Iran Journal of Computer Science (2026) 9:9
https://doi.org/10.1007/s42044-025-00384-z

RESEARCH

A systematic image augmentation framework based on dataset limitations for small object detection

Ume Nisa[1] · Muhammad Syafiq Mohd Pozi[1] · Mohamed Ali Saip[1]

Received: 19 June 2025 / Accepted: 13 October 2025
© The Author(s) 2025

Abstract
Small object detection in aerial images remains a significant challenge due to low-resolution instances, sparse object distribution, and the lack of large, diverse datasets. These limitations reduce the accuracy and generalization ability of modern object detection algorithms. To address this, we first propose a three-tier systematic framework for optimizing data augmentation strategies, which includes (1) dataset analysis, (2) definition of augmentation objectives, and (3) systematic selection of techniques based on identified challenges. Building upon this framework, we propose a data augmentation pipeline designed to enhance small object visibility, improve dataset balance, and increase model robustness. The proposed pipeline consists of two primary stages. The first stage performs image segmentation and cropping, dividing images into smaller segments and applying localized zoom-in and center-cropping techniques. This stage selectively retains only segments with high object density, discarding empty or irrelevant regions, thereby increasing the object-to-image ratio. The second stage involves object-centric image resizing and augmentation, which standardizes input dimensions while maintaining enhanced object clarity and scale, ensuring compatibility with convolutional neural networks (CNNs). The effectiveness of the pipeline is validated using YOLOv8m on two datasets AI-TODv2 and SODA-A. The results demonstrate substantial improvements in performance metrics, including a +15.9% AP@0.5 on AI-TODv2 and a 23.6% increase in SODA-A using YOLOv8m. But only the class 'ship' in AI-TODv2 showed a decline. Notably, in both datasets, the average precision (AP) of the classes has improved, highlighting the effectiveness of the proposed pipeline in enhancing the detection of hard-to-detect classes. These results confirm that our augmentation strategy significantly improves detection accuracy and recall for small objects in aerial scenes, overcoming the limitations of sparse, dense, and imbalanced datasets.

Keywords Small object detection · Aerial images · Augmentation approaches · AI-TOD · YOLOv8

1 Introduction

Image understanding is a core issue in computer vision that can be dealt with classification, recognition, localizing, and, more importantly, detection of objects [1–3]. The term "object detection" refers to an activity that typically consists of several subtasks, like 'facial detection', 'pedestrian detection', and 'small object detection' [4]. Small object detection has significant importance due to its real-world application [5] including autonomous driving [6, 7], surveillance [8], unmanned aerial vehicles [9] and satellite image analysis [10, 11]. Due to its practical application, it is crucial to detect objects quickly and efficiently for effective decision-making. Even though there have been significant advancements in generic object detection, small object detection research, especially in aerial images, has progressed somewhat slowly. More specifically, even for the most advanced detectors, there is still a significant performance gap between the detection of small- and normal-sized objects specifically in aerial images due to many reasons, like incomplete information, occluded objects, lack of authentic definition and the lack of large-scale datasets for the detection of small objects [12–15]. The detection of small objects becomes more challenging in the aerial images dataset [16].

✉ Ume Nisa
kazmi_nisa@hotmail.com; nisa_ume@ahsgs.uum.edu.my

✉ Muhammad Syafiq Mohd Pozi
syafiq.pozi@uum.edu.my

Mohamed Ali Saip
mdali@uum.edu.my

[1] School of Computing, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Published online: 02 January 2026

We have noticed that objects in aerial imagery often exhibit distinct spatial distributions, either sparsely located or densely clustered within specific regions, while large portions of the images frequently contain few or no objects, as can be seen in Fig. 1. This Fig. 1 shows that the dataset has the challenges of scattered and sparse objects; some images contain only a single object or vast areas without any objects, often causing the training model to miss these objects. This uneven distribution poses challenges even for advanced detection models, often leading to misdetections or inaccuracies. Furthermore, publicly available aerial datasets lack the scale and diversity required to effectively train advanced object detection algorithms, limiting their potential for robust performance and generalization. Researchers need large datasets specifically for small-scale object detection to evaluate the performance of object detection algorithms [17–19]. Image augmentation approaches are one of the solutions to deal with data scarcity and dataset quality, but researchers are also believed to pay less attention to generating data augmentation approaches for small object detection compared to other computer vision issues due to their complex structure and the absence of a specific framework on the optimal number or combination of augmentation approaches to apply [20]. Researchers often rely on various traditional augmentation techniques as part of a preprocessing pipeline to improve dataset quality, typically through a trial-and-error approach. However, there is a gap in the literature regarding the understanding of why specific traditional image augmentation methods are chosen. The reasoning behind these choices is often unclear and needs to be explicitly addressed. However, the lack of a standardized framework or best practices for selecting and implementing augmentation approaches creates inconsistency and hinders reproducibility in studies. Therefore, to mitigate this, there must be a well-defined framework for utilizing traditional image augmentation approaches.

Image augmentation approaches are primarily dataset-agnostic and applied without explicit linkage to dataset-specific limitations. Consequently, two major issues persist:

- **Reproducibility:** Without a structured guideline, researche rs replicate prior augmentation strategies inconsistently.
- **Optimization of augmentation choices:** The absence of a principled selection process often results in the use of redundant or suboptimal transformations, wasting computational resources.

1.1 Contribution

To address these challenges, this paper aims to propose a framework for researchers to optimize the use of traditional image augmentation approaches. The framework is based on the common limitations present in the aerial datasets, enabling more informed augmentation choices. The framework introduces a systematic three-tier process that explicitly links dataset limitations to augmentation objectives and technique selection. Additionally, it provides a systematic augmentation pipeline developed for the proposed framework, specifically designed to enhance the detection performance of small objects. The main contributions described in this paper are as follows:

- Provides a framework for researchers to optimize the use of data augmentation approaches, eliminating the reliance on trial-and-error methods. The dataset's limitations are used as a baseline to systematically select the most effective traditional augmentation approaches.
- A data augmentation pipeline is proposed to validate the effectiveness of the framework, demonstrating its applicability in addressing the challenges associated with small object detection dataset and enhancing model performance.
- Extensive experimentation is conducted on the AI-TODv2 [21] and SODA-A [19] image dataset, where the proposed approach demonstrated improved performance. Experiments conducted with YOLOv8m model demonstrate substantial gains in recall and AP@0.50.

1.2 Paper structure

The subsequent sections of this paper are arranged as outlined below. Section 2 reviews previous work, highlighting dataset limitations, and describing the augmentation approaches proposed to address these challenges in the domain of small object detection. Section 3 outlines the paper's methodology. A detailed three-tier systematic framework is presented, along with a data augmentation pipeline designed on the basis of this framework. Section 4 presents the experimental methodology and results. The section details the datasets, evaluation protocols, and implementation environment, followed by performance evaluations, ablation studies, and visual analyses demonstrating the effectiveness of the proposed augmentation pipeline for small object detection. Section 5 summarizes the study, highlighting the proposed three-tier framework and the augmentation pipeline for small object detection (SOD). It reports improved performance on AI-TODv2 and SODA-A, while acknowledging limitations such as hardware constraints and challenging low-resolution images. Finally, it outlines future directions, including incremental learning and strategies to enhance generalization.

2 Previous work

Small object detection (SOD) is inherently challenging due to the limited visual information present in low-resolution



Fig. 1 AI-TODv2 highlights the challenges of scattered and sparse objects, with some images containing only a single object or vast areas without any objects

objects. The definition of 'small' varies across studies, but in the context of images, it typically refers to the number of pixels that an object occupies [22]. According to [23], a vehicle of 20×12 pixels is hard to detect because of its small size in an image. Tiny objects have less than 16×16 pixels in size [24]. Instances that occupy less than 50 pixels in height and width are considered small objects [25, 26]. Compared to larger objects, small objects have a low pixel count, are not interpretable, and easily fuse with their backgrounds [12]. Although various definitions of 'small objects' exist, their effective detection hinges not only on advanced algorithms but also on the availability of datasets specifically designed for this purpose.

