

An improved efficientnet B7 underwater marine organism image classification model based on the spatial group enhancement attention mechanisms

Zhengjie Wang

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China

Corresponding author: 2135070522@st.usst.edu.cn

Abstract. In this paper, we adopt the Efficientnet B7 model, which is an excellent performer in deep learning algorithms, and improve it by combining with the Spatial Group Enhancement attention mechanism to provide an innovative method for underwater biological image classification. After a reasonable division of the dataset, we introduced the improved Efficientnet B7 model for training. By observing the change of the loss curve on the training set, it can be seen that the loss value of the model gradually decreases from 0.6 to 0.05 and tends to be stable, which indicates that the classification prediction of the model is gradually close to the real result, and it can classify the underwater marine organisms images effectively. The accuracy on the training set gradually increases from the initial 80.12% to 98.31%, and the recall also grows from 75.82% to 98.93%, showing that the model predicts well. The performance on the test set shows an accuracy of 84.86%, a recall of 82.40%, and an F1 value of 80.2%, indicating that the model has good generalisation ability. This means that our proposed improved methods and ideas have achieved significant results in the task of underwater biological image classification, providing a useful reference for research in related fields. By comparing the experimental results, it can be seen that the optimisation of the Efficientnet B7 model by introducing the Spatial Group Enhancement attention mechanism achieves better performance in the underwater marine organisms image classification task, which lays the foundation for the future application of deep learning in underwater environments. This study is of positive significance for the promotion of underwater biology research, marine protection and other fields, and provides new ideas and methods for carrying out more in-depth related research.

Keywords: Efficientnet B7, Spatial Group Enhancement Attention Mechanisms, Image Classification.

1. Introduction

Underwater organism image classification refers to the use of computer vision technology to automatically identify and classify different organisms in images captured in the underwater environment. Effective classification of underwater organisms can help scientists better understand the marine ecosystem, monitor changes in the marine environment, and protect marine living resources [1]. However, due to the complex underwater lighting conditions, image blurring, noise interference and other factors, traditional image processing methods are difficult to accurately classify underwater

organism images, so the introduction of deep learning algorithms has become the key to solve this problem.

Deep learning algorithms play an important role in underwater image classification. Compared with traditional machine learning methods, deep learning algorithms are able to learn feature representations from a large amount of data, with more powerful representation and generalisation capabilities. In the task of underwater biological image classification, convolutional neural network (CNN) [2] is usually used as the infrastructure to extract image features through multi-level convolution and pooling operations, and combined with a fully connected layer for classification and identification. Deep learning algorithms can effectively handle complex and variable image information in underwater environments and improve classification accuracy and robustness.

In addition to traditional supervised learning methods, some deep learning models based on techniques such as weakly supervised learning [3], migration learning [4] and meta-learning [5] have appeared in recent years, which further enhance the effectiveness of underwater biological image classification. For example, by introducing the migration learning technique, model parameters pre-trained in other domains can be migrated to the underwater biological image classification task, accelerating model convergence and improving accuracy [6]; while meta-learning enables the model to adapt to new categories or new environments more quickly. These emerging techniques provide new ideas for solving the problems of data scarcity and labelling difficulties in underwater biological image classification tasks.

In summary, in the context of today's rapid technological development, deep learning algorithms show great potential and broad prospects in the field of underwater biological image classification. In this paper, the Efficientnet B7 model, which has a better classification effect in deep learning algorithms, is selected and improved by using the Spatial Group Enhancement attention mechanism, which provides a new method and idea for underwater biological image classification. In the future, with the increasing size of the dataset, the performance of hardware equipment and the continuous improvement of algorithm optimisation, it is believed that deep learning will play a more and more important role in the classification of underwater organisms and bring more innovation and breakthroughs in marine scientific research and conservation work.

2. Data set sources and data analysis

In this paper, we choose an open source underwater marine life dataset, the dataset contains a total of five underwater marine life, namely penguins, puffins, seals, sea rays and sharks, we divide the dataset according to the ratio of 7:3, 70% of the images are classified as the training set, and the penguins, puffins, seals, sea rays and sharks have 334, 353, 286, 323 and 392 images respectively after divided into the training set, and there are 143, 155, 120, 129 and 166 images after divided into the test set. images and after division into test set penguins, puffins, seals, morays and sharks have 143, 155, 120, 129 and 166 images respectively. Example images of penguins, puffins, seals, sea rays and sharks are shown in Fig. 1.



