

Article

Optimizing Fish Feeding with FFAUNet Segmentation and Adaptive Fuzzy Inference System

Yo-Ping Huang ^{1,2,3,4,*}  and Spandana Vadloori ¹ ¹ Department of Electrical Engineering, National Penghu University of Science and Technology, Penghu 88046, Taiwan; spandanamanoj@gmail.com² Department of Electrical Engineering, National Taipei University of Technology, Taipei 10608, Taiwan³ Department of Computer Science and Information Engineering, National Taipei University, New Taipei City 23741, Taiwan⁴ Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung 41349, Taiwan

* Correspondence: yphuang@gms.npu.edu.tw or yphuang@ntut.edu.tw

Abstract: Efficient and optimized fish-feeding practices are crucial for enhancing productivity and sustainability in aquaculture. While many studies have focused on classifying fish-feeding intensity, there is a lack of research on optimizing feeding, necessitating a precise and automated model. This study fills this gap with a hybrid solution for precision aquaculture feeding management involving segmentation and optimization phases. In the segmentation phase, we used the novel feature fusion attention U-Net (FFAUNet) to accurately segment fish-feeding intensity areas. The FFAUNet achieved impressive metrics: a mean intersection over union (mIoU) of 89.39%, a mean precision of 95.07%, a mean recall of 95.08%, a mean pixel accuracy of 95.12%, and an overall accuracy of 95.61%. In the optimization phase, we employed an adaptive neuro-fuzzy inference system (ANFIS) with a particle swarm optimizer (PSO) to optimize feeding. Extracting feeding intensity percentages from the segmented output, the ANFIS with PSO achieved an accuracy of 98.57%, a sensitivity of 99.41%, and a specificity of 99.53%. This model offers fish farmers a robust, automated tool for precise feeding management, reducing feed wastage and improving overall productivity and sustainability in aquaculture.



Citation: Huang, Y.-P.; Vadloori, S. Optimizing Fish Feeding with FFAUNet Segmentation and Adaptive Fuzzy Inference System. *Processes* **2024**, *12*, 1580. <https://doi.org/10.3390/pr12081580>

Academic Editor: Iqbal M. Mujtaba

Received: 7 July 2024

Revised: 25 July 2024

Accepted: 26 July 2024

Published: 28 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: aquaculture; fish-feeding optimization; feature fusion attention U-Net (FFAUNet); adaptive neuro-fuzzy inference system (ANFIS); particle swarm optimizer (PSO)

1. Introduction

Aquaculture, centered around aquatic animals such as fish, has evolved into a crucial global industry, supplying a substantial portion of the world's seafood demand. This sector's importance is continually growing, addressing the rising need for seafood and serving as a pivotal solution to meet this demand. However, managing the feeding of these aquatic species is integral to achieving optimal production and ensuring their well-being. Feed cost in aquaculture operations is a significant variable [1,2]. According to the Food and Agriculture Organization [3], feed costs comprise over 50% of the total production cost in intensive aquaculture [4]. A study on European sea farming [5] revealed that the feeding costs are high and often represent more than 63% of total expenses. Recent research conducted in Egypt [6,7] indicated even much higher feed costs, ranging from 75% to 85% of the annual running costs of fish production. In aquaculture, where feed represents a significant variable cost, inadequate feeding hampers regular fish growth, while excessive feeding results in wastage and potential environmental pollution and fish diseases [8–11]. In addition to the amount of feed administered, factors such as feed palatability and attractiveness significantly impact the feeding process. Palatability and attractiveness promote ingestion and utilization of available nutrients, which is crucial for successful feed

manufacture. For instance, diet palatability and attractiveness in shrimp farming reduce the time shrimp approach the feed, limiting nutrient leaching and feed loss, helping to prevent the deterioration of rearing pond environments from overloaded nutrient input [12]. Measuring fish-feeding intensity is a practical solution to address these challenges and enhance production efficiency. Despite advancements, the issue of overfeeding persists in the aquaculture industry mainly due to challenges associated with accurately monitoring fish feed consumption underwater [13]. Precision feeding, a strategy to ensure optimal food quantity for farm animals, has emerged as a solution to promote growth and performance while minimizing feed wastage [14]. Traditional aquaculture feeding practices relied on subjective, manual methods, incorporating historical practices and direct physical monitoring of fish-feeding behavior [15]. Attempts to integrate temperature sensors into management software for assessing fish appetite proved time-intensive and impacted the accuracy of crucial feeding information [16].

Deep learning has advanced various techniques in aquaculture, including live fish identification, fish species classification, analysis of fish behavior, biomass estimation, fish-feeding intensity classification, and water quality prediction [2,17]. Yang et al. [18] proposed an improved Efficientnet-B2 network for classifying feeding intensity regions and obtained an accuracy of 89.56%. Feng et al. [19] performed feeding intensity classification using 3D ResNet-GloRe and obtained a classification accuracy of 92.68%. Ubina et al. [16] utilized a deep optical flow-based neural network for video interpolation and the last layer for optical flow regression, reaching up to 95% accuracy. Similarly, Zhou et al. [20] classified fish-feeding intensity using a CNN, achieving an accuracy of 90%. Parra et al. [21] implemented a sensor system to oversee and regulate fish-feeding activities in marine fish farms, while Chiang and Chang [22] developed sensors for measuring salinity levels. Zeng et al. [23] proposed an audio spectrum Swin transform model to grade the feeding intensity based on the acoustic signal and achieved 96.16% accuracy. Zhou et al. [24] applied a machine-learning SVM model to classify reflective frames based on image texture. Zhou et al. [25] introduced the flock-feeding intensity index, a nonlinear metric gauging fish-feeding intensity by considering the dispersion of fish during feeding. Zhao et al. [26] implemented a modified DCGAN-based semi-supervised learning model for live fish identification with 83.07% accuracy. Wu et al. [27] suggested an ANFIS utilizing time-series data, resulting in a notable accuracy of 97.89%. Zhao et al. [28] developed an automatic feeding decision-making ANFIS system based on water quality parameters. Wang et al. [29] proposed a model and achieved 95.44% accuracy in feeding activity classification and counting the uneaten feeding pellets. Han et al. [30] proposed a simple CNN to quickly identify the behavior state of fish shoals. Huang et al. [31] proposed an automatic detection method for eating behavior based on audio analysis. Most of the previous studies have focused on classification.

In this context, the current research study aims to address these challenges by developing an innovative framework, namely the feature fusion attention U-Net (FFAUNet). This framework combines deep learning using the VGG16 [32] feature extraction backbones. It constructs a two-layer feature fusion employing the same level composition (SLC) [33–35] interconnected by an attention mechanism followed by a segmentation block. This enhancement is designed to improve feature extraction, emphasizing highly discriminative features of fish-feeding regions and their intensity while de-emphasizing less relevant areas. The FFAUNet is designed to segment fish-feeding areas and their intensity, providing a comprehensive solution to enhance the precision and efficiency of fish feeding in aquaculture. Furthermore, we utilized an adaptive neuro-fuzzy inference system (ANFIS) with a particle swarm optimizer (PSO) for feeding optimization. This study contributes to the limited existing literature on using deep-learning models for identifying fish-feeding intensity, opening up new avenues for technological advancements in aquaculture practices. The primary contributions of this study are outlined as follows:

- (1) The proposed model significantly improves the accurate segmentation of the fish-feeding area compared with the original state-of-the-art VGG16-based U-Net model and ResNet-based U-Net model. This enhancement is critical for precise and reliable segmentation of fish-feeding areas.
- (2) A novel feeding optimization framework was introduced to combine a FFAUNet with an ANFIS. This innovative approach accurately segmented feeding areas and their intensity on water surfaces. Additionally, this study achieved superior accuracy in feeding optimization by integrating the PSO and ANFIS models.
- (3) Beyond the core framework, a comprehensive evaluation of the PSO and ANFIS model was conducted, comparing its performance with other prominent algorithms such as artificial neural networks (ANNs), support vector machines (SVMs), and random forest (RF). This multi-algorithm evaluation provides insights into the robustness and efficacy of the proposed feeding optimization model.
- (4) The current technology can open up new possibilities for more precise approaches to fish-feeding practices in aquaculture.

