ISSN 1533-9211

DOI: 10.5281/zenodo.8375079

EVALUATING SAMPLING-BASED TECHNIQUE FOR IMPROVED SMALL OBJECT DETECTION USING APPROPRIATE SIZES OF THE ANCHOR BOXES

PRASHANT MISHRA

Research Scholar, Department of Computer Science, Om Sterling Global University, Hisar, Haryana.

PARVEEN SEHGAL

Professor, Research Supervisor, Department of Computer Science, Om Sterling Global University, Hisar, Haryana.

Abstract

Small object detection is a challenging task in computer vision due to the limited spatial information and low resolution of these objects. Existing generic object detectors demonstrate satisfactory performance on medium and large-sized objects, but they often struggle to accurately recognize small objects. The challenges arise from the low resolution and simple shape characteristics typically associated with small objects. In this paper, we propose an up sampling-based method for end-to-end tiny object identification that outperforms the existing state-of-the-art methods. We, like other contemporary approaches, first produce suggestions before labeling them. In the event of somewhat little things, we recommend tweaks to both of these procedures.

Keywords: Anchor Box, Small object, Resolution, Performance, Detection

1) INTRODUCTION

Object detection is a crucial part of computer vision since it involves searching through digital pictures for occurrences of a certain class of visual objects (such people, animals, or automobiles). Object detection seeks to provide computational models and methods that supply one of the most fundamental bits of information required by computer vision applications: Which things are located where? Accuracy (in terms of both classification and localisation) and speed are the two most important criteria for object detection.

Many additional computer vision tasks rely on object detection as a foundation, including instance segmentation, picture captioning, and object tracking, and so on. The fast growth of deep learning techniques in recent years has tremendously aided the advancement of object identification, resulting in astonishing discoveries and catapulting it to the forefront of study with unprecedented focus. These days, object detection is employed in a broad variety of practical contexts, from autonomous vehicles to robot vision to video surveillance and beyond.

Small object detection plays a vital role in various computer vision applications, including object recognition, video surveillance, autonomous driving, robotics, and medical imaging. Unlike larger objects, small objects typically exhibit limited spatial extent and low resolution, making them challenging to detect accurately. The detection of small objects is of paramount importance as it enables a wide range of practical applications, enhancing the understanding and analysis of visual data.

ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

In object recognition tasks, small objects often represent critical components of a scene. For instance, in image classification, correctly identifying small objects can significantly impact the overall accuracy and reliability of the classification results. Consider a scenario where a surveillance camera needs to identify a person carrying a prohibited item, such as a knife or a firearm. The accurate detection of these small objects can help prevent potential security threats and ensure public safety.

Moreover, small object detection is crucial in video surveillance systems. Surveillance cameras are deployed in various environments, such as airports, shopping malls, and public spaces, to monitor activities and identify potential security breaches. However, due to their size, small objects can easily go unnoticed, making it difficult to track and analyze them effectively. By improving the detection of small objects, surveillance systems can enhance their ability to detect suspicious behavior, recognize specific objects, and provide more accurate and timely alerts.

Autonomous driving is another field where small object detection is of utmost importance. Self-driving vehicles rely on computer vision systems to perceive their surroundings and make informed decisions. Detecting small objects like pedestrians, cyclists, or road signs is crucial for ensuring the safety of both the autonomous vehicle and other road users. Accurate detection of small objects allows autonomous vehicles to predict and respond to potential hazards in real-time, contributing to safer and more efficient transportation systems.

In robotics, small object detection is vital for object manipulation, grasping, and interaction. Robots equipped with vision capabilities need to detect and localize small objects in their environment to perform tasks such as picking and placing objects, sorting items, or assembling components. Precise detection of small objects enables robots to operate with higher accuracy and efficiency, leading to improved productivity in industrial automation, logistics, and manufacturing sectors.

Medical imaging is another domain that greatly benefits from small object detection. In fields like radiology and pathology, identifying and analyzing small structures within medical images can aid in the diagnosis of diseases, assessment of treatment effectiveness, and monitoring of patient health. Detecting small lesions, tumors, or abnormalities can have a significant impact on the accuracy and early detection of various medical conditions, ultimately saving lives and improving patient outcomes.

