Industrial Engineering and Applications
L.-C. Tang (Ed.)

doi:10.3233/ATDE230049

Hot Work Control Measures Using CNN-Based Object Detection and Projective Geometry for Industrial Surveillance Application

Pakcheera CHOPPRADIT [a] Chaitat UTINTU [a] Kasisdis MAHAKIJDECHACHAI [a]
Vasin SUTTICHAYA [a] Teepakorn TEEPAKORN [a] and Ek THAMWIWATTHANA [a]
[a] *AI and Robotics Ventures Co.,Ltd.*

Abstract. Industrial safety management has been a common challenge for many industries to implement since industrial hazards could cause fatal risks and unscheduled downtime. In this paper, we proposed an alternative approach for hot work control measures using CNN-based object detection and projective geometry, which could be integrated with the existing surveillance system. This method aims to monitor hot work activity and implement the risk assessment policy, which could control by the hazard area control. The dataset for our study consisted of 909 images of hot work activities captured by two closed-circuit television (CCTV) cameras. There are two steps to our methodology, which are the object detection stage and the bird's-eye perspective transform stage. In the first stage, Workers, Welders, and Hot works are localized using an object detection algorithm, which is YOLOv5. To maximize the F1-score performance of object detection, we ran the experiments to train YOLOv5 with three levels of augmentations: low, medium, and high. For the second stage, four points are required in the method of transforming the object's Cartesian coordinates into the new coordination in a bird's-eye perspective. The radius distance threshold has to be manually calibrated for each specific camera point of view. If there is a worker that moves into the hot work radius, the violation alarm is triggered. The results show that medium augmentations produce the best results, with an overall mAP and F1-score of 0.77 and 0.74, respectively. In addition, the predefined distance threshold is also required and can vary in the different scenarios in the bird's-eye perspective transform stage.

Keywords. hot work, control measures, object detection, projective geometry, industrial safety management

1. Introduction

Industrial safety management is crucial, yet it remains challenging for many organizations to implement and maintain. It refers to the safe practices that aim to escort business premises and on-site workers from the risk of fatal hazards. Since poor safety management can lead to accidents and unscheduled downtime, many industries, especially construction sites, have employed intelligent systems to be fully aware of potential high-hazard situations. Hot work operations, such as welding, cutting, soldering, and other ac-

tivities that involve open flames, spark production, or heat, are common cases of dangers that can be disastrous for on-site employees and assets. Various hot work hazards should be considered in the risk assessments, for example, electrical hazards, UV or infrared light radiation, dangerous fumes, and spark combustion. The essential aspect of hot work safety management is to carry out hot work activities properly to minimize risk. Hence, hot work occurrences must be identified, and then effective control measures must be applied. Control measures could be diverse, depending on the uniqueness of each business's operations and work environment. However, wearing appropriate personal protective equipment (PPE), identifying hot work activities in a general-use area, and performing hot work in a restricted location are all frequent habits that could be tracked using the intelligent surveillance system. To monitor hot work activities and limit access to such areas in real-time, we proposed a method based on deep learning and projective geometry algorithms, which could be applied to the existing surveillance system, such as the CCTV system.

This paper proposes the hot work control measure approach based on deep learning and projective geometry technique. YOLOv5 [1], the novel one-stage object detection, is chosen to localize all workers, welders, and hot works positions on the image with three levels of augmentation. Then, the projective geometry algorithm, which is perspective transformation, is applied to reveal all hot work coordinates in a bird's-eye perspective. Thus, violations will be declared when the pair-wise Euclidean distances between hot work and worker exceed the predefined distance threshold. Furthermore, the proposed method might also be utilized to trigger the appearance of hot work activity.

This paper is organized as follows. Section 2 describes the motivation according to related works. The core methodology, including object detection and projective geometry, is presented in Section 3. Section 4 presents the experimental results and the analysis of the results. Finally, the conclusion and the discussion on the limitations are explained in the last section.

