PAPER • OPEN ACCESS

A Classification Algorithm of Fish Feeding Behavior for Automatic Bait Feeding Control

View the article online for updates and enhancements.

A Classification Algorithm of Fish Feeding Behavior for Automatic Bait Feeding Control

Jialin Zhang, Feng Cen and Lihong Xu*

Department of Control Science and Engineering, Tongji University, Shanghai 201804 China

*Corresponding author email: xulihong@tongji.edu.cn

Abstract. In aquaculture, automatic feeding control of fish can effectively improve production efficiency. However, most automatic feeding devices are open-loop control because of lack of information on fish feeding behavior, which could lead to a large amount of bait waste. To address this problem a novel video classification algorithm based on an inter-frame relationship Bayesian estimation network(IRBEN) is proposed in this paper, which provided prior knowledge for automatic feeding control of fish. The IRBEN first employs a VAE encoder to convert the frames of a video clip into multivariate Gaussian distributions(MGDs). Then, two fully connected networks, one is trained on the MGDs associated with the fish eating video clips and the other on the MGDs associated with the fish noneating video clips, are employed to predict the MGDs of the frame after an interval from the MGD of the current frame. The classification is conducted by finding the fully connected network achieving smaller KL distance between the predicted MGD and the actual MGD. The experimental results show that the IRBEN achieves the classification accuracy of 97.5%.
Keywords: Automatic control; Aquaculture; Video classification; Computer vision.

1. Introduction

According to the report of FAO[1], two-thirds of the global fish production, which are 1.7×10^8 tones in 2016, are aquaculture production. The amount of feed is a critical factor that determines the profit of aquaculture. In some fish breeding processes, the expense on the bait accounts for about 40% of the total cost. In order to reduce the unnecessary feeding and achieve the automatic control of bait feeding, the fish feeding behavior is intensively investigated to assess fish appetite and guide the feeding[2][3].The aim of this work is to push further the development of fish feeding behavior understanding in a commercial-scale recirculating aquaculture system (RAS) via video classification.

In recent years many intelligent aquaculture methods have been developed including acoustic methods[4]-[6] and machine vision methods[7]-[10]. Compared with the acoustic methods, the machine vision methods have the characteristics of high efficiency and convenience to use in real production practice[11], and thus, become a technology trend in the field of fish feeding behavior understanding[2].

In the task of fish feeding behavior analyzing, many attempts have been made based on machine vision, such as, detecting residual feeds via segmenting underwater images[8][12][13], estimating feeding intensity by using a modified kinetic energy model[14], and quantifying feeding behavior through Delaunay triangulation[9]. These works, however, only focused on the spatial features of fish feeding activities. For a commercial-scale RAS, where the density of the fishes is usually very high, sometimes, even fishery experts are unable to recognize the fish feeding behavior based merely on a single image. By contrast, besides the spatial features, the video clips can provide the temporal

features that include additional discriminative information about the fish feeding behavior. Therefore, we consider employing the video clips to classify the fish feeding behavior in this paper.

2. Background

The inter-frame relationship of the video sequence has already been utilized in some research works to quantify the fish feeding behavior recently. Duarte, S. et al.[15] introduced an image processing activity index (IPAI) based on frame difference. Liu et al.[7] proposed a computer vision-based feeding activity index (CVFAI) by considering the overlap of the fishes. In these works, however, the experiments were conducted in a simulation site of the commercial-scale RAS and the videos were acquired from the cameras that were deployed above the water surface. In the commercial-scale RAS, however, the density of the fishes is much higher (around 1,000 fishes per tank in our experiment) and illumination condition is more complicated than in the simulation site. Therefore, more challenges will be faced for fish feeding behavior understanding from the images or videos acquired in the commercial-scale RAS. Besides, the complex illumination condition can cause serious light reflection, which undermines the performance of most methods[7][9], for the cameras placed over the water surface. For instance, the light reflection is so serious that many fishes are blocked out. To alleviate the interference of the light reflection, in this work, we place the cameras under water surface to record videos.

For the tanks with dense fishes, Atoum, Y. et al.[8] developed an approach based on reflection frame feature and feed detection. However, in a commercial-scale tank, the light reflection in a frame is strongly affected by bait, and thus, the handcrafted features are unsuited to represent fish feeding behavior[16]. In contrast, this paper employs neural networks to learn the inter-frame relationship of the video clip.

