

A Regression-Based Deep Learning Approach for Fish Fry Counting: Addressing Label Imbalance and Annotation Ambiguity

Melly Damara Chaniago^{1*}, Lukman Zaman², Yosi Kristian³

^{1,2,3}Department of Information Technology, Faculty of Science and Technology, Institut Sains Teknologi Terpadu Surabaya, Surabaya, East Java, Indonesia

E-mail: ^{1*}mellydamara@gmail.com, ²lz@stts.edu, ³yosi@stts.edu

(Received: 25 May 2026, revised: 23 Jun 2026, accepted: 24 Jun 2026)

Abstract

Manual counting of Nile tilapia fry is time-consuming and prone to human error, particularly under high-density conditions where visual overlap creates significant annotation ambiguity. Furthermore, real-world data collection often results in skewed label distributions. To address these challenges, this study proposes a regression-based fry counting system using EfficientNetV2S with transfer learning. Using a dataset collected from a real hatchery environment, targeted count-level balancing and rigorous annotation refinement were applied prior to model training to improve label consistency and data distribution. The novelty of this study lies in the integration of a count-level balancing strategy and a systematic annotation refinement process to address label imbalance and annotation ambiguity, two critical issues that are often overlooked in regression-based fish fry counting. The proposed model was evaluated using standard regression metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), coefficient of determination (R^2), and Mean Absolute Percentage Error (MAPE). Experimental results showed that the best-performing model achieved an optimized MAE of 13.25 fry, RMSE of 16.93 fry, R^2 of 0.985, and MAPE of 7.07%. Comparative experiments further demonstrated that the proposed EfficientNetV2S architecture outperformed EfficientNetB0, ResNet50, and MobileNetV2. These findings indicate that systematically addressing annotation ambiguity and balancing training data significantly contributes to robust fry count estimation performance in practical aquaculture applications.

Keywords: Deep Learning, EfficientNetV2S, Fish Fry Counting, Annotation Refinement, Transfer Learning.

I. INTRODUCTION

Nile tilapia (*Oreochromis niloticus*) is one of the most widely cultivated freshwater fish species due to its rapid growth rate, adaptability, and economic value in aquaculture production[1]. In hatchery activities, fry counting plays an important role in stock estimation, feeding management, and production planning. However, fry counting is still commonly performed manually, requiring considerable time and labor, especially when the number of fry reaches thousands in a single transaction. Under high-density conditions, overlapping fry and continuous movement may increase human counting errors, resulting in reduced accuracy and potential economic losses in aquaculture operations.

Recent developments in computer vision and deep learning technologies have enabled automated object counting from digital images [2], [3]. Various studies have applied convolutional neural networks (CNNs), object detection methods, and regression-based approaches for fish population

estimation and aquatic object counting. Zhang et al. [2] introduced a fully convolutional regression network for shrimp egg counting using synthetic data generated by a Generative Adversarial Network (GAN) and achieved high counting accuracy under relatively uniform visual conditions. This study demonstrates the potential of regression-based approaches for dense small-object counting tasks. Another study by Khai et al. [3] applied Mask R-CNN for underwater fish detection and counting. Although the method achieved promising detection performance, detection-based approaches may experience performance degradation under high-density conditions due to object overlap and occlusion.

Several recent studies have also investigated fish larvae counting using deep learning approaches. Costa et al. [4] proposed an automatic Nile tilapia larvae counting method using smartphone images and a CNN-based multi-stage refinement approach. Meanwhile, Carvalho et al. [5] evaluated several object detection architectures, including YOLOv8, DETR, and RT-DETR, for fish larvae counting tasks.

However, these approaches still rely on individual object localization before counting, which may reduce efficiency and prediction stability in dense fry environments.

Despite the architectural advantages of regression networks, the performance of these models is fundamentally bottlenecked by dataset quality. In practical aquaculture environments, collecting visual data inherently leads to two major challenges: label imbalance and annotation ambiguity. First, hatcheries naturally produce highly skewed data distributions, clustering around specific transaction sizes while leaving intermediate count ranges underrepresented. This imbalance biases the regression model toward heavily represented densities. Second, establishing accurate ground truth in high-density images is exceptionally difficult; human annotators frequently introduce subjective inconsistencies—or annotation ambiguity—when interpreting severe visual occlusion. Without addressing these dataset deficiencies, even advanced Convolutional Neural Networks (CNNs) will fail to generalize reliably.

To overcome these intersecting challenges, this study proposes a regression-based Nile tilapia fry counting system using EfficientNetV2S with transfer learning. Crucially, rather than relying solely on architectural complexity, this study emphasizes a rigorous data preparation pipeline. We introduce a count-level balancing strategy combined with a systematic annotation refinement process to actively correct label imbalance and reduce ground-truth ambiguity before model training.

Unlike previous studies that primarily focus on network architecture improvements, this study emphasizes dataset quality enhancement through count-level balancing and annotation refinement, thereby addressing two fundamental limitations of regression-based fry counting systems. Overall the main contributions of this study are: (1) The development of an improved data preparation pipeline that mitigates label imbalance through targeted sample balancing and resolves annotation ambiguity via systematic ground-truth refinement, (2) The implementation of a regression-based fry counting approach utilizing an EfficientNetV2S backbone with partial fine-tuning, optimized specifically for dense, overlapping aquaculture environments, and (3) A comprehensive comparative evaluation demonstrating the quantitative impact of annotation refinement on model performance, alongside baseline comparisons against standard CNN architectures (EfficientNetB0, ResNet50, and MobileNetV2).

