International Journal of Internet, Broadcasting and Communication Vol.16 No.4 133-140 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.4.133

IJIBC 24-4-15

A study of Strawberry Maturity Classification Using Improved Faster R-CNN

Taewook Kim[1], Heejun Youn[2], Seunghyun Lee[3], Soonchul Kwon[*4]

[1]M.S, Department of Smart System, Kwangwoon University, Seoul, South Korea;
[2]M.S, Department of Plasma Bio Display, Kwangwoon University, Seoul, South Korea;
[3]Professor, Ingenium College Liberal Arts, Kwangwoon University, Seoul, South Korea;
*[*4]Associate professor, Graduate School of Smart Convergence, Kwangwoon University, Seoul, South Korea;*

rpaisxol@kw.ac.kr, hjyun@kw.ac.kr, shlee@kw.ac.kr, [*]ksc0226@kw.ac.kr

Abstract

In strawberry cultivation, maturity classification plays an important role in ensuring the efficiency and quality of harvesting. In this study, we propose an Improved Faster R-CNN model to address these challenges, using MobileNetV3-Large as the backbone network to achieve a lightweight model, and introducing RoI Align to improve the spatial accuracy of the feature map. Experiments are conducted using the KGCV_Strawberry dataset, with precision, recall, F1 score, and mean average precision (mAP) measured for performance evaluation. The experimental results show that the proposed model achieves an average precision of 71.35%, recall of 71.07%, and F1 score of 71.21% across all classes. In particular, the proposed model achieves 63% performance on mAP0.5 and 58% performance on mAP0.5:0.95, which is comparable to existing ResNet-based models while achieving faster inference speed. The proposed model achieves a processing speed of 27.6543 ms, which is about 2 ms faster than existing ResNet-based models. This indicates that the goal of creating a lightweight model with improved image processing capability was achieved with minimal performance degradation. This research is expected to contribute to the development of automated strawberry cultivation systems in greenhouse environments and has the potential to be applied to various agricultural environments in the future.

Keywords: *Convolutional Neural Network (CNN), Faster R-CNN, Image Classification, RoI Align, Strawberry Maturity*

1. Introduction

Strawberries are a fruit harvested worldwide and have received a lot of attention due to their nutritional value and commercial importance [1]. South Korea is one of the leading countries in greenhouse strawberry production, producing a significant number of strawberries annually. However, strawberry cultivation is very labor-intensive, and the climate in South Korea is characterized by a temperate monsoon climate, with hot and humid summers and cold and dry winters. These climatic conditions pose complex challenges for strawberry cultivation, including harvest timing and quality maintenance [2]. One of the main challenges in

Manuscript Received: September. 16, 2024 / Revised: September. 21, 2024 / Accepted: September. 26, 2024
Corresponding Author: ksc0226@kw.ac.kr
Tel:+82-2-940-8637, Fax: +82-50-4174-3258
Associate professor Graduate School of Smart Convergence, Kwangwoon University, South Korea

strawberry cultivation is ensuring the efficiency and quality of the harvest, which is largely dependent on the accurate sorting of strawberries [3]. Traditional strawberry harvesting methods mainly rely on manual labor, which is time-consuming, error-prone, and not sufficient to meet the growing demand for strawberries [4]. To address these issues, it is increasingly recognized that automated systems utilizing machine learning and computer vision techniques are needed in agriculture [5]. The main objective of this research is to improve the model to accurately classify the maturity of strawberries and develop a lightweight model to be embedded on mobile devices or other embedded devices to process data with embedded fast inference speed [6]. With these goals in mind, this study aims to optimize the model for embedding AI systems on devices in indoor growing environments. To this end, we propose an improved Faster R-CNN model, which uses MobileNetV3-Large as a backbone network to realize the lightweight of the model and introduces RoI Align to improve the spatial accuracy of the feature map.

