İSTANBUL UNIVERSITY PRESS

EISSN 2602-473X

AQUATIC SCIENCES AND ENGINEERING

Aquat Sci Eng 2022; 37(4): 220-228 • DOI: https://doi.org/10.26650/ASE202221163202

Research Article

Multi-species Fish Identification using Hybrid DeepCNN with Refined Squeeze and Excitation Architecture

Jansi Rani Sella Veluswami[1], Nivetha Panneerselvam[1]

Cite this article as: Jansi Rani, S.V., & Nivetha, P. (2022). Multi-species fish identification using hybrid DeepCNN with refined squeeze and excitation architecture. *Aquatic Sciences and Engineering, 37*(4), 220-228. DOI: https://doi.org/10.26650/ASE202221163202

ABSTRACT

Fish play a prominent role in the food web and fish farming has value for both human consumption and tourist attractions. Due to the increasing importance of marine biodiversity, recognition of fish species has become a prominent task in monitoring the mislabelling of seafood and extinct species. This problem can be solved using traditional manual annotation on the images. To reduce manpower, cost, and tremendous time, deep learning approaches are used which always require large datasets. Therefore, fish species identification is a challenging task using disproportionately small data sets. In this research, we develop a new method by refining the squeeze and excitation network for the automatic fish species classification model to identify 23 different types of fish species. To achieve this, a hybrid framework using deep learning is proposed on a large-scale dataset and implemented transfer learning for a small-scale dataset. Deep learning methods can be used to identify fish in underwater images. In this study, we have proposed a new method of hybrid Deep Convolutional Neural Network (CNN) along with a Support Vector Machine (SVM) for classification. Additionally, the Squeeze and Excitation (SE) block has been improved for improved feature extraction. The proposed method achieved an accuracy of 97.90%. Then post-training with the small-scale dataset (Croatian) achieved an accuracy of 94.99% with an 11% improvement compared to Bilinear CNN (B-CNN) (Qui et al., 2018) and can be used in any underwater applications to identify fish species and avoid mislabelling of seafood.

Keywords: Deep Learning, Convolutional Neural Network, Squeeze and Excitation, Fish species, Fish4knowledge dataset

ORCID IDs of the author:
J.R.S.V. 0000-0003-2863-5465;
N.P. 0000-0002-4556-1034

[1]Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India

Submitted:
24.08.2022

Revision Requested:
06.10.2022

Last Revision Received:
08.10.2022

Accepted:
09.10.2022

Online Published:
19.10.2022

Correspondence:
Jansi Rani SELLA VELUSWAMI
E-mail:
svjansi@ssn.edu.in

INTRODUCTION

The average fish consumption boosted from 9.0 kg per capita in 1961 to 20.5 kg in 2018. The common fish intake extended from 9kg consistent with capita in 1961 to 20.5 kg in 2018 (Dagoudo, Qiang, & Solevo., 2022). Climatic changes, excess fishing and other human activities are the factors that affect the marine ecosystem and the fisheries. They also pressure the fish and their habitats. This increases the need for monitoring and managing the population of fish. But it is difficult in coastal areas where humans get directly involved. Failing to do so will result in the degradation of marine ecosystems and extinction of a specific fish species. For example, there is a huge shrinkage of salmon species in the Northwest Pacific which contributes to a major part of fisheries (Crozier et al.,2019). Therefore, a warning should be imposed by the government and aquatic management to preserve the endangered species. Fish identification helps biologists, academic researchers, and ocean scientists to determine the geological changes and the biomass level in oceans due to its prominence in marine science. Secondly, people buy seafood by believing the selling person or the label on the food packet. But often, people are cheated by seafood mislabelling (Chen et al.,2020). Often, Tilapia is

mislabelled/ substituted as Snapper (Naaum, Warner, Mariani, Hanner & Carolin.,2016). Many fish species are similar in taste and texture. Hence several retailers sell low-market value fish as high-market value fish. There arises an imbalance in prices due to mislabelling (Pollack et al.,2018) since seafood is a highly traded food commodity (Kroetz et al.,2020). Deep learning approaches can be used to fix these problems instead of the manual fish annotation in the images collected through sea divers.

With the advancement in internet technology, fish species classification uses computer vision technology. In the inception, machine learning algorithms (Fouad et al.,2013) were widely used where feature selection was done manually. Now, many research works are carried out using deep learning models (Villon et al.,2020) where feature selection is done automatically. Xu et al. (2021) proposed a method for small-scale unbalanced fish species identification in which they implemented Transfer learning and SE-ResNet152 on the Fish Pak dataset which has 915 images. The SE-ResNet152 network was employed in this study to extract fish image features of higher quality and improve fish species identification. The body, head, and scale datasets have classification accuracy ratings of 98.80%, 96.67%, and 91.25%, respectively. This study was able to solve the problem of small-scale and unbalanced datasets using their class-balanced focal loss function. The environment is varied and diversified in the actual development and processing, and even specific aspects of the fish are blocked, resulting in fish images with fewer feature information. Their approach still must be optimised for this situation. The suggested approach, on the other hand, does not take into account the impact of a complex environment on fish species identification.

For fish recognition and species identification in underwater habitats, Jalal et al. (2020) used a combination of GMM-YOLO and optical flow-YOLO. For the purpose of automatically identifying fish species present in coral reefs, Villon et al. (2018) analysed the performance of four models developed with the same CNN architecture. For underwater fish detection in the wild, Labao & Naval. (2019) proposed a cascaded deep network system with linked ensemble components for 18 underwater videos using R-CNN. Santos & Goncalves. (2019) proposed a CNN pretrained model Inception which classifies fish species, family, and order of a pantanal image dataset. In another study (Allken et al.,2019) they implemented fish species recognition using InceptionNet which pretrained on the ImageNet Classification dataset. Ovalle et al. (2022) used iObserver (Vilas et al.2020) for the input data collection and these images were annotated manually with the species name and size using Mask R-CNN.

