feature maps to the original size. This bottleneck concept effectively improves features quality for both classification (Xie et al., 2017) and object detection (Redmon & Farhadi, 2017), while expansion is powerful for generating more detailed features with non-linear functions.

We 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. 2

shows the conversion of the feature map size of r × c × M to r × c × N. DSC-BE convolutes a bottleneck of BM, followed by a convolution of expansion to its original size.

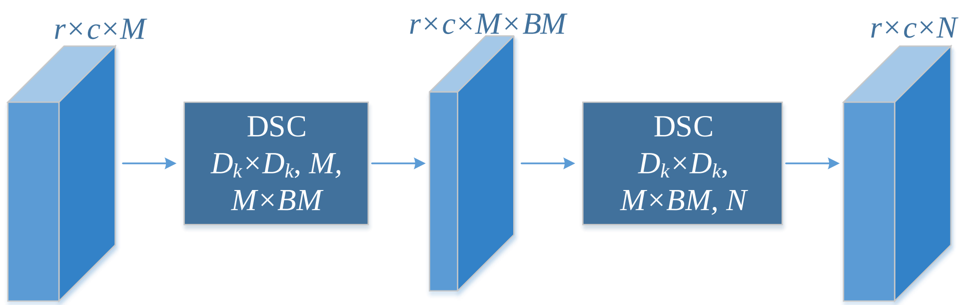

Fig. 2 Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE)

3.3 Residual Network

Residual Network (ResNet) is an architecture that introduces a concept of residual (skip connection) (He et al., 2016). Saturated performance problems during training are resolved by adding residues or remaining feature maps from the previous convolution layer. ResNet's original concept allows a feature map to jump over several layers to join the next layer. This procedure is performed by identity mapping, a feature map to the next layer using the adding operation. If $G(f)$ is a convolutional operation on a feature map, then the residual network is adding identity mapping as follows:

$$f_k = f_{k-1} + G_n(f_{k-1}) \quad (3)$$

Where $G_n(f)$ is n times convolutional skipped by a feature map f_{k-1}. This method is proven to be able to avoid saturated performance problems when the training is conducted with a high number of epochs.

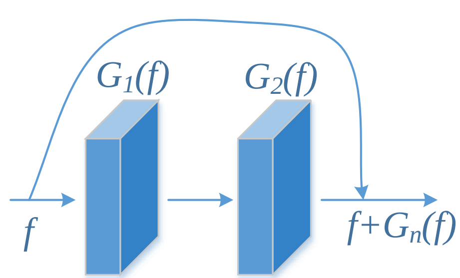

Fig. 3 Identity mapping by ResNet (He et al., 2016)

3.4 Residual Transition

In most CNN architectures, such as Densenet and MobileNet, the model is divided into several blocks where each block contains several layers of convolution. Generally, the size of the feature maps in a block is the same. By convention, the size of the number of feature blocks above is usually larger but smaller resolution. Change the size of the feature map from one block to the next using transitions, as in Densenet (G. Huang et al., 2017). The transitions on Densenet use 1 × 1 convolution and 22/2 max-pooling to obtain half of the resolution than before. So that we can add the skip connection from the previous block to the transition block, we can use additional padding or projection size (Xie et al., 2017).

We propose Residual Transition (RT) for bridging the skip connection in the transition block using 3×3 depthwise convolutions (DC) on the two feature maps. Then, we combine the results using concatenation, and then we convolute 1 × 1 convolution followed by batch normalization, relu activation, and max-pooling. As presented in Fig. 4, the feature map of size r × c × M has convoluted $G_n(f)$ to f, while f' is the feature map of the previous layer. Both f and f' feature maps can be involved in the transition block together to maintain the feature up to the transition block. We use DC and DSC as mechanisms to convolute but keep computational costs down. The RT output will become the input feature map for the next convolutional block.

The RT formula can be shown as follows:

$$f_k = DSC[DC(f_{k-1}) \cup DC(f'_{k-1})] \quad (4)$$

Where the operator ∪ is union operation, f_{k-1} is the current feature map, f'_{k-1} is the residual features from previous block.

