Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19768-7

Image augmentation approaches for small and tiny object detection in aerial images: a review

Ume Nisa[1]

Received: 29 February 2024 / Revised: 21 May 2024 / Accepted: 23 June 2024

Abstract
The task of detecting small and tiny objects has not shown significant performance improvement compared to detecting medium and large objects, even with advanced detection methods. Several factors contribute to this, including the small size of the objects themselves, limited availability of small-scale datasets, clustering of objects, and a low object-to-image ratio, etc. However, various image augmentation techniques have been developed to address these challenges. In this study, previous surveys and reviews have been summarized, offering an overview of the augmentation libraries used to implement various augmentation approaches. It also conducted a brief review of both traditional and state-of-the-art augmentation methods, outlining their respective advantages and disadvantages. The focus is on more to explore the effectiveness of traditional and state-of-the-art augmentation approaches for small and tiny object detection tasks in terms of performance improvement. This review paper distinguishes itself from previous papers by covering a thorough study of image augmentation approaches specially designed for detecting small and tiny object detection tasks in aerial images-a topic that has not been extensively covered on this scale in prior literature. This paper also provided a discussion enabling readers to discern which approach or combination of approaches is suitable for addressing specific challenges related to small and tiny object detection. This will help the readers to establish a good knowledge and develop critical and evaluation skills in this domain. Furthermore, it provides present challenges with possible solutions and outlines future directions.

Keywords Image augmentation · Small object detection · Aerial · dataset · Deep learning

1 Introduction

In computer vision models, especially in deep learning, the fundamental component is data. The performance of these deep learning solutions needs to be improved due to insufficient training data. Therefore, having ample data for training is essential for any deep learning model to function well [1]. Modern machine learning models often need a lot of high-quality, labelled data to ensure decent performance. However, image collection and labeling

✉ Ume Nisa
kazmi_nisa@hotmail.com

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

Published online: 01 August 2024

are time-consuming and costly processes. It is frequently not possible to gather enough training data in many real-world application contexts. At the moment, image augmentation approaches are the best solution for solving this issue. The primary objective of data augmentation is to extend the number, quality, generalizability, and variation of the training dataset. Image augmentation approaches can be used to increase the number of images and instances in the dataset by artificially generating variants of real images. Appropriate image augmentation approaches can also ease the need for complex algorithms and allow for the excellent generalization ability of more straightforward model structures [2].). In computer vision tasks, datasets contain images and videos. Even though there have been significant advancements in generic object detection, small object detection research has moved along 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 tiny objects [3, 4]. Researchers use self-generated datasets for small and tiny instance detection, which did not appear to perform any better in cross-evaluation studies than other methods. It is also believed that researchers have paid less attention to generating data augmentation approaches for small object detection as compared to other computer vision issues due to its complex structure. The identified gap is the lack of a review paper focused on performance analysis, as small and tiny object detection remains a crucial challenge for researchers. There is no comprehensive review paper specifically focused on small and tiny object detection. This review primarily focuses on the significant advancements made in the last five years in small or tiny object detection based on image augmentation approaches in aerial images. In order to be comprehensive and make the text easier to read, several additional relevant works are also included. The main contribution of this work is to compile the maximum amount of papers written on small and tiny object detection, which will assist researchers in understanding the current state of the field, identifying existing challenges, and exploring potential solutions.

1.1 Comparison with previous reviews

1.1.1 Small and tiny object detection reviews

In recent years there have been different reviews and survey papers written about small and tiny object detection that have been summarized Table 1. The paper written by [5] focuses on small object performance evaluation on four models for deep learning: Fast R-CNN, Faster R-CNN, RetinaNet, and YOLOv3. Also, the pros and cons of these models are introduced. Tong and Wu [6] describes in detail the current tiny and small object detection techniques based on deep learning from five perspectives, including multi-scale feature learning, data augmentation, training strategy, context-based detection, and GAN-based detection. But in this survey, researchers just briefly overview the basic augmentation approaches and their performance on small and tiny object detection. Moreover, analyzed experimental results on five datasets, and points out five promising future directions. Chen et al. [7] survey the four pillars (multiscale representation, contextual information, super-resolution, and region-proposal) for deep learning-based small object detection, lists some small object detection datasets, and reports the performance of different methods on three datasets (MS-COCO, PASCAL-VOC, and TT100K), Six possible directions in the future are also provided. Liu et al. [4] listed out four possible issues faced during the detection of small object detection: a: the

Table 1 Small and tiny object detection reviews/surveys

Title of paper	Year	Specifications
An evaluation of deep learning methods for small object detection [10]	2020	Focuses on small object performance evaluation on four models for deep learning: Fast R-CNN, Faster R-CNN, RetinaNet, and YOLOv3. Also, the pros and cons of these models are introduced.
Recent advances in small object detection based on deep learning: A review [11]	2020	Reviews the existing deep learning-based small object detection methods from five aspects, analyses experimental results on five datasets, and points out five promising directions.
A survey of the four pillars for small object detection: multiscale representation, contextual information, super-resolution, and region proposal [7]	2020	Surveys the four pillars for deep learning-based small object detection, lists some small object detection datasets, and reports the performance of different methods on three datasets. Six possible directions for the future are also provided.
A survey and performance evaluation of deep learning methods for small object detection [4]	2021	Aiming at the four challenges of small object detection, it summarizes the corresponding solutions and offers some experimental analysis.
A Survey on Deep Domain Adaptation and Tiny Object Detection Challenges, Techniques and Datasets [12]	2021	Three alternative current techniques for object detection were addressed in this survey. 1) Discrepancy-based, Adversarial-based, Reconstruction-based, Hybrid domain adaptive deep learning approaches are studied general and tiny object detection-related issues and provided solutions by using empirical and comparative analysis. 2) They reviewed 5 strategies of tiny object detection namely: multi-scale feature learning, Data augmentation, Training strategy (TS), Context-based detection, GAN-based detection 3) They also identified various object detection techniques, such as convolutions and convolutional neural networks (CNN), pooling operations.
Deep Learning for UAV-based Object Detection and Tracking: A Survey [8]	2021	This paper provides an in-depth analysis of the state-of-the-art and future potential of deep learning-based UAV object recognition and tracking techniques.
Deep learning-based detection from the perspective of small or tiny objects: A survey [6]	2022	This survey comprehensively discusses small or tiny object datasets, the definitions of small or tiny objects, techniques for small or tiny object detection, detection performance analysis of small or tiny objects, and promising directions for small/tiny object detection.
Image Augmentation Approaches for Small and Tiny Object Detection in Aerial Images: A Review	2024	Ours

information contained in each feature layer is insufficient for small objects; b: a lack of context information for small instances; c: class imbalance; d: small objects contain inadequate positive examples. Besides issues, they enlist their solutions as well as compare performances of YOLOv3 and RCNN family on three datasets regarding small object detection. The solutions offered are 1: feature maps combination 2: putting additional contextual information 3: some solutions of balance classes 4: addition of positive examples. Wu et al. [8] provides an in-depth analysis of the state-of-the-art and future potential of deep learning-based UAV object recognition and tracking techniques. Three alternative current techniques for object detection were addressed in this survey [9]. 1) Discrepancy-based, Adversarial-based, Reconstruction-based, and Hybrid domain adaptive deep learning approaches study general and tiny object detection-related issues and provided solutions by using empirical and comparative analysis. 2) They reviewed 5 strategies of tiny object detection namely: multi-scale feature learning, Data augmentation, Training strategy (TS), Context-based detection, and GAN-based detection 3) They also identified various object detection techniques, such as convolutions and convolutional neural networks (CNN), pooling operations. Although this is a comprehensive survey, it just briefly highlighted image augmentation approaches related to tiny object detection. Tong and Wu [6] This survey comprehensively discusses small or tiny object datasets, the definitions of small or tiny objects, techniques for small or tiny object detection, detection performance analysis of small or tiny objects, and promising directions for small/tiny object detection.

1.1.2 Image augmentation surveys/reviews

Data Augmentation approaches are becoming popular day by day, especially in computer vision tasks. In Tables 2, 3 and 4, discussed a few recent and old surveys/reviews on data augmentation approaches. Zhou and Buyya [13] outlines the fundamental ideas and the difficult problems in making mobile cloud augmentation systems. The taxonomy is divided into two types: a) mobile computing augmentation, and b) mobile storage augmentation. Furthermore, the benefits and drawbacks of each type are addressed, along with the ideal circumstances under which it should be used. Mikołajczyk and Grochowski [14] reported the state-of-the-art data augmentation techniques concisely to improve performance in image classification tasks. Shorten and Khoshgoftaar [15] presented an extensive survey related to image augmentation approaches for 2D image data, and their main challenges during augmentation. Its focus is the image classification task. It does not go into great detail about more modern techniques for data augmentation, such as as neural style transfer (NST), 3D modelling etc. [16–19] presented reviews/survey papers on image augmentation approaches but their primary application domain is medical images. Khosla and Saini [20] cover data wrapping and oversampling data augmentation approaches. Moreover, the overfitting issue and its solutions have been concisely discussed. Wang et al. [21] overview the categories of transformation that supported the face data augmentation. This paper also discussed the background and the concept of face data augmentation, commonly used approaches by comparative analysis. Liu et al. [22] survey application domain is the augmentation of text data in NLP. Naveed et al. [23] is primarily concentrated on region-level data augmentation approaches such as image region deletion, addition, and combination approaches. Kaur et al. [1] presented a comprehensive review of photometric and geometric transformation data augmentation approaches related to object detection but precisely touched small object detection image augmentation methods. Duong and Nguyen-Thi [5] reviewed text enhancing augmentation techniques such as Syntax-Tree Transformation, Back translation, and synonyms replacement for analysis of

