ICIC Express Letters
Part B: Applications
Volume **14**, Number **5**, May **2023**

ICIC International ©2023 ISSN 2185-2766
pp. **483–489**

FISH SEMANTIC SEGMENTATION BASED ON IMPROVED LOSS FUNCTION

Jiaxin Huang, Liang Xu, Yi Guo*, Yangcheng Liu and Boqi Deng

School of Electrical and Electronic Information
Xihua University
No. 999, Jinzhou Road, Jinniu District, Chengdu 610039, P. R. China
{ huangjiaxin; shinyxl; liuyangcheng; dengboqi }@stu.xhu.edu.cn
*Corresponding author: lpngy@vip.163.com

Received October 2022; accepted January 2023

Abstract. *Fish semantic segmentation is the premise to realize intelligent freshwater fish breeding technology. Deep learning-based works in this field has obtained certain achievements, but they neglect small-scale objects segmentation problem. To improve segmentation accuracy of small-scale objects, building up on Mask R-CNN, a fish segmentation method based on weighted cross entropy loss function is proposed (WCE-Mask R-CNN). Experiments on test dataset show that WCE-Mask R-CNN gains 2.1% and 1.7% improvement on MIoU and F1 respectively, which prove the effectiveness of the weighted cross entropy loss function.*

Keywords: Semantic segmentation, Fish body segmentation, Weighted cross entropy loss function, Deep learning, Mask R-CNN

1. **Introduction.** With the improving living standard, the demand for fish is on the increase. The disadvantages of the traditional breeding technology that relies heavily on artificial cultivation are gradually emerging. In the traditional technology, the breeders measure the fish by visual observation after fishing. It is not only inefficient, but also causes accidental death of fish during the fishing, resulting in unnecessary losses. Therefore, a non-contact breeding technology is urgently needed. The development of artificial intelligence technology has brought new ideas to the freshwater fish breeding industry. The use of three-dimensional point cloud technology to assist farmers in measuring fish not only reduces the labor cost, but also improves the breeding efficiency. As the basis of 3D point cloud, semantic segmentation directly affects the results of 3D point cloud computing. Therefore, semantic segmentation is of great research value.

In recent years, the semantic segmentation method of traditional computer vision [1-3] can realize the automatic segmentation of the fish body, but under the real environment, it is affected by light and plankton, and the segmentation effect is not good. With the advent of the CNN network [4], the use of the image semantics of deep learning networks has become a research hotspot. Garcia et al. [5] used Mask R-CNN [6] to automatically segment the fish body and applied it to large and small fish measurement. Chang et al. [7] used Mask R-CNN to segment fish bodies in sonar images. Abe et al. [8] used SegNet [9] to achieve the segmentation of underwater body, which relieves the semantic segmentation of fish in the noise environment to a certain extent, but the problem of sample imbalance is not solved. The segmentation effect of small scale fish body is not ideal. Introducing attention mechanism into the network is one of the methods to improve the semantic segmentation of fish body. Ji [10] introduced SE-Net attention module [11] in Deeplabv3+ network [12] to achieve semantic segmentation of each part of fish. Zhang et al. [13] used DPANet to improve the segmentation of fish. However, the introduction

DOI: 10.24507/icicelb.14.05.483

of attention mechanism improves the accuracy of fish segmentation, and increases the number of parameters in the network model. It is not conducive to model transfer and deployment.

In order to improve the accuracy of the segmentation of small-scale objects and does not introduce additional parameters, this thesis proposes a fish semantic segmentation method based on improved loss function. In the generation stage of Mask R-CNN mask, the weighted cross entropy loss function is introduced to strike a balance between positive and negative samples in fish body semantic segmentation, alleviate the difficulty of small-scale objects segmentation, and improve the overall accuracy of fish semantic segmentation.

Our main contributions can be summarized as follows.

1) We propose a Mask R-CNN fish body semantic segmentation method based on the cross entropy loss function, which is used for fish segmentation tasks for underwater images.
2) Use the cross entropy loss function in the network, without additional parameters.
3) Our method achieves better performance than traditional Mask R-CNN on the self-built dataset.

