A lightweight network for portable fry counting devices

Weiran Li, Qian Zhu, Hanyu Zhang, Ziyu Xu, Zhenbo Li

College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, PR China
National Innovation Center for Digital Fishery, Ministry of Agriculture and Rural Affairs, Beijing, 100083, PR China
Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs, Beijing, 100083, PR China
Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing, 100083, PR China
Key Laboratory of Smart farming for aquatic animal and livestock, Ministry of Agriculture and Rural Affairs, Beijing, 100083, PR China

ARTICLE INFO

Article history:
Received 30 January 2022
Received in revised form 9 October 2022
Accepted 18 February 2023
Available online 23 February 2023

Keywords:
Fry counting
Lightweight Convolutional Neural Network
Channel attention mechanism
Density map regression
Computer vision

ABSTRACT

Estimating the number of fries plays a critical role in the maintenance of fish breeding, transportation, and the preservation of marine resources in aquaculture. Generally speaking, statistics are recorded manually by fishers and government units. Manual recording is time-consuming and increases the workload of fishers. Compared with traditional physical shunt devices, visual-based algorithms have benefits such as non-restriction of labors, minimal equipment installation, and maintenance costs. However, these methods generally come with massive calculations and model parameters, or poor abilities of aggregation handles and counting precision. This paper proposes a fry counting method named MSENet for portable fry counting devices. Firstly, the lightweight network is designed with simpler parameters (Params: 139.46 kB) for portable embedding. The visualized single-channel fry density maps are predicted by feeding the original images and the number of fries is calculated through integration. Then, the Squeeze-and-Excitation block is utilized to strengthen the features of weighty channels. The model training is refined by hyperparameter studies, the shortened preparation stage enhances the portability. What is more, a fry counting dataset NCAUF and an extra set NCAUF-EX are built for verifications of network generalization. The results demonstrate that the lightweight MSENet outperforms in fry counting with higher precision and competently solves the issue of fry aggregation (MAE: 3.33). The source code and pre-trained models are available at: https://github.com/vranlee/MSENet.

1. Introduction

The assessment of fish biomass and stocks is important for a wide range of intelligent aquaculture, to maintain an acceptable breeding density and precision feeding. Fish counting is crucial for ascertaining the optimal development and transportation strategy, as well as for supplying conditions, calculating profits, etc. [1,2]. In scenarios such as proliferation and release, logistics, and transportation, the necessity for counting fry is more pressing than the need for adult fish monitoring in aquaculture. Fry counting is a subproblem of fish counting, which is more challenging to solve due to higher densities. Specifically, higher densities result in hypoxia, predation, and other issues, whereas lower densities require larger aquaculture ponds and higher maintenance costs.

Manual counting methods of fry counting often utilize tools such as a fish sieve (a special tool for separating large and small fry), measuring cups, or weighing directly. These methods with poor calculation efficiency are easily exploited by unscrupulous merchants [3]. Therefore, there is an urgent need for an automatic counting algorithm to develop efficiency and reduce the intensity of labor. More recently, literature about physical diversion methods has emerged that take the place of manual counting: Boys et al. [4] investigated the influence of approach velocity and mesh size on-screen design, reducing the damage caused to fish to some extent. Mesa et al. [5] analyzed the flat-plate fish screens and find the adaptability interval of the screen depth. These methods have reduced the workload of fishery staff efficiently, but they have the drawback of low counting rate or accuracy, causing damage to the fry due to the interaction of the physical device.

Vision-based counting methods [6–8] strategically diminish the physical damage to the fish bodies, and the simple device requirement of a signal camera decreases the application complexity and equipment maintenance costs in aquaculture. Lainez and Gonzales [9] published a fry counter based on image processing. It utilizes the Convolutional Neural Network (CNN) for automatic fish counting and verifies the performance of optimized thresholds in various sizes of fish. Klapp [10] designed
2. Methodology

2.1. Dataset and implement details

Considering that the associated system is applied to fixed locations in small farmers, a portable device equipped with a camera is picked. The main camera should be a top-down view, and the fry roam freely in annular aquaculture ponds. As a result, the dataset scene is set in a white circular basin. Varied density conditions are created by placing up to 300 crucian carp fries in a fixed scenario, which is used to replicate the dispersion of fish fry in aquaculture ponds with varying densities.