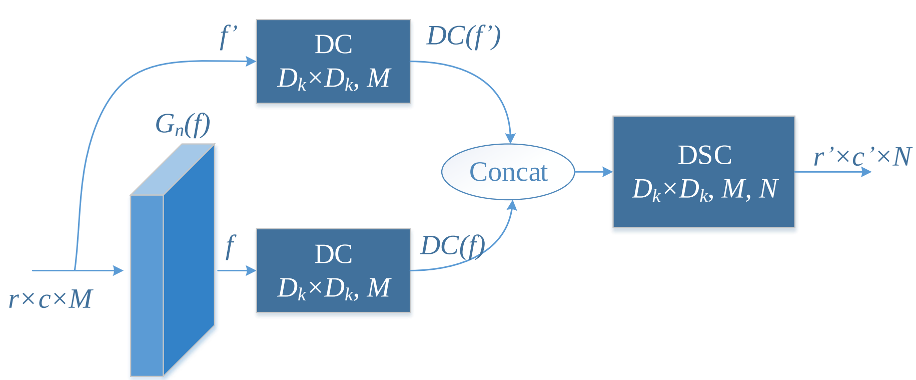

Fig. 4 Residual Transition

3.5 MobileNetV1 with Bottleneck and Expansion (MB-BE)

MobileNet V1 convolution layer is classified into ten blocks; the first block uses standard convolution, which produces 32 features; the next block uses DSC and down-sampling with max-pooling where the feature map increases by binary multiplication up to 1024 features in block ten. MobileNetV1 with Bottleneck and Expansion (MB-BE) is proposed as a new CNN architecture that partially inherits the MobileNet V1 architecture. As presented in Fig. 5, 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 pre-trained MobileNet V1 can be used as initial weights, where pre-trained MobileNet V1 has been trained using Imagenet (millions of images and thousands of classes). The pre-trained weights can recognize various image features, so it becomes a huge advantage if it is trained using the initial weight of pre-trained MobileNet V1.

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 1024, respectively done according to the Depth. As a feature map 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 are removed; then replaced using five new blocks to reduce the number of parameters and recognize fish eye features. As explained before, the fish's eye freshness is difficult to be distinguished because there are not many visual features, both internal within class and external between classes. 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 explained before, which in Fig. 5, it is assumed that BM = 0.5, meaning that in DSC-BE, a bottleneck convolution of 50% is carried out and followed with expansion to the original size. For example, in Fig. 5, DSC-BE 256 is convolved, then DSC-BE evolved bottleneck 128, followed with 256 features expansion. In addition, an ablation study is conducted to investigate the effect of BM on architectural measures and performance. Since MobileNet V1 is the backbone in MB-BE, the Bottleneck Multiplier = 1.0 parameter is also specified.

The depth of DSC-BE is also considered; it determines the depth of the DSC-BE convolution carried out. The Depth = 1 indicates DSC-BE 1 time; Depth = 2 indicates two times, and so on. Three blocks will pass using Depth, as shown in Fig. 5. Furthermore, the effect of the depth of DSC-BE is investigated to determine the adequate Depth value through an ablation study.

The model described earlier is a feature extraction block, whereas a classification block should be added in CNN. A fully connected layer with 1024 neurons is added in the hidden layer and one output layer with the number of neurons according to the experiment's dataset.

