Pomiary Automatyka Robotyka, ISSN 1427-9126, R. 27, Nr 3/2023, 85–94, DOI: 10.14313/PAR_249/85

Deep Learning for Small and Tiny Object Detection: A Survey

Aleksandra Kos, Dominik Belter
Poznan University of Technology, Institute of Robotics and Machine Intelligence, 60-965 Poznań, Poland
Karol Majek
CuFiX Karol Majek, ul. gen. F. Kleeberga 1A, 05-825 Grodzisk Mazowiecki

Abstract: In recent years, thanks to the development of Deep Learning methods, there has been significant progress in object detection and other computer vision tasks. While generic object detection is becoming less of an issue for modern algorithms, with the Average Precision for medium and large objects in the COCO dataset approaching 70 and 80 percent, respectively, small object detection still remains an unsolved problem. Limited appearance information, blurring, and low signal-to-noise ratio cause state-of-the-art general detectors to fail when applied to small objects. Traditional feature extractors rely on downsampling, which can cause the smallest objects to disappear, and standard anchor assignment methods have proven to be less effective when used to detect low-pixel instances. In this work, we perform an exhaustive review of the literature related to small and tiny object detection. We aggregate the definitions of small and tiny objects, distinguish between small absolute and small relative sizes, and highlight their challenges. We comprehensively discuss datasets, metrics, and methods dedicated to small and tiny objects, and finally, we make a quantitative comparison on three publicly available datasets.

Keywords: Deep Learning, Small Object Detection, Tiny Object Detection, Tiny Object Detection Datasets, Tiny Object Detection Methods

1. Introduction

Object detectors can be divided into one-stage [1] and two-stage [2] detectors. Most methods are anchor-based [2, 1], but in recent years, some anchor-free detectors have also been proposed [3]. Two-stage methods tend to have a rather long inference time because of the additional step of region proposal generation [2]. One-stage methods are usually faster, however, they are also less precise [4]. To better handle multi-scale objects, a method presented in [5] proposes the Feature Pyramid Network that combines feature maps from different depths of the pyramid. The main challenges of general object detection include invariance to object scale, inter-class and intra-class appearance differences, and noisy backgrounds. The difference between the precision for small and large instances indicates that small objects are still a big challenge even for the state-of-the-art detectors [1, 6]. In recent years, the collected datasets contain objects of an even smaller scale and refer to them as tiny [7–9]. Numerous datasets, particularly the remote sensing from drones and satellites [10, 11], contain extremely high-resolution images, which poses additional challenges. In this article, we summarize the progress of the research in the field of tiny object detection. The main contribution of this paper is as follows: 1) a detailed discussion of the challenges specific to small and tiny object detection, as well as a comparison of different object size definitions, 2) a comprehensive analysis of publicly available small and tiny object detection benchmarks, 3) a thorough discussion of metrics and methods dedicated to small objects, and extensive quantitative analysis.

Autor korespondujący:
Aleksandra Kos, aleksandra.kos@doctorate.put.poznan.pl

Artykuł recenzowany
nadesłany 23.06.2023 r., przyjęty do druku 30.08.2023 r.

2. Tiny Object Detection

Visual recognition of tiny objects shares many challenges with the problem of generic object detection, however, there are also many small-scale specific issues: 1) objects with a small number of pixels have limited appearance information, making both classification and localization difficult, 2) a low signal-to-noise ratio that appears especially in high-resolution aerial images with complex backgrounds and sparsely distributed objects, 3) most convolutional backbones rely on downsampling, which can cause the small features to disappear or make them highly contaminated by the background. Standard anchor assignment strategies and the Intersection over Union (IoU) have also been shown to be ineffective [8, 12, 13].

Table 1. Thresholds for size classes [px] in selected datasets
Tabela 1. Wartości progów dla różnych kategorii rozmiarów obiektów w wybranych zbiorach danych

Dataset	tiny			small	medium	large
MS COCO [15]	—			0–32	32–96	96–inf
WIDER FACE [16]	—			10–50	50–300	300–inf
TinyPerson [9]	tiny1	tiny2	tiny3	20–32	—	—
	2–8	8–12	12–20			
AI-TOD [7]	very tiny		tiny	16–32	32–64	—
	2–8		8–16			
SODA [11]	eT	rT	gT	32–45	—	—
	0–16	16–24	24–32			

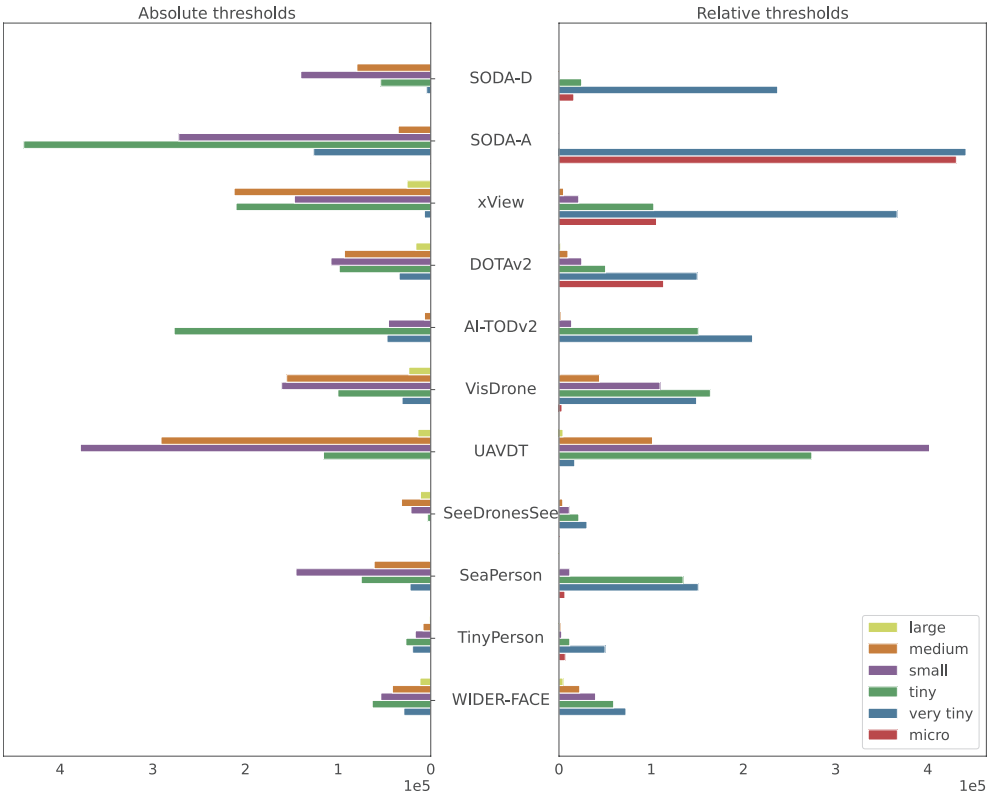

Fig. 1. The number of instances in the 6 predefined size categories in the selected object detection datasets
Rys. 1. Liczba obiektów przypadająca na każdą z 6 kategorii rozmiarów obiektów w wybranych zbiorach danych

