IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 December 2023, accepted 15 January 2024, date of publication 13 February 2024, date of current version 22 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365585

RESEARCH ARTICLE

Fry Counting Method in High-Density Culture Based on Image Enhancement Algorithm and Attention Mechanism

HONGYUAN CHEN[1], YUAN CHENG[2,3,4], YU DOU[1], HUACHAO TAN[1], GUIHONG YUAN[1], HAI BI[5], AND DAN LIU[1]

[1]College of Information Engineering, Dalian Ocean University, Dalian 116000, China
[2]Ningbo Institute, Dalian University of Technology, Ningbo 315000, China
[3]Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China
[4]Key Laboratory of Environment Controlled Aquaculture (Dalian Ocean University) Ministry of Education, Dalian 116000, China
[5]Hangzhou Yunxi Smart Vision Technology Company Ltd., Dalian 116000, China

Corresponding author: Dan Liu (liudan@dlou.edu.cn)

This work was supported in part by the Basic Scientific Research Project for Higher Education Institutions of the Education Department of Liaoning Province under Grant LJKMZ20221089, in part by Dalian Program on Key Science and Technology Research and Development Project under Grant 2021YF16SN013, in part by the Key Laboratory of Environment Controlled Aquaculture (Dalian Ocean University) Ministry of Education Project under Grant 202307, in part by the Agricultural Science and Technology Key Research and Development Program of Taizhou under Grant 22NYJB01, and in part by the Key Research and Development Program of Zhejiang Province under Grant 2023C02029.

ABSTRACT It is important in production to achieve accurate counting and density estimation of high-density culture fry under the environmental conditions of aquaculture scenarios in an efficient and accurate manner. However, none of the current methods for fry counting works well under the high-density and high-overlap conditions of real aquaculture scenarios. Therefore, in this paper, we propose a high-density farming fry monitoring network model, Super-Resolution GAN Density Estimate Attention Network (SGDAN), which incorporating an image enhancement algorithm and an attention mechanism, and we create a high-density farming fry dataset (HD-FryDataset) based on the environmental conditions of real aquaculture scenarios. The network model is designed to improve and optimize the targeted subnetworks for several key aspects of high-density fish fry monitoring work. Four subnetworks are included for image optimization, feature extraction, attention, and density map estimation. The experimental results show that the SGDAN network model achieved an average counting accuracy of 97.57% on the high-density culture fry dataset, which was 8.23% and 2.06% higher than those of MCNN and CSRNet, respectively. Additionally, the MAE and RMSE of the model were reduced by 71.9% and 67.3% and by 34.3% and 33.2% compared with those of MCNN and CSRNet, respectively. The model proposed in this paper also has a better ability to generate predictive density maps. The density maps generated by SGDAN have values of the evaluation metrics PSNR and SSIM of 20.33 and 0.933, respectively, which are 3.31 and 0.037 and 2.63 and 0.031 higher than those of MCNN and CSRNet. In general, the network model proposed in this paper outperforms existing network models in two applications: accurate counting of fry and generation of density maps for high-density culture in aquaculture. It also provides a good solution for digitizing the number of fry and visualizing the density of high-density culture in intelligent aquaculture systems.

INDEX TERMS Aquaculture, fry counting, super resolution, attention mechanism, deep learning.

I. INTRODUCTION

Fry counting is the counting of the target number of fry in a given area to aid production decisions [1], [2], [3]. In aquaculture breeding or production programs, counting the number of fry has a considerable cost in terms of the manpower required [4]. At the same time, the monitoring of high-density culture fry is important for the whole aquaculture industry [5]. On the one hand, it can guarantee the quality of fish fry and optimize breeding benefits. In a high-density

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

culture environment, the culture area is limited, and the concentration of nutrients in the water column is high. This can easily cause disease, resulting in pollution of the breeding environment and the spread of disease, thus reducing the benefits of breeding. Therefore, monitoring of high-density fish fry can detect diseases and abnormalities early and enable targeted prevention and control measures to be taken, which can help protect the quality and vitality of fish fry while reducing the risk of breeding and economic losses and improving the efficiency of breeding. On the other hand, it can improve the image of the industry and ensure food safety. Fish fry is an important source and basis of aquatic products, and the quality of fish fry is related to the food safety of aquatic products. Scientific and intelligent monitoring of high-density farmed fish fry can enhance consumer confidence and acceptance of fish products and ensure that they meet food safety standards. While ensuring consumer food safety, these measures promote the sustainable development of the aquaculture industry.

However, aquaculture counts face many challenges. On the one hand, there are the challenges posed by the objects themselves, such as the transparency of the objects, the differences in the shapes and sizes of the objects, and the problems of overlap caused by motion. On the other hand, difficulties arise from the complexity of the background environment, such as interference problems, current disturbances, and the complexity of the underwater environment [6]. Therefore, it is crucial to solve the problem of accurate counting and density estimation of high-density culture fry under the environmental conditions of real aquaculture scenarios in an efficient and accurate way.

To address and solve these challenges and problems, aquaculture-related researchers and producers are seeking innovative approaches to increasing the efficiency of aquaculture production and improving the quality and survival conditions of cultured fry [7]. The traditional methods for counting artificial fish are weighing and statistical averaging of distribution sampling [8]. The former can cause stress and physical damage to fry, and tested fry take longer to resume normal feeding growth. The latter is time-consuming, laborious, and vulnerable to human subjective factors.

With the rapid development of computer vision technology and deep learning-related research, there is an urgent need to scientifically guide aquaculture and production with the help of advanced technology. Fish fry counting methods using techniques such as visual image processing and machine learning combined with deep learning have gradually attracted the attention of scholars around the world [9], [10], [11], [12], [13], [14], [15]. Fish counting methods based on computer vision [16], [17] have the advantages of high efficiency, ease of operation and accuracy compared to traditional methods. Chen et al. [18] segmented and noise-reduced fry images and used a recursive-based connected component labeling algorithm to obtain the connected regions in binary images. The number of fry in a connected area is calculated by establishing two identification rules related to the average fry area and the size of the area. Shuo et al. [19] proposed a computer vision-based image processing method to solve the fry image adhesion problem and to accomplish accurate counting of fry. Yang et al. [20] used MATLAB software for grayscale, noise reduction and morphological processing of images preprocessed in Photoshop and used the connectivity map counting method and area counting method to count fry. Guo et al. [11] proposed a fish fry counting algorithm based on machine vision tracking.

In addition, related scholars have constructed counting models targeting the estimation of density maps for counting fish populations in aquaculture using multicolumn convolutional neural networks and deeply expanded convolutional neural networks. The results showed that the average counting accuracy could reach 95.06% [14] for different fish densities. The fish counting method based on density map estimation can map the input image into a corresponding density map, and the value of each pixel in the density map can represent the number of targets at that location [21]. The total number of targets in the image is obtained by integrating the density map. In addition to the quantity information, the density map contains location information, which can show the spatial distribution of the fish population and is an important guide for practical production activities. For example, if the density of fish in one small area is much higher than that in another area, it indicates that there may be an anomaly in that area [22].

All of the fish counting methods mentioned above can achieve good accuracy and expected results under the conditions of low density and low overlap application scenarios. However, for actual aquaculture fry, there are random and uneven characteristics of the fry distribution as well as high density and high overlap, and the above methods have difficulty achieving the expected counting accuracy and effect.

Therefore, to better solve the problem of accurate counting and density estimation for high-density culture fry under the environmental conditions of actual aquaculture scenarios, this study proposes a high-density culture fry monitoring network model (SGDAN) based on multiple modules and attention mechanisms based on the creation of a real high-density culture fry dataset (HD-FryDataset). The network model consists of four subnetworks for image optimization, feature extraction, attention, and density map estimation. Among them, the image optimization network consists of the restricted contrast adaptive equalization histogram algorithm (CLAHE) and the Super-Resolution Generative Adversarial Network [23] (SRGAN) for contrast enhancement, noise reduction of the original image and sharpness enhancement of the original image of high-density farmed fish fry fed into the network model. The feature extraction network consists of a multicolumn convolutional neural network (CNN) for extracting the overall feature map of high-density culture fry images. The attention network is based on an attention mechanism that embeds location information into channel attention for more accurate identification and extraction of key information in the overall feature map of high-density culture fry during dense counting and density map generation. The density map estimation network is used for the generation of visual, high-quality density maps for predictive counts and the representation of fry distribution and aggregation levels. The purpose of this study is to develop a high-potential and efficient monitoring model (including the digitalization of quantity and visualization of density) for high-density culture fry and to provide reliable theoretical support for the development of intelligent aquaculture systems.

The network model proposed in this study achieves the highest counting accuracy and the optimal density map quality when oriented to a more realistic working situation in the aquaculture industry, and when compared with various performance parameter indexes of other current classical and mainstream density estimation networks.

II. MATERIALS AND METHODS

A. DATASET

1) DATASET ACQUISITION AND ANNOTATION

The image data used in this experiment were obtained from a real aquaculture environment in a farming plant in Dalian, Liaoning Province, People's Republic of China. The fish were black fish fry, the depth of the water basin was 21 cm, the water depth was 6-7 cm, the diameter of the basin mouth was 60 cm, the diameter of the basin bottom was 50 cm, and the total number of fry was approximately 570. The video of the blackfish fry activity was shot from top to bottom in a vertical view with a resolution of 1920 × 1080 and a frame rate of 25 fps. The experimental data were collected under conditions that did not affect the normal growth and activity of blackfish fry and did not involve animal ethics issues. The images were selected every 25 frames to make up the dataset, and the final image size was 1920 × 1080. Retaining all full-size information enables the future deployment of the network model on mobile terminals for real application scenario environments. The dataset used in the experiment contains 192 annotated high-density blackfish fry images and consists of two parts for training and testing the model, with a ratio of 3:1 between the training and testing sets. All images use the same annotation method and density map generation method, which involve approximately 109,000 accurately annotated targets. Datasets used in the same type of study have approximately 65% of the total number of labeled targets and 20% of the number of labeled targets in a single image, while the total number of images is four times higher than the size of the dataset of this paper [24]. This is good evidence that the dataset created and used in this paper is more suitable for real aquaculture scenarios and has more similar characteristics to the high-density characteristics of fish fry in such scenarios.

