International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Advancements and Challenges in Small Object Detection: A Comprehensive Review of Methods, Applications, and Future Directions

Versha Sharma, Rahul Gupta
Department of Computer Science and engineering,
S.R Institute of management & Technology, Lucknow, India,
svarsha0987@gmail.com

Abstract: Small object detection is a fundamental yet challenging problem in computer vision, with significant applications in areas such as autonomous driving, surveillance, medical imaging, and remote sensing. Detecting small objects is inherently difficult due to their low resolution, poor feature representation, and vulnerability to occlusion and background noise. Traditional object detection models often underperform in this context, primarily due to scale variance and the imbalance of small objects in training datasets. Despite these advances, small object detection still faces challenges in achieving robust generalization, real-time performance, and interpretability. This paper provides an overview of current methods, highlights the persistent limitations, and explores future directions for research to improve the accuracy and efficiency of small object detection systems in real-world scenarios.

Keywords: small object; detection; tracking; computer vision; survey

1. Introduction:

Small object detection is a critical yet challenging area in computer vision, with applications in surveillance, autonomous driving, medical imaging, and remote sensing. The task involves identifying and classifying objects that occupy a relatively small number of pixels in an image, often in cluttered or low-resolution scenarios. Despite advancements in object detection algorithms, small object detection remains a bottleneck due to issues such as low feature representation, scale variance, and occlusion.

2. Challenges

1. **Scale Variance**: Small objects often appear at different scales depending on their distance from the camera. Standard convolutional neural networks (CNNs) struggle to generalize effectively across scales. While multi-scale training and data augmentation help, they are not always sufficient.
2. **Low Resolution and Detail**: Small objects have fewer pixels, making it difficult to capture fine details. This often leads to poor feature representation and lower detection accuracy.
3. **Clutter and Background Noise**: Small objects are more likely to be confused with background textures or occlusions, further complicating the detection task.
4. **Imbalance in Training Data**: Most datasets are skewed toward larger objects, making small object detection a less emphasized task during training.

3. Advances in Small Object Detection

Several innovations have addressed these challenges:

1. **Feature Pyramid Networks (FPNs)**: FPNs enhance multi-scale feature representation by combining low-level and high-level features. This has proven effective for small object detection by leveraging fine-grained details from early layers of CNNs.
2. **Super-Resolution Techniques**: Upsampling strategies, such as those used in generative adversarial networks (GANs), enhance the resolution of small objects before detection. This pre-processing step has shown promise in improving accuracy.
3. **Anchor-Free Detectors**: Models like FCOS and CenterNet eliminate the reliance on predefined anchor boxes, which are often ineffective for small objects due to their size mismatch.
4. **Attention Mechanisms**: Incorporating attention modules helps focus on regions of interest, improving the model's ability to distinguish small objects from background noise.
5. **Data Augmentation**: Synthetic data generation, along with targeted augmentation strategies like random cropping and resizing, increases the representation of small objects during training.

4. Techniques for Enhancing Small Object Detection

Several strategies have been developed to address these challenges:

1. Multi-Scale Feature Fusion: Models like Feature Pyramid Networks (FPN) and architectures leveraging multi-scale feature maps help retain fine-grained features for small objects.
2. Super-Resolution Techniques: Pre-processing steps such as upscaling small object regions using super-resolution methods can improve feature extraction.
3. Data Augmentation and Synthesis: Techniques like oversampling, random cropping, and using generative adversarial networks (GANs) to create synthetic small object instances improve model training.
4. Attention Mechanisms: Attention modules can highlight regions of interest, enabling the model to

focus more on small objects in cluttered backgrounds.

5. Loss Function Customization: Loss functions such as focal loss help mitigate the impact of class imbalance by assigning higher weights to hard-to-detect small objects.

5. Applications

Small object detection finds use in a variety of domains:

- **Autonomous Vehicles**: Detecting pedestrians, traffic signs, and other small but critical objects in complex environments.
- **Surveillance**: Identifying unauthorized drones or tracking individuals in large crowds.
- **Remote Sensing**: Detecting small vehicles, ships, or wildlife in satellite or aerial images.
- **Medical Imaging**: Identifying tiny abnormalities, such as tumors or polyps, in diagnostic scans.

