IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 8, 2020, accepted July 16, 2020, date of publication July 29, 2020, date of current version August 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3012873

Research on Recognition of Motion Behaviors of Copepods

ZHENGRUI SHI, LUJIE CAO, YU HAN, HAIXING LIU, FENGSHOU JIANG, AND YU REN
Department of Mechanical and Electrical Engineering, College of Engineering, Ocean University of China, Qingdao 266100, China

Corresponding author: Lujie Cao (lujiecao@ouc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41376140 and Grant 51276124.

ABSTRACT The motion behaviors of copepods has important scientific research value and there is very little research on recognition of their motion behaviors simultaneously. Recognition of the basic motion behaviors of copepods using deep learning methods can greatly reduce the time cost of distinguishing and statistics, as well as achieve the purpose of improving efficiency. Based on the characteristics of motion of copepods that bring challenges to the extraction of motion fragments from raw video and the establishment of data set, such as instantaneous moving, static status most time, small-scale and high-frequency, this article propose an improved Camshift algorithm for detection of moving targets to overcome these challenges and establish the motion behaviors image acquisition system and a standard data set of motion behaviors, which provides the experience and methods of marine zooplankton behaviors database. Finally, the LRCN network that combines the advantages of CNN and LSTM is adopted to study the impacts of different factors on the model performance, such as the number of frames of sample, preprocessing operations and sample dimensions. Experimental results show that the LRCN network has excellent potential in classification of motion behaviors of copepods, when the number of frames of sample reaches 7, the precison rate, recall rate, f1-score are 0.96, 0.95, 0.95, respectively. In addition, the rise in number of frames and preprocessing has a positive effect on the recognition, the 4D samples (image sequence) is more suitable for the LRCN model than 3D samples (trajectory image).

INDEX TERMS Copepods, motion behaviors, data set, deep learning, LRCN network.

I. INTRODUCTION

Nowadays, the studies and applications of deep learning in marine zooplankton mostly stay in the field of image recognition of species, and there is little research on their behavior recognition. Meanwhile, the research subjects of behavior recognition mainly focus on humans or large livestock [1]–[6]. The algorithms of behavior recognition are divided into two schools: machine learning and deep learning. Machine learning algorithms locate and describe the key points such as Haar-like, Hog, LBP, SIFT, etc., pre-process and cluster the features, and achieve the purpose of classification [7]–[13]. Deep learning methods have an end-to-end extraction of features from raw data for classification, mainly include 3D CNN, RNN, LSTM and other deep neural networks [14]–[17]. Due to the continuous efforts of researchers and the improvement of computational capacity of computer, deep learning algorithms have become more mature and gained fast proliferation from the field of behavior recognition. In deep learning algorithm, the most famous is the CNN network and the researchers have improved the CNN to the 3D-CNN or three-stream CNNs to recognize behaviors and obtained good results [18], [19]. In 1997, Hochreiter and Schmidhuber [20] proposed the long short-term memory network (LSTM) that belongs to the time loop network. It is suitable for sequence data and widely used to process speech, text, images, and video [21]. Peng *et al.* [22] constructed a Recurrent Neural Network (RNN) composed of a Long Short Term Memory (LSTM) unit for the data of Inertial Measurement Unit (IMU) to classify seven cattle behavior patterns, results show that the LSTM-RNN model had better classification capacity than the CNN model, especially in social behaviors and mobility. Majd and Safabakhsh [23] proposed a deep network based the LSTM unit called C^{2}LSTM that utilize convolution and corresponding operators to analysis the spatial and motion structure of video sequence, as well as applied the deep network for classification of human behavior patterns. According to the experiment of the two classical

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

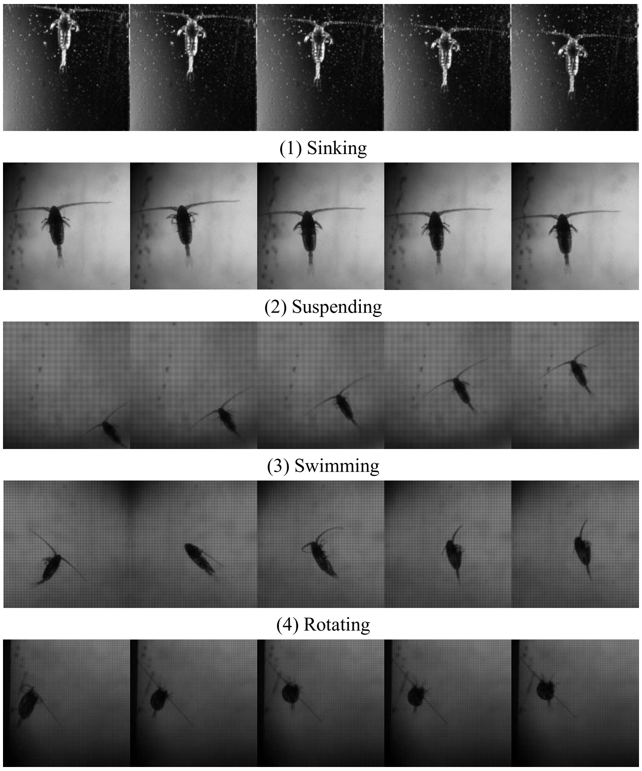
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

benchmarks: UCF101 and HMDB51, the results show that C^{2}LSTM can acquire effectively spatial features and time dependencies.

The Long Term Recurrent Convolutional Network (LRCN) model combines the advantages of CNN model and LSTM model and is widely used in video description, behavior recognition and image recognition [24]. The single-value prediction and sequence output capabilities of LRCN networks can handle single-frame or video sequences. Gautam *et al.* [25] utilized the LRCN network based on transfer learning called MyoNet to classify the lower limb motions and forecast the knee angle simultaneously. Bhuvaneshwari and Manjunathan [26] resorted to the LRCN algorithm to classify different gesture images. Kim *et al.* [27] optimized the LRCN model to prevent nuclear power plant from military or violent invasion by identifying five types of intruder behavior patterns.

The training basis of motion behaviors recognition of copepods is the establishment of data set, which depends on the detection of motion and preprocessing. The general preprocessing flow of behavior data set includes background subtraction, Gaussian blur, morphological operations, down-sampling, etc. To solve the problem of motion detection, a more robust tracking system that combines Camshift and Optical flow is usually be adopted by researchers. Meanwhile, the Gaussian mixture model is a common method to build a visual surveillance system that model each pixel as a GMM model, and an online approximation method are applied to update the system parameters to track moving targets.