The remainder of this paper is organized as follows. In Section 2, we mainly present the dataset used in the experiment. In Section 3, we will present our proposed WCE-Mask R-CNN, including the design details. In Section 4, we mainly introduce the experimental environment and how to design the experiment. In Section 5, we present the experimental results and perform the necessary analysis of our work. In Section 6, we conclude our work.

2. Dataset Preparation.

2.1. Dataset source. In order to speed up the convergence of the model, thesis adopts the training method of transfer learning instead of end-to-end training. Therefore, the experimental datasets contain two parts. The first part for training is the Microsoft large-scale dataset COCO [14], and the other part is the images taken by the ZED binocular camera in real fish breeding environment.

In the first part, COCO dataset is a large-scale dataset funded and annotated by Microsoft in 2014, which contains rich object detection and segmentation data. Dataset is mainly intercepted from complex daily scenes. The targets in the images are calibrated by accurate segmentation. The images include 91 types of targets, 328000 images and 2500000 labels.

In the second part, the images captured by the zed binocular camera are called as fish dataset, which contains 224 fish images. Figure 1 shows dataset sample. The dataset can be divided into small-scale and large-scale according to the proportion of fish area in the

FIGURE 1. Dataset sample

image. The target whose median ratio of bounding box area to image area falls between 0.08% and 0.58% is defined as a small-scale target [15].

2.2. **Data preprocess.** Data enhancement is an effective means to expand dataset, improve image visual varieties, and avoid model overfitting [16]. Due to the limited number of images in fish dataset and continuity of fish state and similar background in the image caused by periodic image capture, the training is doomed to overfit. Through turning, rotation, and other geometric data enhancement operations, the original continuity of fish state is broken, and the overfitting of the model in the training process is avoided. In order to alleviate the impact of image noise to some extent, the gray-scale adjustment method of contrast enhancement is used to make the outline of the fish clearer in the image. After data enhancement, the original fish dataset is expanded to 1344 images, and divided into training set, validation set and test set by the ratio of $6:3:1$.

In order for the self-built set to be trained into the network, it is necessary to have a corresponding image json file that satisfies the network input parameters, so it is manually labeled using LabelMe [17] deep learning image annotation software. Label all fish in the image as a fish class, each fish in order, and the background is labeled as a background class. Save all annotation files with the suffix .json to prepare the input data for subsequent network training.

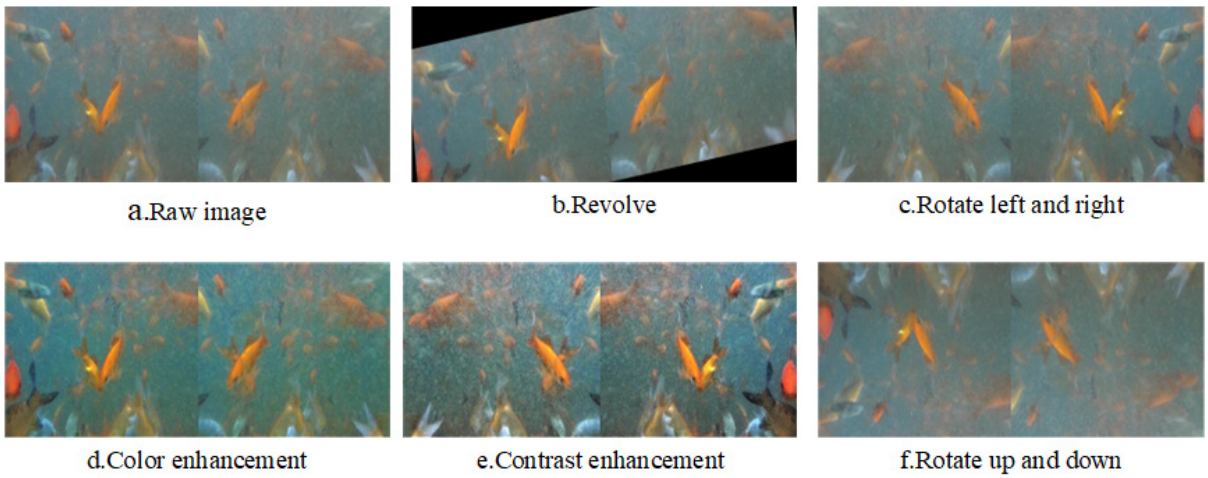

FIGURE 2. Data enhancement sample

3. **Fish Semantic Segmentation Model.** Mask R-CNN [6] is an improved algorithm based on Faster R-CNN [18]. Figure 3 shows the Mask R-CNN algorithm framework. Mask R-CNN generates the feature map through the convolution neural network (CNN), and multiple candidate boxes through the FPN [19]. And then retain the most accurate candidate boxes through post-processing and map them to the corresponding feature map. Through RoIAlign, the uneven size input is changed into a fixed size feature map, and finally the extracted ROI features are fed into classifier and regressor.