Despite the central role of datasets in the advancement of SOD, most widely adopted benchmarks exhibit structural limitations. The COCO dataset [27], though widely used in object detection research, is not tailored for small object detection, as it predominantly features larger objects, limiting its suitability for addressing specific challenges for smaller object instances. The DOTA dataset [3] exhibits an uneven distribution of small objects, highlighting a lack of balanced representation across object scales. The UAVDT dataset [28], with 27.5% small object instances, demonstrates limited applicability to broader aerial scenarios. Similarly, the DIOR dataset [26] faces challenges due to its long-tail distributions and clustered object classes, while the VEDAI dataset [29] lacks diversity in object scales and environmental variations. Although the VisDrone dataset series [30–33] incorporates varying densities and environmental conditions, it remains predominantly drone-centric, limiting its generalizability to other aerial contexts. This study [34] introduces a dataset for aerial person detection in search and rescue tasks. It discusses the key challenges of aerial datasets, such as sparsity, scale variation, and visibility in adverse conditions. More recently, AI-TOD [21] and SODA-A [19] were introduced to specifically address the scarcity of tiny objects. However, even these datasets face sparsity, clustering, and background imbalance issues. The challenges associated with datasets are common and remain a significant barrier to achieving high accuracy in small object detection. For example, Zhang et al. [35] addresses the problem of balancing accuracy and speed for person detection. Furthermore, the creation of new specialized datasets is often prohibitively expensive and time-consuming, making it impractical for many researchers.

In response to these dataset constraints, data augmentation approaches have been extensively employed as a cost-effective solution to address dataset limitations, enhance diversity, and improve detection accuracy. Numerous prior studies have focused on improving the detection performance of small objects through the application of various augmentation approaches. Traditional augmentation approaches and synthetic transformations have been widely utilized to enhance dataset diversity and improve detection performance. For instance, Junos et al. [36] employed brightness transformation, rotation, blurring, and motion blur to synthetically expand a palm oil fruit dataset, increasing the dataset size from 700 to 6300 images. Similarly, Psiroukis et al. [37] applied augmentation techniques to georeferenced ortho-mosaic images, leveraging transformations to classify broccoli heads based on maturity levels. Additionally, Wang [38] utilized scaling, contrast enhancement, flipping, and rotation as preprocessing techniques to address challenges in small object detection, demonstrating significant improvements in accuracy by enriching the dataset and overcoming feature extraction limitations. Zhu et al. [39] combined Mosaic, MixUp, and conventional techniques (cropping, translation, rotation) on TPH-YOLOv5 for the VisDrone2021 dataset. These approaches underscore the effectiveness of traditional and synthetic transformations in addressing dataset scarcity and improving model performance. These methods demonstrate consistent performance gains but remain largely generic and lacking dataset-specific adaptation.

In another paper, the SeaDronesSee Object Detection v2 dataset [40] has the challenges of limited data availability and imbalanced datasets, with a focus on out-of-scale instances. To address these issues, the system generates new inputs considering the dataset imbalance and utilizing existing instances. New images are created by combining instances from different training set images, excluding the majority class instances to avoid further imbalance. Randomly selected minority class instances with matching perspectives are incorporated into the new image. Measures are taken to prevent overlapping instances and position them below the highest instance in the original image, ensuring a realistic image [41]. Cropping and scaling strategies have been effectively used to address challenges related to dataset limitations and overfitting. Deng et al. [42] and Zhao et al. [43] focus on dividing images to address dataset limitations, with [43] integrating test-time augmentation for added performance gains. In contrast, Gao et al. [44] used random cropping and expansion to enhance dataset diversity, achieving notable detection accuracy. Although all approaches tackle dataset constraints, the method [43] emphasizes multiscale testing, while [42, 44] focus primarily on reducing overfitting and increasing instance variety.

Beyond image-level transformations, instance-level augmentation techniques have emerged to address dataset imbalance and small object sparsity more directly. Zhan et al. [33] introduced the ECopy-Paste method, which mitigates issues such as limited annotated data, brightness imbalance, and small target representation in the VisDrone2019-DET dataset, achieving improved detection performance. Kisantal et al. [5] similarly utilized copy-paste augmentation, where tiny objects from segmented masks were duplicated and modified before being randomly placed within the same image, enhancing training data diversity in the MS-COCO dataset. Expanding on these concepts, Chen et al. [45] and Ye et al. [46] proposed the Stitcher technique, which focuses on stitching image parts with a higher density of small objects to address dataset imbalance effectively. These approaches offer targeted augmentation by increasing small object visibility and addressing class imbalance without extensive new data collection.

More sophisticated compositional methods, such as Mosa ic and adaptive cropping, offer additional improvements by synthesizing new scenes from multiple source images. Bochkovskiy et al. [47] pioneered Mosaic and Self-Adversarial Training (SAT), creating composite images and altering features to improve model generalization. Wei et al. [48] extended these ideas by integrating Adaptive Cropping, Mosaic, and Mask Resampling to tackle scale variation and sparsity in aerial datasets. Duan et al. [49] further refined Mosaic, focusing on detecting and merging cluster regions for improved small object detection, while [50] combined Mosaic with Geometric Transformations to enrich datasets for defective chip detection. Collectively, these approaches demonstrate the adaptability of Mosaic techniques across diverse datasets, with each method addressing specific challenges such as sparsity, clustering, and object scale variation. VTSaRNet [51] introduces a transformer-based RGB-T framework with instance-level copy–paste and mosaic augmentation for aerial person detection in complex SaR scenarios, achieving high precision and recall. However, it relies on multimodal data and heavy architectures.

The studies mentioned above have explored data augmentation to improve performance of detectors, highlighting performance gains but lacking guidance on selecting optimal augmentation approaches under data scarce conditions. This study [52] proposes a framework to guide the selection of appropriate image augmentation approaches for quality inspection tasks in data-scarce scenarios, demonstrating significant improvements in detection accuracy and loss reduction using pixel-level transformations. But still [52] uses the error and trial method to figure out the optimal selection of approaches.

While previous studies have explored various augmentation approaches to address the challenges of small object detection, they lack systematic frameworks that explicitly

connect dataset limitations to augmentation choices. This paper builds on previous efforts, addressing this gap by introducing a systematic augmentation framework specifically designed for small object detection in aerial image dataset. By proposing a framework and a pipeline for data augmentation, our approach seeks to optimize dataset preparation and improve detection accuracy, setting a foundation for more robust and reproducible research in this field.

3 Methodology

3.1 Framework

To address the challenges of small object detection, this paper presents a systematic framework for researchers to optimize data augmentation approaches, eliminating the reliance on trial-and-error methods shown in Fig. 3. Aerial datasets' limitations are used systematically as a baseline to ensure that only the most effective techniques from traditional augmentation approaches are applied. The following are the steps:

1. Analyze dataset limitations
   A thorough analysis of the dataset is essential to understand its inherent limitations. Key characteristics such as object size, object-to-image ratio, object density, class distribution, etc. should be evaluated. Common challenges, such as sparse or densely clustered objects, noisy labels, missing annotations, and background clutter, must be identified. These limitations serve as the foundation for selecting augmentation approaches that can address specific weaknesses in the dataset. Dataset limitations are the main cause of the performance degradation. Table 1 shows the common limitations of aerial datasets that effects small object detection accuracy. While, Table 2 highlights how these common problems are distributed across major aerial datasets, demonstrating that most datasets share overlapping challenges such as low resolution, sparsity, scale imbalance, and background clutter. Moreover, an analysis of dataset SODA-A shown in Fig. 2 indicates that the dataset is characterized by small object sizes and high object density.
2. Define Augmentation Objectives
   Clearly defining the objectives of data augmentation is crucial for achieving meaningful improvements. The augmentation objectives should address the challenges identified during the dataset analysis and align with the goals of small object detection. Objectives may include enhancing the visibility and prominence of small objects, increasing dataset diversity by introducing variations in object orientation, lighting, and occlusion, balancing object-to-image ratios by addressing sparse or dense object regions.
3. Systematic Selection of Augmentation Approaches
   The selection of augmentation approaches must be systematic and based on the dataset's limitations and defined objectives. Techniques should be chosen to directly address identified challenges, avoiding redundancy or unnecessary complexity. For instance, methods that improve object visibility or balance the dataset's object-to-image ratio are prioritized. Each selected technique should be evaluated for its scalability and potential to enhance the dataset's utility.