2. Related Works

Control measures could be implemented in different strategies. For example, one efficient way to improve safety performance was to learn from past accidents. Traditionally, hazard records were gathered by managers to investigate their patterns of manifestation, thus they could analyze the root cause and plan the strategic policy to prevent them from reoccurring. However, analyzing such unstructured reports could be time-consuming and inefficient. Many works attempted to tackle this problem. Therefore, the industrial safety analysis was divided into predictive and retrospective methods [2]. Many papers suggest a retrospective method based on text mining and deep learning model [2–4]. Firstly, the topic words in the accidents were extracted using the latent Dirichlet allocation (LDA) to create cause topics, then the accident's cause was predicted using Convolutional Neural Networks (CNNs) and the previous cause topics. Text mining methods could help to find and sort out the cause of the accident faster. These key causes can be utilized to develop the optimal safety measures to minimize the number of industrial accidents as well as hot work hazards [4]. Although these proposed techniques were straightforward, however, the overall performance might depend on the regularity and amount of hazard records dataset. From these aspects, industrial control measures that don't rely on historical data and could be enforced in a real-time scenario might be the alternative approach.

In the last few years, computer vision techniques were proven to be useful in various applications such as control measures in the surveillance system. There were numerous studies on control measures to monitor social distancing during the COVID-19 pandemic outbreak. Recent works proposed a methodology to estimate the distances of the pedestrian by using object detection along with projective geometry to calculate distances between humans in the bird's-eye perspective [5–9]. Therefore, it was possible to compute point correspondences between 2D and 3D worlds with this approach. For industrial safety management, we were motivated by this technique to implement the online risk assessment system for hot work activities, since the proposed method could be integrated into the existing surveillance system and could also be implemented as an efficient control measure in the construction site.

3. Methodology

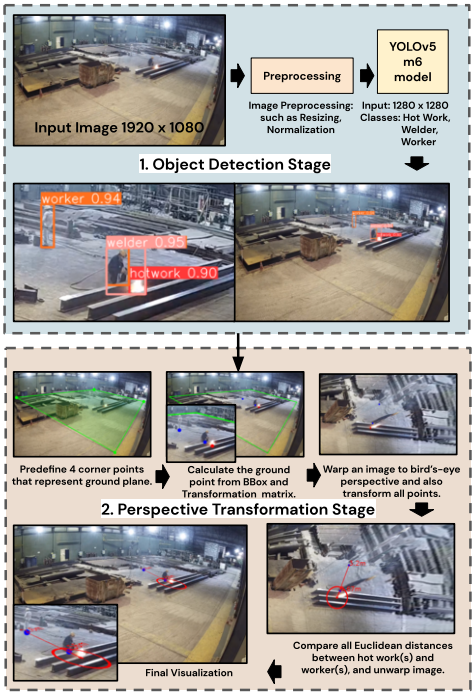

Figure 1. The overview of core methodology.

This section describes the deep learning and projective geometry techniques which were utilized in the proposed method. To integrate effective hot work control measures with the existing surveillance system, our approach must restrict access when some workers carry out hot work activity and attempt to warn the safety officer in real-time when others enter the predefined distance around hot work activity, which could identify as the risk of hazard. The proposed method is divided into two stages, which are object detection and perspective transformation. The overall architecture is illustrated in Fig. 1.

3.1. Hot Work and Worker Detection with Object Detection

Object detection aims to determine the location and type of objects present in the input image. Generally, object detection methods based on CNN are classified into two types: two-stage and one-stage approaches. The two-stage object detection algorithms are designed to follow the standard object detection pipeline, which includes region proposal and classification tasks. In the first stage, region proposals are generated, then each proposal is delivered to the second network to be classified. Although this paradigm yields strong localization and classification results, it has certain limitations, including huge computational time. The examples of two-stage object detectors are RCNN [10], fast-RCNN [11], and faster-RCNN [12]. The one-stage object detection algorithms were proposed to mitigate the real-time bottleneck issue. Instead of splitting tasks, the region proposal and classification are performed simultaneously by a single network. This improvement greatly reduces the computational complexity, resulting in increased speed while still maintaining accuracy when compared to two-stage approaches. There are many well-known works according to this technique, including SSD [13] and YOLO [14].

