**Classification of Fish Eye Freshness Using Deep Learning**

JAN EDGAR E. TUPAS

Student, Mapúa University, jeetupas@mymail.mapua.edu.ph

CHRISTIAN HENRY MIGUEL E. CARUZ

Student, Mapúa University, chmecaruz@mymail.mapua.edu.ph

RAFAELLO JOSE M. VIERA

Student, Mapúa University, rjmviera@mymail.mapua.edu.ph

Further advancements in machine learning and deep learning have enabled the development of effective computer vision models that automate image classification and segmentation tasks, introducing solutions that aim to improve the quality of life. In the domain of aquaculture, computer vision models have been mainly implemented to assess the quality of fish effectively. The thesis presents a deep learning-based method to automate the determination and classification of the freshness of the fish eye. A unified pipeline between a SegFormer model and an EfficientNetV2 model was produced to segment and accurately classify the fish eye region effectively. The EfficientNetV2 model was further improved by integrating Efficient Channel Attention Blocks and Coordinate Attention Blocks, enabling the model to focus on the vital features of the fish eye that provide relevant information in determining freshness. The model was trained and validated on a Sea Bream (Sparus Aurata) fish images dataset with four freshness classes. Image preprocessing techniques were implemented to enhance the quality of the data further. The developed EfficientNetV2 model obtained a final accuracy of 87.27%. Incorporating the ECA and CA blocks into the model enhanced the classification performance by enabling the model to focus on essential features more effectively while maintaining a more efficient architecture, reducing the total number of parameters from the base model by 16.46%. The promising results point toward the potential of deploying the model in real-world scenarios. To further improve model performance, it was recommended that the size and diversity of the dataset be increased, more advanced image preprocessing techniques be explored to enhance data quality, and more advanced attention mechanisms be integrated into the model.

**CCS CONCEPTS •** Applied computing **•** Human-Centered Computing **•** Software and its engineering

**Additional Keywords and Phrases:** Fish Freshness Classification, Image Classification, Image Segmentation, Image Preprocessing, EfficientNetV2

1. **introduction**

Within the fields of artificial intelligence (AI) and computer science, computer vision uses processing power to extract information from images and videos. According to Feng et al., computer vision tasks aim to replicate human vision by automatically enabling a computer system to see, recognize, and grasp the visual world [18]. One of the core tasks in computer vision is image classification. It entails obtaining a picture as input, processing it, and analyzing it to extract features, then using the features to categorize the image. The developments in deep learning have allowed its seamless integration in computer vision.

Various research has been conducted in the fisheries industry to implement computer vision for nondestructive, fast, and accurate fish quality assessment. According to Jayasundara et al., imaging systems and deep learning techniques have greatly enhanced the quality assessment procedures of aqua agriculture [2]. Medeiros et al. stated that fast, objective, and robust measurement and determination of fish freshness can be achieved through computer vision and machine learning, and the manual approach of sensory evaluation methods can be automated to simplify and accelerate the process of determining fish freshness using computer vision methods [35]. Computer vision-based fish freshness classification can be a reliable alternative for assessing and evaluating fish freshness, which is nondestructive and requires no complicated instruments.

In this paper, the researchers developed an image classification detection model to classify fish freshness based on the eye. The eye region of the fish was segmented and classified according to its freshness: Very Fresh, Fresh, Not Fresh, and Spoiled. The researchers utilized the SegFormer to segment the fish eye region and the EfficientNetV2 to classify the freshness of the fish eye. Attention mechanisms were further incorporated into the model to enhance performance and improve efficiency.

The general objective of the study was to develop a deep learning image classification model to classify the freshness of fish based on the eye, which manifests accurate and effective performance while retaining efficiency. The specific objectives of the study are the following: (1) To develop and implement a SegFormer semantic segmentation model for isolating the fish eye region from the rest of the fish, minimizing the influence of extraneous background information during the classification process; (2) To develop and implement an EfficientNetV2 classification model aimed at enhancing the performance and accuracy of fish freshness classification, while optimizing computational efficiency for practical deployment by incorporating advanced attention mechanisms, namely the Efficient Channel Attention and Coordinate Attention, into the EfficientNetV2 model, enabling it to focus on the most salient and relevant features of the segmented fish eye for improved classification performance.