II. METHODOLOGY

This study aims to develop an artificial intelligence-based system for predicting the number of Nile tilapia fry. The proposed system is expected to automatically estimate fish counts through digital image analysis, thereby assisting the counting and data recording processes in aquaculture activities. Figure 1 depicts the overall workflow of the proposed research, including dataset acquisition, data preparation, model training, and performance evaluation

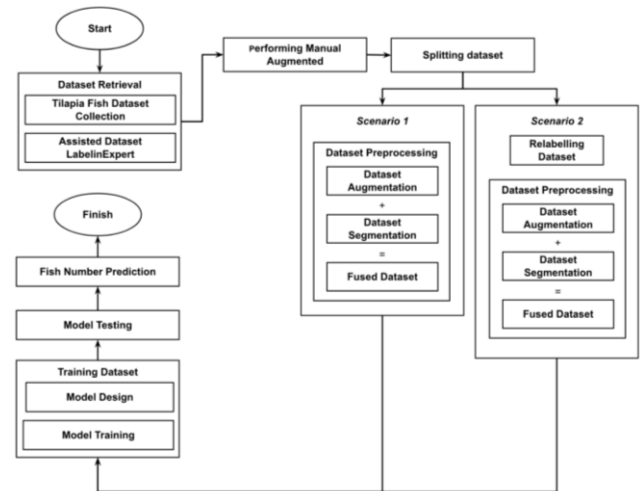


Figure 1. Workflow of the Proposed Fry Counting System.

A. Dataset Acquisition

The dataset used in this study consisted of digital images of Nile tilapia (*Oreochromis niloticus*) fry collected directly from a hatchery environment in Lamongan Regency, East Java, Indonesia. Image acquisition was conducted under controlled conditions using a smartphone camera positioned vertically above a white container filled with clear water to minimize background interference and light reflections. Videos were recorded at a resolution of 1080p and subsequently converted into image frames. The extracted images were manually inspected to ensure adequate visual quality in terms of focus, lighting consistency, and object visibility. The overall experimental setup used during dataset collection is illustrated in Figure 2.

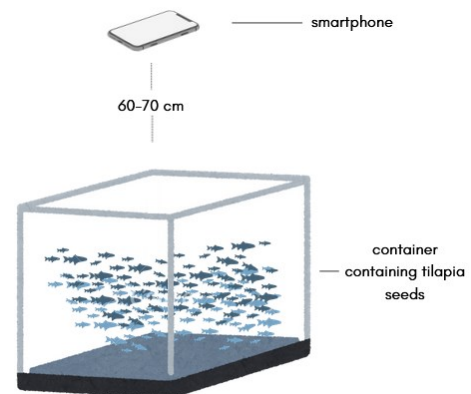


Figure 2. Experimental Setup for Nile Tilapia Fry Image Acquisition.

After the recording process was completed, each video was extracted into static image frames using Adobe Premiere Pro 2022. The extracted frames were then visually inspected to ensure adequate image quality in terms of focus, lighting, and object clarity.

Since Nile tilapia fry continuously moved within the observation container during video recording, successive frames naturally exhibited variations in fry positions, orientations, and overlap patterns. The extracted frames were

manually inspected and selected based on image quality and visual diversity. This process helped reduce redundancy among samples and ensured that the resulting dataset contained sufficiently varied representations of fry distribution conditions.

A total of 435 images were obtained from the video extraction process, containing an overall count of 127,287 Nile tilapia fry across all images. Images that met these quality criteria were subsequently used as the primary dataset for preprocessing, model training, and system evaluation. The collected dataset contains variations in fry density, spatial distribution, and object overlap, representing practical hatchery conditions. Several examples of the collected images are presented in Figure 3.

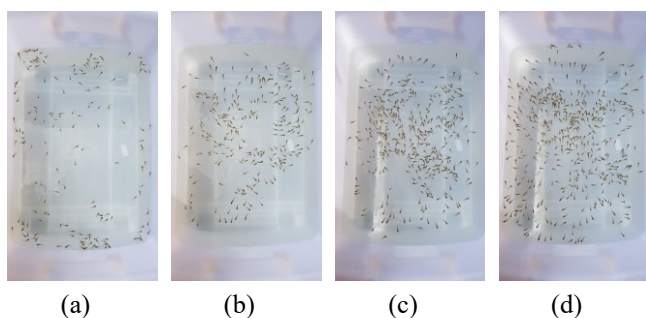


Figure 3. (a) Image With 167 Fry, (B) Image With 243 Fry, (C) Image With 395 Fry, and (D) Image With 455 Fry.

B. Count-Level Balancing Strategy

Initial analysis of the collected dataset revealed an imbalanced distribution of fry count labels, where several count ranges contained substantially more samples than others. In addition, irregular gaps between adjacent count values were observed, which may negatively affect regression learning performance. Under such conditions, the model may become biased toward densely represented count ranges while showing reduced sensitivity to sparsely represented labels.

Rather than generating synthetic images—which risk introducing domain shift—existing high-density images were systematically down-sampled [6]. Individual fry were digitally removed using non-destructive image editing tools (e.g., content-aware deletion) to simulate lower population counts. Crucially, a strict visual inspection protocol was enforced post-editing to ensure the preservation of background continuity, water texture, and lighting consistency. This ensured that no synthetic editing artifacts or spatial anomalies were introduced that the CNN could inadvertently learn as predictive features.

After the balancing process was completed, the dataset increased to 826 images with a total of 246,135 Nile tilapia fry across all images. The generated samples were used exclusively within the training subset to reduce label imbalance and improve training stability. This approach was intended to improve label continuity for regression learning while preserving testing data independence. The distribution of fry count labels before and after the balancing process is presented in Figure 4.

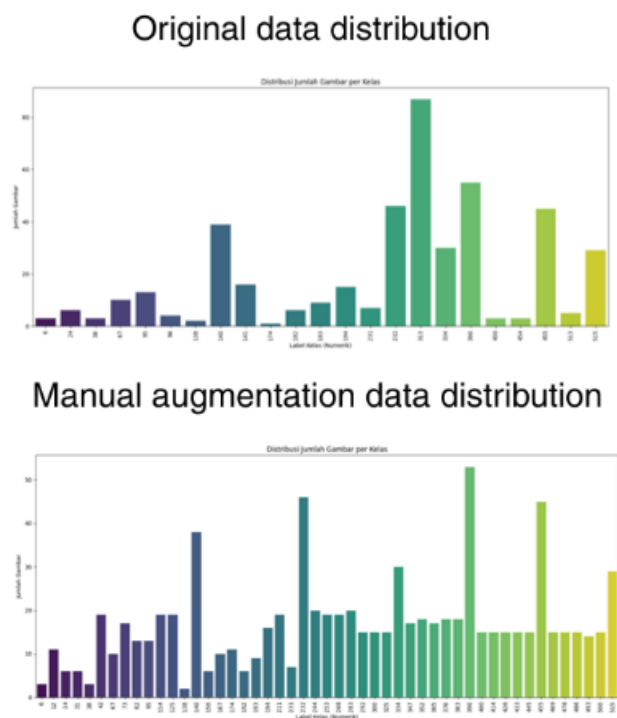


Figure 4. Distribution of Fry Count Labels Before And After The Balancing Process.