You Only Look Once (YOLO) is the well-known object detection method developed by Redmon et al. [14]. As the name implies, YOLO is a one-stage object detector that solves the problems of object location and classification with a single forward propagation. Furthermore, to approach the detection problem as the regression problem, the input image is divided into grid cells, with each grid cell responsible for predicting bounding box confidence and class probability simultaneously. Finally, the final bounding box and classification are predicted by results aggregation. From those properties, YOLO has been widely used and known for its detection accuracy in real-time applications. Furthermore, various works have been proposed to improve the original YOLO algorithm. The first three versions of YOLO, including YOLOv1 [14], YOLOv2(YOLO9000) [15], YOLOv3 [16] were developed by the original YOLO authors. After that, another research group come up with novel ideas and officially published YOLOv4 [17], which was followed by the GitHub repository YOLOv5 [1] developed by the company Ultralytics. In YOLOv5, adaptive bounding box anchors, mosaic data augmentation, and adaptive image filling are integrated to improve preprocessing capability. The architecture consists of three main parts (i.e., backbone, neck, and head) as described in Fig. 2. For the backbone, cross-stage partial network (CSPNet) [18] and spatial pyramid pooling (SPP) [19] are utilized to handle multiscale feature extraction. In the neck, YOLOv5 used the feature pyramid structures of FPN [20] and PANet [21], resulting in improved detection efficiency. Finally, the head output is utilized to predict objects of various sizes on feature maps. There are five different architectures offered by YOLOv5, including YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. These variations relate to the number of feature extraction modules and convolutional kernels used in the network. On top of that, YOLOv5 P6 models, which added the extra large object output layer and pre-trained with a resolution size of 1280×1280, are also available.

For our proposed method, we choose the medium size of YOLOv5 P6 (YOLOv5m6) since the native resolution of our dataset is 1920×1280 and hot work objects are tiny when compare to the whole image. Three hundred epochs and 100 patients with three levels of augmentation were used to train the model. The three object classes used for training are worker, welder, and hot work, where worker class refers to a person working here in the workshop, hot work class refers to the spark produced by the welder, and welder class refers to the worker performing hot work tasks.

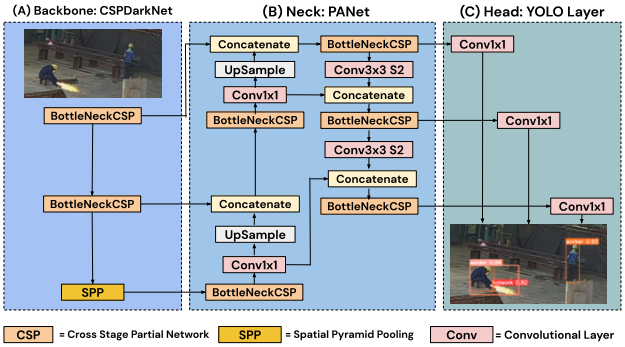

Figure 2. YOLOv5 Architecture.

3.2. Hot Work Control Measures with Projective Geometry

The process of shifting from one perspective to another perspective with matrix multiplication is known as perspective transformation. The popular view of transforming is a bird's-eye view. The technique of transforming from one plane to another plane is known as homography, and it can be used to convert images from any view to bird's-eye views. Brief homography is written according to the Equation (1).

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} \hat{x_1} \\ \hat{x_2} \\ \hat{x_3} \end{bmatrix} \tag{1}$$

where (x_1, x_2, x_3) is coordinates on ground plane, $(\hat{x_1}, \hat{x_2}, \hat{x_3})$ is coordinates on image and H is the homography matrix.