A size of an object is usually classified by its absolute area [14, 11], by the geometric mean of its height and width on the image [9, 7], or by a single dimension [15]. The tiny category was created [9, 7, 11] to distinguish objects whose size is in the lower range of small objects in the COCO dataset, but there is no unambiguous definition as each dataset uses slightly different threshold values (Tab. 1). If possible, in Tab. 1, we convert the thresholds to the size of the object defined as the geometric mean of its height and width. Regarding WIDER FACE, the provided values concern the object's height. We identify two types of datasets regarding the goal of the research – small-only and multi-scale object detection. The first group contains the datasets that define the upper limit on the object size, so the scale variation is greatly reduced. Datasets containing a whole range of scales, from very small to large, but dominated by tiny instances, belong to the second group. The absolute thresholds shown in Tab. I do not take image size into account, so an object of a certain size will be classified in the same group regardless of image size. The very high-resolution data found in some datasets [10, 16, 11] cause additional challenges. Due to memory constraints, it cannot be processed at full resolution, downsampling causes a significant information loss, and patch-by-patch analysis greatly increases the runtime. Therefore, in this paper, we also consider the relative definition of object size.

3. Tiny Object Detection Datasets

We have collected 10 object detection datasets that either directly focus on small or tiny object detection, or contain many small instances. Most of them are aerial datasets with images collected by UAVs [9, 17–20], satellites [16, 11], or both [8, 21]. Their main features are presented in Tab. 2. Average sizes (defined as the geometric mean of height and width) were calculated for the whole image (S_{image}) and for the bounding box of the object (S_{object}). We report the number of images and objects for subsets for which labels are publicly available and we use the following annotation type abbreviations: Horizontal Bounding Box (HBB), Oriented Bounding Box (OBB), Multiple Object Tracking (MOT), Single Object Tracking (SOT), Crowd Counting (CC), Coarse Point (CP). Among the presented datasets, WIDERFACE seems to be the least challenging due to the low image resolution, relatively high object size, and the fact that instances smaller than

Table 2. Comparison of the main characteristics of object detection datasets
Tabela 2. Porównanie głównych cech wybranych zbiorów danych do wykrywania obiektów

dataset	data type	object type	labels	images	objects	classes	S_{image}	S_{object}
WIDER FACE [16]	natural scenes	faces	HBB	16,106	199,132	1	877±132	33±53
TinyPerson [9]	aerial (UAV-based)	people	HBB	1,610	72,651	2	1540±527	19±23
SeaPerson [18]	aerial (UAV-based)	people	HBB, CP	6,279	304,462	1	1456±207	23±13
SeeDronesSee [19]	aerial (UAV-based)	objects at sea	HBB, MOT, SOT	10,477	67,390	5	2644±831	63±71
UAVDT [20]	aerial (UAV-based)	vehicles	HBB, MOT, SOT	40,409	798,795	3	743±3	32±21
VisDrone [21]	aerial (UAV-based)	vehicles, pedestrians	HBB, MOT, SOT, CC	8,599	471,266	10	1200±250	35±33
AI-TODv2 [8]	aerial (mixed)	multi-type	HBB	14,008	376,625	8	800±0	13±6
DOTAv2 [22]	aerial (mixed)	multi-type	HBB, OBB	2,422	349,675	18	3756±3536	33±49
xView [17]	aerial (satellite-based)	multi-type	HBB	847	601,806	60	3148±317	35±40
SODA-A [11]	aerial (satellite-based)	multi-type	OBB	2,513	872,069	9	3627±162	16±8
SODA-D [11]	autonomous driving	pedestrians, vehicles, etc.	HBB	24,828	278,433	9	2790±597	25±10

10 pixels are ignored. SODA [11], AI-TOD [8], and TinyPerson [9] introduce upper object size limits, which are 45, 64, and 32 pixels respectively. More detailed information on the distribution of the object scales can be found in Fig. 1. To distinguish between small absolute size datasets and small relative size datasets, we plotted the number of instances in each size category for two sets of thresholds. We used thresholds from [9] for *very tiny*, *tiny* and *small* combined with thresholds from [14] for *medium* and *large*. To obtain relative thresholds, each absolute value was divided by the average size of the image in the COCO dataset. The absolute thresholds have the following values: *micro* (0–2 px), *very tiny* (2–8 px), *tiny* (8–16 px), *small* (16–32 px), *medium* (32–96 px), *large* (96–∞ px), and the relative thresholds are as follows: *micro* (0–0.38 %), *very tiny* (0.38–1.52%), *tiny* (1.52–3.05 %), *small* (3.05–6.10 %), *medium* (6.10–18.30 %), *large* (18.30–100 %). Satellite data include multiscale objects when the absolute definition is applied. However, due to extremely high resolution most of them become *micro*, *very tiny*, or *tiny* when the relative definition is applied. A similar relationship holds for UAV-based datasets, but, due to the lower resolution, there are rather no micro objects in these datasets. There are some exceptions – AI-TOD, SODA fully dedicated to the tiny object detection, so there are no large instances, and UAVDT in which, regardless of the definition, most objects are *medium*, *small* or *tiny*.

4. Tiny Object Detection Metrics

The same metrics are usually used to evaluate the quality of tiny and small object detection and multi-scale object detection [14, 22]. Older datasets [15, 19, 10] use metrics introduced by the Pascal VOC [22], while newer ones [18, 20, 7, 16, 11] more often apply MS-COCO [14]. Average Precision (AP) is used by both Pascal VOC and MS-COCO. The main differences are the interpolation method, the IoU threshold, and Average Recall (AR) introduced in COCO. For more details, we refer to [22, 14]. AP is also often reported for each class [15, 7, 8, 21, 16, 11] or object size [14, 15, 7–9, 11] separately. AR is sometimes adjusted to better reflect the conditions of a particular dataset, e.g. VisDrone and AI-TOD report AR_{500} and AR_{1500} respectively due to the large average number of objects per image. In Tiny Person, an IoU of 0.25 is used to emphasize that detection is more important than precise localization. In [8, 12, 13] the reduced effectiveness of the IoU in small object detection was shown. This is because the IoU is highly sensitive to position deviation when applied to small objects. The issue was solved by introducing the Normalized Gaussian Wasserstein Distance (NWD), a Dot Distance (DotD), and the Receptive Field based Label Assignment (RFLA), respectively. All three methods are discussed in more detail in the following sections.

5. Tiny Object Detection Methods

In this section, we discuss the following groups of small and tiny object detection methods: Focus-and-Detect, Data-Augmentation, Sampling-Based, Attention-Based, Scale-Aware, Context-Aware, and Feature-Imitation Methods. The groups we use are similar to those defined in [11].