As the analysis of the literature mentioned above, we intend to design a system can identify and track multiple objects of interest through the method of Camshift that are suitable for relative still environments, and establish a motion behaviors data set after fully considering the uniqueness of zooplankton motion. In addition, through summarizing the experience of previous study, the LRCN network is used for classification of 5 motion behavior patterns of copepods.



(1) Sinking

(2) Suspending

(3) Swimming

(4) Rotating

(5) Reversing

FIGURE 1. Five motion behaviors of copepods.

II. MOTION BEHAVIORS AND DATA ACQUISITION

A. BIOLOGICAL SAMPLES AND DEFINITIONS OF MOTION BEAHVIORS OF COPEPODS

The copepods studied in this article mainly consist of *Calanus sinicus* and *Paracalanus parvus*, which are native to the Yellow Sea, China. The Biological samples were obtained by dragging vertically a dense fishing net at the Zhongyuan Wharf in the Southern District of Qingdao in March 2019.

Moison *et al.* [28] divided copepod motion patterns into five behaviors according to the gravity coordinate system and the direction of motion: break, forward swimming, backward swimming, sinking and upward swimming. This article refers to previous experience, the five motion behaviors of ‘sinking’, ‘suspending’, ‘swimming’, ‘rotating’ and ‘reversing’ are studied. As shown in Figure 1, the following five behaviors are described based on kinematic parameters. Under the effect of gravity, the copepod keeps its motion organs relatively still or only swings its antennae, the motion behavior of the body gradually sinking is called ‘sinking’. ‘Suspending’ means that the copepod can gain upward power by continuously drawing appendages to overcome its own gravity, and the spatial coordinates remain basically unchanged or float slightly. ‘Swimming’ behavior is the action of the copepod paddling the appendages so that water flow to generate forward power and a speed of forward motion for it, moving from one place to another. ‘Rotating’ behavior is the motion of copepod rotating around the axis from head to tail. ‘Reversing’ behavior is the motion that the copepod twists its torso and paddles appendages to obtain torque and change its direction.

B. DATA ACQUISITION SYSTEM

The experimental image acquisition system is mainly composed of an optical path system and an operation & storage system. The collection of zooplankton motion behaviors image samples is different from humans and large animals based on the these challenges or characteristics: (1) Zooplanktons are phototaxis, which may interfere with the sampling process, but sampling requires a light source. (2) Zooplanktons represented by copepods were not insensitive to red light. (3) Zooplanktons have a micro-scale body size, which requires high resolution and wide depth of field cameras for sampling. (4) Zooplanktons have a high frequency during motion but do not move most of the time.

Considering these first three points, the optical path system is composed of a CCD digital camera, a collimated red low-power laser light source to prevent light interference and a long focal length beam expansion lens to get a wide depth of field, as shown in Figure 2.

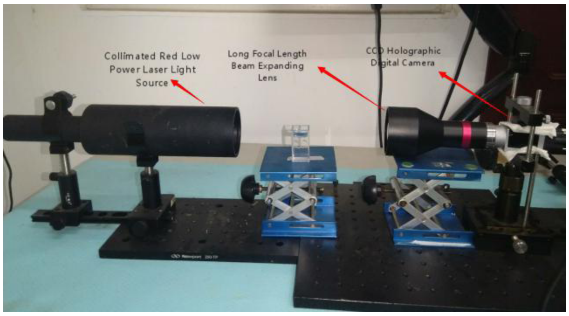

Collimated Red Low Power Laser Light Source
Long Focal Length Beam Expanding Lens
CCD Holographic Digital Camera

FIGURE 2. Experimental sampling optical path system of motion behaviors of copepods.

To solve the third challenge of small-scale of copepods, the setting of image resolution is significant. One organism needs at least 1 pixel to display and the larger *Calanus sinicus* can reach 4000 μm [29], the spatial resolution adopted by the system should not be less than 1 pixel/4000 μm. According to Nyquist Theorem—the signal sampling frequency of the system should not be lower than twice the signal frequency, so it is not less than 2 pixel/4000 μm. In the actual sampling, this value should be at least 5$\sim$10 times the theoretical value, so s $<$ 200 μm/pixel, the resolution used in this article is in the range of 5 μm/pixel$\sim$20 μm/pixel, much higher than the required resolution.

In view of the fourth characteristic and the fact that frame rate is an important parameter of the original video sampling system. The higher the frame rate, the more comprehensive motion information can be obtained, such as the position and posture at different moments. Referring to the measured results in the study and the previous researches, it is concluded that the average swimming speed of copepods is about 3$\sim$6 mm/s and the general body length is about 2 mm. In order to maintain the continuity of motion recording, at least two frames are recorded within a body length during motion, so the frame rate is about 6 fps (2*1/(2 mm/6 mm/s)) [30]. According to Nyquist Theorem, the frame rate should be at least 12 fps. In actual sampling, the minimum frame rate is usually 5$\sim$10 times, the paper select 50 fps to 100 fps. The total number of frames that the CCD camera can record is 1000, so the observation time can last at least 10 seconds.

III. STANDARD DATA SET

A. MOTION DETECTION

The process of building a standard data set is shown in Figure 3. As mentioned above, considering that copepods are almost static most of the time, filtering automatically out some useless and static video fragments can reduce time consumption of selection. To solve this problem, the tracking target algorithm—Camshift algorithm [31] is selected to remove those video fragments where targets have no motion over a certain period of time. The Camshift algorithm detect the motion of the object and extract the video fragment with moving objects. After tracking the targets, these video fragments, which are longer than a certain frames but have no any motion, are eliminated.

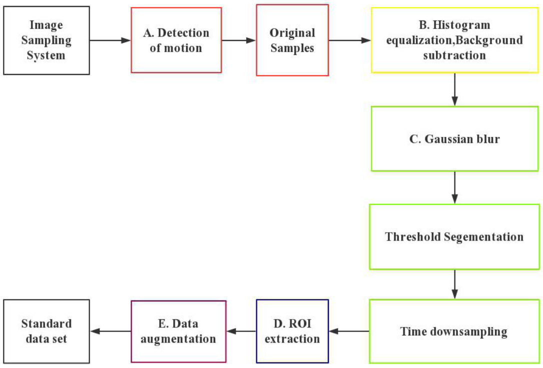


FIGURE 3. The flow of building a standard data of motion behaviors of copepods.