For Morphological based fish species identification on a small-scale dataset (FishPak dataset with 915 images) HT, Rauf et al. (2019) used 32-Layer CNN architecture enhancing the network's extensive feature extraction capabilities using ResNet-50, GoogleNet, AlexNet, and LeNet-5. For fish recognition using an underwater drone with a Panoramic camera that is automated using Deep learning algorithms, Meng, Hirayama & Oyanagi. (2018) analysed the performance of three networks: AlexNet, GoogleNet, and LeNet. Prasetyo et al., (2021) proposed a new residual network strategy called MLR (Multi-level residual) by combining low-level features with high-level features using depth-wise separable convolution (DSC) for the FishKnowledge dataset. Zhang et al. (2021) extracted texture features after reducing the noise from the fish4knowledge dataset and implemented a Deep Neural Network (DNN). In Jin et al.'s (2022) study, they proposed an integrated two-stage spatial pooling method in the squeeze part of the SE Block which consists of a rich descriptor extraction by fusing these descriptors into a C-dimensional channel feature. An accurate re-weight score can be returned for channel attention. Using this method, we can get both local and global informative features, but computational cost is additional in the squeeze part.

For fine-grained fish species identification on small-scale data sets (Croatian fish dataset with 794 images), Qiu et al. (2018) suggested an enhanced transfer learning algorithm with refined SENet. This paper compared the experiments on B-CNNs, B-CNNs plus SE blocks, and B-CNNs plus refined SE blocks, and the highest accuracy reached was BCNN+SE - 71.80%. This paper managed to work on a small data set, but they didn't achieve results with great accuracy. This can be improved by using DeepCNN techniques. Our method outperforms their results without pre-processing for the same dataset i.e., the Croatian dataset.

To summarize, though much work has been done, there are still challenges in improving the accuracy of the classification of different species in a large, unbalanced dataset. The objective is to preserve the aqua ecological species. And hence we have worked with the large dataset Fish4Knowledge(F4K) (Phoenix, Huang & Fishera., 2013) where there are 23 different species. The proposed system aims at developing an automated system to identify fish species with less computation time.

The main contribution of this paper can be summarized as follows:

1. The proposed hybrid framework of the CNN model is combined with refined Squeeze Excitation (SE) and SVM to improve the overall performance of the model
2. This develops an automatic fish species classification method to identify 23 different fish species
3. The hybrid CNN-refined SE-SVM model achieves good classification performance for a large-scale unbalanced dataset using augmentation and also for a small-scale dataset
4. The experimental result comparison with existing works

The rest of the paper is organised as follows: Section 2 describes the methodology and the algorithm used, section 3 talks about the experimental results with comparisons, and section 4 concludes the article with the findings.

MATERIALS AND METHODS

Dataset

The Fish4Knowledge dataset consists of 27,370 images of fish that were generated from underwater fish videos that were taken off the coast of Taiwan. This dataset includes images of 23 spe-

cies. As part of the Fish4Knowledge project, the initial dataset developed includes the snapshots of fish underwater and uses binary masks that separate the fish from their backgrounds.

Proposed methodology

The proposed methodology aims at developing an automated fish species classification for a large dataset using the hybrid framework of CNN with SE and SVM. Approaches based on deep learning can be applied without having feature extraction. By regularly changing the weights, the model automatically learns the features. When there are more layers in a deep learning model, it is referred to as being deeper. In essence, deep CNN is just CNN with more layers.

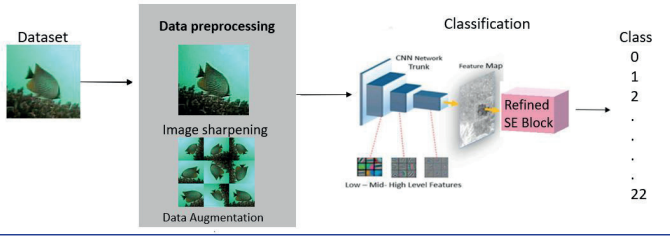


Figure 1. Proposed Architecture

Image preprocessing

As shown in Fig.1, the images from the dataset are subjected to image sharpening in the preprocessing stage. Before sharpening, the resolution of the image is slightly improved using a pre-trained model called LapSRN (Laplacian Pyramid Super-Resolution Network). LapSRN is a sequential super-resolution method that incorporates a coarse-to-fine Laplacian pyramid framework to super-resolve low-resolution images. LapSRN uses the Charbonnier loss function in Equ.1 instead of MSE loss. This loss function is robust enough to handle outliers. rs is assumed to be the s-level pyramid residual image and x_s denotes the image after up-sampling. The corresponding high-resolution (HR) image is $y_s = r_s + x_s$ and Y_s equivalent pyramid level is generated from the high-resolution images after down-sampling and the bicubic interpolation and loss function is Equ.1 and Equ.2.

$$\text{Loss}(Y, y; \theta) = \frac{1}{N}\sum_{i=1}^{i=N} \sum_{S=1}^{S=L} \delta(Y_s - y_s) \quad (1)$$

$$\text{Loss}(Y, y; \theta) = \frac{1}{N}\sum_{i=1}^{i=N} \sum_{S=1}^{S=L} \delta((Y - x_s) - r_s)) \quad (2)$$

where $\delta()$ in Equ.2 is $\delta(x) = \sqrt{X^2 + \varepsilon^2}$; X denotes $\delta((Y - x_s) - r_s))$; ε is a penalty which is a very small value. L represents the number of layer levels in the pyramid (L = 1, 2, 3); i is the image pixels, N denotes the number of pixels in the image. Usually, image sharpening is done using a Laplacian kernel where the sum of elements is 0 giving a binary image. So, we have used a modified kernel whose sum of elements is 1 giving a coloured image. Then, the sharpened images in the classes with a count less than 2000 is augmented. Since the dataset is unbalanced, image augmentation is performed. Images are randomly rotated to an angle of 90°,180° and 270° and flipped horizontally and vertically. After augmenting 1000 images per species, the total number of images increased from 27,370 to 45,360 images. The pre-processed images are sent to the deep convolutional neural network and Refined SENet for feature extraction. Then, classification of the species is done using SVM classifiers.