The significance of small object detection extends beyond the aforementioned applications. It has implications in fields like agriculture, environmental monitoring, quality control, and many others. For instance, in agriculture, detecting pests or diseases on plants at an early stage can help prevent crop losses and optimize resource allocation. In environmental monitoring, identifying small objects like endangered species or pollution sources contributes to biodiversity conservation and ecosystem management.

Despite the widespread importance of small object detection, it remains a challenging task. The limited spatial information, low contrast, occlusion, and scale variations associated with small objects pose significant difficulties for traditional detection algorithms. Consequently,

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

researchers and practitioners are continuously exploring novel methodologies and techniques to improve the accuracy, robustness, and efficiency of small object detection systems.

2) CHALLENGES ASSOCIATED WITH SMALL OBJECT DETECTION

Small object detection presents a set of unique challenges that make it a complex task in computer vision. These challenges stem from the characteristics of small objects and the limitations of traditional detection algorithms. Understanding these challenges is crucial for developing effective solutions to improve small object detection performance. Here are some key challenges associated with small object detection:

Limited Spatial Extent

Small objects occupy a small region within an image, often comprising only a few pixels. The limited spatial extent makes it difficult to extract discriminative features and distinguish small objects from their surrounding background. The lack of visual cues and context poses a significant challenge in accurately localizing and detecting small objects.

Low Resolution

Small objects are frequently captured at lower resolutions compared to larger objects due to factors such as long-range imaging or object distance. The low resolution results in reduced level of detail and makes it challenging to discern fine-grained object characteristics, leading to potential misclassification or false detections. The lack of sufficient information exacerbates the difficulty of detecting small objects accurately.

Scale Variations

Small objects can exhibit significant scale variations within a scene or across different images. They can appear at different sizes due to varying distances from the camera or zoom levels. Handling scale variations is crucial for effectively detecting small objects, as the detection algorithm must be capable of detecting objects at multiple scales and adapt to size discrepancies.

Occlusion

Small objects are more susceptible to occlusion by other objects or scene elements, which can partially or completely obscure them. Occlusion hampers the visibility of small objects, making it challenging to detect and accurately localize them. Overcoming occlusion requires robust algorithms that can reason about occluded parts or leverage contextual information to infer the presence of small objects.

Low Contrast

Small objects often have low contrast compared to their surrounding background. The limited intensity variation can make it challenging for traditional detection algorithms to distinguish small objects from their surroundings. Consequently, distinguishing small objects from clutter or background noise becomes a significant challenge, requiring sophisticated techniques to enhance contrast and improve detection accuracy.

The Seybold REPORT
ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

Computational Efficiency

Achieving real-time performance for small object detection is a demanding requirement, particularly in applications such as robotics or video surveillance. The limited computational resources, combined with the need for accurate detection, pose a challenge in developing efficient algorithms that can handle the high computational demands while maintaining high detection accuracy.

Dataset Bias

Training accurate and robust small object detectors heavily relies on the availability of diverse and representative datasets. However, obtaining large-scale annotated datasets with a significant number of small objects can be challenging. Dataset bias, where the training data does not adequately represent the real-world distribution of small objects, can adversely affect the performance of small object detectors when deployed in different environments or domains.

Addressing these challenges requires the development of advanced methodologies and techniques specifically designed for small object detection. Recent research has explored novel approaches, including feature enhancement techniques, multi-scale detection frameworks, context modeling, and deep learning architectures, to mitigate these challenges and improve the performance of small object detection systems. Continued efforts in this field will contribute to overcoming these challenges and advancing the state-of-the-art in small object detection.