In this research, we utilize the New China Agricultural University Fry (NCAUF) dataset as our basic dataset, including 3200 images with an 800 x 600 or 600 x 800 resolution, and each image contains about 100–300 crucian carp fries, length varies from 1 to 3 cm. NCAUF is augmented from the China Agricultural University Fry (CAUF) dataset, with an original resolution size of 4000 x 3000 or 3000 x 4000, captured by the Redmi Note 7 mobile phone camera. Down-sampling is performed to compress the time of density map generation.

In order to validate the generalizability of the model under various lighting conditions, culture pond backgrounds, and other influencing factors, data enhancement and augmentation are carried out. The basic dataset is enhanced by adjusting brightness, contrast, image smoothing, and Gaussian noise. Four expanded image sets are mixed with the original one to compose NCAUF. Various density distributions of fry in argument images are included in the samples, as shown in Fig. 1. The sparse set contains about 100 fries; the slightly dense set contains about 170 fries; the dense set contains about 280 fries. Structural Similarity (SSIM) is a metric that compares the brightness, contrast, and structure of enhanced and original images. The noise quality of the enhanced images is measured by the Peak Signal to Noise Ratio (PSNR). According to the ratio of 6:2:2, the data is divided into training, validation, and test sets. The samples and division results are enumerated in Fig. 1 and Table 1. The experiments are conducted on one GeForce GTX 1080 Ti GPU. Furthermore, to check the compatibility and lightweight properties of the model, the CPU of the AMD R5 3500U is utilized for test. Finally, the lightweight model is verified to be built on a portable device and achieve real-time fry counts by evaluating it on a Raspberry Pi 4B paired with an HD camera.

1 The sample datasets can be found at https://github.com/vranlee/MSENet.
Table 1
Division result of NCAUF and CAUF dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training set</th>
<th>Validation set</th>
<th>Test set</th>
<th>Total images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>480</td>
<td>160</td>
<td>160</td>
<td>800</td>
</tr>
<tr>
<td>After augmentation</td>
<td>2880</td>
<td>160</td>
<td>160</td>
<td>3200</td>
</tr>
</tbody>
</table>

2.2. Framework

The researchers have found that the multi-scale or multi-task methods can make an effective contribution to the accuracy of the counting model. In this work, we propose a lightweight MSENet for portable fry counting devices. By feeding the target image sequence, the network generates and visualizes the single-channel density map of fry, and predicts the result by the integration of the map.

The processing of datasets should be considered from three perspectives: (1) a lightweight model is able to process data that has uniform size and high contrast, (2) the adhesion occurs when the amount of fry becomes larger, which causes a deviation in counting results, (3) precision and efficiency are crucial priorities in aquaculture applications. MSENet strengthens the robustness of the counting model and enhances the accuracy and speed as well. The network structure diagram is shown in Fig. 2. We utilize the multi-column neural network of MCNN [19] in the front-end network and add the attention mechanism module of Squeeze-and-Excitation Networks [20] in the back-end network. The end network uses a 1 × 1 convolution layer to convert the synthesized map into a single-channel density map, then predicts the counting results of fries through an integral operation.

2.3. Multi-column network structure

MCNN [19] is applied in crowd counting of single-image. Although the network has achieved good results in applications, it leads to an excessive amount of training parameters and calculations in the training process. What is more, the multi-column network structure cannot efficiently predict objects of different sizes in the processing process. MSENet retains the advantages of MCNN for the front-end network. In the back-end network, a channel attention mechanism is introduced to improve the accuracy and efficiency of the model. The end network adopts a 1 × 1 convolution layer to take the place of the fully connected layer, which reduces the parameters and realizes multi-scale input to avoid distortion.

The multi-column structure of the front-end implements a three-column convolutional neural network: Large-scale(L) column uses large-scale convolution kernel as 9 × 9, 7 × 7, 7 × 7; Medium-scale(M) column uses medium-scale convolution kernel as 7 × 7, 5 × 5, 5 × 5, 5 × 5; Small-scale(S) column uses small-scale convolution kernel as 5 × 5, 3 × 3, 3 × 3, 3 × 3, 3 × 3. The parallel sub-networks in each column have the same depth but different sizes of the convolution kernel. Finally, each sub-network gets various receptive fields. The purpose of using multiple-scale convolution kernels is to respond to the scales of fry, which strengthens the capability to acquire the fry features of unequal spatial dimensions. A 2 × 2 max-pooling and ReLU are adopted as the activation function. After the max-pooling operation, the training samples are reduced by 1/4. The generated sub-blocks compress the image size to accelerate the train of the network, and achieve a comparable result as well.
2.4. Attention block

MSENet introduces the Squeeze-and-Excitation module in the back-end from the SENet [20]. It is usually embedded in the network to cooperate with various network components or structures, such as a convolutional layer, ResNet [21], Inception [22], etc. The channel attention mechanism is shown in Fig. 3.