Based on this framework, an augmentation pipeline is proposed, with each stage putting the framework steps into practice as described in the following Sect. 3.2.

3.2 Proposed pipeline for data augmentation for small object detection

This section presents the proposed framework developed based on a three-tier framework comprising: (1) dataset analysis, (2) augmentation objective definition, and (3) systematic selection of augmentation approaches. According to the above framework, the steps for dataset analysis, the study selected the AI-TOD and SODA-A datasets. The AI-TOD dataset is a dataset that is specifically generated for small or tiny object detection, consisting of a total of 28,036 images. The average size of the instances in the images is 12.8 pixels. They are randomly distributed in the aerial images. But most areas of the images have empty spaces where there are either few objects or no objects due to the scattered distribution of objects in the dataset, resulting in a high percentage of empty spaces in the images. This characteristic of the dataset could present challenges for detection algorithms as they may miss certain target objects, particularly in areas with low object density. The high percentage of empty space in the images may also lead to detection algorithms being biased towards detecting objects in denser areas and overlooking objects in sparser areas. This is a critical consideration when designing models for object detection using this dataset. Another notable feature of the AI-TOD dataset is that it is a small-scale dataset that may restrict the accuracy and performance of models. SOTA models need more and more data for training purposes to perform well, as they are significantly rely on the capacity to generalize from a wide range of instances. As such, it may be necessary to supplement the dataset with additional images to ensure optimal performance of algorithms in object detection tasks.

To address the identified limitations of the AI-TOD dataset, the following augmentation objectives have been defined to guide the design of the proposed framework. First, the framework aims to enhance the visibility of small objects by increasing their mean size, thereby making them more detectable by object detection models. Second, it seeks to bal-



Fig. 2 Object size distribution in SODA-A—small and dense objects

Table 1 Common limitations of aerial datasets affecting small object detection

Limitation	Description	Underlying causes	Impact on detection performance
Low object resolution [53]	Small objects occupy very few pixels, often $<16 \times 16$ in size	High-altitude imagery, sensor resolution limitations	Loss of semantic detail; poor feature extraction
Sparse object distribution [54]	Objects are scattered with large background areas	Wide-area aerial views; infrequent target appearances	Underfitting; poor generalization due to insufficient positive samples
Scale imbalance [55]	Disproportionate number of large versus small objects	Unbalanced annotation; natural object diversity	Model bias toward dominant object sizes; degraded small object recall
Background clutter [56]	Complex textures and structures around target objects	Urban environments, vegetation, road markings	Object-background confusion; false positives or missed detections
Background occlusion [57]	Small objects are partially or fully blocked by surrounding structures or objects	Overhead view limitations; dense urban environments; natural occluders (trees, buildings)	Incomplete feature visibility; missed detections or false negatives
Annotation noise or gaps [58]	Incomplete, inconsistent, or inaccurate object annotations	Manual labeling errors; small object visibility challenges	Impaired supervision; model confusion during training
Clustered small objects [59]	High-density object groups in limited spatial regions	Events, traffic scenes, or object congregation scenarios	Overlapping detections; NMS suppression issues
Environmental variability [60]	Illumination changes, fog, haze, and shadows across images	Different times of day, seasons, or weather conditions	Domain shift; decreased robustness under unseen conditions

ance the object-to-image ratios by reducing the influence of empty or sparsely populated regions and focusing on object-dense segments of the images, thus minimizing bias and improving model performance. Third, the framework strives to increase dataset diversity by expanding the dataset size and introducing variations in the orientation, resolution, and content of objects, enhancing the generalizability of the detection models. Finally, it aims to improve dataset utility by standardizing image dimensions while preserving enhancements in object visibility, ensuring compatibility with Convolutional Neural Networks (CNNs). These objectives are carefully designed to directly address the dataset's challenges and align with the overarching goals of small object detection research.

The proposed framework systematically uses augmentation approaches that align with the defined objectives, ensuring targeted and efficient enhancements to the dataset. The framework comprises two main stages.

3.2.1 Image segmentation and cropping

The first stage of the proposed framework focuses on image segmentation and cropping. It is designed to increase dataset

Table 2 Distribution of common dataset problems across benchmark aerial datasets

Dataset	Low resolution	Sparse dist.	Clustered dist.	Class imbalance	Background clutter	Small dataset	Annotation noise	Small object size
COCO [27]	✓	×	×	✓	✓	×	×	×
DOTA [3]	✓	✓	✓	✓	✓	✓	×	✓
UAVDT [28]	✓	✓	✓	×	✓	✓	×	✓
DIOR [26]	✓	✓	✓	✓	✓	×	×	✓
VEDAI [29]	✓	✓	×	✓	×	×	×	✓
VisDrone [30–33]	✓	✓	✓	✓	✓	✓	×	✓
AI-TODv2 [21]	✓	✓	✓	✓	✓	✓	✓	✓
SODA-A [19]	✓	✓	✓	✓	✓	✓	×	✓

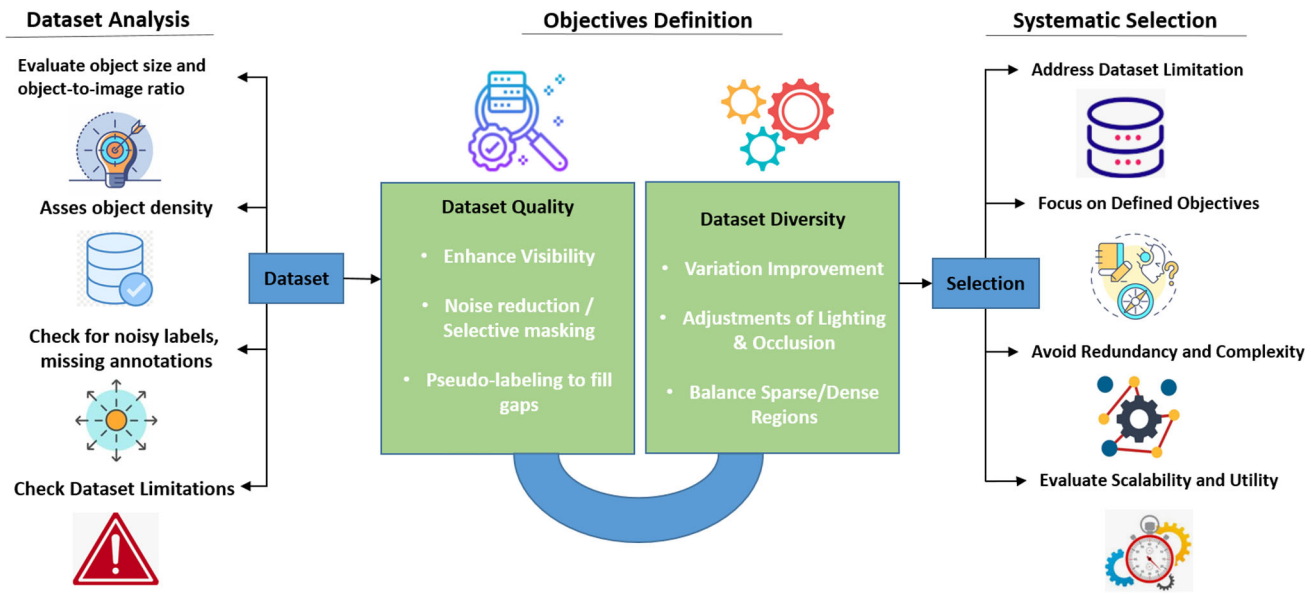


Fig. 3 Three-tier framework for selecting augmentation approaches: dataset analysis, objective definition, and systematic selection