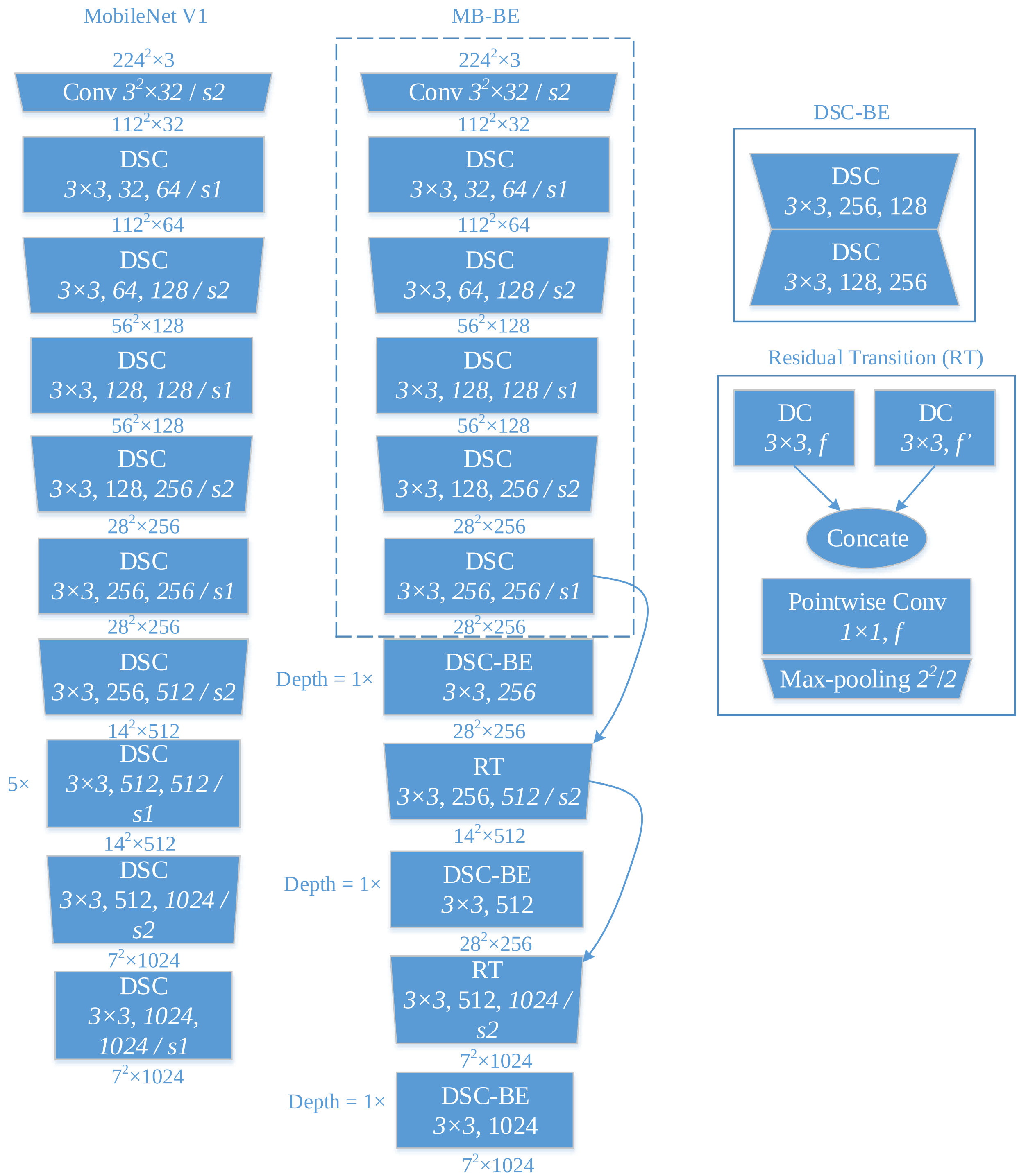

Fig. 5 MobileNetV1 with Bottleneck and Expansion (MB-BE)

3.6 Framework Research Classifying the Freshness of Fish Eyes

The system framework for classifying the freshness of the fish eye, as presented in Figure 5.6, begins with the fish eye's segmentation to focus processing on the fish eye. This section is conducted using the object detection method; accordingly, in this paper, we assume that the images input in the model is the segmented fish eye. For 4392 images in the dataset, we divided them into training, validation, and testing data. This data splitting will be explained in the next section. The proposed model will be trained using training data and validated using data validation in the experimental session. Furthermore, the trained model is used to classify the test data. It is essential to classify the test data for determining the model's reliability in classifying images that have not been seen during the experiment session. Next, we evaluated the classification results on all data with accuracy metrics and in-depth analysis

