Benchmarking performance of object detection under images distortions of an uncontrolled environment

Ayman Beghdadi, Malik Mallem, Lotfi Beji

BENCHMARKING PERFORMANCE OF OBJECT DETECTION UNDER IMAGES DISTORTIONS OF AN UNCONTROLLED ENVIRONMENT

Ayman Beghdadi, Malik Mallem and Lotfi Beji

IBISC Lab, Univ Evry, Paris-Saclay University, Evry, France
aymanaymar.beghdadi@univ-evry.fr, malik.mallem@univ-evry.fr, lotfi.beji@univ-evry.fr

ABSTRACT

The robustness of object detection algorithms plays a prominent role in real-world applications, especially in the uncontrolled environments due to distortions during image acquisition. It has been proved that object detection methods suffer from in-capture distortions to perform a reliable detection. In this study, we present a performance evaluation framework of the state-of-the-art object detection methods on a dedicated dataset containing images with various distortions at different levels of severity. Furthermore, we propose an original strategy of image distortion generation applied to the MS-COCO dataset that combines some local and global distortions to reach a better realism. We have shown that training with this distorted dataset improves the robustness of models by 31.5%. Finally, we provided a custom dataset including the natural images distorted from MS-COCO to perform a more relevant evaluation of the robustness concerning distortions. The database and the generation source codes of the different distortions are publicly available[1,2].

Index Terms— Deep learning, Object detection, Distortion, Robustness, Benchmark

1. INTRODUCTION

Object detection and recognition is one of the most challenging and widely studied problems in computer vision and especially in uncontrolled environment. Deep Neural Networks (DNN) based approaches [1, 2, 3] have been proven very promising paradigms to solve such complex problems in various real-world applications. However, generally, the proposed DNN-based architectures do not consider the sensitivity of the learning models to common data distortions and corruption effects. Indeed, it has been shown by Szegedy et al.[4] that a small perturbation invisible to the human visual system can have significant repercussions on the performance of object recognition models. It is therefore important to conduct ad hoc studies to understand the effects of such perturbations and to be able to generate them at different levels in order to develop more robust object detection networks. The limitations of object detection methods, in terms of robustness against some common signal degradation, data corruptions, or adversarial examples, have been highlighted in several works [5, 6, 7, 8]. A solution would be to perform pre-processing before the high-level phases. But this implies being able to identify the distortion automatically in order to apply the most appropriate solution. A more elegant and efficient solution present in many research [9, 10, 11, 12] is to enrich the database used for learning by simulating the most relevant distortions. It is in this last framework that the proposal described in this work is made. Thereby, we propose a full framework evaluation of robustness for a set of object detection methods [13, 14, 15] through several distortions applied on the MS-COCO dataset [16]. This data augmentation is performed through some common global distortions (noise, motion blur, defocus blur, haze, rain, contrast change, and compression artefacts) in the whole image and some local distortions (local blur motion, BackLight illumination and local defocus blur) in specific areas that include the possible dynamic objects or scene conditions. To the best of our knowledge, our work is the first to consider the local distortions to assess the robustness of object detection models. Finally, we proceeded to a manual selection of the distorted images present in the validation set of the MS-COCO dataset and retained only the annotations of the distorted objects to perform a more relevant evaluation of the impact of distortions on models. The main contributions of our study are summarized in the following:

- A comprehensive evaluation of the robustness of the state-of-the-art object detection methods against global and local distortions at 10 levels of distortion is provided (see section 4.1).
- A dataset dedicated to the study of the impact of local and global distortions on the robustness of object detection is built from MS-COCO dataset [17].
- It is shown that the robustness of object detection process is improved by enriching the training process by using data augmentation based on generated distorted images (see table 3).
- An additional new benchmark dataset constructed for natural in-capture distortions from MS-COCO dataset that allows to better assess the reliability of the object detection models in case of real distortions (see tables 2 and 4) [18].

The remainder of the paper is organized as follows. Section 2 summarizes previous related literature. Section 3 is devoted to detail the methods of evaluation and the distortion generations. Then, section 4 is dedicated to show and discuss results. Finally, concluding remarks and perspectives are provided in section 5.

2. RELATED WORK

The evaluation of object detection methods from images acquired in an uncontrolled environment has been the subject of a few studies [19, 20]. But most of the time, the databases used contains images with limited number of degradation types and therefore cannot be generalized to practical cases where one has to face several distortions. Some papers deal with the impact of various distortions on the detection performance such as in the study [21] that attempts to evaluate the robustness of some backbones of CNN architectures against geometric transformations. A comparative study between

[1] https://github.com/Aymanbegh/Benchmarking-performance
[2] https://github.com/Aymanbegh/Distorted-Natural-COCO

(a) Original (b) Local defocus blur (c) BackLight illumination (BL) (d) Local motion blur

Fig. 1: Illustration of some local distortions.

