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

Enhancing Vision Systems: A Comprehensive Review of Small Object Detection Using Artificial Intelligence

Abhinav, Dr. C.L.P. Gupta
Department of Computer Science and Engineering,
Bansal Institute of Engineering & Technology, Lucknow – India
yabhinav048@gmail.com, clpgupta@gmail.com

Abstract: Small object detection remains a challenging task in computer vision due to the limited pixel information, poor feature representation, and occlusion in cluttered backgrounds. These constraints reduce detection accuracy and reliability, especially in applications such as autonomous driving, surveillance, and remote sensing. Artificial Intelligence (AI), particularly deep learning, has brought remarkable progress to object detection; however, the detection of small objects still lags behind. This review presents a comprehensive overview of AI-based techniques for small object detection, emphasizing the limitations of conventional models, advancements in feature extraction and network architecture, and the role of data augmentation, resolution enhancement, and attention mechanisms. The paper also highlights current challenges, benchmark datasets, and suggests future research directions to improve the performance and robustness of small object detection systems.

Keywords: Small object detection, Artificial intelligence, Deep learning, Convolutional neural networks, Object recognition, Computer vision, Image processing, Feature extraction, YOLO, SSD, Transformer models

1. Introduction

Object detection is a core problem in computer vision with extensive applications ranging from traffic monitoring and autonomous vehicles to medical imaging and aerial surveillance. While the detection of large and medium-sized objects has achieved high accuracy with the advent of deep learning techniques, small object detection remains a persistent challenge. Small objects, typically defined as those occupying less than 1% of the image area, often suffer from inadequate feature representation, scale variation, and background noise interference, leading to poor localization and classification performance.

Recent advancements in Artificial Intelligence (AI), particularly deep learning, have provided powerful tools for learning hierarchical feature representations. However, the architectural design of many state-of-the-art detection models, such as Faster R-CNN, YOLO, and SSD, tends to prioritize global semantics and downsample input data, making them less effective for detecting small targets. This review aims to explore AI-driven solutions specifically tailored to the small object detection problem, evaluate their performance, and identify current limitations and opportunities for improvement.

Object detection is a central task in the field of computer vision, enabling machines to identify and localize objects within images or videos. It plays a pivotal role in a wide range of real-world applications such as autonomous driving, aerial surveillance, industrial inspection, healthcare diagnostics, and security systems. While significant advancements have been made in detecting medium and large-scale objects through deep learning techniques, the accurate detection of small objects continues to present substantial challenges. Small objects, typically defined as those occupying a minimal portion of the image—often less than 1% of the total pixel area—are inherently difficult to detect due to limited visual information, low resolution, and high susceptibility to occlusion and background clutter.

Traditional object detection algorithms, including region-based and one-stage detectors like Faster R-CNN, YOLO, and SSD, often rely on high-level semantic features and downsampled representations. These architectural designs, although effective for larger objects, tend to lose critical fine-grained details that are essential for recognizing small-scale targets. The reduced spatial resolution in deeper network layers exacerbates the difficulty of preserving discriminative features necessary for detecting small objects. As a result, small object detection suffers from lower accuracy, higher false-positive rates, and increased localization errors compared to the detection of larger counterparts.

The emergence of Artificial Intelligence (AI), particularly deep learning and convolutional neural networks (CNNs), has revolutionized object detection by enabling automatic feature extraction and learning from large datasets. However, detecting small objects remains a bottleneck, prompting researchers to explore innovative strategies. These include multi-scale feature fusion, context-aware modules, super-resolution techniques, attention mechanisms, and transformer-based models. Moreover, data-centric approaches such as advanced data augmentation, synthetic dataset generation, and transfer learning are increasingly being adopted to enhance the robustness and generalizability of small object detectors.

Given the critical role that small object detection plays in mission-critical applications—such as detecting pedestrians at a distance in autonomous driving, identifying micro-lesions in medical imaging, or recognizing small vehicles in aerial footage—there is a growing need to develop specialized AI-based models that can overcome these challenges. This paper provides a comprehensive review of current techniques and trends in small object detection using artificial intelligence. It analyzes the limitations of existing methods, evaluates

promising architectures, discusses benchmark datasets and evaluation metrics, and outlines future research directions aimed at improving detection accuracy and system efficiency.