On the one hand, since each convolution layer only extracts the corresponding local features, in other words, it just corresponds to a local receptive field, the features are incapable of utilizing external information. Therefore, a global average pooling is implemented for each channel in each feature. It can be defined as:

\[
Z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \quad (1)
\]

where \(H \times W\) is the dimension of spatial, \(H\) is the height, and \(W\) is the width. The features of each channel are enhanced to make full utilization of the information after the pooling and the dependence between channels. The channel conversion makes the relations not mutually exclusive and keeps a non-linear relationship:

\[
s = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)), W_1 \in \mathbb{R}^{C \times C}, W_2 \in \mathbb{R}^{C \times \hat{C}} \quad (2)
\]

where \(\sigma\) is the Sigmoid function, \(\delta\) is the ReLU function. The parameter \(r\) is used to reduce the dimension of the fully connected layer.

2.5. Features merged

The 1 \times 1 convolution layer is introduced to the end network to take the place of the fully connected layer, which validly reduces the parameters. The block is responsible for merging the three columns of CNNs and predicting a density map. MSENet compresses network parameters by combining the columns of convolutional neural networks and utilizes generated image sub-blocks to speed up training. The MSENet is lighter than the previous models, and the density map determines its performance. In detail, the marking point of a single fry can be recorded as \(x_i\), then \(\delta(x-x_i)\) is used to mark a fish. The frame containing \(N\) fries can be labeled as:

\[
y(x) = \sum_{i=1}^{N} \delta(x-x_i) \quad (3)
\]

To correlate the density map with the number of fries, the network implements the density map based on a geometrically adaptive Gaussian kernel as:

\[
Y(x) = \sum_{i=1}^{N} \delta(x-x_i) * H_\theta(x), \text{ with } \theta = \alpha \bar{F} \quad (4)
\]

\[
\bar{F} = \frac{1}{d} \sum_{j=1}^{d} \bar{f}_j \quad (5)
\]

where \(x_i\) is the pixel coordinate, \(\delta(x-x_i)\) is the function of fry coordinates in the image, \(N\) is the total number of fries. \(H_\theta(x)\) represents the Gaussian convolution kernel, \(\bar{f}\) is the average distance of the closest \(d\) fries to the target fry. In general, we mark the center of fry as the distance between each other, and the distance is approximately equal to the size of the fish heads in dense conditions. In our experiments, the model predicts the best counting result when the \(\alpha\) is set to 0.3. The position of the fry in the density map has a better correspondence to the variance of the Gaussian kernel, which reflects the size of the fry more clearly (Fig. 4).
3. Experiments

3.1. Metrics and visualization

In model training, we utilize Mean Absolute Error (MAE), Mean Squared Error (MSE) to evaluate the counting precision, the size of parameters (Params) to reflect the complexity of networks. Computing costs are expressed in floating-point operations per second (FLOPs). The MAE and MSE are calculated as follows:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |\text{actual}_i - \text{predicted}_i| \\
MSE = \frac{1}{n} \sum_{i=1}^{n} (\text{actual}_i - \text{predicted}_i)^2
\]  

The visualized counting results are shown in Fig. 5. It can be concluded that our MSENet keeps a high counting accuracy under various fry densities, with an error rate under 5%.

We created NCAUF-EX test dataset to investigate the counting effect under various scenarios, fish species, and numbers in order to verify the generalization of MSENet. NCAUF-EX test set includes 100 frames of 1920*1080 resolution images, captured from a fixed camera perspective of an aquaculture pond. Compared to NCAUF dataset, the background implications are relatively complex. Fries are partially or entirely obstructed due to the slanted shooting angle. Natural motion video clips of 10 cm spotted knifejaw fry are captured in this set. The internal light setting is relatively inadequate. The camera captures a portion of the fish shadows as well. On NCAUF-EX dataset, the generalization test of the model is shown in figure 6. The results show that our model maintains superior counting precision in various scenarios with an error rate less than 6%.