Considering the actual culture conditions of fish fry crowding, this study draws on the idea of crowd density estimation in crowding scenarios [25], [26], [27] and marker methods from other similar studies [14]. In this study, for high-density, heavily obscured fry populations, the locations of the targets and the number of fish in the image were determined by marking the head of each fish. This can address the problem that most fry are mislabeled or missed due to insufficiently exposed parts caused by severe occlusion. The image annotation and labeling results are shown in Fig.1 and Fig.2. Fig.1(a) indicates the case of no occlusion, Fig.1(b) indicates the case of general occlusion, and Fig.1(c) indicates the case of severe occlusion.

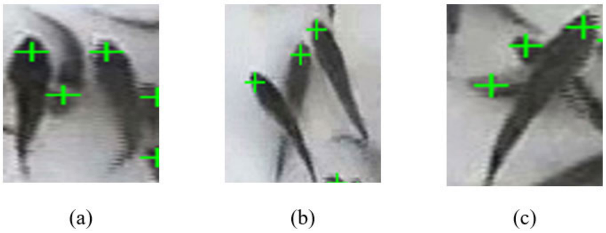


FIGURE 1. Image annotation: (a) No occlusion (b) General occlusion (c) Severe occlusion.

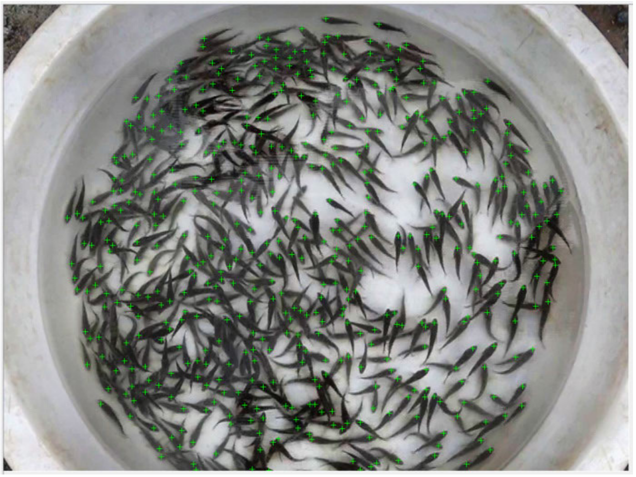
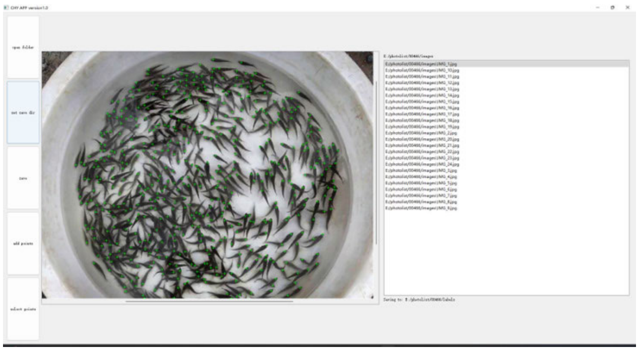


FIGURE 2. Results of image annotation.

Additionally, to better complete the annotation of the dataset, this study uses a self-designed annotation program, CHYAPP version 1.0, for point annotation of density estimation sample images. The generated label files are in .mat format and can provide accurate quantity and location information. The labeling program supports manual saving, loading existing labels, deleting incorrect labels and other functions. The interface of the marker program is shown in Fig.3.

2) DENSITY MAP GENERATION

Many algorithms for counting have been proposed in the literature. However, in detection-based target counting methods, it is often assumed that the target consists of a single entity that can be detected by some given detector [28], [29], [30], [31]. The limitation of these detection-based methods is that in clustered environments or very dense target groups, target-to-target occlusion can severely affect the performance of the detector and hence the final estimation accuracy. Therefore,



E:/photolist/00466/images
E:/photolist/00466/images/IMG_1.jpg
E:/photolist/00466/images/IMG_10.jpg
E:/photolist/00466/images/IMG_11.jpg
E:/photolist/00466/images/IMG_12.jpg
E:/photolist/00466/images/IMG_13.jpg
E:/photolist/00466/images/IMG_14.jpg
E:/photolist/00466/images/IMG_15.jpg
E:/photolist/00466/images/IMG_16.jpg
E:/photolist/00466/images/IMG_17.jpg
E:/photolist/00466/images/IMG_18.jpg
E:/photolist/00466/images/IMG_19.jpg
E:/photolist/00466/images/IMG_2.jpg
E:/photolist/00466/images/IMG_20.jpg
E:/photolist/00466/images/IMG_21.jpg
E:/photolist/00466/images/IMG_22.jpg
E:/photolist/00466/images/IMG_23.jpg
E:/photolist/00466/images/IMG_24.jpg
E:/photolist/00466/images/IMG_3.jpg
E:/photolist/00466/images/IMG_4.jpg
E:/photolist/00466/images/IMG_5.jpg
E:/photolist/00466/images/IMG_6.jpg
E:/photolist/00466/images/IMG_7.jpg
E:/photolist/00466/images/IMG_8.jpg
E:/photolist/00466/images/IMG_9.jpg
Saving to: E:/photolist/00466/labels

FIGURE 3. **Example diagram of the marking program operation interface.**

we consider a density map regression approach in this paper to achieve accurate counting of high-density cultured fish fry and subsequent studies. The method of density map regression is based on the known location of each target, and then estimate the size of the target where the location is located, so that the coverage area of the target can be obtained, through some kind of density map generation strategy, the area is transformed into the probability that the area may be a target in the region, and the probability of the region sums up to 1 (or indicates how many targets there may be in each pixel), and ultimately, a density map of the target can be obtained, and by integrating and summing the density map, the number of targets in the region can be obtained by summing the integrals of the map.

Simply put, a Gaussian kernel is used to simulate the head of a target object at the corresponding position in the original image. After completing such operations for all the corresponding positions in the image, the matrix composed of all these Gaussian kernels is normalized [22]. Typically, there are three strategies for generating density maps. The first is a fixed-size density map [32], the second is a perspective density map [33], and the third is a KNN density map [22]. Among them, the third method is suitable for very crowded scenarios, so this paper adopts this strategy to generate density maps, and the specific implementation process is as follows.

If there is a fish at pixel x_i, suppose there is a delta function $\delta\left(x - x_i\right)$. Then, an image with N fish markers can be represented by Equation (1).

$$H\left(x\right) = \sum\nolimits_{i=1}^{N} \delta\left(x - x_i\right) \tag{1}$$

For each head x_i in a given image, denote the distances to its k nearest neighbors as $\left\{d_1^i, d_2^i, \ldots, d_m^i\right\}$. Therefore, the average distance is $\bar{\boldsymbol{d}}^{\mathbf{i}} = \frac{1}{m}\sum\nolimits_{j=1}^{m} d_j^i$. The pixel associated with x_i corresponds to a region on the image in the scene whose radius is roughly proportional to $\bar{\boldsymbol{d}}^{\mathbf{i}}$. Therefore, to estimate the density of fry around pixel x_i, a Gaussian kernel with variance δ_i proportional to $\bar{\boldsymbol{d}}^{i}$ is needed to convolve $\delta\left(x - x_i\right)$, and more precisely, the density F is shown in Equation (2) below.

$$F\left(x\right) = \sum\nolimits_{i=1}^{N} \delta\left(x - x_i\right) \times G_{\sigma_i}\left(x\right), \sigma_i = \beta \bar{\boldsymbol{d}}^{\mathbf{i}} \tag{2}$$

where $\beta = 0.3$ is the adaptive parameter. The results of the density map generated by the above process are shown in Fig.4. Fig. 4(a) shows the original image of the sample set, and Fig.4(b) shows the corresponding generated real density map.

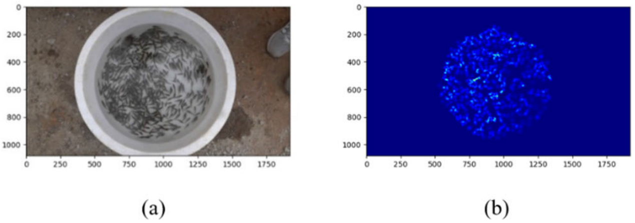


FIGURE 4. **(a) Original image (b) Density map.**

B. NETWORK DESIGN

In this study, a high-density culture fry monitoring network model (SGDAN) incorporating an image enhancement algorithm and attention mechanism is proposed to monitor high-density culture fry (including the digitization of quantity and visualization of density) in aquaculture environments. The model consists of an image optimization module, a feature extraction module, an attention module, and a density map estimation module, which correspond to several key aspects of high-density fish fry monitoring work and solve the corresponding problems. In the whole SGDAN network model, the original image will first be sent to the image optimization module to improve the contrast between the target fry and the original background through CLAHE, and the local features and edge details of the image will be continuously improved through SRGAN. These optimized raw images are then converted into pixel matrices that are used as inputs to the multicolumn convolutional neural network model that follows. The multilinear convolutional neural network model consists of three parallel, identically structured convolutional neural networks, each with a different convolutional kernel size and each with an attention mechanism added at the end of the output to enhance the ability to pay attention to the relevant portion of the training result of the input data. During training, the backbone network will process both the original image and the label density map. The input image is passed through the image processing branch and the generated features are used in the density map branch and the image branch. The output of the convolutional neural network is then compared to the corresponding ground truth density map, and the loss is passed through backpropagation to update the weights of the network. When the training is complete, the trained network is used for testing and prediction. As a result, the input raw images can generate predicted density maps, and by subsequent processing (e.g., integral summation) of these generated predicted density maps on the density map estimation module it is possible to obtain information on the location of the fry and the predicted number of fry. The network structure diagram of SGDAN is shown in

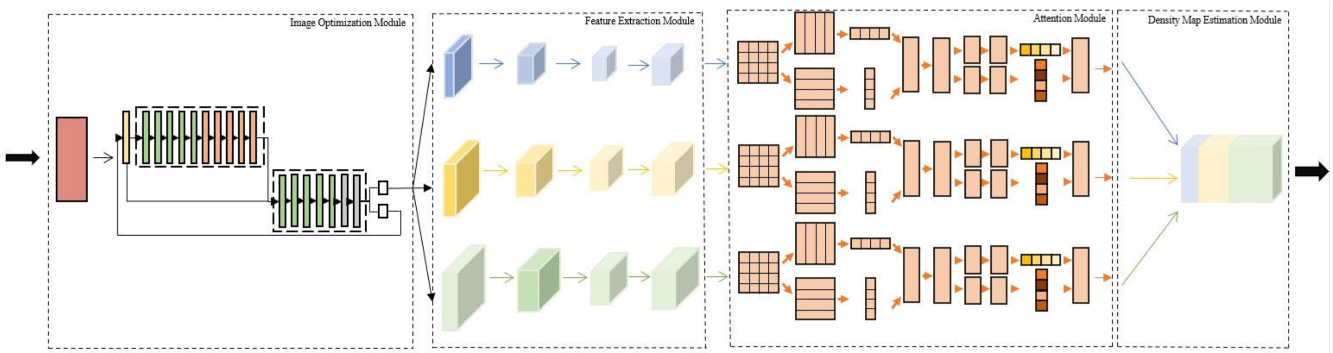


Image Optimization Module
Feature Extraction Module
Attention Module
Density Map Estimation Module

FIGURE 5. The network structure diagram of SGDAN.