To find homography matrix x_1 and x_2 are divided by x_3 as in the Equation (2) and Equation (3). After rearranging the equation, constrained least squares are used to solve H matrix. Due to the degree of freedom of H matrix, four points are the requirement for solving the homography matrix.

$$\frac{x_1}{x_3} = \frac{h_{11}\hat{x_1} + h_{12}\hat{x_2} + h_{13}\hat{x_3}}{h_{31}\hat{x_1} + h_{32}\hat{x_2} + h_{33}\hat{x_3}} \tag{2}$$

$$\frac{x_2}{x_3} = \frac{h_{21}\hat{x_1} + h_{22}\hat{x_2} + h_{23}\hat{x_3}}{h_{31}\hat{x_1} + h_{32}\hat{x_2} + h_{33}x_3} \tag{3}$$

The welder object is used to prevent false positive by confirming the intersection of the bounding box between the welder and hot work object. The original image, as shown in Fig. 3, requires four points of the ground plane for transforming the ground plane into a bird's-eye view.

After that, the homography matrix is calculated and used to transform an image into a bird's-eye view. The center bottom point of the human and hot work coordinates are also transformed to the same perspective of the ground plane. The center bottom points of hot work are used as the center of area control. After transforming the image and

Figure 3. Original image.

human points, finding the radius size of area control that equals one meter in real life is required for the first time in each area. The circles are generated and adjusted to match all four traffic cones that are arranged in a square with a size of 1×1 meter as in Fig. 4.

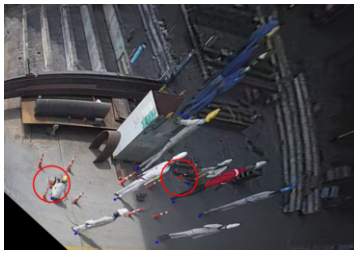

Figure 4. Bird's-eye perspective with fitting circle.

The Euclidean distance, as shown in Equation (4), can be used to compute the distance between the hot work and the worker.

$$D = \sqrt{(\hat{x_w} - \hat{x_h})^2 + (\hat{y_w} - \hat{y_h})^2} \tag{4}$$

where $(\hat{x_w}, \hat{y_w})$ is the transformed center bottom coordinate of worker bounding box and $(\hat{x_h}, \hat{y_h})$ is the transformed center bottom coordinate of hot work bounding box.

If the distance between the hot work and the worker is less than the radius's circle, the number of workers in the hazardous areas is counted and sent the alarm signal to the system. Fig. 5 illustrates an overall process where the red dot represents the hot work object's center bottom point, the blue dot represents the worker object's center bottom point, the red circle represents the area control, and the green line represents the distance between the hot work and worker objects.

4. Experiments

4.1. Dataset Preparation

From July 2021 to August 2022, two closed-circuit television (CCTV) cameras were set up to record the activities of construction site workers. The hot work activities also were captured in 1920×1080 pixels photos and 909 photos were taken from the entire video archive. Then, to prepare for training an object detection model, each worker, hot work,

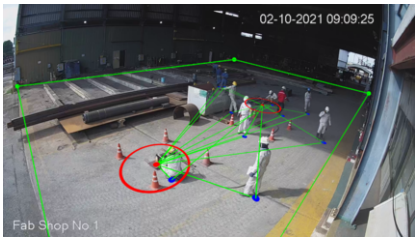

Figure 5. Overall hazardous area monitoring.

and welder object in the scene was labeled. The number of each instance is written in the following Table 1.

Classes	hot work	Welder	Worker	All
Instances	1003	944	5719	7666

Table 1. Number of instances each class in all dataset.

4.2. Experimental Setup

In our dataset, the pictures were taken from videos and the same object could be found in other images. Time series cross-validation in each workshop area with K equal to 5 was consequently chosen and the dates of the recorded image would be segregated in each train, test, and validation for the performance fairness measurement and to avoid model overfitting. The train and the test set were separated into 80 percent and 20 percent. Then, we use the train set to do a time series split with K equal to 5.

Since there aren't as many occurrences of hot work and welder objects, augmentation should be usually noticed in the training model. The YOLOv5 algorithm offers a variety of augmentation methods. For this article, we will train the model in the high, medium, and low levels of image augmentation referring to the YOLOv5 parameter *hyp* in the YOLOv5 framework. The probability of doing image augmentation and how the augmentation changed the images in the train loader stage are described for each level.