C. Annotation Refinement

An annotation refinement process was conducted to improve label consistency in images containing dense fry distributions and visual overlap. In practical hatchery conditions, manual counting may produce inconsistencies due to fry movement, object occlusion, and overlapping spatial distributions [3]. These conditions may affect annotation reliability and subsequently influence regression learning performance.

In high-density hatchery environments, severe physical overlap and visual occlusion introduce significant annotation ambiguity. Initial manual counts frequently contain subjective inconsistencies, which fundamentally cap the learning ceiling of the regression model. To establish a highly reliable ground truth, an annotation refinement protocol was executed utilizing the Roboflow platform [7]. Instead of relying on global visual estimates, this refinement utilized a localized verification strategy. Annotators systematically magnified overlapping fry clusters and utilized point-level annotations to resolve spatial ambiguity. Particular attention was given to partially occluded fry and depth-induced size variations. By standardizing the criteria for what constitutes a "visible" fry, this refinement stage significantly reduced inter-image annotation variance, yielding a robust, high-fidelity ground truth tailored for dense-object regression.

The annotation refinement process was performed following a consistent counting protocol. Each image was manually re-examined using the same annotation criteria to minimize subjective variation during counting. To improve annotation consistency, all corrected labels were carefully

reviewed before being incorporated into the final dataset. This procedure helped ensure that the resulting annotations remained reliable and consistent across different fry density levels and overlap conditions.

It should be noted that the annotation refinement process did not change the number of images in the dataset. The total dataset remained at 826 images, while the refinement process only updated the count labels to obtain more consistent and reliable annotations across different fry density levels.

The annotation refinement process resulted in a more proportional and consistent label distribution across different fry density levels. The distribution of labels after the refinement stage is presented in Figure 5.

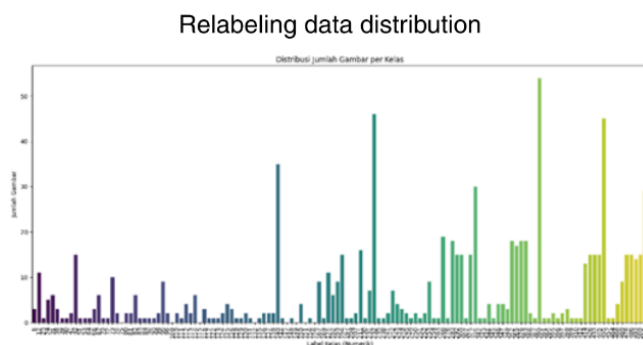


Figure 5. Label Distribution of the Dataset After Relabeling.

D. Scenario 1

In Scenario 1, the dataset was prepared using the original annotations combined with the count-level balancing strategy. Prior to preprocessing, the dataset was divided into training and testing subsets using an 80:20 ratio. The train-test split was performed before any augmentation or segmentation process. Consequently, augmented or segmented versions of testing images were never included in the training dataset, preventing potential data leakage during model evaluation. The testing subset consisted exclusively of original images and was not subjected to additional augmentation or segmentation processes. This strategy was applied to preserve testing data independence and ensure that model evaluation reflects realistic hatchery conditions.

Only the training subset underwent further preprocessing stages, including programmatic image augmentation and image segmentation. Image augmentation was performed using the ImageDataGenerator module from Keras to increase training data diversity and reduce overfitting risk [8]. The applied transformations included rotation, translation, zooming, shear transformation, and horizontal flipping.

In addition to augmentation, image segmentation was performed using a Simple Edge-Morphological Segmentation approach based on Canny edge detection combined with morphological operations [9]. This process was intended to emphasize fry regions while suppressing unnecessary background information, enabling the model to focus on relevant visual features during training.

All augmented and segmented images were subsequently combined with the original training images to form a fused training dataset used exclusively for model training.

E. Scenario 2

Scenario 2 was designed as an extension of Scenario 1 by incorporating the refined dataset generated during the annotation refinement stage. This scenario aimed to evaluate the impact of improved annotation consistency on fry count estimation performance.

Following the annotation refinement process, the updated dataset was divided into training and testing subsets using the same 80:20 ratio applied in Scenario 1. The train-test split was performed before any augmentation or segmentation process. Consequently, augmented or segmented versions of testing images were never included in the training dataset, preventing potential data leakage during model evaluation. Similar to the previous scenario, the testing subset consisted exclusively of original images without additional preprocessing to preserve testing data independence and ensure fair model evaluation under realistic hatchery conditions.

Only the training subset underwent further preprocessing stages, including programmatic image augmentation and image segmentation. Image augmentation was performed to increase visual diversity and improve model robustness against variations in fry orientation, position, and density. Subsequently, image segmentation was applied to emphasize fry regions and reduce background interference during feature extraction.

The resulting augmented and segmented images were subsequently combined with the original training images to form a fused training dataset used exclusively for model training. This experimental configuration enabled direct comparison with Scenario 1 to evaluate the contribution of annotation refinement toward prediction accuracy and model generalization performance.

F. Dataset Training

The dataset training stage was conducted after all data preparation procedures were completed for each experimental scenario. In both Scenario 1 and Scenario 2, the prepared dataset consisted of 826 images. The dataset was initially divided into training and testing subsets using an 80:20 ratio, resulting in 660 training images and 166 testing images.

To preserve testing data independence and prevent data leakage, the testing subset consisted exclusively of original images and was not subjected to augmentation, segmentation, or annotation modification processes. This strategy was adopted to ensure that model evaluation reflects realistic hatchery conditions and accurately measures the model's generalization capability on previously unseen data.