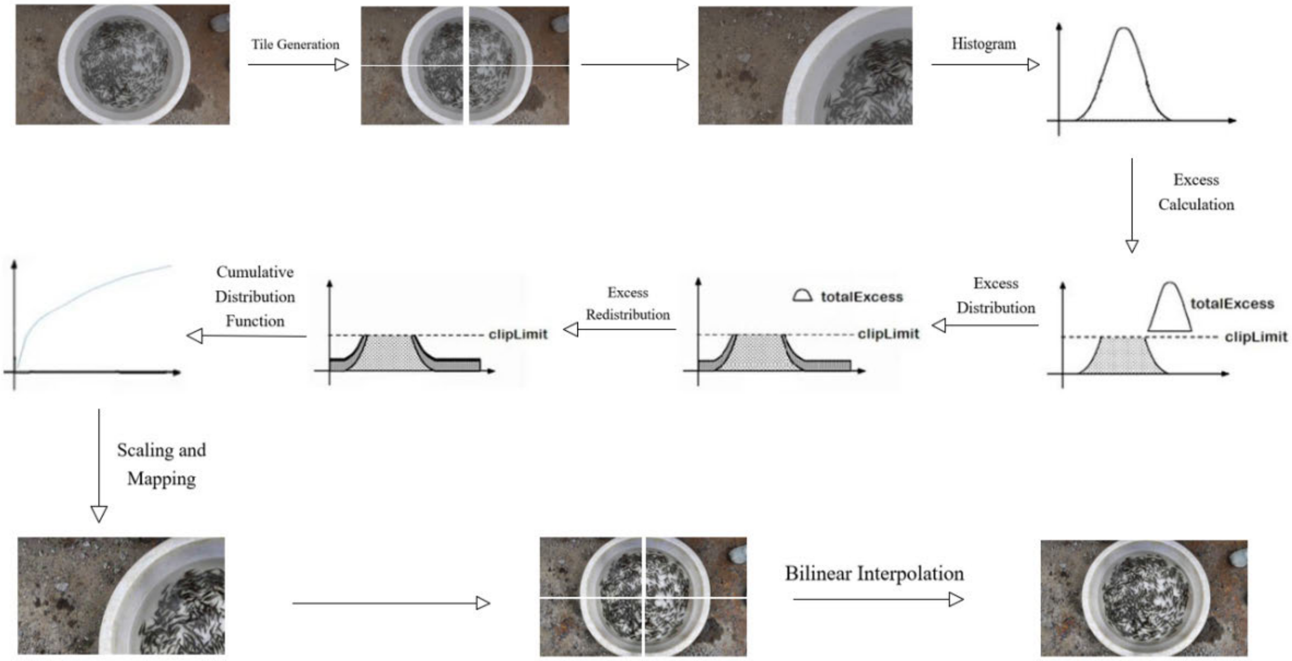


FIGURE 6. The flow diagram of the CLAHE algorithm.

Fig.5. Next, the compositions and the principles of action of the four subnetworks are described in detail.

1) IMAGE OPTIMIZATION MODULE

Light is easily weakened when propagating underwater, thus causing color distortion in the original image of high-density fry in the captured video, and at the same time, the number of fry targets in a single full-size original image of high-density fry causes low clarity of individual fry targets and problems such as rough contours and blurred edges. To address these issues, an image optimization network is designed in the SGDAN network model, which consists of CLAHE and SRGAN.

CLAHE optimizes the contrast of the original image of high-density cultured fry while suppressing the overamplification of noise. The specific implementation process of the CLAHE algorithm is as follows: first, the original fry image is preprocessed, such as by image chunk filling, and then the mapping relationship of each chunk is calculated. The mapping relationship is calculated for the difference between the color of the blackfish fry itself and the white breeding water basin by using a contrast limit. Finally, the optimized image is obtained using the interpolation method. The algorithm flowchart is shown in Fig.6.

SRGAN enhances the resolution of the original image of high-density cultured fry and the edge details of individual cultured fry after image enlargement. The specific implementation of the SRGAN network is as follows: the original images of high-density fish fry are used to train the generative network and the discriminative network of the SRGAN network alternately. In each round of training, the discriminative network determines whether the high-resolution images

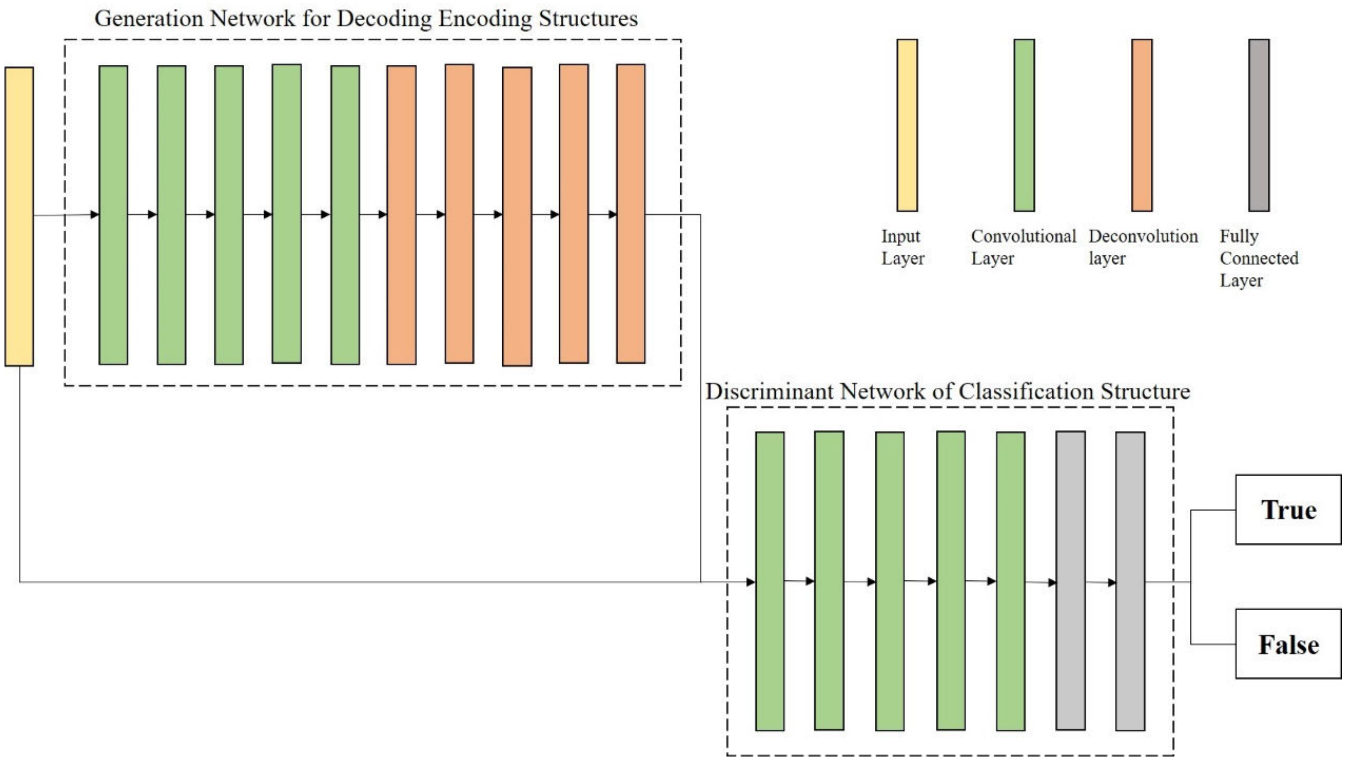

Generation Network for Decoding Encoding Structures
Input Layer
Convolutional Layer
Deconvolution layer
Fully Connected Layer
Discriminant Network of Classification Structure
True
False

FIGURE 7. The network structure diagram of SRGAN.

generated by the generative network based on the original resolution of the high-density farmed fry images are accurate. If the discriminative network thinks that the generated high-resolution image is not realistic enough compared with the original image, the generative network will continue to adjust the feature information in the original image until a more realistic high-resolution image is generated to "fool" the discriminative network. The network structure diagram is shown in Fig.7.

2) FEATURE EXTRACTION MODULE

Due to the high overlap of high-density culture fry, there will be a problem of inconsistent distances between fry in the same culture water (the same culture water basin) because of the camera lens or the inconsistent sizes of the exposed parts of the fry, with the result that some fry in the original image can be clearly identified and some need to be carefully observed. To address this, a feature extraction network is designed for the study, consisting of a three-column CNN. The network structure diagram of the feature extraction network is shown in Fig.8. Each column of the CNN has a filter with a different-size local perceptual field of view (convolutional kernel), which will have different effects for fry of different scales (distance and size).