Feature extraction

Convolutional neural networks (CNN) can handle both feature extraction and classification methods. The performance enhancement with CNN for image classification, as illustrated in Fig.3, was remarkable. Convolution Filters use convolution to combine a kernel with the input image to produce feature maps. The pooling layer handles the down sampling procedure and can either employ Max pooling or Average pooling. Dense layers are utilised to link every neuron from the previous levels to the ones after them.

CNN architecture

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which consists of multiple hidden layers where the convolution layers consist of image maps and filters. A CNN has multiple hidden layers to extract low-level features from the input image. The important layers in a CNN are: (1) Input layer, (2) Convolutional layer, ReLU layer & Pooling layer which are feature extracting layers, (3) Fully Connected layer which is the classification layer.

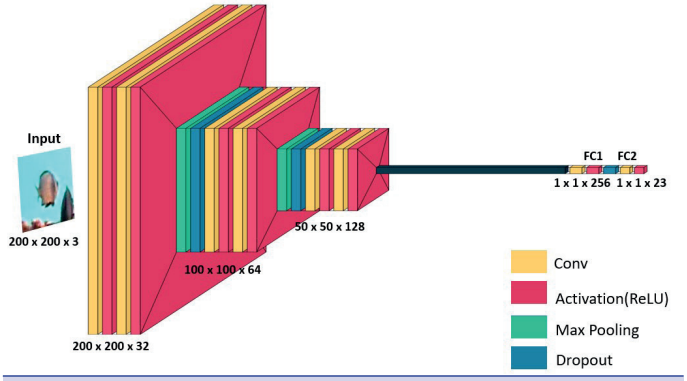


Figure 2. CNN Architecture

1. An RGB image is nothing more than a three-dimensional matrix of pixel values which is the input layer.
2. Convolution Filters use convolution to combine a kernel with the input image to produce feature maps. Lower-level features include edges or colour arrangement whereas higher-level features can recognise distinct fish shapes. The activation function, ReLU, performs an element-wise operation by setting all negative pixel values to zero. The pooling layers reduce the dimension of the feature maps. In the fish species identification, highlighted features are crucial in the images, as a result, we employed the max pooling operation, which chooses the largest element from the feature map region enclosed by the filter.
3. Dropout is an operation that ignores randomly selected neurons during training. It is a computationally cheap operation that prevents overfitting.
4. The pooled feature map is flattened to convert all the resultant 2-D arrays into a single linear vector and fed to a fully

connected layer to classify the image and get the final output.

Refined squeeze and excitation

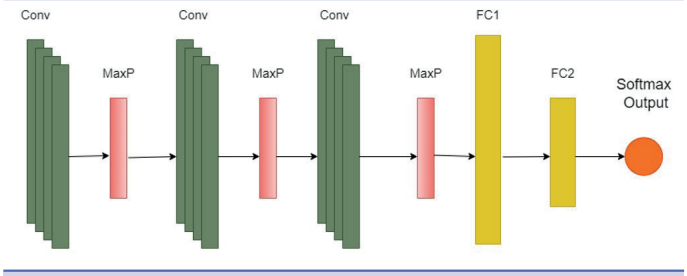


Figure 3. DeepCNN

The Squeeze-and-Excitation Block shown in Fig.5 is an architectural block that allows a network to implement dynamic channel-wise feature recalibration to improve its representational power. The SE block comes after every block of the Baselines and it can be used with any network. To generate an explicit channel relationship using global spatial features, the authors introduced a lightweight module called a squeeze and excitation block or an SE block which consists of a squeeze step and an excitation step (Hu, Shen & Son,2018).To improve the squeeze operation further we refined the squeeze operation by adding global average pooling, Equ.3, and global max-pooling layers, Equ.4, in order to get the benefits from both the layers

$$SQ^{c}_{avg} = \frac{1}{S' \times S'} \sum_{a=1}^{S'} \sum_{b=1}^{S'} (f_{a,b,c}) \quad (3)$$

$$SQ^{c}_{max} = \max_{a,b=1}^{S',S'} (f_{a,b,c}) \quad (4)$$

where S' denotes the modified dimensions and f ∈ RS' ×S' ×B' is the input feature map to the SE block, fa,b,c is the feature at (a,b) position. SQ^{c}_{avg} and SQ^{c}_{max} are the cth channel's squeezed values applying the global average and maximum pooling. The squeeze technique fundamentally extracts the channel-specific information. Further, the maximum pooling will keep the information in a local context, whereas the global pooling will retain the knowledge in a global context. The matrix is aggregated into a Squeeze-and-excite operation to produce a matrix that can emphasise information features and suppress less useful information channel-wise, and it is also proven to improve the image classification performance. For current state-of-the-art CNNs, SE blocks greatly enhance performance at a small additional computational cost (Hu et al., 2018). In order to get high performance and accuracy, we integrated CNN and a Refined SE (Squeeze and Excitation) Block.