Figure 1. Selected data sets.
(Photo credit : Original)

3. Method

3.1. *Efficientnet B7*

EfficientNet B7 is an efficient and powerful convolutional neural network model proposed by Google, and it is the largest and most powerful version of the EfficientNet family [7]. The design philosophy of

EfficientNet B7 is to improve the accuracy and generalisation of the model while maintaining its efficiency. EfficientNet B7 achieves excellent performance by combining three important components of the deep neural network domain: network depth, network width, and resolution, as well as using a composite scaling approach to balance these three aspects. The network structure of EfficientNet B7 is shown in Figure 2.

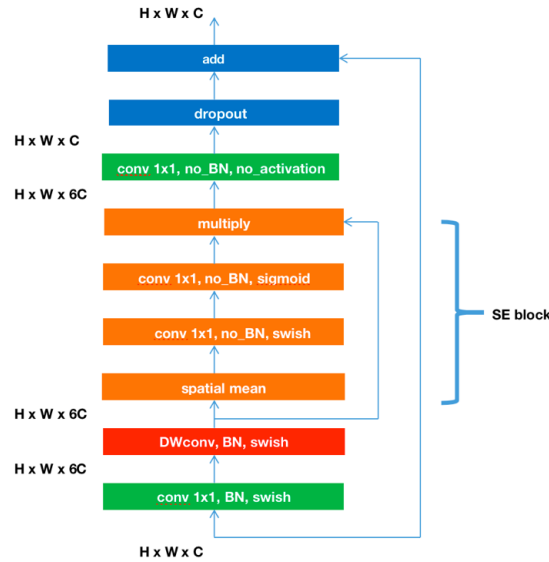


Figure 2. The network structure of EfficientNet B7.
(Photo credit : Original)

First, EfficientNet B7 improves the expressive power of the model by increasing the network depth. By increasing the depth of the layers, the model can learn more complex and abstract feature representations, which improves its ability to fit complex datasets. In EfficientNet B7, the authors used a method called Compound Scaling to simultaneously adjust network depth, width, and resolution to maintain the overall balance of the model [8]. This method can effectively improve the model performance without increasing the computational overhead.

Second, EfficientNet B7 enhances the model's ability to process feature dimensions by increasing the network width. Increasing the network width allows each layer to learn more dimensionally rich and discriminative quality feature representations, which helps to improve the accuracy of the model in recognition tasks. In addition, the balance between the number of parameters and the computational cost is taken into account in the design process to avoid too many parameters leading to overfitting or too high computational cost.

Finally, EfficientNet B7 improves the model's ability to capture detailed information by increasing the input image resolution. With the increase in resolution, the model can better capture the finer and richer information in the image, which improves the recognition accuracy of detail information such as complex scenes and small objects. In practical applications, appropriately increasing the input image resolution can bring better visual perception.

3.2. Spatial Group Enhancement

Spatial Group Enhancement (SGE) is a channel attention mechanism that can effectively increase the model's focus on important features to further optimise model performance [9]. Adding the SGE module to a powerful neural network model such as EfficientNet B7 can make the model more focused on key features and improve recognition accuracy [10]. The structure of EfficientNet B7 after adding the attention mechanism is shown in Figure 3.

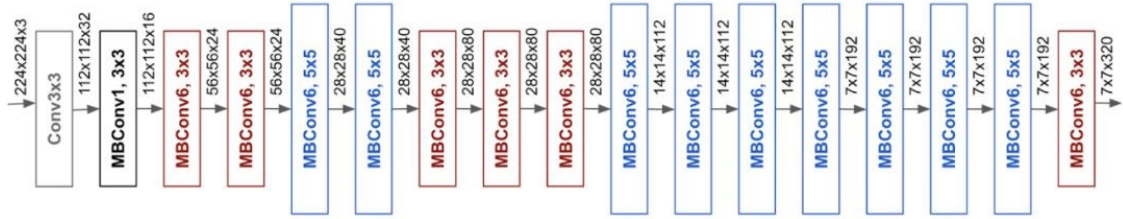


Figure 3. The structure of EfficientNet B7 after adding the attention mechanism.
(Photo credit : Original)

The principle of the SGE channel attention mechanism consists of two key parts: Spatial Attention and Channel Attention. Spatial Attention is used to capture the correlation between different locations in the feature map, and adjusts the importance weight of each channel by learning spatial information. In this way, the model can make better use of feature information at different locations, thus reducing the interference caused by redundant information to the model and improving the diversity and differentiation of feature representation.