Moreover, an ELU is used as the activation function of the network model. This makes the normal gradient of the network model closer to the unit natural gradient and increases its robustness to noise. The details of the feature extraction network are shown in Table 1.

3) ATTENTION MODULE

Related mobile network design studies have demonstrated the significant effectiveness of attention mechanisms in improving model performance, and thus, suitable attention mechanisms are added to the research model in this paper. The input of the attention network corresponds to the outputs of the three parallel networks of the feature extraction network, and the output terminal is connected in parallel with the attention mechanism to form the attention network. More precise identification of the key information of images during dense counting is achieved through attention networks. The flow chart of the attentional network is shown in Fig.9. Attentional networks decompose traditional channel attention into two one-dimensional feature encoding processes that aggregate features along two spatial directions. In this way, remote dependencies can be captured along one spatial direction, while precise location information can be retained along the other spatial direction. The generated feature maps are then encoded into a pair of direction-aware and position-sensitive attention maps, which can be complementarily applied to the input feature maps to enhance the representation of high-density culture fry.

In this case, the attention network contains a nonlinear batch-normalized residual block, a convolutional layer and

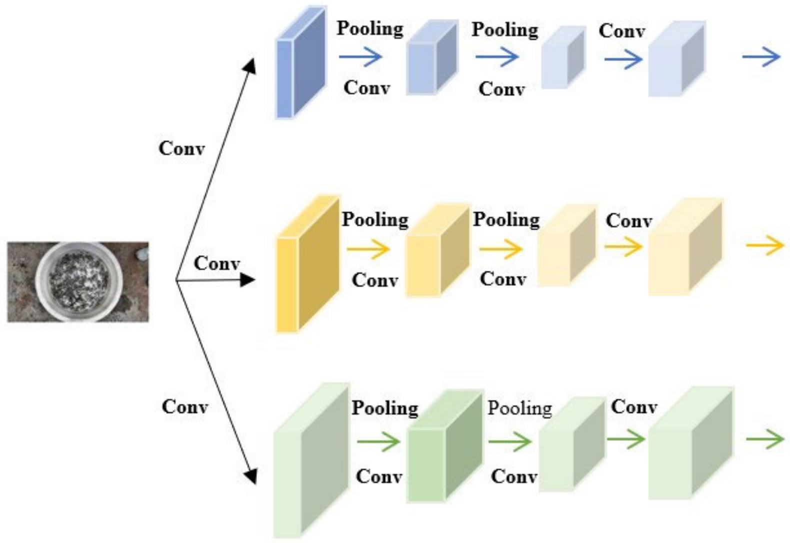


FIGURE 8. The structure diagram of the feature extraction network.

TABLE 1. Details of the feature extraction network.

Module	Layer 1	Layer 2	Layer 3
Feature extraction module	Conv	Conv	Conv
	ELU	ELU	ELU
	Maxpooling	Maxpooling	Maxpooling
	Conv	Conv	Conv
	ELU	ELU	ELU
	Maxpooling	Maxpooling	Maxpooling
	Conv	Conv	Conv
	ELU	ELU	ELU
	Conv	Conv	Conv
	ELU	ELU	ELU

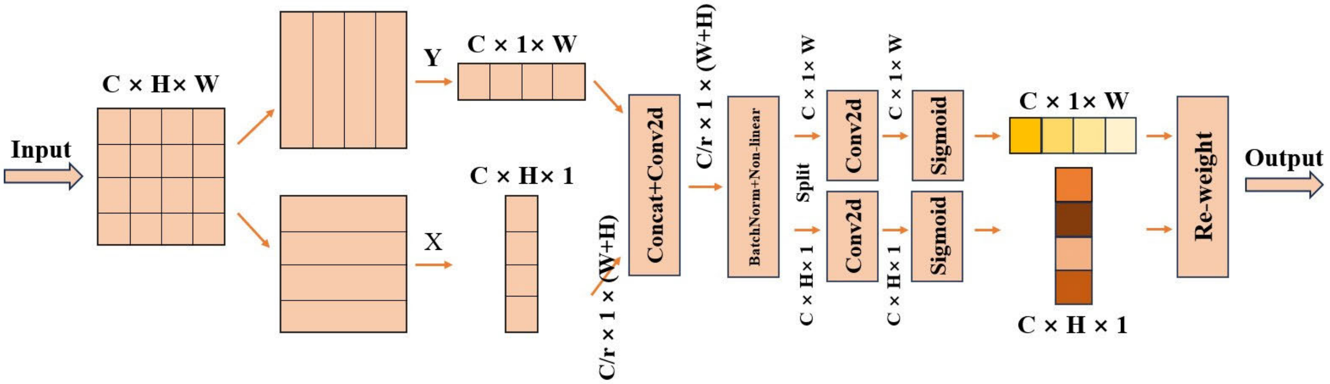


FIGURE 9. The flow diagram of the attention module.

two parallel separate convolutional layers; the network details are shown in Table 2.

The output Y of the attentional network is shown in Equation (3).

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{3}$$

Unlike channel attention, which focuses only on reweighing the importances of different channels, the attention mechanism incorporated in the research model of this paper considers the encoding of spatial information [34]. As shown in Fig.9, attention along the horizontal and vertical directions is applied to the input tensor simultaneously.

TABLE 2. Details of the attention network.

Module	Channel H	Channel W
Attention network	Residual	
	AdaptiveAvgPool	AdaptiveAvgPool
	Concat	
	Conv	
	ReLU	
	BatchNorm (Nonlinear)	
	Split	
	Conv	Conv
	Sigmoid	Sigmoid
	Product	

4) DENSITY MAP ESTIMATION MODULE

The density map estimation network is used to predict counts and obtain an intuitive, high-quality density map representation of the distribution and aggregation of cultured fry in an image. To map stacked feature maps to density maps, a filter of size 1 × 1 is used for convolution operations. The output of the CNN network trained with the attention mechanism added to each column is mapped through a 1 × 1 filter to generate the corresponding two-dimensional density matrix, and the number of fish fry input to the original image can be obtained by performing an integral summation operation on this two-dimensional density matrix, and the corresponding density map image is generated. Three columns of CNN networks containing attention mechanisms and different sized convolutional kernels respectively are finally fused together through concatenation. The network structure diagram of the density map estimation network is shown in Fig.10, and the network details are shown in Table 3.

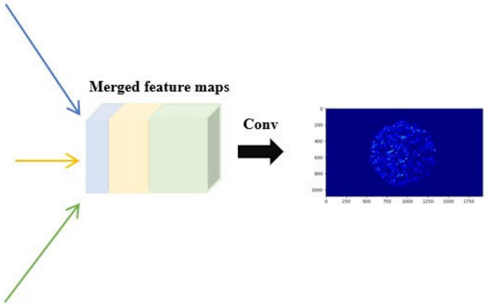


FIGURE 10. The flow diagram of the density map estimation network.

TABLE 3. Details of the density map estimation network.

Module	Layer1	Layer2	Layer3
Density map estimation network	Conv		

C. LOSS FUNCTION

To allow the SGDAN network model to reduce the error between the density map and ground truth at the stage of generating the predicted density map, the density image is processed to present better local details. In this study, the smooth L1 loss is used to train the model instead of the L2 parametric error (MSE loss) and L1 parametric error (MAE loss), which are commonly used in traditional regression problems.

The smooth L1 loss is defined as shown in Equation (4).

$$SL1\left(x\right)=\begin{cases}0.5\left(\delta x\right)^{2} & |x|<\dfrac{1}{\delta^{2}}\\ |x|-\dfrac{0.5}{\delta^{2}} & otherwise\end{cases} \tag{4}$$

The SmoothL1Loss loss function is actually a segmentation function that is smooth at $[-1,1]$, which solves the unsmoothing problem of MAE. The problem of possible outlier points gradient explosion due to MSE is solved in the interval $[-\infty,1)(1,+\infty]$. The functional plots of the L1 parametric loss function (MAE loss), L2 parametric loss function (MSE loss) and smooth L1 loss are shown in Fig. 11.

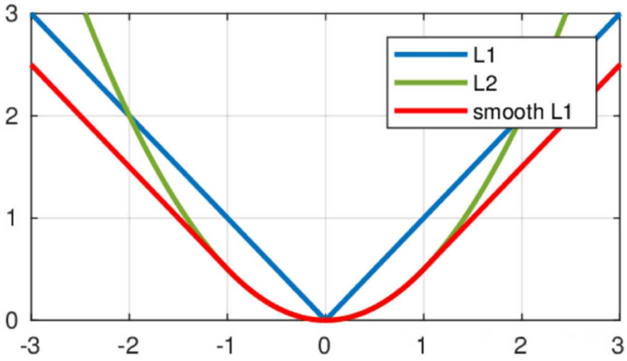


FIGURE 11. A graph of three functions.

D. EVALUATION INDICATORS

In this study, MAE, RMSE and mean accuracy are used to evaluate the performance of the network model, and the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and visual evaluation are used to evaluate the performance of the network model in generating density maps.

MAE and RMSE are both used to measure the performance of network models, and both PSNR and SSIM are widely used to objectively evaluate image quality. MAE characterizes the accuracy of network model estimation, RMSE characterizes the stability of network model estimation, and mean accuracy characterizes the average accuracy of network models for fry counts in real aquaculture environments. PSNR characterizes the error between the generated density map and ground truth based on the error between the corresponding pixel points [35], and SSIM characterizes the similarity between the generated density map and ground truth in terms of luminance, contrast, and structure [36]. The above five evaluation indicators are defined as shown in Equations (5), (6), (7), (8), and (9).

$$MAE = \frac{1}{N}\sum\nolimits_{i=1}^{N}\left|z_i - \hat{z}_i\right| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N}\left(z_i - \hat{z}_i\right)^2} \tag{6}$$

$$MeanAccuracy = \left(1 - \frac{\sum_1^N \left|z_i - \hat{z}_i\right|}{\sum_1^N z_i}\right) \times 100\% \tag{7}$$

$$PSNR = 10 \times log_{10}\frac{(2^n - 1)^2}{MSE} \tag{8}$$

$$SSIM\,(x, y) = l\,(x, y)^{\alpha} \times c\,(x, y)^{\beta} \times s\,(x, y)^{\gamma} \tag{9}$$

where N denotes the number of images in the test set, z_i denotes the actual number of fry contained in the i-th image, and $\hat{z}_i$ denotes the number of fry contained in the images estimated by the algorithm. n is the number of bits per sampled value, and MSE is the mean square error between the original image and the image being processed. x and y distinguish the two images to be compared, and $l\,(x, y)$, $c\,(x, y)$ and $s\,(x, y)$ are the luminance similarity, contrast similarity and structural similarity, respectively. α, β, and γ are the weighting coefficients, which are generally taken as 1. Smaller values of MAE and RMSE and larger values of mean accuracy indicate better counting performance of the network model. The larger the value of PSNR is, the closer the value of SSIM is to 1, indicating the higher quality of the density map generated by the network model.