2. Background Theory

2.1 Faster R-CNN

Faster R-CNN is a deep learning architecture for object detection that includes a region proposal network (RPN) and the Fast R-CNN detector as its main components [7]. The RPN is implemented as a full convolutional network and serves as the core of Faster R-CNN, efficiently generating high-quality region suggestions. The model can handle input images of different sizes and uses a convolutional layer to extract features from the image. The extracted features are utilized in two ways: One is used to suggest regions where objects are likely to be present via RPNs, and the other goes through an additional convolutional layer to generate more complex, high-dimensional features. These processed features are finally fed into a fully connected layer of the network, which performs the regression task of predicting the exact coordinates of the bounding box and the classification task of classifying the object. Figure 1. shows the structure of faster R-CNN.

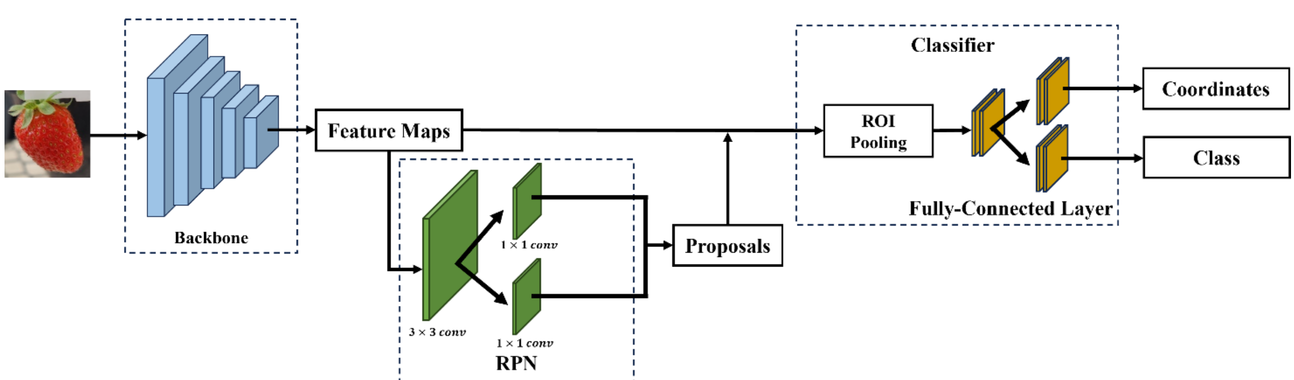

Figure 1. The structure of faster R-CNN

2.2 Related Work

2.2.1 Strawberry Detection

Recent advancements in deep learning techniques have significantly improved the accuracy and efficiency of strawberry detection. Li et al. developed a YOLOv5-ASFF algorithm for detecting and monitoring multistage strawberries in complex scenes, achieving 91.86% mAP and 88.03% F1 score, outperforming other mainstream algorithms [8]. Zhou et al. proposed a novel greenhouse-based system for strawberry detection and plumpness assessment using an improved Faster R-CNN model, which achieved an average fruit extraction accuracy of over 86% [9]. Wang et al. introduced a DSE-YOLO model for multi-stage strawberry detection, addressing issues such as small strawberry size and complex natural environments. Their model achieved F1 score values of up to 81.59% and mAP values of 86.58%, demonstrating effective detection of strawberries at every growth stage in natural scenes [10].

2.2.2 Strawberry Maturity Classification

Accurate classification of strawberry maturity is essential for optimizing harvesting processes and ensuring fruit quality. Various approaches using deep learning techniques have been developed to address this challenge. Yue et al. developed a fast and non-destructive method using smartphones to identify strawberry maturity stages, classifying them into immature, nearly mature, and mature categories [11]. This approach achieved high accuracy while providing a user-friendly solution for farmers. Zhou et al. proposed a more detailed classification system using both UAV and near-ground imaging techniques, coupled with the YOLOv3 deep learning algorithm [12]. Their method successfully classified strawberries into seven maturity stages for near-ground images and three stages for UAV images, demonstrating the potential of multi-scale imaging in maturity classification. Tao et al. further improved upon existing methods by introducing an enhanced YOLOv5 model, named YOLOv5s-BiCE, which incorporated advanced techniques such as the CARAFE module and a dual-attention mechanism [13]. This approach achieved superior accuracy in classifying five maturity stages of strawberries, including malformed fruits, while maintaining computational efficiency.