In aquaculture, deep neural network based methods have been applied to the fields like live fish recognition[17][18], fish behavior analysis and trajectory tracking[19]-[21] and fish species classification[22]. Deep neural networks learns features from a task-specific dataset automatically. However, there is still lack of the datasets acquired under the industrial production environment. Zhou et al.[16] built a fish feeding behavior image dataset from a simulation site of the commercial-scale RAS. However, due to the differences in breeding density and optical environment, the deep learning model trained in the simulation environment has a significant performance degradation in the industrial environment.

To address the problem of the fish feeding behavior understanding in the commercial-scale RAS, in this work, a dataset named underwater video dataset of the Atlantic salmon swimming behavior (UVDASSB), where the videos are captured under an industrial production environment, is first constructed. Then, an inter-frame relationship Bayesian estimation network(IRBEN)-based classification method is proposed to identify the eating behavior and noneating behavior from the video clip. In the UVDASSB, each sample is a 5-second video clip annotated as either eating or noneating. The number of noneating and eating video clips are 3132 and 659, respectively. The IRBEN is a generative model and considers the inter-frame relationship of the video clip as a Bayesian relationship. The video clips of different classes are modeled with different IRBENs. For instance, the eating IRBEN($IRBEN_{eating}$) or noneating IRBEN($IRBEN_{n\_eating}$) are trained by using eating or noneating video clips, respectively. The class of the test video clip is predicted based on these two models.

3. Materials and Methods

The raw video data of UVDASSB were acquired from a commercial fish farm (Oriental Ocean Co. Ltd., Shandong, China). The video capture system is illustrated in Fig.1 four underwater cameras are spread at the edge of the tank from different directions and 0.5m above the bottom. The video data were captured from three tanks, each having around 1000 Atlantic salmons. The installation of the equipment was guided and supervised by the fish feeding experts to alleviate disturbing the activity of the fishes.
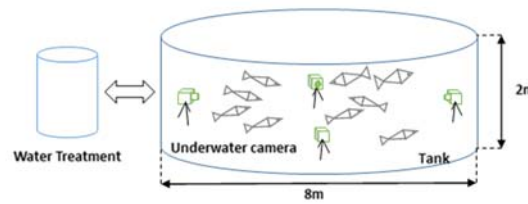
Figure 1. The schematic of experimental system.

The fishes are fed 3 times a day at 8:00, 12:40, and 20:00. During video recording, all of the feeding conditions keep in a production practice environment. An example of the video frame taken by the underwater camera is shown in Fig.2a.

The raw video sequences acquired by the video capture system have a total of 20269 minutes at the frame rate of 30fps with the resolution of 2704x1520. From the raw video sequences, we manually drew 3791 video clips, each 5 seconds, to construct the UVDASSB. Each video clip is resized to 540x304 and annotated as either eating or noneating under the guidance of the fish feeding experts. The video clips are divided into training set, validation set, and test set. In Fig.2, we show the examples of the fish eating video frames and fish noneating video frames of the UVDASSB.

(a)   (b)

Figure 2. Examples of the video frames in the UVDASSB. a) 15 frames of an eating video clip; b) 15 frames of a noneating video clip. The 15 video frames are uniformly sampled from the video clip of 150 frames.

*3.1. Inter-frame Relationship Bayesian Estimation Network Based Classification*

In a video sequence, there is usually a high correlation between the tth frame f_t and $(t+\tau)$th frame $f_{t+\tau}$ for a small interval τ, which is named inter-frame relationship in this paper. We observed that for different classes of the fish feeding behaviors, the inter-frame relationships can be described with different models Let $p_e(f_{t+\tau}|f_t)$ and $p_{ne}(f_{t+\tau}|f_t)$ denote the conditional probability density functions (PDFs) of the eating video clips and noneating video clips, respectively. It is however intractable to figure out these two conditional PDFs in the image space. Inspired by the Variational Auto-Encoder(VAE)[23], where the image is encoded as a multivariate Gaussian distribution(MGD) in a latent space, we encode the video frames as the MGD PDFs by a VAE encoder, and then model the relationship between the MGD PDF of the tth frame and the MGD PDF of the $(t+\tau)$th frame. Let $z_t=p(l|f_t) \sim N(l; \vec{\mu}_t, \vec{\sigma}_t^2)$ denote the MGD PDF of the tth frame. Here l, $\vec{\mu}_t$ and $\vec{\sigma}_t^2$ denote the latent space variable, the mean vector of z_t, and the variance vector of z_t, respectively. The concept is illustrated in Fig.3 Instead of figuring out $p(f_t|f_{t+\tau})$ in the image space, the relationship between f_t and $f_{t+\tau}$ is converted to a fully connected network (FCN) model between z_t and $z_{t+\tau}$ in the latent space.