These contributions collectively mark a significant advancement in aquaculture, showcasing the potential of the proposed framework to revolutionize fish-feeding optimization through improved accuracy, innovative methodologies, and the integration of cutting-edge technologies.

The rest of the paper is organized as follows: Section 2 details the materials and methods applied in this study. Section 3 gives the experimental setup, results, and discussions. The paper is concluded in Section 4 with the conclusions and proposed future work.

2. Materials and Methods

2.1. Dataset Details

The dataset comprised a collection of live video recordings captured by a surveillance system focusing on the water within a fishpond. These recordings captured the entire feeding process of the fish, from the initial splash to the conclusion of the recording. Based on the feeding intensity observed on the water surface, we categorized the feeding as 'Heavy', 'Medium', and 'Normal'. In the 'Heavy' feeding category, fish exhibit strong, vigorous movements and actively swim to the feed, creating noticeable disturbances on the water surface as they compete for food. The 'Medium' feeding category is characterized by moderate activity on the water surface, with fewer feeding sessions than 'Heavy' feeding. In the 'Normal' feeding category, fish do not respond to the feed. The surface remains calm and the fish are either still or swimming very slowly. The primary purpose of utilizing these videos was to meticulously analyze and study the intricate feeding patterns exhibited by the fish.

In this study, a total of 1153 images derived from full-length videos that were converted into frames were used. These images were labeled: 'Background', 'Normal', 'Medium', and 'Heavy' based on the intensity of water splash on the surface. Annotation of these images was performed using LabelMe software version 5.2.1. The data were then split into a training set (90%) and a testing set (10%), with a further subdivision of 10% from the training set for validation purposes. The feeding area on the aquaculture fishpond is depicted in Figure 1.

2.2. Attention Mechanism

Attention mechanisms have become crucial in various deep-learning architectures, including those used for image classification, segmentation, and detection. In this context, attention mechanisms help the model focus on relevant parts of the input image while suppressing irrelevant information by merging channel-wise and spatial features and enhancing the model's ability to accurately identify the regions of interest in an image by integrating the attention mechanisms into the backbone network [36–40]. This enhancement integrates the squeeze-and-excitation (SE) attention [36] mechanism into the proposed model. The introduced SE attention modules bring adaptability to the network by

dynamically adjusting the importance of different channels based on the data context [36] (Figure 2). The SE attention mechanism involves extracting global information through average pooling (squeeze phase) and then learning channel-wise activation coefficients using a neural network (excitation phase). The attention-weighted feature map is obtained by scaling the original features based on the computed activation coefficients. This mechanism allows the model to highlight important channels and suppress less relevant ones adaptively. This enhancement in model performance arises from the SE attention mechanism's ability to guide the network focus toward more informative features while diminishing the influence of less valuable information. As a result, the increased focus on significant details and subtle patterns enhances the network's discriminative capabilities and ability to effectively generalize to previously unseen data, ultimately leading to improved feature representation and heightened model accuracy.



(a)



(b)



(c)

Figure 1. Illustration of different feeding scenarios depicting the intensity of feeding activity in aquaculture. (a) Heavy feeding, (b) Medium feeding, and (c) Normal feeding.

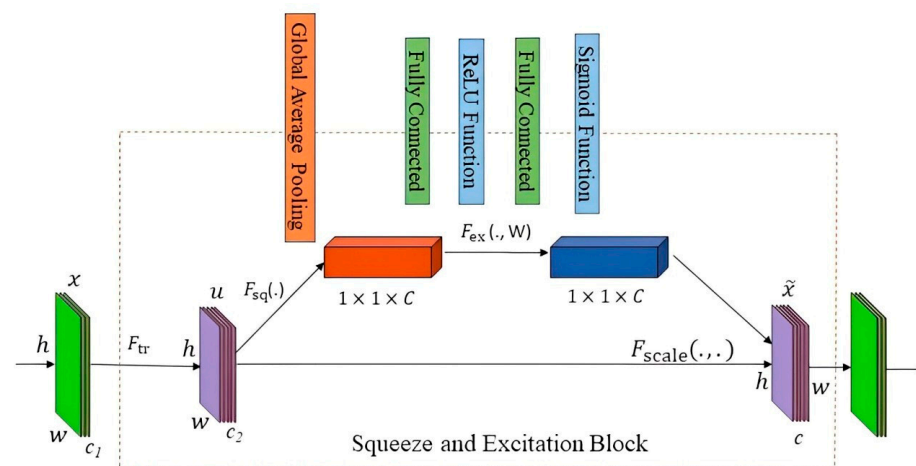


Figure 2. Squeeze and excitation attention block architecture.

Algorithm

Step 1: Squeeze Phase: The SE attention mechanism is applied to each channel of the feature map independently. The global average pooling operation is applied across the spatial dimensions of the input feature map to obtain a channel-wise descriptor vector. Let $X \in \mathbb{R}^C \times H \times W$ be the input feature with C channels and spatial dimensions $H \times W$ of the feature map. The channel-wise descriptor vector z , denoted as z_c , is computed for each channel c , as shown in Equation (1).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_{i,j,c} \quad (1)$$

where $X = \{x_1, x_2, \dots, x_c\}$, H is the height, W is the width of the feature maps, and $x_{i,j,c}$ is the value at position (i, j) in channel c . The first step is to perform global average pooling across spatial dimensions for each channel. This operation produces a channel-wise statistic z representing the average activation for each channel.

Step 2: Excitation Phase: The channel-wise descriptor vector z obtained from the squeeze phase is then passed through a two-layer neural network to compute channel-wise activation coefficients.

Step 3: Squeeze Layer: A squeeze operation defined in Equation (2) involves passing the z through a fully connected (FC) layer with a nonlinear activation function δ :

$$u_c = \delta(W_1 z_c + b_1) \quad (2)$$

where W_1 is the weight matrix of the FC layer, b_1 is the bias vector of the FC layer, and δ is a nonlinear activation function like ReLU.

Step 4: Excitation Layer: Another FC layer is employed to capture channel-wise dependencies and generate scaling factors using Equation (3).

$$s_c = \sigma(W_2 u_c + b_2) \quad (3)$$

W_2 is the weight matrix of the second FC layer, b_2 is the bias vector of the second FC layer, and σ is a sigmoid activation function.

