

Real-Time Instance Segmentation of Metal Screw Defects Based on Deep Learning Approach

Wei-Yu Chen, Yu-Reng Tsao, Jin-Yi Lai, Ching-Jung Hung, Yu-Cheng Liu, Cheng-Yang Liu

Department of Biomedical Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan, cyliu66@nycu.edu.tw

In general, manual methods are often used to inspect defects in the production of metal screws. As deep learning shines in the field of visual detection, this study employs the You Only Look At CoefficientTs (YOLACT) algorithm to detect the surface defects of the metal screw heads. The raw images with different defects are collected by an automated microscopic camera scanning system to build the training and validation datasets. The experimental results demonstrate that the trained YOLACT is sufficient to achieve a mean average accuracy of 92.8 % with low missing and false rates. The processing speed of the trained YOLACT reaches 30 frames per second. Compared with other segmentation methods, the proposed model provides excellent performance in both segmentation and detection accuracy. Our efficient deep learning-based system may support the advancement of non-contact defect assessment methods for quality control of the screw manufacture.

Keywords: Defect inspection, metal screw, deep learning, convolutional neural network.

1. INTRODUCTION

In general, the defect inspections are often performed manually and suffer fluctuations in the inspection accuracy, speed, and a large number of human resources [1]. A less-time and labor-intensive automatized method of defect inspections is required because the small metal components and screws are manufactured in large quantities. The visual detection of production components is increasingly relevant for quality control in manufacturing industry [2]. The exact and constant quality control is crucial, especially for metal screws. Recently, machine learning has been applied in many industries for surface recognition issues [3], [4]. Several machine learning algorithms, such as support vector machines, linear discriminant models, and neighbor-based models, are trained on feature characterizations from pre-processed images [5]. Moreover, several non-deep learning-based methods are also presented, including morphological edge detection, autoregressive models, histograms, co-occurrence matrices, and autocorrelation [6]-[8]. Compared with deep learning, the operation of the above-mentioned methods is very complicated. Deep learning algorithm plays a decisive role in the image recognition, such as face recognition, handwriting recognition, and voice recognition [9]-[11]. The advantage of deep learning algorithm is high detection rate for anomaly objects. Convolutional neural network (CNN) has been widely used to solve the complex image recognition problems of anomaly detection tasks [12]. Various image recognition-based techniques have been proposed in the field of surface detection. A vision inspection system is developed to acquire railway images and the

contour extraction of anomalies [13], [14]. An image recognition system based on threshold segmentation and morphological opening calculation is developed for identifying internal screw threads [15]. However, the strong disturbances of the measured images lead a high under-segmentation and result in the decrease of recognition accuracy. The improvement of the image recognition algorithm is further required.

There are two main ways for the design of object recognition model. Two-stage network has the attention of the recognition accuracy, such as region-based fully convolutional networks (R-FCN) and faster region convolution neural networks (R-CNN) [16]. One-stage network has the attention of the recognition speed, such as You Only Look Once (YOLO) [17] and single shot multiBox detector (SSD) [18]. Mask R-CNN is an instance segmentation model based on the improvement of faster R-CNN, which is two-stage instance segmentation [19]. In the two-stage model, the re-pooling operations are logically serial to map features by the bounding box, and it is an arduous task to speed up. The You Only Look At CoefficientTs (YOLACT) is a one-stage instance segmentation model based on the one-stage object detector [20], [21]. The YOLACT algorithm abandons the implicit feature location step and separates the instance segmentation tasks into two parallel subtasks. A series of prototype masks is generated for covering the entire image. A series of linear combination coefficients is predicted for each instance. During inference, the corresponding predicted mask coefficient for each instance is used to simply multiply and

add with the prototype mask. The final mask for each instance is cropped and thresholded according to the bounding box. In this study, a real-time method based on YOLACT is proposed for identifying different defects in metal screws. An optical microscopic scanning system is developed for automatic surface defect inspection. Three significant advantages of the deep learning-based YOLACT model are fast speed, high mask quality, and strong universality. This model of prototype mask generation and mask coefficients can be applied to many popular products in real industrial processes.