3) REVIEW OF LITERATURE

Wang, Xuan et al., (2023) as AI continues to advance rapidly, smart cities are increasingly interested in remote-sensing image technology. Small dense objects in vast remote sensing environments have been the focus of remote sensing object detection research in recent years. Due to considerations such as picture resolution, item size, quantity, and orientation, small object recognition is still a difficult area of study. In this research, we focus on tiny item recognition in remote sensing and explore how deep learning-based object detection might be used in this context. The purpose of this article is to let the reader fully grasp the purposes of the study. The main datasets and assessment methodologies widely used in modern remote sensing object recognition algorithms are compiled here. The tiny item recognition methods in remote sensing pictures are also reviewed, and the irregularity problem of object detection in remote sensing images is discussed. We also conduct tests and analyses using state-of-the-art techniques for detecting tiny targets. Finally, the difficulties and unfinished business of remote sensing tiny item identification are underlined.

Hua, Wei & Chen, Qili (2023) when evaluating aerial remote sensing photos, small item recognition is a significant problem for the area of computer vision. While deep learning and other detection approaches have seen remarkable advancement in recent years, its transfer to aerial imagery has fallen short of expectations. Small target size, perspective specificity, backdrop complexity, and scale and orientation variability are all major factors in limiting image collection circumstances. Despite the rising popularity of deep learning-based

The Seybold REPORT
ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

algorithms as a solution to these issues, very little research has compared and contrasted the effectiveness of various deep learning methodologies for detecting tiny targets in aerial photos. This research, then, intends to investigate the feasibility of using deep learning techniques for tiny item recognition in aerial photographs. The main difficulties of pinpointing little objects in aerial photographs will be summed up. The next step is a comprehensive examination and classification of the current deep learning optimization algorithms used to overcome the obstacles inherent in aerial picture identification. We next offer a complete overview of the item detection datasets and assessment metrics used to aerial remote sensing photos. We also provide experimental evidence for the currently suggested detection techniques. Finally, prospective future directions and the benefits and drawbacks of various optimization methodologies are examined. Researchers in this area should be able to use it as a resource.

Wang, Xiaobin et al., (2022) the recognition of small items, and notably the detection of small objects in UAV aerial photos, has historically been a challenging direction in the field of object detection. Small items and high density can be easily identified in UAV photos. This article addresses these issues by enhancing object detection capabilities from the perspectives of data and network architecture. When it comes to information, data augmentation and the picture pyramid mechanism are the most common methods of utilization. The data augmentation strategy uses the technique of picture division to enhance the number of little objects, which facilitates thorough training of the algorithm. Due of the greater density of the object, the picture pyramid process is employed. The pictures are up-sampled to three different sizes and transmitted to three separate detectors during training. Finally, we get our final detection findings by combining the data from all three detectors. The little thing has few individual pixels and characteristics. Detection efficiency can be boosted by employing context. In this study, we augment the yolov5 network architecture with an attention mechanism and a detection head to the underlying feature map to increase the network's focus on low-level details. The detection performance of tiny objects may be greatly enhanced by applying data augmentation and enhanced network structure. In this work, we conduct our experiment on the Visdrone2019 dataset and the DOTA dataset. Our suggested strategy has been shown to greatly enhance tiny item identification performance in experimental settings.

Tong, Kang et al., (2020) Detecting small objects is a difficult task in computer vision. It has found several uses in the fields of defense, transportation, manufacturing, and other areas. We examine the current deep learning approaches to small object detection from five perspectives—multi-scale feature learning, data augmentation, training strategy, context-based detection, and GAN-based detection—to help readers gain a deeper appreciation for this important topic. Then, using well-known datasets like MS-COCO and PASCAL-VOC, we conduct a comprehensive evaluation of the effectiveness of many common tiny object recognition techniques. Finally, five potential future research directions are highlighted: emerging small object detection datasets and benchmarks; multi-task joint learning and optimization; information transmission; weakly supervised small object detection methods; and a framework for small object detection.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

Nguyen, Nhat-Duy et al., (2020) a fascinating area of computer vision is the detection of small objects. The lightning-fast progress being made in the field of deep learning has piqued the interest of many academics working on novel techniques. These include suggestions for additional regions, split grid cells, multiscale feature maps, and a different kind of loss function. Consequently, there have been recent notable advances in object detecting speed. One-stage and two-stage state-of-the-art detectors alike have had trouble seeing very small objects. Current state-of-the-art models using deep learning from both perspectives, including Fast RCNN, Faster RCNN, RetinaNet, and YOLOv3, are compared and contrasted in this work. We provide a thorough analysis of the benefits and drawbacks of models. In order to determine which sorts of items work best with each model and backbone, we test the models on a variety of datasets including multiscale objects. Two standard datasets, one including small objects and the other using filtered data from PASCAL VOC 2007, were used for a thorough empirical examination. Finally, we offer our comparative analysis and outcomes.