Step 5: Scale and Reshape: The channel-wise activation coefficient s_c is used to scale the original feature map channels. The scaled feature map $Y_{i,j,c}$ is obtained by element-wise multiplication with the original feature map $x_{i,j,c}$ as shown in Equation (4).

$$Y_{i,j,c} = s_c \cdot x_{i,j,c} \quad (4)$$

The scaled feature map Y is the final attention-weighted feature map.

2.3. Proposed Segmentation Model (FFAUNet Architecture)

The proposed network consists of two identical parallel backbones. This design allows for integrating features at the same hierarchical level from both backbones, enhancing feature refinement. Specifically, features from blocks 1, 2, ..., n in backbone-1 (denoted as X) are combined with features of the same size from the corresponding blocks in backbone-2 (denoted as Y), as shown in Equation (5). This merging process facilitates more precise segmentation by refining the features through their combination. These combined features undergo refinement through the corresponding attention block (Equation (6)) before passing these features to the next stage in backbone-2 (Equation (7)).

$$Z_i = X_i + Y_i \quad (5)$$

$$W = SE_{Attention}(Z) \quad (6)$$

$$Y' = Nextstage(W) \quad (7)$$

where Z_i is the i th element of the combined features Z , X_i is the i th element of the features from backbone-1, and Y_i is the i th element of the features from backbone-2. $SE_Attention(Z)$ is the operation performed by the attention block to refine features. These refined features, denoted as W , are then forwarded to the next stage Y' in backbone-2. The refined features W , obtained from the SE attention block at each stage, are used as skip connections and directed into the up-sampling convolutional blocks (Figure 3). For image segmentation, the VGG16-based U-Net architecture [41] was employed. This variant of the U-Net model uses the VGG16 network as its encoder. The U-Net architecture consists of an encoder path that captures contextual information and a decoder path for precise localization. VGG16 is a well-known CNN model praised for its deep architecture and strong performance in various computer vision tasks [42]. The feature extraction block of this model includes several convolutional and max pooling layers that progressively downsample the input image to capture high-level features. A comparative analysis against three other models was conducted to assess the performance of the proposed model. These three models do not include the attention block: the ResNet50 backbone-based U-Net, the original VGG16-based U-Net model, and the FFUNet. These models represent distinct architectures in image segmentation. ResNet50 [39] was chosen for its proven efficiency in deep neural networks. The original VGG16-based U-Net model represented a benchmark in image segmentation tasks. FFUNet provided a relevant comparison, lacking the attention block present in the proposed model. This comparative analysis helps elucidate the impact of attention mechanisms in enhancing segmentation performance.

2.4. Optimization Phase (ANFIS Architecture)

Furthermore, a comprehensive analysis of the segmented images produced by the proposed FFAUNet model was conducted to identify and categorize the areas within the images as 'Heavy', 'Medium', or 'Normal' regions and, subsequently, implemented a precise ANFIS [43,44] classification system to make informed decisions regarding the course of action for feeding. These decisions were categorized as follows: 'Continue (C)', 'Maintain (M)', 'Reduce (R)', or 'Stop (S)' in the feeding process. In the fish-feeding images, specific labels are placed on different feeding conditions based on the density of feeding observed. When the feeding region in the images was extensive, the decision was labeled as 'C' (indicating heavy feeding), and in cases where the feeding region was moderately heavy, the decision was 'M' (indicating moderate heavy feeding). For images showing a medium level of feeding activity, the decision was labeled as 'R' (indicating medium feeding). Finally, when the water level was back to normal in the feeding region, the decision was 'S' (indicating stop feeding). Figure 4 illustrates the workflow of the proposed model.

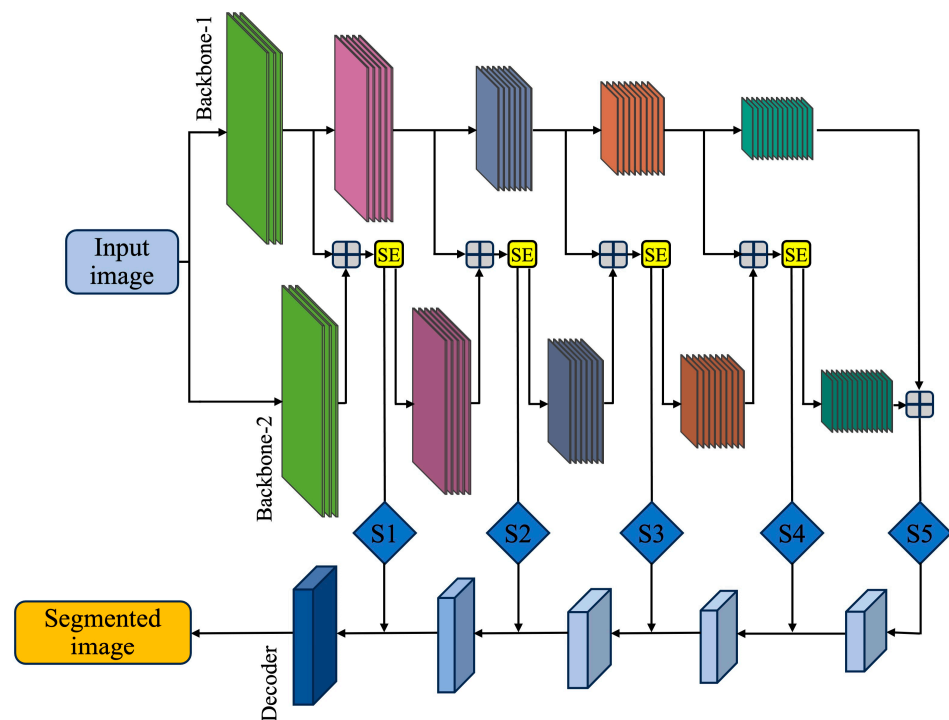


Figure 3. The proposed feature fusion attention U-Net (FFAUNet) model.

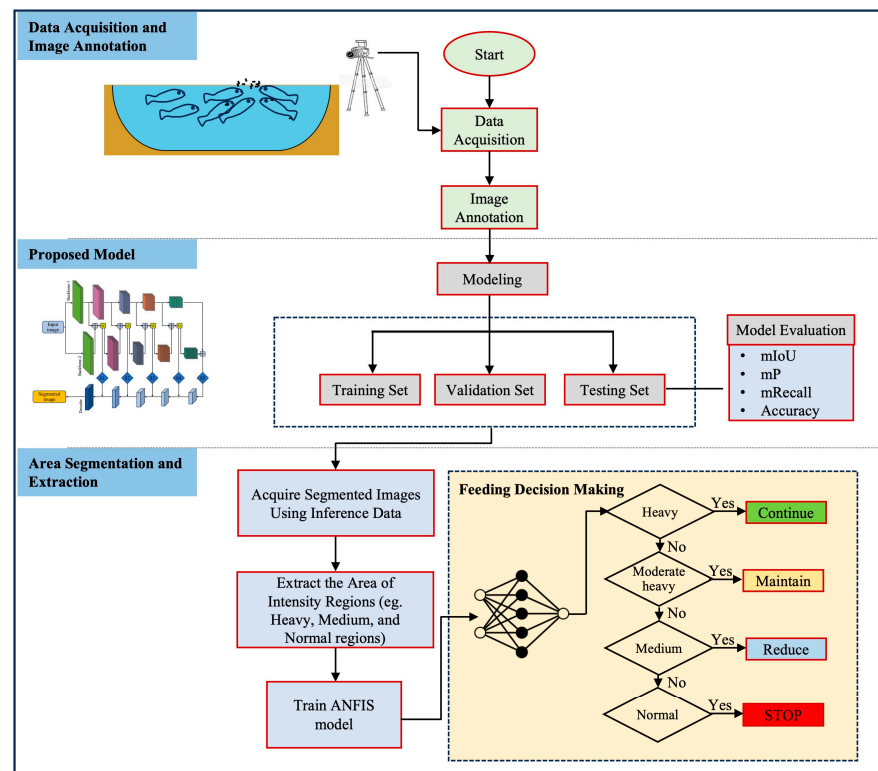


Figure 4. Workflow of the proposed model.