<table>
<thead>
<tr>
<th>Methods</th>
<th>MAE</th>
<th>MSE</th>
<th>Params</th>
<th>FLOPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANet</td>
<td>15.07</td>
<td>17.49</td>
<td>1.4M</td>
<td>4.45G</td>
</tr>
<tr>
<td>CMTL</td>
<td>5.22</td>
<td>6.92</td>
<td>2.4M</td>
<td>5.98G</td>
</tr>
<tr>
<td>SFCN</td>
<td>3.39</td>
<td>4.61</td>
<td>38.6M</td>
<td>30.97G</td>
</tr>
<tr>
<td>CSRNet</td>
<td>3.33</td>
<td>4.58</td>
<td>0.1M</td>
<td>1.38G</td>
</tr>
</tbody>
</table>

3.2. Comparative studies

In this section, we conduct comparative experiments with typical crowd counting models, including SANet [23], CMTL [18], SFCN [24], CSRNet [17] and the fry counting model FCNet,\(^2\) based on density map regression.

We verify the effectiveness of the network by comparing it with mainstream density map regression-based crowd counting methods. The result is shown in Fig. 7 and Table 2.

The results show that SANet has a smaller model parameter size, but the counting precision is terrible. Similarly, CMTL performs poorly in precision and model complexity compared to ours. The parameters of SFCN and CSRNet achieve 16.3M or more. The FLOPs are up to 30.97G, implying excessive requirements for computing. However, the counting precision of SFCN and CSRNet is poorer than ours. Our MSENet outperforms others on MAE (3.33) and MSE (4.58), which indicates that MSENet keeps a high counting accuracy in the various densities of fry. Besides, the remarkably slight size of model parameters (0.1M) and the least FLOPs (1.38G) compared to other models. The results show that MSENet outperforms other density map regression-based

\(^2\) Proposed by Zheng Miao, master degree thesis of China Agricultural University.
counting methods in both accuracy and simplicity, indicating that it can be implemented in portable processors.

In comparison to crowd counting, the size of the fry is smaller in the images combined with single appearance information, the coloration contrast is more perceptible as well. The occlusion problem is particularly remarkable in dense scenarios. On the one hand, the fish tend to aggregate when they encounter external disturbance, resulting in visible adhesion. On the other hand, the head of a fish is easily occluded in high-density situations and usually the heads are adopted as the center point of the annotation. We compare our model with the fry counting method FCNet, which is based on the density map regression method as well. The results are shown in Table 3. According to the statistics, MSENet has superior counting effectiveness and precision. The parameter size is only 1% of FCNet, which makes it more appropriate to be integrated into portable counting devices.

### 3.3. Optimal network learning strategy

An appropriate learning strategy helps the training process reach a stable convergence point faster. Optimizing the hyperparameters in this section further maximizes the mobility of the model, facilitating application processes of related breeding scenarios. In our experiments, we set the initial parameter variables, including batch_size, learning rate, decay rate, and optimizer as Table 4.

We implement various learning rates in the training phase. The convergence curves are shown in Fig. 8. The horizontal and vertical coordinates represent the number of training epochs and the MAE, MSE values, respectively. The MAE and MSE corresponding to the various learning rates after convergence are listed in Table 5.

The experiments implement various learning rates between 0.00001 and 0.00100. Fig. 8 indicates that when the learning rate is set to 0.00050, the model reaches relatively smooth learning progress and achieves 3.70 MAE and 4.94 MSE. Table 5 shows that MAE begins to shrink when the learning rate is set as 0.00010. The minimum MAE and MSE are 3.47 and 4.69, respectively. A higher learning rate (larger than 0.00100) destroys the training convergence of MSENet, while a lower learning rate (smaller than 0.00010) slows down the training effect. It can be concluded from the results that when the learning rate keeps the interval between 0.00010 and 0.00050, the model achieves a better training function.

Similarly, we experiment with the determination of learning rate decay and set the learning rate at 0.00010. As is shown in Fig. 9, the experiments are implemented at five decay rates of 0.980, 0.990, 0.995, 0.998, and 1.000. Table 6 shows the best MAE and MSE corresponding to various learning decay rates.