3. Experiment and Methods

3.1 Image Acquisition

In this study, we utilized the publicly available dataset KGCV_Strawberry for data collection [14]. The reason for choosing this dataset is that it allows for the detection and maturity classification of strawberries. For the purpose of collecting images for maturity classification, strawberries grown in a greenhouse were randomly selected. A total of 1477 images were collected, of which 1205 were taken indoors and 272 were taken outdoors. Each image contains more than 5000 strawberry objects, labeled across different growth stages. The labeled growth stages consist of background, flower, small green, green, white, turning red, and overripe. For each strawberry, a bounding box (BBox) has been labeled, which plays an important role in helping the model learn accurate detection and maturity classification of strawberries.

3.2 Experiments Environment

The hardware environment featured an AMD Ryzen 5 5600G processor with Radeon Graphics, an NVIDIA GeForce RTX 3060 graphics card, and 56GB of memory. The software environment utilized Windows 10 as the operating system and Python 3.8 as the programming language. For deep learning tasks, the PyTorch 2.1 framework was employed along with Torchvision 0.16 for computer vision operations. To leverage GPU acceleration, CUDA 12.1 and cuDNN 9.1 were used.

3.3 Proposed Method

3.3.1 Fruit Maturity Classification Framework

We propose a framework for analyzing the maturity of strawberries. This framework consists of two main networks. A Detection Network for strawberry detection and maturity classification, and a Trait Estimation Network for maturity classification after image patching. The Detection Network is based on an improved Faster R-CNN. This network identifies individual strawberries in the input image and creates a BBox that accurately represents the location of each strawberry. The BBox generated by the Detection Network is used to extract individual strawberry image patches from the original image. These patches are used as input to the Trait Estimation Network, which is based on the Densenet-121 structure and analyzes the maturity of each strawberry from its patches. With this two-step approach, we can automatically analyze the maturity of individual strawberries over time. Figure 2. shows the overall Fruit Maturity Classification Framework.

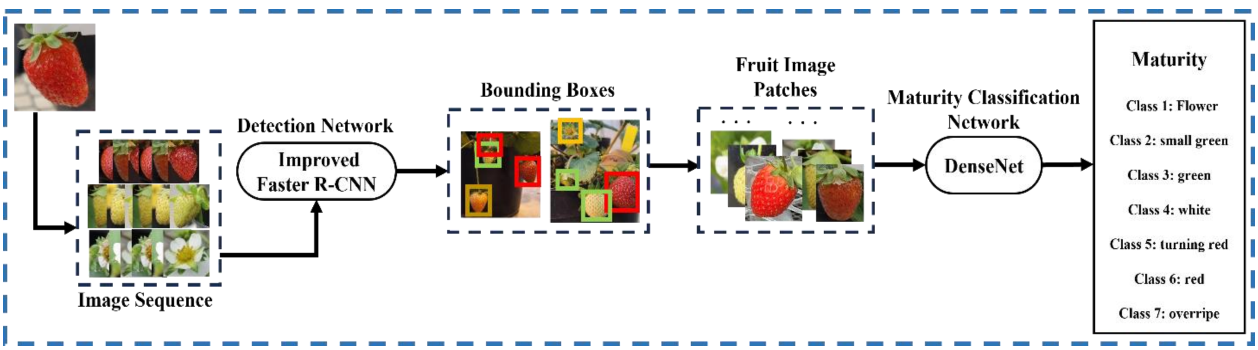

Figure 2. Fruit Maturity Classification Framework

3.3.2 Improved Faster R-CNN

In this study, we propose two main modifications to improve the performance of Faster R-CNN. First, we replace the existing ResNet50 backbone network with MobileNetV3-Large. MobileNetV3-Large is designed for efficient operation on mobile and embedded devices, utilizing depth-wise separable convolutions and the hard-swish activation function [15,16]. This replacement significantly reduces computational cost while maintaining detection performance, enabling effective object detection even in resource-constrained environments. Secondly, we implement RoI Align in place of RoI Pooling. RoI Align addresses the spatial misalignment issue inherent in RoI Pooling by eliminating quantization and using bilinear interpolation. This modification preserves more accurate spatial information, which is particularly crucial for detecting small objects like strawberries. These enhancements collectively improve both the inference speed and detection accuracy of the model, making it more suitable for real-time applications on low-power devices [17]. Figure 3. shows the architecture of our Improved Faster R-CNN.