It is worth noting that the original Camshift is not suitable for this study, so we need to improve it slightly. Because of the high frame rate of video, the time interval between two adjacent frames is small, and the displacement is small when the copepod has a certain speed. However, the small displacement does not mean that there is the motion of copepod, because it may also be caused by accidental factors,such as instrument or camera vibration. If the copepod continues to move for a period of time, the larger displacement can be obtained by superimposing these smaller displacements. Therefore, the frame by frame detection is unscientific and inefficient, we detect two end frame of one video to determine whether one video has moving objects. However, the displacement of some motions, such as ‘suspending’, even the time is enough long, may not be very large. Since this article does not involve precise mathematics analysis, it only make qualitative analysis.

$$f(x) = \sum_{k=1}^{k=2} \alpha_k \frac{\frac{1}{\sigma_k}\phi(\frac{x-\mu_k}{\sigma_k})}{\Phi(\frac{b-\mu_k}{\sigma_k}) - \Phi(\frac{0-\mu_k}{\sigma_k})}, \quad 0 < x < b$$

$$\alpha_1 + \alpha_2 = 1; \quad \mu_1 = 0; \; \mu_2 = \frac{a+b}{2}; \tag{1}$$

According to the common sense of probability, in a still water environment, the displacement/distance (x) of fixed time interval in [0 mm, a mm]∪[a mm, b mm] present the two mixed truncated normal distribution as shown in the formula (1). Furthermore, the motion probability and the displacement present an step function relationship in a video fragment with fixed number of frames, but considering these previous point mentioned above, it may need to be modified at the initial scope (a mm), but the formula (2) is generally satisfied. $\phi(\cdot)$ is the standard normal function, $\Phi(\cdot)$ is the cumulative distribution function

$$p(x) = \begin{cases} 0, & x \leq 0 \\ \dfrac{1}{a}x, & 0 < x < a \\ 1, & a \leq x \leq d \end{cases} \tag{2}$$

Therefore, the algorithm increase the detection density for the video sequence section with larger displacement and reduce relatively the detection density for the video sequence section with smaller displacement in a whole video. In this article, the algorithm of number of frame of video fragment detected is set to two dimensions: level (l) and number of video fragment divided in video fragment of the previous level, i.e., density (k), in order to ensure the accuracy and effectiveness of the detection, set the two lower detection length threshold (the number of frames of single video fragment) and introduce reward and penalty mechanism.

In the first level, the whole video is divided into k_1 fragments, this level has no penalty and only rewards. In other words, when entering the 2nd level, the k value of these fragments, will add 1, whose displacement is greater than the displacement threshold setting in 1st level, and the others remain unchanged. After the 2nd level, the penalty and reward are co-existed and brought into the next level, the algorithm keeps the detection of fragment until $k = 1$ or the second lower detection length threshold is met. With the iterations, the number of frames of detected video fragment gradually decreases, if the first lower detection length threshold is satisfied, it will stop the detection of those intermediate fragments with displacement (i.e. >threshold of displacement) whose fragments on both sides have displacement (i.e. >threshold of displacement). When the second lower detection length threshold is satisfied, it will stop detecting the video fragments with displacement (i.e. >threshold of displacement), only detect these fragments (<threshold of displacement) whose nearest fragments have a displacement (i.e. >threshold of displacement), from the nearest to itself.

At last, the algorithm outputs those fragments considered to have displacement, fuses the fragments meeting the requirements as a whole video sequence that their interval between two ends is less than the 1/10 of the number of frames of the longer fragment, from the left-most frame to the right-most frame. The algorithm flow is shown in Figure 4.

Considering that the two frames to be detected by the mechanism mentioned are far apart not frame by frame, the Camshift algorithm may fail, we initialize the search box size to the entire image size for determining the new centroid in new frame, and move one vertex of the current search box along the direction and distance of vector from the centroid of the previous frame to the current centroid, and reduce box size simultaneously. If the moving distance is greater than the set threshold, recalculate the adjusted box centroid and perform a new round of box position and size adjustment until the moving distance of the box center and the centroid is less than the threshold, or the number of iterations reaches a certain maximum. After the convergence condition is satisfied, we keep reducing the current search box size until the error between the grayscale average and variance of search boxes of the current frame and the previous frame is less than the threshold, the search for this frame is ended, the working process diagram of search box is shown in the Figure 5. If the video contains multiple targets, we can utilize the common Kalman+Camshift algorithm to track multiple objects, and finally extract video samples with moving target from raw video.

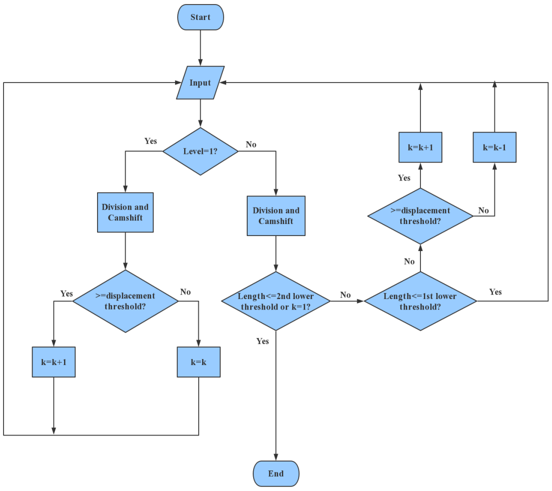

Start
Input
Level=1?
Yes
No
Division and Camshift
Division and Camshift
>=displacement threshold?
Yes
No
k=k+1
k=k
Length<=2nd lower threshold or k=1?
No
Yes
Length<=1st lower threshold?
Yes
No
>=displacement threshold?
Yes
No
k=k+1
k=k-1
End

FIGURE 4. The algorithm flow of motion detection of whole video.

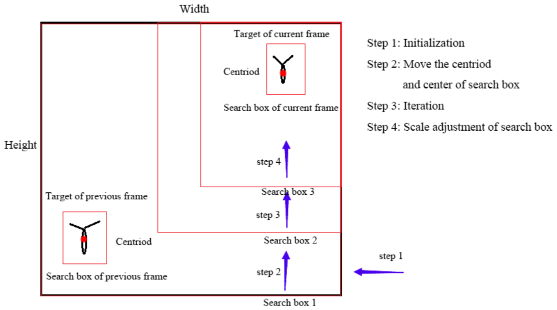


FIGURE 5. The working process diagram of search box in the improved Camshift.