2. EXPERIMENTAL SYSTEM & METHODS

It is important to obtain high-quality raw images for building the dataset and reliable defect detection. However, a regular camera has difficulty in capturing the surface defects with adequate resolution. Screw is usually a metal material and the screw surface may reflect light randomly, which affects the image quality. Moreover, some screw surface defects are narrow, which cannot be seen by naked human eyes. In this study, a monocular microscope scanning imaging system with an industrial camera and coaxial ring light is utilized for raw image acquisitions. Fig.1. shows the experimental configuration of the microscope scanning imaging system. An industrial camera (Hayear HY-5200) with resolution of 4608×3456 pixels and a 12 mm lens (Hayear HM2012-10M) is employed in the imaging system. The distance between the screw and lens is about 90 mm. A ring light (Hayear HY-209-144B) is beneficial for eliminating the shadows of metal rough surface.

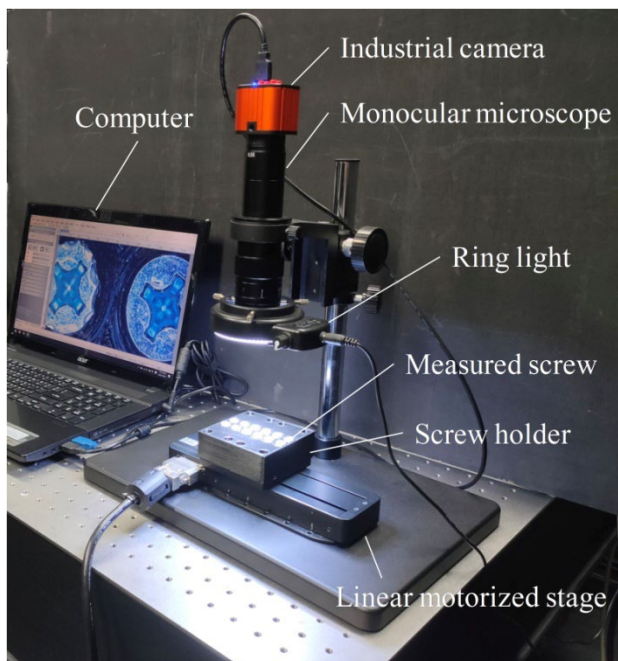


Fig.1. Experimental monocular microscope scanning imaging system for raw screw image acquisitions.

It is a challenge to receive enough images for training the YOLACT model. If training datasets are insufficient, overfitting could quickly occur and the model learns the data by memorization. To avoid this, the basic concept is to create more different training images by data augmentation [22].

Fig.2. shows several samples of the raw images of the screw surface defects captured by the imaging system. The metal screws M6 with GB819 cross countersunk head are shown in Fig.2.a). The geometrical profiles of the metric screws are defined by the International Organization for Standardization. The defect types of screw surfaces include stripped screws, surface-damaged screws, unprocessed screws, and surface-dirty screws. The raw images are collected under various defect types and lighting conditions. The image resolution is 550×300 pixels and is saved in JPG format. The 1440 qualified raw images are selected for building up the dataset. The image rotation, shear, flip, zoom and shift changes are implemented to increase the diversity of the train images. The total number of image datasets is 3200 after data augmentation. 90 % of images are randomly picked for training model, and the validation dataset is produced by 10 % of images.