On the other hand, Channel Attention is used to learn the correlation between different channels and adaptively adjust the weight of each channel. Through the Channel Attention mechanism, the model can better mine the interactions and dependencies between different channels, so that important features are given more weight while non-important features are suppressed. This mechanism helps to improve the model's ability to perceive key features in the dataset and reduce the impact of invalid information on the final output results.

In EfficientNet B7, the inclusion of the SGE channel attention mechanism after the last convolutional layer can make the model more intelligent and precise in performing the final feature fusion and classification. By introducing the SGE module, EfficientNet B7 can automatically learn and adjust the weight distribution of each channel in space and depth, which improves the network's ability to perceive and express key features in complex datasets.

4. Result

After dividing the dataset, the Efficientnet B7 underwater marine organisms image classification model improved based on the Spatial Group Enhance attention mechanism is imported, and this experiment is carried out based on python3.10, using the pytorch framework, with the machine memory of 32G, the batch size set to 12, the learning rate set to 0.0001, epoch is set to 85, record the change of the loss value during the training process and draw the loss curve change image, as shown in Figure 4.

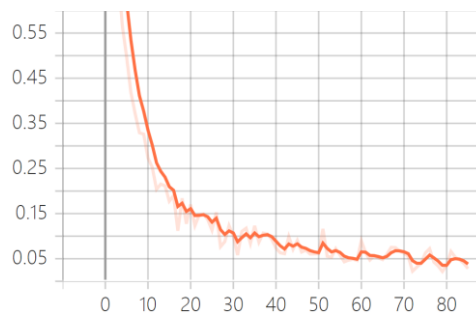


Figure 4. The change of the loss value.
(Photo credit : Original)

As can be seen from the change of the loss curve in the training set, the loss value gradually decreases from 0.6 to 0.05 and tends to converge, indicating that the prediction results of the model are getting closer and closer to the real classification results, and that the model is able to classify the images of underwater marine organisms very well.

The changes of accuracy and recall during the output training process are shown in Fig. 5.

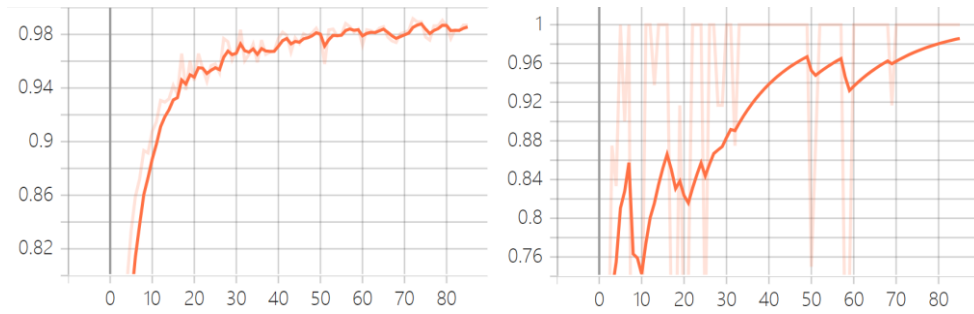


Figure 5. The changes of accuracy and recall.
(Photo credit : Original)

As can be seen in Fig. 5, the accuracy of the training set gradually increases from the initial 80.12% to 98.31% and tends to be stable, and the recall increases from the initial 75.82% to 98.93% and tends to be stable, and the model has a better prediction effect.

In the testing session, the heatmap image of the prediction results is output and three parameters such as accuracy, recall and F1 are calculated for the test set. The heatmap image is shown in Fig. 6 and the results of accuracy, recall and F1 are shown in Table 1.

Table 1. Selected data sets.

