Journal of Korea Multimedia Society Vol. 25, No. 7, July 2022(pp. 903-911)
https://doi.org/10.9717/kmms.2022.25.7.903

Shrimp Quality Detection Method Based on YOLOv4

Tao Xingyi[†], Feng Yiran[††], Lee Eung-Joo[†††], Tao Xueheng[††††]

ABSTRACT

A shrimp quality detection model using YOLOv4 deep learning algorithm is designed, which is superior in terms of network architecture, data processing and feature extraction. The shrimp images were taken and data expanded on their own, the LableImage platform was used for data annotation, and the network model was trained under the Darknet framework. Through comparison, the final performance of the model was all higher than other common target detection models, and its detection accuracy reached 93.7% with an average detection time of 47 ms, indicating that the method can effectively detect the quality of shrimp in the production process.

Key words: Deep Learning, Convolutional Neural Networks, Shrimp, YOLOv4, Target Detection

1. INTRODUCTION

In recent years, China's shrimp production has gradually increased, and the peeling of shrimp balls from whole shrimp is an important part of its primary processing, and identifying whether the shrimp is qualified or not is an important part of it [1-2]. At present, the identification of shrimp quality is generally done manually, which takes a lot of time and labour costs and has a higher possibility of misjudgement. Research on quality detection based on the degree of curling of shrimps has not been reported, but target detection techniques have been used in the food sector, where deep learning-based target detection techniques have demonstrated their superiority in many fields [3-7] and are favoured in food processing and inspection [8-10]. Song Chao [11] used unsupervised K-means combined with support vector machines in a traditional method and a deep learning method to detect egg cracks, and the results showed that the traditional method had poor environmental adaptability, and the defect recognition rate based on deep learning. Fan et al. [12] used a method based on candidate defect region counting and support vector machines to detect defects in apple appearance online, and its detection accuracy was much lower than that of the deep learning method. Yin Hongpeng et al. [13] found that deep learning is the best visual recognition method today, and has more room for development than traditional methods.

YOLOv4 is a recently proposed deep convolutional neural network with higher detection efficiency. To ensure the efficiency of the shrimp detection model, the paper proposes to design a

※ Corresponding Author : Lee Eung-Joo, Address: (48520) Sinseon-ro, Nam-gu, Busan, Korea, TEL : +82-51-629-1143, FAX : +82-51-629-1143, E-mail : ejlee@tu.ac.kr
※ Corresponding Author : Tao Xueheng, Address: (116 034) No.1 Qinggongyuan, Ganjingzi District, Dalian, China, TEL : +82-411-86323773, FAX : +82-411-86323773, E-mail : xhtao@dlpu.edu.cn
Receipt date : Jun. 22, 2022, Approval date : Jul. 18, 2022
[†] School of Management, Liaoning university of International Business and Economics, China
(E-mail : taohouse@163.com)
[††] Dept. of Information and Communication Engineering, Tongmyong University, Korea
(E-mail : fengyr@dlpu.edu.cn)
[†††] Dept. of Information and Communication Engineering, Tongmyong University, Korea
(E-mail: ejlee@tu.ac.kr)
[††††] School of Mechanical Engineering and Automation, Dalian Polytechnic University, China
(E-mail: xhtao@dlpu.edu.cn)
※ This work was supported by the Dalian Young Stars of Science and Technology Project (NO.2021RQ088)

shrimp quality inspection method based on YOLOv4 deep learning network, by making various forms of shrimp data images such as adding noise and rotation to shrimp images, using YOLOv4 network model to learn and train the data set, using industrial CCD cameras to collect and detect the quality category of shrimp online, verifying that the classification of the model The accuracy and detection speed of the model can meet the production requirements and can improve the efficiency of shrimp production.

2. TECHNICAL METHODS

The production of shrimp balls requires that the shrimps are cooked to a semi-cooked state, when the shells are bright red and the curling angle <90° is qualified shrimps. The quality of the shrimps is judged by their curling characteristics.