In contrast, only the training subset underwent additional preprocessing stages, including programmatic image augmentation and image segmentation. The resulting augmented and segmented images were subsequently combined with the original training images to form a fused training dataset containing 1,977 images. Meanwhile, the testing dataset remained unchanged with a total of 166 original images.

Based on this training strategy, a total of 2,143 images were used in this study. The distribution of training and testing datasets is summarized in Table 1.

Table 1. Summary of Training and Testing Datasets.

Dataset Type	Number of Images
Training Data	1977
Testing Data	166
Total Data	2143

G. Model Architecture

This study proposes a convolutional regression network designed to automatically estimate the number of Nile tilapia fry from digital images. The proposed architecture employs EfficientNetV2S as the primary feature extraction backbone within a transfer learning framework. EfficientNetV2S was selected due to its computational efficiency and strong feature representation capability achieved through compound scaling strategies balancing network depth, width, and input resolution [10].

The overall architecture of the proposed regression model is illustrated in Figure 6. Input images were resized to $224 \times 224 \times 3$ before being processed by the EfficientNetV2S backbone to extract high-level visual features related to fry distribution, density, and spatial patterns. The extracted feature maps were subsequently forwarded to a fully connected regression head to estimate the fry count value.

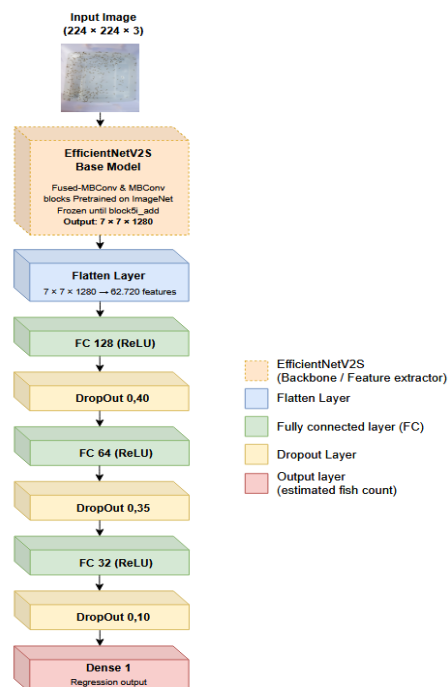


Figure 6. Architecture of the Proposed Convolutional Regression Model

1) Backbone – Efficientnetv2s

The EfficientNetV2S backbone consists of convolutional blocks combining Fused-MBConv and MBConv modules to efficiently extract hierarchical visual features from input images [11]. In this study, the classification layer of EfficientNetV2S was removed, and the network was initialized using ImageNet pre-trained weights. Consequently,

the backbone functioned exclusively as a feature extractor for fry counting tasks.

The use of transfer learning enables the model to leverage low-level visual representations learned from large-scale datasets while adapting higher-level features to the specific characteristics of Nile tilapia fry images [12].

2) Regression Head

The output feature maps generated by the EfficientNetV2S backbone were flattened into one-dimensional feature vectors before being processed by the regression head. The regression head consisted of three sequential Dense layers containing 128, 64, and 32 neurons, respectively, each using the ReLU activation function [13].

To improve training stability and reduce overfitting risk, Batch Normalization was applied after the first Dense layer [14], followed by Dropout layers with rates of 0.40, 0.35, and 0.10, respectively [15]. The final output layer consisted of a single neuron with linear activation to produce continuous fry count predictions.

3) Output Layer

The output layer consisted of a single Dense neuron with linear activation to produce continuous fry count predictions [16]. Since the proposed task is formulated as a regression problem, classification-based activation functions such as sigmoid and softmax were not applied.

4) Fine-Tuning Strategy

A partial fine-tuning strategy was applied to adapt the pre-trained EfficientNetV2S model to the fry counting task [12]. During initialization, convolutional layers below the block5i_add layer were frozen, while deeper layers were unfrozen and trained together with the regression head.

This strategy preserves low-level visual representations learned from ImageNet while enabling higher-level layers to adapt to fry-specific visual patterns, including object overlap, density variation, and fry orientation.

5) Training Configuration

The model was compiled using Mean Absolute Error (MAE) as the primary loss function and Mean Squared Error (MSE) as an additional evaluation metric. Optimization was performed using the Adam optimizer with a learning rate of 0.0003 and AMSGrad enabled to improve convergence stability [17].

Model training was conducted for 40 epochs using shuffled training data. During training, both training and validation losses were monitored to evaluate convergence behavior and detect potential overfitting. The model achieving the lowest validation loss was selected as the final model for performance evaluation.

Overall, the proposed regression model processes normalized fry images through the EfficientNetV2S backbone to extract high-level visual features, which are subsequently forwarded to the regression head for fry count estimation. By leveraging transfer learning and selective fine-tuning, the proposed architecture is able to capture relevant spatial

patterns and produce stable fry count predictions under varying density conditions.

H. Dataset Evaluation

The model evaluation stage was conducted using testing data that were never utilized during the training process in both experimental scenarios. This strategy was adopted to ensure that performance assessment reflects the model’s generalization capability on previously unseen images under realistic hatchery conditions.

During evaluation, each test image underwent the same normalization procedure applied during training before being processed by the trained regression model to generate fry count predictions. Model performance was assessed using several regression-based evaluation metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), coefficient of determination (R^2), and Mean Absolute Percentage Error (MAPE) [17].

MAE was used to measure the average absolute difference between predicted and actual fry counts [18], providing an intuitive representation of prediction error in the same unit as the original data. Meanwhile, MSE and RMSE were employed to evaluate prediction consistency and penalize larger estimation errors more heavily. The coefficient of determination (R^2) was used to evaluate how well the model explains the variance in fry count data, whereas MAPE was applied to represent relative prediction error in percentage form[17].

Lower MAE, MSE, RMSE, and MAPE values indicate better prediction performance, while higher R^2 values indicate stronger agreement between predicted and actual fry counts [19]. Collectively, these metrics provide a comprehensive evaluation of model accuracy, robustness, and predictive reliability for automated fry count estimation.