DNN-based architectures (ResNet-152, VGG-19, GoogLeNet) and human observer performance was conducted in [22]. In this study twelve common distortions are considered in the performance evaluation of object recognition methods. It has been shown that these DNN-based architectures perform remarkably well in the case of the distortions on which the training was done. However, this performance decreases, compared to that of the human observer, for unseen distortions. Another interesting study on the evaluation of the robustness of classification models of images from the imagenet database and subjected to fifteen artificial distortions at five levels of severity, revealed the effect of image quality degradation on the stability of the training process for various DNN-based architectures [6]. Recently, a benchmark study on the robustness of many object detection models against 15 types of global distortions and stylized images through data augmentation process has been done in [5].

3. PROPOSED BENCHMARKING FRAMEWORK

In order to evaluate the robustness of the considered models against image distortions in an uncontrolled environment, we propose 10 types of distortions (global and local) at 10 linearly increasing levels of severity [17]. In figures 1 (b), (c), and (d), distortion levels are 8, 7, and 10 respectively.

3.1. Image distortions

Global distortions affect the image as a whole and come from different sources related in general to the acquisition conditions. Some are directly dependent on the physical characteristics of the camera and are of photometric or geometric origin. Among the most common distortions that affect the quality of the signal are defocus blur (Defocus-Blur), photon noise, geometric or chromatic aberrations, and blur (Motion-Blur) due to the movement of the camera. The other types of degradation are related to the environment and, more particularly, the lighting and atmospheric disturbances in the case of outdoor scenes. Compression and image transmission artifacts are another source of degradation that is difficult to control. These common distortions have been already considered in benchmarking the performance of some models [5, 6, 8].
Local distortions are undesirable signals affecting one or more localized areas in the image (see figure 1). A typical case is the Motion Blur (Loc. MBlur) due to the movement of an object of relatively high speed. Another photometric distortion is the appearance of a halo around the object contours due to the limited sensitivity of the sensors or backlight illumination (BackLight). The artistic blur (Loc. Defoc.) affecting a particular part of the targeted scene, the object to be highlighted by the pro-shooter, is another type of local distortion. Thus, integrating the local distortions in the database increases its size and makes it richer and more representative of scenarios close to real applications. It is worth noticing that these local distortions are made possible thanks to the annotations provided in the MS-COCO database and, in particular, the details of the shape and location of the object of interest.

3.2. Description of the considered models

In this study we focus on few models of state-of-the-art object detection methods which have the best performance or notoriety, namely YOLOv4 [13], Mask R-CNN [15] and EfficientDet [14] with distinct backbone architectures that highlights the correlation between the backbone and their robustness. The table 1 summarizes the different models and their components which are used in this benchmarking.

Table 1: State-of-the-art object detection models.

Methods	Architecture	Size	Add-blocks
Mask R-CNN [15]	Resnet-101	512	FPN, RPN [23, 24]
YOLOv4 [13]	CSPDarknet53 YOLOv4-tiny	512 416	PAN [25]
EfficientDet [14]	EfficientNet-B0 EfficientNet-B1 EfficientNet-B2 EfficientNet-B3 EfficientNet-B4	512 640 768 896 1024	Bi-FPN [14]

FPN: Feature Pyramid Network, RPN: Region Proposal Network, PAN: Path Aggregation Network, Bi-FPN: Bi-directional FPN.

3.3. Robustness Evaluation against Natural Distortions

We manually identified the natural distortions present in the MS-COCO validation set to better assess the robustness in real scenarios. Indeed, in the existing methods, robustness evaluations were performed on synthetic distortions or specific datasets, limiting those studies' extension in real-world applications. It is worth noticing that the selected sub-sets from MS-COCO contain images with global distortions and the associated object annotations. In contrast, only affected objects are considered in the case of local distortions (see figure 2). This benchmarking framework for robustness evaluation uses one of the most widespread databases for object detection, restricted to a limited set of seven natural distortions types (see table 2). In this way, we are laying the foundations for a future common

(a) Original annotations (b) Sorted annotations

Fig. 2: Natural distortions selection in the MS-COCO dataset.

evaluation method for evaluating object detection robustness against real-world distortions. It is important to notice that some natural

Table 2: Features of distorted natural sub-sets.