5.1. Focus-and-Detect Methods

Focus-and-detect methods are used to guide the detector to focus on the specific areas of high-resolution images. To increase the relative size of the object, high-resolution images can be processed using a sliding window [23–25]. In [23], the input image and the tiles obtained by dividing the input image are fed to the detector to maintain the detection quality for larger objects. In the end, the predictions from all windows are combined. The method presented in [25] integrates a super-resolution GAN into a simple slidingwindow pipeline. The use of a sliding window, however, is computationally sub-optimal. In aerial datasets, objects are often clustered in selected image areas, and a large portion of an image does not contain any objects. Therefore, [24] supports the sliding-window detection pipeline with additional neural networks that remove the background-only tiles. Many methods find the regions that are analyzed in detail in additional steps. The main differen-

ces between these approaches are how the tiles are generated, and whether the coarse and fine predictions are fused. [26, 27] divide the original images into several uniformly sampled tiles, and then select tiles for fine detection. Regular grid sampling, however, can lead to errors, since objects can lie on multiple tiles at the same time. In [28] a fine detection based on the prior coarse cluster proposals is introduced. Each selected tile can be divided or padded before resizing to maintain the scale of the objects. Zhang et al. [29] also use a coarse-to-fine approach, and the tiles are predicted by the Difficult Region Estimation Network (DREN), but no scale prediction is performed. Similar approaches are proposed in [30–32], where either binary semantic segmentation [30] or density maps [31, 32] support the tiling process. Unlike [28, 29, 31], in [32, 30] only tiles are passed to the detector. Koyun et al [33] propose the Incomplete Box Suppression (IBS) algorithm to eliminate the influence of truncation on the tile-based detection quality. In [34], same as in [31, 29, 28, 23], the final detections are obtained by fusing the predictions from a global image and selected tiles. However, in [34] the initial detections are directly used to generate tiles in an unsupervised manner. Then, the anchorfree detectors are utilized to eliminate the negative impact of high variability in object sizes. In [35] tiles are generated with a deep reinforcement learning strategy to handle variations in scales and sizes. During testing, final detections are obtained by merging the global image and tile outputs. Deng et al. [36] combine focus--and-detect with feature imitation, but unlike in [25], it scales only selected tiles, and the coarse detections are fused with predictions from tiles. The tiles are selected non-uniformly to deal with differences in scales and various aspect ratios.

5.2. Data Augmentation Methods

Data augmentation is commonly used in generic object detection to diversify the samples and reduce the risk of overfitting [1]. In this section, we have collected data augmentation methods developed specifically for detecting small objects. In [37], the authors duplicate images with small objects and perform a copy-paste operation on each of them to increase the number of small instances. This solution can increase the number of underrepresented samples, however, to avoid introducing extra noise, a semantic mask is required to precisely crop an object. Also, in many tasks, such as aerial imagery, objects tend to occupy specific areas in the image, and pasting them in random locations could degrade the results. To address these issues, [38] uses an additional semantic segmentation network to extract the road map from the UAV-taken image, and an object scale is also properly handled. In [39], input images are obtained by cropping a new image for each object. An even simpler cropping strategy is used in [40], where the original image is split into several parts using a regular grid. Both of these methods also use multi-model fusion. In [40] two detectors are trained, one for classes with many instances, and the other for classes with sparse representation in the dataset. In [39], the division is made based on the image resolution. Another data augmentation method has been proposed in [41], which uses Downsampling-GAN (DS-GAN) to generate small synthetic objects from larger ones. The generated objects are then added and blended with the background.

5.3. Sampling-Based Methods

Sampling-based methods address the issues of anchor sampling strategies in the detection of small objects. As pointed out in [42], the feature map for a small object contains too little information, and the anchors are too large. Small objects also get very few matches, which has been linked to a problem in the anchor matching strategy. Due to multiple negative anchors, there is also an increased risk of false positives. In [42], the anchor association is performed at different depths of the backbone network to ensure proper representation and scale of anchors for different object sizes. An anchor-matching strategy is also adjusted by lowering the IoU threshold. A new anchor design, together with the Expected Max Overlapping (EMO) score, was proposed in [43] to increase the average IoU between objects and anchors. The authors reduce the stride of the anchor by enlarging the feature map, use shifted anchors, and randomly shift objects during training. In [44], the feature up-sampling, multi-level features, and an inception module are all combined to improve anchor sampling. [12, 13] introduce new metrics designed to replace an IoU in the label assignment process. The IoU has been observed to be very sensitive to location variations when applied to small objects. Thus, [12] proposed a Dot Distance (DotD) defined as a normalized Euclidean distance between two rectangles. Both [14] and [8] model the Gaussian Distributions of the bounding boxes. Xu et al. [13] measure the distance between these distributions with Kullback-Leibler divergence, while [8] uses the Normalized Wasserstein distance.

5.4. Attention-Based Methods

When dealing with tiny object detection in high-resolution images, the signal-to-noise ratio is particularly low, so many attention-based methods [44–49] were developed to suppress the background noise and highlight relevant features. These methods are often combined with the multi-level feature fusion [44, 46–49] to enhance the feature representation of small objects. Yi et al. [45] use a Recurrent Neural Network (RNN) with attention to make the detector more focused on the relevant image areas, such as a road for cars, and a roadside for traffic signs. Attention-Guided Balanced Pyramid (ABP) introduced in [49] fuses features at different levels of the feature pyramid. The fusion is made adaptively with a two-part attention-based sub-net. The Level-Based Attention method (LA) is used to learn weights for each pyramid level, while the Spatial Attention Network (SA) highlights regions that do contain objects. [44, 46, 47] adapt the channel attention mechanism based on the Squeezeand-Excitation (SE) Block [50] to highlight channels relevant to detection and suppress noise. Other attention-based blocks are pixel attention in [44], and spatial attention in [46]. [48] models the spatial relationship between pixel pairs.

5.5. Scale-Aware Methods

In tiny object detection, because of downsampling, the last feature map that is usually used for detection has little or no representation of these objects. To combine the rich semantic information of deep features with the spatial information present in shallow layers, and to prevent the disappearance of small objects features, feature maps from different layers are used [51–54]. The scale-aware methods for detecting small objects are often based on the Feature Pyramid Network (FPN) [5], but introduce additional modifications to better handle the small-scale objects. In [51], global features are combined with multi-scale ROI features. Liu et al. [52] handle the misalignment between deep and shallow features by introducing the Image Pyramid Transformation Module (IPGT). [53] proves that the simple FPN can have a negative impact on tiny object detection, and thus introduce a statistically estimated fusion factor to control how deep and shallow features are combined. In [54], an attention module is combined with a feature-fusion approach. The Context Attention Module (CAM) generates multi-scale attention heatmaps, while the Scale Enhancement Module (SEM) makes the detector focus on specific object scales in different layers. Features in subsequent layers are fused. Similar approaches that combine feature-fusion and attention-based techniques are [44, 46–49, 55]. [56] uses a feature pyramid pooling to reduce false positives in high-resolution satellite images. In [57] a set of feature maps from different layers is processed and a new set of feature maps is fed to the detector head. DSFD

combines feature fusion with a sampling method to increase the number of positive anchors. Another approach is presented in [58], where instead of a feature pyramid, the authors proposed multiple detection modules, each for a feature map with a different stride. QueryDet [59], in a multi-stage process, fuses the lower-resolution features with the higher-resolution features to create a sparse feature map and filter-out background pixels. In [39] a multi-model fusion is applied – the authors use three detectors, each trained with images at different resolutions.