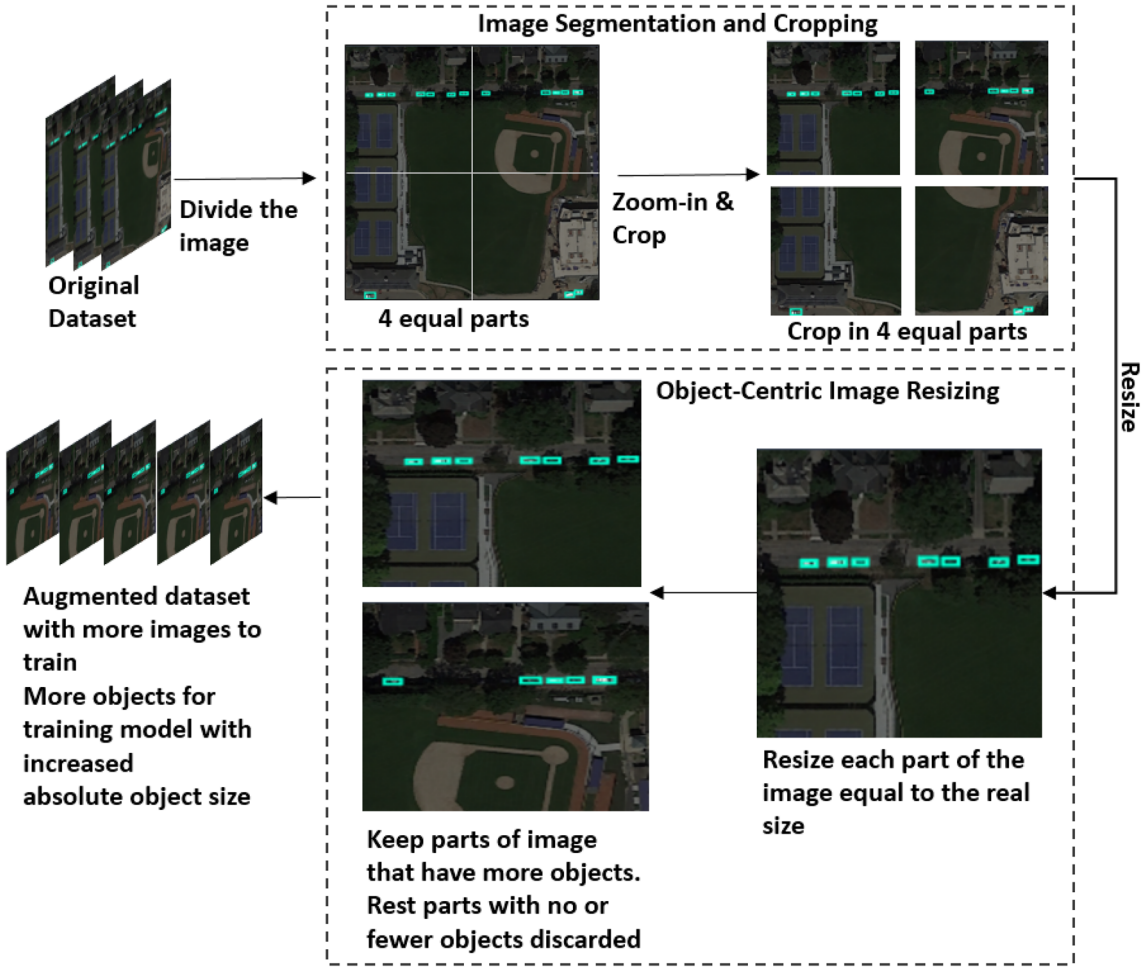


Fig. 4 Proposed data augmentation pipeline for AI-TODv2: (1) Image segmentation and cropping, (2) Object-centric image resizing

size and the prominence of the object. This is achieved by dividing each image into four equal segments based on its dimensions. For an image with dimensions $X \times Y$, this process involves halving the dimensions along both axes. For example, an image of 1000×1000 pixels is divided into four segments of 500×500 pixels each. The segmentation process involves splitting the image horizontally at its midpoint to create two parts of $(X \times Y)/2$ pixels. Further splitting each of these parts vertically at their midpoints yields four parts of $X/2 \times Y/2$ pixels. These smaller segments allow localized application of augmentation approaches, specifically zoom-in and crop augmentation. The zoom-in operation magnifies each segment by 25%, increasing the resolution and enhancing the visibility of smaller objects.

Following this, a center-cropping operation is applied to retain only the most object-dense regions within the magnified segments. Segments containing no objects are discarded. The image was divided into four equal parts to generate additional training samples and to ensure that no region of the original image was omitted. This strategy also increases the object-to-image ratio, allowing small objects to occupy a larger proportion of the cropped images, thus improving their visibility during training. A fixed 25% zoom-in factor was applied to further enlarge small objects while maintaining sufficient contextual information. This is done to ensure that a dataset is enriched with meaningful content. This selective cropping also increases the object-to-image ratio, improving the dataset's suitability for small object detection tasks.

3.2.2 Object-centric image resizing and augmentation

The second stage involves object-centric image resizing and augmentation. This stage standardizes the dataset and emphasizes regions of interest. After segmentation and augmentation, the retained image segments are resized to fixed dimensions compatible with the input requirements of Convolutional Neural Networks (CNNs). This resizing preserves the relative magnification achieved in the zoom-in phase. It also ensures consistent input sizes for model training.

Crucially, only segments with a higher density of small objects are retained for resizing, while those with no objects are excluded. This targeted approach optimizes the dataset by increasing the representation of small objects. As a result, it effectively raising the small object ratio from its current 79% to an anticipated 90%–95%. Furthermore, the resizing operation amplifies the absolute mean size of objects. It is particularly critical for datasets with small object instances. This ensures that the model can better focus on and learn from small objects during training.

Figure 4 serves as a graphical representation of the proposed pipeline. It illustrates the steps involved in generating an augmented dataset. This visual aid offers valuable information about the step-by-step process. It helps readers understand the procedures involved in generating the augmented dataset. The proposed data augmentation pipeline for small object detection is applied to the two datasets: AI-TODv2 and SODA-A. The pipeline consists of two main stages: "Image Segmentation and Cropping" and "Object-Centric Image Resizing". These steps are repeated for each image to generate an augmented or synthetic dataset for better model training.

4 Experiment details

In this section, we address the datasets, evaluation metrics, and implementation details to validate the proposed approach.

4.1 Dataset

The experiments are conducted on AI-TOD and SODA-A, two benchmark datasets for small-object detection in aerial imagery.

4.1.1 AI-TOD

Tiny Object Detection in Aerial Images [12] is specially released for small and tiny object detection consisting of 28,036 high resolution aerial images with 700,621 objects for 8 classes (airplane, bridge, storage tank, ship, swimming pool, vehicle, person and wind mill). The average instance dimension in this dataset is roughly 12.8 pixels, which is significantly less than other available datasets for detection of small and tiny objects in aerial images. Furthermore, this dataset has been relabelled due to noisy labels, missing annotations, severe class imbalance, and location errors [21]. Due to the unavailability of the test set, the results are reported in the validation set.

4.1.2 SODA-A

Small Object Detection dAtasets (SODA)-Aerial [19] consists of 2510 high-resolution aerial images with around 2.5M annotated objects across 9 traffic-related categories (people, rider, bicycle, motor, vehicle, traffic sign, traffic light, traffic camera, and warning cone). suffers from class imbalance (e.g., few helicopter samples), extreme object density/occlusion, and limited classes (9 only). Its orientation annotations can be ambiguous at tiny scales, and medium/large objects are ignored.