ANFIS (Figure 5) belongs to the category of adaptive networks and shares functional equivalence with fuzzy inference systems (FIS). Its architectural design integrates both a fuzzy system and a neural network. The Takagi–Sugeno (T-S) fuzzy model has garnered significant attention in the engineering field. It offers a systematic approach to modeling nonlinear systems, combining fuzzy logic principles with rigorous mathematical theory.

This integration makes the T-S fuzzy model a powerful tool for researching fuzzy systems. The antecedent parameters and consequent parameters of T-S models are represented as particles and are simultaneously acquired during the training phase [45].

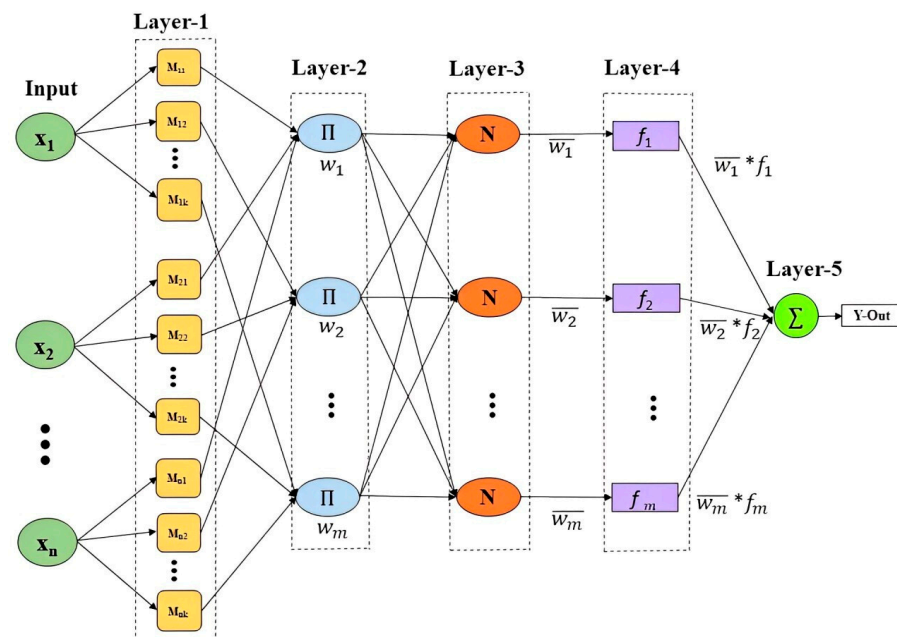


Figure 5. ANFIS model architecture.

In parallel, the ANFIS has demonstrated its versatility across numerous research domains, consistently delivering powerful results. The ANFIS is renowned for its ability to synergize with various algorithms, effectively reducing training phase errors [46]. The ANFIS structure with n input and one output is shown in Figure 5.

The scheme consists of five layers and m IF-THEN fuzzy rules. To make predictions using the fuzzy model featuring n inputs ($x_1, x_2, \dots, x_n$) with linguistic labels $M_{1k}, M_{2k}, \dots, M_{nk}$, where $k = 1, 2, \dots, K$, one output Y , and n fuzzy rules, the fuzzy rule base for the first order T-S model can be formed as follows using Equation (8):

If x_1 is M_{1k} , x_2 is M_{2k} , $\dots$, and x_n is M_{nk} , then

$$f_j = p_j x_1 + q_j x_2 + \dots + r_j x_n + t_j \quad (8)$$

where j represents the j th fuzzy rule; the whole rule base may contain at most $m = K^n$ fuzzy rules. t_j stands for the bias term, which is crucial as it allows for adjusting the output f_j .

The learning process in the ANFIS aims to optimize parameters by applying a learning algorithm using input and output datasets. The consequent parameters, $p_j, q_j, \dots, r_j$, and t_j are determined during the training process within the fuzzy region specified by the fuzzy rules. During training, the learning algorithm adjusts these parameters to minimize the error between the predicted output of the ANFIS model and the actual output from the training data.

The initial layer of the architecture facilitates the transmission of input data from each of its nodes to subsequent layers.

Layer 1: Referred to as the fuzzification layer. In this layer, the fuzzy sets of each input variable are determined. For the input variables x_i , the output of each node (O_i) is the membership grade μ calculated using fuzzy sets M_{ik} shown in Equations (9) and (10).

$$O_i = \mu M_{ik}(x_i), i = 1, 2, \dots, n; k = 1, 2, \dots, K \quad (9)$$

where bell-shaped functions represent the membership functions:

$$\mu M_{ik}(x_i) = \frac{1}{1 + \left| \frac{x_i - c_{ik}}{a_{ik}} \right|^{2b_{ik}}} \quad (10)$$

where a_{ik} , b_{ik} , and c_{ik} are the parameters to be learned, known as the premise parameters of the bell-shaped membership function.

Layer 2: Designated as the rule layer, where it performs the AND operator. In this layer, the firing strength w_j of each rule is computed based on the fuzzy sets in the premise part of the fuzzy rule, as determined in the fuzzification layer. The firing strength for the j th rule is calculated using the product operator as indicated in Equation (11):

$$w_j = \prod_{j=1}^m O_j, \quad j = 1, 2, \dots, m. \quad (11)$$

These firing strengths are then normalized to obtain $\bar{w}_j$.

Layer 3: The normalized firing strengths ($\bar{w}_j$) are calculated using Equation (12):

$$\bar{w}_j = \frac{w_j}{\sum_{j=1}^m w_j} \quad (12)$$

Layer 4: This layer is referred to as the defuzzification layer. In each node of this layer, the weighted values of rules are computed by Equation (13):

$$f_j = \bar{w}_j(p_j x_1 + q_j x_2 + \dots + r_j x_n + t_j), \quad j = 1, 2, \dots, m. \quad (13)$$

Layer 5: The final output (Y_{out}) is computed using Equation (14) by combining the individual rule outputs:

$$Y_{out} = \sum_{j=1}^m \bar{w}_j f_j \quad (14)$$

In the present study, three inputs of the ANFIS ‘Heavy region’, ‘Medium region’, and ‘Normal region’ are selected with a single output classified as ‘C’, ‘M’, ‘R’, or ‘S’. Each of the three inputs is associated with three generalized bell membership functions such as ‘High’, ‘Medium’, and ‘Low’. As a result, the generalized rule base consists of 27 fuzzy rules.