The results show that when the decay rate is set between 0.980 and 0.990, MSENet achieves a more stable training process. The 0.995 to 1.000 decay rates bring training results with greater volatility in recognition accuracy. When the decay rate is 1.000, MSENet gets the best result of 3.33 MAE and 4.58 MSE. Combined with the results of the compaction of various learning rates and decay rates in MSENet training, we prefer the learning rate of 0.00050 and the decay rate of 0.980 for model migration application scenarios.

### 3.4. Ablation studies

#### 3.4.1. Backbone

To employ our MSENet into removable embedded devices, we sort out various networks and calculate the size of parameters and FLOPs, including AlexNet [25], ResNet [21], VGG [26].

![Fig. 6. Generalization test of MSENet on NCAUF-EX dataset.](image)

![Fig. 7. Performance of typical crowd counting methods.](image)
The results show that VGG, VGG-Decoder, and ResNet-50 have parameters under 10M, the parameter sizes of ResNet-101, AlexNet are larger than 20M, which greatly increases the complexity of models. The FLOPs of VGG, VGG-Decoder, ResNet-50, and ResNet-101 are larger than 7G, which brings higher computation costs. MCNN has the smallest parameter size of just 0.1M, which can be embedded into mobile services. Then we compare the counting accuracy of AlexNet and MCNN specifically. AlexNet and MCNN are utilized as the backbone networks and trained on NCAUF dataset. Table 8 shows the detailed performance. From MAE and MSE we can conclude that the MCNN has a better counting accuracy performance. The AlexNet has

<table>
<thead>
<tr>
<th>Backbone</th>
<th>Params (M)</th>
<th>FLOPs (G)</th>
<th>Training time (s)</th>
<th>Val time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG</td>
<td>7.7M</td>
<td>13.99G</td>
<td>55.55</td>
<td>1.81</td>
</tr>
<tr>
<td>VGG-Decoder</td>
<td>8.4M</td>
<td>15.63G</td>
<td>60.47</td>
<td>2.03</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>8.7M</td>
<td>7.43G</td>
<td>57.64</td>
<td>2.11</td>
</tr>
<tr>
<td>ResNet-101</td>
<td>27.7M</td>
<td>22.35G</td>
<td>145.01</td>
<td>4.27</td>
</tr>
<tr>
<td>AlexNet</td>
<td>61.1M</td>
<td>0.67G</td>
<td>26.33</td>
<td>1.70</td>
</tr>
<tr>
<td>MCNN</td>
<td><strong>0.1M</strong></td>
<td>1.36G</td>
<td>26.28</td>
<td><strong>1.60</strong></td>
</tr>
</tbody>
</table>

Fig. 8. The MAE and MSE curves correspond to various learning rates.

Fig. 9. The MAE and MSE curves correspond to various decay rates.

Fig. 10. The statistics of parameter size and FLOPs of various backbones.

The experiments show the complexity and portability of the backbones, as shown in Fig. 10 and Table 7.
an extremely large size of parameters (61.1M) compared to the MCNN (0.1M), so the increased expense of embedded devices is challenging to put into practice on a large scale in fishery production. Both MCNN and AlexNet have tinier FLOPs, however, the MCNN has a simpler training and test process. The MAE and MSE curves are shown in Fig. 11.

### 3.4.2. Attention block

We conduct a comparative experiment with MCNN to verify the effectiveness of the introduced channel attention mechanism module. The models are trained on NCAUF datasets to verify the improvement in counting accuracy when the attention block is embedded into our network. The result is shown in Fig. 12 and Table 9.

The result indicates that the MSENet with attention block achieves the smallest MAE (3.33) and MSE (4.58), bringing a 28.2% and 22.6% error reduction on MAE and MSE, respectively. The parameter size and FLOPs show that our attention block brings an obvious performance improvement without any obvious complexity or calculation multiplying. The reason is that the embedded SENet block improves the recognition performance of the network for small-scale fish targets. Channel features with the largest amount of information are emphasized, while the irrelevant features are simultaneously suppressed.

### 3.4.3. Data augmentation

We prove the applicability of data transformation through further experiments. To confirm the outperformance of MSENet on both NCAUF and CAUF dataset, we utilize MAE, MSE, parameter size, and FLOPs to test the accuracy, model complexity, and calculation cost of MCNN and MSENet. The result is shown in Fig. 13 and Table 10.