III. RESULTS AND DISCUSSION

This section presents the experimental results and discussion of the proposed Nile tilapia fry counting system under two experimental scenarios. The evaluation was conducted to analyze the impact of different data preparation strategies, including count-level balancing and annotation refinement, on regression-based fry count estimation performance.

Model performance was evaluated using MAE, MSE, RMSE, R^2 , and MAPE metrics. In addition to quantitative evaluation, further analysis was conducted using scatter plots, MAE distribution analysis across fry density levels, and learning curves to examine prediction behavior, model consistency, and generalization capability under varying hatchery conditions.

A. Scenario 1

In Scenario 1, the model was trained using the original dataset combined with the count-level balancing strategy, followed by augmentation and segmentation processes applied exclusively to the training subset. This configuration

was intended to improve training data diversity and enhance the model’s ability to learn fry distribution patterns under varying visual conditions. The evaluation results for Scenario 1 are summarized in Table 2.

Table 2. Performance Evaluation Results under Scenario 1.

Evaluation Metrics	Data
MAE	22.19
MSE	759.25
RMSE	27.55
R^2	0.962
MAPE	12.28%

The model achieved an MAE of 22.19 fry and an RMSE of 27.55 fry, indicating that the average prediction deviation remained within several tens of fry per image. In addition, the model obtained an R^2 value of 0.962, demonstrating that the proposed regression model was able to explain approximately 96.2% of the variance in fry count data. The MAPE value of 12.28% further indicates that the relative prediction error remained within an acceptable range for automated fry count estimation tasks.

Overall, the results of Scenario 1 demonstrate that the proposed regression model was capable of producing stable fry count predictions under varying density conditions. The combination of count-level balancing, augmentation, and segmentation contributed to improved visual diversity in the training data, enabling the model to learn more robust feature representations related to fry distribution and overlap patterns. Figure 7 presents the scatter plot comparison between actual fry counts and predicted values for Scenario 1.

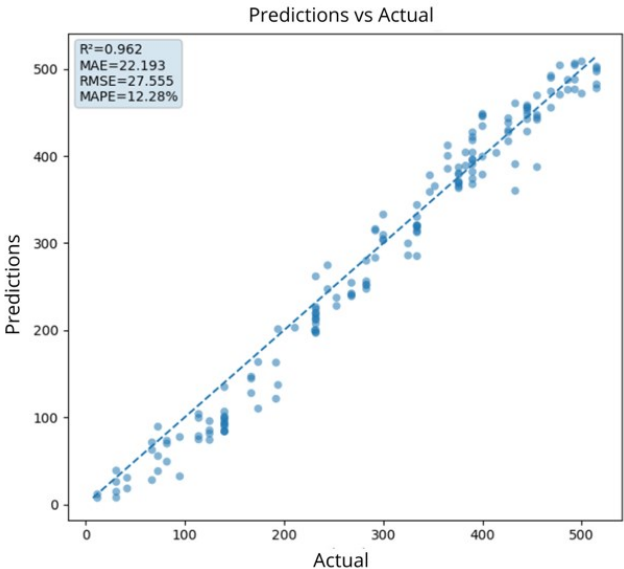


Figure 7. Scatter Plot Comparison Between Actual and Predicted Fry Counts for Scenario 1.

Most prediction points are distributed close to the diagonal reference line, indicating strong agreement between predicted and actual fry counts across the testing dataset. However, a larger prediction dispersion can be observed in several

intermediate and high-density fry groups, indicating that prediction errors tend to increase as fry overlap and spatial complexity become more significant [2],[3].

Despite this dispersion, the overall prediction trend remained proportional across different fry count ranges. This result indicates that the proposed regression model maintained relatively stable prediction capability even under varying fry density conditions, although several prediction deviations were still observed in images containing irregular overlap patterns.

To further analyze prediction consistency across different fry density levels, the distribution of Mean Absolute Error (MAE) was evaluated for multiple fry count intervals. The MAE distribution for Scenario 1 is presented in Figure 8.

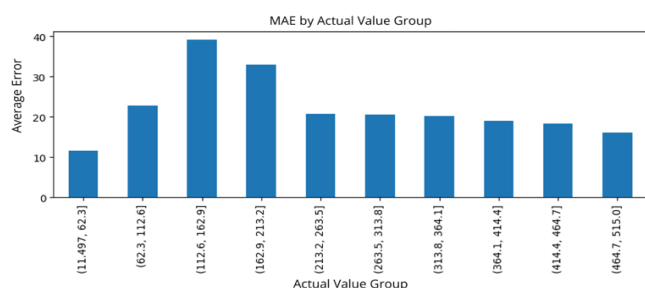


Figure 8. Histogram of Mean Absolute Error (MAE) Distribution for Scenario 1.

The MAE analysis shows that lower prediction errors were generally observed in images containing relatively small fry populations, where object overlap remained limited. Interestingly, the highest prediction error was observed in the intermediate fry count interval (112–162), indicating that the model experienced greater difficulty in estimating fry counts under moderate-density conditions. This behavior may be associated with increased spatial variation and irregular overlap patterns that are less visually consistent compared to extremely dense fry distributions [2].

In contrast, prediction errors in higher-density intervals remained relatively stable without extreme deviations. This finding suggests that the proposed regression model was still able to maintain consistent prediction behavior despite increased fry density and overlap complexity. The learning curve for Scenario 1 is presented in Figure 9.

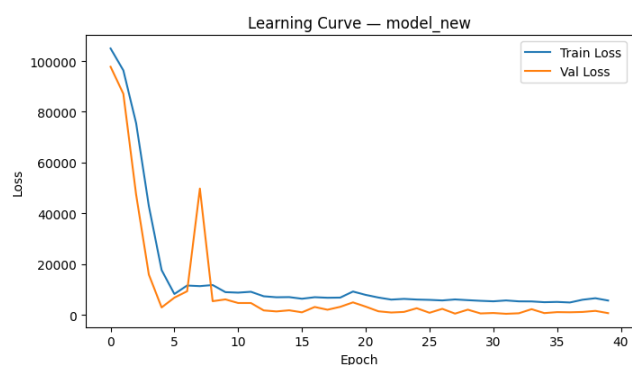


Figure 9. Learning Curve of Training and Validation Loss for Scenario 1.