4.2 Evaluation metrics

We adopt the COCO [27] evaluation protocol, using Average Precision (AP) and Average Recall (AR) to assess detec-

Table 3 Augmentation pipeline settings

Framework	Operation	Setting	Rationale
Stage 1	Segments	Split each image into a 2×2 grid (tiles of size $X/2 \times Y/2$); stride = tile size; no overlap	Increase dataset size and object size
	Zoom-in	Scale factor $s = 1.25$ about tile center; interpolation = bilinear (OpenCV `INTER_LINEAR`); anti-aliasing = off	Increase small objects viability
	Crop	Crop back to original tile size ($X/2 \times Y/2$) centered; padding mode = `reflect` only if needed	Restore tile dims; avoid border artifacts
	Density metric	$\text{density} = \frac{\#\text{boxes with center inside tile}}{\text{tile area}}$	Dataset-agnostic criterion
	Retention rule	Keep tile if #boxes ≥ 1 objects/MP; otherwise discard. Max kept per image $K = 4$ (keep all that pass)	Remove empty/near-empty tiles
Stage 2	Resize to model input	Letterbox to 640×640 (YOLOv8m default), stride= 32; keep aspect ratio with padding; interpolation = bilinear (OpenCV); border value = 114	Match training pipeline exactly
	Normalization	Scale pixels [0, 255] $\rightarrow$ [0, 1]; no mean/std normalization (YOLO default)	Align with Ultralytics defaults
	Shuffling	Random seed = 42; deterministic tile ordering; ties broken by tile index (row, col)	Exact repeatability

Table 4 Hardware and software configurations used in the experiment

Component	Specification
GPU type	NVIDIA Tesla T4 (16 GB VRAM)
GPU runtime limit	Max 12 h per session
CPU	2 vCPUs
Memory	16 GB RAM
Storage	~70–80 GB
Batch size	4
Python version	Python 3.10
PyTorch	Pre-installed (PyTorch 2.0+)
MMCV	Installed manually
CUDA	Default with Colab
Optimizer	Stochastic gradient descent (SGD)
Learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Training epochs	100
Dataset access	Mounted from Google Drive
Pre-trained weights	Initialized from YOLOv8m

tion performance. The overall AP is the mean of AP scores across IoU thresholds from 0.5 to 0.95 (in 0.05 steps). Simply, AP represents the average precision across multiple IoU thresholds, while overall AR measures maximum recall with a fixed number of detections per image, and AP50 represent AP at IoU thresholds of 0.5. AP@0.5 measures the model's prediction accuracy at an IoU threshold of 0.5, meaning the predicted bounding boxes must overlap with the ground truth by at least 50%.

4.3 Implementation environment

We trained YOLOv8m [61, 62] using Google Colab's GPU, utilizing the AI-TODv2 and SODA-A dataset for evaluation. To address hardware limitations, we employed the AI-TODv2 subset of the dataset for experiments, ensuring manageable training and inference times. The model YOLOv8m is tested in our experiments to validate the proposed framework both with and without augmentation (for baseline). For better performance YOLOv8m was trained for 100 epochs with a fixed input image size of 40×640 pixels for both training and validation. The batch size was consistently set to 4 across all experiments. The dataset AI-TODv2 subset comprised 600 images (augmented to approximately 2400 through data augmentation) for experiments using augmentation and 180 images for validation and total 2214 images were used to set baseline results (without augmentation). The dataset SODA-A, consisting of 2513 images, was divided into three subsets: 1067 images for training, 576 images for validation, and 870 images for testing. When applying the framework the training set will increase to (4000 images). Optimization was conducted using the default settings provided by the YOLO framework. The results obtained without augmentation on YOLOv8 serve as baseline. We evaluate performance using AP (Average Precision), AP50 (AP at IoU = 0.50), and AR (Average Recall).

Table 3 provides the full specification of the augmentation pipeline used in this work. The same configuration is applied to both AI-TODv2 and SODA-A datasets. During the training phase, data augmentation strategies were applied on both the dataset to enhance the YOLOv8' generalization capabilities, resulting in improved performance across the validation scenarios. As shown in Table 4, the model was trained with the listed configurations. We applied the same settings for baseline ensured the validity of the comparative analysis.

4.4 Performance analysis

4.4.1 AI-TODv2 dataset

The experimental results shown in Table 5, demonstrate measurable improvements in YOLOv8m's performance using the proposed framework. The most notable enhancement is in recall, which increased by 31% points, indicating significantly fewer missed detections. The F1-score improved by 19 percentage points, reflecting better balance between precision and recall. AP50 increased by 15.9% points, showing stronger detection performance at the standard IoU threshold of 0.5. Our proposed framework outperforms the baseline YOLOv8m across all categories, achieving top results in five, reinforcing the visual outcomes shown in Table 6. The framework enhances detection for Airplane, improving from 17.5% to 17.7% with a slight increase of 0.2, Bridge, improving from 2.6% to 11.27% with a significant increase of 9.2%, Storage-Tank, improving from 78.9% to 80.1% with a slight increase of 1.2%, and Person, improving from 42.6% to 47.6% with a substantial increase of 5%. However, decline was observed for the ship class (–40.1%), which provides valuable insight into how categories with large homogeneous textures, scale variability, and complex water backgrounds respond differently to augmentation strategies. The validation results are also shown an improvement compared to the baseline in Table 5. Moreover, Fig. 5 shows curve maintains high recall at low-to-moderate confidence thresholds.

4.4.2 SODA-A

The experimental results on the SODA-A dataset, summarized in Tables 8 and 9, demonstrate that the proposed framework yields consistent improvements across all evaluation metrics and object classes when compared with the baseline YOLOv8m model.

The proposed framework shows a 23.6% gain in AP50 and an

Table 5 YOLOv8 validation results on AI-TODv2

Method	Publication	AP	AP50	Recall	F1
Faster RCNN [63]	2015	16.2	35.7	42.0	34.5
RoI Transformer [64]	2019	17.0	37.2	44.5	36.0
Oriented RCNN [65]	2021	17.5	37.8	45.2	37.2
SSPNet [66]	2022	16.8	36.5	43.1	35.0
YOLOv8m [62]	2023	**17.9**	38.1	46.0	39.0
YOLOv8m with proposed framework	–	17.3	**54.0**	**77.0**	**58.0**

The improved results are in **bold** as %

Table 6 Class-wise AP comparison across different detectors

Method	Airplane	Bridge	Storage tank	Ship	Vehicle	Person
Faster RCNN [63]	12.4	3.1	22.5	18.7	15.2	2.9
RoI Transformer [64]	11.7	9.7	22.9	32.3	15.7	17.7
Oriented RCNN [65]	13.7	2.9	17.9	23.6	18.2	4.7
SSPNet [66]	15.2	6.7	43.2	22.8	31.2	28.9
YOLOv8m [62]	17.5	2.6	78.9	**82.9**	42.6	0.0
YOLOv8m with proposed framework	**17.7**	**11.27**	**80.1**	42.8	**47.5**	**71.6**

The best results are in **bold** as %. Improvement = With proposed framework − Baseline (No Aug.)

Table 7 SODA-A dataset split (images)

Split	# Images	% of total images (%)
Train	1067	∼42.5
Val	576	∼22.9
Test	870	∼34.6
Total	2513	100

AP to 0.7%. Both recall and F1 improve +25.7% and 15.7% respectively. Table 9 shows improvement in all the classes AP results. Overall, these results show that the proposed framework substantially improves YOLOv8m's ability to detect small objects on SODA-A (Fig. 5).

4.5 Ablation study

To further examine the contribution of each component in the proposed augmentation framework, we performed an ablation study on the SODA-A dataset using YOLOv8m as the baseline detector. The baseline model was trained on the original SODA-A dataset without any augmentation. As shown in Table 7, this configuration achieved 11.2% AP50, 36.7% AP, and 48.5% recall, indicating limited ability to detect sparsely distributed and occluded small objects.

4.5.1 Stage I: image segmentation and cropping

Applying segmentation into four sub-regions and discarding empty patches yielded a noticeable improvement. AP50 increased to 53.6% (+1.9%) and recall rose to 65.8% (+17.3%), demonstrating that reducing background bias and focusing on object-dense regions strengthens the model's ability to capture small objects.