In order to measure the difference between the predicted value and the ground truth value in the training process, the loss function is usually introduced into the convolutional neural network. The better the loss function is, the better the model performs. In the traditional Mask R-CNN network, when generating a mask, the full convolution network (FCN) [20] uses the cross entropy as the loss function. However, when segmenting small-scale target fish, the segmentation performance of small objects is poor because of the obvious imbalance between positive samples and negative samples. In order to improve the segmentation performance of small objects, the weighted cross entropy loss function

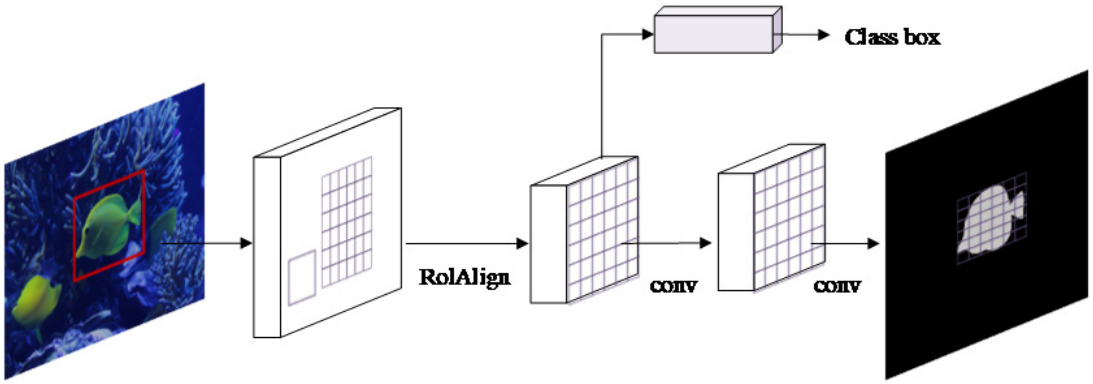

FIGURE 3. Mask R-CNN algorithm framework

is introduced in the full convolution network stage to replace the plain cross entropy loss function, as shown Equation (1),

$$WCE = -\frac{1}{N}\sum_{i=1}^{N}\lambda_i \sum_{j\in(0,1)} y_{ij}\log p_{ij} \tag{1}$$

where the WCE represents weighted cross entropy, λ_i is the weight of sample i, and y_{ij} represents signum function. If the category of the sample is a certain category, $y_{ij} = 1$, otherwise $y_{ij} = 0$, and p_{ij} is the probability of sample i being category j.

In this paper, the weights of samples are calculated by median frequency balancing [7], as shown in Equation (2) and Equation (3),

$$freq(i) = \frac{sum\_pixel\_i}{sum\_image\_pixel\_i} \tag{2}$$

$$\lambda_i = \frac{median\_of\_freq(i)}{freq(i)} \tag{3}$$

where $freq(i)$ is the frequency of category i, sum_pixel_i is the number of pixels that fall in category i, $sum\_image\_pixel\_i$ is the pixels number of an image, $median\_of\_freq(i)$ is median value of frequencies of all categories.

In the full convolution network, the network will consistently extract features through down sampling to reduce the amount of memory costs and computation, and increase the receptive field to obtain more rich semantic features. The small-scale fish itself contains a tiny number of pixels, and the number will get fewer after down sampling. In such case, remained pixels carry too few semantic information to deliver good classification and regression performance. Therefore, the weighted cross entropy loss function is introduced into FPN in this paper to boost the difference between the ground truth value of each batch and the predicted value of the model during semantic segmentation, and then Adam is used to optimize the loss function to get a smaller loss value.