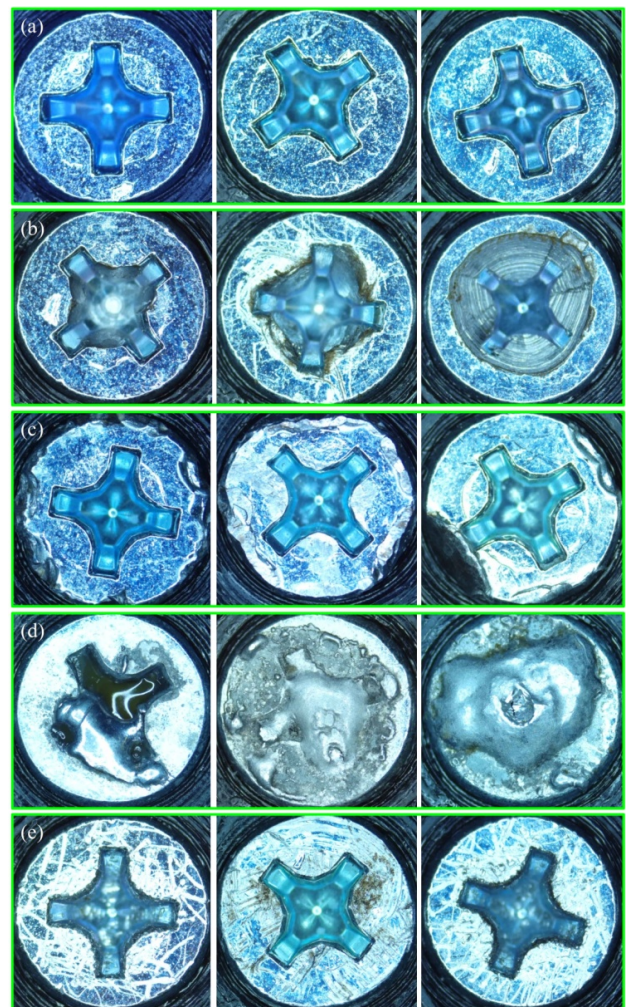


Fig.2. Raw surface images of the metal screw head: a) defect-free screws, b) stripped screws, c) surface-damaged screws, d) unprocessed screws, and e) surface-dirty screws.

In this work, the YOLACT model is used to extract the morphological contours of the screw defects. The network architecture of the YOLACT model is shown in Fig.3. The default size of input raw images is 550×300 pixels. The default backbone network is the deep residual learning

combined with feature pyramid networks [23], [24]. The feature maps extracted by the backbone network are transmitted into two parallel branches. The first branch of the prediction head receives all-size feature maps, and predicts the object location, category, and mask coefficients. The redundant objects are removed by fast non-maximum suppression (NMS) [25]. The second branch of the protonet receives the largest-size feature map, and creates multiple prototype masks. The prototype masks are linearly mixed mask coefficients to obtain the mask conformable to each object. The object detection box is used to crop the corresponding masks. The final mask is obtained by thresholding, and can be used to extract the outer contours of the screw defects. In the experiment, a core processor (Intel i9-10920X) is employed in conjunction with a graphics card (NVIDIA GeForce RTX 3080 Ti). The operating system is Windows 10, the Python version is 3.7, and the NVIDIA parallel computing platform is CUDA 1.1. The model is conducted using the open source deep learning framework of Pytorch 1.9.0.

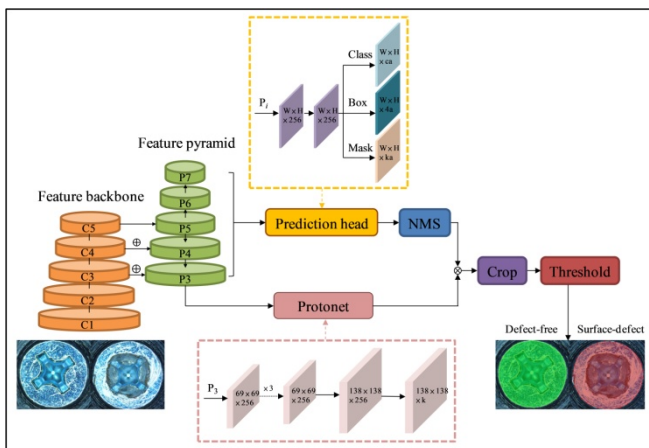


Fig.3. Network architecture of the screw defect extraction based on the YOLACT model.

3. RESULTS & DISCUSSIONS

The experimental dataset is a large image database collected by our automatic scanning imaging system. The selected screw images include the variations in lighting and posture. Four categories of screw defects are labeled in order to extract the key contours of surface defects. The LabelMe tool is used to complete the labeling of surface defects in the images [26]. The 2880 screw images are selected as the training set, and the 320 screw images are used as the verification set. The YOLACT model is trained by the above dataset. Table 1. shows the training parameters. The size of input image in the network is $\text{max_size} = 550 \times 300$ pixels. The number of iterations is lr_steps , where the learning rate decays during the training. The maximum number of training iterations is max_iter . The number of images processed in the same batch is batch_size .