The IRBEN structure is shown in Fig.4. The input is a video frame f_t. After encoded by the VAE encoder, f_t is converted to the MGD PDF z_t. Then, the FCN is employed to predict the MGD PDF of the $(t+\tau)$th frame according to z_t, i.e., $\hat{z}_{t+\tau} = fcn(z_t)$.

The training of the IRBEN includes two stages. In stage one, the VAE encoder is trained via a VAE network by using all of the training video frames. In stage two, for each class of fish feeding behavior, an FCN model is trained by using the MGD PDFs associated with that class of fish feeding behavior. For instance, in the experiments of this work, two classes of fish feeding behaviors, eating and noneating, are investigated. Therefore, two FCN models, fcn_{eating} and fcn_{n_eating}, are trained by using z_t's associated with the eating behavior and noneating behavior, respectively.
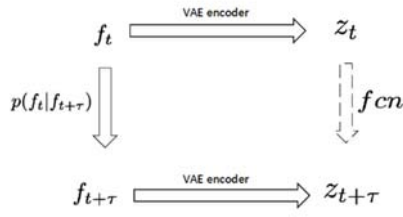
Figure 3. Converting the interframe relationship in the image space to the relationship between MGD PDFs in the latent space.
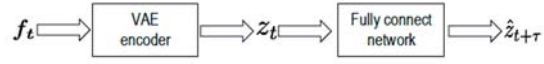

Figure 4. The structure of the IRBEN.

For a test video clip, the classification is achieved by comparing the prediction error of each FCN model for the $(t+\tau)$th frame. The prediction error is given by the distances between the MGD PDF generated by the VAE encoder and the MGD PDF predicted by the FCN model. For instance, for two classes: eating behavior and noneating behavior, the test video frame f_t and $f_{t+\tau}$ are first encoded as z_t and $z_{t+\tau}$ by the VAE encoder, and then, by inputting z_t, two FCN models, fcn_{eating} and fcn_{n_eating}, generate $\hat{z}_{t+\tau}^e$ and $\hat{z}_{t+\tau}^{ne}$ as the predictions of the MGD PDFs for the $(t+\tau)$th frame, respectively. The class of the test video clip is determined by comparing the distance between $z_{t+\tau}$ and $\hat{z}_{t+\tau}^e$ and the distance between $z_{t+\tau}$ and $\hat{z}_{t+\tau}^{ne}$.

In the following, we detail the IRBEN-based classification approach.

The VAE[23], as shown in Fig.5, consists of an encoding network and a decoding network. The encoding network converts the video frame f_t into a MGD PDF z_t in the latent space and the decoding network reconstructs the frame from z_t. The dimension of the latent space is a hyper-parameter and set to 50 in this paper.
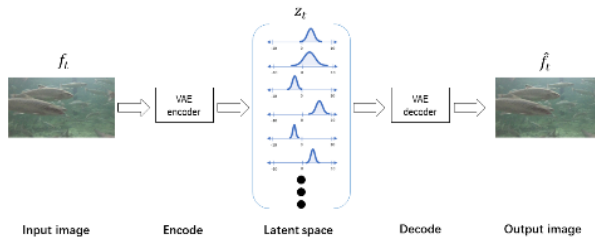


Figure 5. The VAE structure.
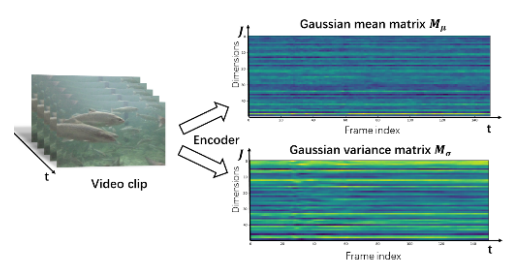
Figure 6. Illustration of converting a video clip into two matrices.