6. Related Work:

In the study by Habibi et al. [18], tracking was integrated with a super-resolution technique wherein a high-resolution image was created from multiple low-resolution images. Given that super-resolution enhanced the visual quality of small objects, the process provided more tracking information, thereby increasing precision. The tracking process was then conducted through an adaptive Particle filter, as proposed in Huang et al. [17].

Liu et al. [19] put forth an approach grounded in super-resolution, using convolutional neural network (CNN) to track small objects. A deep-learning network was deployed to enhance the visual quality of small objects, subsequently improving the tracking performance. A Particle filter was then employed for tracking [20].

In their research,Wu et al. [21] proposed an enhanced kernel correlation filter (KCF)- based approach for tracking small objects in satellite videos. Occlusion presents a significant hurdle in object tracking, especially apparent in satellite videos due to the minute size of objects, making them more susceptible to occlusion. The methodology in this study used the average peak correlation energy and the peak value of the response map to determine potential object occlusion. The object's subsequent location was forecasted employing a Kalman filter.

Notable among these methods is the algorithm proposed by Blostein et al. [22], dubbed Multiple Hypothesis Testing (MHT). This method operates under the assumption that the intensity values of background and noise are lower than the mean target intensity. In this approach, the track tree roots are chosen from a predetermined number of points with the highest intensity value. For each root, the algorithm selects neighboring points in the subsequent frame to construct a track tree. Within MHT, there are two thresholds—T_1 and T_2—against which each point on the track is compared. If the new point surpasses T_2, the algorithm records the track and proceeds to the next frame. If the point falls below T_1, the track is rejected. However, if the new point lies between T_1 and T_2, the algorithm defers the decision to the next frame. Ultimately, the tree is pruned to yield a desired number of tracks. Nonetheless, this method faces challenges when tracking fast-moving small objects, as the search area increases exponentially. This makes current MHT algorithms computationally impractical for objects moving at speeds exceeding 1 pixel/frame. In response to this problem, Ahmadi et al. [23] utilized the Multi-Objective Particle Swarm Optimization algorithm (MOPSO) [23] to identify the most optimal track within each root.

Salari et al. [24] presented an effective algorithm for tracking dim targets within digital image sequences. The algorithm operates in two stages: noise removal and tracking. Initially, the Total Variation (TV) filtering technique is employed to improve the Signal Noise Ratio (SNR) and eliminate the image's noise. Subsequently, to detect and track dim tiny targets, a genetic algorithm with associated genetic operators and encoding is used.

In the study by Shaik et al. [25], Bayesian techniques were deployed for the detection and tracking of targets in infrared (IR) images. The algorithm begins by applying preprocessing to incoming IR targets to reduce noise and segmentation. The initial position of the object is ascertained utilizing ground truth (GT) data. Subsequently, a grid composed of segments around the target's position in the ensuing frame is chosen, and regions with high-intensity within this segment are highlighted. Employing Bayesian probabilistic methodologies, the likelihood of the object shifting its position from the current frame to any high-intensity location within this grid is then calculated. The position suggesting the highest probability is chosen, and the object's position in the following frame is established. Given that an object's intensity may not necessarily be the highest in a frame, the position and intensity of the object in the previous frame are considered in the Bayesian probabilistic equation to determine its position in the next frame.

An alternative methodology was introduced by Srivastav et al. [28], which incorporated three-frame differencing and background subtraction for detecting moving objects in videos. The procedure commences with the selection of three successive frames from the image sequence. Subsequently, the difference between the first and second frames is computed, denoted as D_1. Similarly, the outcome of the difference between the second and third frames is labeled as D_2. If DB signifies the result of subtracting the background from the current frame, moving objects are detected by implementing a pixel-wise logical OR operation on D_1, D_2, and DB. Finally, background noise is eliminated by utilizing a median filter.