4. **Experimental Environment and Design.**

4.1. **Experiment environment.**

Experiments are conducted with AMD Ryzen 5 5600x CPU, NVIDIA Geforce RTX3080ti GPU and 32G memory. At the same time, CUDNN11.0 is used to speed up the computation of convolutional neural networks, and Adam optimizer with default initial parameters is used for the iterative process of the model. Use Keras [21] high-level application program interface for Tensorflow. Model construction, training and prediction are carried out with Tensorflow. The specific version of environment used in the experiment is Tensorflow GPU = 1.13.1, Keras = 2.1.5.

4.2. **Experiment design.** In order to verify the effectiveness of the proposed method, four estimators are used, including MIoU (mean intersection over union), precision, recall and comprehensive evaluation index (F1). A comparison experiment is designed to prove the effectiveness of weighted cross entropy for fish semantic segmentation, as shown in Table 1.

TABLE 1. Hyperparameter of comparison experiment

Model	Epoch	Iteration/epoch	Initial learning rate
Mask R-CNN	50	725	0.0001
WCE-Mask R-CNN	50	725	0.0001

Transfer learning is used during the training process. Mask R-CNN and WCE-Mask R-CNN are separately pre-trained on COCO dataset before they are trained on the fish dataset for 50 epochs with initial learning rate set as 0.0001. During the training process on fish dataset, each epoch has 725 iterations.

5. Results.

5.1. **Analysis of loss curves.** According to designed experiments, both Mask R-CNN and WCE-Mask R-CNN are trained 50 epochs, and their loss curves are shown in Figure 4. It can be seen that the loss keeps going down with the increase of epochs and converges in the end. It is worth noting that WCE-Mask R-CNN loss function converges faster.

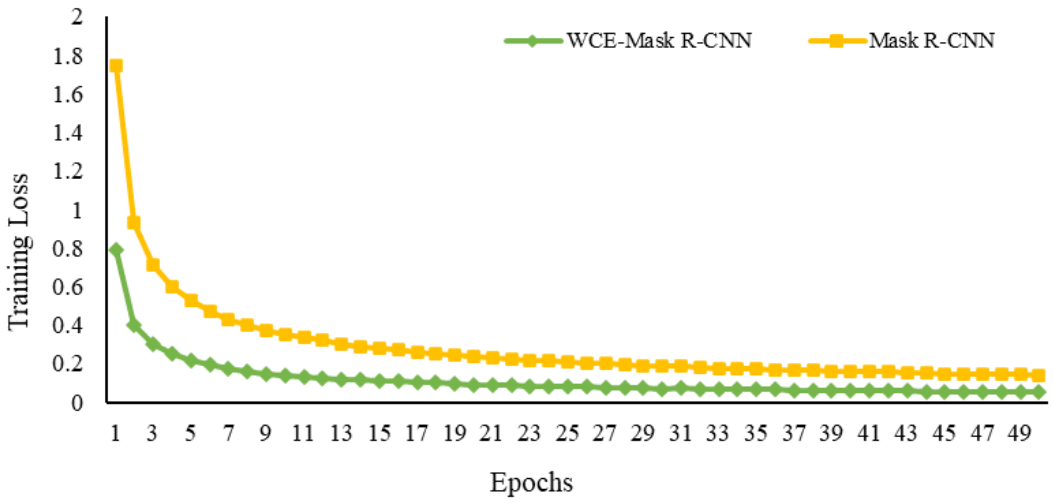

FIGURE 4. Loss curves of the training process

5.2. **Analysis of segmentation performance.** In order to further evaluate the optimization effect of the weighted cross entropy loss function on the model, trained models of Mask R-CNN and WCE-Mask R-CNN are used to predict the test set. Fuzzy contour of small-scale fish in the image makes it difficult to segment, so in the full convolution stage, the weighted cross entropy loss function is used to improve the weight of positive samples, which helps the Adam optimizer to optimize the model better. As shown in Figure 5, it can be observed that WCE-Mask R-CNN has a better segmentation performance for small-scale fish, especially for some small parts of fish bodies, such as tails.