2.5. Particle Swarm Optimization

In this study, a population-based optimization method known as particle swarm optimization (PSO) [47,48] is employed to determine the optimal value of the regularization parameter. The PSO method assists the ANFIS in adjusting the membership function (MF) parameters and minimizing a cost function. The main goal of the PSO algorithm is to explore the search in the search space for the positions that approximate the global minimum or maximum solution(s). In PSO, a swarm initiates with a group of randomly selected solutions called particles. Each particle possesses three parameters, position (x_i) at every iteration, velocity (v_i) representing the rate of position change, and the previous best positions P_i . During each iteration, the most suitable position is determined for each particle by considering the best positions of its neighboring members. The position of the particle with the best fitness value is termed the global best position P_g , thereby iteratively updating their velocities. The velocity update is defined by Equation (15) below:

$$v_i(t+1) = \mathcal{W} \cdot v_i(t) + c_1 \cdot r_1 \cdot (P_i(t) - x_i(t)) + c_2 \cdot r_2 \cdot (P_g - x_i(t)) \quad (15)$$

where $\mathcal{W}$ is the inertia weight, controlling the impact of the previous velocity on the current velocity. c_1 and c_2 are scaling factors. r_1 and r_2 are random numbers in the range [0, 1]. P_i denotes the best solution of the i th particle. P_g is the best-known position among the

neighboring particles. $v_i(t)$ signifies the current velocity of particle i at iteration t , while $v_i(t + 1)$ indicates the updated velocity of particle i for the subsequent iteration $t + 1$.

2.6. Performance Evaluation

The evaluation of classification performance involves utilizing accuracy, precision, recall, mean intersection over union (mIoU), mean pixel accuracy (mPA), mean precision (mP), and mean recall (mR) obtained from the confusion matrix, which includes true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Specifically, TP represents predictions where the predicted class label on the test data aligns with the class of the ground truth image. FP pertains to predictions in which the predicted class does not correspond to the class of the ground truth image. TN denotes predictions where the model correctly anticipates that a given test image does not belong to a specific class. FN signifies predictions where the test image belongs to the ground truth image class, but the model inaccurately predicts it as not belonging to that class.

(1) Accuracy:

This metric represents the ratio of correctly predicted classifications to the overall number of predictions made and is defined by Equation (16).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

(2) Precision:

Precision refers to the proportion of accurately classified samples in relation to the TP and FP values. It computes the percentage of correctly classified samples (Equation (17)).

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

(3) Recall/Sensitivity:

Recall is the ratio of TP to the sum of TP and FN and is defined by Equation (18).

$$Recall = \frac{TP}{TP + FN} \quad (18)$$

(4) Specificity:

Equation (19) indicates how well a test can identify negative cases.

$$Specificity = \frac{TN}{TN + FP} \quad (19)$$

(5) mIoU:

mIoU is a widely used evaluation metric for image segmentation tasks. It measures the average of the Intersection over Union (IoU) values for each class. IoU for a specific class is calculated as the ratio of the intersection of predicted and ground truth regions to their union. Let n be the number of classes. For each class i , IoU is calculated using Equation (20):

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (20)$$

Then, the mIoU is computed as the average IoU over all classes (Equation (21)):

$$mIoU = \frac{1}{n} \sum_{i=1}^n IoU_i \quad (21)$$

(6) mPA (mean Pixel Accuracy):

Let N be the total number of pixels. Correct_Pixels represents the total number of correctly classified pixels. mPA is calculated using Equation (22):

$$mPA = \frac{\text{Correct_Pixels}}{N} \quad (22)$$

(7) mP (mean Precision):

Let n be the number of classes. For each class i , the mP is computed as the average precision over all classes using Equation (23):

$$mP = \frac{1}{n} \sum_{i=1}^n \text{Precision}_i \quad (23)$$

(8) mR (mean Recall):

Let n be the number of classes. For each class i , the mR is computed as the average recall over all classes using Equation (24):

$$mR = \frac{1}{n} \sum_{i=1}^n \text{Recall}_i \quad (24)$$

3. Results and Discussion

The input images to the model are resized to $512 \times 512 \times 3$. The proposed models were performed in Python 3.10 on a Windows 10 operating system with a configuration of Intel Core CPU i7-10700F and 8 GB RAM, supported by GPU (NVIDIA GeForce RTX 3060 Ti, Santa Clara, CA, USA). Python programming and the hyperparameters used for training are batch size of 2, Adam optimizer, with an initial learning rate of 1×10^{-6} and a minimum learning rate of $\text{Init_lr} * 0.0001$.

The FFAUNet model yielded impressive results on the test dataset across multiple evaluation metrics. The mean intersection over union (mIoU) was 89.39%, reflecting the model's robust segmentation performance, and exhibited a mean pixel accuracy (mPA) of 95.12%, underlining its precision in pixel-level classification. The model also demonstrated reasonable precision of 95.07% and recall of 95.08%. The obtained accuracy with the proposed model was 95.61% on the test dataset. These values highlight the model's ability to identify and classify relevant features. The model's performance details are provided in Table 1. The comparative analysis of the four models employed in this study—the original Resnet50, VGG16-based U-Net, the feature fusion U-Net without attention (FFUNet), and the proposed FFAUNet model—is given in Table 2.

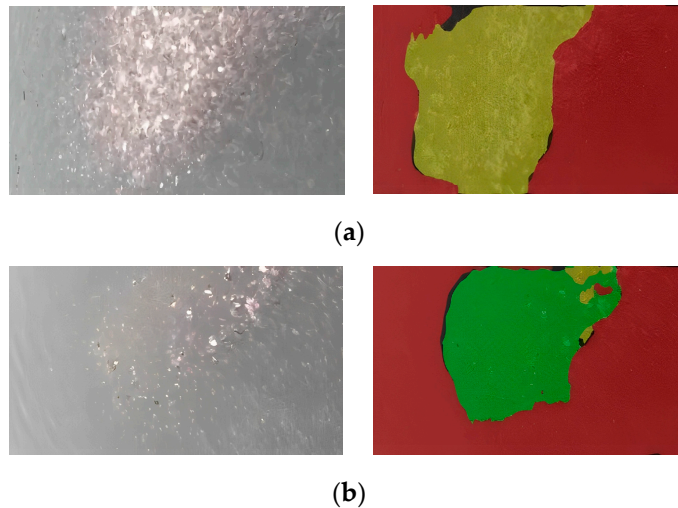
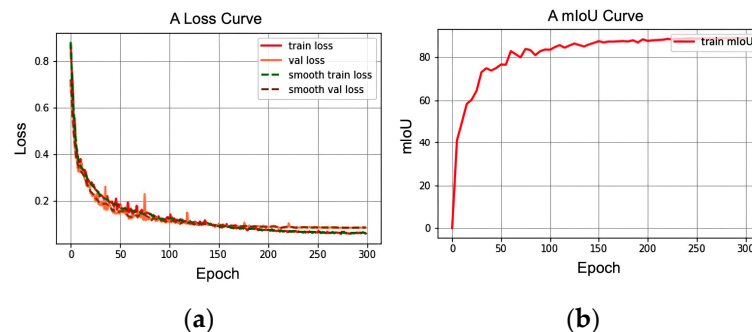
Table 1. Performance metrics of each class by the proposed model.