5.6. Context-Aware Methods

Since small objects contain little information, especially in the deep layers of the convolutional network, many efforts have been made to aid the detection using the contextual clues carried by their surroundings [60–66]. In [60], a simple two-branch pipeline, built on top of R-CNN [2], was proposed. The object region proposal and its surrounding area are both encoded into a feature vector, concatenated, and then passed to the classification head. The Inside-Ouside Net (ION) [61] extracts the context features using the RNN, and fuses them with the multi-scale ROI features for each region, which makes it both scale and context-aware. The method presented in [62] developed for tiny face detection tasks, also combines the scale-aware and context-aware approaches. Multiple scales are handled by a coarse image pyramid and separate detectors for each scale that share the backbone weights. The context features are incorporated by enlarging receptive fields and introducing more background information. A receptive field can be increased by using skip connections as done in [67]. The authors also enhance the spatial information lost by using skip connections and thus help in the localization task. Tang et al. [63] use a semi-supervised learning strategy to additionally learn the important context classes, e.g. human body for face detection. The authors of [65] prove that ROI Pooling used in twostep methods has a negative impact on contextual information, therefore they propose a Context-Aware ROI Pooling method based on deconvolutions. The method is also scale-aware as it uses separate detection heads for each object size. In [66], a Context-Aware Detection Network (CAD-Net) was introduced. It combines global context features with a pyramid of local context features and the attention module to detect small objects in satellite images. In [64], a spatial context information is used for the re-detection of low-confidence objects. This approach is based on the observation that in UAV images, distinctive classes are often clustered in separate areas of the image, so the class probabilities for low-confidence detections are updated based on the weighted distance from high-confidence detections.

5.7. Feature-Imitation Methods

To enhance the poor representation of small objects, some methods [68–73] leverage recent advances in Generative Adversarial Networks (GANs). The authors of the Perceptual Generative Adversarial Network (Perceptual GAN) [68] point out the fact that simple feature enhancement made by adding the features extracted from the shallow layers is not always beneficial to the detection task. Instead of using a multi-level pyramid, feature, or image upsampling, a super-resolution approach has been proposed to make the features of small objects similar to those of larger objects. A similar feature-level GAN approach is presented in [71], however, it also introduces direct supervision for the training process. [69] and [70] work directly on image regions extracted by baseline detectors. In [69] small, blurred faces are super-resolved by the generator. The classification loss is back-propagated to the generator. [70] extends this method to multi-class detection. The discriminator additionally outputs the class probabilities and offsets of the bounding boxes. Unlike [69], both the classification loss and the regression loss are used to train the generator. In [72, 73] the image-level super-resolution methods are explored for detecting objects in satellite images. In addition to super-resolution methods at the image, feature, or region level, some methods perform feature imitation in other ways. In [74], a Knowledge Distillation approach, called Self-Mimic Learning (SML), was introduced to improve the weak representation of small pedestrians. It uses the "mimic loss" to teach the features of small objects to be similar to those of larger objects. [75] handles the small pedestrian detection by mimicking the cued recall process in humans. It uses the embedding learning to help detect small objects by recalling the appearance of large objects. Unlike [74], clues from bigger instances can also be used during inference. A self-supervised learning approach was used in Self-supervised Feature Augmentation Network (SFANet) [76]. During training, the backbone takes a pair of images (upsampled and downsampled) as input and uses features extracted from the larger image as guidance.

6. Quantitative Results

Among the datasets described earlier (Tab. 2), we selected VisDrone [20], and AI-TOD [7] due to their frequent use by authors of small object detection methods. In the case of VisDrone, there are slight differences in labels between the 2018 and 2019 versions. Additionally, it has become a common practice to use the validation subset for conducting tests, due to the delayed publication of the test subset and earlier utilization of the validation set by other authors. Therefore, we present metric values for each version separately – VisDrone18-val (Tab. 3) and VisDrone19-val (Tab. 4). For AI-TOD, the reported values relate to the test dataset (Tab. 5). In all comparisons, we used the metric values provided by the authors of each respective method.

For the VisDrone2018 dataset (Tab. 3), among the three methods using ResNet50 or ResNet101 as a backbone, CDMNet achieves the best values for almost all metrics. CDMNet does not use coarse predictions, which may explain its lower APl. DMNet performs better than ClusDet, except for AP_{50} which is higher for ClusDet. The supreme value of AP_{50}, with other metrics lowered, may indicate that ClusDet has some difficulties with precise location. F&D [33] achieves the best results on the VisDrone2018 validation dataset, except for APl. Like CDMNet, F&D does not utilize coarse predictions, however the obtained AP_l value is much higher than CDMNet. Most likely, Incomplete Box Suppression used in F&D reduces the negative impact of region cropping on the quality of large object detection. The second best result (in terms of AP, AP_{50} and AP_{75}) is achieved by SAIC-FPN, which does not report AP broken down by object size, so it is difficult to assess its effectiveness for small objects. Similar to CDMNet and F&D, SAIC also does not use coarse detections, however, due to the lack of AP_l, it is impossible to assess how this affects the detection of large objects. Second best AP_s and AP_m are achieved by CRENet with Hourglass as a backbone network. With a relatively lightweight backbone (DLA-34), CRENet still achieves competitive results. In Tab. 4 the VisDrone2019 results are shown, and AGDN has the highest metrics except for AP_{50}. This method, however, uses a variety of components to improve detection quality. AGDN is the only one to report AP at different scales, so it is not possible to compare methods with respect to different object sizes. Most of the methods in Tab. 4 perform inference with a different input resolution, so it is difficult to isolate the effect of the method itself from the effect that an input size may have on the detection quality. Of the three metrics that have been introduced to replace the IoU, RFLA shows the highest quality on the AI-TOD test subset (Tab. 5), except for AP_m. The main difference between RFLA and NWD is the function used to measure the distance between Gaussian Distributions. As pointed