Table 2 Image augmentation surveys/reviews

Title of Paper	Year	Application domain	Specifications
Augmentation techniques for mobile cloud computing: A taxonomy, survey, and future directions [13]	2018	Cloud computing	It discusses the most important ideas and toughest problems in implementing mobile cloud augmentation systems. Moreover, it organizes the cutting-edge methods into two categories and creates a taxonomy for mobile compute augmentation and mobile storage augmentation, respectively. The benefits and drawbacks of each sort of approach are addressed, along with the ideal circumstances under which it should be used. Additionally, it also covers future directions and assesses the gaps that mobile cloud augmentation systems still have.
Data augmentation for improving deep learning in image classification problem [14]	2018	Classification	This paper confers a new perspective on the texture style and style-transfer technique together with a novel application concept and pertinent examples, narrating advanced approaches used for data augmentation in classification tasks. Also, provides findings after discussing the benefits and drawbacks of the methodologies under consideration.
A survey on image data augmentation for deep learning [31]	2019	Extensive (2D images)	A comprehensive analysis of numerous data augmentation techniques for 2D image data is provided by the authors. Their assessment addresses a number of important difficulties that come up while using data augmentation such as changes in image resolution and dataset sizes growing rapidly.
Data augmentation for brain-tumor segmentation: a review [16]	2019	Medical Images	They highlight augmentation approaches like affine, elastic and, pixel-wise on medical images (brain tumor). Besides, GAN and Cycle GAN are also discussed concisely.

Table 2 continued

Title of Paper	Year	Application domain	Specifications
Enhancing performance of deep learning models with different data augmentation techniques: A survey [20]	2020	Extensive (data warping, oversampling, the overfitting problem)	This survey article focuses on various data augmentation (GAN) methods regarding data warping and oversampling. Moreover, overfitting issue and its solutions have been discussed briefly.
A survey on face data augmentation for the training of deep neural networks [21]	2020	Face Recognition	A literary review of data augmentation approaches demonstrates the broad range of applications for various face detection and identification challenges. This paper describes various types of data augmentation methods for face detection and recognition by comparing techniques.
A survey of text data augmentation [22]	2020	Text Estimation	Text data augmentation in NLP is uncommon. The authors present a detailed survey of text data augmentation methods and their application in different domains.
Survey: Image mixing and deleting for data augmentation [32]	2021	Classification, image recognition and object detection	In this survey, the primary concentration is the subclass of data augmentation techniques: region-level (deletion, addition, combination) data augmentation methods. Additionally, it empirically tests these methods for classification, fine-grained image recognition, and object detection, demonstrating how this type of data augmentation enhances deep neural networks' general performance.
Data augmentation for object detection: A review [1]	2021	Object Detection	This is an exhaustive review of data augmentation approaches used in the field of object detection. Furthermore, they stated the class imbalance problem and outlined some solutions in the literature. In the end, time augmentation methods have been elaborated.

Table 2 continued

Title of Paper	Year	Application domain	Specifications
GAN-based data augmentation and anonymization for skin-lesion analysis: A critical review [17]	2021	Medical Images	The methods utilized for GAN-based augmentation are summarized following a thorough examination of the literature. In that review, they enumerate problems with experimental planning that might produce overly positive outcomes.

sentiments. Shorten et al. [24] has discussed a number of techniques for using data augmentation in text data, including label augmentation, neural augmentation, and symbolic augmentation. Khalifa et al. [25] presented a comprehensive survey about image augmentation in different application domains - specifically, medical imaging, remote sensing, and agriculture. Yang et al. [26] briefly presented basic approaches (Image Manipulation Image Erasing Image Mix) and advance approaches (Auto Augment Feature Augmentation Deep Generative Models) of data augmentation in three application domains image classification, object detection and segmentation. Mumuni and Mumuni [27] presented a detailed survey that covers almost every data augmentation method. Although it is a general and comprehensive survey but does not deal with any specific application domain. Oubara et al. [28] precisely reviewed the augmentation methods adopted in remote sensing for different computer vision tasks but did not address object detection or small object detection. Lewy and Mańdziuk [29] briefly described and compared two data augmentation topics presented in the last five years: automated selection of augmentation approaches and mixing images. They also subdivided them into pixel-wise, patch-wise, pairs of images, and images used to generate an augmented sample. Li et al. [30] just over-viewed image augmentation basic approaches as data preprocessing methods for object detection.

1.2 Contribution

The major contribution of this review is to analyze and assess the role of image augmentation approaches in addressing the challenge of small and tiny object detection. This paper has summarized survey and review papers from the last five years. Tables 1, 2, 3, and 4 show that no survey or review paper has been published yet regarding image augmentation approaches for small and tiny object detection. It also highlights some basic and state-of-the-art image augmentation approaches, including geometric and photometric transformations, region-wise, pixel-wise, and sample-wise transformations, as well as GANs, Neural Style Transfer, Meta Learning, and feature space transformations, and their performance for small and tiny object detection tasks. This study is based on extensive research and covers a wide range of subjects, some of which are novel contributions to the knowledge. Moreover, these additions create an updated, comprehensive, and complete assessment that notably sets itself apart from earlier reviews and surveys. This research is considered an important addition to the image augmentation and small object detection domain. Additionally, it is anticipated that this review will inspire new ideas for research on image augmentation and help researchers

Table 3 Image augmentation surveys/reviews (continued)

Title of paper	Year	Application domain	Specifications
A review of medical image data augmentation techniques for deep learning applications [18]	2021	Medical Image	It shows a critical review of data augmentation (fundamental, deformable, deep learning, etc.) approaches trained deep learning models using medical images limited to CT and MRI. This review seeks to provide an understanding of these methods and confidence in the accuracy of the models generated.
A review: preprocessing techniques and data augmentation for sentiment analysis [5]	2021	Analysis of Sentiments	It reviews Syntax-free Transformation, Synonyms-Replacement, Back Translation augmentation approaches for enhancing the data for sentiment analysis.
Text data augmentation for deep learning [24]	2021	Text Data	In this survey, numerous methods for applying data augmentation approaches including label augmentation, neural augmentation, and symbolic augmentation have been discussed. Additionally, they also elaborate the use of data augmentation to simulate distribution shift and test generalization. They also highlight the main differences and similarities in data augmentation NLP and computer vision fields.
A comprehensive survey of recent trends in deep learning for digital images augmentation [25]	2022	Extensive (medical imaging, remote sensing, and agriculture)	This survey provides a systematic review of 2D image augmentation approaches and encompasses a wider range of fields such as medical imaging, remote sensing, and agriculture. Nevertheless, contemporary cutting-edge techniques like neural style transfer, meta-learning, and Generative adversarial network methods are only explained concisely.
Image data augmentation for deep learning: A survey [26]	2022	Computer Vision Tasks	They provide a succinct overview of image data augmentation approaches along with comparisons of the performance of standard augmentation methods for computer vision tasks such as classification, object detection, and segmentation.

Table 3 continued

Title of paper	Year	Application domain	Specifications
Data augmentation: A comprehensive survey of modern approaches [27]	2022	Computer Vision Tasks	This is a comprehensive survey of modern approaches and applications of data augmentation approaches to date. They review basic and advanced methods of deep learning, feature learning, and meta-learning augmentation approaches. Moreover, it presents topics like 3D graphics modeling, neural rendering, and generative adversarial networks. Finally, a comparative analysis is performed on different advanced augmentation approaches.
Generative adversarial networks in medical image augmentation: a review [19]	2022	Medical image	This paper is divided and reviewed into three groups: classification, segmentation, and imaging modality in terms of GAN-based augmentation methods. It also presents merits, demerits, and potential future research for segmenting medical images.
Survey on remote sensing data augmentation: Advances, challenges, and future perspectives [28]	2022	Remote Sensing	Basic Techniques, Imaging Simulation System Based Methods, Deep Learning-based Methods briefly. Precisely reviewed the augmentation methods adopted in remote sensing for different computer vision tasks but did not address object detection.

better comprehend the detection of small and tiny objects. The following is a summary of contributions to this paper:

1. The recent and older reviews and survey papers on small and tiny object detection, as well as image augmentation, have been summarized. It has been observed that, to date, there is no comprehensive paper specifically addressing image augmentation for the detection of small and tiny objects. While a few review/survey papers have briefly touched on this concept, none have provided a thorough analysis. The novelty of this paper lies in presenting the first comprehensive review of image augmentation approaches specifically used for detection of small and tiny objects. This work systematically compiles and assesses the scattered literature, providing a detailed analysis that has been notably absent in previous research.
2. Additionally, This paper has also provided an overview of augmentation libraries utilized for different approaches.
3. Systematically review image augmentation approaches such as geometric and photometric transformation, region-wise, pixel-wise, sample-wise, GAN, Neural Style Transfer, Meta Learning, and Feature Space with their advantages and disadvantages.