Shrimp inspection is done in 3 parts: image acquisition, image processing and shrimp sorting. The detection system model is shown in Fig. 1. The use of industrial cameras to capture images of the shrimp production line, industrial light sources help to obtain clearer images of the production, which helps to improve the accuracy of the system detection.

The trained YOLOv4 model is transplanted to an industrial computer and used for real-time inspection of the shrimp production line; shrimp sorting is carried out by a robotic arm or other sorting device to pick out the detected unqualified shrimp.

(a) (b)

Fig. 1. Shrimp quality classification. (a) Qualified and (b) Unqualified.

2.1 YOLOv4 Network Framework

YOLO network is a classification algorithm that uses regression networks to achieve target detection. Compared with traditional region candidate networks [14-15], YOLO network integrates two phases of candidate region generation and detection, and treats the detection task directly as a regression problem, so it has faster detection speed and shows good results in most target detection tasks, and can achieve The overall network structure of YOLOv4 is shown in Fig. 2.

The YOLOv4 target detection network uses the CSPDarknet53 as the backbone and contains five CSP modules with 3×3 convolutional kernels in front of each module and a step size of 2 to further enhance the learning capability of the network; the Path Aggregation Network (PANet) is used as the neck, with additional Spatial Pyramid Pooling (SPP) modules with 1×1, 5×5, 9×9 and 13×13 maximum pooling. The pooling method can increase the perceptual area and separate more important contextual features; the YOLOv3 detection head is used as the head. yOLOv4 uses Mosaic and CutMix data enhancement strategies to increase the input image variability and enrich the image feature information, and the designed target detection model can obtain higher robustness, while label smoothing, learning rate cosine annealing attenuation and other techniques are used to optimize the network training process [17].

2.2 Detection Principle

The YOLO network divides the input shrimp image into an N×N grid. The grid is responsible for detecting the shrimp target if the centre coordinates of the shrimp to be measured fall in one of the grids. In the detection process, each grid cell predicts B bounding boxes, and each bounding box contains five prediction values: t_x, t_y, t_w, t_h and the confidence level, which reflects the confidence level and prediction accuracy of the predicted target box [18]. The centre coordinates (b_x, b_y) and the

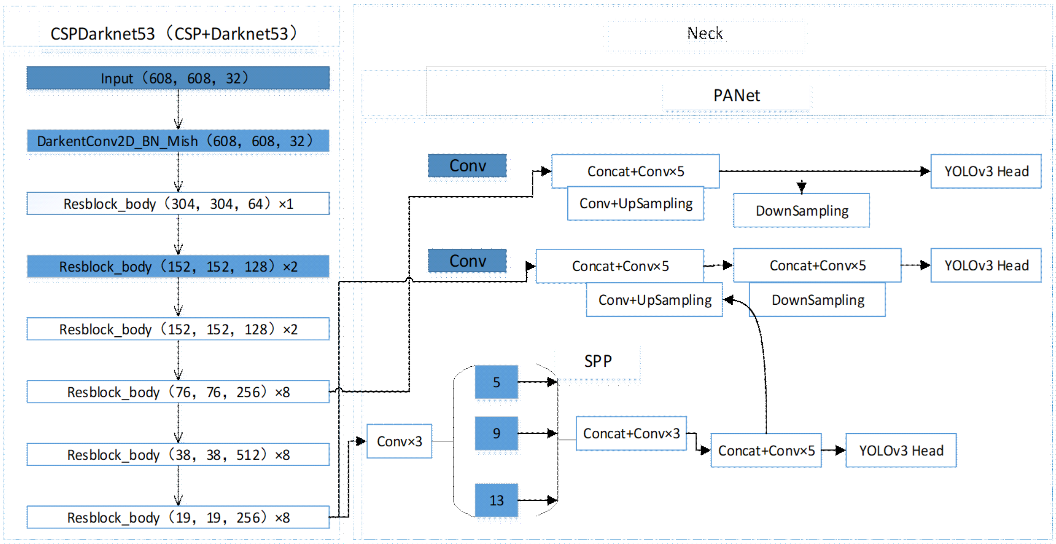


Fig. 2. Yolov4 network structure diagram.

width and height b_w, b_h of the predicted boxes are calculated according to equation (1).