Class	mIoU (%)	mPA (%)	mP (%)	mR (%)
Normal	93.15	97.00	96.02	96.02
Medium	85.87	94.05	93.04	93.17
Heavy	91.17	96.25	97.11	97.05
Background	87.38	93.17	94.12	94.09

The accuracy resulting from Resnet50 was 93.69%. The original VGG16-based U-Net achieved an accuracy rate of 94.21%, while the FFUNet got slightly better at 94.80%. The proposed FFAUNet model achieved an outstanding accuracy rate of 95.61%. The proposed model learned the features better on the limited dataset. The SE attention block focused on essential features by suppressing the less important features. The segmentation images produced by the FFAUNet model are shown in Figure 6a, b. The loss and mIoU curves of the proposed method are given in Figure 7.

Table 2. Comparative results of the four models used for this study.

Models	mIoU (%)	mPA (%)	mP (%)	mR (%)	Accuracy (%)
Resnet50	86.54	92.61	92.96	92.61	93.69
VGG16Unet	87.55	93.33	93.35	93.33	94.21
FFUNet	88.34	93.85	93.73	93.85	94.80
FFAUNet	89.39	95.12	95.07	95.08	95.61

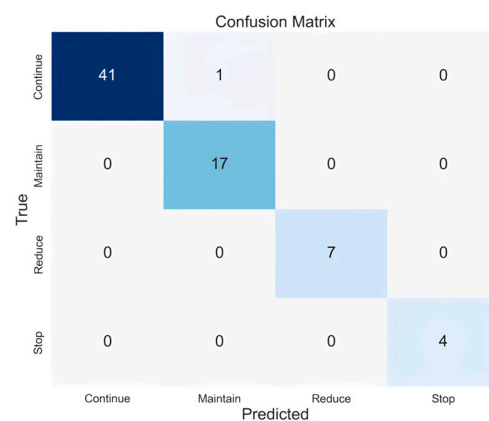
**Figure 6.** Segmented output images from the FFAUNet model. (a) Heavy and (b) Medium.**Figure 7.** (a) The proposed model loss curve and (b) the mIoU curve.

An inference dataset comprising 233 images extracted from five videos was utilized to optimize feeding. Initially, the FFAUNet model was employed for image segmentation. The pixel counts within color-coded regions were quantified and represented as percentages to determine intensity levels (Heavy, Medium, and Normal). An example breakdown indicated that the heavy feeding region, marked in yellow, covered 41.14% of the image; the medium feeding region, marked in green, covered 0.04%; and the normal feeding region, marked in red, occupied 56.23%, as illustrated in Figure 6a. Similarly, from Figure 6b, the heavy feeding region covered 1.28% of the image, the medium feeding region was 31.71%, and the normal feeding region occupied 65.32%. The heavy feeding region is much more extensive in Figure 6a compared with that in Figure 6b. The medium feeding region is almost negligible in Figure 6a but substantial in Figure 6b. It is noted that the summation of these regions did not reach 100% due to misclassification of some pixels along the segmented boundaries. Sample illustrations highlighting identified feeding intensity regions from segmented images are provided in Table 3.

Table 3. Extracted feeding intensity areas from the segmented images.

Image Name	Heavy (%)	Medium (%)	Normal (%)	Error (%)	Actual Class
F55.jpg	41.14	0.04	56.23	2.59	C
F24.jpg	21.29	3.31	72.81	2.59	M
F10.jpg	1.61	25.68	70.50	2.21	R
frame1.jpg	0	0	97.24	2.76	S

The data were categorized into four distinct groups based on the feeding intensity observed in the images. These categories include 'Continue (C)', 'Maintain (M)', 'Reduce (R)', and 'Stop (S)'. Initially, the ANFIS model without PSO was trained using these data, with a 70% training set and 30% testing set split. However, even without the PSO optimizer, it achieved an accuracy of 93.56%. Optimal results were achieved by carefully tuning PSO parameters and ANFIS model configurations. PSO parameters included a population size (nPop) of 40 individuals, an acceleration coefficient (phi) of 2.05, using the 'RB' constriction factor type, a velocity factor (vel_fact) set to 0.5, cross-entropy cost function was used, and the algorithm running for 500 epochs. In the ANFIS model, three Gaussian membership functions were employed for each input variable, adopting the Takagi–Sugeno type of fuzzy inference system to process fuzzy rules. The ANFIS with PSO achieved 98.57% accuracy. This enhanced the model's ability to capture complex relationships within the dataset. Prediction results, presented as a confusion matrix, are depicted in Figure 8.

**Figure 8.** Confusion matrix on the test dataset.

Furthermore, to assess the performance of the ANFIS model, specific evaluation models such as random forest (RF), support vector machines (SVMs), and artificial neural networks (ANNs) were employed. These algorithms, which have been chosen for their suitability in evaluating the ANFIS model, encompass a set of techniques and procedures designed to measure their effectiveness in making decisions or predictions based on linguistic variables and rules.

The ANN model consisted of three dense layers with 200, 100, and 64 neuron counts, respectively. The final output layer utilized the Softmax activation function. The Adam optimizer was employed to train the model. The learning rate for the Adam optimizer was set to 0.01. The optimal accuracy is achieved using these parameters in the ANN model.

Training sessions were conducted for various SVM kernel functions to determine the most effective one. The SVM classifier trained encompassed linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian variations, each with different gamma values. Results of the analysis revealed that the medium Gaussian and coarse Gaussian SVM classifiers both yielded the highest accuracy of 95.71%. Specifically, the 'rbf' kernel function with a gamma value of 0.11 was used. For the RF model, the number of estimators

was set to 30, and the random_state was set to 5. The average accuracy obtained was 94.28%, 95.71%, and 97.14% for the RF, SVM, and ANN models, respectively. In comparing these methods with the ANFIS, outstanding performance was obtained with an accuracy of 98.57%, sensitivity of 99.41%, and specificity of 99.53% (Figure 9).

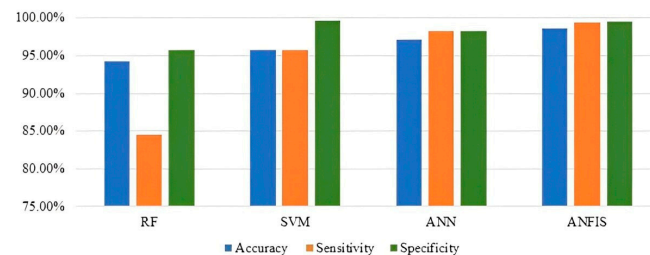


Figure 9. Model's performance matrix.

There are also a few studies that employed the ANFIS tuned with PSO in various problem domains, including healthcare, agricultural domains, etc. Nazari et al. [49] used the ANFIS and PSO to calculate saffron yield based on terrain characteristics in the Siminehrood catchment, located south of Urmia Lake, Iran. In human resource management, Liu et al. [50] utilized an algorithm based on the fusion of PSO with random weight and the ANFIS to accurately predict the human resource structure. Most studies have utilized the ANFIS (tuned with PSO) as a tool for multi-class classification, while some have applied it for binary classification. The literature shows numerous examples of using the ANFIS for predictive model development. A key advantage of the ANFIS is its ability to generate accurate predictions even with limited input data [51].

ANFIS models excel in capturing nonlinear relationships within data, leveraging their fuzzy logic-based approach combined with neural network adaptability. In contrast, SVMs optimally performed when the data were linearly separable, adjusting their parameters to suit the dataset characteristics. Compared with ANNs, ANFIS models possessed fewer parameters and reduced susceptibility to overfitting. They demonstrate robustness even with limited and imbalanced data, surpassing ANNs, SVMs, and RF in these scenarios.