The learning curves show that both training and validation losses decreased significantly during the early training epochs before gradually stabilizing in later epochs. Although minor fluctuations in validation loss were observed, the overall trend indicates stable convergence behavior without severe overfitting.

The relatively small gap between training and validation loss suggests that the proposed model achieved good generalization capability on previously unseen testing images. These results demonstrate that the training strategy applied in Scenario 1 was effective in learning relevant fry counting features while maintaining stable optimization behavior.

B. Scenario 2

In Scenario 2, the data preparation pipeline followed the same general procedure as in Scenario 1, with the addition of a re-labeling stage performed prior to programmatic augmentation and segmentation. All images resulting from count-level balancing were re-counted, and their labels were corrected using the Roboflow platform through a more rigorous annotation procedure. This step was intended to reduce subjectivity and minimize annotation errors inherent in purely visual manual counting, which was identified as a potential limitation in Scenario 1.

After re-labeling, the images were processed through the same augmentation and segmentation steps as in Scenario 1 and subsequently combined into a new fused dataset. This dataset was then split into training and testing subsets and used to train the regression model for Scenario 2. The quantitative evaluation results for Scenario 2 are summarized in Table 3.

Table 3. Performance Evaluation Results under Scenario 2.

Evaluation Metrics	Data
MAE	13.25
MSE	286.88
RMSE	16.93
R ²	0.985
MAPE	7.07%

The model achieved an MAE of 13.25 fry and an RMSE of 16.93 fry, indicating that the average prediction deviation was substantially reduced compared to Scenario 1. In addition, the model obtained an R² value of 0.985, demonstrating that the proposed regression model was able to explain approximately 98.5% of the variance in fry count data. The MAPE value of 7.07% further indicates that the relative prediction error remained low and within an acceptable range for practical automated fry count estimation.

Overall, the results of Scenario 2 demonstrate that incorporating annotation refinement contributed positively to prediction accuracy and model consistency. More reliable annotations enabled the model to learn a more accurate relationship between visual image features and fry count labels, resulting in more stable prediction behavior across different fry density conditions.

Figure 10 presents the scatter plot comparison between actual fry counts and predicted values for Scenario 2.

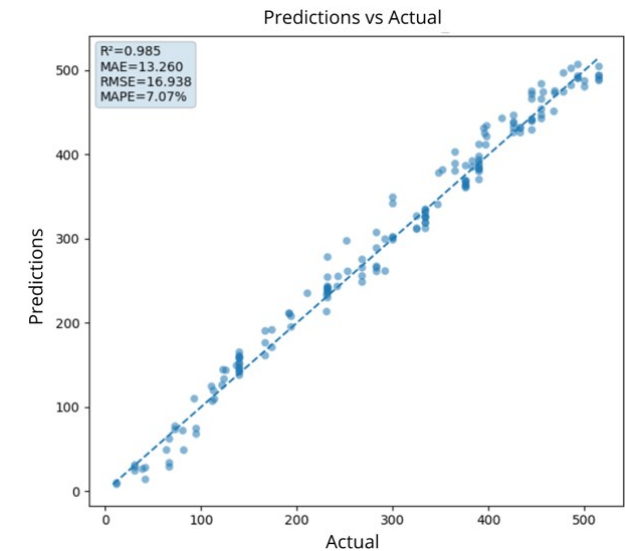


Figure 10. Scatter Plot Comparison Between Actual and Predicted Fry Counts for Scenario 2.

Most prediction points are distributed very close to the diagonal reference line, indicating strong agreement between predicted and actual fry counts across the testing dataset. Compared to Scenario 1, prediction dispersion became noticeably smaller, particularly within intermediate fry density intervals that previously exhibited larger prediction deviations.

This improvement suggests that annotation refinement helped reduce label ambiguity during training, enabling the regression model to learn more consistent visual representations of fry distribution patterns. Although several prediction deviations remained visible in images containing substantial fry overlap, the overall prediction trend remained highly proportional across different fry count ranges.

To further evaluate prediction consistency across fry density intervals, the distribution of Mean Absolute Error (MAE) for Scenario 2 is presented in Figure 11.

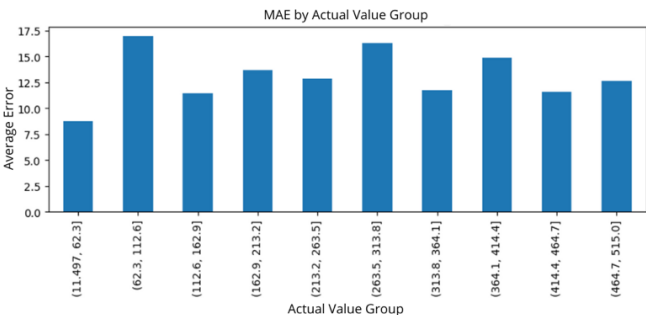


Figure 11. Histogram of Mean Absolute Error (MAE) Distribution for Scenario 2.

The MAE distribution analysis shows that prediction errors remained relatively stable across different fry count intervals, with most MAE values distributed within a relatively narrow range. Compared to Scenario 1, intermediate fry density groups exhibited lower prediction errors, indicating that the annotation refinement stage

contributed positively to improving prediction consistency under moderate-density conditions.

Although slightly higher MAE values were still observed in several dense fry groups due to increased overlap complexity and visual ambiguity, the error distribution did not exhibit extreme deviations. This finding suggests that the proposed regression model maintained stable prediction capability even under challenging fry density conditions. The learning curve for Scenario 2 is presented in Figure 12.

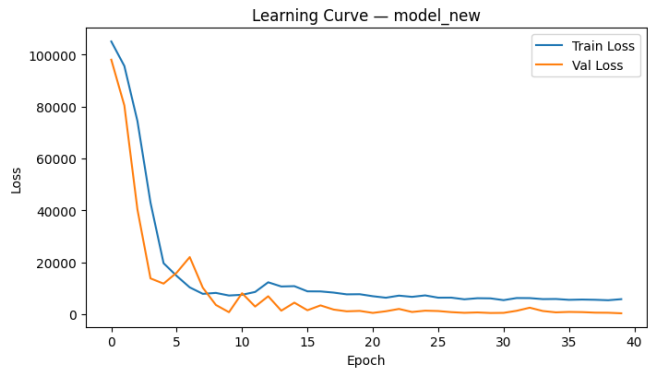


Figure 12. Learning curve of training and validation loss for Scenario 2.