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases} \tag{1}$$

where

$\sigma(x)$ – the logistic function.

c_x, c_y – the coordinates of the upper left corner of each grid in the feature map.

p_w, p_h – the width and height of the a priori box relative to the feature map.

t_x, t_y, t_w, t_h – the coordinates of the centre of the model prediction and the width and height.

The centre coordinates and width of the predicted box are divided by the corresponding feature image size and multiplied by the original input image size to obtain the actual value of the predicted bounding box coordinates relative to the original image of the shrimp.

Each grid still needs to predict one category of information, recorded as category C, so the output is a tensor of size $N \times N(5 \times B + C)$. At the same time, the detection system calculates the cross-merge ratio according to equation (2).

$$IoU = \frac{D_R \cap G_T}{D_R \cup G_T} \tag{2}$$

where

IoU – Intersection over Union.

D_R – Prediction box.

D_T – Ground truth box.

IoU is an important indicator of the accuracy of the prediction frame; the larger the IoU, the more accurate the location of the prediction frame.

The specific prediction process for shrimp is shown in Fig. 4. The YOLOv4 algorithm first resets the input shrimp image to 608×608 size; uses the CSP-Darknet53 network to extract the image features; sends the feature vectors to SPP and PANet for prediction; and applies the non-maximum value suppression algorithm to eliminate repeated predictions to obtain the final prediction results.

2.3 Loss Function

The loss function for the YOLOv4 network training consists of 3 parts: regression loss L_{ciou}, confidence loss L_{conf}, and classification loss L_{class}. If no target exists in a certain bounding box, only the confidence loss is calculated, and if a target exists, the 3 losses are calculated. The expression of the loss function is:

$$L_{ciou} = \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{obj}[1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\frac{16}{\pi^4}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^4}{1 - IoU + \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2}] \quad (3)$$

$$L_{conf} = -\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{obj}[\widehat{C_i^j} log(C_i^j) + (1 - \widehat{C_i^j})\log(1 - C_i^j)] - \lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{obj}[\widehat{C_i^j} log(C_i^j) + (1 - \widehat{C_i^j})\log(1 - C_i^j)] \quad (4)$$

$$L_{class} = -\sum_{i=0}^{S^2} I_{i,j}^{obj} \sum_{c \in classes} [\widehat{P_i^j} log(P_i^j) + (1 - \widehat{P_i^j})\log(1 - P_i^j)] \quad (5)$$

$$L_{oss} = L_{ciou} + L_{conf} + L_{class} \quad (6)$$

where

S^2, B – feature map scales and a priori boxes.

λ_{noobj} – the weighting factor.

$I_{i,j}^{obj}, I_{i,j}^{noobj}$ – if there is a target at the jth a priori frame of the ith grid, take 1 and 0 respectively, if there is no target, take 0 and 1 respectively.

$\rho(\cdot)$ – the Euclidean distance.

c – the diagonal distance between the predicted box and the closed region of the actual box.

b, w, h – the centre coordinates and width and height of the prediction box.

b^{gt}, w^{gt}, h^{gt} – coordinates of the centre of the actual box and its width and height.

$C_i^j, \widehat{C_i^j}$ – the confidence level of the predicted frame and the labeled frame.