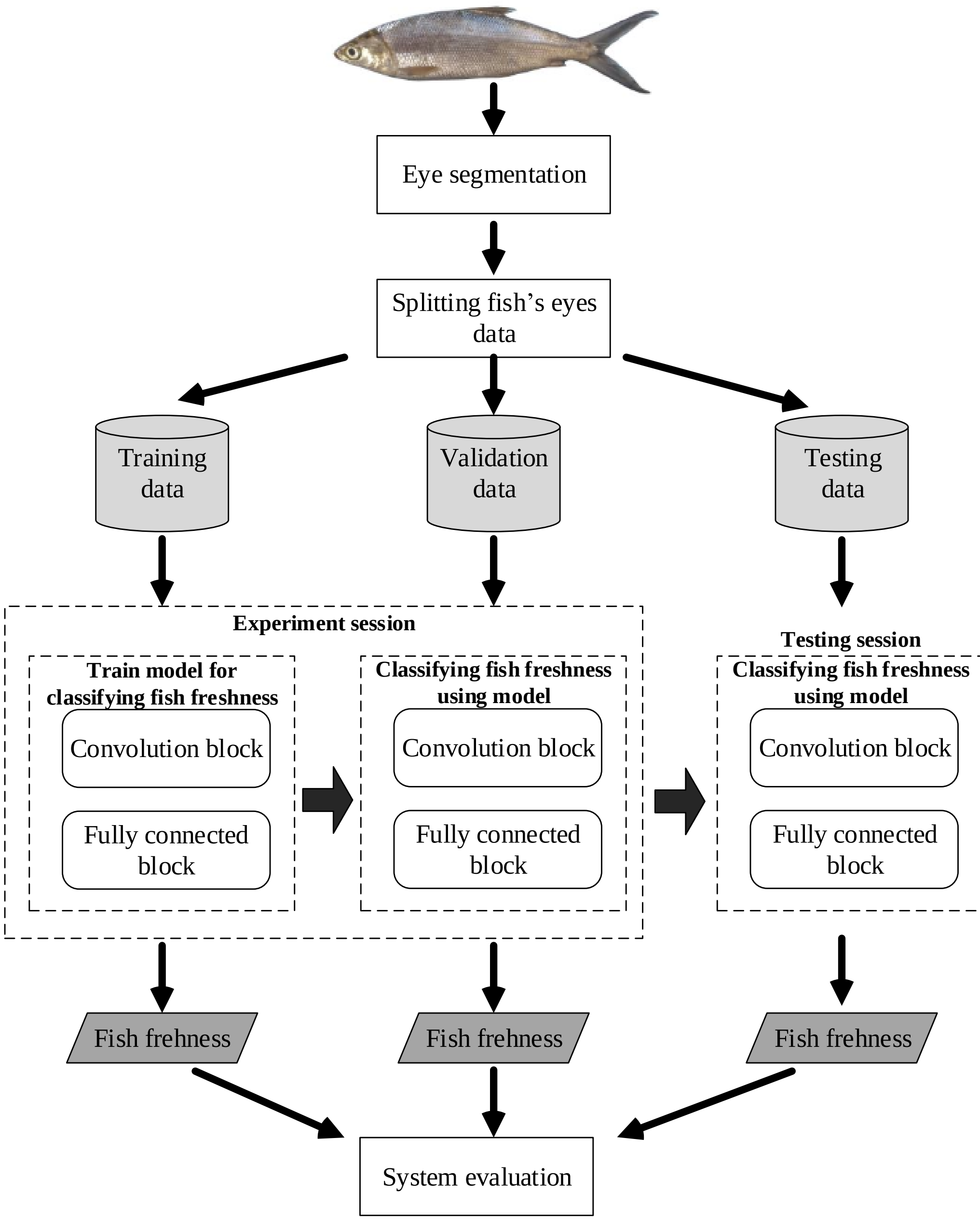

Fig. 6 Framework Research Classifying the Freshness of Fish Eyes

4. Result and Discussions

4.1 Dataset

We used the Freshness of the Fish Eyes (FFE) dataset (Prasetyo, Adityo, Suciati, & Fatichah, 2020) to evaluate our proposed model's performance. This dataset consists of 4392 images of fish eyes, consisting of eight fish species; each species consists of highly fresh (day 1 and 2), fresh (day 3 and 4), and not fresh (day 5 and 6) (https://data.mendeley.com/datasets/xzyx7pbr3w/draft?a=f24e65b8-e5bf-4f08-913a-749634eb383b). The eight fish species as follows Chanos Chanos (500 images), Johnius Trachycephalus (240 images), Nibea Albiflora (421 images), Rastrelliger Faughni (769 images), Upeneus Moluccensis (792 images), Eleutheronema Tetradactylum (240 images), Oreochromis Mossambicus (625 images), and Oreochromis Niloticus (805 images). It is not easy to differentiate the freshness of fish eyes because there are not adequate features internal within classes and external between classes. We also conducted an ablation study on our model using the FFE dataset to evaluate how effective our model solves classifying the freshness of fish eyes. We also used Caltech-101 (Fei-Fei et al., 2004) and Coil-100 (Nene et al., 1996) as other datasets to evaluate our model's ability to classify other cases.