1. **review of related literature**
   1. **Fish Freshness Classification using Deep Learning**

Convolutional Neural Networks (CNNs) dominate the computer vision landscape. As such, the reviewed literature on the classification of fish freshness using deep learning is dominated by using CNN-based architectures for automatically estimating the freshness levels of fish images. Because of its network architecture, CNNs can handle high-dimensional pictures more effectively by reducing model complexity and weights [1]. CNN architectures are effective for image classification tasks as they balance effective classification performance and efficient use of computing resources. Because CNNs can automatically extract characteristics from images, perform well, and need less complexity when used in fish freshness classification [2]. Some CNN architectures that were utilized for fish freshness classification include VGG-16 [2] [4] [5] [2] [6], MobileNetV1 [7] [8] [6] [9], and YOLO [10] [11]. However, several works related to fish freshness classification using CNN architectures result in limited classification performance, such as in the case of [2] [8] [10] [6] [9] [12], indicating an opportunity for further improvement. The studies of Yildiz et al. [12] and Anas et al. [10] suggested exploring and using other deep learning models to improve the performance of models for fish freshness classification. The work of Rodrigues et al. [13] provided insights into the use of the Vision Transformer for classifying fish freshness and presented improved performance. However, it comes at the cost of significant and expensive computing resources and data. Vision Transformers utilize more parameters and require more training time and data than efficient CNN architectures, introducing challenges with limited datasets and preventing the deployment of the model in resource-limited devices such as mobile devices. There is an opportunity to identify and develop efficient and effective CNN architectures that have yet to be used in recent works relating to fish freshness classification. This would improve classification performance while minimizing computing resources and data use.

* 1. **EfficientNetV2**

Following the successful EfficientNet classification model, Tan and Le improved upon the original model and developed the EfficientNetV2, incorporating faster training speed and better parameter efficiency [15]. One drawback of the original model was slow training when image sizes were large. This could be addressed by decreasing the batch size, but it is not optimal. Another drawback was the use of depthwise convolutions, which are slow in the early layers of the architecture. Lastly, the compound scaling method scales depth, width, and resolution equally at all stages, which is sub-optimal and does not equally contribute to training speed and parameter efficiency. According to Tan and Lee, EfficientNetV2 incorporates the training-aware neural architectural search (NAS) to achieve better accuracy, less training step time, and a network with fewer parameters [15]. The search process focuses on finding the best combination of layers and scaling parameters to improve training speed and accuracy. The improved architecture uses the fused MBConv and MBConv in the early layers, smaller expansion ratios for less memory access overhead, smaller 3x3 kernel sizes while increasing the layers to compensate for the reduced receptive, and removes the last stride-1 stage due to the large parameter size and memory access overhead. Fused MBConv layers are improved versions of the Mobile Inverted Bottleneck Convolution (MBConv) layers, which efficiently combine convolutions and linear bottleneck layers. EfficientNetV2 achieves high accuracy with fewer parameters. Because of its efficiency, it can be implemented on devices with limited resources. It could be useful for real-world applications, such as systems for classifying fish freshness in different environments. EfficientNetV2 can collect significant features from photos with competitive performance with its balanced scaling method, essential for reliable fish freshness classification based on eye images. Furthermore, EfficientNetV2 has been trained on large-scale image datasets such as ImageNet, facilitating transfer learning. The model can be fine-tuned to the specific dataset of fish eye images to enhance performance.