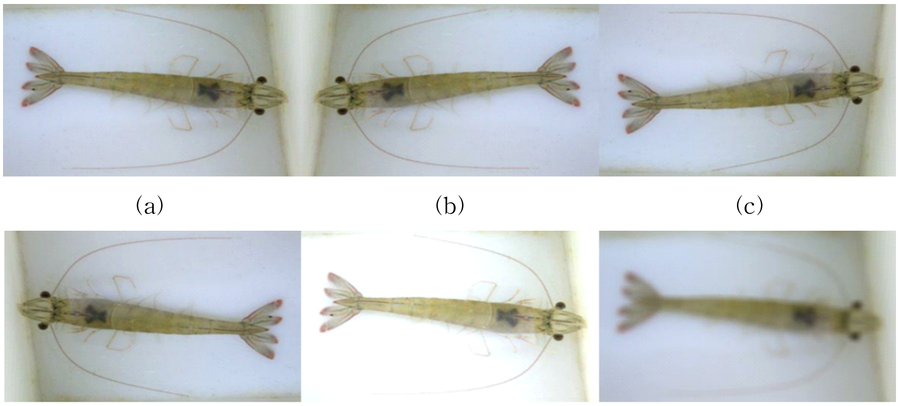
$P_i^j, \widehat{P_i^j}$ – the category probabilities of the prediction frame and the labeled frame.

Confidence loss and classification loss are calculated by the cross-entropy method, and the bounding box regression loss is calculated by the CIoU loss function. Compared with the traditional mean square error loss function, CIoU effectively avoids the problem of being sensitive to the scale of the target object and can better focus on the relationship between the position of the predicted box and the actual box, reflecting the connection between the two.

3. DATA PRE-PROCESSING

3.1 Data Collection

The dataset used for the experiments is taken and annotated by the autonomy. To make the images taken by the autonomy have higher robustness, the dataset images are subjected to dataset augmentation operations and random scale changes, cropping, flipping, and rotation of the images to in-



(a) (b) (c)

(d) (e) (f)

Fig. 3. Image preprocessing. (a) Original picture, (b) horizontal flip, (c) vertical flip, (d) vertical and horizontal flip, (e) Gaussian blur, and (f) enhanced contrast.

crease the variability of the input images, so that the training model has higher robustness to the images taken from different scenes. The pre-processed images are shown in Fig. 3.

A total of 10,716 images of shrimps of different sizes, shapes and colours were obtained from the data expansion. There are two types of shrimp samples in the picture data, with a total of 4986 qualified samples and 5730 unqualified samples, so that the size and richness of the data can be guaranteed.

3.2 Data Labeling

The data set needed to be annotated before training the YOLOv4 model, using the, LableImage data annotation tool, setting the qualified shrimp category to 0 with a label of qualified and the unqualified category to 1 with a label of unqualified, saving the text file in YOLO data format by default after annotation, and the sample labelling information for both categories of shrimp is shown in Fig. 4.

The content of the text file saved after data annotation is shown in Table 1. Each row in the table represents the location information of a shrimp target, the label represents the category of the specific target annotation, and x, y, w, h are floating point numbers from 0 to 1, which are normalized values relative to the scale of the whole image. Eighty percent of the annotated images were randomly divided as the training set and 20% as the test set. As shown in Table 1.

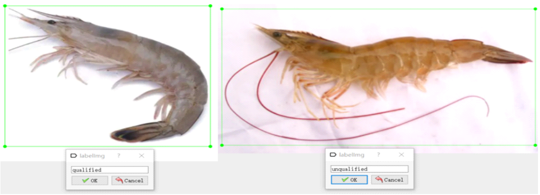


Fig. 4. Sample marker chart.

Table 1. Dataset labeling.

Category	x	y	w	h
0	0.783	0.380	0.215	0.148
1	0.177	0.391	0.336	0.159

x and y represent the x-axis and y-axis coordinates of the centre of the marker box; w and h represent the width and height of the marker box respectively.

4. TESTING AND ANALYSIS

4.1 Test Configurations

Deep learning network model training usually requires high training platform configuration. YOLOv4 network can be trained on CPU or GPU, and GPU is chosen for its low training cost due to its much higher computing power than CPU. The labelled and segmented shrimp dataset was used as the training sample, and the specific configuration information of the YOLOv4 network model training platform is shown in Table 2.