Proposed CNN-SE architecture

The architecture of the CNN-SENet is depicted above in Fig.4. The input image of size 200 x 200 with 3 RGB channels is fed as input to the convolution block. The first and second iterations of this block have 32 filters in 3 x 3 each, followed by a ReLU activation function and then max pooling, followed by a dropout layer. The third and fourth iterations have 64 filters in 3 x 3 each, followed by a ReLU activation function and then max pooling, followed by a dropout layer. The fifth iteration has a convolution layer containing 128 filters in 3 x 3 followed by a ReLU and then a final iteration with 128 filters. This output is batch normalized and then enters the refined SE Block which is depicted in Fig.6

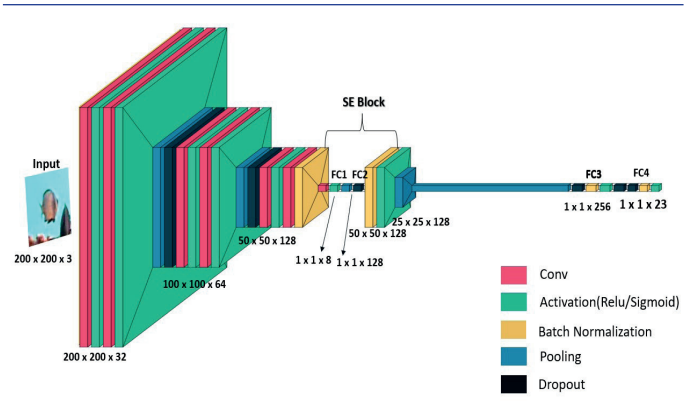


Figure 4. CNN-SE Architecture

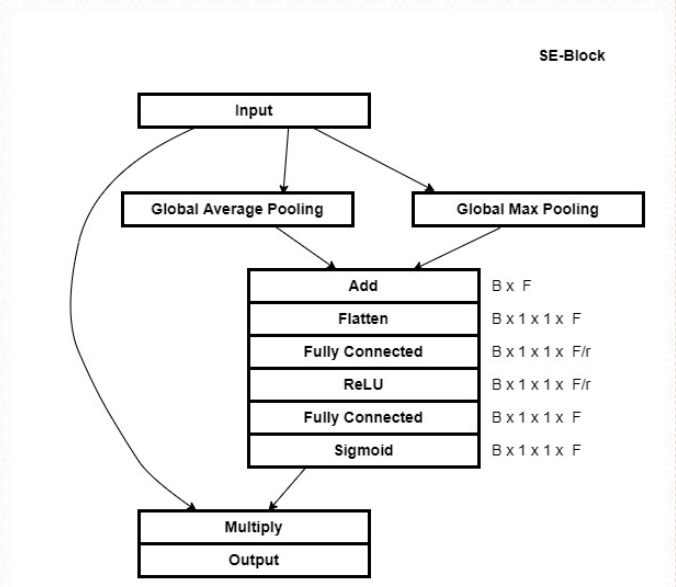


Figure 5. Refined SE Block

1. The input for the Refined SE block in Fig. 5 is a convolutional block.
2. Using global average pooling and global max pooling, each channel is "squeezed" into a single numeric value.
3. Two squeezing channels are taken into consideration to produce the excitation scores through the addition of global average pooling and global max pooling respectively.
4. Two fully connected layers with ReLU and Sigmoid activation are used:

- We use two fully connected layers to get non-linear relations
- We use independent sigmoid activation to get non-mutually exclusive channel relations
- The intermediate layers' node size is reduced for better generalization and to reduce computation overhead (F / r)

5. The output of the SE block (channel attention) is multiplied channel-wise to the original input (Hu et al.,2018)

This offers a CNN building block, shown in Fig.2, that improves network dependencies at nearly no cost in terms of computation. With the dimension size unchanged, this multiplied outcome is then extended to a rectified linear layer that performs element-wise activation. The result is then dimensionally reduced and resized by passing it through a Max Pooling layer (2x2). The network also comprises two fully connected (FC) levels. The first FC layer, which has 256 neurons, is flattened after max pooling. After batch normalising, the output from this fully connected layer, a reduction function, is applied. Before the final fully connected layer is executed, a 20% dropout layer is employed, which has 23 neurons. Softmax is the final layer, which uses a classifier function to calculate the probability distribution for each class. The Adam optimizer applies a categorical cross-entropy to the data.