Table 3. Comparison of selected object detection methods on the VisDrone18-val dataset. We mark the first, second, and third best value for each metric
Tabela 3. Porównanie wybranych metod na zbiorze danych VisDrone18-val. Oznaczamy pierwszą, drugą i trzecią najlepszą wartość każdej metryki

method	backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
ClusDet [29]	ResNet50	26.7	50.6	24.7	17.6	38.9	51.4
DMNet [32]	ResNet50	28.2	47.6	28.9	19.9	39.6	55.8
CDMNet [33]	ResNet50	29.2	49.5	29.8	20.8	40.7	41.6
ClusDet [29]	ResNet101	26.7	50.4	25.2	17.2	39.3	54.9
DMNet [32]	ResNet101	28.5	48.1	29.4	20.0	39.7	57.1
CDMNet [33]	ResNet101	29.7	50.0	30.9	21.2	41.8	42.9
ClusDet [29]	ResNeXt101	28.4	53.2	26.4	19.1	40.8	54.4
DREN [30]	ResNeXt101	27.1	—	—	—	—	—
DMNet [32]	ResNeXt101	29.4	49.3	30.6	**21.6**	41.0	**56.9**
CDMNet [33]	ResNeXt101	30.7	51.3	32.0	22.2	**42.4**	44.7
F&D [34]	ResNeXt101	**42.0**	**66.1**	**44.6**	**32.0**	**47.9**	54.5
SAIC-FPN [26]	ResNeXt101	**35.7**	**63.0**	**35.1**	—	—	—
CRENet [35]	Hourglass	**33.7**	54.3	**33.5**	**25.6**	**45.3**	**58.7**
RRNet [39]	Hourglass	—	**61.1**	32.9	—	—	—
CRENet [35]	DLA-34	30.3	53.7	29.2	**21.6**	41.9	50.6

Table 4. Comparison of selected object detection methods on the VisDrone19-val dataset. We mark the first, second and third best value for each metric (⋆ methods using Cascade R-CNN)
Tabela 4. Porównanie wybranych metod na zbiorze danych VISDRONE19-VAL. Oznaczamy pierwszą, drugą i trzecią najlepszą wartość każdej metryki (⋆ metody korzystające z Cascade R-CNN

method	backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
AGDN [40]	CSPDarknet53	**41.8**	**66.1**	**43.6**	**33.7**	**54.2**	**59.4**
AdaZoom [36]	ResNet50	36.2	63.5	36.1	—	—	—
GLSAN [37]	ResNet50	30.7	55.4	30.0	—	—	—
GLSAN⋆ [37]	ResNet50	32.5	55.8	33.0	—	—	—
DeForm [41]	ResNet50	30.1	58.0	27.5	—	—	—
RFEB⋆ [68]	ResNet50	33.7	58.6	33.9	—	—	—
GLSAN [37]	ResNet101	30.7	55.6	29.9	—	—	—
MPFPN⋆ [56]	ResNet101	29.1	54.4	27.0	—	—	—
AdaZoom [36]	ResNeXt101	**37.6**	**66.3**	**37.3**	—	—	—
AdaZoom⋆ [36]	ResNeXt101	**40.3**	**66.9**	**41.8**	—	—	—

Table 5. Comparison of selected tiny object detection metrics on the AI-TOD test dataset
Tabela 5. Porównanie wybranych metryk do wykrywania bardzo małych obiektów na zbiorze testowym AI-TOD

method	backbone	AP	AP_{50}	AP_{75}	AP_{vt}	AP_t	AP_s	AP_m
DotD [12]	ResNet50	14.9	38.5	9.3	7.2	16.1	17.9	23.7
RFLA [14]	ResNet50	21.1	51.6	13.1	9.5	21.2	26.1	31.5
NWD [78]	ResNet50	17.8	43.8	11.0	2.5	17.0	26.1	34.3

out in [13], NWD is not scale-invariant, which may explain the relatively low AP_{vt} and AP_t, and its superior performance for medium objects. DotD shows the worst quality among the three compared methods. It is also the simplest as IoU is replaced by the Normalized Euclidean Distance.

7. Conclusions and Future Work

In this work, we collected various size definitions and metrics used for evaluation of small and tiny object detection. We also described the challenging datasets, and compared them

taking into account both relative and absolute object size definitions. We gathered multiple methods targeting small and tiny object detection. We divided them into seven groups: focus-and-detect, data augmentation, sampling-based, attention-based, scale-aware, context-aware, and feature-imitation, and discussed each approach extensively. We also presented quantitative results for selected methods. Focus-and-detect systems are successfully applied to drone-based datasets as shown in Tab. 3 and Tab. 4. These coarse-to-fine approaches are mainly used for high-resolution data with many tiny instances, however, due to the truncation effect, they tend to deteriorate the quality of larger objects. Based on the reported values, methods that support the detection process by using a global image tend to report higher AP_l. Incomplete Box Suppression used in [33] seems to significantly improve the quality of small and medium objects. RRNet [38], which introduces data augmentation focused on small objects, also shows promising results. The improvement in the detection quality for tiny objects is also observed for methods replacing IoU with other metrics, such as DotD [12], NWD [77] and RFLA [13]. Despite all efforts, it is difficult to accurately and reliably compare the methods based on the results presented by the authors. Different approaches use different backbones, input resolutions or additional components. Many of the methods studied use self-collected datasets, and even when a publicly available dataset is used, there are still variations in the versions and splits used for evaluation. In the future, we plan to analyze the most promising approaches on a common benchmark, taking into account the inference time, which is crucial in practical applications such as mobile robotics.

Acknowledgements

The research was supported by the Ministry of Education and Science as part of the "Doktorat Wdrożeniowy" program (DWD/5/0203/2021).