Table 4 Image augmentation surveys/reviews (continued)

Title of paper	Year	Application domain	Specifications
A review: Data pre-processing and data augmentation techniques [33]	2022	Pre-Processing Data	This paper discusses various pre-processing (Data Transformation, Information gathering, Data generation) and Augmentation techniques (Symbolic augmentation, Rule-based augmentation, Graph-structured augmentation, Mix-up augmentation, Feature space augmentation, Neural augmentation) for improving the performance and outcomes of machine learning designed models. For increasing the accuracy of the prediction of the models' various techniques of an image, data augmentation is discussed with the concepts of Data Wrapping and Data Oversampling.
A survey on GAN-based data augmentation for hand pose estimation problem [34]	2022	Pose Estimation	This survey presents a detailed synopsis of pose estimation algorithms that benefitted from GAN-based data augmentation approaches, hand-pose datasets, and their comparative analysis. The limitations of the current approaches and potential future paths are highlighted in the conclusion.
A review on remote sensing imagery augmentation using deep learning [35]	2022	Remote Sensing	Images This work aims to provide an overview of the most recent and cutting-edge image augmentation using deep learning based on CNN applications for improving remote sensing images.
An overview of mixing augmentation methods and augmentation strategies [36]	2022	Multiple domains	This survey briefly describes and compares two data augmentation topics presented in the last five years: automated selection of augmentation approaches and mixing images. They also subdivided them into pixel-wise, patch-wise, pairs of images, and images used to generate an augmented sample.

Table 4 continued

Title of paper	Year	Application domain	Specifications
Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey [30]	2022	Object Detection for Remote Sensing Images	This is one of the comprehensive surveys regarding object detection for aerial images. They present RSI object detection approaches, processes, datasets, metrics, and their comparative analysis. They briefly describe basic approaches of data augmentation as data preprocessing methods.
Data Augmentation in Classification and Segmentation: A Survey and New Strategies [37]	2023	Classification and Segmentation	The paper surveys data augmentation techniques in computer vision, with a focus on segmentation and classification. It introduces the "random local rotation strategy," a novel approach that mitigates the shortcomings of traditional rotation methods.
Image Data Augmentation Approaches: A Comprehensive Survey and Future directions [38]	2023	General Overview	This survey comprehensively reviews SOTA data augmentation methods to combat overfitting in computer vision tasks with limited data. It provides a detailed taxonomy and results across tasks like image classification, object detection, and semantic segmentation for both supervised and semi-supervised learning.
Image Augmentation Approaches for Small and Tiny Object Detection in Aerial Images: A Review	2024	Small and Tiny Object Detection	Ours

4. This survey work stands out since it simultaneously reviews many performances of image augmentation approaches in terms of small and tiny object detection.
5. Finally, discussed challenges, future direction, and conclusion.

This review paper is organized into four sections. In Section 1, it summarizes previous reviews and survey papers concerning small and tiny object detection (see Table 1) as well as data augmentation papers within various application domains (see Tables 2, 3,4) and and explains how this paper differs from them. Furthermore, Section 2 thoroughly reviews existing operation libraries for geometric and photometric transformation, as well as basic and state-of-the-art approaches of image augmentation, discussing their advantages and disadvantages. Next, Section 3 presents the performance and application of various image augmentation approaches for small and tiny object detection task. Successively, it discusses challenges and future research direction in this domain.

2 Image augmentation approaches

For training a deep learning model, data is a fundamental and essential component. Deep learning performance limitations are occasionally attributable to limited training data. Therefore, to ensure optimal performance of any deep learning model, a substantial amount of data must be used for training [1]. There are a lot of issues in gathering large amounts of data such as cost, time consumption, limited availability, and privacy concerns. So, image augmentation is the solution to enlarge the datasets to overcome the issue of limited data. Image augmentation uses existing data to create modified copies of datasets, which are then used to artificially expand the datasets. It entails modifying the dataset only slightly or creating new points of information using deep learning techniques. In other words, image augmentation approaches are used to increase the amount of the dataset needed by deep learning models. This is achieved by intentionally modifying existing real images within the dataset. The objective of augmentation is to increase the number of class instances while maintaining the original characteristics. This data can be used for both training and testing of the model. By and large, the performance of deep learning models depends heavily on the availability of a significant amount of data. For instance, the performance of small and tiny object detection can be enhanced by increasing the dataset's instances and images as shown in Fig. 1.

2.1 Existing libraries for geometric and photometric transformations image augmentation operations

Traditionally, geometric and photometric transformations in image augmentation are time-consuming, tiring, and humanoid tasks. These restrictions make picture and video editing tools hardly ever employed in real-world situations. In deep learning models, there are specially designed code libraries such as Augmentor [39], Imgaug [40] and Albumenatation [41] used for image augmentation to make this task efficiently. Each DL framework comes with a set of augmentation techniques, and libraries such as TensorFlow [42], MXNet [43], Caffe [44], Keras [45], PyTorch [46]. This paper has enlisted some of the frequently used packages in Tables 5 and 6 with specifications. Additionally, there are several software tools like Octave [47]), MIPAV [48], LabView [49], Matlab [50], Mathematica [51] and ERDAS Imagine [52] are being used for image transformation.

2.2 Performance evaluation metrics

Performance evaluation metrics are critical in computer vision tasks because they provide a comprehensive assessment of a model's accuracy, efficiency, and overall effectiveness. Key metrics typically include:

2.2.1 Average precision

Average Precision (AP) is the evaluation metric that is the most frequently used in research studies. It is the most popular metric among the various labeled datasets like COCO and DIOR for the evaluation of computer vision tasks like object detection. It is mostly used to measure the accuracy of detection. Before defining the Average precision, its main concepts should be reviewed as its variation shares these concepts frequently.

- True Positive (TP) predicts correct objects in the bounded boxes.

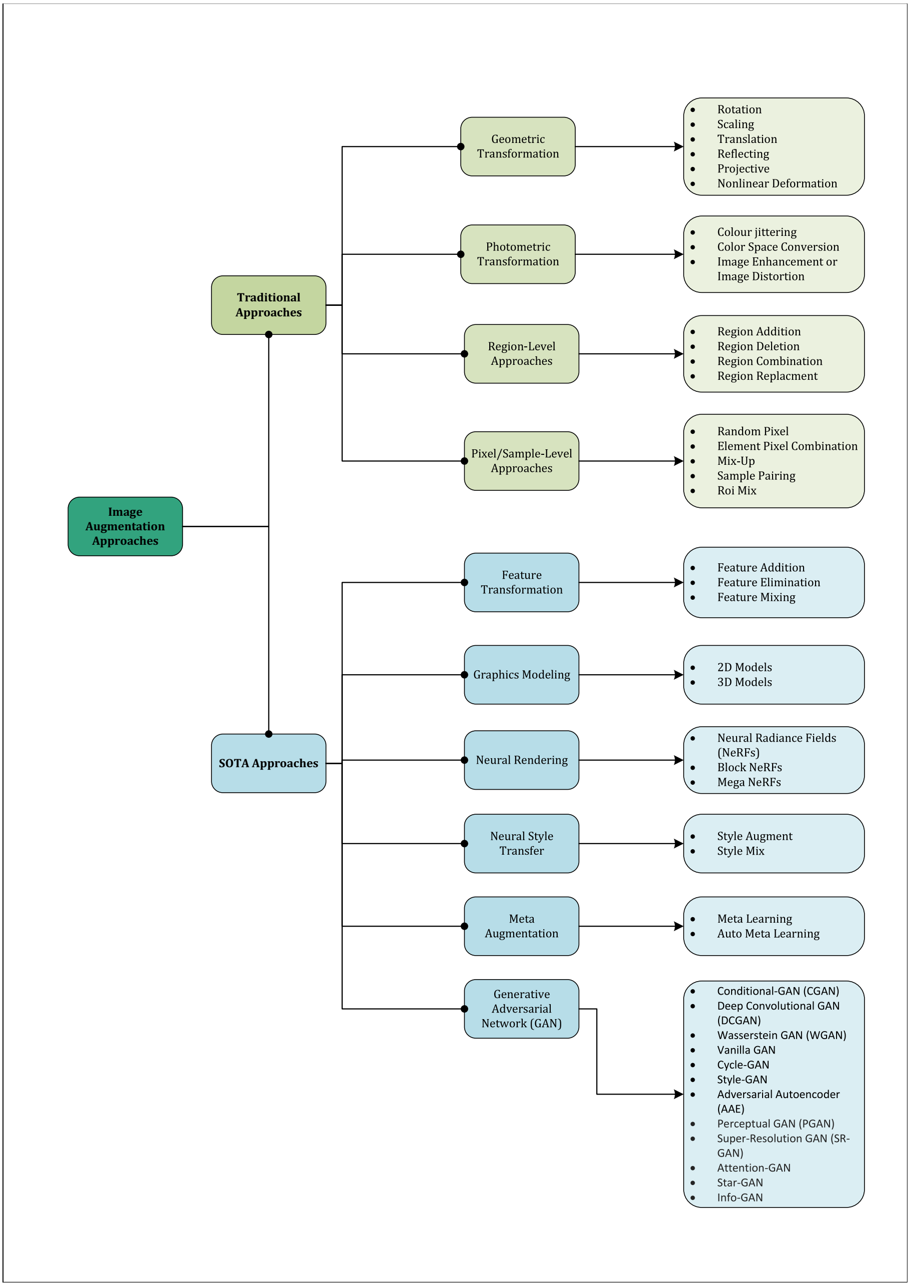

Fig. 1 The taxonomy of image augmentation approaches

Table 5 Summary of image augmentation libraries

Package	Code	Specification
Scikit-image	From scikit image import transform	• open-source library • scipy.ndimage • scientific image analysis • image transformation
Augmentor	Import Augmentor	• Image augmentation library • Image augmentation library for ML and DL models • A fast and quick library for performing various transformation tasks.
Torchvision	Import Trochvision.Tranformer as T	• Open-source Python machine learning framework •Common image transformation augmentation techniques for all CV tasks.
Albumentations	Import albumentations as A	• Various methods for image augmentation • Applied to all CV tasks.
Imgaug	From imgaug import augmenters as iaa	• supports various augmentation techniques. • Powerful interface that can augment images. • Used for bounded boxes, segmentation task.
OpenCV	Import cv2	• Powerful library for image augmentation. • Used for all CV tasks. • Availability of geometric and photometric transformation methods.