We take 50 groups of original videos, a total of 109 motion fragments in experiment to test the accuracy of the algorithm, the experimental results are shown in Table 1. Frame error represents the error ratio of the number of frames between the detected and true. The working example of the algorithm is shown in Figure 6. Experimental results show that the algorithm proposed can detect copepods motion well.

B. STANDARD HISTOGRAM EQUALIZATION AND BACKGROUND SUBTRACTION

These RGB images with a size of 1280 × 1024 pixels are converted to grayscale images for reducing the computational burden of training and highlight key elements required for machine learning. Some copepods are too close to the

TABLE 1. Experimental results of detection algorithm.

Frame error	<10%	10%~30%	30%~50%	>50%	total
Number	83	11	7	8	109

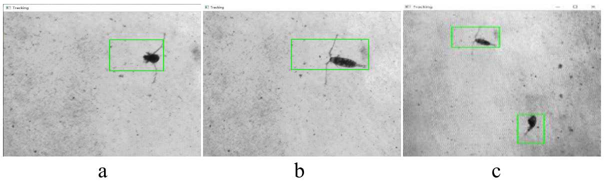


FIGURE 6. The algorithm work demonstration of motion detection of -single object-a&b; mutiple objects-c).

background grayscale due to their own translucency, the standard histogram equalization is used to divide the image into two parts so that the image contrast is opened. However, the histogram equalization changing the original grayscale structure may cause information loss, so this operation cannot be applied to all images. Next, unnecessary background elements be eliminated through Gaussian mixture background modeling [32].

C. GAUSSIAN BLUR

Gaussian blur and threshold segmentation (Otsu algorithm) are performed for each image. Because the laser equipment in this experiment sometimes does not provide a high uniform light source, and the grayscale distribution of some images is not enough uniform, a single threshold is not suitable for some images. These images can use local threshold segmentation method. In addition, there are some small objects such as moving impurities in the water. In order to obtain a clearer image, we perform morphological operations to eliminate environmental noise and highlight the subject, and then down-sample in the time dimension to obtain samples of different number of frames (T).

D. ROI DETECTION

After performing boundary detection on each image to calculate the boundary of the target or obtaining the maximum connected region of target, we offset the origin of x-y coordinates to build an image boundary box around the biological boundary and the range of motion and perform cropping to obtain region of interest (ROI), the images are normalized to the fixed resolution of 128 × 128 pixels by Gaussian pyramid downsampling.

E. DATA AUGMENTATION

In this article, 530 original video samples of five types of motion behaviors are obtained through experiments and image processing. After downsampling in the time dimension, data augmentation is performed. Data augmentation includes image geometry transform and other methods, such as rotation, cropping, mirroring, scaling and Gaussian noise, the distribution of samples of motion behaviors of copepods is shown in Table 2.

TABLE 2. Distribution of the samples of motion behaviors of copepods.

	Reversing	Rotating	Sinking	Suspending	Swimming
Raw video samples	60	60	150	110	150
T=3	600	600	1500	1100	1500
T=5	600	600	1500	1100	1500
T=7	600	600	1500	1100	1500
T=9	600	600	1500	1100	1500

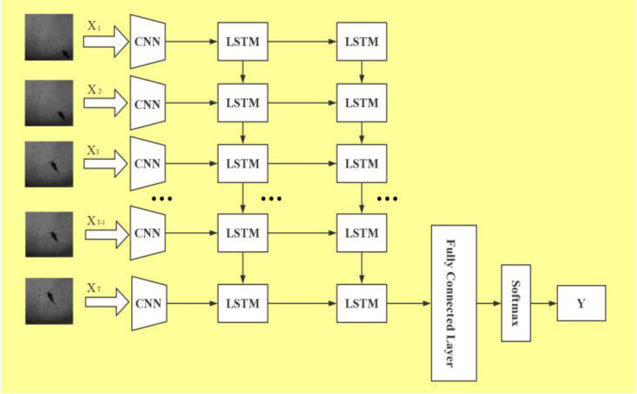


FIGURE 7. The structure of the LRCN network adopted in this paper.

IV. EXPERIMENTS NETWORKS AND SETTING

In this article, the long-term recurrent convolutional network (LRCN) is used to identify the motion behaviors of copepods. The entire structure of the long-term recurrent convolutional network is divided into two parts including CNN network and LSTM network, as shown in Figure 7, the CNN part on the left and the LSTM part on the right. The CNN mainly extracts features from frames and the LSTM deal with the time sequence. The length input of LRCN is consisted of T consecutive frames. Because the samples to be predicted have to involve time dimension, the input with five dimensions (i.e., number of samples, time step, width, length, channel) are fed into entire network. Moreover, the time distributed layer is applied to solve the problem that the CNN is not suitable for sequence input. While training the CNN model, a flattening layer is added at the end of the CNN model to change the 4D tensor into the 2D tensor that conforms to the input of LSTM.

The CNN part is constructed of the convolutional layers, pooling layers as well as dropout layer in turn and alternately. In view of the facts that the filter size should not be too large to facilitate computation and several small convolution kernels is better than a single large convolution kernel, hence we apply a convolutional filter with a size of 3 × 3. Meanwhile, we utilize zero padding, stride 1 as convolution parameter, ReLU function as the hidden layer nonlinear activation function. The filter size of max pooling is 2 × 2. In order to learn the 64 final feature maps with a size of 4 × 4, a five-layer convolution structure is designed for the CNN module. The first convolution layer inputs images of 128 × 128, outputs feature maps with a size of 64 × 64 ×8, and the second convolution layer outputs feature maps with a size of 32×32 ×16. The output dimension is reduced, the third layer is 16 × 16 ×32, the fourth layer is 8 × 8 ×64, and the fifth layer is 4 × 4 ×64.

The LSTM part of the model consists of two hidden LSTM layers and a fully connected layer for single value prediction. The T continuous vectors are fed into the first LSTM layer that has 1024 units. Furthermore, the first LSTM layer is a many-to-many structures to realize sequence-to-sequence learning corresponding to the sequence of the second layer. The second layer with 256 units is a many-to-one structure to deal the relationship of sequence to one output. At last, the second LSTM layer produces an output for a fully connected layer, which is connected to the softmax layer that output a vector of size 5 corresponding to the probability of 5 types of copepod behaviors.