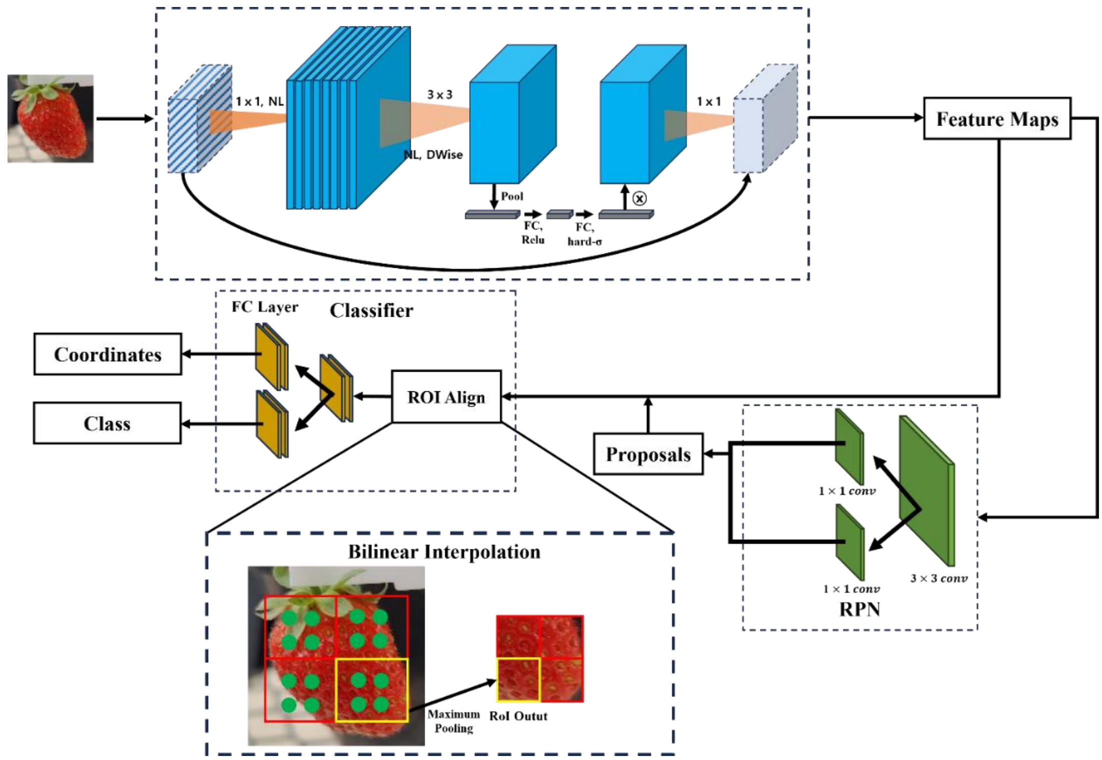

Figure 3. Improved Faster R-CNN Architecture

3.3 Evaluation Metrics

Several evaluation metrics are used to assess the performance of detection and classification models based on a dataset. The main metrics are precision, recall, F1 Score, mAP50, and mAP50:95. These metrics are calculated based on the results of True Positive (TP), True Negative (TN), False Positive (FP), and False Negatives (FN). The following shows the important metrics to evaluate the performance of the model during model training.