- False Positive (FP) predicts outcome incorrectly like missed objects or mistaken identity of objects.
- False Negative (FN) predicts undetected bounded boxes. Bounded boxes that are hard to detect in any image.

$$AveragePrecision(AP) = \int_0^1 p(r)\,dr \tag{1}$$

$$mAP = \frac{1}{N}\sum_{n=1}^{N} AP_n \tag{2}$$

Equation 1 calculates the Average Precision (AP), which represents the average precision of relevant items among the top-ranked ones in a given list. It assesses the system's retrieval performance by considering precision values (p) at different rank positions (r) ranging from 0 to 1. The AP formula aggregates these precision values across various rank positions, yielding a single score that reflects the system's object detection capability. On the other hand, equation 2 represents the mean Average Precision (mAP) for object detection evaluation. Its parameters include "N," representing the total number of classes, and "Average Precision (n)," which

Table 6 Summary of image augmentation libraries (continued)

Package	Code	Specification
ImageDataGenerator	from tensorflow.keras. preprocessing.image import ImageDataGeneratoren	• Specially built library for image augmentation by Keras
		• Produces tensor image data in batches while adding real-time data augmentations.
SOLT: Streaming Over Lightweight data Transformations	from kafka import KafkaConsumer	• Image augmentation library for DL.
		• Quick and integrated with PyTorch.
Kornia	import kornia	• Offers a variety of image augmentation functions.
		• Built to make use of GPUs for faster processing.
		• Differentiable augmentations are supported for efficient optimization and regularisation.
Open3D	import open3d as o3d	• 3D data processing and visualization.
		• Basic photometric transformations on 3D data.
Pillow	• from PIL import Image	• Supports a wide range of techniques such as rotation, flipping, cropping, and resizing.
	• from PIL import ImageEnhance	• Provides a simple and intuitive interface for performing image augmentations.
	• from PIL import ImageOp	• Allows for fine-grained control over image augmentation parameters.

is the Average Precision value for each class "n.". With these above-mentioned concepts, the definition of Average Precision (AP) is the mean detection of precision under various recalls. The variation of AP which is mostly used is mean Average Precision (mAP) which determines the overall AP shown in (2).

2.2.2 Precision and recall

Equation (3) presents precision. This spots pertinent instances, and it measures the mean accuracy of correct predictions. On the other hand, (4) represents recall which locates all bounding boxes, and measures the percentage of correct positive predictions among all given ground truths. According to the formula :

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{3}$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{4}$$

2.2.3 F1-score

The F1 -score is widely used to evaluate the performance of models by detecting as many positives (recall) as possible and ensuring the detections are accurate (precision) as shown in (5).

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

The F1-Score is essentially the harmonic mean of precision and recall. It tends to skew toward the lower number and makes it more sensitive to the lower value between precision and recall. The 2 in the numerator of the formula ensures that if either precision or recall is 0, the F1-Score will also be 0. It effectively balances the two metrics, so that improving either precision or recall will improve the F1-Score. Overall, it forces the model to optimize both detecting all relevant objects (recall) and minimizing false alarms (precision).

Above metrics are not just useful for evaluating the performance of models but are also critical for model selection, the optimization process, benchmarking for comparing the model's performance, etc.

2.3 Basic image augmentation approaches

2.3.1 Geometric image augmentation

Geometric transformations are the basic approaches to image augmentation. These transformations move the pixels from their initial place without changing the value of pixels in an image. These transformations are further divided into two classes affine and non-affine transformations. Affine transformations need a small number of parameter modifications to complete a variety of augmentation actions. The most common actions of affine transformations are rotation, translation, shearing, zooming, cropping, reflection, scaling, resizing, rotation, etc. On the other hand, non-affine transformations convert input image points to equivalent ones in a differing view frame using a pre-established relationship. Examples of these transformations are projective transformations, and nonlinear deformation. See Reference 7

2.3.2 Photometric image augmentation

Photometric image augmentation techniques alter an image's pixel content while maintaining its spatial composition as well as numerous of these methods alter the RGB channel composition to fit certain parameters. Initially, photometric modification is accomplished by altering the image's visual characteristics, including colour, brightness, sharpness, contrast, and saturation level. Color jittering, colour space conversion, and picture enhancement and distortion techniques are some of the most often utilized photometric data augmentation methods. See table as reference 7

2.3.3 Region-level image augmentation

Region-level augmentation approaches transform some areas of the image rather than the complete image. The main objective of region-level approaches is to alter particular characteristics like the location, and content of the sub-regions of the input image so can to avoid the overfitting problem of DL models. These approaches are beneficial for all computer vision tasks, especially object detection tasks. The evidence will be presented in Section 3 in detail. These approaches are further divided into Region-deletion, Region-Replacement and Region-Combination. Region deletion is basically a data disposal approach that is used to delete some information at the input level. Region deletion approaches remove a few random locations from the original images during training, so a variety of visual elements can be

Table 7 Advantages and disadvantages of augmentation approaches

Approaches	Advantages	Disadvantages
Geometric Augmentation	• Due to their simple yet effective nature, they are a popular choice as the primary form of augmentation in many computer vision tasks, frequently preceding the use of more sophisticated techniques.	• The task of determining the best strategy for data augmentation presents a formidable challenge.
	• Enables the creation of diverse image variations, expanding the training dataset and enriching the learning process.	• If the original dataset contains inherent biases, the augmented data is subject to the same biases, exacerbating the complexity of data augmentation.
	• Facilitates the discovery of new patterns and features in images, enhancing the performance of machine learning models.	• It is critical to develop evaluation mechanisms to assess the quality of augmented datasets. This is a critical task because it is critical to ensure that the augmented data is reliable and verifiable.
	• The ability to increase model robustness to image variations such as lighting and orientation ensures more consistent and accurate predictions across different scenarios.	
Photometric Augmentation	• Deep learning models' resilience to image variations caused by environmental conditions and imaging devices can be improved by using photometric data augmentation, which involves applying various transformations to the data, resulting in improved image recognition and classification performance.	• The need to ensure that the parameters and methods align with the network application.
	• Photometric data augmentation can increase the diversity of the training data by generating additional training images with different lighting conditions and camera settings, resulting in better generalization of the model.	• Objects may become indistinguishable from the background, potentially misguiding the network during training.

added to the training images dataset [53–57]. These approaches can be helpful in those tasks where objects are occluded. Most frequently used augmentation techniques are Cutout [53], Random Erasing [57], Hide-and-Seek [54], Gridmask [58] GridCut and mix [59] . The cutout technique removes areas randomly from the input image by using square matrix that makes the pixel value zero. Random Erasing also arbitrarily chooses a rectangle area in the image and then erases the pixels in that area with random values. Region-aware Random Erasing [55] is an extension of the Random Erasing approach. Hide-and-Seek (HaS) makes the input image into grids and removes random areas by awarding zero value to pixels of that image.

Gridmask and GridCut are two advanced methods of region-level image augmentation. These methods create a grid and change the region pixels. Region-replacement methods cut out some areas of the input image, make some alterations, and paste them back arbitrarily. Cut-Paste [60], CutMix [61], CutBlur [62], PatchShuffle [63], Cut-Thumbnail [64] are some examples of Region-replacement augmentation approaches. Cut-Paste and Cut-Thumbnail methods outperformed all the previous methods by alleviating the problem of loss-information. There are some advanced methods of Region-Replacement such as FocusMix [65], Attentive Cut-Mix [66], AttributeMix [67]. These literature methods are used to copy and paste patches semantically. Region-combination image augmentation approaches use information of the patches from the same or different input image and combine this information pixel-wise or fuse smaller patches of different images to make a new sample image. Random Image Cropping and Patching (RICAP) [68], Patch-Level Augmentation [69], SaliencyMix [70], PuzzleMix [71], SuperMix [72], TokenMix [73], Self-Replacement-and-Keeping-Augment [74] are some of the approaches combine different or same images' patches to augment input samples. The advantages and disadvantages of this approach have been shown in the Table 8. Table 9 provides an overview of multiple approaches, each supplemented by brief descriptions to facilitate understanding.

Table 8 Advantages and disadvantages of augmentation approaches

Region-Level Image Augmentation	•Region elimination is a useful technique for creating the effect of occlusion in computer vision tasks.	• Region deletion approaches have a limitation in their random deletion process, which can lead to the deletion of important features, thus negatively affecting performance.
	• Occlusion can occur when tracking objects visually due to their interaction with other objects or changes in the environment. Elimination of regions can help to account for this.	• The random nature of the deletion process does not allow for the selection of appropriate regions for deletion.
	• Information dropping is a technique for removing unnecessary visual features while retaining the more useful ones.	• These techniques may not be applicable or effective for all types of images or object classes. This may limit the generality and effectiveness of the model in certain scenarios.
	• This technique is particularly useful when appearance information is not critical to the recognition task at hand.	
Pixel-Level /Sample Level Augmentation	• It is effective in combating adversarial attacks, which are deliberate attempts by malicious actors to trick or deceive deep learning models.	• The introduction of artificial, out-of-distribution samples into the original dataset is a limitation. This can happen when the mixed images are too dissimilar to the original dataset, resulting in poor model performance and accuracy on new and unknown data.
	• It has the ability to improve deep learning model interpretability by generating augmented images that highlight specific features or patterns in the input data.	• It can make interpreting the model's behavior and identifying areas for improvement more difficult.