Wang, Zijie et al., (2019) Research into artificial intelligence has accelerated the development of numerous new applications, like autonomous driving, in recent years due to the continued popularity of deep learning. Despite the extensive research into ways for detecting larger objects like cars and people in self-driving technology, no viable solution exists for detecting smaller objects like stones on the road. However, the stability of an autonomous driving system is severely disrupted by tiny targets on the road. That's why it's crucial to conduct roadside detection of tiny targets.

Lim, Jeong-Seon, et al., (2019) Attempting to use an object detection method in any given setting has a number of significant caveats. Low resolution and sparse data make it more difficult to recognize tiny objects. To better recognize tiny objects, we offer a context-based approach to object identification. The suggested technique concatenates multi-scale characteristics to provide extra features from various levels as context. We also suggest an attention technique for object recognition that may zero down on a specific region of an image and include contextual data from the target layer. The experimental findings demonstrate the superior accuracy of the suggested technique compared to that of traditional SSD for detecting tiny objects. Our Mean Average Precision (mAP) on the PASCAL VOC2007 test set was 78.1% for 300$times$300 input.

Hu, Guo et al., (2018) In order to extract deep semantic characteristics from a picture, the current object recognition technique based on a deep convolution neural network must perform multilayer convolution and pooling operations on the whole image. The detection models perform better when dealing with larger objects. Existing models are unable to recognize low-resolution objects that are heavily impacted by noise because the features after multiple convolution processes do not adequately capture the main properties of the tiny objects. In this study, we are able to extract features from many convolution layers of the object and use these multiscale features to accurately recognize tiny objects. In our detection model, we use the third, fourth, and fifth convolutions to extract picture features, which are then combined into a single dimensional vector. Classifiers employ the vector to sort objects into categories, and location data is determined using bounding box regression. Testing has shown that our model

ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

outperforms state-of-the-art methods by 11% in terms of detection accuracy for tiny objects. We also applied the model to the problem of identifying airplanes in remote sensing photos, with encouraging results.

Zhong, Yuanyi, et al., (2018) in this research, we present a generic method for improving object recognition through anchor box optimization. At the moment, anchor boxes are frequently used in cutting-edge detection systems. All of these systems, however, establish anchor box shapes in advance using heuristics and then lock in the size during training. We propose dynamically learning the shapes, which enables the anchors to automatically adapt to the data distribution and the network's learning power, to boost accuracy and save design time for the anchor boxes. You may simply construct the learning strategy using stochastic gradient descent and integrate it into any existing anchor box-based detection system. There is hardly any influence on the inference time cost from the additional training expense. Extensive studies further show that the suggested anchor optimization technique regularly outperforms the baseline method by a large margin ($ge 1%$ mAP absolute gain) on a number of benchmark datasets, such as Pascal VOC 07+12, MS COCO, and Brainwash. The complexity of the anchor box design issue is considerably reduced by the simultaneous verification of robustness against many initialization strategies for the anchor boxes.

4) PROPOSED DATASET

The Small Object Dataset provides the basis for our investigations. Microsoft's COCO and the SUN dataset have contributed to this collection of 4,925 photos. The objects in each of the ten selected categories have an average size of less than 30 centimeters. Only photographs with objects taking up a relatively modest percentage of the frame were selected from the set of images containing these classes.

We employ the standard performance measure in detection, which states that if the Intersection over Union (IoU) overlap between the predicted and ground-truth bounding boxes is more than 0.5, the detection is accurate. Mean Average Precision (mAP) is a metric for evaluating detection algorithms as a whole, since it represents the region under the precision-recall curve.