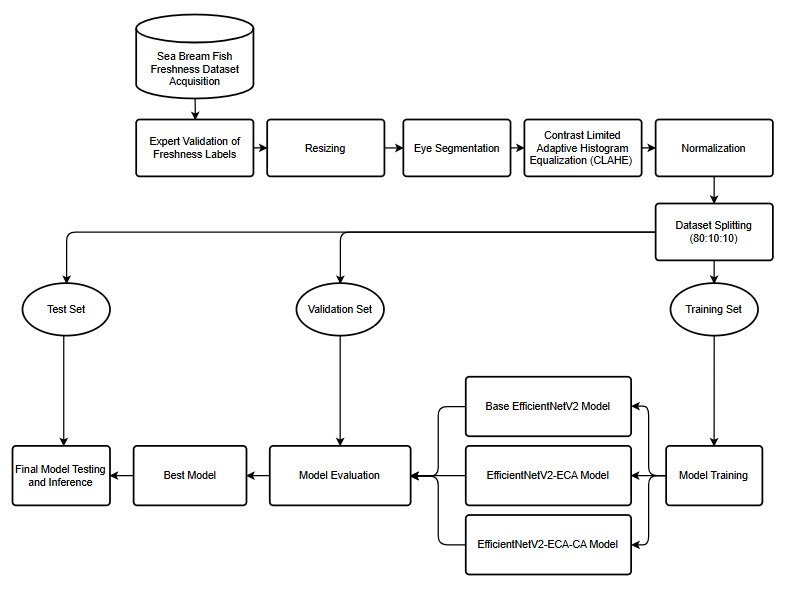
* 1. **Efficient Channel Attention**

The Efficient Channel Attention (ECA) block was introduced in the work of Wang et al. to improve the performance of deep CNNs by effectively learning channel attention while keeping model complexity low [16]. Unlike older approaches, such as the Squeeze-and-Excitation (SE) block, which frequently reduces the dimensionality of channel representations, the ECA block preserves the original channel dimensions. The direct relationship between channels and their weights enables more effective learning of channel attention. The ECA block offers a local cross-channel interaction method based on adaptive 1D convolutions to capture channel dependencies without global pooling operations. The approach allows the ECA block to be effortlessly integrated into various CNN designs, offering a considerable speed improvement while not introducing significant complexity. The ECA block is lightweight, requiring fewer parameters than other attention mechanisms [7]. The combination of efficiency and efficacy makes the ECA block a desirable addition to the toolkit for improving model performance in computer vision applications.

* 1. **Coordinate Attention**

To further improve the efficiency of attention mechanisms while improving or maintaining performance, the work of Hou et al. introduced the coordinate attention block [17]. It aimed to enhance feature representation in mobile networks by combining channel interactions with positioning information. Unlike standard channel attention algorithms, which typically condense spatial information into a single channel descriptor via 2D global pooling, the coordinate attention block takes a more complex approach. It starts by applying two distinct one-dimensional global pooling operations on the input feature tensor, aggregating features along the vertical and horizontal axes. This produces two direction-aware feature maps that preserve the spatial structure of the input data, which is critical for accurately capturing the context of objects in the picture. These direction-aware feature maps are converted into attention maps, encoding long-term dependencies along their respective spatial directions. The produced attention mappings are then applied to the original feature tensor via element-wise multiplication, emphasizing the characteristics of interest while accounting for the relevance of other channels. The dual attention technique enables the model to focus on relevant areas, which boosts object recognition and placement accuracy. The coordinate attention block is lightweight and easily integrates into existing mobile network infrastructures, making it ideal for applications that require efficient processing [7].

1. **methodology**
   1. **Conceptual Framework**



**Figure 1.** Conceptual Framework of the Study

The conceptual framework of the study, as shown in Figure 1, outlines a thorough, structured, and theoretical approach to the development of an effective and efficient model for fish freshness classification based on the fish eye using fish images. The process of the study was conducted through the following steps: Dataset acquisition; Expert validation of the dataset; Image preprocessing through resizing, eye segmentation, contrast limited adaptive histogram equalization (CLAHE), normalization, and data splitting; Model training and evaluation; Final model testing and inference.

* 1. **Data Acquisition and Validation**

A dataset of Sea Bream (Sparus Aurata) fish from the work of Rodrigues et al. [14] was used in the study. It originally consisted of 4293 images of Seabream fish tagged with the time elapsed since their capture. A total sample of 40 Sea Bream fish was acquired. The images were documented for 8 to 326 hours or 14 days since fish capture. The images were captured in portrait and landscape layouts with a dimension of 3000x4000 pixels and vice versa. The images were captured in JPEG file format and RGB color format. The images were captured at various angles, focusing on the fish eye region and the whole fish. To ensure the integrity and quality of the dataset, it was subjected to careful evaluation and verification from two fish physiology and processing experts. The experts were assistant professors from the Institute of Fish Processing Technology at the University of the Philippines Visayas, both with degrees in Fisheries. The extensive experience and knowledge of the experts in the field allowed for the validation of the dataset, evaluating whether the fish eye from the images manifested features of their corresponding freshness level. The freshness level of the fish represents the classes for the dataset, which encompass four levels: Very Fresh, Fresh, Not Fresh, and Spoiled. Following the experts' recommendations, fish images whose eye features strongly and accurately manifested their corresponding freshness label were retained in the dataset, while those that did not and were inaccurate were removed. This resulted in a dataset of 2194 images for developing the classification model. **Table 1** shows the class distribution of the dataset. The dataset is considered balanced, with differences of only a few samples between classes. It was essential to ensure a balanced dataset to prevent bias toward a specific class during training. The images of fish labeled Very Fresh were taken within 48 hours or two days after the fish were captured. The images of fish labeled as Fresh were taken after 48 hours and before 120 hours since the fish were captured. The images of fish labeled as Not Fresh were taken after 120 hours and before 240 hours since the fish were captured. The images of fish labeled as Spoiled were taken after 240 hours since the fish were captured.