The YOLACT algorithm is compared with the traditional mask R-CNN algorithm for verifying the superiority of the proposed method. The identical dataset is employed to train these two models. Fig.4. shows the training accuracy and loss of the YOLACT model and traditional mask R-CNN with

20000 iterations. Experimental results indicate that the accuracy and loss tend to be stable after 20000 iterations. It is clearly stated that the accuracy of the proposed YOLACT method is much higher than the traditional mask R-CNN method. The accuracy of the YOLACT training is close to 100 % after 4000 iterations. It can be seen from Fig.4.b) that the loss rate of the proposed YOLACT reduces rapidly, which is better than that of the traditional mask R-CNN. The loss rate of the proposed YOLACT is close to 0, when the number of iterations is about 20000. The YOLACT model has achieved excellent training without relying too much on the dataset and overfitting.

Table 1. Training parameters.

Parameter	Value
max_size	550×300 pixels
lr_steps	(280000, 600000, 700000, 750000)
max_iter	800000
batch_size	5

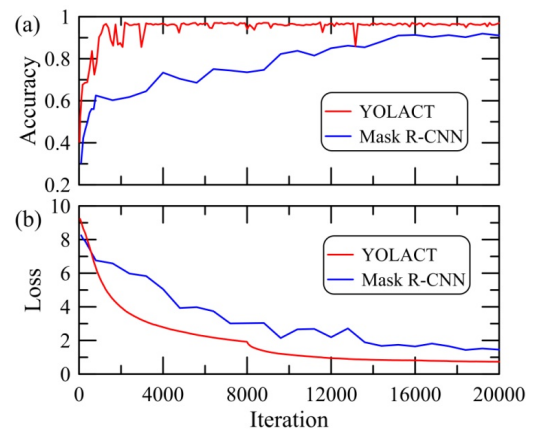


Fig.4. a) Accuracy and b) Loss of the trained YOLACT model and traditional mask R-CNN model with 20000 iterations.

Table 2. Average precision of the screw defect detection.

IOU	All	0.50	0.55	0.60	0.65	0.70
Box	40.11	44.46	44.46	44.46	44.46	44.46
Mask	41.37	44.46	44.46	44.46	44.46	44.46
IOU	0.75	0.80	0.85	0.90	0.95	
Box	44.46	43.97	41.65	33.17	15.62	
Mask	44.36	44.12	43.04	38.58	21.36	

Table 2. shows the average precisions (AP) of the box data and mask data performed by the YOLACT model. The intersection over union (IOU) is a measurement based on the Jaccard Index that estimates the overlap between two bounding boxes [27]. A ground truth bounding box and a predicted bounding box are required for the calculation. By setting the IOU threshold, we can easily evaluate if the detection is correct or incorrect. For the accuracy of the mask and bounding boxes, the high AP values are displayed between 0.50 overlap and 0.85 overlap. The lowest AP values of the bounding box and mask are 15.62 and 21.36 at 0.95 overlap, respectively. The overall values of mean average precision (mAP) are 40.11 and 41.37 for the bounding box and mask. These training results clarify that the YOLACT

model is well recognized for classifying screw defect types with high accuracy and high reliability. Fig.5. shows the detection results of the proposed YOLACT model. The label displayed on the bounding box is analyzed to distinguish the defect types of screw surfaces. These detection results can be used well in recognition due to the high accuracy and high reliability of classifying screw defect types. It can be considered in Fig.5.b) that the wear condition of the screw surface is the reason that influences the accuracy of the model. In addition, the unknown patterns also affect the coverage of the bounding box of the screw surfaces.