4.2 Experimental Setting

We used pre-trained MobileNet V1 (Howard et al., 2017), publicly available, as a backbone architecture with trained weights from Imagenet to complete classifications of fish eye freshness. Our model also used an input image of 224 × 224 pixels in size. We connected our proposed feature extraction model block with a classification block consisting of a fully connected layer with 1024 neurons as the hidden layer. Next, we added an output layer with the number of neurons according to the number of classes in the dataset. We also added a 0.5 dropout. We used RMSProp with a learning rate of 1e-5 and a loss function categorical cross-entropy for the optimizer. We divided the dataset into training, validation, and testing with the respective percentages of 60, 20, and 20, respectively. For the FFE dataset, we used batch-size 24 and 22 for training and validation, respectively. For Caltech-101, we used batch-sizes 30 and 31; while for Coil-100, we used batch-sizes 44 and 14.

4.3 Ablation Study

We conducted an ablation study to determine the appropriate configuration of the MB-BE parameters to complete the freshness classification of fish eyes. MB-BE uses MobileNet V1 as the backbone model, and then it is connected to three times DSC-BE and two times RT. In each DSC-BE block, we investigated the depth of BSC-BE, i.e. how many times the DSC-BE convolution was carried out in order to achieve high performance. We recorded MB-BE (first second third) as the notation of the depth of DSC-BE (convolution was conducted on the block). The higher the depth of DSC-BE, the more parameters were utilized (deeper convolution was conducted). We also investigated the appropriate BM values by evaluating BM = 0.3, 0.5, 0.7, and 0.9. We combined these two parameters in the ablation study to obtain an adequate configuration for achieving optimal performance.

The results of the ablation study are presented in Table 1, where we evaluated the BM from 0.3 to 0.9 at the same depth of DSC-BE, i.e. (1 1 1). The results show that the performance of training, validation, and testing has a similar pattern, where accuracy is low for small BM, where for BM = 0.3, the system achieves an accuracy of 50.30%, 51.03%, and 52.96% for training, validation, and testing, respectively. For BM between 0.3 and 0.9, the performance peaked at 0.5, when the accuracy reached 56.75%, 63.21%, and 60.12% for training, validation, and testing, respectively. When we raise the BM, the MB-BE performance contracts considerably falling to approximately 55%. As explained earlier, BM determines the bottleneck-level; the smaller the BM, the higher the bottleneck occurs. BM also affects the number of parameters; the higher the bottleneck (small BM), the smaller parameters, and vice versa. Our ablation study results reveal that too large or small a bottleneck does not ensure better performance. Our findings suggest that the best bottleneck-level is BM = 0.5.

We additionally conduct an ablation study for the same BM with various Depths; this is to discover the Depth that provides the most optimal performance. Ablation also pays attention to the number of parameters, regarded that MobileNet is identic with small parameters; this means that the greater the Depth, the more convolution is carried out (more parameters are required). Therefore, we additionally consider the simplicity of the architecture (small number of parameters), in order to the model is proper for application on mobile devices. The results of the ablation study at the Depth level (2 5 1), (1 5 1), and (1 2 2) show that the accuracy of all Depth alternatives does not exceed Depth (1 1 1), where the highest accuracy is 56.72%. We discovered that increasing Depth also increases the number of parameters, which more than 4 million parameters are utilized, but it does not improve performance. Therefore, the results of the ablation study also reveal that the best Depth is (1 1 1).

Table 1. Ablation study of MB-BE

Model	**Parameters (million)**	**BM**	**Accuracy (%)**		
			Train	**Val.**	**Test**
MB-BE (1 1 1)	2.33	0.3	50.30	51.03	52.96
MB-BE (1 1 1)	**3.16**	**0.5**	**56.75**	**63.21**	**60.02**
MB-BE (1 1 1)	3.44	0.7	55.69	59.00	59.23
MB-BE (1 1 1)	4.00	0.9	55.73	59.00	59.00
MB-BE (2 5 1)	4.05	0.5	44.99	49.09	48.41
MB-BE (1 5 1)	4.50	0.5	44.46	47.95	49.43
MB-BE (1 2 2)	4.52	0.5	51.63	56.72	55.81

We compared the performance of MB-BE in classifying the freshness of fish eyes using the FFE dataset. The results presented in Table 2 show that ResNet50 reaches the best accuracy of training data, validation, and testing with an accuracy of 84.86%, 78.47%, and 78.82%; while MB-BE achieved an accuracy of 56.75%, 63.21%, and 60.02% for training, validation, and testing, respectively. However, the performance of ResNet50 is higher than MB-BE, but the parameters utilized by ResNet50 are a vast number than MB-BE, approximately seven times larger, where ResNet50 uses 23.59 million parameters while MB-BE is only 3.16 million parameters.