The Keras framework is used as the experimental operating environment. We utilize the NVIDIA GTX1080 graphics card, one 4T hard disk, one 16G DDR4 memory module and one 4-cores Intel i7700 CPU as experimental hardware environment. In this paper, the data set is randomly divided into three groups for avoid overfitting: 80% training set, 10% validation set and 10% test set.

We select cross entropy as the loss function, because cross entropy is affected by the difference between the prediction and the actual when using gradient descent method rather than mean squared error and it has better convergence characteristics for the general classification task. For the activation function, all hidden layers adopt the ReLU activation function and the output layer resort to softmax as a classifier to complete the classification task. The experiment adopt the widely used mini batch gradient descent method, use the Adam optimizer for optimization. The batch size is 64, the learning rate is 0.0001, the dropout rate is set to 0.2, the L1-L2 regularization is used to prevent overfitting, and the training period is set to 300 epochs. Finally, the samples are input into the network for experiment and the parameters of the network are shown in Table 3.

TABLE 3. Network model structure and parameters.

Main structure	Output Shape	Param #
Conv1+ MaxPool1+ Dropout1	(None, 5, 64, 64, 8)	80
Conv2+ MaxPool2+ Dropout2	(None, 5, 32, 32, 16)	1168
Conv3+ MaxPool3+ Dropout3	(None, 5, 16, 16, 32)	4640
Conv4+ MaxPool4+ Dropout4	(None, 5, 8, 8, 64)	18496
Conv5+ MaxPool5+ Dropout5	(None, 5, 4, 4, 64)	36928
Flattening	(None, 5, 1024)	0
LSTM1+Dropout6	(None, 1024)	8392704
LSTM2+Dropout7	(None, 256)	1311744
Fully connected layer	(None, 64)	16488
Softmax	(None, 5)	325
Total		9782533

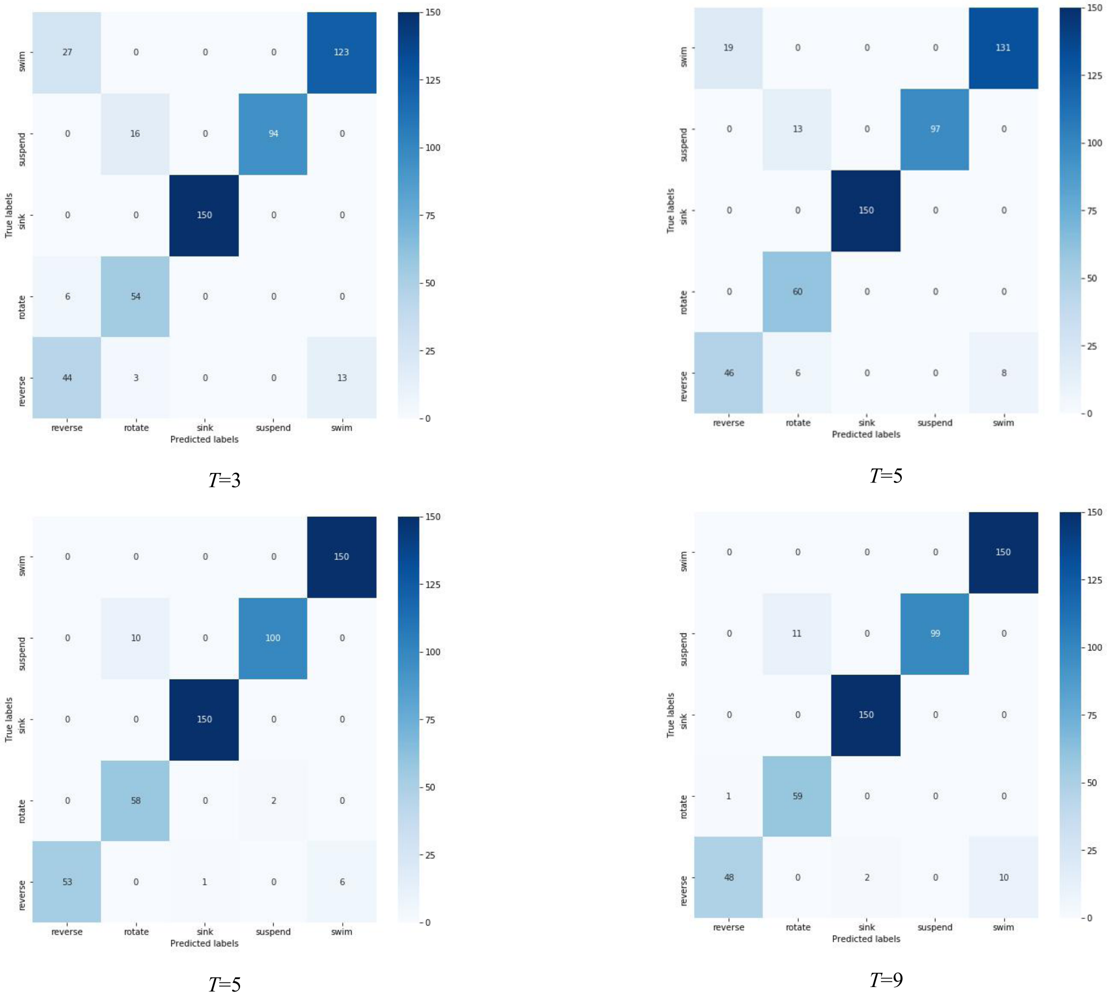
V. EXPERIMENT OF CRUCIAL FACTORS IN THE LRCN NETWORK

A. NUMBER OF FRAMES OF SAMPLE

In order to test the impact of the number of frames of a single sample (T) on the motion behaviors recognition of LRCN model, the experiment was divided into four groups according to the number of frames of sample, i. e., $T = 3$, $T = 5$, $T = 7$ and $T = 9$. The data used for training and testing the model is manually labeled and the training set is fed into the LRCN network for training. After the experiment, the evaluation metrics is shown in Table 4 and the confusion matrix shown in Figure 8. The comparison found that with the increase of the number of frames, the overall evaluation metrics of the model has been significantly improved, reaching the maximum at $T = 7$, and the average F1 score is 0.95. Experimental results show that the rise in T makes contribution to improving the performance of the model by increasing the spatial and temporal features available for learning. It is worth noting that after T reaches a certain amount, the model performance is limited by the factors of model such as depth and structure, and the data itself, the positive effect of the rise in the number of frames on the model performance will be slowed down and even decrease slightly. The main reason for the slight decline in the evaluation metrics after $T = 7$ is that the rise in the number of frames leads to some samples with insufficient frames. Those samples with sufficient number of frames are augmented with geometry transform to compensate for the disappeared samples due to too few frames, which will cause the overall sample diversity to decrease.