Zhu et al. [29] incorporated three-frame differencing and operations such as "AND" and "XOR" for swift detection of moving objects. The difference image, p_1, is obtained by calculating the difference between the initial two frames, and p_2 is obtained from the difference between the second and third frames. Subsequently, a new image, p_3, is created by performing p_1 AND p_2. The next step involves obtaining p_2 XOR p_3, resulting in a new image, p_4. Ultimately, the detection image is derived from p_1 AND p_4. Following detection, noise is mitigated using post-processing algorithms. In their research, Yin et al. [30] proposed an algorithm known asMotionModeling Baseline (MMB), designed to detect and track small, densely clustered moving objects in satellite videos. The process commences with the extraction of candidate slow-moving pixels and region of interest proposals using accumulative multi-frame differencing (AMFD). The full targets are then efficiently detected using low-rank matrix completion (LRMC). Lastly, the motion trajectory-based false alarm filter mitigates false alarms by compiling the trajectory over time, underlining that authentic moving targets are more likely to exhibit continuous trajectories.

Zhou et al. [31] presented a study that utilized an efficient and unsupervised approach, employing background subtraction for object delineation inWide Area Motion Imagery (WAMI). Initially, background subtraction is used to detect low contrast and small objects, leading to the extraction of objects of interest. Following this, a convolutional neural network (CNN) is trained to reduce false alarms by considering both temporal and spatial data. Another CNN is subsequently trained to forecast the positions of several moving targets within a specified area, thus reducing the complexity of the necessary multi-target tracker. A Gaussian Mixture-Probability Hypothesis Density (GM-PHD) filter is finally employed to correlate detections over time.

Teutsch et al. [32], proposed an algorithm for detecting moving vehicles in Wide Area Motion Imagery that enhanced object detection by utilizing two-frame differencing along with a model of the vehicle's appearance. The algorithm amalgamates robust vehicle detection with the management of splitting and merging, and applies an appearancebased similarity measure to estimate assignment likelihoods among object hypotheses in consecutive frames.

Aguilar et al. [33] proposed a multi-object tracking (MOT) technique for tracking small moving objects in satellite videos. They used a patch-based CNN object detector with a three-frame difference algorithm to concentrate on specific regions and detect adjacent small targets. To improve object location accuracy, they applied the Faster Region-based convolutional neural network (Faster R-CNN) [34] since the three-frame difference algorithm neither regularizes targets by area nor captures slow-moving targets. Furthermore, they applied a direct MOT data-association approach facilitated by an improved GM-PHD filter for multi-target tracking.

This approach was advanced by Aguilar et al. [35], where the performance of Faster R-CNN's object detection was significantly boosted by merging motion and appearance data on extracted patches. The new approach comprises two steps: initially obtaining rough target locations using a lightweight motion detection operator and, then, to enhance the detection results, combining this information with a CNN. An online track-by-detection methodology is also applied during the tracking process to convert detections into tracks based on the Probability Hypothesis Density (PHD) filter.

In the research conducted by Lyu et al. [36], a real-time tracking algorithm was introduced, specifically designed for ball-shaped, fast-moving objects, leveraging frame difference and multi-feature fusion. The process initiates by applying frame difference between two consecutive frames, after which the resulting differential image is segmented into smaller contours. A multi-feature-based algorithm is then used to determine if these are moving areas with ball-shaped objects.

Hongshan et al. [37] proposed a wiener filter-based infrared tiny object detection and tracking technique that optimizes filtering under stable conditions based on the least mean square error metrics. Given that the background is distributed in the image's low-frequency part and the high-frequency part primarily encompasses small objects, an adaptive background suppression algorithm is performed, taking advantage of the low-pass Wiener filter's characteristics. Appropriate segmentation then reveals potential targets. The relationship between multiple frames, including the continuity and regularity of target motion, is utilized for detection and tracking.

In the research conducted by Deshpande et al. [38], they applied max-mean and maxmedian filters on a series of infrared images for the detection of small objects. The initial step involves applying either the max-mean or max-median filter to the unprocessed image. Subsequently, the filtered image is subtracted from the original one to highlight potential targets. A thresholding step, which is guided by the image's statistical characteristics, limits the quantity of potential target pixels. Finally, the output images are cumulatively processed to track the target. The post-processing algorithm is equipped to detect the continuous trajectory of the moving target.