Evaluation parameters	Value
Accuracy	84.86%
F1	80.20%
Recall	82.40%

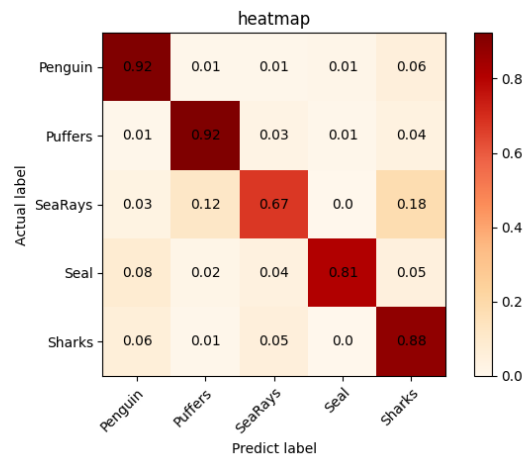


Figure 6. The heatmap image.
(Photo credit : Original)

From the test results of the test set, it can be seen that accuracy reaches 84.86%, recall reaches 82.40% and F1 reaches 80.2%, the model also shows good generalisation ability in the test set.

5. Conclusion

In this paper, an efficient deep learning algorithm, EfficientNet B7 model, is used and improved with Spatial Group Enhancement attention mechanism to provide an innovative approach for underwater biological image classification. After a reasonable division of the dataset, we successfully trained the EfficientNet B7 model based on the improved EfficientNet B7 model, and observed that the loss curves

on the training set showed a trend of gradually decreasing and converging to 0.05, which indicated that the model's ability to classify the images of underwater marine organisms was gradually improved, and the prediction results were closer to the real situation. The accuracy on the training set improves from the initial 80.12% to 98.31%, and the recall also improves from 75.82% to 98.93%, showing that the model has made significant progress in the classification task.

On the test set, we observe that the model achieves an accuracy of 84.86%, a recall of 82.40% and an F1 value of 80.2%, which indicates that the improved model has good generalisation ability on unseen data. By comparing the results of the training set and the test set, we can confirm that the model not only performs well on the training data, but also effectively recognises and classifies images of underwater marine organisms in practical applications.

In conclusion, this study optimises EfficientNet B7 by introducing the Spatial Group Enhancement attention mechanism, and achieves satisfactory results in the task of classifying underwater biological images. This approach not only improves the performance of the model on the training and test sets, but also provides a novel and effective technical tool for the field of underwater biology research. In the future, we can further explore how to combine more advanced techniques to further improve the model performance to cope with more complex and diversified underwater environments, and make more contributions to the protection of marine life and research work.

References

- [1] Zhou, Z., et al. "Improving the classification accuracy of fishes and invertebrates using residual convolutional neural networks." *ICES Journal of Marine Science* 80.5 (2023): 1256-1266.
- [2] Nandika, Muhammad Rizki, et al. "Assessing the Shallow Water Habitat Map** Extracted from High-Resolution Satellite Image with Multi Classification Algorithms." *Geomatics and Environmental Engineering* 17.2 (2023).
- [3] Lopez-Vazquez, Vanesa, et al. "Deep learning based deep-sea automatic image enhancement and animal species classification." *Journal of Big Data* 10.1 (2023): 37.
- [4] Ahmad, Usman, et al. "Large Scale Fish Images Classification and Localization using Transfer Learning and Localization Aware CNN Architecture." *Comput. Syst. Sci. Eng.* 45.2 (2023): 2125-2140.
- [5] Fan, Jianchao, and Chuan Liu. "Multitask gans for oil spill classification and semantic segmentation based on sar images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023): 2532-2546.
- [6] Aziz, Rabia Musheer, et al. "CO-WOA: novel optimization approach for deep learning classification of fish image." *Chemistry & Biodiversity* 20.8 (2023): e202201123.
- [7] Danqing Ma, , Shaojie Li, Bo Dang, Hengyi Zang, Xinqi Dong. "Fostc3net:A Lightweight YOLOv5 Based On the Network Structure Optimization." (2024).
- [8] Zongqing Qi, , Danqing Ma, Jingyu Xu, Ao Xiang, Hedi Qu. "Improved YOLOv5 Based on Attention Mechanism and FasterNet for Foreign ObjectDetection on Railway and Airway tracks." (2024).
- [9] Lu, Liangdong, Jia Xu, and Jiuchang Wei. "Understanding the effects of the textual complexity on government communication: Insights from China's online public service platform." *Telematics and Informatics* 83 (2023): 102028.
- [10] Kumari, Rina, et al. "Identifying multimodal misinformation leveraging novelty detection and emotion recognition." *Journal of Intelligent Information Systems* 61.3 (2023): 673-694.