The learning curves show that both training and validation losses decreased significantly during the early training epochs before gradually stabilizing in later epochs. Compared to Scenario 1, the validation curve in Scenario 2 exhibited smaller fluctuations and more stable convergence behavior, indicating improved training consistency after the annotation refinement stage.

The relatively small gap between training and validation loss suggests that the proposed model achieved good generalization capability on previously unseen testing images without severe overfitting. These results demonstrate that improving annotation consistency contributed positively to training stability and prediction reliability in regression-based fry counting tasks.

The results obtained in Scenario 2 indicate that annotation refinement contributed positively to prediction consistency and overall regression performance. To further analyze the impact of different data preparation strategies on regression-based fry count estimation, comparative evaluations were conducted using the original dataset, Scenario 1, and Scenario 2 configurations. The comparison results are presented in Table 4.

Table 4. Comparison Of Model Performance Under Different Data Preparation Strategies						
Data Preparation Strategy	Total Data	MAE	MSE	R ²	RMSE	MAPE
Original Dataset	1128	44.20	2304.46	0.868	48.00	20.66%
Scenario 1 (Count-Level Balancing)	2143	22.19	759.25	0.962	27.55	12.28%

Data Preparation Strategy	Total Data	MAE	MSE	R ²	RMSE	MAPE
Scenario 2 (Count-Level Balancing + Annotation Refinement)	2143	13.25	286.88	0.985	16.93	7.07%

As shown in Table 4, the count-level balancing strategy applied in Scenario 1 substantially improved model performance compared to the original dataset. Increasing training data diversity through balancing, augmentation, and segmentation enabled the regression model to learn more robust visual representations of fry distribution patterns under varying density conditions.

Further performance improvements were observed in Scenario 2 after incorporating the annotation refinement stage prior to preprocessing. Compared to Scenario 1, Scenario 2 achieved lower prediction errors and stronger agreement between predicted and actual fry counts across all evaluation metrics. These improvements indicate that more consistent annotations helped reduce label ambiguity during training and enabled the model to learn more reliable relationships between image features and fry count values.

Overall, Scenario 2 achieved the best performance among all evaluated configurations. The combination of count-level balancing and annotation refinement produced a dataset that was both more visually diverse and more annotation-consistent, enabling the proposed regression model to generate more accurate and stable fry count predictions under varying hatchery conditions.

In addition to evaluating different data preparation strategies, comparative experiments using multiple CNN architectures were also conducted to assess the effectiveness of the proposed EfficientNetV2S backbone for regression-based fry counting tasks. The comparative performance results across different CNN architectures are presented in Table 5.

Table 5. Comparison of Model Performance Across Different CNN Architectures

Model	Total Data	MAE	MSE	R ²	RMSE	MAPE
EfficientNetV2S	2143	13.25	286.88	0.985	16.93	7.07%
EfficientNetB0	2143	17.39	467.48	0.976	21.62	9.86%
ResNet50	2143	16.59	425.48	0.978	20.62	7.85%
MobileNetV2	2143	35.65	2142.45	0.891	46.28	17.89%

As presented in Table 5, EfficientNetV2S achieved the best overall performance across all evaluation metrics. The proposed model obtained the lowest MAE, MSE, RMSE, and MAPE values, while also achieving the highest R² score among all evaluated architectures. These results indicate that EfficientNetV2S was more effective in learning complex fry distribution patterns and estimating fry counts under varying density conditions.

Compared to EfficientNetB0 [20], the EfficientNetV2S architecture demonstrated superior prediction accuracy, suggesting that the improved scaling strategy and architectural optimization of EfficientNetV2S contributed positively to feature extraction performance in dense fry counting tasks. Similarly, although ResNet50 [21] achieved relatively competitive results, its prediction errors remained higher than those produced by EfficientNetV2S.

Meanwhile, MobileNetV2 [22] exhibited substantially lower performance compared to the other evaluated architectures. This result may be associated with the limited feature representation capability of lightweight architectures when handling images containing dense object overlap and complex spatial distributions.

Overall, the comparative analysis demonstrates that EfficientNetV2S provides a favorable balance between feature representation capability and regression performance for automated Nile tilapia fry counting tasks. These findings support the suitability of EfficientNetV2S as the primary backbone architecture for regression-based fry count estimation under practical hatchery conditions.

Collectively, the experimental evaluations validate the core hypothesis of this study: while selecting a highly capable architecture like EfficientNetV2S is important, regression-based counting networks are fundamentally bottlenecked by dataset quality. The substantial jump in performance between Scenario 1 (MAE 22.19) and Scenario 2 (MAE 13.25) isolates the exact value of mitigating annotation ambiguity. By establishing a balanced label distribution and a verified ground truth, the regression model was able to learn stable, reliable mappings between overlapping visual features and fry counts, effectively solving the undercounting issues that plague traditional detection-based methods in high-density aquaculture.

IV. CONCLUSION

This study proposed a regression-based approach for automated Nile tilapia fry counting from digital images using EfficientNetV2S and transfer learning. The proposed system was developed to address the limitations of manual fry counting processes, particularly under dense fry distribution and overlapping object conditions commonly encountered in hatchery environments.

Experimental results demonstrated that the combination of count-level balancing, annotation refinement, augmentation, and segmentation contributed positively to prediction performance and model stability. Among all evaluated configurations, Scenario 2 achieved the best overall performance, obtaining an MAE of 13.25 fry, RMSE of 16.93 fry, R² of 0.985, and MAPE of 7.07%. These results indicate that improving annotation consistency significantly enhanced regression learning performance and reduced prediction errors.