**Table 1. Class Distribution of the Dataset**

|  |  |  |
| --- | --- | --- |
| Class | Number of Images | Time Since Capture |
| Very Fresh | 543 | Within 48 Hours |
| Fresh | 542 | After 48 and Before 120 Hours |
| Not Fresh | 576 | After 120 and Before 240 Hours |
| Spoiled | 533 | After 240 Hours |
| Total | 2194 |  |

* 1. **Data Preprocessing**

Expert validation of the dataset and allowed the removal of inconsistencies. This process ensured that the freshness labels associated with each image were accurate and true, establishing the integrity of the dataset for use. After validation, the dataset was subjected to preprocessing tasks, encompassing resizing, segmentation, cropping, CLAHE, and normalization to enhance data quality and relevance, which was essential for analysis and model training.

* + 1. **Resizing**

The original images were captured in dimensions of 3000x4000 pixels (portrait) and 4000x3000 pixels (landscape). It was vital to resize the image to be consistent with the inputs used in pre-training the SegFormer model to ensure that the data meets the input requirements of the model, which allows it to function correctly. In this case, the images were resized into 512x512 pixels in preparation for fish eye segmentation, reducing the images to a more manageable size and maintaining the aspect ratio to prevent features from being distorted.

* + 1. **Segmentation Using SegFormer**

The specific regions were the fish eye, fish body, and background. The fish masks were used to isolate the fish eye region from the body region and image background. The images were trained, evaluated, and tested in the SegFormer model to segment the specified regions. After training, validation, and testing, the SegFormer model was deployed to segment the fish eye region from the fish images effectively. The SegFormer deep learning architecture utilizes a transformer-based backbone to capture local and global image features. The input image, which is the fish image, is processed through a series of encoder layers that extract hierarchical features at multiple resolutions. The features are passed to a decoder for refinement, where a lightweight attention mechanism is used to determine the boundaries of the eye region accurately. The ability of SegFormer to integrate contextual information allows it to identify the eye region from the surrounding areas, translating to precise segmentation. The final output of the model is the segmented fish eye. It is further resized to 300x300 pixels.

* + 1. **Contrast Limited Adaptive Histogram Equalization**

The clarity, color, and texture features in the fish eye contain essential information that is key to determining fish freshness. Enhancing the contrast of these features allows them to be more visible to the image classification model. To improve the quality of the fish eye images, particularly the differentiation of the color and texture features of the fish eye between various freshness stages, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied. The images were converted to LAB color space to apply CLAHE to the RGB segmented fish eye images. CLAHE was applied to the L (lightness) channel using the OpenCV library. This enhanced the contrast of the brightness of the image while preserving the color balance between the A and B channels. A clip limit of 2.0 and tile grid size of 8.0 was used for a balanced localization and enhancement of contrast. After applying CLAHE, the image is converted back to the RGB color space.

* + 1. **Normalization**

The images were further normalized to ensure that each input channel had a similar distribution, helping the model avoid bias towards any color channel and enhancing training efficiency and model convergence. The pixel values of each image in the dataset were normalized for each color channel, which, in this case, was Red, Green, and Blue. Equation 3.1 was used to normalize the images. A mean of 0.485 and a standard deviation of 0.229 were used for the Red channel. A mean of 0.456 and a standard deviation of 0.224 were used for the Green channel. A mean of 0.406 and a standard deviation of 0.225 were used for the Blue channel. These normalization processes and values were used considering the models developed in the study were pre-trained on ImageNet. The input photos are guaranteed to have a pixel distribution consistent with the training conditions used by the models pre-trained on ImageNet.