4.3. Result Analysis

The performance of the proposed system focuses on object detection performance because the failure of the object detection process will affect the system's overall efficiency. The overall performance of object detection was evaluated by mean average precision at threshold 0.5 (mAP0.5) and the average of mean average precision at 0.5 to 0.95 step by 0.05 (mAP0.5:0.95). Note that, mAP is famously used to measure overall object detection since it measures the performance in various thresholds. The results were summarized in Table 2.

The results suggest that the medium augmentation at mAP0.5 is mostly superior to the other augmentation. For mAP50:95, although the medium augmentation is not particularly superior to other augmentations, the performance is near the highest overall. Therefore, we prefer medium augmentation in this experiment.

We select the best confident threshold from the F1-score curve of the final validation fold. Note that, the confidence threshold can vary at each level of augmentation. Preci-

Metrics	Classes	Levels of augmentation		
		Low	Medium	High
mAP50	hot work	0.656	0.663	**0.699**
	Welder	0.787	**0.816**	0.754
	Worker	0.842	**0.844**	0.833
	Overall	0.762	**0.774**	0.762
mAP50:95	hot work	0.289	0.283	**0.29**
	Welder	**0.537**	0.518	0.486
	Worker	0.569	0.55	**0.57**
	Overall	**0.465**	0.45	0.449

Table 2. mAP0.5 and mAP0.5:0.95 results of the test set.

sion, Recall, and F1-score are the metrics for evaluating the results. F1-score is calculated as the average of Precision and Recall. Precision is the ratio of accurately predicted positive data to all positively predicted data. Recall is the ratio of accurately predicted positive data to all really positive data. The results are summarized in Table 3.

Table 3 shows that the medium augmentation is superior to others, according to F1-score and Recall. The low augmentation also received the highest Precision. However, the false negative will have less effect than the false positive in a real-world application. Thus, medium augmentation is recommended.

Metrics	Classes	Levels of augmentation		
		Low	Medium	High
Precision	hot work	0.772	0.765	**0.78**
	Welder	**0.908**	0.858	0.835
	Worker	**0.88**	0.848	0.864
	Overall	**0.853**	0.824	0.826
Recall	hot work	0.53	0.525	**0.561**
	Welder	0.639	**0.716**	0.639
	Worker	0.75	**0.779**	0.751
	Overall	0.64	**0.674**	0.65
F1-score	hot work	0.629	0.623	**0.653**
	Welder	0.750	**0.781**	0.724
	Worker	0.810	**0.812**	0.804
	Overall	0.731	**0.741**	0.728

Table 3. Precision, Recall and F1-score results of the test set.

5. Conclusion

For industrial surveillance applications, we presented a hot work control measures technique utilizing CNN-based object detection and projective geometry. Our methodology consists of two stages, which are the object detection stage and the bird's-eye perspective transform stage. The object detection stage is responsible for detecting three classes of objects, including Worker, Welder, and Hot work. The perspective transform stage uses four reference points to transform the coordination into the new coordination in the

bird's-eye perspective. After the workers and the hot works are localized, the hazard risk distance can be calculated from a birds-eye perspective.

The YOLOv5 model is applied in the object detection stage. We conducted experiments to train the YOLOv5 model using three levels of augmentation, including low, medium, and high augmentation. The experiment shows that the medium augmentation yields the best result with 0.77 mAP and 0.74 F1-score.

The perspective transform stage applies homography transform to the images from CCTV. The radius distance threshold has to be manually calibrated for each specific camera point of view. If there is a worker that moves into the hot work radius, the violation alarm is triggered. In addition, the predefined distance threshold is also required and can vary in different scenarios.

Acknowledgment

AI and Robotics Ventures Co., Ltd. (ARV) and Thai Nippon Steel Engineering & Construction Corporation Ltd. (TNS) provided support for this paper. We gratefully thank Thai Nippon Steel Engineering & Construction Corporation Ltd. (TNS) for giving us the dataset. We would like to give special thanks to our Machine Learning team, who always give full support and open for an advice in any time.

References

[1] *ultralytics/yolov5*. Oct. 2020. DOI: 10.5281/zenodo.4154370.