The table shows that NCAUF dataset obviously improves the training effect of the models and increases the robustness of the model, potentially due to the larger size and complexity of the images. Furthermore, our MSENet increases counting precision on both NCAUF and CAUF dataset, as seen by the decreased MAE and MSE.

### 4. Conclusion

Compared with the poor efficiency and physical injury of traditional fry counting applications, the utilization of a lightweight vision-based model realizes automatic fry counting with high precision. In this paper, we have proposed a lightweight fry-counting network, MSENet, with outperforming portability. The Squeeze-and-Excitation block is introduced into the model, combined with the destiny map regression method. An obstacle fry dataset, NCAUF, containing various condition factors has been organized for training and experiment with various hyperparameters to compress the training stage for migrating applications. To verify the outperformance of our model, we compared MSENet with several typical counting methods and conducted ablation studies to show the effectiveness of each module. The results show that our MSENet significantly improves the counting efficiency and precision both in sparse and aggregation scenarios. And the minimal complexity and weak computation costs of our model make it embeddable into the mobile fry counting equipment in aquaculture. Unfortunately, the existing method is limited to a scene with a fixed viewpoint, which makes it suitable...
for counting fry inside a circular culture tank. On the one hand, the unsupervised training strategy may be extended to overcome the problem of limited density map annotation datasets and the cost of fish dataset annotation in future works. Further research should be undertaken to investigate the counting technique within a constrained field of vision with the implementation of a mix of hardware equipment designed for the proliferation and release stage. Based on the benefits of density map regression in dealing with occlusion, the solution of iterative counting in stream scenes has practical significance.

**CRediT authorship contribution statement**

**Weiran Li:** Methodology, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing.  
**Qian Zhu:** Methodology, Investigation, Resources, Data curation, Visualization.  
**Hanyu Zhang:** Visualization, Investigation.  
**Ziyu Xu:** Data curation, Visualization.  
**Zhenbo Li:** Conceptualization, Validation, Formal analysis, Investigation, Writing – review & editing, Supervision.

Table 9  
The performance of MCNN and MSENet.

<table>
<thead>
<tr>
<th>Network</th>
<th>Attention block</th>
<th>MAE$\downarrow$</th>
<th>MSE$\downarrow$</th>
<th>Params$\downarrow$</th>
<th>FLOPs$\downarrow$</th>
<th>Training time (s)</th>
<th>Val time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCNN</td>
<td></td>
<td>4.64</td>
<td>5.92</td>
<td>0.1M</td>
<td>1.4G</td>
<td>26.59</td>
<td>1.61</td>
</tr>
<tr>
<td>MSENet</td>
<td>$\checkmark$</td>
<td>3.33 ($\downarrow28.2%$)</td>
<td>4.58 ($\downarrow22.6%$)</td>
<td>0.1M</td>
<td>1.4G</td>
<td>27.27 (+0.68)</td>
<td>1.93 (+0.32)</td>
</tr>
</tbody>
</table>

Table 10  
The performance of MCNN and MSENet trained on NCAUF or CAUF dataset.

<table>
<thead>
<tr>
<th>Networks</th>
<th>Data augmentation</th>
<th>Dataset</th>
<th>MAE$\downarrow$</th>
<th>MSE$\downarrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCNN</td>
<td>$\checkmark$</td>
<td>CAUF</td>
<td>12.9</td>
<td>15.4</td>
</tr>
<tr>
<td>MCNN</td>
<td>$\checkmark$</td>
<td>NACUF</td>
<td>4.64</td>
<td>5.92</td>
</tr>
<tr>
<td>MSENet</td>
<td>$\checkmark$</td>
<td>CAUF</td>
<td>11.7</td>
<td>14.4</td>
</tr>
<tr>
<td>MSENet</td>
<td>$\checkmark$</td>
<td>NACUF</td>
<td>3.33</td>
<td>4.58</td>
</tr>
</tbody>
</table>

Fig. 13. The MAE and MSE curves of MCNN and MSENet.
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the Key-Area Research and Development Program of Guangdong Province - Ecological engineering breeding technology and model in seawater ponds (2020B020201009); National Key R&D Program of China - Integrated demonstration of seawater and brackish water fish ecological intensive intelligent breeding model and processing and circulation (2020YFD0900204).

References