* + 1. **Dataset Splitting**

The dataset was split into a training, validation, and testing set following the 80:10:10 ratio, respectively, as shown in Table 2. This split ensures that most of the data was used to train the model to learn patterns, features, and relationships within the dataset. It also ensured the improvement in generalization, especially with the limited dataset, while maintaining sufficient data for tuning, validating, and testing the model. The training set consisted of 1775 images. The validation set encompassed 219 images. The test set consisted of 220 images. The dataset of segmented fish eye images was used to train and evaluate the performance of the developed EfficientNetV2 models.

**Table 2. Dataset Distribution**

|  |  |
| --- | --- |
| Set | Number of Images |
| Training | 1775 |
| Validation | 219 |
| Test | 220 |
| Total | 2194 |

* 1. **EfficientNetV2 Training**

The Google Colab platform was used to train the models utilizing a T4 GPU. The EfficientNetV2S (Small) variant was used because it performs effectively with lower parameters than the other variants. It was pre-trained on the ImageNet-21k dataset. The hyperparameters of the models were selected and tuned through the baby-sitting method. Initial training runs were conducted to determine the optimal hyperparameters for training the models. Close observations of the Training and Validation Losses and other performance metrics accompanied this. Considering the size of the dataset and the intricacy of the features from the fish eye that provide essential information regarding freshness, a batch size of 8 was used. Although training time was increased, a lower batch size allowed the model to update its weights more frequently. This further allows the model to generalize and makes it less prone to overfitting. The Cross-Entropy Loss Function was used as it is designed for multi-class classification tasks. The Stochastic Gradient Descent (SGD) optimizer with momentum was used to train the models. Unlike the Adam optimizers, SGD avoids overly aggressive updates. This results in a smoother and more stable training process. Although it leads to a longer training time, it results in a more stable convergence and better generalization, which is vital considering the size of the dataset. The SGD optimizer had a learning rate of 0.00001, momentum of 0.9, and weight decay of 0.0001. A low learning rate allows the models to make gradual and precise updates throughout the training. This reduces the risk of overshooting the optimal solution. The momentum and weight decay also allowed for stable updates and convergence while ensuring regularization, preventing overfitting, and allowing the models to train effectively. To allow sufficient time for the models to learn during training and considering the combination of the SGD optimizer with lower learning, the models were trained for 300 epochs. The Training and Validation Losses and other performance metrics were closely monitored during the training of the models to detect signs of overfitting and instabilities and assess model performance over the course of the training. The best weights, which resulted in the lowest Validation Loss, were recorded.

* 1. **Model Evaluation**

During training, the performance of the models was closely monitored and evaluated using the Training Loss and Validation Loss. The Training Loss measures how well the model fits the training data by assessing the error of the model on the training set. It is calculated by taking the sum of errors for each sample in the training set. Meanwhile, the Validation Loss measures how well the trained model performs on the validation set unseen during training. The Validation Loss evaluates how well the model generalizes to new data. It is calculated by taking the sum of errors for each sample in the validation set. The performance of the SegFormer model for the segmentation task was evaluated using Accuracy (AC) and Intersection over Union (IoU). The performance of the EfficientNetV2 models for the classification task was assessed using AC, Precision (PR), Recall (RE), F1-Score, and the Confusion Matrix.

1. **results and discussion**
   1. **Segmentation of the Fish Eye using the SegFormer**

## **Table 3. SegFormer Performance Metrics Results**

|  |  |  |
| --- | --- | --- |
|  | IoU | Accuracy |
| Training | 94.18% | 99.24% |
| Validation | 94.01% | 99.28% |
| Test | 94.14% | 99.32% |

The results of the SegFormer's overall performance metrics are presented in Table 3. The model obtained a test IoU of 94.14%, closely matching the training IoU and validation IoU. This indicates that the model effectively generalized to unseen data without overfitting. The robustness of the model in accurately segmenting the fish eye and body regions is supported by the consistency across the training, validation, and test IoU. The model maintained an IoU above 94% even in the test set, validating its effectiveness in learning information for fish segmentation. Additionally, the model achieved a 99.32% accuracy on the test set, a slightly higher score than the training and validation accuracies, which can be attributed to diverse samples in the dataset. The test accuracy confirms that the model can effectively classify pixels into the correct segments in the image with minimal misclassifications. Inference was performed to observe the outputs of the SegFormer model. Images from the test set were passed into the body and eye segmentation model. Figure 1 shows the fish eye and body mask produced by the model after segmenting the regions where the red highlight indicated the fish body and the blue highlight indicated the fish eye. The segmented eye image is the final output of the model.