Table 9 Region-Level augmentation approaches

Region replacement approaches	Brief description
PA-AUG	Replaces a region of an image with a patch from another image with similar content.
CutBlur	Replaces a region of an image with a blurred version of that region.
CutMix	Replaces a region of an image with a patch from another image, and blends the two images.
CutPaste	Replaces a region of an image with a patch from another image, and preserves the original context.
AttributeMix	Replaces a region of an image with a patch from another image, and preserves the original attributes (e.g. color, texture).
Attentive CutMix	Similar to CutMix, but uses an attention mask to select the patch from the other image.
Region Swapping Approaches	Brief Description
InPS	Replaces a region of an image with a patch that has a similar texture.
CutMix	Replaces a region of an image with a patch from another image, and blends the two images.
PatchShuffle	Divides the image into patches, shuffles them, and reassembles them.
PuzzleMix	Divides the image into square patches, shuffles them, and reassembles them.
Region Combination Approaches	Brief Description
MixUp	Combines two images by linearly interpolating their pixel values.
GridMix	Divides the image into a grid and replaces each cell with a patch from another image.
RICAP	Randomly crops two or more images, and pastes them together in a new image.
SaliencyMix	Replaces a region of an image with a patch from another image, and preserves the original saliency (i.e. importance).
SuperMix	Combines two or more images by blending their saliency maps and pixel values.
Region Deletion Approaches	Brief Description
Cutout	Randomly cut out a square region of an image.
Random Erasing	Randomly erases a rectangular region of an image and replaces it with noise or another background.
Hide and Seek	Deletes multiple, smaller square regions from the image.
KeepAugment	Similar to Cutout, but instead of replacing the cut region with a background color, it keeps the surrounding pixels and fills the cut region with the average color of the image.
GridCut	Divides the image into a grid and randomly removes one or more cells.
Grid Mask	Similar to GridCut, but instead of removing cells, it masks them with random noise.
FenceMask	Removes a rectangular region of an image and replaces it with another image.

2.3.4 Pixel-level/ sample level image augmentation

Pixel-level image augmentation approaches combine the pixels of two or more images by creating samples by manipulating the sensitivities of the chosen pixels of images. The objective of these approaches is to create new sample data by applying pixel-wise adjustments and modifications to the given image dataset. Table 10 shows some of the examples like Mix-Up [75], CoMixup [76],Sample-Pairing [77], Smooth-Mix [78], Ada-MixUp [79], Mixup without hesitation [80], RoiMix [81], Augmix [82], Noisy Mixup and Random Pixels [83], Smart Augment [84]. For a overview of the advantages and disadvantages associated with this approach, please see Table 8.

2.4 State-of-the-art image augmentation approaches

2.4.1 Deep feature transformation

Deep Feature Transformation (DFT) is a technique used to enhance the performance of computer vision models by augmenting existing images with additional features. This technique is used to increase the diversity of the data set, which can help the model better generalize to new data. DFT works by transforming the image data into a higher-dimensional feature space. This transformation is achieved by applying a set of convolutional filters to the image. The resulting feature space contains more complex representations of the data that are more meaningful to the model. These can be divided into feature addition, elimination, and combination. Addition approaches include increasing the amount of an existing dataset by adding additional features, which gives the model more data to learn from. Some of the examples are Attention Augment [85], interpolation methods [53, 86–88]. Another strategy is to use attention or feature fusion techniques to merge features taken from various input samples like smart augmentation [67] creating an enriched feature vector. By giving the net-

Table 10 Pixel-level/ Sample Level Image augmentation

Approaches	Brief description
Mix-Up	Creates new samples by linearly interpolating between pairs of samples and their labels
CoMixup	A variant of Mix-Up that combines two or more input images by interpolating the corresponding feature maps.
Sample-Pairing	Augments the training set by pairing each sample with another sample from the same class.
Smooth-Mix	A variant of Mix-Up that uses Gaussian noise to smooth the interpolation weights Ada-MixUp A variant of Mix-Up that adapts the interpolation weights based on the difficulty of the samples.
Noisy Mixup and Random Pixels	A variant of Mix-Up that adds noise to the interpolated samples and randomly changes some of their pixels in purpose.
Mixup without hesitation	A variant of Mix-Up that uses a deep reinforcement learning framework to learn when to apply Mix-Up.
RoiMix	Applies Mix-Up to the region of interest (RoI) of object detection datasets Element-wise pixel combinations combine the corresponding pixels of two or more images using element-wise operations.
Augmix	Uses a cascade of augmentation operations to produce augmented images.

work additional robust and discriminative characteristics, these techniques hope to boost its performance by enabling the network to handle variances and generalize to new or undiscovered inputs. A prominent regularization method for enhancing deep learning models is feature elimination (removal). To enhance the overall performance of the model, this method involves removing specific characteristics from the intermediate convolutional neural network (CNN) layers. Mixout and its derivatives [89, 90], Shakeout [91], and Zoneout [92], etc, are commonly used to randomly delete a subset of features and their synaptic connections from intermediate layers. Table 11 outlines the various benefits and drawbacks, providing a clear comparison to help understand the impact of the approach.

2.4.2 Digital visualization

A potent method for creating synthetic data that can be used to train computer vision models is computer graphics modelling or digital visualization. For this, complex 2D and 3D models of the objects can be produced using CAD tools. It is possible to produce 3D data using CAD modelling techniques, which can then be rendered and rasterized into 2D photos or used to train models for 3D visual recognition tasks. For the purpose of training deep learning models across multiple domains, a large number of substantial datasets like shapenet [93], SceneNet RGB-D [94], Looking beyond appearances [95], Hypersim [96] developed utilizing CAD modelling techniques are accessible. More and more, data synthesis and visual augmentation are being done using computer graphics modelling approaches based on 3D gaming engines.

Table 11 Advantages and disadvantages of augmentation approaches

Deep Feature Transformation	• Manipulating feature vectors or generating new features is more appealing than manipulating entire images because it is easier and less computationally expensive.	• These are not based on domain knowledge, which means that it may miss important attributes that are critical to some tasks.
	• Changing features at the vector level also allows for smoother transformations than modifying the input images.	• These are only applicable to image data, so they cannot be used with other types of data.
	• Combining feature space techniques with traditional data augmentation techniques can help improve performance.	• It is important to use caution when adding too many features since this can lead to overfitting, where the model becomes too specialized to the training data and performs poorly on new data. Computer Graphic Modeling
Computer Graphic Modeling	• More advanced information about the objects being represented can be included by using computer-generated 3D datasets, allowing for the solution of more complex problems.	•The computer-generated images may not accurately reflect the real-world data, leading to discrepancies between the model and the actual environment.
	• It is possible to create scenarios that would be difficult or impossible to create in the real world, allowing for more thorough testing and training of computer vision systems.	• The cost and time required to create high-quality CGM can be high, making it difficult to scale up or apply to large datasets.

With the help of these techniques, larger, more intricate scenes with interactive elements may be created, resulting in datasets for more difficult tasks like augmented and virtual reality, autonomous driving, and semantic scene interpretation in complex contexts [97–101]. For the advantages and disadvantages of this approach, please refer to Table 11. This table illustrate the various benefits and drawbacks associated with the approach.

2.4.3 Neural rendering

To produce lifelike photos and videos from 3D models or settings, the technology known as "neural rendering" uses deep learning algorithms. In order to teach a neural network how to replicate realistic lighting, shading, texture, and other visual properties, a huge dataset of images must be used during the technique's training phase. According to [102] the goal of neural rendering is to utilize deep neural networks to acquire knowledge about the 3D scene and obtain parameters that can be used to complete the rendering process. These methods have created large-scale datasets like RTMV [103] and Synscapes [104], focused on complex applications such as outdoor scene understanding and autonomous driving. In Table 12, the advantages and disadvantages of this approach are outlined, and Table 13 presents a summary of state-of-the-art methods, highlighting the approaches currently utilized by researchers.

Table 12 Advantages and disadvantages of augmentation approaches

Neural Rendering	• These approaches can produce highly realistic synthetic images with a wide range of variations, such as lighting, pose, and texture, which may not be possible with traditional data augmentation techniques.	• The utilization of neural networks to execute this technique was an arduous task as a result of impediments arising from restricted computational resources.
	• Generating synthetic images can significantly reduce the need for manual data collection, which can be expensive and time-consuming.	• This approach suffers from the disadvantage of demanding extensive memory, thereby constraining the resolution of pixels and the ability to handle voluminous scenes.
	• Synthetic images can be labeled with high accuracy and consistency since the ground truth information is available by design, unlike real-world images that require manual annotation.	
Generative Adversarial Networks (GANs)	•Generative modeling techniques are commonly utilized to accomplish domain adaptation tasks, which involve adapting a model trained on one domain to perform well on a different, but related, domain.	•These are vulnerable to adversarial attacks, where small changes to the input can cause large changes in the generated output.
	•GANs are capable of learning complex data distributions that may be difficult to model with other machine-learning techniques.	•Require high computational resources to train and generate data, making it difficult for small-scale applications.

Table 13 Some state-of-the-art methods of Neural Rendering

Method	Description
Neural Radiance Fields [105]	Generates photorealistic 3D scenes by modeling light transportation through different points in a 3D space.
Block-NeRF [106]	Generates very large dynamic scenes of outdoor environments in the form of videos from a collection of sparse images by dividing scenes into smaller units and employing multiple networks to process different scene parts.
Mega-NeRF [107]	Capable of generating high-quality, large-scale 3D scenes by leveraging efficient hierarchical representations and multi-resolution voxel grids.