In the latent space, the video clip is represented as two matrixes: Gaussian mean matrix $M_\mu = [\vec{\mu}_0 ... \vec{\mu}_t ... \vec{\mu}_\delta]$ and Gaussian variance matrix $M_\mu = [\vec{\sigma}_0^2 \ \vec{\sigma}_t^2 ... \vec{\sigma}_\delta^2]$ as shown in Fig.6. The number of the rows of the matrixes equals 150 which are the length of the video clip and the number of the columns of the matrixes equals the dimension of the latent space. In our experiments, the size of each matrix is 150×50.

The structure of FCN is shown in Fig.7. The network input is z_t and the network output is $\hat{z}_{t+\tau}$. The FCN is trained by using the KL divergence between z_t and $\hat{z}_{t+\tau}$ as the loss function, which is defined as equation (1).

$$loss_t = D_{KL}[z_{t+\tau}||\hat{z}_{t+\tau}] = -\frac{1}{2}\sum_j^J[log\frac{\sigma_j^2}{\hat{\sigma}_j^2} - \frac{\sigma_j^2}{\hat{\sigma}_j^2} - \frac{(\mu_j-\hat{\mu}_j)^2}{\hat{\sigma}_j^2} + 1] \tag{1}$$

Where μ_j and σ_j are the mean and variance of the jth components of $z_{t+\tau}$, respectively, and $\hat{\mu}_j$ and $\hat{\sigma}_j$ are the mean and variance of the jth components of $\hat{z}_{t+\tau}$, respectively. Note that we suppose the dimension of the latent space is J.
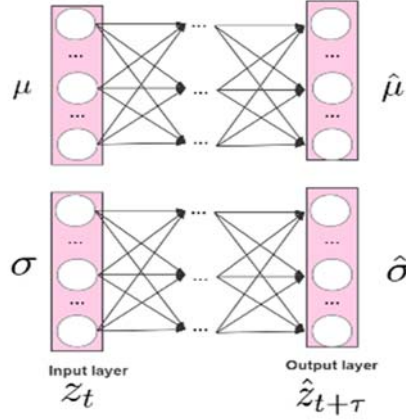
Figure 7. FCN structure.
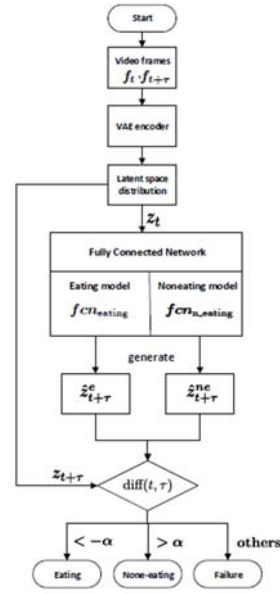
Figure 8. Diagram of the IRBEN-based classification on the UVDASSB.

In our implementation, instead of using the KL divergence for a specific frame, we employ an average of the KL divergences over the frames in a video clip as the loss function, which is defined as equation (2).

$$L_t = \frac{1}{\delta} \sum_{t=1}^{\delta} loss_t \tag{2}$$

Where δ is the length of the video clip. τ is set to 1 in this paper. With the average KL divergence, the model can take the video clip, instead of two video frames, as training samples.

The fish feeding behavior classification based on the IRBEN for the UVDASSB is outlined in Fig.8. First, two video frames, f_t and $f_{t+\tau}$, are encoded as z_t and $z_{t+\tau}$ through the VAE encoder, respectively. Then, two MGD PDFs $\hat{z}_{t+\tau}^e$ and $\hat{z}_{t+\tau}^{ne}$, are generated by *inputting* z_t into two FCN models, *fcn_eating* and *fcn_n_eating*, which are trained with the eating and noneating video clips, respectively. Finally, the class of the video clip is determined by the difference between the prediction errors of *fcn_eating* and *fcn_n_eating*, which is defined as equation (3).

$$diff(t,\tau) = D_{KL}[\hat{z}_{t+\tau}^e || z_{t+\tau}] - D_{KL}[\hat{z}_{t+\tau}^{ne} || z_{t+\tau}] \tag{3}$$

Where τ is the frame interval.

To improve the robustness, we adopt the error difference averaged over the frames in the video clip for the classification, which is given as as equation (4).

$$diff_\delta(\tau) = \sum_{t=0}^{\delta-\tau} diff(t,\tau) \tag{4}$$

The rule of the classification is defined as as equation (5).

$$prediction = \begin{cases} eating & diff_\delta < -\alpha \\ noneeating & diff_\delta > \alpha \\ fai & others \end{cases} \tag{5}$$

Where, $\alpha > 0$ is the minimum confidence distance which can be set according to the feeding strategy of the system. The robustness of the prediction increases with α.