4.5.2 Stage II: object-centric resizing and augmentation

Only resizing and augmentation of object-dense regions were applied. Standardizing image dimensions while amplifying the absolute mean object size improved network compatibility and learning focus. Compared to baseline, AP50 rose to 62.3% (+10.6%) and recall improved to 68.4% (+19.9%). This confirms that resizing alone contributes significantly. AP score also improved slightly.

4.5.3 Stage III: full framework

Finally, applying the complete pipeline achieved the best performance. AP50 rose to 74.8% (+23.6%),Ap to 37.4% improved slightly, recall to 74.2% (+25.7%), and F1-score improved by +15.7% compared to baseline. The complete framework combining segmentation, zoom-in cropping, and object-centric resizing provided the best results shown in Table 10.

Table 8 Test results on SODA-A

Model	Publication	AP	AP50	Recall	F1
Faster RCNN	2015 [63]	27.6	45.21	39.7	32.5
RoI transformer	2019 [64]	36.9	73.2	59.8	50.45
Oriented RCNN	2021 [65]	34.5	68.7	47.6	40.1
SSPNet	2022 [66]	31.2	65.3	47.8	38.0
YOLOv8m	2023 [62]	36.7	51.2	48.5	40.3
YOLOv8m with proposed Framework	–	**37.4**	**74.8**	**74.2**	**56.0**

The best results are shown in bold as %

Table 9 Category-wise AP comparison between baseline detectors and proposed framework

Method	Airplane	Helicopter	s-vehicle	l-vehicle	Ship	Container	s-tank	s-pool	Windmill
Faster RCNN [63]	48.3	19.7	23.7	21.6	42.4	39.8	43.8	35.2	22.9
RoI Transformer [64]	52.8	21.7	37.1	26.8	45.7	36.4	45.6	36.8	27.5
Oriented RCNN [65]	52.7	21.2	33.9	25.5	46.3	33.1	42.8	37.2	21.8
SSPNet [66]	42.1	21.1	32.1	15.6	37.3	27.2	28.6	34.8	22.5
YOLOv8m [62]	51.7	20.4	36.1	25.9	43.5	35.7	44.6	36.9	23.5
YOLOv8m with proposed framework	**63.2**	**35.0**	**48.5**	**39.2**	**55.8**	**50.1**	**58.0**	**49.5**	**37.0**

The best results are shown in bold as %

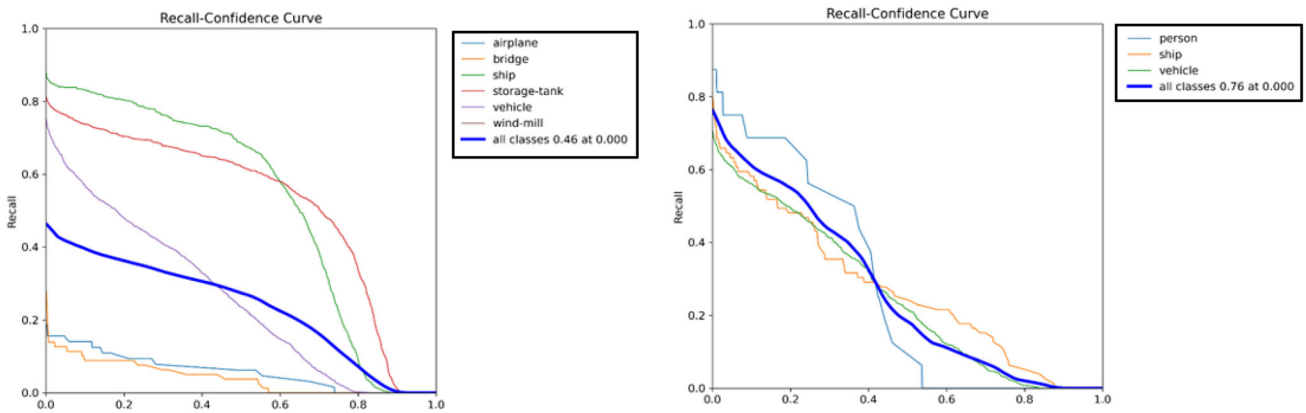


Fig. 5 Comparison analysis of recall-confidence curves for YOLOv8 with and without data augmentation on AI-TODv2: The row represents YOLOv8 (left: without augmentation approaches, right: with augmentation approaches)

Table 10 Ablation results on the dataset SODA-A

Configuration	AP50 (%)	AP (%)	Recall (%)	F1 (%)
Baseline (No Aug)	51.7	36.7	48.5	40.3
Stage 1: segmentation & cropping	53.6	36.9	65.8	49.2
Stage 2: object-centric resizing only	62.3	37.1	68.4	51.0
Full pipeline (Stage 1 + Stage 2)	**74.8**	**37.4**	**74.2**	**56.0**

The best results are shown in bold as %



Fig. 6 The visualized results comparison of AI-TODv2 dataset using YOLOv8m. The left image shows results without augmentation approaches applied, missing several objects. The right image demonstrates the proposed framework with augmentation, accurately detecting more objects

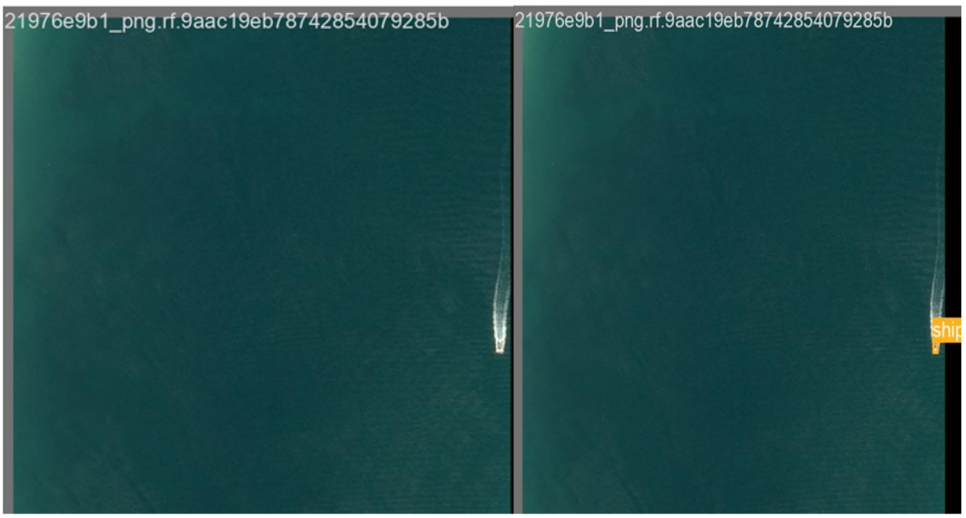


Fig. 7 The visualized results compare SODA-A (right: without augmentation), which missed the object, with our proposed framework (left), using YOLOv8m, that successfully detects the object

4.6 Visual representation of proposed framework

In addition to the quantitative analysis previously presented, we now provide a comparative visualization that illustrates the qualitative results. Visual comparisons between the proposed framework applied to AI-TODv2 and SODA-A using the YOLOv8m baseline model are provided in Figs. 6 and 7, respectively. Furthermore, Fig. 8 present a significant increase in the number of detected objects, highlighting the improved detection capabilities of our framework.