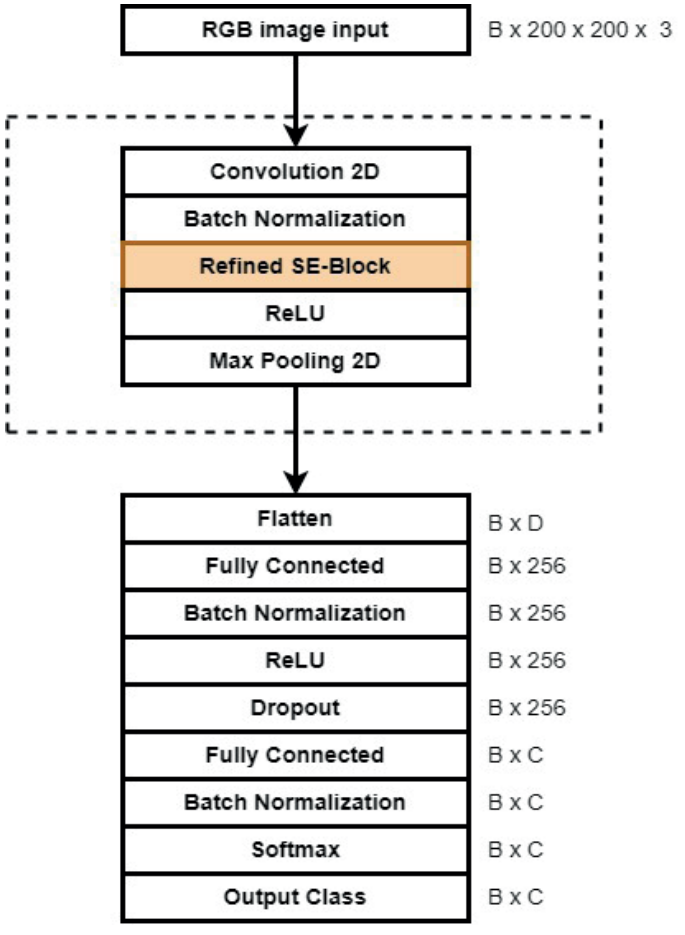


Figure 6. CNN SE Block

Training procedure

Adam was used as the optimiser and a batch size of 24 on various epochs with a default learning rate of 0.001. We divide the training set of images into batches of 24. It takes 1276 iterations to complete 1 epoch. Experiments are run on a system with an AMD Ryzen 3250U processor running at 260 GHz using 4 GB of RAM, running Windows version 10. The Keras and TensorFlow frameworks are used to implement the proposed work. The platform we employed for the execution was Google Collaboratory in which a GPU hardware accelerator is enabled.

Performance evaluation parameters

Accuracy, precision, f1-score and recall are the performance metrics used for performance comparison.

Accuracy: Accuracy is a popular metric in multi-class classification that may be calculated straight from the confusion matrix. Accuracy is a metric that indicates how well the model predicts the whole collection of data accurately. To validate the proposed hybrid model, training and testing is carried out using the Fish-4Knowledge dataset (Phoenix et al., 2013).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

True Negative (TN): the actual is not fish and predicted is also not fish
True Positive (TP): the actual is fish and predicted is also fish
False Negative (FN): the actual is fish and predicted is not fish
False Positive (FP): the actual is not fish and predicted is fish

Precision: A model's precision describes how many of the identified items are relevant. It is defined as the ratio of true positives to the sum of true positives for each class.

$$Precision = \frac{TP}{TP + FP}$$

Recall/Sensitivity: The number of positive class predictions made from all positive cases in the dataset is calculated.

$$Recall = \frac{TP}{TP + FN}$$

F1 score: The F1 score, which ranges from 0.0 to 1.0, is a weighted harmonic mean of recall and precision. The scores for each class indicate how accurate the classifier was in classifying the data points in that class in comparison to all other classes.

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Support: Support is the total number of actual count of each class in the test set. For example, in Fig.8, under the support column, fish_01 has 1197 test images.

Confusion Matrix: A confusion matrix, Fig.7, is a table that lists how many predictions a classifier made correctly and incorrectly. It is employed to evaluate a classification model's effectiveness. There are 23 classes in the reported study and hence the confusion matrix shown in Fig.7 has 23* 23 values.

RESULTS AND DISCUSSION

Experimental Results for the proposed method using the Fish4Knowledge dataset

After training the model through 50 epochs, we achieved sufficient accuracy and further training did not improve accuracy on the validation set. The model was then evaluated on the test dataset and the results are shown in Fig.8. The performance values of all 23 classes are listed. The classes, 10, 16, 19, 20, and 21, have an f1 score of 100%, and all others range from 93% to 99% except for class 8. The accuracy obtained is 98%.

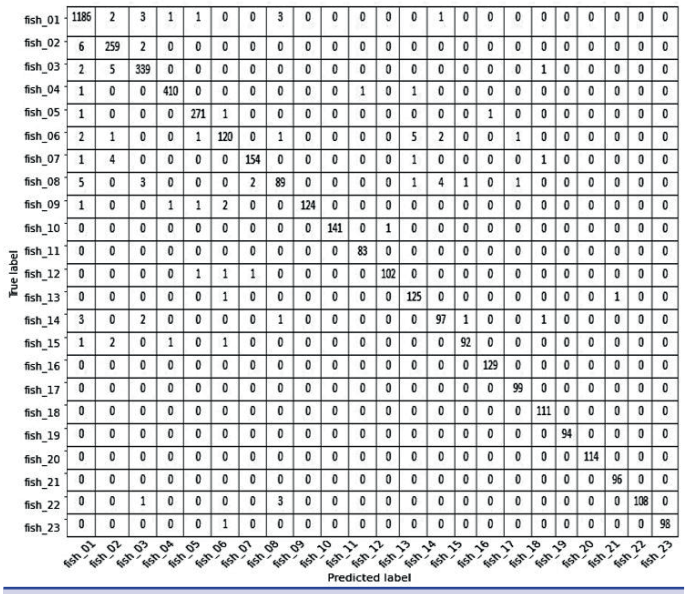


Figure 7. Confusion Matrix for the deepCNN-SENet with SVM

Table 1. Parameter sizes at different r for SE Block and Refined SE Block

r	'p' size with SE	'p' size with Ref-SE
8	22.778M	22.778M
16	22.776M	22.776M

where 'r' denotes the reduction ratio and 'p' denotes the parameter size. As we can see from Table.1, increasing the r value has decreased the parameter size slightly.

As we train the model through multiple epochs, we can see in Fig.9a the train set accuracy (blue) is steadily increasing, and the validation set accuracy also increases along with it. Since the validation set accuracy has not decreased compared to the training set, we find that variance is low, and the model has not overfit the training set. The final validation accuracy is 97.36%. Since the validation set accuracy is high, we find that bias is also low.