3.2. Performance Evaluation
To assess the performance of the proposed method, the IRBEN is evaluated by accuracy, precision, recall, and specificity, which are defined as equation (6)-(9).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \qquad (6)$$

$$precision = \frac{TP}{TP+FP} \times 100\% \qquad (7)$$

$$recall = \frac{TP}{TP+FN} \times 100\% \qquad (8)$$

$$specificity = \frac{TN}{TN+FP} \times 100\% \qquad (9)$$

Where TP, FP, TN, and FN denote the number of true positives, false positives, true negatives, and false negatives, respectively. Here, we define the eating sample as the positive sample and the noneating sample as the negative sample. For the prediction of failure, the eating sample is counted into FN and the noneating sample into FP.

4. Results and Discussion

4.1. Video Frame Pre-processing

As mentioned above, the video of the UVDASSB is recorded under a very complicated luminous environment. That means the captured video frames have low SNR. Therefore, it is necessary to improve the frame quality by using some image pre-processing approaches. The underwater video frames suffer from poor visibility. Therefore, we enhance the luminance and increase the SNR of the video frames by scaling the blue, green, and red channels independently as equation (10).

$$p = \gamma v \qquad (10)$$

Where v is the original pixel value of a channel, p is the pixel value after processing, and γ is the scaling parameter. In our experiments, all samples in the UVDASSB are scaled with $\gamma=2$. In Fig.9 an example of the video frame after image preprocessing is shown.
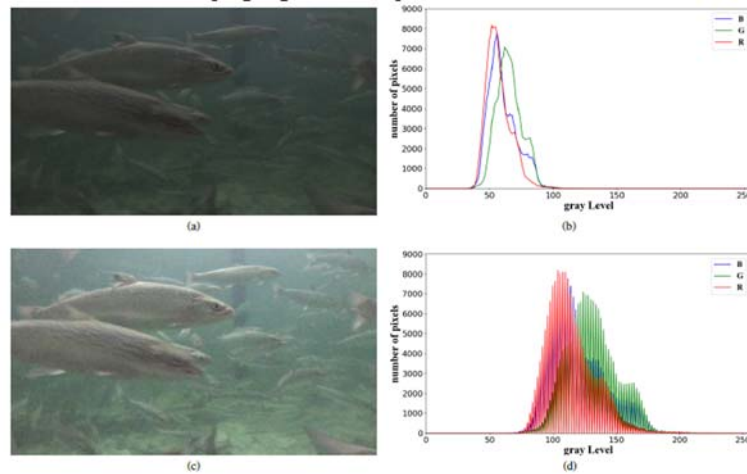

Figure 9. Example of image preprocessing with =2. a)Original video frame; b)Original video frame histogram of RGB; c)Pre-processed video frame; d)Pre-processed video frame histogram of RGB.

4.2. The Performance of Inter-frame Relationship Bayesian Estimation Network

In UVDASSB each video clip is encoded into two matrixes in the latent space and two FCN models, fcn_{eating} and fcn_{n_eating}, are trained with the matrixes associated with eating and noneating in the latent space, respectively. In the experiments, all the hyper-parameters of the proposed method are set as follow: J=50, δ=150, and τ=1. And the number of neurons in six layers of the FCN are 50, 60, 80, 80, 60 and 50. In Fig.10, we shown that after 1700 epochs of training, the validation losses of fcn_{eating} and fcn_{n_eating} converge to a stable value around 18.
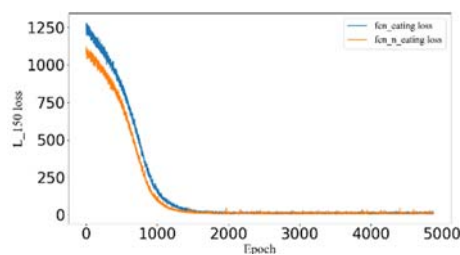
Figure 10. Training loss of fcn_{eating} and fcn_{n_eating} on validation set.
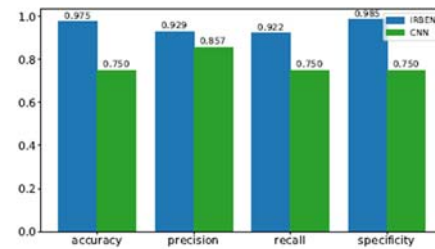
Figure 11. The performance comparison between IRBEN and CNN-based method.