5 Conclusion and future work

It is one of the major issues in small object detection regarding the availability of large-scale datasets, which is important for training, especially in aerial images. Data augmentation approaches have been used by researchers to deal with the scarcity of datasets, but which data augmentation methods and combinations are optimal is still a question. To address this, we have proposed a framework based on dataset limitations that used only traditional augmentation approaches. For experiments, this study used the AI-TODv2 and SODA-A datasets, which are crucial datasets for small object detection and share common to different known limitations. We found that the trial-and-error method is time consuming and sometimes does not align with the objective of the

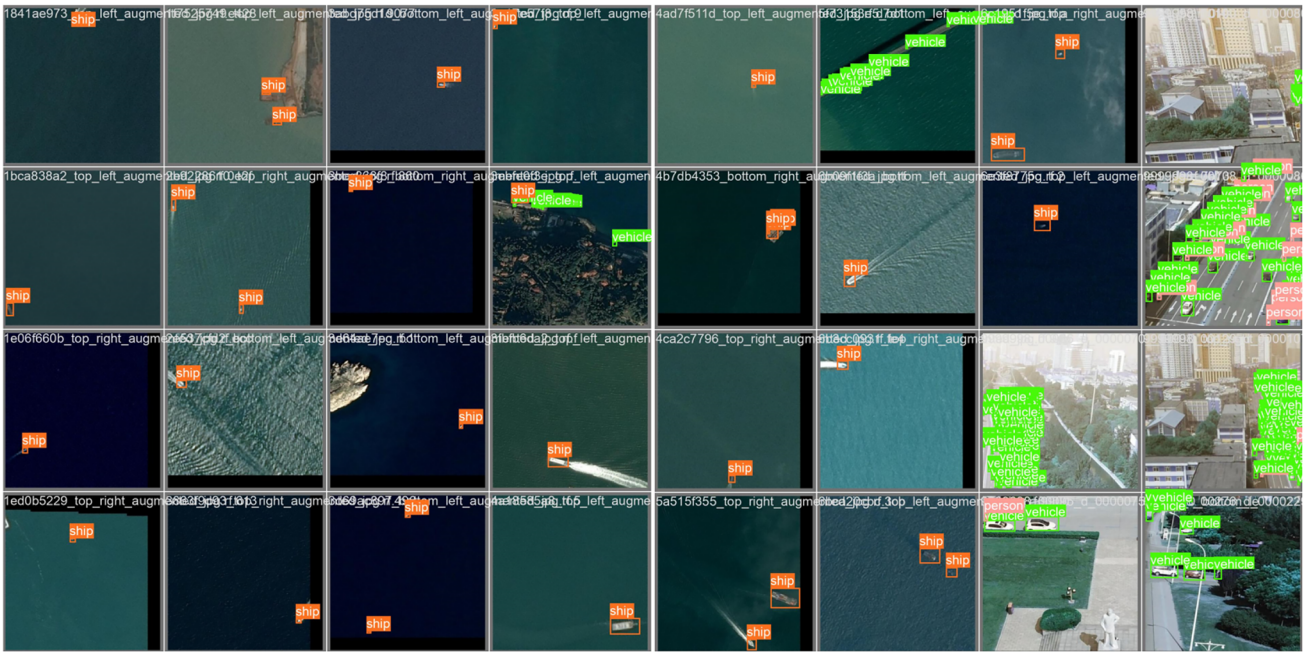


Fig. 8 Improved object detection results using YOLOv8 with the proposed augmentation framework

task, resulting in suboptimal outcomes. This study proposes a systematic three-tier framework for selecting augmentation approaches from a wide range of options, eliminating the need for trial and error. Based on this framework, a pipeline has been developed to detect small objects. We used two datasets: AI-TODv2 and SODA-A. The rigorous experiments using YOLOv8 on AI-ToDv2 and SODA-A with the proposed framework demonstrate better performance on the popular datasets. The proposed framework has shown improved results, despite certain limitations. Due to hardware constraints, experiments on the AI-TODv2 dataset were conducted using a subset rather than the full dataset. Extreme cases, such as very low-resolution or noisy images, remain challenging, suggesting that future work should incorporate complementary strategies, such as incremental learning. Additionally, we intend to conduct experiments on the full AI-TODv2 dataset in the future.

Author Contributions Ume Nisa performed all experiments and prepared the manuscript, including all figures and tables. Muhammad Syafiq Mohd Pozi and Mohamed Ali Saip provided critical revisions and contributed to the overall refinement of the manuscript.

Funding Open access funding provided by The Ministry of Higher Education Malaysia and Universiti Utara Malaysia Open access funding provided by The Ministry of Higher Education Malaysia and Universiti Utara Malaysia. No funding was received for this study.

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1627–1645 (2009)
2. Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M.: Deep learning for generic object detection: a survey. Int. J. Comput. Vis. **128**, 261–318 (2020)
3. Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M.Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M., et al.: Object detection in aerial images: a large-scale benchmark and challenges. IEEE Trans. Pattern Anal. Mach. Intell. **44**(11), 7778–7796 (2022)

4. 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, pp. 9725–9734 (2019)
5. Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., Cho, K.: Augmentation for Small Object Detection. arXiv:1902.07296 (2019)
6. Cai, Y., Luan, T., Gao, H., Wang, H., Chen, L., Li, Y., Sotelo, M.A., Li, Z.: YOLOv4-5D: an effective and efficient object detector for autonomous driving. IEEE Trans. Instrum. Meas. **70**, 1–13 (2021)
7. Mahaur, B., Mishra, K.: Small-object detection based on YOLOv5 in autonomous driving systems. Pattern Recogn. Lett. **168**, 115–122 (2023)
8. Yundong, L., Han, D., Hongguang, L., Zhang, X., Zhang, B., Zhifeng, X.: Multi-block SSD based on small object detection for UAV railway scene surveillance. Chin. J. Aeronaut. **33**(6), 1747–1755 (2020)
9. Han, W., Li, J., Wang, S., Wang, Y., Yan, J., Fan, R., Zhang, X., Wang, L.: A context-scale-aware detector and a new benchmark for remote sensing small weak object detection in unmanned aerial vehicle images. Int. J. Appl. Earth Obs. Geoinf. **112**, 102966 (2022)
10. Bhil, K., Shindihatti, R., Mirza, S., Latkar, S., Ingle, Y., Shaikh, N., Prabu, I., Pardeshi, S.N.: Recent progress in object detection in satellite imagery: a review. In: Sustainable Advanced Computing: Select Proceedings of ICSAC, vol. 2021, pp. 209–218 (2022)
11. Zeng, S., Yang, W., Jiao, Y., Geng, L., Chen, X.: SCA-YOLO: a new small object detection model for UAV images. Vis. Comput. **40**(3), 1787–1803 (2024)
12. 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), pp. 3791–3798. IEEE (2021)
13. Liu, Y., Sun, P., Wergeles, N., Shang, Y.: A survey and performance evaluation of deep learning methods for small object detection. Expert Syst. Appl. **172**, 114602 (2021)
14. Hu, M., Li, Z., Yu, J., Wan, X., Tan, H., Lin, Z.: Efficient-lightweight YOLO: improving small object detection in YOLO for aerial images. Sensors **23**(14), 6423 (2023)
15. Xie, X., Cheng, G., Li, Q., Miao, S., Li, K., Han, J.: Fewer is more: efficient object detection in large aerial images. Sci. China Inf. Sci. **67**(1), 112106 (2024)
16. Tong, K., Wu, Y.: Deep learning-based detection from the perspective of small or tiny objects: a survey. Image Vis. Comput. **123**, 104471 (2022)
17. Raj, A., Gandhi, K., Nalla, B.T., Verma, N.K.: Object detection and recognition using small labeled datasets. In: Computational Intelligence: Theories, Applications and Future Directions-Volume II: ICCI-2017, pp. 407–419. Springer, Berlin (2019). https://doi.org/10.1007/978-981-13-1135-2_31
18. Cheng, B., Girshick, R., Dollár, P., Berg, A.C., Kirillov, A.: Boundary IoU: Improving object-centric image segmentation evaluation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15334–15342 (2021)
19. Cheng, G., Yuan, X., Yao, X., Yan, K., Zeng, Q., Xie, X., Han, J.: Towards large-scale small object detection: survey and benchmarks. IEEE Trans. Pattern Anal. Mach. Intell. **45**(11), 13467–13488 (2023). https://doi.org/10.1109/tpami.2023.3290594
20. Nisa, U.: Image augmentation approaches for small and tiny object detection in aerial images: a review. Multimed. Tools Appl. **84**, 21521–21568 (2025)
21. 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 J. Photogramm. Remote. Sens. **190**, 79–93 (2022)
22. Ma, P., He, X., Chen, Y., Liu, Y.: ISOD: improved small object detection based on extended scale feature pyramid network. Vis. Comput. **41**(1), 465–479 (2025)
23. Liu, K., Mattyus, G.: Fast multiclass vehicle detection on aerial images. IEEE Geosci. Remote Sens. Lett. **12**(9), 1938–1942 (2015)
24. 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, pp. 3974–3983 (2018)
25. 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 J. Sel. Top. Appl. Earth Obs. Remote Sens. **14**, 2148–2161 (2020)
26. Li, K., Wan, G., Cheng, G., Meng, L., Han, J.: Object detection in optical remote sensing images: a survey and a new benchmark. ISPRS J. Photogramm. Remote. Sens. **159**, 296–307 (2020)
27. 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: Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13, pp. 740–755. Springer, Berlin (2014)
28. 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), pp. 370–386 (2018)
29. Razakarivony, S., Jurie, F.: Vehicle detection in aerial imagery: a small target detection benchmark. J. Vis. Commun. Image Represent. **34**, 187–203 (2016)
30. Zhu, P., Wen, L., Du, D., Bian, X., Ling, H., Hu, Q., Nie, Q., Cheng, H., Liu, C., Liu, X., et al.: Visdrone-DET2018: the vision meets drone object detection in image challenge results. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 437–468 (2018)
31. Du, D., Zhu, P., Wen, L., Bian, X., Lin, H., Hu, Q., Peng, T., Zheng, J., Wang, X., Zhang, Y., et al.: VisDrone-DET2019: the vision meets drone object detection in image challenge results. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 213–226 (2019). https://doi.org/10.1109/ICCVW.2019.00030
32. Cao, Y., He, Z., Wang, L., Wang, W., Yuan, Y., Zhang, D., Zhang, J., Zhu, P., Van Gool, L., Han, J., et al.: VisDrone-DET2021: the vision meets drone object detection challenge results. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2847–2854 (2021). https://doi.org/10.1109/ICCVW54120.2021.00319
33. Zhan, W., Sun, C., Wang, M., She, J., Zhang, Y., Zhang, Z., Sun, Y.: An improved YOLOv5 real-time detection method for small objects captured by UAV. Soft. Comput. **26**, 361–373 (2022)
34. Zhang, X., Feng, Y., Wang, N., Lu, G., Mei, S.: Aerial person detection for search and rescue: survey and benchmarks. J. Remote Sens. **2025**, 0474 (2025)
35. Zhang, X., Feng, Y., Zhang, S., Wang, N., Lu, G., Mei, S.: Robust aerial person detection with lightweight distillation network for edge deployment. IEEE Trans. Geosci. Remote Sens. **62**, 1–16 (2024)
36. Junos, M.H., Mohd Khairuddin, A.S., Thannirmalai, S., Dahari, M.: Automatic detection of oil palm fruits from UAV images using an improved YOLO model. Vis. Comput. **38**(7), 2341–2355 (2022)
37. Psiroukis, V., Espejo-Garcia, B., Chitos, A., Dedousis, A., Karantzalos, K., Fountas, S.: Assessment of different object detectors for the maturity level classification of broccoli crops using UAV imagery. Remote Sens. **14**(3), 731 (2022)
38. Wang, S.: An augmentation small object detection method based on NAS-FPN. In: 2020 7th International Conference on Information Science and Control Engineering (ICISCE), pp. 213–218 (2020). IEEE
39. Zhu, X., Lyu, S., Wang, X., Zhao, Q.: TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detec-