The training parameters of the network learning model are: the input image size is 608×608, the total number of input images for each iteration is 64, the training is divided into 8 batches, the momentum value is 0.9, the weight decay factor is 0.0005, the maximum number of iterations is 10 000, the initial learning rate is 0.001, and the mosaic The data augmentation strategy was used, and the learning rate started to decay at 8,000 and 9,000 iterations.

4.2 Model Assessment

Loss value The loss value is the error between the predicted value and the true value of the sample, and is calculated by the loss function; the smaller the loss value, the better the prediction effect. A complete training log is kept during the training of the network, and the loss value is plot-

Table 2. Training platform configuration.

Name	Configurations
Operating system	ubuntu 16.04.06
Framework	Pytorch
CPU	Intel i7-9700K
GPU	NVIDIA GeDorce RTX 2080S

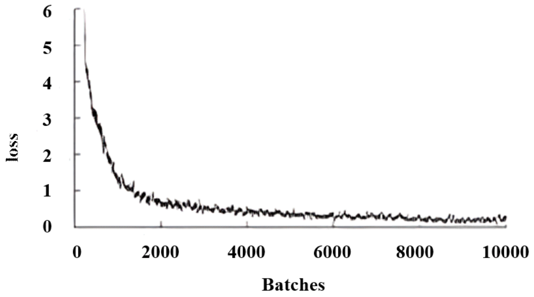


Fig. 5. Loss value curve.

ted visually according to the log information after the training is completed, and the loss value line graph is shown in Fig. 5.

As can be seen from Fig. 5, the loss value decreases as the number of iterations increases, and after 10,000 iterations, the loss value is generally <0.4 and fluctuates around 0.35, proving that the network model is well trained.

Test results The mean average precision (mAP) is an important measure of target detection efficiency and is determined by the precision and recall rates. The area under the P-R curve is called the mean accuracy value. The larger the value of the mean accuracy value for all target categories, the more effective the neural network model is.

The trained network model was used for shrimp test set prediction and the precision (P) and recall (R) were calculated separately according to equation (7).

$$\begin{cases} R = \dfrac{T_P}{T_P + F_N} \\ P = \dfrac{T_P}{T_P + F_P} \end{cases} \tag{7}$$

where

T_P - the positive class judged to be a positive class.

F_P - negative class judged to be a positive class.

F_N --- positive class judged to be a negative class.

The test set contained a total of 2489 qualified shrimp samples and 2366 unqualified shrimp samples. As can be seen from Table 3, there were a few false positive samples for both categories of shrimps, but the precision and recall rates for each category were >93% and the model predicted better performance.

Comparative analysis After training and testing with the YOLOv4 model on the shrimp dataset, the configuration information of the training platform remained unchanged, and the Faster RCNN [19], EfficientDet [20] and YOLOv3 deep learning models were used for training and analysis on the same dataset, and the mean accuracy means and frames per second (FPS) of the different models were compared in Table 4.

As can be seen from Table 4, compared with the FasterRCNN with ResNet101 [21] as the backbone, the EfficientDet with EfficientNet-B0 as the back-

Table 3. Test results of two types of shrimp.

Prediction Results of Qualified Products		Prediction Results of Unqualified Products		Accuracy Rate/%		Recall/%	
qualified	unqualified	qualified	unqualified	qualified	unqualified	qualified	unqualified
1532	35	30	1372	97.77	97.86	98.10	97.44

Table 4. Performance comparison of different models.

Model	Backbone Network	mAP/%	FPS/frame
Faster R-CNN	ResNet-101	86.3	20
EfficientDet	EfficientDet-B0	87.1	17
YOLOv3	Darknet-53	81.7	26
YOLOv4	CAPDarknet-53	89.2	31

bone and the YOLOv3 with Darknet53 as the backbone, the mean accuracy of the YOLOv4 model has improved by 2.9%, 2.1% and 7.5% respectively. The number of frames per second has increased by 11, 14 and 5 frames per second respectively, which has improved the prediction accuracy and detection speed.