2. Challenges in Small Object Detection

Small object detection presents unique challenges that are distinct from those faced in detecting larger objects. The primary issues include:

- Low Resolution and Fewer Pixels: Small objects have limited pixel coverage, making it harder to extract meaningful features.
- Loss of Spatial Information: Downsampling layers in convolutional networks often eliminate crucial details of small objects.
- High Background Clutter: Small targets are often indistinguishable from complex backgrounds, leading to false positives.
- Scale Variation: Objects of interest may appear in vastly different scales depending on distance or camera angle.
- Occlusion and Overlap: In many real-world scenarios, small objects are partially hidden or overlap with others, complicating their identification.

3. Artificial Intelligence Techniques for Small Object Detection

3.1 Convolutional Neural Networks (CNNs)

CNNs have been the backbone of object detection, enabling the automatic learning of spatial hierarchies. However, typical architectures downsample images to reduce computational complexity, which disproportionately affects small objects. Recent improvements include:

- Multi-scale feature fusion (e.g., Feature Pyramid Networks) to preserve fine-grained information.
- Context-aware modules to capture surrounding regions that provide cues for object presence.
- Dilated convolutions to expand the receptive field without reducing resolution.

3.2 Region Proposal Networks (RPN)

Models like Faster R-CNN generate object proposals using anchor boxes of various scales. For small objects, introducing smaller anchors and fine-tuning anchor ratios improves proposal generation. However, performance remains limited when objects are densely packed.

3.3 One-Stage Detectors

YOLO and SSD models have been optimized for real-time performance but often underperform on small objects. Enhancements such as YOLOv5's PANet and YOLOv7's re-parameterization modules aim to improve feature localization for small objects. SSD variants also incorporate additional low-level feature maps for better resolution.

3.4 Transformer-Based Models

Transformer architectures (e.g., DETR and its variants) are emerging alternatives in object detection. While powerful in modeling global context, their performance on small objects remains under investigation. Hybrid CNN-transformer models offer promising results by combining local detail with global attention.

4. Data-Centric Approaches

In addition to architectural advancements, data-centric techniques play a critical role in enhancing small object detection:

- Super-resolution techniques enhance the input resolution to improve feature richness.
- Data augmentation strategies such as CutMix, Mosaic, and Copy-Paste artificially increase the presence of small objects in training data.
- Synthetic datasets and adversarial training can supplement training with diverse examples of small objects.

5. Benchmark Datasets and Evaluation Metrics

Evaluation of small object detection models requires datasets that include sufficient small-object instances. Common datasets include:

MS COCO: Contains a large number of small object instances, with AP_{small} as a specific metric.

DOTA and xView: Used for aerial and satellite image analysis with many small objects.

VisDrone: Focused on small object detection in drone imagery.

Open Images: Offers a diverse set of images with objects of varying scales.

Evaluation metrics typically include Average Precision (AP) at different IoU thresholds and AP_{small} for isolating small-object performance.

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. Limitations and Research Gaps

Despite progress, several limitations remain:

- Trade-off between accuracy and speed: Enhancing detection resolution increases computational cost.
- Generalization to unseen data: Models trained on specific datasets may struggle with domain shift.
- Insufficient labeled data: Small objects are often underrepresented in training datasets.
- Model interpretability: Lack of transparency in AI models makes it hard to explain false detections or failures.

8. Future Directions

Future research should focus on:

- Designing lightweight models optimized for detecting small objects on edge devices.
- Developing cross-domain transfer techniques to handle varied imaging conditions.
- Leveraging multimodal data (e.g., infrared and radar) to improve detection in low-visibility environments.
- Incorporating self-supervised learning to exploit unlabeled data for better generalization.

9. Conclusion

Small object detection is a critical yet challenging task within computer vision, where standard AI models often fall short. This review highlights the importance of designing specialized architectures and training techniques that cater to the unique constraints of small object scenarios. Through a combination of high-resolution processing, enhanced feature extraction, and intelligent learning frameworks, AI holds the potential to significantly improve the reliability and accuracy of small object detection across diverse real-world applications. Continued research in this domain will not only benefit autonomous systems and surveillance but also pave the way for smarter, safer AI-powered vision systems.

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.
[17] Raghawend, Anurag, "Detect Skin Defects by Modern Image Segmentation Approach, Volume 20, Issue 1, 2020
[18]. Pimentel, M.A.; Clifton, D.A.; Clifton, L.; Tarassenko, L. A review of novelty detection. Signal Process. 2014, 99, 215–249.
[18]. Xu, Y., Y. Guo, L. Xia, and Y. Wu, "An support vector regression based nonlinear modeling method for Sic mesfet," Progress In Electromagnetics Research Letters, Vol. 2, 103–114, 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]