B. PREPROCESSING

In order to study the impact of preprocessing operations such as background subtraction, the samples are divided into two categories, one without any preprocessing operation—only its size scale is normalized so that the samples can be input into the network; the other one with preprocessing operation including background subtraction through the Gaussian mixture background modeling method, downsampling, etc. The number of frames is set to 7, the two types of training sets are fed into the LRCN network model for training. The resulting confusion matrix is shown in



T=3

T=5

T=5

T=9

FIGURE 8. Confusion matrix of the experiment of number of frames of sample.

TABLE 4. Evaluation metrics of the experiment of number of frames of sample.

T=3	Reversing	Rotating	Sinking	Suspending	Swimming	Average
Precision	0.57	0.74	1.00	1.00	0.90	0.84
Recall	0.73	0.90	1.00	0.85	0,82	0.86
F1	0.64	0.81	1.00	0.92	0.86	0.85
T=5						
Precision	0.71	0.76	1.00	1.00	0.94	0.88
Recall	0.76	1.00	1.00	0.88	0,87	0.90
F1	0.73	0.96	1.00	0.94	0.90	0.89
T=7						
Precision	1	0.85	0.99	0.98	0.96	0.96
Recall	0.88	0.97	1.00	0.91	1.00	0.95
F1	0.94	0.91	1.00	0.94	0.98	0.95
T=9						
Precision	0.98	0.84	0.99	1.00	0.94	0.95
Recall	0.80	0.98	1.00	0.90	1.00	0.94
F1	0.88	0.90	0.99	0.95	0.97	0.94

Figure 9 and the evaluation metrics are shown in Table 5. Experimental results show that the average F1 score without preprocessing is only 0.67, because the lack of necessary preprocessing operations can not remove unnecessary background element and highlight the information to be learned.

Through the experimental results, it is found that a series of preprocessing can improve the recognition potency of the

TABLE 5. Evaluation metrics of the experiment of preprocessing operations.

	Reversing	Rotating	Sinking	Suspending	Swimming	Average
Precision	0.42	0.54	1.00	0.96	0.65	0.71
Recall	0.37	0.83	1.00	0.44	0.80	0.69
F1	0.39	0.66	1.00	0.60	0.71	0.67

TABLE 6. Evaluation metrics of the experiment of sample dimensions.

	Reversing	Rotating	Sinking	Suspending	Swimming	Average
Precision	0.53	0.73	1.00	1.00	0.70	0.79
Recall	0.67	0.92	1.00	0.54	0.80	0.78
F1	0.59	0.81	1.00	0.70	0.75	0.77

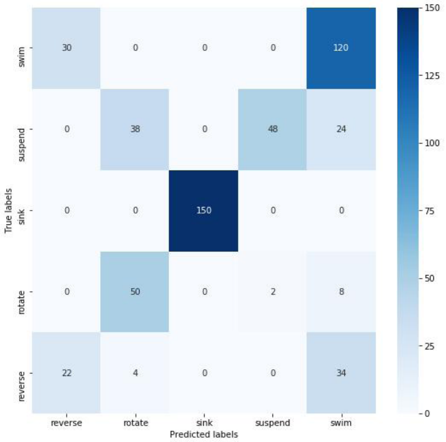


FIGURE 9. Confusion matrix of the experiment of preprocessing operations (image background subtraction,ROI, etc.,).

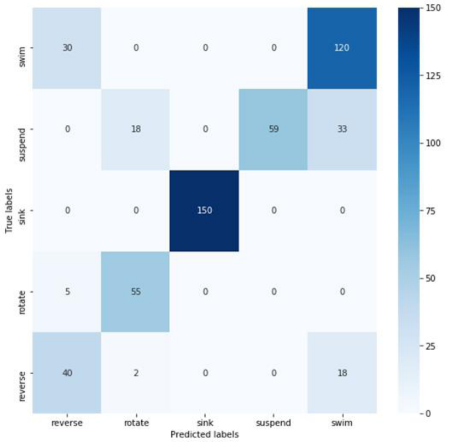


FIGURE 11. Confusion matrix of the experiment of sample dimension.

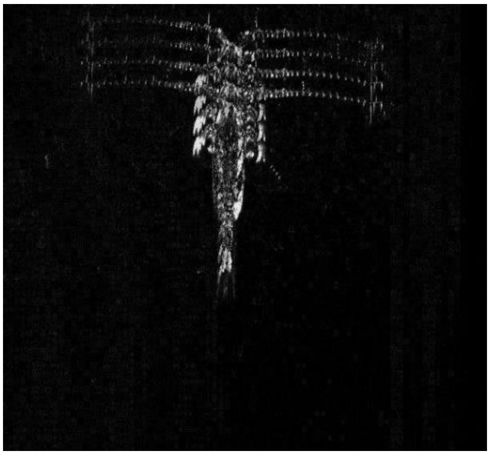


FIGURE 10. 3D samples (trajectory image) of sinking.

model, remove a large amount of information and features unrelated to the recognition target, strengthen the ability of selection and learning of useful information and features of the model, and reduce the data dimension. The data dimension of natural pictures obey a certain distribution, minus the common parts such as background, preprocessing can highlight the differences and characteristics between sample individuals, as well as prevent the occurrence of overfitting.

C. SAMPLE DIMENSION

The frame images in the original sequence samples are superimposed into an trajectory image to form a 3D sample, that is, length, width and channel, as shown in Figure 10. Considering that the image superimposed by too many frames is blurred, the recognition effect is not good, so the number of frames of samples (T) to be superimposed sets to 5. Finally, the input dimensions of the LRCN model in this article are modified to facilitate the input of single-frame trajectory images. The confusion matrix of the experimental results is shown in Figure 11, the evaluation metrics are shown in Table 6, and the average F1 score is 0.77. The model performs well, but the evaluation metrics have decreased relative to original sequence sample experiment ($T = 5$). This is because the trajectory image samples are stacked by original frame images, the features in time domain such as speed direction are difficult to extract, and some spatial features such as body details are blurred due to superimposition of different body parts.