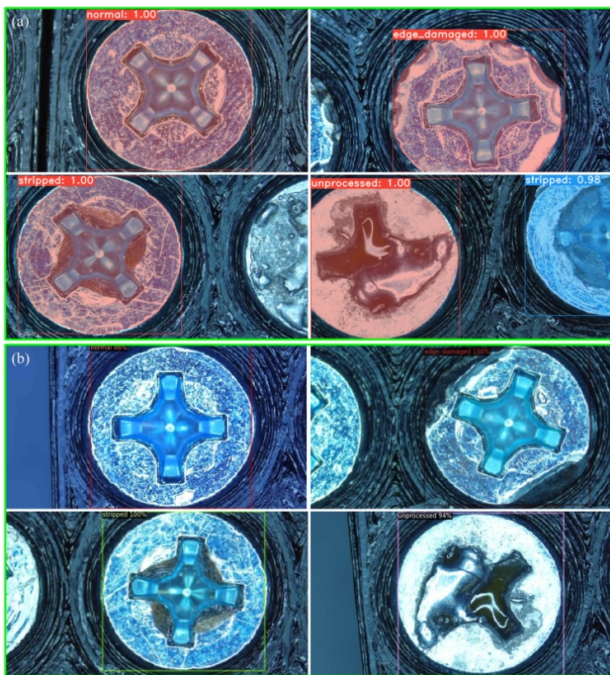


Fig.5. Detection results of a) the trained YOLACT model and b) the traditional mask R-CNN model for screw defect surfaces.

4. CONCLUSION

The surface defects of the metal screws yield great threats to mechanical system, which lead to serious strength reduction. Several current image-level defect classification methods only provide qualitative defect data, but they cannot estimate various surface defects in a reliable manner. In this work, a robust deep learning-based technique using a state-of-the-art YOLACT model is proposed to automatically detect the defects on screw surfaces on pixel-level. The scanning microscopic camera system equipped with coaxial ring light is developed to acquire high-quality raw images. The experimental results confirm that such network architecture has excellent potential in detecting the defect features of screw head without the needs of any pre-processing treatment. A satisfactory performance is achieved as exhibited by a mean average accuracy of 92.8%. In comparison with the mask R-CNN algorithm, the YOLACT model has better performance for both segmentation and detection accuracy. However, further works are still required to combine a lightweight optical microscopic system with a robot manipulator for realizing onsite screw defect inspection. The pixel-level segmentation information with spatial position distributions and defect sizes will be

incorporated in the analysis model for evaluating the defect-caused strength degradation and judging the probability of fracture.

ACKNOWLEDGMENT

This work is funded by the Ministry of Science and Technology of Taiwan (MOST 108-2221-E-010-012-MY3, MOST 109-2923-E-010-001-MY2) and the Yen Tjing Ling Medical Foundation (CI-110-28).

REFERENCES

- [1] Wang, T., Chen, Y., Qiao, M., Snoussi, H. (2018). A fast and robust convolutional neural network-based defect detection model in product quality control. *The International Journal of Advanced Manufacturing Technology*, 94 (9), 3465-3471. <https://doi.org/10.1007/s00170-017-0882-0>.
- [2] Chen, J., Liu, Z., Wang, H., Núñez, A., Han, Z. (2018). Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Transactions on Instrumentation and Measurement*, 67 (2), 257-269. DOI: 10.1109/TIM.2017.2775345.
- [3] Kumar, A. (2008). Computer-vision-based fabric defect detection: A survey. *IEEE Transactions on Industrial Electronics*, 55 (1), 348-363. DOI: 10.1109/TIE.1930.896476.
- [4] Min, Y., Xiao, B., Dang, J., Yue, B., Cheng, T. (2018). Real time detection system for rail surface defects based on machine vision. *EURASIP Journal on Image and Video Processing*, 3, 1-11. <https://doi.org/10.1186/s13640-017-0241-y>.
- [5] Xie, X. (2008). A review of recent advances in surface defect detection using texture analysis techniques. *Special Issue on Quality Control by Artificial Vision*, 7 (3), 1-22. <https://doi.org/10.5565/rev/elcvia.268>.
- [6] Mukherjee, A., Chaudhuri, S., Dutta, P., Sen, S., Patra, A. (2006). An object-based coding scheme for frontal surface of defective fluted ingot. *ISA Transactions*, 45 (1). [https://doi.org/10.1016/S0019-0578\(07\)60060-3](https://doi.org/10.1016/S0019-0578(07)60060-3).
- [7] Haralick, R., Shanmugam, K., Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3 (6), 610-621. DOI: 10.1109/TSMC.1973.4309314.
- [8] Hoseini, E., Farhadi, F., Tajeripour, F. (2013). Fabric defect detection using auto-correlation function. *International Journal of Computer Theory and Engineering*, 5 (1), 114-117. <https://doi.org/10.7763/IJCTE.2013.V5.658>.
- [9] Guo, G., Zhang, N. (2019). A survey on deep learning based face recognition. *Computer Vision and Image Understanding*, 189, 102805. <https://doi.org/10.1016/j.cviu.2019.102805>.
- [10] Eltay, M., Zidouri, A., Ahmad, I. (2020). Exploring deep learning approaches to recognize handwritten arabic texts. *IEEE Access*, 8, 89882-89898. DOI: 10.1109/ACCESS.2020.2994248.
- [11] Hu, H., Chang, S., Wang, C., Li, K., Cho, H., Chen, Y., Lu, C., Tsai, T., Lee, O. (2021). Deep learning