2.4.4 Generative adversarial network (GAN)

The GAN [108] approaches [109] are used for generating synthetic image data by implementing two sub-approaches: a: to generate realistic data and b: to distinguish between real and artificially generated data. Figure 2 provides a visual representation of the GAN concept, elucidating the workings and components of the Generative Adversarial Network. Ref 12

The following Table 14 shows some of the example papers of the GAN with their brief description.

2.4.5 Neural style transfer (Fig. 3)

Neural style transfer (NST) [125] is a method that creates new images by blending features from different layers of deep convolutional neural networks (CNNs). The technique aims to produce images that have similar visual characteristics as a reference image while preserving the underlying structure of the original image. Several techniques for improving the visual quality of stylized images have been proposed, such as using whitening and colouring transforms [126] and introducing training loss terms to optimize for photorealism [127, 128]. These methods are designed to reduce image distortions and improve the overall photometric quality of generated data. Some approaches [129] emphasize memory efficiency when stylizing high-resolution images. Traditional style transfer methods have a limitation in that they primarily focus on texture while ignoring image shape. StyleAugment [130] and StyleMix [131] are style transfer techniques that prioritize shape as well as texture to address this shortcoming. Reference Table 15 lists the pros and cons of this approach.

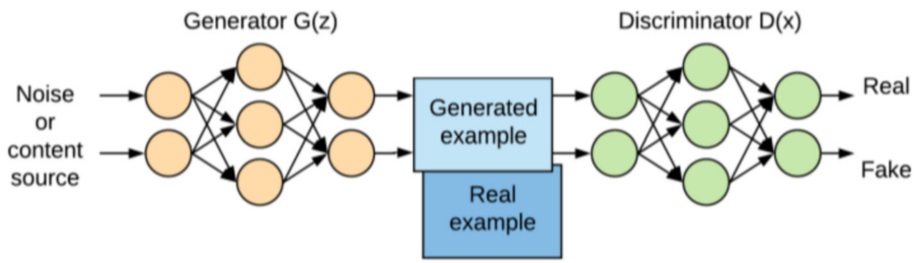

Fig. 2 Visual representation of the Generative Adversarial Network (GAN) concept as explained by [14]

Table 14 Some state-of-the-art approaches of Generative Adversial Network (GAN)

GAN Approaches	Brief Description
Conditional-GAN (CGAN) [110]	Uses additional input for conditional information for generating images with specific attributes.
Deep Convolutional GAN (DCGAN) [111]	Uses convolutional layers in the generator and discriminator networks for generating high-quality images.
Wasserstein GAN (WGAN) [112]	Uses Wasserstein distance for generating diverse and realistic images.
Vanilla GAN [113]	Basic GAN architecture for generating simple images.
Cycle-GAN [114]	Image-to-image translation between two domains, used for generating images in with different attributes.
Style-GAN [115]	Generates diverse and high-quality images with controllable style and structure.
Adversarial Autoencoder (AAE) [116]	Uses autoencoder as the generator network for generating images.
Super-Resolution GAN (SR-GAN) [117]	Generates high-resolution images from low-resolution inputs.
Attention-GAN [118]	Uses attention mechanisms to focus on specific regions of the input images to generate high-quality dataset.
Star-GAN [119]	Image-to-image translation between multiple domains with a single generator and discriminator, used for generating images with multiple attributes or styles.
Info-GAN [120]	Learns interpretable and disentangled representations of images, allowing control over specific attributes without labeled data.
BigBi-GAN [121]	Learns a bi-directional mapping between images and their corresponding text descriptions, used for generating images from text and vice versa.
ESR-GAN [122]	Combines perceptual loss with adversarial loss for improved visual quality in super-resolution tasks.
LS-GAN [123]	Uses mean squared error as the adversarial loss, used for generating high-quality images with sharp edges.
VQ-GAN [124]	Uses vector quantization to learn a discrete representation of images for generating images dataset.

2.4.6 Meta augmentation (Fig. 4)

Meta augmentation is a method of equipping machine learning models with the ability to acquire valuable concepts that assist in the extension of acquired knowledge to new tasks. It has two categories in terms of augmentation:

- Meta-learning techniques use meta-data to train deep learning models that can generalize across tasks and adapt to new ones. These methods incorporate data augmentation and few-shot learning methods to improve sample efficiency [132–136]. Meta-learning surpasses transfer learning in flexibility, particularly during training phases. Extracting meta-knowledge greatly enhances the efficiency of few-shot learning in testing environments [137].

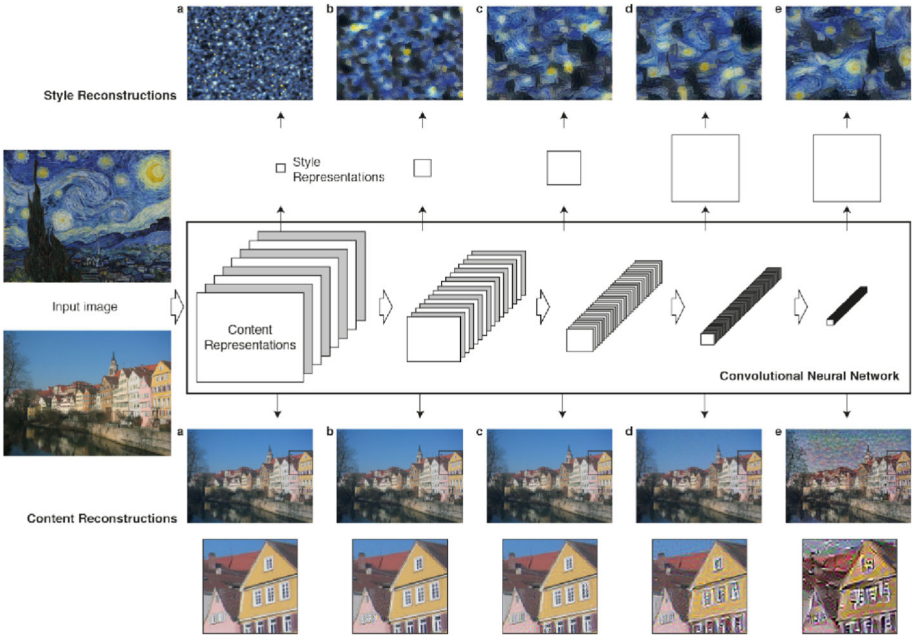

Fig. 3 Visualization of the reconstructions of style and content in Neural Style Transfer [125]

- AutoML-based data augmentation [138–145] is a meta-learning technique that employs optimization methods to learn the best model hyper-parameters for improved generalization. This method automates the process of creating effective data augmentations and training a neural network for optimal task performance. AutoML approaches facilitate the creation of systems without requiring extensive expert knowledge and prolonged research. These approaches are employed to automate the design of neural networks, aiming to identify the best architectures for specific problems.

Ref 15

2.5 Definitions of small objects

The small objects in an image are referred to by how many few pixels they occupy. Computer vision researchers provide various definitions depending on various applications and environments. Liu and Mattyus [147] concluded that vehicles take few pixels in images that is why it is difficult to detect. According to [148] all object instances are no larger than 0.2% of the image size, and a sizable proportion of them are between 0.1% and 0.15%. Another researcher [149] stated that DOTA objects are separated into three categories and small and tiny objects range in height of bounded boxes between 10 and 50. By definition, tiny objects are less than 16 x 16 pixels in size. As compared to larger objects small objects have low pixel count, do not interpret, and easily fuse with their backgrounds [3]. Instances that occupy less than 50 pixels in terms of height and width are considered small objects [150, 151]. Figure 5 shows some representation of small objects in the aerial images.

Table 15 Advantages and disadvantages of augmentation approaches

Neural Style Transfer	• It leverages the capabilities of a neural network to extract significant visual characteristics from an image that embodies a particular artistic style and subsequently applies these features to another image that lacks the same style. This process results in a transformation of the second image, which now bears the artistic appearance of the original image.	•The generated dataset images' quality may not always be consistent or realistic.
	• Neural Style Transfer (NST) can provide an advantage in image restoration or enhancement by leveraging the ability to transfer the style of a high-quality image to a low-quality or degraded image. This process enables the low-quality or degraded image to adopt the visual appearance of the high-quality image, resulting in improved visual quality and the restoration of some of the lost details.	
Meta Augmentation	• It enables few-shot learning, particularly useful when data is scarce or expensive to obtain.	• It may introduce noise or variability into the training data.
	• It expands the support dataset and range of tasks to diversify data for each class and create more diverse tasks for each training sample. This enhances the model's adaptability to new tasks and improves its overall performance.	• It necessitates substantial computational resources, such as processing power, memory, and time.
		• To implement effectively, a high level of skill and expertise is required, including knowledge of machine learning algorithms, data processing, and computer programming.