	Noise	Contrast	Blur	Defoc.	Rain	O.Blur	BackLight
Images	44	42	32	201	21	127	128
Objects	289	312	224	1299	110	464	934
Ratio	1.0	0.83	0.97	0.72	0.95	0.34	0.68

Ratio: Number of retained objects/annotated objects in images.

distortions such as Defocus, Object Blur, and BackLight have low ratios (see table 2). This results from the nature of local distortions, which only affect specific objects in the images (see figure 2).

4. RESULTS AND DISCUSSIONS

4.1. Evaluation of the robustness

We evaluated the robustness of models against distortions on the MS-COCO 2017 validation set, which contains 5K annotated images that we have corrupted through our 10 types of distortions [17].

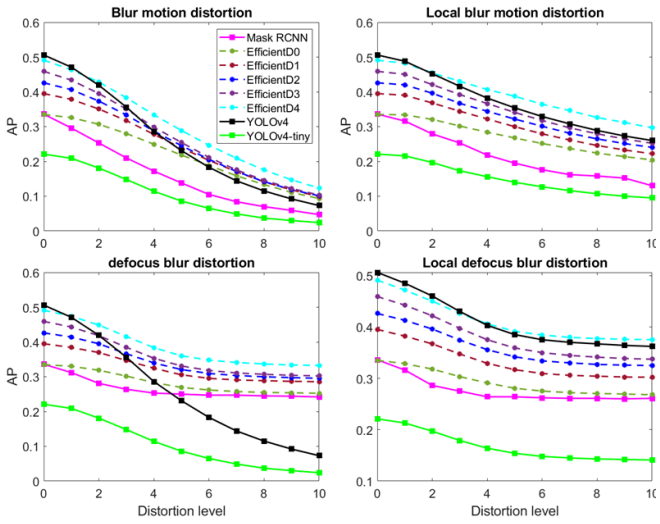


Fig. 3: Robustness of object detection models against global(left)/local (right) distortions.

Note that level 0 of distortion represents the COCO AP score for non-distorted images. We assessed the object detection performance of models through the AP (Average Precision) from the COCO metrics that provides a relevant performance indicator. Figures 3 and 4

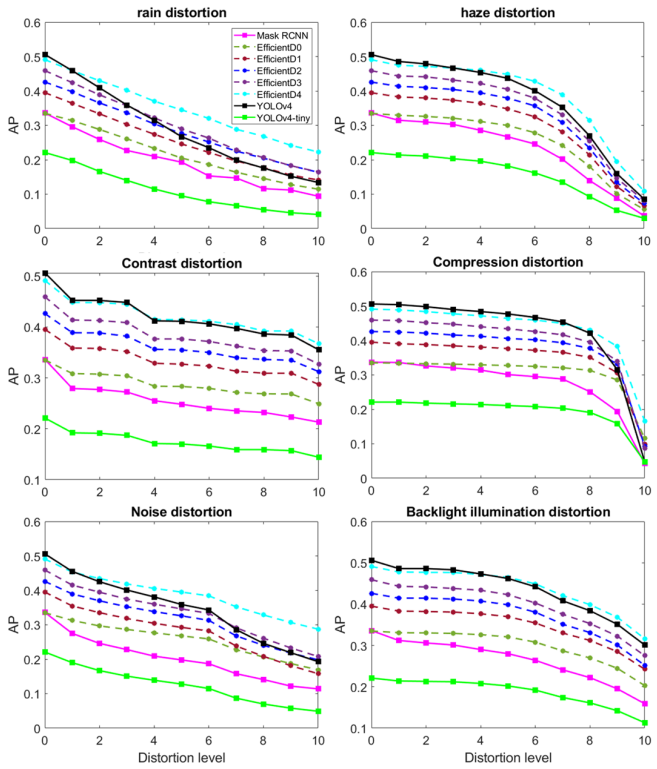
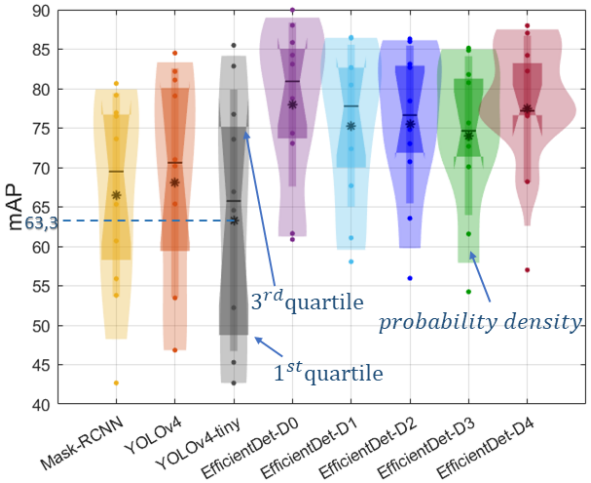