Compared to other CNN architectures, MB-BE further outperforms MobileNet, Densenet, and VGG16, where all other CNNs reach an accuracy of approximately 34% to 59%. In terms of the number of parameters, MB-BE is more excellent than other CNNs except for MobileNet2, where MB-BE uses 3.16 million parameters while the other CNNs use more than 3.2 million parameters, and MobileNet2 uses 2.25 million parameters; nevertheless, the accuracy reached by MB-BE is higher than MobileNet2. We further investigated this experiment using other classification datasets, namely Caltech-101 and Coil-100.

Table 2. Performance evaluation using FFE dataset

Model	**Parameters (million)**	**Accuracy (%)**		
		Train	**Val.**	**Test**
MobileNet V1	3.22	53.87	57.97	59.11
MobileNet V2	2.25	54.29	55.35	53.87
ResNet50	23.59	**84.86**	**78.47**	**78.82**
DenseNet 121	7.04	35.96	43.05	42.37
VGG16	14.71	34.26	41.00	43.85
MB-BE (1 1 1)	3.16	**56.75**	**63.21**	**60.02**

In the end, the model performance in session testing determined the system's robustness when implemented in real cases. The higher the testing performance, the more robust the system completes the classification. Our experimental results show that the highest performance on the test data is ResNet50 and MB-BE; both outperform all state-of-the-art. ResNet50 outperforms the accuracy of MB-BE, but MB-BE utilizes highly smaller parameters than ResNet50 so that MB-BE is proper for classifying the freshness of fish eyes.

4.2 MB-BE for Image Classification

To prove MB-BE's performance in solving other classification problems, we experimented on the Caltech-101 and Coil-100 datasets. Caltech-101 is a very familiar object classification dataset that consists of 9144 images, divided into 102 classes. The number of images ranges from 31 to 800 pictures for each class; it means Caltech-101 is an imbalanced dataset. The experimental results on the Caltech-101 dataset presented in Table 3 show that MB-BE was not robust in completing the imbalanced dataset, where the performance achieved in the training session was 78.56%. This performance was superior to Densenet and VGG16 but outperformed by MobileNet and Resnet50. Resnet50 achieved the highest performance with an accuracy of 93.78%. It stands to reason that Resnet also uses a vast number of parameters. The performance of MB-BE in session validation and testing was also similar; MB-BE achieved an accuracy of 70.11% and 68.56% for validation and testing, respectively. This result outperformed VGG16 but also outperformed MobileNet, Densenet, and Resnet50. The highest accuracy was achieved by Resnet50 with an accuracy of 92.73% and 92.67% for validation and testing.

Table 3. Performance evaluation using Caltech-101

Architecture	**Train**	**Val.**	**Test**
Mobilenet V1	80.51	89.40	89.67
Mobilenet V2	81.99	88.63	89.50
Densenet	71.38	83.06	83.98
Resnet50	**93.78**	**92.73**	**92.67**
VGG16	42.71	59.67	51.94
MB-BE	78.56	70.11	68.56

Columbia University Image Library (Coil-100) is a color image dataset of 100 objects that have been arranged against a black backdrop on a motorized turntable. Concerning a fixed color camera, the turntable was rotated by 360 degrees to vary the object pose. Object images were taken at 5-degree pose intervals, which correspond to 72 poses per object (7200 images in total). The experimental results on the Coil-100 dataset presented in Table 4 show that all models achieve high accuracy. Only Densenet and VGG16 achieved accuracy below 90% for training sessions, while others achieved accuracy above 90%, and MB-BE achieved the best performance with 99.36% accuracy. In the validation and testing session, all models achieved accuracy above 96%, Resnet achieved 100% accuracy, while MB-BE achieved 99.93% accuracy. Although the MB-BE performance is not the best, the difference with Resnet is only 0.07%. With an accuracy of 99.93% outperforming other state-of-the-art, MB-BE has achieved very optimal performance. Other models such as MobileNet, Densenet, and VGG16 have achieved optimal results as well, where all of them are more than 96%, but MB-BE is still superior to others. 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.