4.2 Testing Results

To verify the feasibility of the YOLOv4 model and the online detection capability of the shrimp inspection system, the model was ported to a shrimp real-time inspection platform with a Z390 motherboard, an i7-9700K CPU at 3.6GHz, an NVIDIA RTX 2080Ti GPU with 11G of video memory and 32G of RAM. an industrial CCD camera was used to online The shrimp production pipeline images are acquired, and the network model performs scale cropping and classification feature extraction on the acquired images to detect the quality category of shrimp in the high-speed pipeline in real time, and the effect of the images collected by the camera in real time and completed detection is shown in Fig. 6.

Through a series of experiments, it can be seen that the improved yolov4 algorithm has high detection accuracy in shrimp quality detection task, and has good robustness in different environments, which improves the application range of the algorithm. As can be seen from Figure 6, the model is capable of completing the task of detecting shrimps in a variety of forms with an accuracy of 93.7% and an average detection time of 47ms, which can meet the task of detecting shrimps in high-speed production processes.

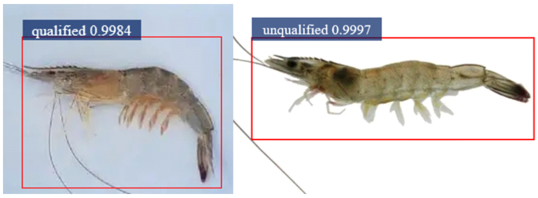


Fig. 6. Detection effect.

5. CONCLUSION

A shrimp quality detection method based on YOLOv4 neural network is proposed. After autonomously collecting and annotating shrimp images, the detection model obtained after training is better than networks such as Faster RCNN and Efficient Det with better robustness through image expansion strategy to meet the richness of shrimp morphology in the dataset. Due to some difficulties in collecting a large amount of complex data independently, the home-made shrimp data images have a single background, and the model is poor at detecting shrimp categories in complex backgrounds with a certain false detection rate. The subsequent work will focus on overcoming and improving this problem in order to further enhance the generalisation performance of the detection model.

REFERENCE

[1] T. Xueheng, W. Xuejun, C. Jinshi, Y. Jixin, Z. Peiwei, Z. Jiucheng, W. Huihui, and Z. Liping, "Sea Cucumber Processing Line With Grading Function," *Application CN20131010-30508A Events*, pp. 1-8, 2013.

[2] T. Xueheng, W. Xuejun, and H. Qiaosheng, "Automatic Weighing and Grading Device for Seafood," *Application CN2015208979320A Events*, pp. 1-10, 2015.

[3] S. Villon, D. Mouillot, M. Chaumont, E.S. Darling, G. Subsol, T. Claverie, and S. Villeger, "A Deep Learning Method for Accurate and Fast Identification of Coral Reef Fishes in Underwater Images," *Elsevier of Ecological Informatics*, Vol. 48, No. 1, pp. 238-244, 2018.

[4] C. Suxia, Z. Yu, W. Yonghui, and Z. Lujun, "Fish Detection Using Deep Learning," *Applied Computational Intelligence & Soft Computing*, Vol. 2020, No. 1, pp. 1-13, 2020.

[5] F. Yiran, T. Xueheng, and L. Eung-Joo,

"Classification of Shellfish Recognition Based on Improved Faster R-CNN Framework of Deep Learning," *Mathematical Problems in Engineering*, Vol. 2021, No. 1, pp. 1-10, 2021.

[6] W. Shuqing, H. Jianfeng, Z. Pengfei, and W. Juan, "Crayfish Quality Detection Method based on YOLOv4," *Food and Machinery*, Vol. 37, No. 3, pp. 120-124, 2021.

[7] Y. Dong-Eon, L. Hyo-Sang, and O. Am-Suk, "Development of Hazardous Food Notification Application Using CNN Model," *Journal of Korea Multimedia Society*, Vol. 25, No. 3, pp. 461-467, 2022.

[8] L. Tao, C. Qiao, and L. Wen, "Research on Intelligent Grading of Beef Marbling Based on Deep Learning," *Food Safety and Quality Detection Technology*, Vol. 9, No. 5, pp. 1059-1064, 2018.