Since the visual characteristics of the human eye are not considered, the evaluation results are often inconsistent with the subjective perceptions of people. The final density map is generated to help and guide producers or researchers to better understand the real spatial distribution of fry in high-density culture. Therefore, visual evaluation was also introduced in this study to compare the density maps generated by the network model and the real density maps based on the subjective perception of the human eye.

III. RESULTS AND DISCUSSION

The experimental environment is a computer equipped with a Windows 11 operating system. The CPU is an Intel Core i5-12600KF, whose main frequency is 3.7 GHz. The GPU is an NVIDIA GeForce RTX3060 with 12 GB of video memory. The experimental platform is PyCharm (version 2022), and the deep learning framework used is PyTorch. The parameter settings for the training process are shown in Table 4.

A. THE PERFORMANCE IF IMAGE OPTIMIZATION NETWORKS

1) THE PERFORMANCE OF THE CLAHE ALGORITHM

A comparison of the color degrees of images before and after processing by the CLAHE algorithm is shown in Fig.12 and Fig.13.

As shown in Fig.12 and Fig.13, the image processed by the CLAHE algorithm is significantly different in terms of color level compared to the original image. The original image was processed by the CLAHE algorithm to reduce the reflection and refraction effect of the water surface of the breeding water basin and enhance the color contrast between the black fish fry and the white breeding water basin.

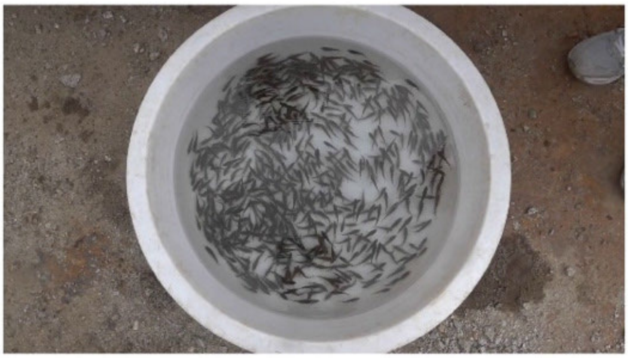


FIGURE 12. Image before CLAHE.

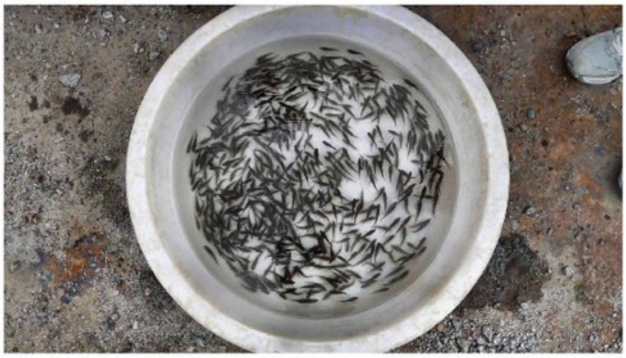
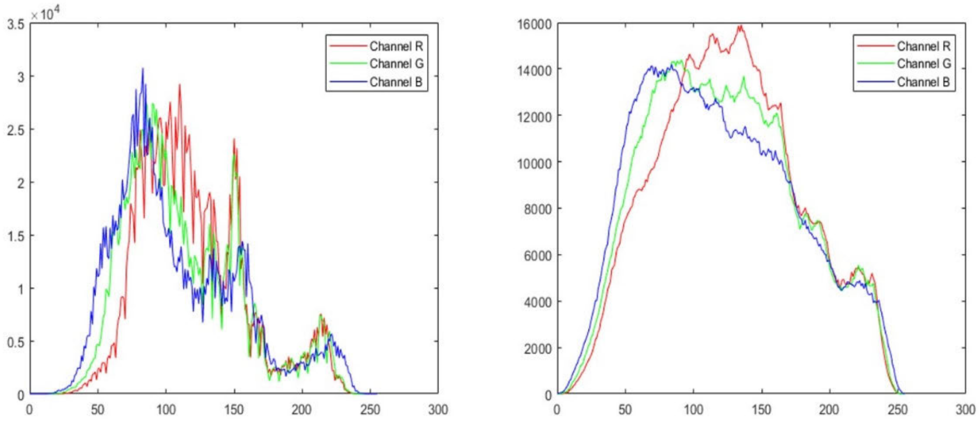


FIGURE 13. CLAHE image.

Additionally, as shown in Fig.14, Fig.14(a) represents the three-channel histogram of the image before CLAHE processing, and Fig.14(b) represents the three-channel histogram of the image after CLAHE processing. The CLAHE algorithm not only enhances the color level of the original image but also suppresses the overamplification of the original image noise. The CLAHE processed image has smoother fold lines for red, green, and blue in the RGB three-channel histogram compared to the preprocessing image.

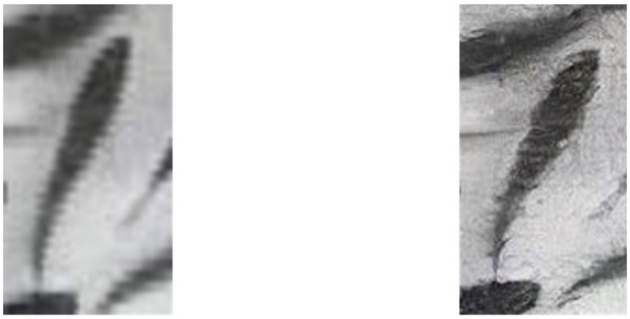
TABLE 4. The parameter settings in the training process.

Parameter name	
optimization algorithm	SGD
initialization of weights	Gaussian initialization with 0.01 standard deviation
epoch	2000
learning rate	1e-6
loss function	Smooth L1
batch size	1 (online learning)



(a) (b)

FIGURE 14. (a) Three-channel histogram of the image before CLAHE. (b)Three-channel histogram of the CLAHE image.



(a) (b)

FIGURE 15. (a) Single fry in the original image. (b) Single fry in the super resolution image.

2) THE PERFORMANCE OF THE SRGAN NETWORK

A comparison of images before and after super resolution processing by the SRGAN network is shown in Fig.15. Fig.15(a) shows a single fry in the original image, and Fig.15(b) shows a single fry in the image after super resolution network processing.

Additionally, to verify the superiority of SRGAN networks over other classical super resolution networks in terms of their performance in image optimization, this study also compares SRGAN networks with Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network [37] (ESPCN) in a cross-sectional manner. The comparison of the two super resolution networks for processed images is shown in Fig.16. Fig.16(a) shows a single fry in the image after SRGAN network processing, and Fig.16(b) shows a single fry in the image after ESPCN network processing.

Fig.15 and Fig.16 show that the effect of the SRGAN network is more realistic in terms of image resolution and the edge details of individual blackfish fry by comparing images before and after processing by the SRGAN network and images after processing by different super resolution networks.

Second, the super resolution network of the SGDAN network model proposed in this paper is changed from SRGAN to ESPCN, and the same dataset is used for a training comparison under the same training parameters. The results of the training are shown in Table 5.

It is evident from Table 5 that the MAE and RMSE of the overall network model decreased by 17.6% and 10.1%, respectively, and the average counting accuracy improved by 0.53% when using the SRGAN network as the super resolution network for the image optimization network compared to those when using the ESPCN network. Moreover, the quality of the generated density maps was 8% and 1.7% higher in terms of PSNR and SSIM, respectively.

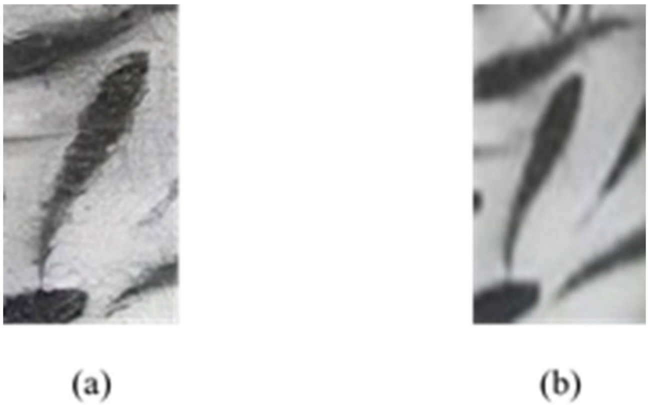


FIGURE 16. (a) Single fry from an SRGAN image. (b) Single fry from an ESPCN image.

In summary, the image optimization network designed here can solve common problems for datasets of high-density culture fry samples, and it has good feasibility and applicability.

3) THE IMPROVEMENT OF DENSITY MAP GENERATION QUALITY BY SUPER RESOLUTION NETWORKS

In this paper, we found through research and experiments that adding super resolution networks can improve the ability of network models of the density estimation type in the prediction phase of generating density maps.

The test network model Density Estimate Attention Network (DAN) is formed by removing the Super-Resolution GAN network from the image optimization module in the SGDAN network model proposed in this paper. It was trained with SGDAN using the same dataset and the same parameter conditions to compare the differences in density maps generated by the final predictions of the two models. The experimental results are shown in Fig.17 and Table 6. Fig.17(a) represents the ground truth of the original image, Fig.17(b) represents the density map generated by the DAN network model, and Fig.17(c) represents the density map generated by the SGDAN network model.