[2] Botao Zhong et al. "Deep learning and network analysis: Classifying and visualizing accident narratives in construction". In: *Automation in Construction* 113 (2020), p. 103089. DOI: 10.1016/j.autcon.2020.103089. URL: https://doi.org/10.1016%2Fj.autcon.2020.103089.

[3] Jianfeng Qiao et al. "Construction-Accident Narrative Classification Using Shallow and Deep Learning". In: *Journal of Construction Engineering and Management* 148.9 (2022). DOI: 10.1061/(asce)co.1943-7862.0002354. URL: https://doi.org/10.1061%2F%28asce%29co.1943-7862.0002354.

[4] Hui Xu et al. "Cause analysis of hot work accidents based on text mining and deep learning". In: *Journal of Loss Prevention in the Process Industries* 76 (2022), p. 104747. DOI: 10.1016/j.jlp.2022.104747. URL: https://doi.org/10.1016%2Fj.jlp.2022.104747.

[5] Yew Cheong Hou et al. "Social Distancing Detection with Deep Learning Model". In: *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*. IEEE, 2020. DOI: 10.1109/icimu49871.2020.9243478. URL: https://doi.org/10.1109%2Ficimu49871.2020.9243478.

[6] Mahdi Rezaei and Mohsen Azarmi. "DeepSOCIAL: Social Distancing Monitoring and Infection Risk Assessment in COVID-19 Pandemic". In: *Applied Sciences* 10.21 (2020), p. 7514. DOI: 10.3390/app10217514. URL: https://doi.org/10.3390%2Fapp10217514.

[7] Dongfang Yang et al. "A Vision-Based Social Distancing and Critical Density Detection System for COVID-19". In: *Sensors* 21.13 (2021), p. 4608. DOI: 10.3390/s21134608. URL: https://doi.org/10.3390%2Fs21134608.

[8] Sreetama Das et al. *Computer Vision-based Social Distancing Surveillance Solution with Optional Automated Camera Calibration for Large Scale Deployment*. 2021. DOI: 10.48550/ARXIV.2104.10891. URL: https://arxiv.org/abs/2104.10891.

[9] Prateek Khandelwal et al. *Using Computer Vision to enhance Safety of Workforce in Manufacturing in a Post COVID World*. 2020. DOI: 10.48550/ARXIV.2005.05287. URL: https://arxiv.org/abs/2005.05287.

[10] Ross Girshick et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 580–587. DOI: 10.1109/CVPR.2014.81.

[11] Ross Girshick. "Fast R-CNN". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169.

[12] Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.6 (2017), pp. 1137–1149. DOI: 10.1109/TPAMI.2016.2577031.

[13] Wei Liu et al. "SSD: Single Shot MultiBox Detector". In: *Computer Vision – ECCV 2016*. Ed. by Bastian Leibe et al. Cham: Springer International Publishing, 2016, pp. 21–37. ISBN: 978-3-319-46448-0.

[14] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016). DOI: 10.1109/cvpr.2016.91.

[15] Joseph Redmon and Ali Farhadi. "YOLO9000: Better, Faster, Stronger". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 6517–6525. DOI: 10.1109/CVPR.2017.690.

[16] Joseph Redmon and Ali Farhadi. *YOLOv3: An Incremental Improvement*. 2018. URL: http://arxiv.org/abs/1804.02767.

[17] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-yuan Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection". In: (Apr. 2020).

[18] Chien-Yao Wang et al. "CSPNet: A New Backbone that can Enhance Learning Capability of CNN". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2020). DOI: 10.1109/cvprw50498.2020.00203.

[19] Kaiming He et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.9 (2015), pp. 1904–1916. DOI: 10.1109/tpami.2015.2389824.

[20] Tsung-Yi Lin et al. "Feature Pyramid Networks for Object Detection". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017). DOI: 10.1109/cvpr.2017.106.

[21] Shu Liu et al. "Path Aggregation Network for Instance Segmentation". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018). DOI: 10.1109/cvpr.2018.00913.