The training loss steadily reduces to below 1.0 in Fig.9b throughout the span of 8 training epochs and a lower loss value on the validation set indicates that the model successfully fit the training data. As a result, the model's overall performance is satisfactory.

	precision	recall	f1-score	support
fish_01	0.98	0.99	0.99	1197
fish_02	0.95	0.97	0.96	267
fish_03	0.97	0.98	0.97	347
fish_04	0.99	0.99	0.99	413
fish_05	0.99	1.00	0.99	272
fish_06	0.95	0.91	0.93	132
fish_07	0.98	0.96	0.97	161
fish_08	0.92	0.84	0.88	106
fish_09	1.00	0.96	0.98	129
fish_10	1.00	0.99	1.00	142
fish_11	0.99	1.00	0.99	83
fish_12	0.99	0.97	0.98	105
fish_13	0.94	0.97	0.95	129
fish_14	0.93	0.94	0.94	103
fish_15	0.98	0.96	0.97	96
fish_16	1.00	1.00	1.00	129
fish_17	0.99	1.00	0.99	99
fish_18	0.97	1.00	0.99	111
fish_19	1.00	1.00	1.00	94
fish_20	1.00	1.00	1.00	114
fish_21	1.00	1.00	1.00	96
fish_22	1.00	0.96	0.98	112
fish_23	1.00	0.99	0.99	99
accuracy			0.98	4536
macro avg	0.98	0.97	0.98	4536
weighted avg	0.98	0.98	0.98	4536

Figure 8. Performance evaluation of proposed frameworks

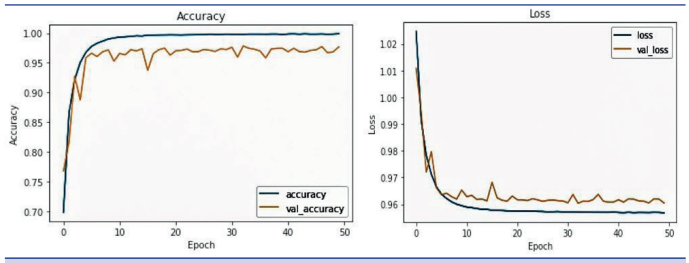


Figure 9. Accuracy and Loss plot of CNN-Refined SENet-SVM on the Fish4knowlege dataset

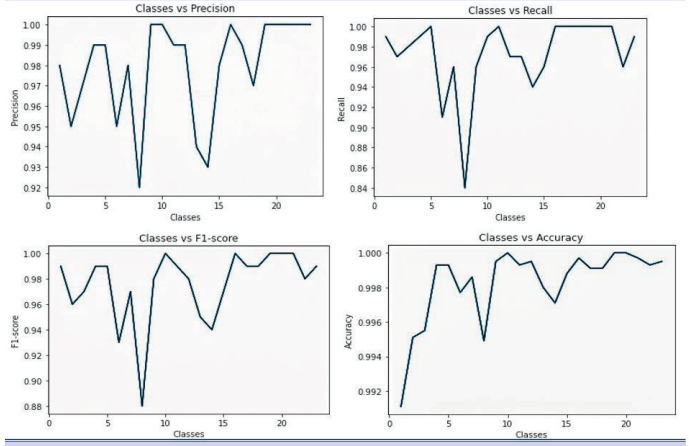


Figure 10. Metrics vs Classes

Fig.10 demonstrates the results of our model under different evaluation metrics. In Fig.10d we can see that the performance is particularly outstanding for species 22 and 23 with an accuracy of 99.95%. Fig.7 shows, for species 22, that 111 images out of 112 are correctly identified, and for species 23, all the images from the test data are correctly identified.

In Table.2, we can see that the CNN with SE model using the Fish4Knowledge dataset with 15 epochs achieved an accuracy of 97.15%. The best results are highlighted. CNN's TPE (Time per epoch) for 5 epochs is 144 sec and CNN+SENet took 157 sec for the same. This slight increase in the computational cost can be justified by its model's performance improvement (accuracy of CNN is 89% and CNN plus SE is 96.25% for 5 epochs).

In Table.3, we can see that CNN with SE model using the Fish4Knowledge dataset after pre-processing the images with 50 epochs achieved an accuracy of 97.77%. The hybrid CNN+SE model with SVM Classifier for the same dataset with squared-hinge loss has achieved an accuracy of 97.83% for 50 epochs. Finally, an accuracy of 97.90% for 50 epochs for the Refined SE model is recorded. This proves that refined SE improves the performance of the hybrid model.

The Cross-entropy loss is used for CNN-SENet and squared-hinge loss for CNN-SE with the SVM classifier. Table.4 proves that CNN-SE with SVM classifier has better performance.

Table 5 shows the comparison of the proposed model with the existing works. The proposed model has an improved accuracy of 97.9%, but Ensemble of Google InceptionNet and SVM achieved 95.37%, DNN achieved 96% and MLR_VGG19 achieved 97.09%. The proposed model has improved by 0.89% when compared with MLR_VGG19.

Experimental Results for the proposed method using small

Table 2. Experimental Result for Fish4Knowledge

Methodology	Epochs	Accuracy	Precision	Recall	F1-score
CNN	5 epochs	89%	0.85	0.60	0.68
	10 epochs	92.06%	0.80	0.60	0.66
	50 epochs	94.70%	0.94	0.95	0.64
CNN+SE	5 epochs	96.25%	0.93	0.82	0.85
	10 epochs	97.08%	0.94	0.83	0.87
	50 epochs	**97.15%**	**0.92**	**0.83**	**0.88**

Table 3. Experiment & Result for Fish4Knowledge after Preprocessing

Methodology	Epochs	Accuracy	Precision	Recall	F1-score
CNN+SE	15 epochs	97.46%	0.97	0.98	0.97
	50 epochs	97.77%	0.98	0.98	0.98
CNN+SE+SVM	15 epochs	97.28%	0.97	0.97	0.97
	50 epochs	97.83%	0.98	0.98	0.98
CNN+Refined SE+SVM	15 epochs	97.70%	0.98	0.97	0.97
	50 epochs	**97.90%**	**0.98**	**0.97**	**0.98**