4. Conclusions

This study aimed to leverage advanced technology to optimize fish-feeding practices in aquaculture by developing and implementing the FFAUNet, which integrates powerful feature extraction capabilities with SE attention mechanisms to accurately segment fish-feeding intensity regions. The proposed FFAUNet model demonstrated precise segmentation results. For further optimization of feeding, the ANFIS was used to quantify the areas corresponding to feeding intensity regions within finely segmented images. The ANFIS with PSO showed a remarkable accuracy rate of 98.57% in optimizing fish feeding. This methodology underscores its reliability and effectiveness in optimizing fish-feeding routines and making critical decisions.

This study presents significant progress in aquaculture, showcasing the potential of advanced technology to optimize fish-feeding practices. It sets the stage for a future with data-driven, highly efficient feeding protocols that enhance sustainability. Future efforts will aim to use underwater fish-feeding behavior data to develop an accessible and user-friendly platform. This platform will allow fish farmers to monitor feeding intensity in real time, improve productivity, and ensure the well-being of the fish.

Author Contributions: Conceptualization, Y.-P.H. and S.V.; methodology, Y.-P.H. and S.V.; software, S.V.; validation, Y.-P.H. and S.V.; formal analysis, Y.-P.H. and S.V.; investigation, Y.-P.H. and S.V.; resources, Y.-P.H.; data curation, Y.-P.H. and S.V.; writing—original draft preparation, Y.-P.H. and S.V.; writing—review and editing, Y.-P.H. and S.V.; visualization, Y.-P.H. and S.V.; supervision, Y.-P.H.; project administration, Y.-P.H.; funding acquisition, Y.-P.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially funded by the National Science and Technology Council, Taiwan, under grants MOST108-2221-E-346-006-MY3 and MOST111-2221-E-346-002-MY3, and in part by Fongyu Co., Ltd. under grant 112A0028-02.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of this study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Fore, M.; Alver, M.; Alfredsen, J.A.; Marafioti, G.; Senneset, G.; Birkevold, J.; Willumsen, F.V.; Lange, G.; Espmark, A.; Terjesen, B.F. Modelling growth performance and feeding behaviour of Atlantic salmon (*Salmo salar* L.) in commercial-size aquaculture net pens: Model details and validation through full-scale experiments. *Aquaculture* **2016**, *464*, 268–278. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, D.L.; Wang, Z.H.; Wu, S.Y.; Miao, Z.; Du, L.; Duan, Y.Q. Automatic recognition methods of fish feeding behavior in aquaculture: A review. *Aquaculture* **2020**, *528*, 735508. [\[CrossRef\]](#)
- Wenning, R. The state of world fisheries and aquaculture (Sofia) 2020 Report. *Integr. Environ. Asses.* **2020**, *16*, 800–801.
- Rana, K.J.; Siriwardena, S.; Hasan, M.R. *Impact of Rising Feed Ingredient Prices on Aquafeeds and Aquaculture Production*; FAO: Rome, Italy, 2009; Volume 541.
- Arru, B.; Furesi, R.; Gasco, L.; Madau, F.A.; Pulina, P. The introduction of insect meal into fish diet: The first economic analysis on European sea bass farming. *Sustainability* **2019**, *11*, 1697. [\[CrossRef\]](#)
- Dickson, M.; Nasr-Allah, A.; Kenawy, D.; Kruijssen, F. Increasing fish farm profitability through aquaculture best management practice training in Egypt. *Aquaculture* **2016**, *465*, 172–178. [\[CrossRef\]](#)
- Shaaan, M.; El-Mahdy, M.; Saleh, M.; El-Matbouli, M. Aquaculture in Egypt: Insights on the current trends and future perspectives for sustainable development. *Rev. Fish Sci. Aquac.* **2018**, *26*, 99–110. [\[CrossRef\]](#)
- Ahmed, N.; Thompson, S.; Glaser, M. Global Aquaculture Productivity, Environmental Sustainability, and Climate Change Adaptability. *Environ. Manage.* **2019**, *63*, 159–172. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hardy, R.W.; Kaushik, S.J. *Fish Nutrition*; Academic Press: Cambridge, MA, USA, 2021.
- Jones, H.A.C.; Noble, C.; Damsgård, B.; Pearce, G.P. Evaluating the effects of a short-term feed restriction period on the behavior and welfare of Atlantic salmon, parr using social network analysis and fin damage. *J. World Aquacult. Soc.* **2017**, *48*, 35–45. [\[CrossRef\]](#)
- Remen, M.; Sievers, M.; Torgersen, T.; Oppedal, F. The oxygen threshold for maximal feed intake of Atlantic salmon post-smolts is highly temperature-dependent. *Aquaculture* **2016**, *464*, 582–592. [\[CrossRef\]](#)
- Chutima, T. Feed palatability and the alternative protein sources in shrimp feed. *Songklanakarin J. Sci. Technol.* **2013**, *36*, 51–55.
- Gomes, A.S.; Zimmermann, F.; Hevroy, E.M.; Soyland, M.A.L.; Hansen, T.J.; Nilsen, T.O.; Ronnestad, I. Statistical modelling of voluntary feed intake in individual Atlantic salmon (L.). *Front. Mar. Sci.* **2023**, *10*, 1127519. [\[CrossRef\]](#)
- Zhang, Y.R.; Lu, R.H.; Qin, C.B.; Nie, G.X. Precision nutritional regulation and aquaculture. *Aquacult. Rep.* **2020**, *18*, 100496. [\[CrossRef\]](#)
- Zhou, C.; Xu, D.M.; Chen, L.; Zhang, S.; Sun, C.H.; Yang, X.T.; Wang, Y.B. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. *Aquaculture* **2019**, *507*, 457–465. [\[CrossRef\]](#)
- Ubina, N.; Cheng, S.C.; Chang, C.C.; Chen, H.Y. Evaluating fish feeding intensity in aquaculture with convolutional neural networks. *Aquacult. Eng.* **2021**, *94*, 102178. [\[CrossRef\]](#)
- Ma, Z.; Li, H.X.; Hu, Y.; Fan, J.Z.; Liu, Y. Growth performance, physiological, and feeding behavior effect of under different culture scales. *Aquaculture* **2021**, *534*, 736291. [\[CrossRef\]](#)
- Yang, L.; Yu, H.; Cheng, Y.; Mei, S.; Duan, Y. A dual attention network based on efficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture. *Comput. Electron. Agric.* **2021**, *187*, 106316. [\[CrossRef\]](#)
- Feng, S.X.; Yang, X.T.; Liu, Y.; Zhao, Z.X.; Liu, J.T.; Yan, Y.J.; Zhou, C. Fish feeding intensity quantification using machine vision and a lightweight 3D ResNet-GloRe network. *Aquacult. Eng.* **2022**, *98*, 102244. [\[CrossRef\]](#)
- Zhou, C.; Xu, D.; Lin, K.; Chen, L.; Zhang, S.; Sun, C.; Yang, X. Evaluation of fish feeding intensity in aquaculture based on near-infrared machine vision. *Smart Agric.* **2019**, *1*, 76.
- Parra, L.; García, L.; Sendra, S.; Lloret, J. The use of sensors for monitoring the feeding process and adjusting the feed supply velocity in fish farms. *J. Sens.* **2018**, *2018*, 1060987. [\[CrossRef\]](#)
- Chiang, C.T.; Chang, C.W. Design of a calibrated salinity sensor transducer for monitoring salinity of ocean environment and aquaculture. *IEEE Sens. J.* **2015**, *15*, 5151–5157. [\[CrossRef\]](#)
- Zeng, Y.H.; Yang, X.T.; Pan, L.; Zhu, W.T.; Wang, D.H.; Zhao, Z.X.; Liu, J.T.; Sun, C.H.; Zhou, C. Fish school feeding behavior quantification using acoustic signal and improved Swin Transformer. *Comput. Electron. Agr.* **2023**, *204*, 107580. [\[CrossRef\]](#)
- Zhou, C.; Sun, C.H.; Lin, K.; Xu, D.M.; Guo, Q.; Chen, L.; Yang, X.T. Handling water reflections for computer vision in aquaculture. *T. Asabe* **2018**, *61*, 469–479. [\[CrossRef\]](#)