VI. CONCLUSION

This article firstly discusses the establishment of standard data motion behaviors of copepods from the image sampling system to subsequent image processing, in particular, the proposed motion detection algorithm performs an important role in acquiring of raw video fragments of motion. Next, the impact of different factors on the accuracy of the LRCN model is studied by three experiments. Experimental results shows that the increase of the number of frames of sample can improve the recognition ability of the model, but after a certain number of times ($T = 7$), the growth rate slows down or even becomes negative, because the increase in the number of frames leads to the less effective samples, for example, a 50-frame video of a certain motion behaviors, there are 10 sets of samples of 5 frames, but there are only 5 sets of sample of 9 frames. Hence the data is supplemented by other means but highly repeatable (geometry augmentation or supplement the number of frames with previous or later frame images irrelevant to motion). In addition, preprocessing has a significant and positive impact on recognition, because it eliminate the unnecessary element for recognition. At last, the 3D trajectory image samples have a relatively high recognition effect, but their overall performance is not as good as 4D sequence image samples for the LRCN network.

In the future, we hope to collect more data in the ocean and expand the species to be recognized from only copepods to other marine zooplankton. Moreover, we intend to develop an academic software or applications with multifunctional purposes, such as species identification, movement parameter measurement and behavior recognition, which can facilitate the use of scientific researchers.

REFERENCES

[1] M. Riekert, A. Klein, F. Adrion, C. Hoffmann, and E. Gallmann, ''Automatically detecting pig position and posture by 2D camera imaging and deep learning,'' *Comput. Electron. Agricult.*, vol. 174, Jul. 2020, Art. no. 105391.

[2] W. Li, Z. Dou, and L. Qi, ''Communication protocol classification based on LSTM and DBN,'' *IEEE Access*, vol. 8, pp. 91818–91828, 2020.

[3] X. Qin, Y. Ge, J. Feng, D. Yang, F. Chen, S. Huang, and L. Xu, ''DTMMN: Deep transfer multi-metric network for RGB-D action recognition,'' *Neurocomputing*, vol. 406, pp. 127–134, Sep. 2020.

[4] H. C. Kang, ''Cow horn detection system based and on behaviors pattern recognition, has approval detection sensor for transmitting detected approval information to database server, and approval detection sensor for detecting estrused information of cow,'' U.S. Patent KR 211,092 B1, May 29, 2020.

[5] B. Achour, M. Belkadi, R. Aoudjit, and M. Laghrouche, ''Unsupervised automated monitoring of dairy cows' behavior based on inertial measurement unit attached to their back,'' *Comput. Electron. Agricult.*, vol. 167, Dec. 2019, Art. no. 105068.

[6] A. Nasirahmadi, B. Sturm, S. Edwards, K.-H. Jeppsson, A.-C. Olsson, S. Müller, and O. Hensel, ''Deep learning and machine vision approaches for posture detection of individual pigs,'' *Sensors*, vol. 19, no. 17, p. 3738, Aug. 2019.

[7] R. Rastgoo, K. Kiani, and S. Escalera, ''Hand sign language recognition using multi-view hand skeleton,'' *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113336.

[8] K. P. Sanal Kumar and R. Bhavani, ''Human activity recognition in egocentric video using HOG, GiST and color features,'' *Multimedia Tools Appl.*, vol. 79, nos. 5–6, pp. 3543–3559, Feb. 2020.

[9] V. Kráger, D. Kragic, A. Ude, and C. Geib, ''The meaning of action: A review on action recognition and mapping,'' *Adv. Robot.*, vol. 21, no. 13, pp. 1473–1501, Apr. 2012.

[10] J. K. Aggarwal and M. S. Ryoo, ''Human activity analysis: A review,'' *ACM Comput Surv.*, vol. 43, no. 3, pp. 1–47, Apr. 2011.

[11] M. Rabbani, Y. L. Wang, R. Khoshkangini, H. Jelodar, R. Zhao, and P. Hu, ''A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing,'' *J. Netw. Comput. Appl.*, vol. 151, Feb. 2020, Art. no. 102507.

[12] I. Laptev, ''On space-time interest points,'' *Int. J. Comput. Vis.*, vol. 64, nos. 2–3, pp. 107–123, Sep. 2005.

[13] G. Willems, T. Tuytelaars, and L. Gool, ''An efficient dense and scale-invariant spatio-temporal interest point detector,'' in *Proc. EECV*, 2008, pp650-663.

[14] T. Ozcan and A. Basturk, ''Transfer learning-based convolutional neural networks with heuristic optimization for hand gesture recognition,'' *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8955–8970, Dec. 2019.

[15] S. A. R. Abu-Bakar, ''Advances in human action recognition: An updated survey,'' *IET Image Process.*, vol. 13, no. 13, pp. 2381–2394, Nov. 2019.

[16] R. O. Oyeleke, C. K. Chang, and J. Margrett, ''Situation–centered goal reinforcement of activities of daily living in smart home environments,'' *Expert Syst.*, vol. 37, no. 1, pp. 1–25, Feb. 2020.

[17] K. Simonyan and A. Zisserman, ''Two-stream convolutional networks for action recognition in videos,'' in *Proc. NIPS*, 2014, pp. 1–7.

[18] B. Mishra, D. Garg, P. Narang, and V. Mishra, ''A hybrid approach for search and rescue using 3DCNN and PSO,'' *Neural Comput. Appl.*, Jun. 2020, doi: 10.1007/s00521-020-05001-7.

[19] L. Wang, ''Three-stream CNNs for action recognition,'' *Pattern Recognit. Lett.*, vol. 92, pp. 33–40, Apr. 2017.

[20] S. Hochreiter, and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp1735-1780, Nov. 1997.

[21] J. Zhang, F. Chen, and Q. Shen, ''Cluster-based LSTM network for short-term passenger flow forecasting in urban rail transit,'' *IEEE Access*, vol. 7, pp. 147653–147671, 2019.

[22] Y. Peng, N. Kondo, T. Fujiura, T. Suzuki, S. Ouma, Wulandari, H. Yoshioka, and E. Itoyama, ''Dam behavior patterns in japanese black beef cattle prior to calving: Automated detection using LSTM-RNN,'' *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105178, doi: 10.1016/j.compag.2019.105178.