Further, in order to evaluate the influence of different loss functions on the network more objectively, MIoU, accuracy rate, recall rate and F1 are used to evaluate on the training set and test set, respectively. The segmentation results of models that adopt different loss functions are shown in Table 2. On training set, MIoU, Pre, Rec and F1

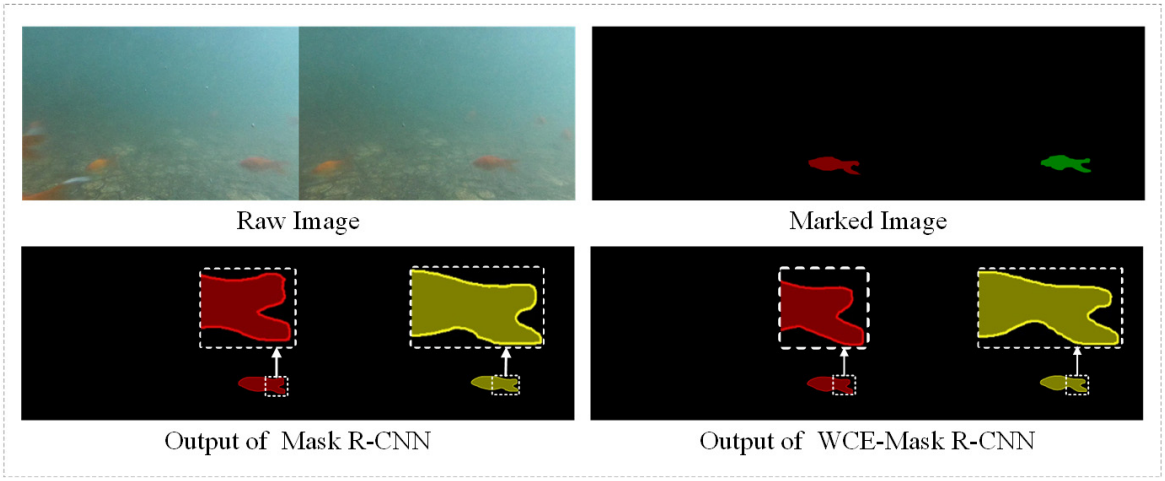

FIGURE 5. Outputs of WCE-Mask R-CNN and Mask R-CNN

TABLE 2. Comparison of results by different loss functions

Model	Train set				Test set				Parm
	MIoU	Pre	Rec	F1	MIoU	Pre	Rec	F1	
Mask R-CNN	0.916	0.943	0.893	0.917	0.843	0.871	0.824	0.847	24MB
WCE-Mask R-CNN	0.935	0.964	0.912	0.937	0.864	0.889	0.841	0.864	24MB

Note: Pre and Rec represent precision and recall, respectively. Parm represents the model parameter quantity.

improved by 1.9%, 2.1%, 1.9% and 2%, respectively. On the test set, MIoU, Pre, Rec and F1 improved by 2.1%, 1.8%, 1.7% and 1.7%, respectively. The results show that WCE-Mask R-CNN is better than the original Mask R-CNN in every indicator, and the weighted cross entropy loss function boosts the segmentation of small-scale fish bodies. In other word, experiments demonstrate the effectiveness of this method.

6. **Conclusions.** In this thesis, WCE-Mask R-CNN model based on weighted cross entropy loss function is proposed to solve the problem of unbalanced positive and negative samples in small-scale target fish segmentation. By introducing the weighted cross entropy loss function in FCN stage, samples with fewer pixels are given higher weights. It can temporarily enlarge the value of the loss function at the initial training stage, magnify the difference between the ground truth and the predicted value, and then use Adam optimizer to make the loss converge faster and finally get a smaller loss value.

The experiment compares WCE-Mask R-CNN with Mask R-CNN from two aspects: training loss curve and segmentation performance. And it is proved that WCE-Mask R-CNN has a faster convergence speed. At the same time, WCE-Mask R-CNN performs better in segmentation of some small parts of small-scale fish bodies. Above all, various indicators on training set, validation set and test set also prove the superiority of WCE-Mask R-CNN. In summary, experiment demonstrates the effectiveness of the proposed method in this thesis, which achieves more refined segmentation of small-scale fish, and provides more accurate results for the three-dimensional modeling of fish body and the research of semantic point cloud.

The weighted loss function is helpful to improve the precision of fish body semantic segmentation. However, it is worth noting that there is a problem of overfitting when the amount of data is small. Therefore, future research can consider designing a better loss function to reduce overfitting. At the same time, further improving the accuracy of fish body semantic segmentation is also the focus of future research.