References

1. Bochkovskiy A., Wang C.-Y., Liao H.-Y.M., *Yolov4: Optimal speed and accuracy of object detection*, arXiv preprint arXiv:2004.10934, 2020, DOI: 10.48550/arXiv.2004.10934.
2. Girshick R., Donahue J., Darrell T., Malik J., *Rich feature hierarchies for accurate object detection and semantic segmentation*, [In:] Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, 580–587, DOI: 10.1109/CVPR.2014.81.
3. Zhou X., Wang D., Krähenbühl P., *Objects as points*, arXiv preprint arXiv:1904.07850, 2019, DOI: 10.48550/arXiv.1904.07850.
4. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A.C., *Ssd: Single shot multibox detector*, [In:] Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, 2016, 21–37, DOI: 10.1007/978-3-319-46448-0_2.
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, 2017, 2117–2125, DOI: 10.1109/CVPR.2017.106.
6. Kos A., Majek K., *CNN-based traffic sign detection on embedded devices*, [In:] Proceedings of the 3rd Polish Conference on Artificial Intelligence, April 25-27, 2022, Gdynia, Poland, 108–111, [Online]. Available: https://wydawnictwo.umg.edu.pl/pp-rai2022/pdfs/25_pp-rai-2022-016.pdf.
7. Wang J., Yang W., Guo H., Zhang R., Xia G.-S., *Tiny object detection in aerial images*, [In:] 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021, 3791–3798, DOI: 10.1109/ICPR48806.2021.9413340.
8. Xu C., Wang J., Yang W., Yu H., Yu L., Xia G.-S., *Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark*, "ISPRS Journal of Photogrammetry and Remote Sensing", Vol. 190, 2022, 79–93, DOI: 10.1016/j.isprsjprs.2022.06.002.
9. Yu X., Gong Y., Jiang N., Ye Q., Han Z., *Scale match for tiny person detection*, [In:] Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2020, 1257–1265, DOI: 10.1109/WACV45572.2020.9093394.
10. Xia G.-S., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Datcu M., Pelillo M., Zhang L., *DOTA: A large-scale dataset for object detection in aerial images*, [In:] Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 3974–3983.
11. Cheng G., Yuan X., Yao X., Yan K., Zeng Q., Han J., *Towards large-scale small object detection: Survey and benchmarks*, arXiv preprint arXiv:2207.14096, 2022, DOI: 10.1109/TPAMI.2023.3290594.
12. Xu C., Wang J., Yang W., Yu L., *Dot Distance for Tiny Object Detection in Aerial Images*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2021, 1192–1201, DOI: 10.1109/CVPRW53098.2021.00130.
13. Xu C., Wang J., Yang W., Yu H., Yu L., Xia G.-S., *RFLA: Gaussian receptive field based label assignment for tiny object detection*, [In:] Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX. Springer, 2022, 526–543, DOI: 10.1007/978-3-031-20077-9_31.
14. Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick C.L., *Microsoft COCO: Common objects in context*, [In:] European Conference on Computer Vision. Springer, 2014, 740–755, DOI: 10.1007/978-3-319-10602-1_48.
15. Yang S., Luo P., Loy C.-C., Tang X., *WIDER FACE: A face detection benchmark*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 5525–5533, DOI: 10.1109/CVPR.2016.596.
16. Lam D., Kuzma R., McGee K., Dooley S., Laielli M., Klaric M., Bulatov Y., McCord B., *xView: Objects in context in overhead imagery*, arXiv preprint arXiv:1802.07856, 2018, DOI: 10.48550/arXiv.1802.07856.
17. Yu X., Chen P., Wu D., Hassan N., Li G., Yan J., Shi H., Ye Q., Han Z., *Object localization under single coarse point supervision*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, 4868–4877, DOI: 10.1109/CVPR52688.2022.00482.
18. Varga L.A., Kiefer B., Messmer M., Zell A., *SeaDronesSee: A maritime benchmark for detecting humans in open water*, [In:] Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, 3686–3696, DOI: 10.1109/WACV51458.2022.00374.
19. Du D., Qi Y., Yu H., Yang Y., Duan K., Li G., Zhang W., Huang Q., Tian Q., *The unmanned aerial vehicle benchmark: Object detection and tracking*, [In:] Proceedings of the European conference on computer vision (ECCV), 2018, 375–391, DOI: 10.1007/978-3-030-01249-6_23.
20. Zhu P., Wen L., Du D., Bian X., Fan H., Hu Q., Ling H., *Detection and tracking meet drones challenge*, "IEEE Transactions on Pattern Analysis and Machine Intelligence", Vol. 44, No. 11, 2021, 7380–7399, DOI: 10.1109/TPAMI.2021.3119563.
21. Ding J., Xue N., Xia G.-S., Bai X., Yang W., Yang M. Y., Belongie S., Luo J., Datcu M., Pelillo M., Zhang L., *Object detection in aerial images: A large-scale benchmark and challenges*, IEEE transactions on pattern analysis and machine intelligence, Vol. 44, No. 11, 2021, 7778–7796, DOI: 10.1109/TPAMI.2021.3117983.

22. Everingham M., Van Gool L., Williams C.K., Winn J., Zisserman A., *The PASCAL Visual Object Classes* (*VOC*) *challenge*, "International Journal of Computer Vision", Vol. 88, No. 2, 2010, 303–338, DOI: 10.1007/s11263-009-0275-4.
23. Özge Ünel F., Özkalayci B.O., Cigla C., *The power of tiling for small object detection*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, DOI: 10.1109/CVPRW.2019.00084.
24. Xie X., Cheng G., Li Q., Miao S., Li K., Han J., *Fewer sis more: Efficient object detection in large aerial images* , arXiv preprint arXiv:2212.13136, 2022, DOI: 10.1007/s11432-022-3718-5.
25. Zhou J., Vong C.-M., Liu Q., Wang Z., *Scale adaptive image cropping for UAV object detection*, "Neurocomputing", Vol. 366, 2019, 305–313, DOI: 10.1016/j.neucom.2019.07.073.
26. Růžička V., Franchetti F., *Fast and accurate object detection in high resolution 4K and 8K video using GPUs*, [In:] 2018 IEEE High Performance extreme Computing Conference (HPEC). IEEE, 2018, DOI: 10.1109/HPEC.2018.8547574.
27. Plastiras G., Kyrkou C., Theocharides T., *Efficient ConvNet-based Object Detection for Unmanned Aerial Vehicles by Selective Tile Processing*, [In:] Proceedings of the 12th International Conference on Distributed Smart Cameras, 2018, DOI: 10.1145/3243394.3243692.
28. Yang F., Fan H., Chu P., Blasch E., Ling H., *Clustered object detection in aerial images*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, 8311–8320, DOI: 10.1109/ICCV.2019.00840.
29. Zhang J., Huang J., Chen X., Zhang D., *How to fully exploit the abilities of aerial image detectors*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, DOI: 10.1109/ICCVW.2019.00007.
30. Kos A., Majek K., Belter D., *Where to look for tiny objects? ROI prediction for tiny object detection in high resolution images*, [In:] 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV). IEEE, 2022, 721–726, DOI: 10.1109/ICARCV57592.2022.10004372.
31. Li C., Yang T., Zhu S., Chen C., Guan S., *Density map guided object detection in aerial images*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, 190–191, DOI: 10.1109/CVPRW50498.2020.00103.
32. Duan C., Wei Z., Zhang C., Qu S., Wang H., *Coarse-grained density map guided object detection in aerial images*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, 2789–2798, DOI: 10.1109/ICCVW54120.2021.00313.
33. Koyun O.C., Keser R.K., Akkaya I.B., Töreyin B.U., *Focus-and-Detect: A small object detection framework for aerial images*, "Signal Processing: Image Communication", Vol. 104, 2022, DOI: 10.1016/j.image.2022.116675.
34. Wang Y., Yang Y., Zhao X., *Object detection using clustering algorithm adaptive searching regions in aerial images*, [In:] European Conference on Computer Vision. Springer, 2020, 651–664, DOI: 10.1007/978-3-030-66823-5_39.
35. Xu J., Li Y., Wang S., *AdaZoom: adaptive zoom network for multiscale object detection in large scenes*, arXiv preprint arXiv:2106.10409, 2021, DOI: 10.48550/arXiv.2106.10409.
36. Deng S., Li S., Xie K., Song W., Liao X., Hao A., Qin H., *A global-local self-adaptive network for drone-view object detection*, "IEEE Transactions on Image Processing", Vol. 30, 2020, 1556–1569, DOI: 10.1109/TIP.2020.3045636.
37. Kisantal M., Wojna Z., Murawski J., Naruniec J., Cho K., *Augmentation for small object detection*, arXiv preprint arXiv:1902.07296, 2019, DOI: 10.48550/arXiv.1902.07296.
38. Chen C., Zhang Y., Lv Q., Wei S., Wang X., Sun X., Dong J., *RRNet: A hybrid detector for object detection in drone-captured images*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, DOI: 10.1109/ICCVW.2019.00018.
39. Wang X., Zhu D., Yan Y., *Towards efficient detection for small objects via attention-guided detection network and data augmentation*, "Sensors", Vol. 22, No. 19, 2022, DOI: 10.3390/s22197663.
40. Zhang X., Izquierdo E., Chandramouli K., *Dense and small object detection in UAV vision based on cascade network*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, DOI: 10.1109/ICCVW.2019.00020.
41. Bosquet B., Cores D., Seidenari L., Brea V.M., Mucientes M., Del Bimbo A., *A full data augmentation pipeline for small object detection based on generative adversarial networks*, "Pattern Recognition", Vol. 133, 2023, DOI: 10.1016/j.patcog.2022.108998.
42. Zhang S., Zhu X., Lei Z., Shi H., Wang X., Li S.Z., *S^3FD: Single Shot Scale-invariant Face Detector*, [In:] Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, 192–201, DOI: 10.1109/ICCV.2017.30.
43. Zhu C., Tao R., Luu K., Savvides M., *Seeing small faces from robust anchor's perspective*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, 5127–5136, DOI: 10.1109/CVPR.2018.00538.
44. Yang X., Yang J., Yan J., Zhang Y., Zhang T., Guo Z., Sun X., Fu K., *SCRDet: Towards more robust detection for small, cluttered and rotated objects*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, 8232–8241, DOI: 10.1109/ICCV.2019.00832.
45. Yi K., Jian Z., Chen S., Zheng N., *Feature selective small object detection via knowledge-based recurrent attentive neural network*, arXiv preprint arXiv:1803.05263, 2018.
46. Lu X., Ji J., Xing Z., Miao Q., *Attention and feature fusion SSD for remote sensing object detection*, "IEEE Transactions on Instrumentation and Measurement", Vol. 70, 2021, DOI: 10.1109/TIM.2021.3052575.
47. Ran Q., Wang Q., Zhao B., Wu Y., Pu S., Li Z., *Lightweight oriented object detection using multiscale context and enhanced channel attention in remote sensing images*, "IEEE Journal of Selected Topics Applied Earth Observations and Remote Sensing", Vol. 14, 2021, 5786–5795, DOI: 10.1109/JSTARS.2021.3079968.
48. Li Y., Huang Q., Pei X., Chen Y., Jiao L., Shang R., *Cross-layer attention network for small object detection in remote sensing imagery*, "IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing", Vol. 14, 2020, 2148–2161, DOI: 10.1109/JSTARS.2020.3046482.
49. Fu J., Sun X., Wang Z., Fu K., *An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images*, "IEEE Transactions on Geoscience and Remote Sensing", Vol. 59, No. 2, 2020, 1331–1344, DOI: 10.1109/TGRS.2020.3005151.
50. Hu J., Shen L., Sun G., *Squeeze-and-Excitation Networks*, [In:] Proceedings of the IEEE/CVF Conference on Com-