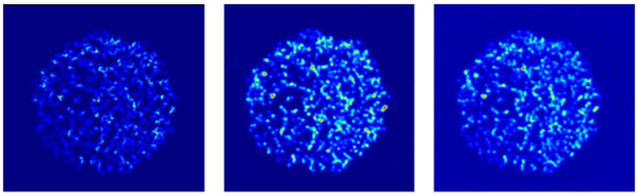

(a) Ground truth
(b) DAN
(c) SGDAN

FIGURE 17. Comparison of generated density maps.

From the actual performance in Fig.17 and the evaluation metrics in Table 6, it is obvious that the density map generated by the network model with the addition of the super resolution network is of higher quality, and the distribution and aggregation reflected by the density map are closer to those of the real density map.

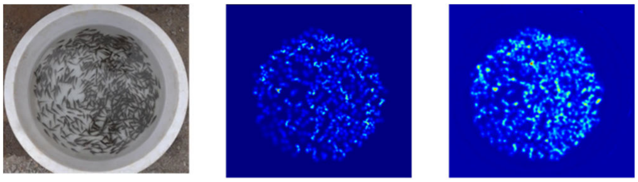

(a) Original image
(b) Ground truth 564
(c) Estimated value 570

FIGURE 18. The counting results and density map of fish.

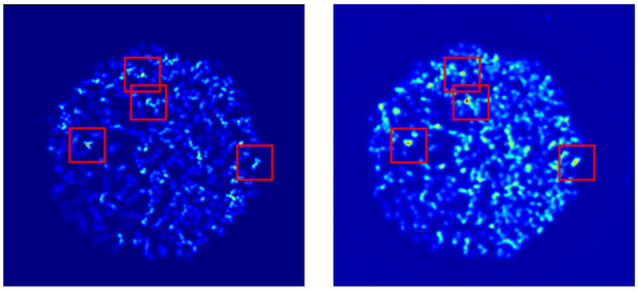

(a) Ground truth
(b) SGDAN

FIGURE 19. Comparison of key areas in density maps.

B. COUNTING RESULTS AND PERFORMANCE EVALUATION OF THE SGDAN MODEL NETWORK

The density map is visualized by color mapping, as shown in Fig.17. When a region has high fish density, the color of that region in the density map is more red. Conversely, if the density of fish is low, the color of the area is bluer. Fig.18 also shows the density maps obtained by the SGDAN network model and the corresponding count results, and it can be seen that the density maps generated by the predictions of the SGDAN network model are highly similar in terms of fry distribution and density compared to the ground truth. Fig.18(a) represents the original image, Fig.18(b) represents the ground truth of the original image and the true number of fry, and Fig.18(c) represents the density map generated by the SGDAN network and the predicted number of fry. As shown in Fig.19, several key regions in the ground truth that indicate the concentration of fry density (the parts marked by red boxes in Fig.19) are present in the density maps generated by the SGDAN network. Fig.19(a) represents the ground truth of the original image, and Fig.19(b) represents the density map generated by the SGDAN network model.

Table 7 and Fig.20 show the performance of SGDAN for different performance evaluation metrics and the line graph of counting accuracy for each high-density culture fry image in the test set, respectively. For the images in test_data, the MAE of the SGDAN network model proposed in this paper is 13.82, and the RMSE is 17.67. The predicted counting accuracy was above 93% for all tested images. Additionally, the predicted generated density map performed well in the image quality evaluation index, with a PSNR of 22.33 and SSIM of 0.933. Combining the above indices, it can be concluded that the SGDAN network model has good stability,

TABLE 5. Results of training evaluation indicators for different super resolution networks.

Super resolution network	MAE	RMSE	Mean Accuracy	PSNR	SSIM
ESPCN	16.78	19.66	97.04%	20.67	0.917
SRGAN	**13.82**	**17.67**	**97.57%**	**22.33**	**0.933**

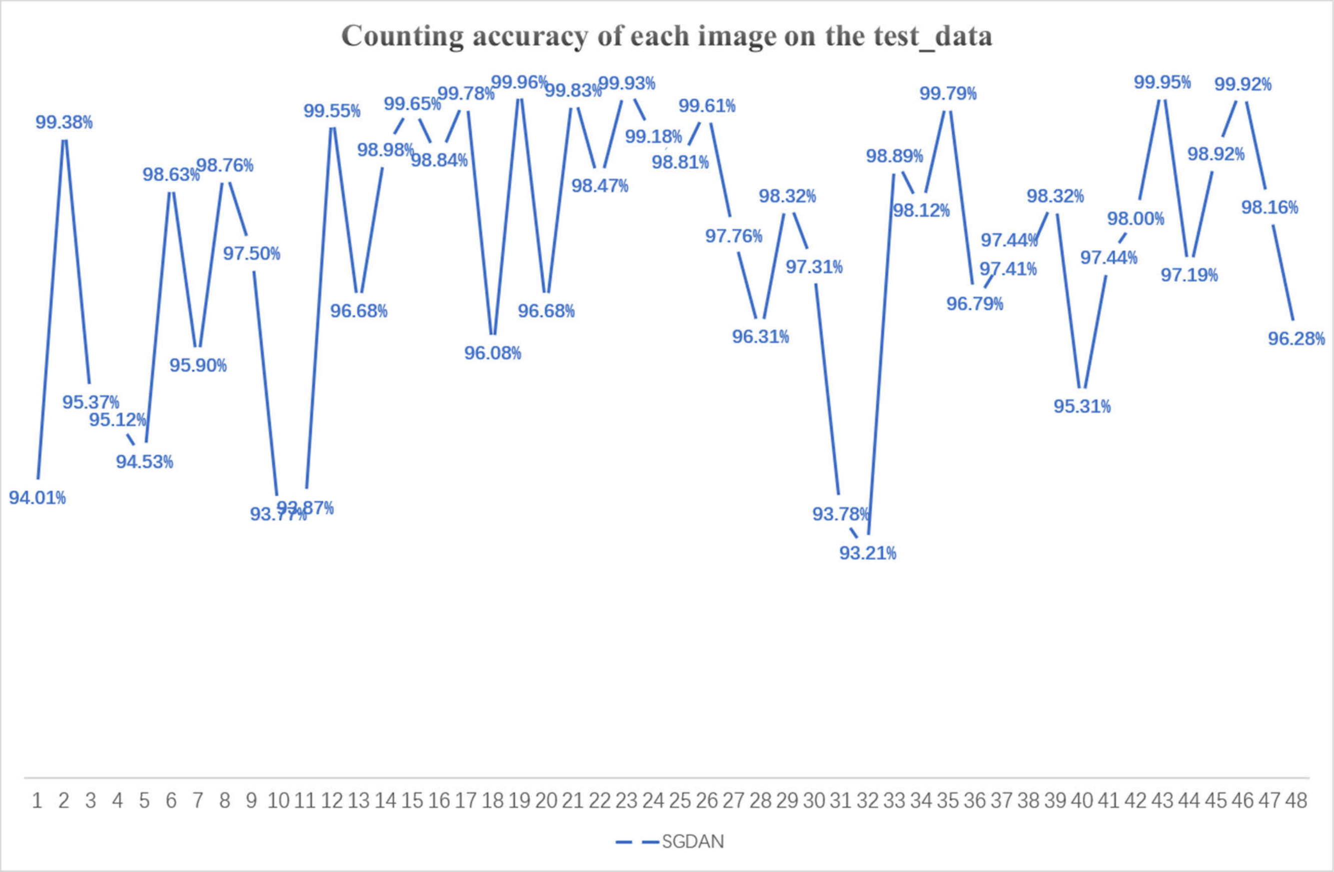


FIGURE 20. Accuracy of the SGDAN model in test_data.

TABLE 6. Quality evaluation indicators for density maps.

Model	PSNR	SSIM
DAN	20.23	0.918
SGDAN	**22.33**	**0.933**

high counting accuracy and good density map generation capability.

Table 7 and Fig.20 show the performance of SGDAN for different performance evaluation metrics and the line graph of counting accuracy for each high-density culture fry image in the test set, respectively. For the images in test_data, the MAE of the SGDAN network model proposed in this paper is 13.82, and the RMSE is 17.67. The predicted counting accuracy was above 93% for all tested images. Additionally, the predicted generated density map performed well in the image quality evaluation index, with a PSNR of 22.33 and SSIM of 0.933. Combining the above indices, it can be concluded that the SGDAN network model has good stability, high counting accuracy and good density map generation capability.

C. COUNTING RESULTS AND PERFORMANCE EVALUATION OF THE DIFFERENT MODELS

1) COMPARISON OF THE COUNTING PERFORMANCE OF DIFFERENT NETWORK MODELS

To further demonstrate the performance of the proposed model, the datasets proposed in this paper are used to train two classical density estimation network models, Multi-Column Convolutional Neural Network [22] (MCNN) and

TABLE 7. Results of assessing the indicators for the SGDAN model.