Comparative experiments with multiple CNN architectures further showed that EfficientNetV2S outperformed EfficientNetB0, ResNet50, and MobileNetV2

across all evaluation metrics. This finding demonstrates that EfficientNetV2S provides a favorable balance between feature extraction capability and regression performance for automated fry count estimation tasks.

Overall, the proposed approach demonstrates strong potential for practical implementation in aquaculture hatchery operations by providing a more efficient, objective, and consistent fry counting process. Future work may focus on improving real-time deployment capability and extending the proposed method to other aquaculture species and more diverse environmental conditions.

REFERENCES

- [1] P. A. Indriati and H. Hafiludin, "Manajemen Kualitas Air Pada Pembenihan Ikan Nila (*Oreochromis niloticus*) Di Balai Benih Ikan Teja Timur Pamekasan," *Juvenil: Jurnal Ilmiah Kelautan dan Perikanan*, vol. 3, no. 2, pp. 27–31, Aug. 2022, doi: 10.21107/juvenil.v3i2.15812.
- [2] J. Zhang *et al.*, "Shrimp egg counting with fully convolutional regression network and generative adversarial network," *Aquac. Eng.*, vol. 94, no. June, p. 102175, 2021, doi: 10.1016/j.aquaeng.2021.102175.
- [3] T. H. Khai, S. N. H. S. Abdullah, M. K. Hasan, and A. Tarmizi, "Underwater Fish Detection and Counting Using Mask Regional Convolutional Neural Network," *Water (Switzerland)*, vol. 14, no. 2, pp. 1–23, 2022, doi: 10.3390/w14020222.
- [4] C. S. Costa *et al.*, "Counting tilapia larvae using images captured by smartphones," *Smart Agricultural Technology*, vol. 4, Aug. 2023, doi: 10.1016/j.atech.2022.100160.
- [5] D. O. de Carvalho, L. F. T. Monteiro, F. M. Bazilio, G. T. H. Higa, and H. Pistori, "Livestock Fish Larvae Counting using DETR and YOLO based Deep Networks," Aug. 2024, [Online]. Available: <http://arxiv.org/abs/2408.05032>
- [6] A. Mumuni and F. Mumuni, "Data augmentation with automated machine learning: approaches and performance comparison with classical data augmentation methods," Mar. 2025, doi: 10.1007/s10115-025-02349-x.
- [7] A. S. Dyah Kurniasari, S. Fauziati, R. Hartanto, and A. Akmaluddin, "Nannofossil Image Enhancement for Improved Detection Using Modified Contrast Stretching Method and YOLO-v8," in *2024 16th International Conference on Information Technology and Electrical Engineering (ICITEE)*, IEEE, Oct. 2024, pp. 177–182. doi: 10.1109/ICITEE62483.2024.10808703.
- [8] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2351–2377, Mar. 2022, doi: 10.1007/s10462-021-10066-4.
- [9] Z. Li, X. Zhu, H. Yu, Q. Zhang, and Y. Jiang, "Edge-Aware Monocular Dense Depth Estimation with Morphology," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, Jan. 2021, pp. 2935–2942. doi: 10.1109/ICPR48806.2021.9412578.
- [10] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114, 2019.
- [11] R. Ali Dihin, "Breast Cancer Detection and Diagnosis Using Gabor Features and EfficientNetV2 Model," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 16, no. 4, Dec. 2024, doi: 10.29304/jqcs.2024.16.41791.
- [12] P. Ye *et al.*, "Partial Fine-Tuning: A Successor to Full Fine-Tuning for Vision Transformers," *arXiv preprint arXiv:2312.15681*, 2023. [Online]. Available: <http://arxiv.org/abs/2312.15681>
- [13] S. Eberle, A. Jentzen, A. Riekert, and G. Weiss, "Normalized Gradient Flow Optimization in the Training of ReLU Artificial Neural Networks," *arXiv preprint arXiv:2207.06246*, 2022. [Online]. Available: <http://arxiv.org/abs/2207.06246>
- [14] I. Al-Shourbaji *et al.*, "A Deep Batch Normalized Convolution Approach for Improving COVID-19 Detection from Chest X-ray Images," *Pathogens*, vol. 12, no. 1, Jan. 2023, doi: 10.3390/pathogens12010017.
- [15] Z. Liu, Z. Xu, J. Jin, Z. Shen, and T. Darrell, "Dropout Reduces Underfitting," in *Proc. 40th Int. Conf. Mach. Learn. (ICML)*, *Proc. Mach. Learn. Res.*, vol. 202, pp. 22233–22248, 2023.
- [16] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, Aug. 2023, doi: 10.1007/s00521-023-08568-z.
- [17] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
- [18] R. Hidayat *et al.*, "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket," *SIMKOM*, vol. 10, no. 1, pp. 101–109, Jan. 2025, doi: 10.51717/simkom.v10i1.703.
- [19] F. A. Rangkuti, Khairunnisa, and S. Sundari, "Implementasi Gradient Boosting Machines Untuk Prediksi Harga Rumah Pada Jakarta Selatan," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 2, pp. 164–172, May 2025, doi: 10.69916/jkbt.v4i2.318.
- [20] A. A. Ahmad and N. Aman, "Improving Colorectal Cancer Detection: Leveraging EfficientNet-B0 and Explainable AI for High-Accuracy Classification," in *2025 IEEE 2nd International Conference on Computing, Applications and Systems, COMPAS 2025*,

- Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/COMPAS67506.2025.11381829.
- [21] G. Gao, Z. Sun, G. Mu, H. Yin, and Y. Ren, "Research on marine fish classification and recognition based on an optimized ResNet50 model," *Marine and Coastal Fisheries*, vol. 16, no. 6, e10317, Dec. 2024, doi: 10.1002/mcf2.10317.
- [22] F. A. A. Harahap, A. N. Nafisa, E. N. D. B. Purba, and N. A. Putri, "Implementasi Algoritma Convolutional Neural Network Arsitektur Model Mobilenetv2 Dalam Klasifikasi Penyakit Tumor Otak Glioma, Pituitary Dan Meningioma," *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTika)*, vol. 5, no. 1, pp. 53–61, 2023, doi: 10.29303/jtika.v5i1.234.