25. Zhou, C.; Zhang, B.H.; Lin, K.; Xu, D.M.; Chen, C.W.; Yang, X.T.; Sun, C.H. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. *Comput. Electron. Agr.* **2017**, *135*, 233–241. [\[CrossRef\]](#)
26. Zhao, J.; Li, Y.H.; Zhang, F.D.; Zhu, S.M.; Liu, Y.; Lu, H.D.; Ye, Z.Y. Semi-Supervised Learning-Based Live Fish Identification in Aquaculture Using Modified Deep Convolutional Generative Adversarial Networks. *T. Asabe* **2018**, *61*, 699–710. [\[CrossRef\]](#)
27. Wu, T.H.; Huang, Y.I.; Chen, J.M. Development of an adaptive neural-based fuzzy inference system for feeding decision-making assessment in silver perch (*Bidyanus bidyanus*) culture. *Aquacult Eng.* **2015**, *66*, 41–51. [\[CrossRef\]](#)
28. Zhao, S.Q.; Ding, W.M.; Zhao, S.Q.; Gu, J.B. Adaptive neural fuzzy inference system for feeding decision-making of grass carp () in outdoor intensive culturing ponds. *Aquaculture* **2019**, *498*, 28–36. [\[CrossRef\]](#)
29. Wang, Y.Q.; Yu, X.N.; Liu, J.C.; An, D.; Wei, Y.G. Dynamic feeding method for aquaculture fish using multi-task neural network. *Aquaculture* **2022**, *551*, 737913. [\[CrossRef\]](#)
30. Han, F.F.; Zhu, J.C.; Liu, B.; Zhang, B.F.; Xie, F.H. Fish Shoals Behavior Detection Based on Convolutional Neural Network and Spatiotemporal Information. *IEEE Access* **2020**, *8*, 126907–126926. [\[CrossRef\]](#)
31. Huang, J.D.; Zhang, T.M.; Cuan, K.X.; Fang, C. An intelligent method for detecting poultry eating behaviour based on vocalization signals. *Comput. Electron. Agr.* **2021**, *180*, 105884. [\[CrossRef\]](#)
32. Simonyan, A.; Zisserman, K. Very deep convolutional networks for large scale image recognition. *Comput. Vis. Pattern Recognit.* **2015**, 1409.1556. Available online: <https://api.semanticscholar.org/CorpusID:14124313> (accessed on 6 July 2024).
33. Zhao, Z.X.; Liu, Y.; Sun, X.D.; Liu, J.T.; Yang, X.T.; Zhou, C. Composited FishNet: Fish detection and species recognition from low-quality underwater videos. *IEEE Trans. Image Process* **2021**, *30*, 4719–4734. [\[CrossRef\]](#)
34. Liang, T.T.; Chu, X.J.; Liu, Y.D.; Wang, Y.T.; Tang, Z.; Chu, W.; Chen, J.D.; Ling, H.B. CBNet: A composite backbone network architecture for object detection. *IEEE Trans. Image Process* **2022**, *31*, 6893–6906. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Huang, Y.P.; Khabusi, S.P. A CNN-OSELM multi-layer fusion network with attention mechanism for fish disease recognition in aquaculture. *IEEE Access* **2023**, *11*, 58729–58744. [\[CrossRef\]](#)
36. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E.H. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal.* **2020**, *42*, 2011–2023, 2547–2560. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Fu, J.; Liu, J.; Jiang, J.; Li, Y.; Bao, Y.; Lu, H. Scene Segmentation With Dual Relation-Aware Attention Network. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 2547–2560. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Su, B.Y.; Chen, H.Y.; Liu, K.; Liu, W.P. RCAG-Net: Residual channelwise attention gate network for hot spot defect detection of photovoltaic farms. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3510514. [\[CrossRef\]](#)
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
40. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
41. Ronneberger, P.F.O.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention, Munich, Germany, 5–9 October 2015.
42. Huang, Y.P.; Vadloori, S.; Chu, H.C.; Kang, E.Y.C.; Wu, W.C.; Kusaka, S.; Fukushima, Y. Deep learning models for automated diagnosis of retinopathy of prematurity in preterm infants. *Electronics* **2020**, *9*, 1444. [\[CrossRef\]](#)
43. Jang, J.-S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst.* **1993**, *23*, 665–686. [\[CrossRef\]](#)
44. Afnan, A.A.; Omar, E.; Uvais, Q.; Somaya, A.-M. ANFIS-Net for automatic detection of COVID-19. *Sci. Rep.* **2021**, *11*, 17318.
45. Wiktorowicz, K.; Krzeszowski, T. Training high-order Takagi-Sugeno Fuzzy systems using batch least squares and Particle swarm optimization. *Int. J. Fuzzy Syst.* **2020**, *22*, 22–34. [\[CrossRef\]](#)
46. Rathnayake, N.; Dang, T.L.; Hoshino, Y. A novel optimization algorithm: Cascaded adaptive neuro-fuzzy inference system. *Int. J. Fuzzy Syst.* **2021**, *23*, 1955–1971. [\[CrossRef\]](#)
47. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995.
48. Bashir, Z.A.; Hawary, M.E.E. Applying wavelets to short-term load forecasting using PSO-based neural networks. *IEEE Trans. Power Syst.* **2009**, *24*, 20–27. [\[CrossRef\]](#)
49. Nazari, H.; Mohammadkhani, N.; Servati, M. Saffron yield estimation by adaptive neural-fuzzy inference system and particle swarm optimization (ANFIS-SCM-PSO) hybrid model. *Arch. Agron. Soil Sci.* **2021**, *69*, 461–475. [\[CrossRef\]](#)
50. Liu, X.; Zhou, Z. A novel prediction model based on particle swarm optimization and adaptive neuro-fuzzy inference system. *J. Intell. Fuzzy Syst.* **2017**, *33*, 3137–3143. [\[CrossRef\]](#)
51. Pasha, A.; Ahmed, S.T.; Painam, R.K.; Mathivanan, S.K.; Karthikeyan, P.; Mallik, S.; Qin, H. Leveraging ANFIS with Adam and PSO optimizer for Parkinson’s disease. *Heliyon* **2024**, *10*, e30241. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.