Efficient fine-grained object detection for robot-assisted WEEE disassembly

Ioannis Athanasiadis, Athanasios Psaltis, Apostolos Axenopoulos, and Petros Daras

Centre for Research and Technology Hellas

Abstract. In the current study, a region-based approach for object detection is presented that is suitable for handling very small objects and objects in low-resolution images. To address this challenge, an anchoring mechanism for the region proposal stage of the object detection algorithm is proposed, which boosts the performance in the detection of small objects with an insignificant computational overhead. Our method is applicable to the task of robot-assisted disassembly of Waste Electrical and Electronic devices (WEEE) in an industrial environment. Extensive experiments have been conducted in a newly formed device disassembly segmentation dataset with promising results.

Keywords: human-robot collaboration · WEEE disassembly · small object detection

1 Introduction

Human-robot collaboration has been recently introduced in industrial environments, where the fast and precise, but at the same time dangerous, traditional industrial robots have started being replaced with industrial collaborative robots for disassembling of WEEE in WEEE recycling plants [1] [1]. The rationale behind the selection of the latter is to combine the endurance and precision of the robot with the dexterity and problem-solving ability of humans. An advantage of industrial collaborative robots (or cobots) is that they can coexist with humans without the need to be kept behind fences. Cobots can be utilised in numerous industrial tasks for automated parts assembling. In the context of the EU-funded project HR-Recycler, industrial collaborative robots are introduced for disassembling of WEEE in WEEE recycling plants. Due to the complexity and high variability of devices, the fully-robotised disassembly is not feasible, thus, a human-robot collaboration scenario is much appreciated. The robot will be able to unscrew, cut or grasp the constituting parts of an electronic device and put them to separate baskets, depending on their type (e.g. capacitors, batteries, PCBs, etc). In this direction, Computer Vision is necessary to assist the robot's perception of the surrounding environment.The Deep Learning (DL) era brought great advancements in numerous of Computer vision domains, mainly because

[1] https://www.hr-recycler.eu/

DL-based methods are capable of grasping complex relations and handling huge amount of data. Specifically, deep Convolutional Neural Networks (CNNs) have been utilised for the task of object detection. These approaches fall into two categories, namely the two-stage and the one-stage methods. Modern two-stage object detection methods such as Faster R-CNN [9] and Mask R-CNN [3] make use of a trainable network, called Regional Proposal Network (RPN), to propose regions which potentially enclose ground truth objects in. On the other hand, in the one-stage methods, the regions are generated and classified in a single forward pass. The YOLO [8] and the SSD [7] algorithms are the most representative one-stage object detection approaches. In [11] the High Possible Regions Proposal Network is introduced in which a feature map with empowered edge information is formed and passed as an additional feature map in the RPN for more accurate region proposing. Object detection has evolved significantly the latest years, nevertheless, the accuracy in detecting small objects is still limited. To address such limitations, in [10] the use of context information is adapted focusing on detecting small objects. A Generative adversarial networks (GAN) based approach is presented in [4], which generates super resolved representations of small objects, similar to the bigger ones, where the object detection algorithms perform better at detecting. The motivation behind this work is to effectively boost the performance of current two-stage object detection methods in detecting small objects by implementing a more sophisticated anchoring technique. A suitable baseline method is chosen which we gradually enhance by combining cascade architecture and a proposed anchoring mechanism targeted at small-sized objects. Finally in this work, we showcase the ability to amplify DL-based object detection approaches by complementing them with additional handcrafted features, targeted explicitly at compensating for the poor visual representation of the small objects.

2 Methods

State of the art two-stage approaches consist of two discrete modules responsible for region proposing and classifying respectively. At the first stage, a set of *candidate regions* of predefined shape and size, called anchors, is uniformly generated across the image. Thereafter, each anchor is validated on its probability of containing a ground truth object by the RPN. The most confident, in terms of objectness, anchors constitute the *proposed regions* which are then passed to the second stage for further classification.