|  |  |
| --- | --- |
| A fish on a black sign  Description automatically generated |  |

**Figure 1. Segmentation of Sea Bream Fish**

* 1. **Freshness Classification of the Fish Eye using EfficientNetV2 ECA-CA Model**

Evaluation of the performance of the EfficientNetV2-ECA-CA model was conducted on the test set, and the results are summarized in Table 4. An accuracy of 87.27% was achieved by the model, reflecting the percentage of instances from the test set that were correctly classified. The obtained high accuracy suggests the ability of the model to learn and generalize the patterns from the features extracted from the data effectively. The model utilized the data to accurately predict the freshness of the fish eye based on the freshness classes. Incorporating channel-wise and spatial attention mechanisms in the model, manifested in introducing the ECA and CA blocks, likely contributed to the effective and robust performance. The ECA and CA block aided the model in capturing the essential features from the fish eye images while also focusing on their spatial distribution within the images.

**Table 4. EfficientNetV2-ECA-CA Test Results**

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1-Score |
| 87.27% | 87.93% | 87.40% | 87.41% |

The confusion matrix of the EfficientNetV2-ECA-CA on the test set is presented in Figure 2. 46 Very Fresh fish eye samples were correctly classified. Furthermore, and more importantly, no instances of the Very Fresh fish eye samples were misclassified into Not Fresh or Spoiled, indicating the effective capability of the model to distinguish between the freshness and degradation of the fish eye. However, there were still misclassifications where nine Very Fresh fish eye samples were predicted as Fresh. This can be attributed to the close similarities in the visual features and cues of the Very Fresh and Fresh fish eyes, especially in the moments leading to the transition from Very Fresh to Fresh. This likely explains the occasional confusion the model commits between the two categories. Incorporating the ECA and CA blocks enabled the model to capture the subtle features and spatial cues. However, there is still a challenge in differentiating between the slight freshness variations, indicating room for improvement. The model demonstrated a strong performance for the Fresh fish eye samples, where the model correctly classified 49 Fresh eye instances. However, misclassifications were still committed, where one Fresh eye sample was misclassified as Very Fresh. Additionally, four Fresh fish eye instances were misclassified as Not Fresh. This indicates that in some cases, the model may have over-penalized slight degradations in freshness. It may have interpreted them as more significant than they were. Additionally, the misclassifications can be attributed to the minor textural or color changes in the fish eye between the adjacent classes, especially in the moments leading to the transition of the fish eye from Very Fresh to Fresh and from Fresh to Not Fresh. It is important to note that the model did not misclassify any Fresh fish eye sample as spoiled, suggesting that it successfully differentiated between freshness and extreme degradation. The model mostly classified Not Fresh fish eye samples reflected in the 48 correctly classified instances. However, there were still misclassifications, with four instances misclassified as Fresh and six misclassified as Spoiled. The misclassifications as Fresh and Spoiled indicate that the model still exhibits a slight struggle in the ability of the model to completely differentiate between subtle variations in the fish eye features, especially with the adjacent classes, likely due to the similarities between Fresh and Not Fresh fish eye during the freshness transition. Additionally, Not Fresh fish eye samples may have shown early signs of spoilage. However, the extent of freshness degradations may have been overestimated by the model, resulting in the misclassifications as Spoiled. Similar visual cues, such as discoloration and cloudiness, may have confused, leading to inaccurate interpretations. The model performed exceptionally well for the Spoiled class, with 49 out of 53 instances correctly classified. This indicates that the model effectively determined the visual features that describe spoilage in the fish eye. Only four instances were misclassified as Not Fresh. This suggests that there were still a few cases where the model struggled to interpret the early stages of spoilage, making it difficult to distinguish from the Not Fresh class. This may be further attributed to the overlap in the visual cues and features of loss of freshness and ultimate spoilage. Indicators such as color degradation, cloudiness, and sunken fish eye samples can manifest in both Not Fresh and Spoiled samples, confusing the model when differentiating between the two classes. The overall solid performance still suggests that the model, with the incorporated blocks, effectively focused on the features of the fish eye most indicative of spoilage.