tion on drone-captured scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2778–2788 (2021)
40. 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, pp. 2260–2270 (2022)
41. Ruiz-Ponce, P., Ortiz-Perez, D., Garcia-Rodriguez, J., Kiefer, B.: Poseidon: a data augmentation tool for small object detection datasets in maritime environments. Sensors **23**(7), 3691 (2023)
42. Deng, Z., Sun, H., Zhou, S., Zhao, J., Zou, H.: Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **10**(8), 3652–3664 (2017)
43. Zhao, H., Zhang, H., Zhao, Y.: Yolov7-sea: Object detection of maritime UAV images based on improved YOLOv7. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 233–238 (2023)
44. Gao, C., Tang, W., Jin, L., Jun, Y.: Exploring effective methods to improve the performance of tiny object detection. In: Computer Vision-ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16, pp. 331–336. Springer, Berlin (2020)
45. Chen, Y., Zhang, P., Li, Z., Li, Y., Zhang, X., Meng, G., Xiang, S., Sun, J., Jia, J.: Stitcher: Feedback-Driven Data Provider for Object Detection (2020). https://doi.org/10.48550/arXiv.2004.12432
46. Ye, Y., Ren, X., Zhu, B., Tang, T., Tan, X., Gui, Y., Yao, Q.: An adaptive attention fusion mechanism convolutional network for object detection in remote sensing images. Remote Sens. **14**(3), 516 (2022)
47. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934 (2020)
48. Wei, Z., Duan, C., Song, X., Tian, Y., Wang, H.: AMR-NET: Chips Augmentation in Aerial Images Object Detection. arXiv:2009.07168 (2020)
49. 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, pp. 2789–2798 (2021)
50. Huang, H., Tang, X., Wen, F., Jin, X.: Small object detection method with shallow feature fusion network for chip surface defect detection. Sci. Rep. **12**(1), 3914 (2022)
51. Zhang, X., Feng, Y., Wang, N., Lu, G., Mei, S.: Transformer based person detection in paired RGB-T aerial images with VTSaR dataset. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens , 1–18 (2025). https://doi.org/10.1109/JSTARS.2025.3526995
52. Herdea, I.-A., Tiwari, D., Oyekan, J., Tiwari, A.: Data augmentation framework for improved classification in object detectors. IEEE Access , 1–1 (2025). https://doi.org/10.1109/ACCESS.2025.3539455
53. Wang, D., Gao, Z., Fang, J., Li, Y., Xu, Z.: Improving UAV aerial imagery detection method via super-resolution synergy. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens , 1–15 (2025). https://doi.org/10.1109/JSTARS.2024.3525148
54. Dhakal, S., Qu, D., Carrillo, D., Dehghani Tezerjani, M., Yang, Q.: Virtualpainting: addressing sparsity with virtual points and distance-aware data augmentation for 3d object detection. Sensors **25**(11), 3367 (2025)
55. Sun, Z., Leng, X., Zhang, X., Zhou, Z., Xiong, B., Ji, K., Kuang, G.: Arbitrary direction SAR ship detection method for multi-scale imbalance. IEEE Trans. Geosci. Remote Sens. **63**, 1–1 (2025). https://doi.org/10.1109/TGRS.2025.3559701
56. Ehrnsperger, M.G., Noll, M., Siart, U., Eibert, T.F.: Background and clutter removal techniques for ultra short range radar. In: 2020 17th European Radar Conference (EuRAD), pp. 78–81. IEEE (2021)
57. Pradhan, P.K., Baruah, U.: Object detection under occlusion in aerial images: a review. In: Contemporary Issues in Communication, Cloud and Big Data Analytics: Proceedings of CCB, vol. 2020, pp. 215–227 (2021)
58. Al-Shammary, A.A., Zaghden, N., Bouhlel, M.S.: Improving automatic image annotation system to handle spatial analysis of aerial imagery. In: 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 153–158. IEEE (2022)
59. 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, pp. 8311–8320 (2019)
60. Schembri, M., Seychell, D.: Small object detection in highly variable backgrounds. In: 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 32–37. IEEE (2019)
61. Ultralytics: YOLOv8 Docs. https://docs.ultralytics.com/. Accessed: 27 April 2023 (2023)
62. Jocher, G., Chaurasia, A., Qiu, J.: YOLO by Ultralytics (Version 8.0.0). Computer software, GitHub (2023). https://github.com/ultralytics/ultralytics
63. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
64. Ding, J., Xue, N., Long, Y., Xia, G.-S., Lu, Q.: Learning roi transformer for oriented object detection in aerial images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2849–2858 (2019)
65. Xie, X., Cheng, G., Wang, J., Yao, X., Han, J.: Oriented R-CNN for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3520–3529 (2021)
66. 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 Geosci. Remote Sens. Lett. **19**, 1–5 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.