Formula (1) shows Precision, which measures the percentage of cases that the model predicts as positive that are positive, and indicates the model's ability to avoid FP.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

Equation (2) shows Recall, which measures the proportion of true positive cases that the model correctly predicts as positive, and evaluates the model's ability to identify TP.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Equation (3) shows F1-Score, which is a harmonic average of precision and recall, providing a single performance metric that balances the two metrics.

$$\mathrm{F1\ Score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{3}$$

Finally, Equation (4) shows the Mean Average Precision (mAP), which is used to evaluate the overall performance of the model, calculating the average value of Average Precision (AP) for each class. Where N is the total number of classes, and $\mathrm{AP_i}$ is the AP of class i.

$$\mathrm{mAP} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{AP_i} \tag{4}$$

mAP50 shows the average precision at an Intersection over Union (IoU) threshold of 0.5, which evaluates the accuracy of object detection. mAP50-95 is the average value of the average precision calculated by varying the IoU threshold from 0.5 to 0.95 in 0.05 intervals, which comprehensively evaluates the performance of the model at different IoU thresholds.

4. Result

4.1 Strawberry Classification Maturity Stage

The study implemented a two-stage approach for strawberry maturity classification. First, an improved Faster R-CNN detection network was used to identify and localize strawberries in the input images, generating bounding boxes around detected fruits. These detected regions were then extracted as fruit image patches. In the second stage, a DenseNet-121 classification network was employed to classify the maturity of each detected strawberry based on these image patches. Table 1 shows a comparison of different backbone networks for the detection stage (ResNet18, ResNet34, ResNet50, ResNet101) and our improved model. For each model, the average precision, average recall, F1 score, mAP0.5, mAP0.5:0.95, and inference time in milliseconds (ms) were evaluated to assess their performance in detecting strawberries across different maturity stages.

Table 1. Performance on Fruit Maturity Classification

	Average Precision	Average Recall	F1 Score	mAP0.5	mAP 0.5:0.95	Inference Time(ms)
Resnet18+FPN	0.6969	0.6899	0.6919	0.58	0.55	29.7557
Resnet34+FPN	0.7063	0.6981	0.7014	0.62	0.58	29.8223
Resnet50+FPN	0.7157	0.7091	0.7108	0.655	0.60	30.1787
Resnet101+FPN	0.7251	0.7180	0.7196	0.67	0.62	30.2366
Ours	0.7135	0.7107	0.7121	0.63	0.58	**27.6543**

The experimental results show that models using ResNet family backbones perform well overall, with ResNet101 having the highest performance and accuracy. Our improved Faster R-CNN-based model, while designed for lightweight efficiency, maintains competitive performance across all classes. While achieving a similar level of performance to ResNet50, our proposed model attains a processing time of 27.6543 ms, which is about 7% faster than the fastest ResNet model. This demonstrates that our lightweight structure, enhanced by RoI Align technology, successfully balances high performance with fast processing speed in strawberry maturity classification.

5. Conclusion

In this study, we propose an improved Faster R-CNN model for strawberry maturity classification. By using MobileNetV3-Large as the backbone network and introducing RoI Align, we improve the inference speed of the model while minimizing the performance degradation. Compared to the existing Faster R-CNN model, the inference speed is improved with a slight loss of accuracy. We show that we can effectively classify various strawberry maturity stages in a greenhouse environment. This can provide important information for determining the harvesting time of strawberries in a greenhouse environment. Therefore, it is necessary to study the inference of the model on various hardware-based embedded devices as well as mobile devices by making it lightweight. Future research should increase the speed of inference while improving the generalization ability of the model to increase accuracy on embedded devices. In addition, the applicability of the model to other varieties of strawberries or similar fruits should be explored to increase the generality of the model and optimize it for embedded systems in agricultural environments. This will allow the results of this study to be more effectively utilized in agricultural environments.

6. Acknowledgement

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-RS-2023-00258639) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

References

[1] F. Giampieri, S Tulipani, M. Alvarez-Suarez, L. Quiles, B. Mezzetti and M. Battino., "The strawberry: composition, nutritional quality, and impact on human health," Nutrition, Vol. 28, No. 1, pp. 9-19, 2012. DOI: https://doi.org/10.1016/j.nut.2011.08.009

[2] N. Wang, G. Li, J. Yang, Y. Zhang, and H. Wang, "Effects of Climate Change on Strawberry Production and Quality in East Asia," in Horticultural Science & Technology, Springer, vol. 33, no. 4, pp. 515-525, 2015. DOI: https://doi.org/10.1007/s10341-015-0249-2.