In the proposed approach, a minimum confidence distance α is used to improve the robustness. The performances of the IRBEN with different α 's are evaluated. The results are tabulated in Table.1. It is obvious that all of the performance indicators go down as α goes up. When α =5, the IRBEN achieves the highest accuracy of 97.5% and it's considered as our baseline. Note that α has the greatest impact on precision, people should pay more attention to this indicator when applying the proposed approach.

Table 1. IRBEN performance under different α.

α	accuracy (%)	precision (%)	recall (%)	specificity (%)
5	97.5	92.9	92.2	98.5
6	94.1	79.7	83.3	95.2
7	88.2	62.4	77.8	90.0
8	79.4	44.0	70.5	81.2

4.3. Comparison Experiment

In this experiment, we compared the proposed approach with the convolutional neural network (CNN) based fish feeding behavior classification method proposed in [16]. The CNN-based method achieves a classification accuracy of 90% on an image dataset that is constructed from a simulation site of the RAS. In [16], the fish feeding behaviors are divided into none eating, weak eating, medium eating, and strong eating. To apply the CNN-based method to the UVDASSB, we consider the category none as the noneating behavior in the UVDASSB and the other three categories: weak eating, medium eating, and strong eating, as the eating behavior in the UVDASSB.

In order to train the network of the CNN-based method, an image dataset, which contains 4000 eating samples and 4000 noneating samples, is constructed from the video frames of the UVDASSB. 70%, 15%, and 15% of the images in the image dataset are used as the training set, validation set, and test set, respectively.

In Fig.11 the experimental results on the four performance indicators are shown. The IRBEN apparently reaches better performance. The poor performance of the CNN-based method on the UVDASSB can be contributed to the following facts: a) the image of UVDASSB is acquired from a commercial-scale RAS where the density of the cultured fish is high; b) the images in the UVDASSB are captured by the underwater cameras, which leads to the images of low SNR. In a high-density RAS, it is very difficult to identify whether the fish is feeding or not according to a single image, whereas the IRBEN considers multiple images so as to gain more information.

5. Conclusion

In order to improve the economic benefits of aquaculture, this paper proposes a classification algorithm of fish feeding behavior for automatic feeding control of fish called IRBEN. The IRBEN first employs a VAE encoder to convert the frames of a video clip into MGDs. Then, two fully connected networks, one is trained on the MGDs associated with the fish eating video clips and the other on the MGDs associated with the fish noneating video clips, are employed to predict the MGDs of the frame after an interval from the MGD of the current frame. The classification is conducted by

finding the fully connected network achieving smaller KL distance between the predicted MGD and the actual MGD.

The experimental results show that the IRBEN algorithm proposed in this paper achieves an accuracy rate of 97.5%. The classification algorithm proposed in this paper provides a prior knowledge for achieving a fully automatic bait feeding control system. Based on the classification results of fish feeding behavior, the automatic feeding control system can easily draw the control decision of feeding or not feeding, which will greatly reduce the waste of bait.

In the future work the UVDASSB will be further optimized and open source for all researchers, and the IRBEN algorithm will be further analysed and tested on the other fish species. In addition, in the future we plan to design a set of automatic feeding control methods based on the classification results of the IRBEN algorithm to achieve feedback-based fish feeding.