Baseline: As a baseline approach, Mask R-CNN [3] was chosen for both its state of the art performance and its efficiency in cases where heavy overlapping between the relevant objects occurs. By utilising the Feature Pyramid Network (FPN) approach of [5], Mask R-CNN becomes more appealing to the detection of small objects.

Cascade: The method as described in [2] is applied on the baseline architecture, with the purpose of simultaneously training multiple classifiers optimized at different Interception over Union (IoU) thresholds. Given that each classifier refines

the input regions to align better with their corresponding targets, a sequence of cooperative classifiers is deployed to progressively increase their quality as well as the quality of the regions in a single pass.

EA-CNN$_{\mathbf{P}}$: Although the uniform anchoring has proven to be quite effective in most cases, generating anchors in discrete pixel intervals with fixed shape and size often results in misalignment between the anchors and their respective ground truth targets, which can be critical in the case of small objects detection. This can be overcome by generating additional candidate regions through applying the maximally stable extremal regions (MSER) algorithm on an image edge-enhanced by simplified Gabor wavelets (SGWs). From this step, only the small-sized edge information regions, referred to as *edge anchors*, are considered. Finally, all the edge anchors are passed to the second stage classifier along with the regions proposed by the RPN.

EA-CNN$_{\mathbf{C}}$: The integration of the edge anchors into the RPN should retain the scale-specific feature maps of the FPN and the mapping between bounding box regressors and identical-shaped anchors. To this end, the edge anchors are refined to match the closest available shape, size and location configurations, as dictated by the hyper-parameters of ratio, scale, and input image dimension respectively. In order to minimize the refinement, additional enlarged feature maps dedicated to the edge anchors, called *edge maps*, are introduced, which correspond to different scales relevant to small objects. After the modifications described above the RPN is able to evaluate regions given both edge and regular anchors as input. Based on that, MSER is applied to both grayscale input image and its edge-enhanced version resulting in more but less precise edge anchors.

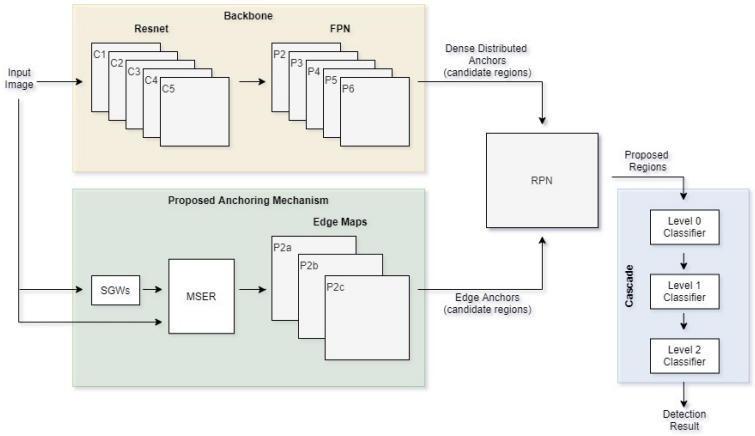


Fig. 1: The architecture combining Cascade with EA-CNN$_{\mathrm{C}}$.

3 Experiments

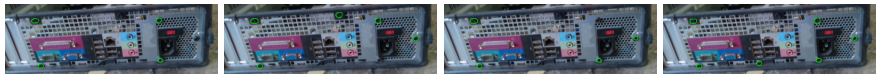
Within the context of HR-Recycler project, a new image dataset was created by capturing multiple PC-Tower devices during their disassembly procedure. The dataset was appropriately annotated by manually segmenting its components at various disassembly stages, which was then used for experimental testing.
Implementation Details: The input images are rescaled such as their biggest side is 512 pixels wide, while their aspect ratio is retained; although feeding images of a higher resolution would most likely result in better detection performance, the application's need for real-time operation is limiting. So as to effectively locate small ground truth targets, we use scales of 10, 12, 15 and 20 pixels with the first one corresponding to the uniformly generated anchors and the rest of them to the edge maps. When the cascade method is used, the base classifier is followed by two extra classification heads of 0.55 and 0.60 IoU thresholds. Finally, for the purpose of generating the edge anchors, the regions produced by the MSER algorithm occupying an area larger than 15^2 pixels or have their aspect ratio falling outside the $[0.5, 2]$ interval, are filtered out.
Results: To evaluate the impact of each method on the detection performance, we report the standard COCO [6] metrics in Table 1. Notably, utilising the cascade approach seems to boost the metrics of the strict IoU threshold. Additionally, improved performance is achieved, in terms of AP_S, when edge anchors are used combined with high-quality cascade classifiers. Moreover, the results shown in Fig. 2, referring to screws, are indicative of the edge anchors' capability to detect significantly small objects.