5) EXPERIMENTAL RESULTS

Proposal Generation

We compare the anchor box sizes you choose with the ones Faster RCNN suggests. With VGG16 as the backbone, we use the typical faster RCNN architecture. The anchor nodes are connected to VGG16's conv5 layer. Aspect ratios of 1:1, 1:2, and 2:1 are used, the same as in faster-RCNN. After 50000 iterations, the network has been fine-tuned with a 0.001 learning rate and a 0.1 gamma. We evaluate the mAP by comparing the first one thousand suggestions for each test picture based on the confidence that the proposal belongs to a non-background class. Faster RCNN is something we'd want to play around with more. Faster-RCNN is advantageous because it saves time and memory during testing and training by eliminating the need to temporarily store created suggestions. We also observed that, when training end-to-

ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

end, it is more effective to link the anchor boxes to conv5 rather than conv4 or conv3.

Our optimized set of anchors outperforms the standard, quicker RCNN. This is to be anticipated, given even the smallest anchor box, with a size of 128, is larger than each and every instance in the collection. We suggest anchor box sizes that account for the whole spectrum of tiny item sizes included in the dataset.

We further demonstrate the strength of these anchors by adding two more anchor boxes, of sizes 40 and 100. We see that even with extra anchor boxes, performance drops to 21.9%. This is because there will be a greater total number of suggestions, which will inevitably contain more proposals of generic objects that exist in the background. The comparatively basic forms of tiny objects make it possible that they may be mistaken for classes in our dataset.

Number of Proposals

Next, we evaluate how many ideas are considered for each test picture and how many options there are for anchor box sizes in relation to the created quality of the proposals. We are still operating using the quicker RCNN framework. The classifier score is used to rank the proposals that do not belong to the background class, and the top-k are selected to determine the mAP.

As can be seen in Figure 1, our preferred anchor boxes outperform the default Faster RCNN across all numbers of proposal settings. Since most genuine positives occur in the top few proposals alone, we conclude that the increase in the number of false positives is to blame for the decline in performance as more proposals are evaluated. Because common background objects have some structural similarities with the classes of interest, this leads to an increase in false positives.

Figure 1: Exemplar results of our method on the small object dataset

ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

Green squares represent the findings. Example errors are included in the last column. A chair's handle looks so much like a mouse that it's mistaken for one in the first picture. The clock was too faint to be picked up in the second picture. The third picture shows a phone that has two parts that are too far apart for our proposal generating approach to treat as a whole.

Super-resolution

In this case, we use the RCNN architecture, in which our trained RPN is used in conjunction with anchor box sizes selected by us to propose regions. We reorder the proposals' scores by using our trained classifier and an up sampling technique. Table 1 summarizes the findings and shows that a super-resolution network yields better outcomes. The filters taught by convolutional layers don't work as well on low-resolution pictures, which is why this enhancement was made possible. When low-resolution photos are enlarged, they become pixelated and have hazy edges. These datasets, however, rely heavily on medium- to high-resolution pictures during training, and as a result, their filters perform best at these resolutions. Super-resolution smooths out this resolution disparity.

Table 1: Results of Our End-To-End Method on the 10-Class Small Object Dataset

Method	Mouse	Phone	Switch	Outlet	Clock	T. paper	T. box	Faucet	Plate	Jar	Average
Faster RCNN	57.5	14.1	15.6	22.3	26.2	31.4	8.4	35.3	11.4	3.2	22.7
RPN and up scaling	57.0	16.6	31.0	29.7	31.7	29.2	23.0	31.8	9.5	4.5	24.9
RPN and super-resolution	60.3	16.7	16.8	23.0	30.5	34.5	12.6	37.6	15.0	4.3	25.0

The accuracy weighted average is shown in the last column. The quicker rcnn network that we trained from the ground up using our anchors appears in the first column. The results of running the rcnn pipeline and upscaling are shown in the second row. The enhanced clarity brought by super-resolution is seen in the third row.