2.5.1 Applications of image augmentation in small object detection

2.5.2 Geometric transformations

Geometric image augmentation enhances and alters images by applying geometric transformations. These transformations involve modifying the spatial properties of the image, such as its position, size, rotation, and perspective. It is used to increase the number of images in the dataset. In the field of object detection, the focus is primarily on locating and identifying objects that appear in various positions within an image. When it comes to aerial image detection, there is a particular emphasis on detecting small objects. During the training phase, incorporating random or group of geometric augmentation approaches enables the generation of an augmented dataset, which greatly benefits the study of small and tiny object detection. By applying transformations like scaling, rotation, translation, and affine transformations to

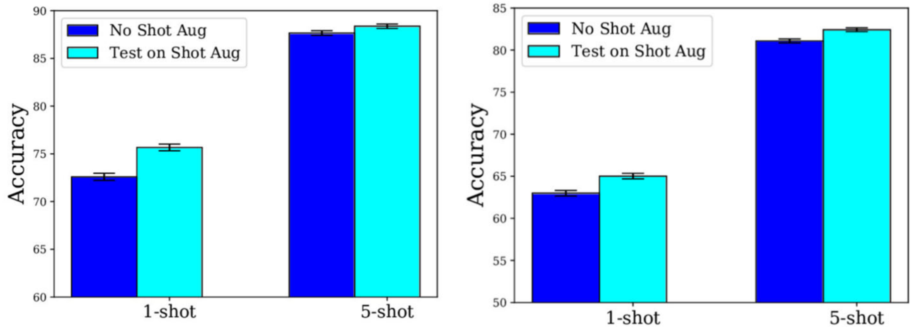

Fig. 4 Achieving enhanced performance through shot augmentation employing MetaOptNet, trained with the proposed Meta-MaxUp technique. The top results showcase 1-shot and 5-shot outcomes on CIFAR-FS, while the bottom results exhibit 1-shot and 5-shot performance on mini-ImageNet [146]

the training images, the dataset becomes more varied and better reflects real-world situations. The researchers [152] proposed two augmentation approaches to address the issue of poor performance in detecting tiny objects due to a lack of such objects in the training set: Firstly, oversampling images with tiny objects during training: By randomly selecting images that contain tiny objects and oversampling them during training, the model can learn to detect tiny objects more effectively. Secondly, copy-pasting tiny objects from segmented masks: This approach involves extracting tiny objects from images using a segmented mask and then copying and pasting them into new images. These new images are then added to the training set, providing the model with more examples of tiny objects to learn from.

One notable approach, as discussed in [153], involves the use of horizontal flipping for data augmentation while maintaining the original size of the image or sub-image. This method avoids any zooming, thereby preserving the integrity of the image data. To further tackle the issue of varying object sizes and to prevent information loss for smaller objects, a scale-matching strategy was implemented. This technique involves cutting images based on the object size to minimize the disparity between objects of different sizes. Another significant contribution in this field is from researchers [154], who observed that the Single Shot Detector (SSD) faces challenges in classifying small objects without resampling. They introduced an innovative data augmentation strategy that significantly improves performance, especially in small datasets. This strategy includes generating random crops as a "zoom-in" operation and a "zoom-out" operation, creating smaller training examples. This augmentation trick not only increased training iterations but also consistently improved the mean Average Precision (mAP) by 2%-3%. To address the UAV object detection challenge due to numerous small

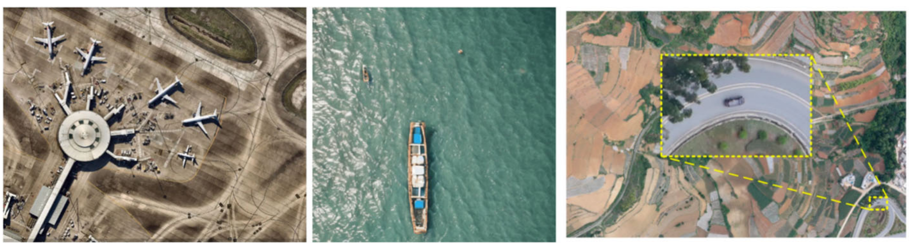

Fig. 5 Examples of some small objects in aerial images

objects in the images, [155] used a memory-efficient strategy employed during training by dividing the images into four cropped parts. This not only conserves memory but also expands the dataset, enhancing the network's sensitivity to small objects by emphasizing their visibility in the enlarged images. During the inference phase, performance is further boosted through the application of test time augmentation (TTA). This involves scaling the image and conducting tests at multiple scales, while also introducing random flips. The resulting predictions from the differently scaled images are merged using non-maximum suppression (NMS) to enhance overall detection accuracy. A novel geometric augmentation, termed "jumble-up," was introduced by [156]. This method, inspired by jigsaw puzzles, involves the non-intersecting split and swap of image parts horizontally and vertically. It aims to enhance the robustness of the YOLOrs network, particularly in scenarios with limited labeled remote sensing (RS) training datasets. Furthermore, [157] demonstrated that incorporating "Crop" augmentation for black edge padding can further enhance performance, especially for rare classes. In another study, [158] researchers incorporated density crops as a new class during training to augment the dataset. This innovative approach involved detecting objects from this new class alongside the base class objects during testing. This method not only enhances the dataset but also improves the model's accuracy in small object detection. It represents a strategic move to address the challenges of limited training data and the need for improved detection capabilities. Additionally, [159] utilized various data augmentation approaches to enhance the training process of their machine learning model. They applied random horizontal flipping with a probability of 0.5 to the training data, a technique that has become relatively standard in image-based machine learning tasks. Beyond this, they employed random cropping and expanding techniques to further increase the diversity of the training data. Their efforts yielded significant results, as evidenced by their impressive APtiny50 score of 71.53%, securing them the second position in their challenge.

2.5.3 Photometric transformations

One notable approach is AUGMIX, introduced in a study [160], which employs a range of photometric transformations applied randomly. This method is designed to enhance the model's ability to detect small objects by improving its robustness to variations in image quality and object appearance. The AUGMIX technique is particularly effective in scenarios where small objects are present with limited details or are partially occluded, making them difficult to detect using conventional methods. Another significant contribution in this field is the augmented small object detection method proposed in [161]. This method utilizes a Neural Architecture Search (NAS-FPN) based feature pyramid architecture. It incorporates various transformations such as scaling, contrast enhancement, flipping, brightness alteration, and random angle rotation. These augmentations are specifically tailored to improve the detection effectiveness of small objects, which often get lost or misclassified due to their size and the lack of distinct features. Study [162] focuses on enhancing the diversity and robustness of the reference patch in their analysis by employing flipping and color jittering as data augmentation methods. These techniques are particularly useful in scenarios where the reference patch lacks variability, thereby limiting the model's ability to generalize across different instances of small objects.

2.5.4 Region-wise image augmentation

One innovative approach is the "Stitcher" approach, introduced in [163], which leverages feedback-driven resizing and stitching of images. This method, focusing on smaller objects, uses loss statistics as feedback to guide subsequent updates, enhancing detection accuracy. Similarly, [164] applied the same technique to tackle the unbalanced distribution of small and tiny objects, thereby boosting detection performance. The cutout approach, as used in [165], combines traditional random masking with targeted cutouts based on the extremities within ground truth boxes. This strategy aims to mitigate occlusion issues while preserving essential semantic information for tiny objects. In contrast, [166] employed a variety of image augmentation methods, including random flipping, Gaussian noise addition, and random cropping, erasing, and softening, specifically for enhancing features during the training stage of a Feature Enhancement Network. Further diversifying the augmentation landscape, [167] and [168] utilized cutout and mosaic augmentation approaches, with [168] specifically applying Random-Cutout image augmentation to simulate occlusion in wheat images. This was achieved by selectively erasing rectangles based on the number and size of wheat ears. Complementarily, [169] enhanced performance by combining Random Perceptive and Random Erasing techniques with Mosaic techniques. Addressing the challenges in gigapixel-level image object detection, [170] introduced a novel framework that combines a Patch Filter Module (PFM) for patch selection and a Patch Packing Module (PPM) for efficient patch arrangement. This approach significantly reduces computation time while maintaining detection efficacy. Similarly, [171] employed copy and restricted paste techniques, alongside three augmentation approaches, to enhance the relevance of objects in scenes for small object detection. In the realm of model training and convergence, [172] employed a combination of Copy-Paste and Mixup augmentation strategies to expedite convergence and smooth the training curve. This facilitated faster model execution. Additionally, [173] and [174] utilized combinations of Mosaic and Mixup approaches, with [174] specifically applying these in YOLOv3 for spatial transformations ((random flipping, cropping) and classification task blending, creating virtual training examples during model training. Zhou et al. [175] study employed the Mix-Up model for small object detection in aerial images, utilizing it for multiple reasons. By fusing two images with a specified ratio, the approach aims to enrich background information, enhance sample diversity, and improve prediction capability, while concurrently mitigating training fluctuations. In a similar vein, [176] upgraded the YOLOv4 algorithm by incorporating Mixup with Mosaic, introducing a multi-attention block for improved feature extraction, and adjusting loss functions for optimized small target detection. Lastly, [177] introduced Corner CutMix, a novel data augmentation approach for deep convolutional neural networks (CNNs). This method combats overfitting by randomly replacing a corner region in an image with a region from a distractor image during training. It also incorporates an auxiliary self-supervised loss function, enhancing the learned representation's transferability and generalizability.