Fig. 4: Robustness of object detection models against distortions.

represent the evolution of the AP score of models according to each distortion type and level. These results show that distortions reduce object detection performance of models by 20% to 89.2%. Moreover, they highlight the weaknesses of YOLOv4 methods against distortions (strongly decreasing curves in figures 4 and 3), which need a deeper study. The evaluations in figure 3 indicated that global distortions have more impact than local distortions on the models. While local and atmospheric disturbances have a roughly uniform impact on methods despite their different architectures, the other distortions produce effects separate and independent. Indeed, overall models performance are greatly affected by distorted images but not uniformly. Each type of distortion has a different rate of impact on models performance depending on the distortion level and object detection method. In order to evaluate the impact of distortions on models relevantly, we computed the robustness rate that represents the ratio between mAP scores of models through each distortion type and level, and non-distorted images (level 0). This robustness rate is illustrated (see figure 5) with a data distribution visualization including boxplots and violin plots that contribute to extract richer information such as the mean, median, and probability density. The diversity of architectures with different network depths made it possible to estimate a possible correlation between this depth and the robustness of the models. Results from figures 3, 4 and 5 indicate that the EfficientDet method's robustness decreases despite its network depth and input size increasing through its different backbone architectures (D0 to D3). Furthermore, the YOLOv4-tiny model was the least robust to distortions with an average decrease in its AP score of



*/-: Mean/Median of mAP for all distortion type and level, Light/dark surrounding shape: probability density/percentiles.

Fig. 5: Average robustness rate of object detection models against the distortions.

36.7% between the distortion levels 0 and 10 (lowest average robustness rate at 63.3%, see mAP score in figure 5). Therefore, it seemed relevant to us to choose this model to evaluate the improvement of robustness through a data augmentation with our distorted dataset.

4.2. Training on distorted dataset

This section presents experimental results on MS-COCO 2017 evaluation set with 5K images. YOLOv4-tiny model is trained following the protocol in MS-COCO challenge, i.e., using the trainval35k set for training that includes 80k images from the train set and 35k images from the validation set. We first performed this training with

Table 3: Object detection percentage performance of the YOLOv4-tiny trained model on our distorted dataset.

Distortion Type	Distortion level (%) 2	4	6	8	10	mAP Average
Noise	9.58	22.3	37.9	80.3	114	47.2%
Contrast	-0.54	1.8	2.47	3.23	5.71	1.99%
Compression	-2.36	-0.48	-0.5	4.84	35.4	4.26%
Rain	12.7	40.4	71.4	101.8	114.3	61.7%
Haze	-0.49	3.14	10.5	29.8	51.7	15.8%
Motion-Blur	6.11	39.5	95.3	151.4	183.3	86.3%
Defoc-Blur	1.02	14.2	22.4	26.3	26.7	16.5%
Loc. MBlur	10.3	31.6	46.9	59.3	67.0	39.8%
Loc. Defoc.	3.08	16.0	23.3	25.5	25.9	17.2%
BackLight	1.45	5.85	19.1	40.3	76.8	24.1%
Average per level	4.11%	17.9%	32.9%	52.3%	70.1%	**31.5%**

the original MS-COCO trainval35k set to obtain reference values, then training with the distorted dataset, which allowed us to evaluate the contribution of the data augmentation on the robustness of the model. Each distortions type represents 5% of the database used for the training process, i.e., 5.9K images for each distortion. In our case, the hyper-parameters are taken from the YOLOv4 method but with some tuning. The training step is 500,050, with the initial learning rate multiplied by 0.1 at 400,040 step then 450,045 step. Evaluation of the training improvement is achieved by comparing the evaluation scores of both models trained on original and distorted train sets, respectively, on the distorted MS-COCO validation sets according to each distortion level. These results are formulated in the table 3 (Columns 2-6) by computing the ratio between AP COCO scores of the distorted model on the natural model. We observed that training has a significant impact on the robustness improvement with an average increase of 31.5% of performance (see last column in table 3). Furthermore, the study highlighted that the training improves more robustness for high-level distortions (see last row in table 3). According to the results of the table 4, training has a weak impact

Table 4: Object detection performance of the YOLOv4-tiny trained model on our distorted natural sub-sets.