Table 4. Performance evaluation using Coil-100

Architecture	**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
MB-BE	**99.36**	99.86	99.93

4.3 Analysis of MB-BE Performance for Image Classification

Classifying the fish eyes' freshness is a classification problem where the fresh and not fresh of fish eyes are slightly different; there are not many adequate features can be observed closely. The images of fish eyes, both fresh and not fresh, cannot vary as in the general classification dataset, so the proper CNN model for solving it may not be as complicated as models in general, such as Resnet, Densenet, and VGG16. We require a lightweight model to be implemented on mobile devices; therefore, MobileNet is a proper alternative because of its small parameter compared to other CNNs. However, MobileNet could not classify fish eyes' freshness optimally because of the straightforward convolution flow architecture and not adequate features to distinguish fresh and not fresh fish eyes. MB-BE, as our proposed CNN architecture, 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 (1 1 1) 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%. Although this performance is still below 70%, MB-BE achieved the most optimal performance with a lighter architecture compared to the state-of-the-art.

5. Conclusions

Our experimental results show that MobileNetV1 with Bottleneck and Expansion (MB-BE) can relatively classify the freshness of the fish eye with the most optimal accuracy, i.e., 63.21% exceeding other models such as MobileNet, VGG16, and Densenet. The application of Depthwise Separable Convolution Bottleneck with Expansion (DSC-BE) and Residual Transition (RT) has proven to help MB-BE reach its best performance learning slightly different features using smaller parameters. Despite this study, we also employed augmentation to grow data variation, but MB-BE's performance was still below 70%. We require to improve performance using other data variation adding strategies, for example, adding variations with Variational Auto Encoder or others.

Acknowledgment

We thank the Research and Community Service Directorate (DRPM) DIKTI Indonesia for supporting the research of authors in the 2020 Basic Research Scheme at the University of Bhayangkara Surabaya under the contract number 008/SP2H/AMD/LT/MULTI/L7/2020 on 10 June 2020 and 048/VI/AMD/LPPM/2020/UBHARA on 11 June 2020.

Declaration of interests

none.

References

Abu Mallouh, A., Qawaqneh, Z., & Barkana, B. D. (2019). Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images. *Image and Vision Computing*, *88*, 41–51. https://doi.org/10.1016/j.imavis.2019.05.001

Badan Pusat Statistik Indonesia. (2015). *Ringkasan Eksekutif Pengeluaran dan Konsumsi Penduduk Indonesia*. Retrieved from https://media.neliti.com/media/publications/48424-ID-ringkasan-eksekutif-pengeluaran-dan-konsumsi-penduduk-indonesia-berdasarkan-hasi.pdf

Fei-Fei, L., Fergus, R., & Perona, P. (2004). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPR.2004.383

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.90

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … Adam, H.

(2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. Retrieved from http://arxiv.org/abs/1704.04861

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. https://doi.org/10.1109/CVPR.2017.243

Huang, J. C., Huang, H. C., & Liu, H. H. (2020). Research on the parallelization of image quality analysis algorithm based on deep learning. *Journal of Visual Communication and Image Representation*, *71*, 102709. https://doi.org/10.1016/j.jvcir.2019.102709

Ismail, N., & Malik, O. A. (2021). Real-time Visual Inspection System for Grading Fruits using Computer Vision and Deep Learning Techniques. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2021.01.005

Jalal, A., Salman, A., Mian, A., Shortis, M., & Shafait, F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*, *57*, 101088. https://doi.org/10.1016/j.ecoinf.2020.101088

Kasinathan, T., Singaraju, D., & Uyyala, S. R. (2020). Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2020.09.006

Kunjulakshmi, S., Harikrishnan, S., Murali, S., D'Silva, J. M., Binsi, P. K., Murugadas, V., … Samuel, M. P. (2020). Development of portable, non-destructive freshness indicative sensor for Indian Mackerel (Rastrelliger kanagurta) stored under ice. *Journal of Food Engineering*, *287*, 110132. https://doi.org/10.1016/j.jfoodeng.2020.110132

Mahmood, A., Bennamoun, M., An, S., Sohel, F., & Boussaid, F. (2020). ResFeats: Residual network based features for underwater image classification. *Image and Vision Computing*, *93*, 103811. https://doi.org/10.1016/j.imavis.2019.09.002