7. Conclusion

Small object detection remains a vibrant research area with high stakes and opportunities for innovation. Continued efforts to improve scale invariance, feature representation, and model robustness are key to unlocking its full potential. As datasets grow richer and computational resources improve, the field is poised for breakthroughs that will significantly impact a wide range of industries.

References:

[1]. Zhou, J.T.; Du, J.; Zhu, H.; Peng, X.; Liu, Y.; Goh, R.S.M. AnomalyNet: An Anomaly Detection Network for Video Surveillance.
IEEE Trans. Inf. Forensics Secur. 2019, 14, 2537–2550. [CrossRef]

[2]. Zhu, L.; Yu, F.R.; Wang, Y.; Ning, B.; Tang, T. Big Data Analytics in Intelligent Transportation Systems: A Survey. IEEE Trans.
Intell. Transp. Syst. 2019, 20, 383–398. [CrossRef]

[3]. Hua, S.; Kapoor, M.; Anastasiu, D.C. Vehicle Tracking and Speed Estimation from Traffic Videos. In Proceedings of the IEEE
Computer Society Conference on Computer Vision and Pattern RecognitionWorkshops, Salt Lake City, UT, USA, 18–22 June 2018;
Volume 2018.

[4]. Hagiwara, T.; Ota, Y.; Kaneda, Y.; Nagata, Y.; Araki, K. Method of Processing Closed-Circuit Television Digital Images for Poor Visibility Identification. Transp. Res. Rec. 2006, 1973, 95–104. [CrossRef]

[5]. Crocker, R.I.; Maslanik, J.A.; Adler, J.J.; Palo, S.E.; Herzfeld, U.C.; Emery, W.J. A Sensor Package for Ice Surface Observations
Using Small Unmanned Aircraft Systems. IEEE Trans. Geosci. Remote Sens. 2012, 50, 1033–1047. [CrossRef]

[6]. Zhang, F.; Du, B.; Zhang, L.; Xu, M. Weakly Supervised Learning Based on Coupled Convolutional Neural Networks for Aircraft Detection. IEEE Trans. Geosci. Remote Sens. 2016, 54, 5553–5563. [CrossRef]