puter Vision and Pattern Recognition, 2018, 7132–7141, DOI: 10.1109/CVPR.2018.00745.
51. Zhang H., Wang K., Tian Y., Gou C., Wang F.-Y., *MFR-CNN: Incorporating multi-scale features and global information for traffic object detection*, "IEEE Transactions on Vehicular Technology", Vol. 67, No. 9, 2018, 8019–8030, DOI: 10.1109/TVT.2018.2843394.
52. Liu Z., Gao G., Sun L., Fang L., *IPG-Net: Image pyramid guidance network for small object detection*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, DOI: 10.1109/CVPRW50498.2020.00521.
53. Gong Y., Yu X., Ding Y., Peng X., Zhao J., Han Z., *Effective fusion factor in FPN for tiny object detection*, [In:] Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021, 1160–1168, DOI: 10.1109/WACV48630.2021.00120.
54. Hong M., Li S., Yang Y., Zhu F., Zhao Q., Lu L., *SSPNet: Scale selection pyramid network for tiny person detection from UAV images*, "IEEE Geoscience and Remote Sensing Letters", Vol. 19, 2021, DOI: 10.1109/LGRS.2021.3103069.
55. Liu Y., Yang F., Hu P., *Small-object detection in UAV-captured images via multi-branch parallel feature pyramid networks*, "IEEE Access", Vol. 8, 2020, 145 740–145 750, DOI: 10.1109/ACCESS.2020.3014910.
56. Pang J., Li C., Shi J., Xu Z., Feng H., *R^2-CNN: Fast tiny object detection in large-scale remote sensing images*. arXiv preprint arXiv:1902.06042, DOI: 10.1109/TGRS.2019.2899955.
57. Li J., Wang Y., Wang C., Tai Y., Qian J., Yang J., Wang C., Li J., Huang F., *DSFD: Dual Shot Face Detector*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 5060–5069, DOI: 10.1109/CVPR.2019.00520.
58. Najibi M., Samangouei P., Chellappa R., Davis L.S., *SSH: Single Stage Headless Face Detector*, [In:] Proceedings of the IEEE International Conference on Computer Vision, 2017, 4885–4894, DOI: 10.1109/ICCV.2017.522.
59. Yang C., Huang Z., Wang N., *QueryDet: Cascaded sparse query for accelerating high-resolution small object detection*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, 13 658–13 677, DOI: 10.1109/CVPR52688.2022.01330.
60. Chen C., Liu M.-Y., Tuzel O., Xiao J., *R-CNN for small object detection*, [In:] ACCV 2016: 13th Asian Conference on Computer Vision, Revised Selected Papers, Part V 13. Springer, 2017, 214–230, DOI: 10.1007/978-3-319-54193-8_14.
61. Bell S., Zitnick C.L., Bala K., Girshick R., *Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 2874–2883, DOI: 10.1109/CVPR.2016.314.
62. Hu P., Ramanan D., *Finding tiny faces*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 1522–1530, DOI: 10.1109/CVPR.2017.166.
63. Tang X., Du D.K., He Z., Liu J., *PyramidBox: A context-assisted single shot face detector*, [In:] Proceedings of the European Conference on Computer Vision (ECCV), 2018, 812–828, DOI: 10.1007/978-3-030-01240-3_49.
64. Liang X., Zhang J., Zhuo L., Li Y., Tian Q., *Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis*," IEEE Transactions on Circuits and Systems for Video Technology", Vol. 30, No. 6, 2019, 1758–1770, DOI: 10.1109/TCSVT.2019.2905881.
65. Hu X., Xu X., Xiao Y., Chen H., He S., Qin J., Heng P.-A., *SINet: A scale-insensitive convolutional neural network for fast vehicle detection*, "IEEE Transactions on Intelligent Transportation Systems", Vol. 20, No. 3, 2018, 1010–1019, DOI: 10.1109/TITS.2018.2838132.
66. Zhang G., Lu S., Zhang W., *CAD-Net: A context-aware detection network for objects in remote sensing imagery*, "IEEE Transactions on Geoscience and Remote Sensing", Vol. 57, No. 12, 2019, 10 015–10 024, DOI: 10.1109/TGRS.2019.2930982.
67. Wang H., Wang Z., Jia M., Li A., Feng T., Zhang W., Jiao L., *Spatial Attention for Multi-Scale Feature Refinement for Object Detection*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, DOI: 10.1109/ICCVW.2019.00014.
68. 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, 2017, 1222–1230, DOI: 10.1109/CVPR.2017.211.
69. Bai Y., Zhang Y., Ding M., Ghanem B., *Finding tiny faces in the wild with generative adversarial network*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 21–30, DOI: 10.1109/CVPR.2018.00010.
70. Bai Y., Zhang Y., Ding M., Ghanem B., *SOD-MTGAN: Small object detection via multi-task generative adversarial network*, [In:] Proceedings of the European Conference on Computer Vision (ECCV), 2018, 210–226, DOI: 10.1007/978-3-030-01261-8_13.
71. Noh J., Bae W., Lee W., Seo J., Kim G., *Better to Follow, Follow to Be Better: Towards Precise Supervision of Feature Super-Resolution for Small Object Detection*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, 9725–9734, DOI: 10.1109/ICCV.2019.00982.
72. Jiang K., Wang Z., Yi P., Wang G., Lu T., Jiang J., *Edge-Enhanced GAN for Remote Sensing Image Superresolution*, "IEEE Transactions on Geoscience and Remote Sensing", Vol. 57, No. 8, 2019, 5799–5812, DOI: 10.1109/TGRS.2019.2902431.
73. Ji H., Gao Z., Mei T., Ramesh B., *Vehicle detection in remote sensing images leveraging on simultaneous super-resolution*, "IEEE Geoscience and Remote Sensing Letters", Vol. 17, No. 4, 2019, 676–680, DOI: 10.1109/LGRS.2019.2930308.
74. Wu J., Zhou C., Zhang Q., Yang M., Yuan J., *Self-mimic learning for small-scale pedestrian detection*, [In:] Proceedings of the 28th ACM International Conference on Multimedia, 2020, 2012–2020, DOI: 10.1145/3394171.3413634.
75. Kim J.U., Park S., Ro Y.M., *Robust small-scale pedestrian detection with cued recall via memory learning*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, 3050–3059, DOI: 10.1109/ICCV48922.2021.00304.
76. Pan X., Tang F., Dong W., Gu Y., Song Z., Meng Y., Xu P., Deussen O., Xu C., *Self-Supervised Feature Augmentation for Large Image Object Detection*, "IEEE Transactions on Image Processing", Vol. 29, 2020, 6745–6758, DOI: 10.1109/TIP.2020.2993403.
77. Wang J., Xu C., Yang W., Yu L., *A normalized gaussian Wasserstein distance for tiny object detection*, arXiv preprint arXiv:2110.13389, 2021, DOI: 10.48550/arXiv.2110.13389.