Model	MAE	RMSE	Mean Accuracy	PSNR	SSIM
SGDAN	13.82	17.67	97.57%	22.33	0.933

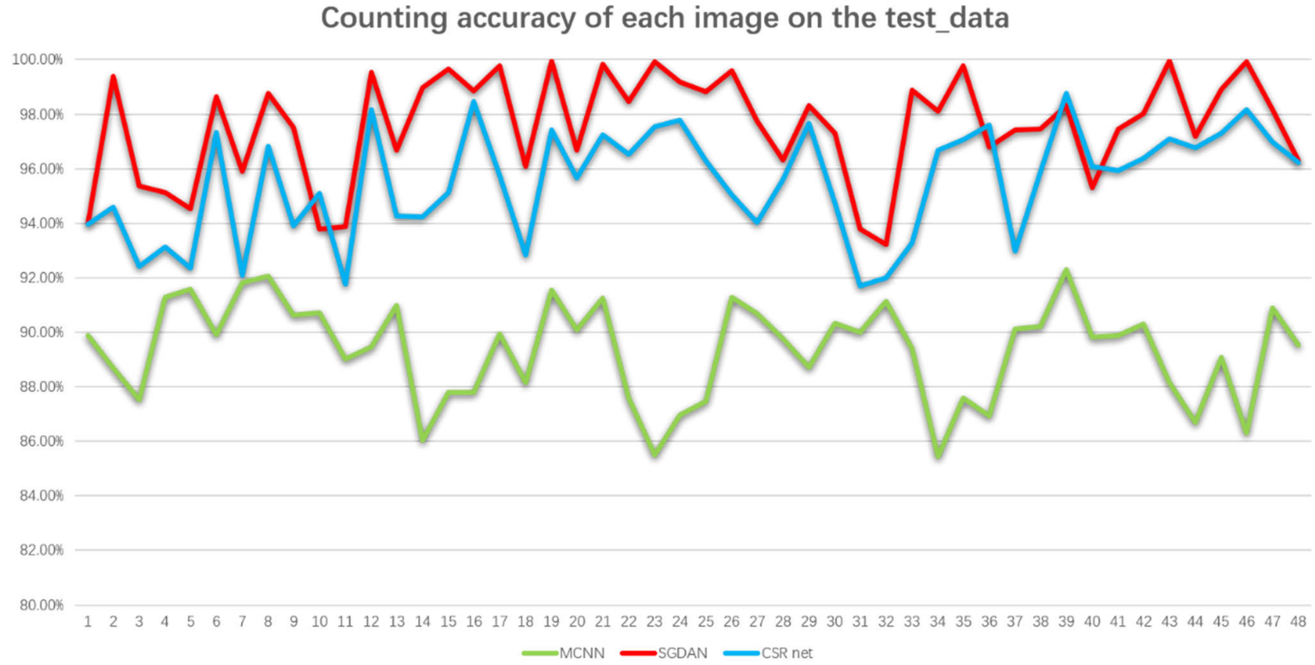


FIGURE 21. Accuracy of different models in test_data.

TABLE 8. Results of assessing indicators for different models.

Model	MAE	RMSE	Mean Accuracy
SGDAN	**13.82**	**17.67**	**97.57%**
MCNN	49.20	54.02	89.34%
CSRNet	21.02	26.45	95.51%

TABLE 9. Quality of the density maps generated by different network models.

Model	PSNR	SSIM
MCNN	19.02	0.896
CSRNet	19.70	0.902
SGDAN	**22.33**	**0.933**

Dilated Convolutional Neural Networks [33] (CSRNet), and the results of the training experiments are compared with those of the SGDAN network model proposed in this paper. The experimental results are shown in Table 8. The proposed method achieves the best performance in terms of MAE, RMSE and Mean Accuracy.

Compared with MCNN and CSRNet, SGDAN improved the MAE and RMSE by 71.9% and 67.3% and by 34.3% and 33.2%, respectively. It is shown that the SGDAN network model proposed in this paper has better accuracy and stability on the high-density aquaculture fry dataset collected based on a real aquaculture environment.

As shown in Fig.21, the SGDAN network model, represented by the red line, is much more accurate in counting on all images of test_data than the MCNN network model, represented by the green line, and the CSRNet network model, represented by the blue line, and the floating range of counting accuracy is also smaller than those of the other two classical network models. In terms of mean accuracy, the SGDAN network model improved by 8.23% and 2.06% compared to MCNN and CSRNet, respectively. All the above results show that the SGDAN network model has more accurate and stable counting performance than MCNN and CSRNet.

2) COMPARISON OF THE ABILITY OF DIFFERENT NETWORK MODELS TO GENERATE DENSITY MAPS

In actual aquaculture, achieving accurate fry counts is only one core challenge. Another central challenge is to accurately control the density of fry in the culture environment. Therefore, in addition to analyzing the performance of the models through evaluation metrics, this paper uses two image quality evaluation metrics, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), as well as visual assessment to compare the actual effects of the density maps generated by different network model predictions.

As seen from Table 9, the quality of the density map generated by the SGDAN network model is higher than that of the other two classical network models in both PSNR and SSIM evaluation metrics, with increases of 3.31 and 0.037 and 2.63 and 0.031, respectively.

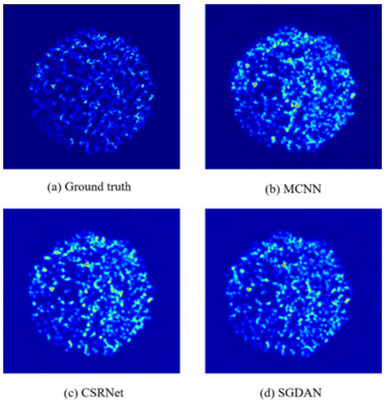


FIGURE 22. Comparison between different network models and real density maps.

In Fig.22, Fig.22(a) represents the ground truth of the original image, Fig.22(b) represents the density map generated by the MCNN network, Fig.22(c) represents the density map generated by the CSRNet network, and Fig. 22(d) represents the density map generated by the SGDAN network. From the visual evaluation, the density map generated by the MCNN-based network model is rougher and blurrier than those of the other two network models—less smooth and clear—and there are also many dense anomalies (red pixels) that do not exist in the real density map. While the density map generated by the CSRNet network-based model is visually superior to that generated by the MCNN network-based model, it is too sticky between the predicted individual fry points, which tends to result in an illusion that the fry are clustered locally. This is particularly evident by comparison with the density map generated based on the SGDAN network model.

The above results show that the SGDAN network model has better performance and capability than MCNN and CSRNet in predicting the generated density maps.

IV. CONCLUSION

In this study, the monitoring of high-density fish fry (including digitalization of quantity and visualization of density) in real aquaculture scenarios is addressed. We propose a high-density fry monitoring network model (SGDAN) incorporating an image enhancement algorithm and attention mechanism, and we collect a high-density fry dataset (HD-FryDataset) under real aquaculture environment scenario conditions for SGDAN training. The network model includes four subnetworks for image optimization, feature extraction, attention and density map estimation. Among them, the image optimization network implements preprocessing (color enhancement and noise reduction), resolution enhancement and image detail optimization of the original image of farmed fish fry. The feature extraction network acquires the overall feature map of the fry image. The attention network focuses on the key information in the overall feature map for identification and extraction. The density map estimation network implements the final fry prediction counts and generates predicted density maps containing information on the spatial distribution of fry. The average counting accuracy of the SGDAN network model can reach 97.57%, which is 8.23% and 2.06% higher than that of MCNN and CSRNet, respectively. Additionally, SGDAN achieves the best performance in comparison with MCNN and CSRNet in terms of the MAE, RMSE, PSNR, SSIM, and visual evaluation indices. In summary, the algorithm model proposed in this study has high counting accuracy and a good predictive ability to generate density maps. It can be used to monitor fish fry for high-density culture under the conditions of real aquaculture scenarios. The algorithm can also be applied to other aquaculture organisms by changing the sample types in the dataset, providing more possibilities for the intelligent and technological development of the whole aquaculture industry.

Nevertheless, our study has many shortcomings. For example, the generalization ability of the model in this paper has not yet been able to be fully validated on other farmed fish due to the difficulty in producing the dataset. Moreover, the model proposed in this paper can only run on still images. We need to do more work if we want to achieve the same results on real-time dynamic video streams.

Since the large-scale high-density fish fry dataset suffers from labeling difficulties, small sample sizes, and labeling errors, our next step will be to investigate how to introduce unsupervised or semi-supervised labeling [38], [39], with the aim of being able to further expand the size and quality of the high-density fish fry dataset. Meanwhile, it is planned to combine the target detection counting method with the density estimation counting method at a later stage to realize real-time target detection based on result-oriented density maps [40], [41]. It helps mainstream target detectors to detect aquaculture fry (especially high-density fry) more effectively.

REFERENCES

[1] J. R. Martinez-de Dios, C. Serna, and A. Ollero, "Computer vision and robotics techniques in fish farms," *Robotica*, vol. 21, no. 3, pp. 233–243, Jun. 2003, doi: 10.1017/s0263574702004733.

[2] J. Zhang, G. Yang, L. Sun, C. Zhou, X. Zhou, Q. Li, M. Bi, and J. Guo, "Shrimp egg counting with fully convolutional regression network and generative adversarial network," *Aquacultural Eng.*, vol. 94, Aug. 2021, Art. no. 102175, doi: 10.1016/j.aquaeng.2021.102175.

[3] D. N. Gonçalves, P. R. Acosta, A. P. M. Ramos, L. P. Osco, D. E. G. Furuya, M. T. G. Furuya, J. Li, J. M. Junior, H. Pistori, and W. N. Gonçalves, "Using a convolutional neural network for fingerling counting: A multi-task learning approach," *Aquaculture*, vol. 557, Aug. 2022, Art. no. 738334, doi: 10.1016/j.aquaculture.2022.738334.

[4] K. M. Babu, D. Bentall, D. T. Ashton, M. Puklowski, W. Fantham, H. T. Lin, N. P. L. Tuckey, M. Wellenreuther, and L. K. Jesson, "Computer vision in aquaculture: A case study of juvenile fish counting," *J. Roy. Soc. New Zealand*, vol. 53, no. 1, pp. 52–68, Jan. 2023, doi: 10.1080/03036758.2022.2101484.

[5] B. Zion, ‘‘The use of computer vision technologies in aquaculture—A review,’’ *Comput. Electron. Agricult.*, vol. 88, pp. 125–132, Oct. 2012, doi: 10.1016/j.compag.2012.07.010.

[6] D. Feng, J. Xie, T. Liu, L. Xu, J. Guo, S. G. Hassan, and S. Liu, ‘‘Fry counting models based on attention mechanism and YOLOv4-tiny,’’ *IEEE Access*, vol. 10, pp. 132363–132375, 2022, doi: 10.1109/ACCESS.2022.3230909.

[7] F. Antonucci and C. Costa, ‘‘Precision aquaculture: A short review on engineering innovations,’’ *Aquaculture Int.*, vol. 28, no. 1, pp. 41–57, Feb. 2020, doi: 10.1007/s10499-019-00443-w.