- application for vocal fold disease prediction through voice recognition: Preliminary development study. *Journal of Medical Internet Research*, 23 (6), e25247. <https://doi.org/10.2196/25247>.
- [12] Zhang, J., Shao, K., Luo, X. (2018). Small sample image recognition using improved convolutional neural network. *Journal of Visual Communication and Image Representation*, 55, 640-647. <https://doi.org/10.1016/j.jvcir.2018.07.011>.
- [13] Li, Q., Ren, S. (2012). A real-time visual inspection system for discrete surface defects of rail heads. *IEEE Transactions on Instrumentation and Measurement*, 61 (8), 2189-2199. DOI: 10.1109/TIM.2012.2184959.
- [14] Feng, H., Jiang, Z., Xie, F., Yang, P., Shi, J., Chen, L. (2014). Automatic fastener classification and defect detection in vision-based railway inspection systems. *IEEE Transactions on Instrumentation and Measurement*, 63 (4), 877-888. DOI: 10.1109/TIM.2013.2283741.
- [15] Chen, Y., Zhu, X., Wang, Y., Guo, J., Chen, L. (2014). Research on image recognition of internal thread. *Tool Engineering*, 12, 77-79.
- [16] Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (6), 1137-1149. DOI: 10.1109/TPAMI.2016.2577031.
- [17] Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, 157, 417. <https://doi.org/10.1016/j.compag.2019.01.012>.
- [18] Zhai, S., Shang, D., Wang, S., Dong, S. (2020). DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion. *IEEE Access*, 8, 2169-3536. DOI: 10.1109/ACCESS.2020.2971026.
- [19] Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., Schneider, D. (2020). Automated cattle counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*, 171, 10530. <https://doi.org/10.1016/j.compag.2020.105300>.
- [20] Bolya, D., Zhou, C., Xiao, F., Lee, Y. (2019). YOLACT: Real-time instance segmentation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 9157-9166. DOI: 10.1109/ICCV.2019.00925.
- [21] Lin, S., Zhu, K., Feng, C., Chen, Z. (2021). Align-Yolact: A one-stage semantic segmentation network for real-time object detection. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-021-03340-4>.
- [22] Shorten, C., Khoshgoftaar, T. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>.
- [23] Wu, S., Zhong, S., Liu, Y. (2018). Deep residual learning for image steganalysis. *Multimedia Tools and Applications*, 77, 10437-10453. <https://doi.org/10.1007/s11042-017-4440-4>.
- [24] Wang, C., Zhong, C. (2021). Adaptive feature pyramid networks for object detection. *IEEE Access*, 9, 107024-107032. DOI: 10.1109/ACCESS.2021.3100369.
- [25] Qiu, S., Wen, G., Deng, Z., Liu, J., Fan, Y. (2018). Accurate non-maximum suppression for object detection in high-resolution remote sensing images. *Remote Sensing Letters*, 9 (3), 237-246. <https://doi.org/10.1080/2150704X.2017.1415473>.
- [26] Russell, B., Torralba, A., Murphy, K., Freeman, T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77, 157-173. <https://doi.org/10.1007/s11263-007-0090-8>.
- [27] Padilla, R., Passos, W., Dias, T., Netto, S., Silva E. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, 10 (3), 279. <https://doi.org/10.3390/electronics10030279>.

Received December 14, 2021

Accepted February 28, 2022