Table 4. Metrics vs Loss

Metrics	Cross Entropy Loss	Squared Hinge Loss
Accuracy(%)	97.77	97.83
Recall(%)	97.78	97.78
Precision (%)	97.56	97.52
F1-score	97.69	97.73

Table 5. Comparison with existing state of art for Fish4Knowledge Dataset

Reference	Model	Accuracy
(Murugaiyan et al.,2021)	Ensemble of Google InceptionNet and SVM	95.37%
(Zhang et al.,2021)	DNN	96%
(Prasetyo et al.,2021)	MLR _VGG19	97.09%
Proposed model	**CNN-Refined SE-SVM**	**97.90%**

scale Croatian dataset

In a paper (Qiu et al.,2018), they used a small-scale Croatian dataset for post-training and achieved an accuracy of 83.56% for B-CNN with SE. They primarily pre-trained the model on the ImageNet dataset, then on the Fish4Knowledge dataset (Phoenix et al.,2013), and finally on a Croatian dataset to fine-tune it (small-scale fine-grained dataset). The Croatian Fish Dataset has a total of 794 images with 12 classes and after augmenting 500/1000 images per species, the total number of images increased to 10,794 images. The results are tabulated below by applying the proposed model on this small-scale dataset with and without augmentation.

the model, so it is a good fit for the training data. After 20 epochs, it maintains a uniform value in the validation accuracy.

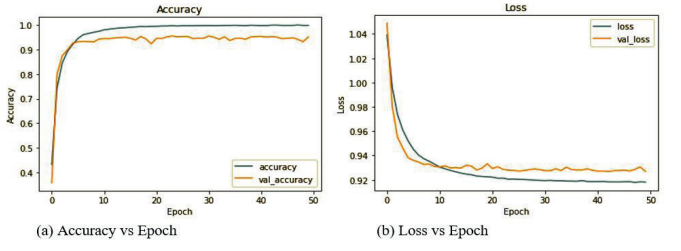
Table.7 shows the accuracy achieved by Qui et al. (2018), where they pre-processed the Croatian dataset with SRGAN, and augmentation was also performed. They also employed pre-training with ImageNet and the Fish4Knowledge dataset in their B-CNN with the refined SENet model, which consists of 10 convolution blocks, while the proposed model consists of 3 convolution blocks. The maximum accuracy produced by Qui et al. is 83.92%. However, the accuracy of the proposed model is 94.99%. It is well proven that there is a performance improvement of 11% when compared to BCNN-SE.

Table 7. Comparison with existing state of the art for Croatian Dataset using transfer learning

Method	Model	Accuracy
(Qui et al.,2018)	BCNN	83.52%
	BCNNs-SE blocks	83.78%
	Improved BCNNs-SE blocks	83.92%
Proposed model	**CNN-refined SE-SVM**	**94.99%**

Table 6. Experimental Result for Croatian Dataset after Preprocessing (Wo_A(Without Augmentation),W_A(With Augmentation)

Methodology	Epoch	Accuracy
CNN (Wo_A)	50 epochs	31.5%
CNN+SE(Wo_A)	50 epochs	56.6%
CNN+SE+SVM (W_A)	30 epochs	79.2%
CNN+SE+SVM (W_A)	**50 epochs**	**94.99%**



(a) Accuracy vs Epoch (b) Loss vs Epoch

Figure 11. Accuracy and Loss plot of CNN-SENet-SVM on Croatian dataset

From Table 6, it is seen that the CNN with SENet model using the Croatian dataset with 30 epochs achieved an accuracy of 71.37%, and then the hybrid CNN integrated with the SVM classifier achieved an accuracy of 79.2%. The lower accuracy could be due to the dataset's limited size and the lack of pre-training. Therefore, the weights from the pretraining step (CNN-SE-SVM) with the Fish4Knowledge dataset are loaded before initiating post-training with this small-scale dataset (Croatian), and the accuracy of the model improves from 79.2% to 94.99% in comparison with the previous result. Hence, transfer learning has a greater impact on performance. From Fig.11 we can see that the accuracy and loss plots are stable, therefore the transfer learning process has not overfit

CONCLUSION

In this study, we have proposed a hybrid framework comprising of CNN, Refined SE, and SVM for identifying 23 different fish species and increased model performance on the Fish4knowledge dataset and achieved an accuracy of 97.90%. It can work on small-scale and large-scale datasets by enhancing transfer learning and squeeze-and-excitation networks for fish image classification on small-scale datasets due to the non-informative channel suppression property of SE blocks, and enhanced classification using SVM. By post-training this model on the Croatian small-scale dataset, we achieved 94.99% accuracy. Thus, the proposed method, CNN-Refined SE with SVM, shows an 11% improvement over the existing method BCNN-SE. This model has achieved better generalisation and distinguishes the fish species well. In future work, this model will be modified to identify the absence of fish and some super-resolution techniques can be used to handle ocean images with varying lighting conditions.

Conflict of interest: There is no conflict of interest.

Ethics committee approval: This study does not need ethical approval.

Financial disclosure: NIL

Acknowledgment: We would like to acknowledge the support provided by Ms. R. P. Lilian Shirley and Ms. R. Nandhini, Department of CSE, Sri Sivasubramaniya Nadar College of Engineering.

REFERENCES

Allken, V., Handegard, N. O., Rosen, S., Schreyeck, T., Mahiout, T., & Malde, K. (2019). Fish species identification using a convolutional neural network trained on synthetic data. ICES Journal of Marine Science, 76(1), 342-349.