Acknowledgment. This research is supported by Grant SCITLAB-1021 of Intelligent Terminal Key Laboratory of Sichuan Province, the National Natural Science Foundation of China under Grant (61973257, 61901394), Central Government Funds of Guiding Local Scientific and Technological Development for Sichuan Province (2021ZYD0034).

REFERENCES

[1] A. Bali and S. N. Singh, A review on the strategies and techniques of image segmentation, *2015 5th International Conference on Advanced Computing & Communication Technologies*, pp.113-120, 2015.

[2] H. Yao, Q. Duan, D. Li et al., An improved K-means clustering algorithm for fish image segmentation, *Mathematical and Computer Modelling*, vol.58, nos.3-4, pp.790-798, 2013.

[3] M.-C. Chuang, J.-N. Hwang, K. Williams et al., Automatic fish segmentation via double local thresholding for trawl-based underwater camera systems, *2011 18th IEEE International Conference on Image Processing*, pp.3145-3148, 2011.

[4] H. C. Shin, H. R. Roth, M. Gao et al., Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, *IEEE Trans. Medical Imaging*, vol.35, no.5, pp.1285-1298, 2016.

[5] R. Garcia, R. Prados, J. Quintana et al., Automatic segmentation of fish using deep learning with application to fish size measurement, *ICES Journal of Marine Science*, vol.77, no.4, pp.1354-1366, 2020.

[6] K. He, G. Gkioxari, P. Dollár et al., Mask R-CNN, *Proc. of the IEEE International Conference on Computer Vision*, pp.2961-2969, 2017.

[7] C. C. Chang, Y. P. Wang and S. C. Cheng, Fish segmentation in sonar images by Mask R-CNN on feature maps of conditional random fields, *Sensors*, vol.21, no.22, 7625, 2021.

[8] S. Abe, T. Takagi, S. Torisawa et al., Development of fish spatio-temporal identifying technology using SegNet in aquaculture net cages, *Aquacultural Engineering*, vol.93, 102146, 2021.

[9] V. Badrinarayanan, A. Kendall and R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.39, no.12, pp.2481-2495, 2017.

[10] X. Ji, *Research on the Target Detection of Underwater Fish Based on Deep Learning and the Study of Fish Body Segmentation Algorithm*, Master Thesis, Tianjin University of Science and Technology, 2021 (in Chinese).

[11] J. Hu, L. Shen and G. Sun, Squeeze-and-excitation networks, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132-7141, 2018.

[12] L. C. Chen, Y. Zhu, G. Papandreou et al., Encoder-decoder with atrous separable convolution for semantic image segmentation, *Proc. of European Conference on Computer Vision (ECCV 2018)*, pp.833-851, 2018.

[13] W. Zhang, C. Wu and Z. Bao, DPANet: Dual pooling-aggregated attention network for fish segmentation, *IET Computer Vision*, vol.16, no.1, pp.67-82, 2022.

[14] T. Y. Lin, M. Maire, S. Belongie et al., Microsoft COCO: Common objects in context, *Proc. of European Conference on Computer Vision (ECCV 2014)*, pp.740-755, 2014.

[15] M. Kampffmeyer, A. B. Salberg and R. Jenssen, Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.1-9, 2016.

[16] C. Shorten and T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, *Journal of Big Data*, vol.6, no.1, pp.1-48, 2019.

[17] B. C. Russell, A. Torralba, K. P. Murphy et al., LabelMe: A database and web-based tool for image annotation, *International Journal of Computer Vision*, vol.77, no.1, pp.157-173, 2008.

[18] S. Ren, K. He, R. Girshick et al., Faster R-CNN: Towards real-time object detection with region proposal networks, *Advances in Neural Information Processing Systems*, vol.28, 2015.

[19] J. Long, E. Shelhamer and T. Darrell, Fully convolutional networks for semantic segmentation, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.3431-3440, 2015.

[20] F. Zammori and R. Gabbrielli, ANP/RPN: A multi criteria evaluation of the risk priority number, *Quality and Reliability Engineering International*, vol.28, no.1, pp.85-104, 2012.

[21] N. Ketkar, Introduction to Keras, in *Deep Learning with Python*, Berkeley, CA, Apress, 2017.