Mohammadi Lalabadi, H., Sadeghi, M., & Mireei, S. A. (2020). Fish freshness categorization from eyes and gills color features using multi-class artificial neural network and support vector machines. *Aquacultural Engineering*, *90*, 102076. https://doi.org/10.1016/j.aquaeng.2020.102076

Murakoshi, T., Masuda, T., Utsumi, K., Tsubota, K., & Wada, Y. (2013). Glossiness and Perishable Food Quality: Visual Freshness Judgment of Fish Eyes Based on Luminance Distribution. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0058994

Nene, S., Nayar, S., & Murase, H. (1996). Columbia Object Image Library (COIL-20). *Technical Report*.

Prasetyo, E., Adityo, R. D., & Purbaningtyas, R. (2019). Classification of segmented milkfish eyes using cosine K-nearest neighbor. *Proceedings of ICAITI 2019 - 2nd International Conference on Applied Information Technology and Innovation: Exploring the Future Technology of Applied Information Technology and Innovation*, 93–98. https://doi.org/10.1109/ICAITI48442.2019.8982124

Prasetyo, E., Adityo, R. D., Suciati, N., & Fatichah, C. (2020). *The freshness of fish eyes dataset*. https://doi.org/http://dx.doi.org/10.17632/xzyx7pbr3w.1

Prasetyo, E., Purbaningtyas, R., & Dimas Adityo, R. (2020). Cosine K-Nearest Neighbor in Milkfish Eye Classification. *International Journal of Intelligent Engineering and Systems*, *13*(3). https://doi.org/10.22266/ijies2020.0630.02

Razzaghi, P., Abbasi, K., & Bayat, P. (2020). Learning spatial hierarchies of high-level features in deep neural network. *Journal of Visual Communication and Image Representation*, *70*, 102817. https://doi.org/10.1016/j.jvcir.2020.102817

Redmon, J., & Farhadi, A. (2017). Yolov3. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2017.690

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation Mark. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00474

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.308

Taheri-Garavand, A., Nasiri, A., Banan, A., & Zhang, Y. D. (2020). Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish. *Journal of Food Engineering*, *278*, 109930. https://doi.org/10.1016/j.jfoodeng.2020.109930

Wang, C. Y., Mark Liao, H. Y., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPRW50498.2020.00203

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. https://doi.org/10.1109/CVPR.2017.634

Xiong, L., Zhang, D., & Zhang, Y. (2020). Water leakage image recognition of shield tunnel via learning deep feature representation. *Journal of Visual Communication and Image Representation*, *71*, 102708. https://doi.org/10.1016/j.jvcir.2019.102708

Zheng, M., Lei, Z., & Zhang, K. (2020). Intelligent detection of building cracks based on deep learning. *Image and Vision Computing*, *103*, 103987. https://doi.org/10.1016/j.imavis.2020.103987

Ubhara Surabaya

Eko Prasetyo <eko@ubhara.ac.id>

Decision on submission to Information Processing in Agriculture

1 message

Information Processing in Agriculture <em@editorialmanager.com> 18 January 2022 at 21:16
Reply-To: Information Processing in Agriculture <ipa_cau@163.com>
To: Eko Prasetyo <eko@ubhara.ac.id>

Manuscript Number: IPA-D-21-00070R8

Combining MobileNetV1 and Depthwise Separable Convolution Bottleneck with Expansion for Classifying the Freshness of Fish Eyes

Dear Mr. Prasetyo,

Thank you for submitting your manuscript to Information Processing in Agriculture.

I am pleased to inform you that your manuscript has been accepted for publication.

My comments, and any reviewer comments, are below.
Your accepted manuscript will now be transferred to our production department. We will create a proof which you will be asked to check, and you will also be asked to complete a number of online forms required for publication. If we need additional information from you during the production process, we will contact you directly.

We appreciate you submitting your manuscript to Information Processing in Agriculture and hope you will consider us again for future submissions.

Kind regards,
Xiangyun Guo
Editorial Office

Information Processing in Agriculture

Editor and Reviewer comments:

More information and support

FAQ: When and how will I receive the proofs of my article?
https://service.elsevier.com/app/answers/detail/a_id/6007/p/10592/supporthub/publishing/related/

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

---