A chart with numbers and a number of words

Description automatically generated with medium confidence

**Figure 2. Confusion Matrix of EfficientNetV2-ECA-CA on Test Set**

Inference was further performed on the test set to assess the capability of the model for the fish eye freshness classification task. Ensuring that the model performs well during inference confirms the reliability and efficacy of the model for determining the freshness category of fish eye samples and the generalizability of the model beyond the training data. Sample inference results are shown in Figure 3 for the Very Fresh, Fresh, Not Fresh, and Spoiled classes.

|  |  |  |  |
| --- | --- | --- | --- |
| A close up of a fish eye  Description automatically generated | A close up of a fish eye  Description automatically generated | A close up of a fish eye  Description automatically generated | A close up of a eye  Description automatically generated |

**Figure 3. Sample Inference Results for Each Freshness Class**

1. **conclusion and recommendations**
   1. **Conclusion**

The objectives of the study focused on developing a nondestructive method for classifying fish eye freshness. It developed a SegFormer model for effectively segmenting and isolating the fish eye region and an improved EfficientNetV2 model for accurately classifying the freshness of the fish eye. The model excelled, achieving an accuracy of 87.27%, with a precision of 87.93%, recall of 87.40%, and F1-score of 87.41%. These high scores verified the ability of the model to generalize effectively to previously unseen data, delivering accurate predictions of fish freshness. Moreover, the confusion matrix revealed that the model maintained high specificity and sensitivity, demonstrating its reliability in accurately identifying all categories of fish freshness. The model also exhibited efficient inference times, making it suitable for deployment in practical applications requiring real-time fish freshness assessment, such as in fish markets, processing facilities, and quality control settings. The inclusion of attention mechanisms, specifically the ECA blocks and CA blocks, was instrumental in enhancing the performance of the model. During feature extraction, the ECA blocks facilitated a more effective focus on critical features by adjusting channel-wise weights, allowing the model to prioritize essential details, such as subtle variations in eye color, clarity, and cloudiness. This led to a more streamlined model with a reduced parameter count, ensuring computational efficiency without sacrificing accuracy. The CA blocks further improved the ability of the model to capture spatial dependencies, enabling focused attention on specific regions within the fish eye that are crucial for determining freshness. These blocks amplified the capability of the model to discern localized features, such as discoloration, pupil dilation, and changes in eye shape, which are vital indicators of fish freshness. The primary objectives of the study were achieved by developing the SegFormer model for the segmentation task and the EfficientNetV2-ECA-CA model for the classification task. The high-performance metrics achieved by both models and their integration into a single pipeline allowed for a nondestructive and highly effective fish eye freshness classification method that provides robust potential for the deployment of the model in the fisheries industry for ensuring the consistent quality and freshness of fish products.

* 1. **Recommendations**

Several recommendations are proposed after the completion of the study to guide future research aiming to improve the accuracy and generalization of the EfficientNetV2 model for classifying fish eye freshness based on segmented fish eye images. The size and diversity of the dataset used for training is one of the significant factors affecting and limiting the performance of machine learning models. Although utilizing the current dataset produced promising results, it is recommended that future research efforts significantly increase the size of the dataset. This can provide the model with more samples to learn from, translating to improved generalization across all scenarios while reducing the likelihood of overfitting. The current dataset used in the study focuses on a specific fish species, Sea Bream (Sparus Aurata). However, many fish species are evaluated for freshness in practical applications such as fish processing and food quality control. Future research should expand the dataset to include images of different fish species. Each species may prevent distinct visual features and cues of freshness due to the physiological differences across the various fish species. Training the model on a more diverse dataset can further improve the generalizability of the model, enabling it to classify freshness across multiple fish species accurately. The EfficientNetV2-ECA-CA model demonstrated an effective performance, but there is still room for improvement to further strengthen its accuracy and efficiency. This can be achieved through further refinements. Future research should focus on experimenting with alternative model architectures and further integrate hybrid attention mechanisms. Additional attention mechanisms can further aid the model in capturing global and local dependencies. This translates to a more accurate determination of freshness features from the fish eye. Additionally, image preprocessing plays a vital role in enhancing the quality of the input images that the model uses. This allows for improved feature extraction that translates to higher classification accuracy. Future research can further experiment with various preprocessing techniques to enhance the dataset used for training the model. The study specifically utilized the CLAHE technique, but future research may further incorporate additional techniques to improve the visibility of vital features in the fish eye.