[9] W. Dandan and H. Dongjian, "Channel Pruned YOLO V5s-Based Deep Learning Approach for Rapid and Accurate Apple Fruitlet Detection Before Fruit Thinning," *Elsevier of Biosystems Engineering*, Vol. 210, No. 1, pp. 271-281, 2021.

[10] Y. Jingyao, "Study on Common Shellfish Feature Recognition Technology," *DaLian Polytechnic University*, 2013.

[11] T. Xueheng, L. Hongjie, L. Jinshi, and W. Xuejun, "Computer Vision Technology for Seafood Quality Evaluation," *2012 International Conference on Computer Science and Service System*, pp. 1888-1891, 2012.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[13] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517-6525, 2017.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition,* pp. 580-587, 2014.

[15] R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision*, pp. 1440-1448, 2015.

[16] R. Shaoqing, H. Kaiming, R. Girshick, and S. Jian, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149, 2017.

[17] H. Kaiming, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 42, No. 2, pp. 386-397, 2020.

[18] Z. Xiaofeng and W. Gang, "Data Augmentation Method Based on Generative Adversarial Network," *Computer Systems & Applications*, Vol. 28, No. 10, pp. 201-206, 2019.

[19] W. Ruixi and X. Qinkun, "Multi-Object Images Recognition Based on Deep Networks and Data Augmentation," *Foreign Electronic Measurement Technology*, Vol. 38, No. 5, pp. 86-90, 2019.

[20] S. Meister, N. Moller, J. Stuve, and R.M. Groves, "Synthetic Image Data Augmentation for Fibre Layup Inspection Processes: Techniques to Enhance the Data Set," *Journal of Intelligent Manufacturing*, Vol. 32, No. 6, pp. 1767-1789, 2021.

[21] K.A. Maula, M. Ralf, and R. Markus, "A Benchmark Data Set to Evaluate the Illumination Robustness of Image Processing Algorithms for Object Segmentation and Classification," *PLoS ONE*, Vol. 10, No. 7, pp. 1-9, 2015.

Tao Xingyi

Tao Xingyi is an undergrad (Bachelor Student). Major in information management and Information Systems, Grade 2020. She participated in the research team of marine food processing equipment and supporting technology of Dalian Polytechnic University. She has published 2 papers at KMMS conference and won the best paper award.

Feng Yiran

Feng Yiran is a doctoral candidate majoring in electronic information and Communication at Tongmyong University, South Korea.He received the B.E. degree and M.E. degree from Dalian Polytechnic University, CHINA in 2013 and 2015 respectively. He works as a teacher at Dalian Polytechnic University and member of Korea Multimedia Association. He is mainly engaged in image recognition and robot technology research.

Eung-Joo Lee

Eung-Joo Lee received his B. S., M. S. and Ph. D. in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, and Aug. 1996, respectively. Since 1997 he has been with the Department of Information & Communications Engineering, Tongmyong University, Korea, where he is currently a professor. From 2000 to July 2002, he was a president of Digital Net Bank Inc. From 2005 to July 2006, he was a visiting professor in the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests include biometrics, image processing, and computer vision.

Tao Xueheng

Tao Xueheng graduated from Dalian University of Technology in June 1997 with a doctorate in mechanical engineering from the Department of Mechanical Engineering. He was appointed as a professor at Dalian Polytechnic University in August 2000 and is currently the academic leader of National Marine Food Engineering Technology Research Center and the director of Key Laboratory of Marine Food Processing Technology and Equipment in Liaoning Province. He is a member of the editorial board of the Journal of Combined Machine Tools and Automatic Processing Technology, a member of the Teaching Steering Committee of Automation in Liaoning General Higher Education Schools, the executive director of the Liaoning Light Industry Society and the president of the Light Industry Machinery Branch, a national, provincial and municipal science and technology reviewer, and an innovation and entrepreneurship mentor of the Ministry of Science and Technology and Dalian City.