Acknowledgments

References
[1] FAO. The state of world fisheries and aquaculture 2018-meeting the sustainable development goals, 2018.
[2] Chao Zhou, Daming Xu, Kai Lin, Chuanheng Sun, and Xinting Yang. Intelligent feeding control methods in aquaculture with an emphasis on fish: a review. Reviews in Aquaculture, 10(4):975–993, 2018.
[3] Jian Zhao, Wei Jun Bao, Feng Deng Zhang, Zhang Ying Ye, Ying Liu, MingWei Shen, and Song Ming Zhu. Assessing appetite of the swimming fish based on spontaneous collective behaviors in a recirculating aquaculture system. Aquacultural Engineering, 78:196–204, 2017.
[4] John E Ehrenberg and Tracey W Steig. A method for estimating the "position accuracy" of acoustic fish tags. ICES Journal of Marine Science, 59(1):140–149, 2002.
[5] Chris Rillahan, Michael Chambers, W Huntting Howell, and Winsor H Watson III. A self-contained system for observing and quantifying the behavior of atlantic cod, gadus morhua, in an offshore aquaculture cage. Aquaculture, 293(1-2):49–56, 2009.
[6] Pallab Kumer Sarker, Dominique P Bureau, Katheline Hua, Murray D Drew, Ian Forster, Keith Were, Brad Hicks, and Grant W Vandenberg. Sustainability issues related to feeding salmonids: a c anadian perspective. Reviews in Aquaculture, 5(4):199–219, 2013.
[7] Ziyi Liu, Xian Li, Liangzhong Fan, Huanda Lu, Li Liu, and Ying Liu. Measuring feeding activity of fish in ras using computer vision. Aquacultural engineering, 60:20–27, 2014.
[8] Yousef Atoum, Steven Srivastava, and Xiaoming Liu. Automatic feeding control for dense aquaculture fish tanks. IEEE Signal Processing Letters, 22(8):1089–1093, 2014.
[9] Chao Zhou, Baihai Zhang, Kai Lin, Daming Xu, Caiwen Chen, Xinting Yang, and Chuanheng Sun. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. Computers and Electronics in Agriculture, 135:233–241, 2017.
[10] Chao Zhou, Kai Lin, Daming Xu, Lan Chen, Qiang Guo, Chuanheng Sun, and Xinting Yang. Near infrared computer vision and neuro-fuzzy modelbased feeding decision system for fish in aquaculture. Computers and electronics in agriculture, 146:114–124, 2018.
[11] Boaz Zion and Assaf Barki. Ranching fish using acoustic conditioning: Has it reached a dead end? Aquaculture, 344:3–11, 2012.
[12] Huanyu Liu, Lihong Xu, and Dawei Li. Detection and recognition of uneaten fish food pellets in aquaculture using image processing. In Sixth International Conference on Graphic and Image Processing (ICGIP2014), volume 9443, page 94430G. International Society for Optics and Photonics, 2015.
[13] Jiaheng Cao and Lihong Xu. Research on counting algorithm of residual feeds in aquaculture based on machine vision. In 2018 IEEE 3rd International Conference on Image, Vision and

Computing (ICIVC), pages 498–503. IEEE, 2018.

[14] Jian Zhao, Zhaobin Gu, Mingming Shi, Huanda Lu, Jianping Li, Mingwei Shen, Zhangying Ye, and Songming Zhu. Spatial behavioral characteristics and statistics-based kinetic energy modeling in special behaviors detection of a shoal of fish in a recirculating aquaculture system. Computers Electronics in Agriculture, 127:271–280, 2016.

[15] Sonia Duarte, Lourdes Reig, and Joan Oca. Measurement of sole activity by digital image analysis. Aquacultural Engineering, 41(1):22–27, 2009.

[16] Chao Zhou, Daming Xu, Lan Chen, Song Zhang, Chuanheng Sun, Xinting Yang, and YanboWang. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. Aquaculture, 507:457–465, 2019.

[17] Lin Meng, Takuma Hirayama, and Shigeru Oyanagi. Underwater-drone with panoramic camera for automatic fish recognition based on deep learning. IEEE Access, 6:17880–17886, 2018.

[18] Hongwei Qin, Xiu Li, Jian Liang, Yigang Peng, and Changshui Zhang. Deepfish: Accurate underwater live fish recognition with a deep architecture. Neurocomputing, 187:49–58, 2016.

[19] Cigdem Beyan, Vasiliki-Maria Katsageorgiou, and Robert B Fisher. Extracting statistically significant behaviour from fish tracking data with and without large dataset cleaning. IET Computer Vision, 12(2):162–170, 2017.

[20] Shuo Hong Wang, Jing Wen Zhao, and Yan Qiu Chen. Robust tracking of fish schools using cnn for head identification. Multimedia Tools and Applications, 76(22):23679–23697, 2017.

[21] XU Zhiping and Xi En Cheng. Zebrafish tracking using convolutional neural networks. Scientific reports, 7:42815, 2017.

[22] Jing Hu, Daoliang Li, Qingling Duan, Yueqi Han, Guifen Chen, and Xiuli Si. Fish species classification by color, texture and multi-class support vector machine using computer vision. Computers and electronics in agriculture, 88:133–140, 2012.

[23] Jason Tyler Rolfe. Discrete variational autoencoders. stat, 1050:7, 2016.