[23] M. Majd and R. Safabakhsh, ''Correlational convolutional LSTM for human action recognition,'' *Neurocomputing*, vol. 396, pp. 224–229, Jul. 2020, doi: 10.1016/j.neucom.2018.10.095.

[24] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, ''Long-term recurrent convolutional networks for visual recognition and description,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 677–691, Apr. 2017.

[25] A. Gautam, M. Panwar, D. Biswas, and A. Acharyya, ''MyoNet: A transfer-learning-based LRCN for lower limb movement recognition and knee joint angle prediction for remote monitoring of rehabilitation progress from sEMG,'' *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–10, Feb. 2020.

[26] C. Bhuvaneshwari and A. Manjunathan, ''Advanced gesture recognition system using long-term recurrent convolution network,'' in *Proc. ICONEEEA*,2019, pp. 1–8.

[27] S. H. Kim, S. C. Lim, and D. Y. Kim, ''Intelligent intrusion detection system featuring a virtual fence, active intruder detection, classification, tracking, and action recognition,'' *Ann. Nucl. Energy*, vol. 112, pp. 845–855, Feb. 2018.

[28] M. Moison, F. G. Schmitt, S. Souissi, L. Seuront, and J.-S. Hwang, ''Symbolic dynamics and entropies of copepod behaviour under non-turbulent and turbulent conditions,'' *J. Mar. Syst.*, vol. 77, no. 4, pp. 388–396, Jun. 2009.

[29] S. Uye, ''Temperature-dependent development and growth of Calanus sinicus (Copepoda, Calanoida) in the laboratory,'' *Hydrobiologia.*, vol. 167, pp. 285–293, Oct. 1988.

[30] G. Chardon, A. Leblanc, and L. Daudet, ''Plate impulse response spatial interpolation with sub-nyquist sampling,'' *J. Sound Vib.*, vol. 330, no. 23, pp. 5678–5689, Nov. 2011.

[31] H. Wang, S. K. Nguang, and J. Wen, "Robust video tracking algorithm: A multi-feature fusion approach," *IET Comput. Vis.*, vol. 12, no. 5, pp. 640–650, Aug. 2018.

[32] G. S. Morrison, "A comparison of procedures for the calculation of forensic likelihood ratios from acoustic–phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model–universal background model (GMM–UBM)," *Speech Commun.*, vol. 53, no. 2, pp. 242–256, Feb. 2011.

ZHENGRUI SHI was born in Anhui, China, in 1995. He received the B.S. degree in energy and power engineering from the Central South University of Forestry and Technology, Changsha, China, in 2017, and the M.S. degree in power engineering from the Ocean University of China, Qingdao, China, in 2020.

From 2017 to 2020, he worked as a Student Research Assistant at the College of Engineering, Ocean University of China. His research interests include image processing, deep learning, and big data. He has participated in the National Natural Science Foundation (research on the coupling relationship between feeding behaviors of marine zooplankton and marine physical factors based on holographic technology and pattern recognition) and the National Natural Science Foundation, these topics are mainly to study the interaction of kinematics of marine zooplankton and ocean flow. He is the main author of one national inventions and one international conference paper published, in 2019.

Dr. Shi is a member of the Institute of Marine Engineering, Science and Technology (IMarEST). He was awarded three school scholarships, from 2017 to 2020. He received one software copyright invention.

LUJIE CAO was born in Xinjiang, China, in 1970. She received the B.S. degree in optical engineering from Tianjin University, Tianjin, China, in 1992, the M.S. degree in fluid dynamics from Beijing University, Beijing, China, in 1999, and the Ph.D. degree in mechanical engineering from the State University of New York, Buffalo, USA, in 2007.

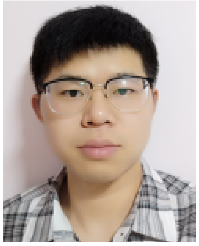
She is currently an Associate Professor with the Department of Mechanical and Electrical Engineering, College of Engineering, Ocean University of China. Her research interests include biotechnology, zooplankton behavioral ecology, turbulent flow, fluid dynamics, and aerosol optical diagnostic techniques. She has participated in the Natural Science Foundation of China. She is the author of three articles indexed by SCI(E), eight national invention patent of China, and a Reviewer of some famous journals, such as *Ship Engineering*.

Dr. Cao is a member of IMarEST.

YU HAN was born in Jiangsu, China, in 1996. He received the B.S. degree in new energy science and engineering from the Huaiyin Institute of Technology, Huaiyin, China. He is currently pursuing the M.S. degree with the Ocean University of China.

His research interests include biotechnology, deep learning, computational fluid dynamics, and multiple objects tracking. He has participated in the Natural Science Fund Project.

Mr. Han is a member of IMarEST. He was a scholarship recipient for excellent course grades and volunteer work in public service in the Ocean University of China, from 2016 to 2019. He is one of the authors of one National Software Copyright.



HAIXING LIU was born in Henan, China, in 1994. He received the B.S. degree in building environment and energy application engineering from Zhengzhou University, Zhengzhou, China, in 2017. He is currently pursuing the M.S. degree with the Ocean University of China.

His research interests include machine learning, deep learning, and computational fluid dynamics.

Mr. Liu is a member of IMarEST. During his time in the Ocean University of China, he received excellent course grade results and participated in the Natural Science Foundation. He is one of the authors of one National Software Copyright.

FENGSHOU JIANG was born in Shandong, China, in 1996. He received the B.S. degree in energy and power engineering from the Qingdao University of Technology, Qingdao, China, in 2019, where he is currently pursuing the M.S. degree.

His research interests include image recognition, computational fluid dynamics, and ocean engineering.

Mr. Jiang is a member of IMarEST. During his time in the Ocean University of China, he received excellent course grade results and participated in the Natural Science Foundation. He is one of the authors of National Software Copyright.

YU REN was born in Shanxi, China, in 1994. He received the B.S. degree in energy and power engineering from Shanxi University, Taiyuan, China, in 2016. He is currently pursuing the M.S. degree with the Ocean University of China.

His research interests include computational fluid dynamics, marine biology, image processing, and big data.

Mr. Ren is a member of IMarEST. During his time in the Ocean University of China, he received two patents for utility model and participated in the Natural Science Foundation. He is one of the authors of one National Software Copyright.