Przegląd metod uczenia głębokiego w wykrywaniu małych i bardzo małych obiektów

Streszczenie: W ostatnich latach, dzięki rozwojowi metod uczenia głębokiego, dokonano znacznego postępu w detekcji obiektów i innych zadaniach widzenia maszynowego. Mimo że ogólne wykrywanie obiektów staje się coraz mniej problematyczne dla nowoczesnych algorytmów, a średnia precyzja dla średnich i dużych instancji w zbiorze COCO zbliża się odpowiednio do 70 i 80 procent, wykrywanie małych obiektów pozostaje nierozwiązanym problemem. Ograniczone informacje o wyglądzie, rozmycia i niski stosunek sygnału do szumu powodują, że najnowocześniejsze detektory zawodzą, gdy są stosowane do małych obiektów. Tradycyjne ekstraktory cech opierają się na próbkowaniu w dół, które może powodować zanikanie najmniejszych obiektów, a standardowe metody przypisania kotwic są mniej skuteczne w wykrywaniu instancji o małej liczbie pikseli. W niniejszej pracy dokonujemy wyczerpującego przeglądu literatury dotyczącej wykrywania małych i bardzo małych obiektów. Przedstawiamy definicje, rozróżniamy małe wymiary bezwzględne i względne oraz podkreślamy związane z nimi wyzwania. Kompleksowo omawiamy zbiory danych, metryki i metody, a na koniec dokonujemy porównania ilościowego na trzech publicznie dostępnych zbiorach danych.

Słowa kluczowe: uczenie głębokie, wykrywanie małych obiektów, wykrywanie bardzo małych obiektów, zbiory danych bardzo małych obiektów, metody wykrywania bardzo małych obiektów

Aleksandra Kos, MSc, Eng.
aleksandra.kos@doctorate.put.poznan.pl
ORCID: 0000-0001-9726-4472

PhD student at Poznan University of Technology, Deep Learning Developer at Cufix. She holds a Master's degree in Automatic Control and Robotics, and completed her studies at the Warsaw University of Technology. Professionally, since 2019, she has been working on deep neural networks in image analysis. She is particularly interested in the detection of small objects in high-resolution images – this is the subject of her Applied Doctorate, which she has been pursuing since 2021 in cooperation with Poznan University of Technology and Cufix company.

Karol Majek, PhD
karolmajek@cufix.pl
ORCID: 0000-0002-1351-8496

Currently working at Cufix on Computer Vision problems such as object detection and tracking. Previously researcher at NASK PIB, Fraunhofer FKIE and Institute of Mathematical Machines. His main research interests are learning based computer vision techniques from classification to object detection.

Dominik Belter, PhD, DSc
dominik.belter@put.poznan.pl
ORCID: 0000-0003-3002-9747

Graduated from the Poznan University of Technology (2007). He received a Ph.D. degree in robotics from the same University in 2012. Since 2012, he has been an Assistant Professor at the Institute of Control and Information Engineering of the Poznan University of Technology. He has been an Associate Professor at the Poznan University of Technology since 2021. He received a DSc. degree in robotics from the same University in 2020. He spent a year working as a postdoc in the Intelligent Robotics Laboratory at the University of Birmingham in the years 2013-2016. Dominik Belter has been taking part as an investigator in 3 EC and 6 national projects. He serves as a member of the scientific board of the CLAWAR and EMCR conferences. He is the author or co-author of over 90 technical papers in the fields of robotics and computer science. His research interests include walking robots, machine learning, vision, robot manipulation, and soft computing. He has been involved in Erasmus+ and IAESTE internships and supervised students from multiple countries. Since 2023, he has been a member of the Committee on Automatic Control and Robotics Committee of the Polish Academy of Sciences