Natural distortions	AP	rel. AP	mIoU	rel.mIoU
Noise	0.003	0.99%	0.005	0.65%
Contrast	0.011	5.47%	0.007	0.95%
Motion-Blur	-0.012	-5.17%	-0.004	-0.54%
Defocus	0.015	12.9%	0.0	0.0%
Rain	-0.009	-2.81%	0.006	0.79%
Loc. MBlur	-0.006	-3.13%	0.0	0.0%
BackLight	0.011	8.94%	0.005	0.68%

rel.: ratio between scores from distorted and original images, mIoU: mean Intersection over Union.

on robustness to natural distortions with an average improvement of 2.5% for the mAP score. However, these results need perspective due to the distortion levels in the distorted natural datasets [18], which are primarily low, and, therefore, not impacted enough by data augmentation from training according to the previously highlighted observation (see average per level of column 2 in table 3).

5. CONCLUSIONS AND FUTURE WORK

This study revealed the impact of the quality of the image frames in the development of object detection methods. Indeed, through this study we have shown that increasing the training database with complex scenarios containing different distortions improves the performance of object detection models. We have also shown that this is also true even for the best state-of- the-art methods. We, particularly, propose an original strategy of image distortion generation applied to the MS-COCO dataset that combines some local and global distortions to reach a better realism. We have shown that training with this distorted dataset improves the robustness of models by up to 31,5%. We, also, provide a custom dataset including the natural images distorted from MS-COCO dataset to perform a more relevant evaluation of the robustness concerning distortions. The database and the generation source codes of the different distortions is made publicly available on github plateform. Considering dataset with local and global distortions is an interesting way for future research.

6. REFERENCES

[1] Zhengxia Zou, Zhenwei Shi, Yuhong Guo, and Jieping Ye, "Object detection in 20 years: A survey," *arXiv preprint arXiv:1905.05055*, 2019.

[2] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[3] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen, "Deep learning for generic object detection: A survey," *International journal of computer vision*, vol. 128, no. 2, pp. 261–318, 2020.

[4] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, D. Erhan, Ian J. Goodfellow, and Rob Fergus, "Intriguing properties of neural networks," *CoRR*, vol. abs/1312.6199, 2014.

[5] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker, Matthias Bethge, and Wieland Brendel, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," *CoRR*, 2019.

[6] Dan Hendrycks and Thomas Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *Proceedings of the International Conference on Learning Representations*, 2019.

[7] Debang Li, Junge Zhang, and Kaiqi Huang, "Universal adversarial perturbations against object detection," *Pattern Recognition*, vol. 110, pp. 107584, 2021.

[8] Philipp Benz, Chaoning Zhang, Adil Karjauv, and In So Kweon, "Revisiting batch normalization for improving corruption robustness," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 494–503.

[9] Mehran Khodabandeh, Arash Vahdat, Mani Ranjbar, and William G Macready, "A robust learning approach to domain adaptive object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 480–490.

[10] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel, "Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness," *arXiv preprint arXiv:1811.12231*, 2019.

[11] Sebastian Cygert and Andrzej Czyżewski, "Robustness in compressed neural networks for object detection," in *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1–8.

[12] Omid Poursaeed, Tianxing Jiang, Harry Yang, Serge Belongie, and Ser-Nam Lim, "Robustness and generalization via generative adversarial training," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15711–15720.

[13] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[14] Mingxing Tan, Ruoming Pang, and Quoc V Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10781–10790.

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[16] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[17] Ayman Beghdadi, "Benchmarking performance of object detection," `https://github.com/Aymanbegh/Benchmarking-performance`, 2022.

[18] Ayman Beghdadi, "Natural distorted ms-coco dataset for object detection," `https://github.com/Aymanbegh/Distorted-Natural-COCO`, 2022.

[19] Sebastian Cygert and Andrzej Czyżewski, "Toward robust pedestrian detection with data augmentation," *IEEE Access*, vol. 8, pp. 136674–136683, 2020.

[20] Xiangning Chen, Cihang Xie, Mingxing Tan, Li Zhang, Cho-Jui Hsieh, and Boqing Gong, "Robust and accurate object detection via adversarial learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16622–16631.

[21] Aharon Azulay and Yair Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?," *J. Mach. Learn. Res.*, vol. 20, pp. 184:1–184:25, 2019.

[22] Robert Geirhos, Carlos RM Temme, Jonas Rauber, Heiko H Schütt, Matthias Bethge, and Felix A Wichmann, "Generalisation in humans and deep neural networks," *Advances in neural information processing systems*, vol. 31, 2018.

[23] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[24] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[25] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759–8768.