[8] B. Chatain, L. Debas, and A. Bourdillon, ‘‘A photographic larval fish counting technique: Comparison with other methods, statistical appraisal of the procedure and practical use,’’ *Aquaculture*, vol. 141, nos. 1–2, pp. 83–96, May 1996, doi: 10.1016/0044-8486(95)01206-0.

[9] E. A. Awalludin, M. Y. M. Yaziz, N. R. A. Rahman, W. N. J. H. W. Yussof, M. S. Hitam, and T. N. T. Arsad, ‘‘Combination of Canny edge detection and blob processing techniques for shrimp larvae counting,’’ in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2019, pp. 308–313.

[10] A. M. Grigoryan, G. Hostetter, O. Kallioniemi, and E. R. Dougherty, ‘‘Simulation toolbox for 3D-FISH spot-counting algorithms,’’ *Real-Time Imag.*, vol. 8, no. 3, pp. 203–212, Jun. 2002, doi: 10.1006/rtim.2001.0280.

[11] J. Guo, D. Zheng, J. H. Chen, and N. F. Liu, ‘‘Counting algorithm based on machine vision tracking,’’ *J. Transducer Microsyst. Technol.*, vol. 2, pp. 154–157, Jan. 2018. [Online]. Available: http://en.cnki.com.cn/Article_en/CJFDTOTAL-CGQJ201802043.html

[12] S. Shah, ‘‘Image enhancement for increased dot-counting efficiency in FISH,’’ *J. Microsc.*, vol. 228, no. 2, pp. 211–226, Nov. 2007, doi: 10.1111/j.1365-2818.2007.01842.x.

[13] F. Tajeripour and S. H. Fekri-Ershad, ‘‘Porosity detection by using improved local binary pattern,’’ in *Proc. 11th WSEAS Int. Conf. Signal Process., Robot. Autom. (ISPRA)*, vol. 1, 2012, pp. 116–121. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/2183379.2183403

[14] S. Zhang, X. Yang, Y. Wang, Z. Zhao, J. Liu, Y. Liu, C. Sun, and C. Zhou, ‘‘Automatic fish population counting by machine vision and a hybrid deep neural network model,’’ *Animals*, vol. 10, no. 2, p. 364, Feb. 2020, doi: 10.3390/ani10020364.

[15] M. Gao, M. Shi, and C. Li, ‘‘Research and implementation of image recognition of tea based on deep learning,’’ in *Proc. 21st ACIS Int. Winter Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distributed Comput. (SNPD-Winter)*, Jan. 2021, doi: 10.1109/SNPDWinter52325.2021.00021.

[16] I. Klapp, O. Arad, L. Rosenfeld, A. Barki, B. Shaked, and B. Zion, ‘‘Ornamental fish counting by non-imaging optical system for real-time applications,’’ *Comput. Electron. Agricult.*, vol. 153, pp. 126–133, Oct. 2018, doi: 10.1016/j.compag.2018.08.007.

[17] R. T. Labuguen, E. J. P. Volante, A. Causo, R. Bayot, G. Peren, R. M. Macaraig, N. J. C. Libatique, and G. L. Tangonan, ‘‘Automated fish fry counting and schooling behavior analysis using computer vision,’’ in *Proc. IEEE 8th Int. Colloq. Signal Process. Appl.*, Mar. 2012, pp. 255–260.

[18] A. Chen, Z. Li, and B. Zhang, ‘‘Automated fry counting method based on image processing,’’ in *Proc. 1st Int. Conf. Electron. Instrum. Inf. Syst. (EIIS)*, Jun. 2017, pp. 1–4, doi: 10.1109/EIIS.2017.8298769.

[19] W. Shuo, F. Liangzhong, and L. Ying, ‘‘Research on the counting method of turbot fry based on computer vision,’’ *Fishery Modernization*, vol. 42, no. 1, pp. 16–19, 2015. [Online]. Available: https://kns.cnki.net/kns8

[20] H. Z. Yang, Y. Lin, Z. S. Tang, Y. D. Zhang, Z. Chen, Y. Huang, and X. Gan, ‘‘Research on fish egg counting method based on image processing,’’ *J. Hydroecol.*, vol. 32, no. 5, pp. 138–141, 2011. [Online]. Available: https://kns.cnki.net/kns9

[21] N. Liu, Y. Long, C. Zou, Q. Niu, L. Pan, and H. Wu, ‘‘ADCrowdNet: An attention-injective deformable convolutional network for crowd understanding,’’ in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3220–3229.

[22] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, ‘‘Single-image crowd counting via multi-column convolutional neural network,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 589–597.

[23] Q. Hou, D. Zhou, and J. Feng, ‘‘Coordinate attention for efficient mobile network design,’’ 2021, *arXiv:2103.02907*.

[24] X. Yu, Y. Wang, D. An, and Y. Wei, ‘‘Counting method for cultured fishes based on multi-modules and attention mechanism,’’ *Aquacultural Eng.*, vol. 96, Feb. 2022, Art. no. 102215, doi: 10.1016/j.aquaeng.2021.102215.

[25] X. Ding, Z. Lin, F. He, Y. Wang, and Y. Huang, ‘‘A deeply-recursive convolutional network for crowd counting,’’ in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1942–1946, doi: 10.1109/ICASSP.2018.8461772.

[26] L. Boominathan, S. S. S. Kruthiventi, and R. V. Babu, ‘‘CrowdNet: A deep convolutional network for dense crowd counting,’’ in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 640–644, doi: 10.1145/2964284.2967300.

[27] H. Lin, Z. Ma, R. Ji, Y. Wang, and X. Hong, ‘‘Boosting crowd counting via multifaceted attention,’’ 2022, *arXiv:2203.02636*.

[28] H. Idrees, K. Soomro, and M. Shah, ‘‘Detecting humans in dense crowds using locally-consistent scale prior and global occlusion reasoning,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 1986–1998, Oct. 2015.

[29] T. Zhao, R. Nevatia, and B. Wu, ‘‘Segmentation and tracking of multiple humans in crowded environments,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1198–1211, Jul. 2008, doi: 10.1109/TPAMI.2007.70770.

[30] M. Li, Z. Zhang, K. Huang, and T. Tan, ‘‘Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection,’’ in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4, doi: 10.1109/ICPR.2008.4761705.

[31] W. Ge and R. T. Collins, ‘‘Marked point processes for crowd counting,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2913–2920, doi: 10.1109/CVPR.2009.5206621.

[32] V. A. Sindagi and V. M. Patel, ‘‘A survey of recent advances in CNN-based single image crowd counting and density estimation,’’ *Pattern Recognit. Lett.*, vol. 107, pp. 3–16, May 2018, doi: 10.1016/j.patrec.2017.07.007.

[33] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, ‘‘Photo-realistic single image super-resolution using a generative adversarial network,’’ 2016, *arXiv:1609.04802*.

[34] O. Keleğ, M. A. Yılmaz, A. M. Tekalp, C. Korkmaz, and Z. Dogan, ‘‘On the computation of PSNR for a set of images or video,’’ 2021, *arXiv:2104.14868*.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, ‘‘Image quality assessment: From error visibility to structural similarity,’’ *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[36] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, ‘‘Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,’’ 2016, *arXiv:1609.05158*.

[37] Y. Li, X. Zhang, and D. Chen, ‘‘CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes,’’ 2018, *arXiv:1802.10062*.

[38] X. Wang, L. Lian, and S. X. Yu, ‘‘Unsupervised selective labeling for more effective semi-supervised learning,’’ in *Computer Vision—ECCV*, vol. 13690, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-031-20056-4_25.

[39] B. Xiao and C. Lu, ‘‘Semi-supervised medical image classification combined with unsupervised deep clustering,’’ *Appl. Sci.*, vol. 13, no. 9, p. 5520, Apr. 2023, doi: 10.3390/app13095520.

[40] C. Li, T. Yang, S. Zhu, C. Chen, and S. Guan, ‘‘Density map guided object detection in aerial images,’’ 2020, *arXiv:2004.05520*.

[41] L. Zhao, Z. Bao, Z. Xie, G. Huang, and Z. U. Rehman, ‘‘A point and density map hybrid network for crowd counting and localization based on unmanned aerial vehicles,’’ *Connection Sci.*, vol. 34, no. 1, pp. 2481–2499, Dec. 2022, doi: 10.1080/09540091.2022.2130878.

HONGYUAN CHEN received the B.S. degree in electronic information engineering from Zhengzhou University of Light Industry, in 2020. He is currently pursuing the M.S. degree in control science and engineering with Dalian Ocean University, China. His research interests include computer vision, deep learning, and aquatic informatization.

YUAN CHENG received the B.S. degree in fishery resources and management from the Ocean University of China, in 2001, and the M.S. degree in rural and regional management from Dalian Ocean University, in 2015. Since 2021, he has been an Institute Researcher with Ningbo Research Institute, Dalian University of Technology. His research interests include aquaculture, ocean engineering, and aquatic informatization. He has won the First Prize of the National Marine Science and Technology Award and the First Prize of Shanghai Marine Science and Technology Award.

YU DOU received the B.S. degree in electrical engineering and its automation from Cangzhou Normal University, China, in 2018, and the M.S. degree in control science and engineering from Dalian Ocean University, China, in 2021. His research interests include deep learning and computer vision.

HUACHAO TAN received the B.S. degree in electronic information engineering from Shandong University of Technology, in 2020, and the M.S. degree in control science and engineering from Dalian Ocean University, China, in 2022. His research interests include deep learning, computer vision, and object detection.

GUIHONG YUAN received the B.S. degree in information and computing science from Dalian Ocean University in 2022. He is currently pursuing the M.S. degree in electronic information. His research interests include computer vision, deep learning, and few-shot learning.

HAI BI received the B.S. degree in electronic information engineering from Changchun Institute of Technology, in 2007, and the M.S. degree in electronics and communication engineering from Harbin Institute of Technology. His research interests include the development of object detection and recognition and image segmentation and analysis.

DAN LIU received the Ph.D. degree from Jilin University, China, in 2010. She is currently an Associate Professor with the College of Electrical Information Engineering, Dalian Ocean University, Dalian, China. Her current research interests include the theory and application of marine-related intelligent image processing and wireless networks and communication systems.