[3] Y. Chen, J. Bin, and C. Kang, "Application of machine vision and convolutional neural networks in discriminating tobacco leaf maturity on mobile devices," Smart Agricultural Technology, vol. 5, p. 100322, 2023. DOI: https://doi.org/10.1016/j.atech.2023.100322.

[4] Q. Feng, G. Zhou, Y. Wang, L. Yang, and H. Wang, "Strawberry harvesting robot for elevated-trough culture," International Journal of Agricultural and Biological Engineering, Vol. 12, No. 1, pp. 27-37, 2019. DOI: https://doi.org/10.25165/j.ijabe.20191201.4634

[5] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70-90, 2018. DOI: https://doi.org/10.1016/j.compag.2018.02.016.

[6] Y. Xiong, Y. Ge, L. Grimstad, and P. J. From, "An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation," Journal of Field Robotics, Vol. 37, No. 2, pp. 202-224, 2020. DOI: https://doi.org/10.1002/rob.21889

[7] R. Girshick, "Fast R-CNN", Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448, 2015. DOI: https://doi.org/10.1109/ICCV.2015.169

[8] Y. Li, J. Zhang, H. Zhang, C. Liu, and Z. Wang, "YOLOv5-ASFF: A Multistage Strawberry Detection Algorithm Based on Improved YOLOv5," Agronomy, Vol. 13, No. 7, 1901, 2023. DOI: https://doi.org/10.3390/agronomy13071901

[9] C. Zhou, J. Lin, S. Liu, Y. Liu, and Y. Wang, "A Novel Greenhouse-Based System for the Detection and Plumpness Assessment of Strawberry Using an Improved Deep Learning Technique," Frontiers in Plant Science,

Vol. 11, 559, 2020.
DOI: https://doi.org/10.3389/fpls.2020.00559

[10] Y. Wang, Z. Zhang, H. Li, L. Li, and Y. Liu, "DSE-YOLO: Detail Semantics Enhancement YOLO for Multi-Stage Strawberry Detection," Computers and Electronics in Agriculture, Vol. 200, 107057, 2022.
DOI: https://doi.org/10.1016/j.compag.2022.107057

[11] X. Q. Yue, Y. X. Liu, L. H. Zhang, and Z. Y. Zhang, "A smart data-driven rapid method to recognize the strawberry maturity," Information Processing in Agriculture, Vol. 7, No. 4, pp. 575-584, 2020.
DOI: https://doi.org/10.1016/j.inpa.2019.10.005

[12] X. Zhou, Z. Zeng, Y. Wang, J. Cao, and L. Zhang, "Strawberry Maturity Classification from UAV and Near-Ground Imaging Using Deep Learning," Smart Agricultural Technology, Vol. 1, 100001, 2021.
DOI: https://doi.org/10.1016/j.atech.2021.100001

[13] Z. Tao, Y. Wu, Q. Huang, X. Li, and Y. Zhang, "Strawberry Maturity Recognition Based on Improved YOLOv5," Agronomy, Vol. 14, No. 3, 460, 2024.
DOI: https://doi.org/10.3390/agronomy14030460

[14] Q. Yang, L. Zhang, Y. Wang, Y. Liu, and J. Chen, "Predicting the growth trajectory and yield of greenhouse strawberries based on knowledge-guided computer vision," Computers and Electronics in Agriculture, Vol. 220, 108911, 2024.
DOI: https://doi.org/10.1016/j.compag.2024.108911

[15] R. Avenash and P. Viswanath, "Semantic Segmentation of Satellite Images using a Modified CNN with Hard-Swish Activation Function," Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), pp. 413-420, 2019.
DOI: https://doi.org/10.5220/0007469604130420

[16] A. Howard, M. Sandler, G. Chu, L. C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1314-1324, 2019.
DOI: https://doi.org/10.1109/ICCV.2019.00140

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961-2969, 2017.
DOI: https://doi.org/10.1109/ICCV.2017.322