**REFERENCES**

1. Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review,* Volume 40, 100379. https://doi.org/10.1016/j.cosrev.2021.100379
2. Taheri‐Garavand, A., Nasiri, A., Banan, A., & Zhang, Y. (2020). Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish. *Journal of Food Engineering*, Volume 278, 109930. https://doi.org/10.1016/j.jfoodeng.2020.109930
3. Jayasundara, J., Ramanayake, R., Senarath, H., Herath, H., Godaliyadda, G. M. R. I., Ekanayake, M. P. B., Herath, V., & Ariyawansa, S. (2023). Deep learning for automated fish grading. *Journal of Agriculture and Food Research*, Volume 14, 100711. https://doi.org/10.1016/j.jafr.2023.100711
4. Mahendran, R., & Seneviratne, G. P. (2022). Using CNN to identify the condition of edible fish*. 2022 7th International Conference on Information Technology Research (ICITR)*, 1-4. doi: 10.1109/ICITR57877.2022.9993419.
5. Rayan, M. A., Rahim, A., Rahman, A., Marjan, M. A., & Ali, U. a. M. E. (2021). Fish freshness classification using combined Deep learning model. *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*. https://doi.org/10.1109/acmi53878.2021.9528138
6. Kato, D., Gu, C., Yu, J., & Zhang, C. (2023). Improving fish freshness classification with fish eye segmentation guidance. *IEEE 12th Global Conference on Consumer Electronics (GCCE)*. 10.1109/GCCE59613.2023.10315518
7. Guo, X., Gu, S., & Wei, D. (2022). Fish freshness evaluation based on MobileNetV1 and attention mechanism. *EITCE ’22: Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering*. https://doi.org/10.1145/3573428.3573475
8. Prasetyo, E., Purbaningtyas, R., & Adityo, R. D. (2020). Performance Evaluation of Pre-trained Convolutional Neural Network for Milkfish Freshness Classification. *6th Information Technology International Seminar (ITIS)*. <https://doi.org/10.1109/itis50118.2020.9321049>
9. Prasetyo, E., Purbaningtyas, R., Adityo, R. D., Suciati, N., & Fatichah, C. (2022). Combining MobileNetV1 and Depthwise Separable convolution bottleneck with Expansion for classifying the freshness of fish eyes. Information Processing in Agriculture, 9(4), 485–496. https://doi.org/10.1016/j.inpa.2022.01.002
10. Anas, D. F., Jaya, I. K. N. A., & Nurjanah, N. (2021). Design and implementation of fish freshness detection algorithm using deep learning. *IOP Conference Series: Earth and Environmental Science,* Volume 944(1), 012007. https://doi.org/10.1088/1755-1315/944/1/012007
11. Akgül, İ., Kaya, V., & Tanır, Ö. Z. (2023). A novel hybrid system for automatic detection of fish quality from eye and gill color characteristics using transfer learning technique. *PLOS ONE,* Volume 18(4). https://doi.org/10.1371/journal.pone.0284804
12. Yildiz, M. B., Yasin, E. T., & Koklu, M. (2024). Fisheye freshness detection using common deep learning algorithms and machine learning methods with a developed mobile application. European Food Research & Technology. https://doi.org/10.1007/s00217-024-04493-0
13. Rodrigues, J., Pacheco, O., & Correia, P. L. (2023). Seabream freshness classification using vision transformers. In *Lecture Notes in Computer Science*, 510–525. https://doi.org/10.1007/978-3-031-49018-7\_36
14. Rodrigues, J., Pacheco, O., & Correia, P. L. (2023). Seabream freshness classification using vision transformers. *Proceedings of the Iberoamerican Congress on Pattern Recognition*.
15. Tan, M., & Le, Q., V. (2021, April 1). EfficientNetV2: Smaller models and faster training. arXiv.org. https://arxiv.org/abs/2104.00298
16. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *Conference on Computer Vision and Pattern Recognition (CVPR),* 11531-11539. 10.1109/CVPR42600.2020.01155.
17. Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. *Conference on Computer Vision and Pattern Recognition (CVPR)*. 13708-13717. 10.1109/CVPR46437.2021.01350

**Authors’ background**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Title\*** | **Research Field** | **Personal website** |
| Jan Edgar E. Tupas | Undergraduate Student | Computer Science | N/A |
| Christian Henry Miguel E. Caruz | Undergraduate Student | Computer Science | N/A |
| Rafaello Jose M. Viera | Undergraduate Student | Computer Science | N/A |
| Dr. John Paul Q. Tomas | Full Professor | Computer Science | N/A |