[7]. Zhou, H.;Wei, L.; Lim, C.P.; Creighton, D.; Nahavandi, S. Robust Vehicle Detection in Aerial Images Using Bag-of-Words and
Orientation Aware Scanning. IEEE Trans. Geosci. Remote Sens. 2018, 56, 7074–7085. [CrossRef]
[8]. de Vries, E.T.; Tang, Q.; Faez, S.; Raoof, A. Fluid Flow and Colloid Transport Experiment in Single-Porosity Sample; Tracking of Colloid Transport Behavior in a Saturated Micromodel. Adv. Water Resour. 2022, 159, 104086. [CrossRef]
[9]. Deliba¸so˘ glu, ˙I. Moving Object Detection Method with Motion Regions Tracking in Background Subtraction. Signal Image Video Process. 2023, 17, 2415–2423. [CrossRef]
[10]. Tsai, C.Y.; Shen, G.Y.; Nisar, H. Swin-JDE: Joint Detection and Embedding Multi-Object Tracking in Crowded Scenes Based on Swin-Transformer. Eng. Appl. Artif. Intell. 2023, 119, 105770. [CrossRef]
[11]. Luo, W.; Xing, J.; Milan, A.; Zhang, X.; Liu, W.; Kim, T.K. Multiple Object Tracking: A Literature Review. Artif. Intell. 2021,
293, 103448. [CrossRef]
[12]. Desai, U.B.; Merchant, S.N.; Zaveri, M.; Ajishna, G.; Purohit, M.; Phanish, H.S. Small Object Detection and Tracking: Algorithm, Analysis and Application. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3776. [CrossRef]
[13]. Rout, R.K. A Survey on Object Detection and Tracking Algorithms; National Institute of Technology Rourkela: Rourkela, India, 2013.
[14]. Yilmaz, A.; Javed, O.; Shah, M. Object Tracking: A Survey. Acm Comput. Surv. (CSUR) 2006, 38, 13-es. [CrossRef]
[15]. 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 Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2014; Volume 8693.
[16]. Naik, B.T.; Hashmi, M.d.F. YOLOv3-SORT: Detection and Tracking Player/Ball in Soccer Sport. J. Electron. Imaging 2022,
32, 011003. [CrossRef]
[17]. Huang, Y.; Llach, J. Tracking the Small Object through Clutter with Adaptive Particle Filter. In Proceedings of the ICALIP 2008—2008 International Conference on Audio, Language and Image Processing, Shanghai, China, 7–9 July 2008.
[18]. Habibi, Y.; Sulistyaningrum, D.R.; Setiyono, B. A New Algorithm for Small Object Tracking Based on Super-Resolution Technique. In Proceedings of the AIP Conference Proceedings; AIP Publishing: Long Island, NY, USA, 2017; Volume 1867.
[19]. Liu, W.; Tang, X.; Ren, X. A Novel Method for Small Object Tracking Based on Super-Resolution Convolutional Neural Network. In Proceedings of the 2019 2nd International Conference on Information Systems and Computer Aided Education, ICISCAE 2019, Dalian, China, 28–30 September 2019.
[20]. Mahmoodi, J.; Nezamabadi-pour, H.; Abbasi-Moghadam, D. Violence Detection in Videos Using Interest Frame Extraction and 3D Convolutional Neural Network. Multimed. Tools Appl. 2022, 81, 20945–20961. [CrossRef]
[21]. Wu, D.; Song, H.; Yuan, H.; Fan, C. A Small Object Tracking Method in Satellite Videos Based on Improved Kernel Correlation
Filter. In Proceedings of the 2022 14th International Conference on Communication Software and Networks, ICCSN 2022, Chongqing, China, 10–12 June 2022.
[22]. Blostein, S.D.; Huang, T.S. Detecting Small, Moving Objects in Image Sequences Using Sequential Hypothesis Testing. IEEE Trans. Signal Process. 1991, 39, 1611–1629. [CrossRef]
[23]. Ahmadi, K.; Salari, E. Small Dim Object Tracking Using a Multi Objective Particle Swarm Optimisation Technique. IET Image Process. 2015, 9, 820–826. [CrossRef]
[24]. Salari, E.; Li, M. Dim Target Tracking with Total Variation and Genetic Algorithm. In Proceedings of the IEEE International
Conference on Electro Information Technology, Milwaukee, WI, USA, 5–7 June 2014.
[25]. Shaik, J.; Iftekharuddin, K.M. Detection and Tracking of Targets in Infrared Images Using Bayesian Techniques. Opt. Laser Technol. 2009, 41, 832–842. [CrossRef]
[26]. Archana, M.; Geetha, M.K. Object Detection and Tracking Based on Trajectory in Broadcast Tennis Video. Procedia Comput. Sci. 2015, 58, 225–232. [CrossRef]
[27]. Zhang, R.; Ding, J. Object Tracking and Detecting Based on Adaptive Background Subtraction. Procedia Eng. 2012, 29, 1351–1355 [CrossRef]
[28]. Srivastav, N.; Agrwal, S.L.; Gupta, S.K.; Srivastava, S.R.; Chacko, B.; Sharma, H. Hybrid Object Detection Using Improved Three Frame Differencing and Background Subtraction. In Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering, Noida, India, 12–13 January 2017.
[29]. Zhu, M.; Wang, H. Fast Detection of Moving Object Based on Improved Frame-Difference Method. In Proceedings of the 2017 6th International Conference on Computer Science and Network Technology, ICCSNT, Dalian, China, 21–22 October 2017; Volume 2018, pp. 299–303. [CrossRef]
[30]. Yin, Q.; Hu, Q.; Liu, H.; Zhang, F.; Wang, Y.; Lin, Z.; An, W.; Guo, Y. Detecting and Tracking Small and Dense Moving Objects in Satellite Videos: A Benchmark. IEEE Trans. Geosci. Remote Sens. 2022, 60, 1–18. [CrossRef]
[31]. Zhou, Y.; Maskell, S. Detecting and Tracking Small Moving Objects in Wide Area Motion Imagery (WAMI) Using Convolutional Neural Networks (CNNs). In Proceedings of the FUSION 2019 22nd International Conference on Information Fusion, Ottawa, ON, Canada, 2–5 July 2019.
[32]. Teutsch,M.; Grinberg,M. Robust Detection ofMoving Vehicles inWide AreaMotion Imagery. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern RecognitionWorkshops, Las Vegas, NV, USA, 26 June 2016–1 July 2016.
[33]. Aguilar, C.; Ortner, M.; Zerubia, J. Small Moving Target MOT Tracking with GM-PHD Filter and Attention-Based CNN. In

Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing, MLSP, Gold Coast, Australia, 25–28
October 2021; pp. 1–6. [CrossRef]
[34]. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks. Adv. Neural Inf. Process. Syst. 2017, 39, 1137–1149. [CrossRef] [PubMed]
[35]. Aguilar, C.; Ortner, M.; Zerubia, J. Small Object Detection and Tracking in Satellite Videos With Motion Informed-CNN and
GM-PHD Filter. Front. Signal Process. 2022, 2, 827160. [CrossRef]
[36]. Lyu, C.; Liu, Y.; Li, B.; Chen, H. Multi-Feature Based High-Speed Ball Shape Target Tracking. In Proceedings of the 2015 IEEE International Conference on Information and Automation, ICIA 2015—In conjunction with 2015 IEEE International Conference on Automation and Logistics, Lijiang, China, 8–10 August 2015.
[37]. Hongshan, N.; Zhijian, H.; Jietao, D.; Jing, C.; Haijun, L.; Qiang, L. AWiener Filter Based Infrared Small Target Detecting and Tracking Method. In Proceedings of the 2010 International Conference on Intelligent System Design and Engineering Application, ISDEA 2010, Changsha, China, 13–14 October 2010; Volume 1.
[38]. Deshpande, S.D.; Er, M.H.; Venkateswarlu, R.; Chan, P. Max-Mean and Max-Median Filters for Detection of Small Targets. In Signal and Data Processing of Small Targets 1999; SPIE: Cergy, Germany, 1999; Volume 3809, pp. 74–83.
[39]. Ahmadi, K.; Salari, E. Small Dim Object Tracking Using Frequency and Spatial Domain Information. Pattern Recognit. 2016,
58, 227–234. [CrossRef]
[40]. Dong, X.; Huang, X.; Zheng, Y.; Shen, L.; Bai, S. Infrared Dim and Small Target Detecting and Tracking Method Inspired by
Human Visual System. Infrared Phys. Technol. 2014, 62, 100–109. [CrossRef]
[41]. Dong, X.; Huang, X.; Zheng, Y.; Bai, S.; Xu, W. A Novel Infrared Small Moving Target Detection Method Based on Tracking
Interest Points under Complicated Background. Infrared Phys. Technol. 2014, 65, 36–42. [CrossRef]
[42]. Zhang, F.; Li, C.; Shi, L. Detecting and Tracking Dim Moving Point Target in IR Image Sequence. Infrared Phys. Technol. 2005, 46, 323–328. [CrossRef]
[43]. Shaik, J.S.; Iftekharuddin, K.M. Automated Tracking and Classification of Infrared Images. In Proceedings of the International
Joint Conference on Neural Networks, Portland, OR, USA, 20–24 July 2003; Volume 2.
[44]. Liu, C.; Ding,W.; Yang, J.; Murino, V.; Zhang, B.; Han, J.; Guo, G. Aggregation Signature for Small Object Tracking. IEEE Trans. Image Process. 2020, 29, 1738–1747. [CrossRef] [PubMed]
[45]. Tzannes, A.P.; Brooks, D.H. Temporal Filters for Point Target Detection in IR Imagery. In Proceedings of the Infrared Technology and Applications XXIII, Orlando, FL, USA, 20–25 April 1997; Volume 3061.