Table 1: Object detection results on PC-Tower dataset in all classes (all its *components*).

Method	**Cascade**	$\boldsymbol{AP_{0.5:0.95}}$	$\boldsymbol{AP_{0.5}}$	$\boldsymbol{AP_{0.75}}$	$\boldsymbol{AP_S}$	$\boldsymbol{AR_{0.5:0.95}}$	$\boldsymbol{AR_S}$
Baseline	-	40.3	69.8	39.5	23.1	47.3	28.0
	✓	39.5	66.2	40.1	23.6	**48.8**	29.0
EA-CNN$_{\mathrm{P}}$	-	38.7	68.2	38.0	22.9	46.9	28.8
	✓	**40.9**	**70.5**	**42.3**	24.6	48.2	30.1
EA-CNN$_{\mathrm{C}}$	-	39.6	69.5	40.1	**25.5**	47.2	31.4
	✓	40.6	69.5	41.9	25.2	48.4	**32.3**

4 Conclusions

In this work an anchoring mechanism utilising heuristic information is proposed; its ability to generate candidate regions that align better with the small-sized ground-truth targets discloses the potential of improving the detection of small objects that current two-stage state-of-the-art algorithms struggle with. The experiments on the PC-Tower disassembly dataset consisted of challenging small-



$AP_{0.5:0.95}$: 29.5 (a) $AP_{0.5:0.95}$: 25.4 (b) $AP_{0.5:0.95}$: 29.1 (c) $AP_{0.5:0.95}$: 37.8 (d)

Fig. 2: Detection performance and AP scores for the class: *screws*. **(a)** Baseline **(b)** Cascade **(c)** EA-CNN$_{\mathrm{P}}$ w/ Cascade **(d)** EA-CNN$_{\mathrm{C}}$ w/ Cascade

sized components (e.g. screws), exhibit promising results. As future work, integrating the edge anchors into an attention-like mechanism, is considered, while experiments will be further extended using different types of WEEE devices.

Acknowledgments. This work was supported by the European Commission under contract H2020-820742 HR-Recycler.

References

1. Axenopoulos, A., Papadopoulos, G., Giakoumis, D., Kostavelis, I., Papadimitriou, A.P., Sillaurren, S., Bastida, L., Oguz, O., Wollherr, D., Garnica, E., et al.: A hybrid human-robot collaborative environment for recycling electrical and electronic equipment. In: 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). pp. 1754–1759. IEEE (2019)
2. Cai, Z., Vasconcelos, N.: Cascade r-cnn: High quality object detection and instance segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019)
3. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask r-cnn. In: The IEEE International Conference on Computer Vision (ICCV) (Oct 2017)
4. Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., Yan, S.: Perceptual generative adversarial networks for small object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1222–1230 (2017)
5. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
6. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
7. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European conference on computer vision. pp. 21–37. Springer (2016)
8. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)

10. Ren, Y., Zhu, C., Xiao, S.: Small object detection in optical remote sensing images via modified faster r-cnn. Applied Sciences **8**(5), 813 (2018)
11. Shao, F., Wang, X., Meng, F., Zhu, J., Wang, D., Dai, J.: Improved faster r-cnn traffic sign detection based on a second region of interest and highly possible regions proposal network. Sensors **19**(10), 2288 (2019)