6) CONCLUSION

Here, we investigate several options for anchor sizes when proposing regions around tiny objects. We also demonstrated that deep super-resolution techniques provide superior results when used to the categorization of small objects. Accurate detection of small objects enhances the capabilities of computer vision systems, leading to improved safety, efficiency, and decision-making. Advancements in small object detection techniques have the potential to revolutionize a wide range of industries, opening up new opportunities for innovation and societal benefits.

References

1) Wang, Xuan & Wang, Aoran & Yi, Jinglei & Song, Yongchao & Chehri, Abdellah. (2023). Small Object Detection Based on Deep Learning for Remote Sensing: A Comprehensive Review. Remote Sensing. 15. 3265. 10.3390/rs15133265.

2) Hua, Wei & Chen, Qili. (2023). A Survey of Small Object Detection Based on Deep Learning in Aerial Images. 10.21203/rs.3.rs-3074407/v1.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8375079

3) Wang, Xiaobin & Zhu, Dekang & Yan, Ye. (2022). Towards Efficient Detection for Small Objects via Attention-Guided Detection Network and Data Augmentation. Sensors. 22. 7663. 10.3390/s22197663.

4) Mohamed, Elhassan & Sirlantzis, Konstantinos & Howells, Gareth & Hoque, Sanaul. (2022). Optimisation of Deep Learning Small-Object Detectors with Novel Explainable Verification. Sensors. 22. 5596. 10.3390/s22155596.

5) Duan, Shuyong & Lu, Ningning & Lyu, Zhongwei & Liu, G.R. & Cao, Bin. (2022). An anchor box setting technique based on differences between categories for object detection. International Journal of Intelligent Robotics and Applications. 6. 10.1007/s41315-021-00176-1.

6) Tong, Kang & Wu, Yiquan & Zhou, Fei. (2020). Recent advances in small object detection based on deep learning: A review. Image and Vision Computing. 97. 103910. 10.1016/j.imavis.2020.103910.

7) Nguyen, Nhat-Duy & Do, Tien & Duc, Thanh & Le, Duy-Dinh. (2020). An Evaluation of Deep Learning Methods for Small Object Detection. Journal of Electrical and Computer Engineering. 2020. 1-18. 10.1155/2020/3189691.

8) Zhong, Yuanyi & Wang, Jianfeng & Peng, Jian & Zhang, Lei. (2020). Anchor Box Optimization for Object Detection. 1275-1283. 10.1109/WACV45572.2020.9093498.

9) Wang, Zijie & Fang, Jianwu & Dou, Jian & Xue, Jianru. (2019). Small Object Detection on Road by Embedding Focal-Area Loss. 10.1007/978-3-030-34120-6_53.

10) Lim, Jeong-Seon & Astrid, Marcella & Yoon, Hyun-Jin & Lee, Seung-Ik. (2019). Small Object Detection using Context and Attention.

11) Kisantal, Mate & Wojna, Zbigniew & Murawski, Jakub & Naruniec, Jacek & Cho, Kyunghyun. (2019). Augmentation for small object detection. 119-133. 10.5121/csit.2019.91713.

12) Hu, Guo & Yang, Zhong & Hu, Lei & Huang, Li & Han, Jia. (2018). Small Object Detection with Multiscale Features. International Journal of Digital Multimedia Broadcasting. 2018. 1-10. 10.1155/2018/4546896.

13) Zhong, Yuanyi & Wang, Jianfeng & Peng, Jian & Zhang, Lei. (2018). Anchor Box Optimization for Object Detection.

14) A. R. Pathak, M. Pandey, and S. Rautaray, "Deep learning approaches for detecting objects from images: A review," in Advances in Intelligent Systems and Computing, 2018, vol. 710, pp. 491–499, doi: 10.1007/978-981-10-7871-2_47.

15) Pham, Phuoc & Nguyen, Duy & Do, Tien & Duc, Thanh & Le, Duy-Dinh. (2017). Evaluation of Deep Models for Real-Time Small Object Detection. 516-526. 10.1007/978-3-319-70090-8_53.