B. B. B. Phoenix X. Huang and R. B. Fishera.Fish4KnowledgeDataset: https://homepages.inf.ed.ac.uk/rbf/Fish4Knowledge/GROUNDTRUTH/RECOG/. 2013.

Chen, P. Y., Ho, C. W., Chen, A. C., Huang, C. Y., Liu, T. Y., & Liang, K. H. (2020). Inves-tigating seafood substitution problems and consequences in Taiwan using molecular barcoding and deep microbiome profiling. Scientific reports, 10(1), 1-9.

Crozier, L. G., McClure, M. M., Beechie, T., Bograd, S. J., Boughton, D. A., Carr, M.,... & Willis-Norton, E. (2019). Climate vulnerability assessment for Pacific salmon and steelhead in the California Current Large Marine Ecosystem. PloS one, 14(7), e0217711.

Dagoudo, M., Qiang, J., & Solevo, M. P. (2022). Status in science and technology devel- opments in Benin's aquaculture industry: a review. Aquaculture International, 1-15.

Data Science Glossary. https://c3.ai/glossary/data-science/.

Deep, B. V., & Dash, R. (2019, March). Underwater fish species recognition using deep learning techniques. In 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 665-669). IEEE.

Dos Santos, A. A., & Goncalves, W. N. (2019). Improving Pantanal fish species recogni- tion through taxonomic ranks in convolutional neural networks. Ecological Informatics, 53, 100977.

Du, J., Zhou, H., Qian, K., Tan, W., Zhang, Z., Gu, L., & Yu, Y. (2020). RGB-IR cross input and sub-pixel upsampling network for infrared image super-resolution. Sensors, 20(1), 281.

Fouad, M. M. M., Zawbaa, H. M., El-Bendary, N., & Hassanien, A. E. (2013, December). Automatic nile tilapia fish classification approach using machine learning techniques. In 13th international conference on hybrid intelligent systems (HIS 2013) (pp. 173-178). IEEE.

Hu, J., Shen, L., & Sun, G. (2018).'Squeeze-and-excitation networks.' In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

ImageNet Dataset. https://image-net.org/download.php. 2013.

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.

Jin, X., Xie, Y., Wei, X. S., Zhao, B. R., Chen, Z. M., & Tan, X. (2022). Delving deep into spatial pooling for squeeze-and-excitation networks. Pattern Recognition, 121, 108159.

Kroetz, K., Luque, G. M., Gephart, J. A., Jardine, S. L., Lee, P., Chicojay Moore, K., ... & Donlan, C. J. (2020). Consequences of seafood mislabeling for marine populations and fisheries management. Proceedings of the National Academy of Sciences, 117(48), 30318-30323.

Labao, A. B., & Naval Jr, P. C. (2019). Cascaded deep network systems with linked ensemble components for underwater fish detection in the wild. Ecological Informatics, 52, 103-121.

Ovalle, J. C., Vilas, C., & Antelo, L. T. (2022). On the use of deep learning for fish species recognition and quantification on board fishing vessels. Marine Policy, 139, 105015.

Meng, L., Hirayama, T., & Oyanagi, S. (2018). Underwater-drone with panoramic camera for automatic fish recognition based on deep learning. Ieee Access, 6, 17880-17886.

Murugaiyan, J. S., Palaniappan, M., Durairaj, T., & Muthukumar, V. (2021). Fish species recognition using transfer learning techniques. International Journal of Advances in Intelli- gent Informatics, 7(2), 188-197.

Naaum, A. M., Warner, K., Mariani, S., Hanner, R. H., & Carolin, C. D. (2016). Seafood mislabeling incidence and impacts. In Seafood Authenticity and Traceability (pp. 3-26). Academic Press.

Pollack, S. J., Kawalek, M. D., Williams-Hill, D. M., & Hellberg, R. S. (2018). Evaluation of DNA barcoding methodologies for the identification of fish species in cooked products. Food Control, 84, 297-304.

Prasetyo, E., Suciati, N., & Fatichah, C. (2021). Multi-level residual network VGGNet for fish species classification. Journal of King Saud University-Computer and Information Sciences.

Qiu, C., Zhang, S., Wang, C., Yu, Z., Zheng, H., & Zheng, B. (2018). Improving transfer learning and squeeze-and-excitation networks for small-scale fine-grained fish image classification. IEEE Access, 6, 78503-78512.

Rauf, H. T., Lali, M. I. U., Zahoor, S., Shah, S. Z. H., Rehman, A. U., & Bukhari, S. A. C. (2019).Visual features based automated identification of fish species using deep convolutional neural networks. Computers and electronics in agriculture, 167, 105075.

Vilas, C., Antelo, L. T., Martin-Rodriguez, F., Morales, X., Perez-Martin, R. I., Alonso, A. A., ... & Barral-Martinez, M. (2020). Use of computer vision onboard fishing vessels to quantify catches: The iObserver. Marine Policy, 116, 103714.

Villon, S., Iovan, C., Mangeas, M., Claverie, T., Mouillot, D., Villéger, S., & Vigliola, L. (2021).'Automatic underwater fish species classification with limited data using few-shot learning. Ecological Informatics, 63, 101320.

Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T., & Villéger, S. (2018). A deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecological informatics, 48, 238-244.

Xu, X., Li, W., & Duan, Q. (2021). Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification. Computers and Electronics in Agriculture, 180, 105878.

Zhang, Y., Zhang, F., Cheng, J., & Zhao, H. (2021). Classification and Recognition of Fish Farming by Extraction New Features to Control the Economic Aquatic Product.' Complexity, 2021.