2.5.5 Pixel-wise or sample-wise image augmentation

A notable study [178] explored two distinct augmentation techniques to enhance the generalizability of object detection algorithms. The first strategy involved substituting backgrounds to add negative samples, thereby diversifying the backgrounds encountered during training. This approach is grounded in the rationale that exposure to a wider range of backgrounds can prevent the model from overfitting to specific scenarios, thus enhancing its ability to generalize across diverse environments. To enhance the generalizability of object detection algorithm. The second strategy adopted by [178] was the incorporation of noise into the training process. This technique is designed to mitigate the risk of the model becoming overly tailored to the specific objects in the training set. By introducing noise, the model is encouraged to focus on the salient features of the objects, rather than the idiosyncrasies of the training data, which may not be present in real-world scenarios. Interestingly, the study observed that the simultaneous application of both these augmentation techniques did not yield as effective results as when they were employed individually. This finding is significant as it suggests that while each method independently contributes to improving the model's performance, their combined use might not be synergistic. In fact, it could potentially lead to a diminution of the effectiveness of the algorithm. In a different approach, study [179] adopts the Pixel Level Balancing (PLB) method, which aims to adapt the training loss with dynamic weight adjustments based on the detection box pixel count. This technique specifically targets the challenge of detecting small objects, a common issue in object detection tasks. By dynamically adjusting the weights in the loss function based on the size of the object, the PLB approach ensures that smaller objects receive adequate attention during the training process, thereby improving the model's accuracy in detecting them. The contrast between the strategies employed in [178] and [179] highlights the diverse methodologies that can be adopted in enhancing object detection algorithms. While [178] focuses on augmenting the input data to improve generalizability, [179] emphasizes the modification of the training process itself to address specific challenges such as small object detection. Furthermore, study [180] addresses the challenge of detecting small objects in images dominated by sky regions. In such scenarios, traditional methods like uniform random cropping often fail to capture crucial elements such as birds, while also misclassifying hard-negative samples like leaves and turbine edges. To overcome this, the study introduces a selective cropping strategy. This strategy involves randomly yet intensively sampling from areas containing key features, utilizing known bounding boxes to intensify the training focus on these important regions. This targeted approach ensures that the model is exposed to and learns from the most relevant parts of the image, thereby improving its accuracy in detecting small objects.

2.5.6 Generative modeling image augmentation

GANs are the preferred data augmentation approach due to their strong image generation capabilities, allowing for the creation of new target images and versatility across applications. However, they face challenges in terms of training complexity, high computational requirements, and occasional image quality issues. The use of GANs for effective training often involves supervised learning, but GANs themselves require a substantial amount of image data for optimal performance as deep learning models. This poses a paradox in the context of image data augmentation, as using GANs to generate additional target images assumes an initial availability of an ample quantity of target images. One innovative approach is the DS-GAN augmentation method [181], which enhances standard object detectors by boosting small object samples during training. This method avoids the introduction of unrealistic artifacts in resized images, leading to an increase in average precision (AP) by 4 to 10 points. Similarly, a novel end-to-end multi-task GAN [182] integrates a super-resolution generator with a multi-task discriminator for aerial image detection. The generator incorporates a gradient guide and edge-enhancement strategy, while the discriminator employs a faster region-based CNN for object detection with detection task losses back propagated into the generator during training for improved optimization. This approach upsamples small objects, enhancing their detectability without compromising image quality. The EESRGAN model [183, 184] represents a significant leap in this domain. It combines an Edge-Enhanced Super-Resolution GAN (EESRGAN) with a Detection network, employing residual-in-residual dense blocks (RRDB) for both GAN and Edge Enhancement Network (EEN) components. This architecture is integrated with Faster R-CNN and SSD for detection, where the detector loss is backpropagated into EESRGAN, synergistically enhancing image quality and detection performance. Rabbi et al. [184] used the same method with improvement for another aerial dataset. The Centered Multi-task Generative Adversarial Network (CMTGAN) [185] adopted that merges small object detection with gaze estimation. This method employs a GAN with a generator for high-resolution imaging and a discriminator for two-stage object detection, focusing on central image areas where small objects commonly appear. Stachoń and Pietroń [186] study adopted a multi-phase approach. It began with augmenting the VOC Pascal dataset using oversampling strategies and Generative Adversarial Networks (GANs), followed by employing the Faster R-CNN model for evaluation. Finally, a Perceptual GAN enhanced small object representation by generating super-resolved, large-object-like images, significantly boosting object recognizability and improving the performance of small object detection. This research [187] focused on improving small object detection in aerial and satellite images using a targeted super-resolution method within a neural network's training phase. The technique is efficient, necessitating fewer network layers, and is tailored to the object's scale and size. The network's learning utilizes deep residual blocks in a Wasserstein Generative Adversarial Network (GAN) framework, optimizing detection in remote sensing imagery. Courtrai et al. [188] study explores enhancing object detection in satellite and aerial remote sensing imagery by utilizing super-resolution to amplify spatial resolution, thereby enlarging object size and clarity. It details the development of a super-resolution framework, initially based on a Generative Adversarial Network (GAN) with residual blocks and its subsequent integration into a cycle model. Significantly, the incorporation of an auxiliary network specifically designed for object detection markedly boosts both the learning process and the efficacy of the super-resolution architecture, leading to notably improved object detection performance. The work of [189] enhances small object detection by integrating a GAN for unsupervised super-resolution with a jointly trained detection network. This approach, utilizing the CycleGAN strategy, increases detection accuracy without relying on low-/high-resolution image pairs. A novel framework proposed in [190] integrates a Super Resolution GAN (SRGAN) and a multitask Wasserstein GAN (WGAN) with a standard detection model like RetinaNet. This framework enhances global image resolution before detection, while the WGAN, applied to anchor regions post-detection, improves local resolution. The unique aspect of this approach is the WGAN's ability to perform generation, classification, and regression tasks, thereby enhancing detection accuracy by refining the generator with classification and regression loss. This method has shown effectiveness in reducing false positives and improving bounding box accuracy on the DOTA aerial image dataset. The Classification-Oriented Super-Resolution Generative Adversarial Network (CSRGAN) [191] represents another innovative approach in UAV aerial imagery. By integrating a classification branch into the SRGAN framework and employing VGG19-based content loss, CSRGAN effectively reconstructs high-quality images from low-resolution inputs, improving object contours and category discrimination to solve small object detection issues. In addressing the challenges of object identification and recognition in remote sensing images, a hybrid approach combining GANs and Convolutional Neural Networks (CNNs) [192] leverages the synthetic data generation capability of GANs to enhance limited labeled data. This method accommodates the diverse backdrops and variances typical in these images, while the integration of CNNs ensures precise and robust small object detection. The FPN-GAN method [193], combining Generative Adversarial Networks with Resnet-50 and a Feature Pyramid Network, offers an innovative solution for detecting small objects in low-resolution, noisy images. This unified end-to-end model enhances image resolution and detail recognition, significantly boosting small object detection in the remote sensing DIOR dataset. In the field of infrared image dataset enhancement, researchers [194–196] have combined GANs with transfer learning and other novel techniques for improved training efficiency and detection accuracy. These approaches include refining models like YOLOv4 with a negative sample focus mechanism, integrating a Dilated CBAM into CSPDarknet53, and employing Conditional Generative Adversarial Networks (CGANs) to augment datasets. Particularly, the novel intensity modulation network [196] generates realistic target objects, facilitating the selection of optimal models through extensive experimentation and improved performance.

In short, the integration of GANs with object detection frameworks has marked a significant advancement in the field of small object detection in aerial and satellite imagery. These methods not only enhance image quality but also improve the accuracy and efficiency of object detection models, addressing the inherent challenges of limited sample sizes, diversity, and feature representation in remote sensing datasets.

2.5.7 Neural style transfer image augmentation

An eminent study is the use of transfer learning with the VGG model, as introduced in [197]. This approach, coupled with data augmentation through style transfer algorithms, has significantly enhanced the capability of convolutional neural networks (CNNs) in image recognition tasks. By incorporating a broader range of prior knowledge through training samples modified by style transfer, this method has shown a marked improvement in recognition precision. In the realm of generative adversarial networks (GANs), a novel generative network loss function has been developed [198], transforming satellite images into aerial images. This transformation is further refined by employing advanced algorithms like D2-Net and LoFTR, which aid in the enhanced matching of satellite-aerial image pairs. This method demonstrates an adept use of style transfer in improving the quality and utility of remote sensing data. Another groundbreaking method [199] involves the intelligent registration of heterogeneous remote sensing images using style transfer via GANs. This technique ensures precise alignment and content preservation, a critical factor in remote sensing applications. The validation of this method through a custom-designed dataset highlights its effectiveness and potential for broader applications. In agricultural monitoring, particularly in orchard monitoring via UAVs, challenges such as scarce open access labeled datasets and class imbalances have been addressed innovatively. The study [200] utilizes YOLOv5 for object detection, enhancing its efficacy through a transfer learning approach with pre-training on a synthetic dataset created via object-based augmentation. This strategy has significantly improved the detection of rare and anomalous vegetation states, showcasing the potential of synthetic data in overcoming data scarcity and imbalance issues. The use of CycleGAN for style transfer between distinct datasets [201] is another noteworthy development. By maintaining the labels of generated data, this approach facilitates effective domain adaptation. The enhancement of the CycleGAN model with multi-scale convolution and an attention mechanism in both generative and discriminative models represents a significant improvement in image quality post-transfer. In the specific area of soybean leaf disease detection, the methodology used in [202] involves evaluating four deep neural network models. Techniques such as fine-tuning, transfer learning, and data augmentation have been employed to enhance accuracy. The use of the SLIC method for precise segmentation of plant leaves in top-view images captured by UAVs further underscores the importance of specialized image processing techniques in agricultural applications. Recent developments in plant leaf disease dataset generation [203] has leveraged advanced techniques such as GANs and Neural Style Transfer (NST). When evaluated with transfer learning models like VGG16, ResNet, and InceptionV3, these innovative augmentations have surpassed traditional methods in accuracy, marking a significant improvement over basic image manipulation methods in dataset enhancement.

2.5.8 Meta augmentation

Meta-augmentation transforms a single task into multiple tasks [132], introducing diversity to enhance the model's adaptability. This technique expands the learning capacity by exposing the model to varied instances of the same task. Researcher [204] adopted a meta-learning approach called Dynamic Resolution Adaptation Transfer to augment training images with diverse smaller resolutions ensuring consistent performance across object sizes with shifted feature aggregation and an anchor relation module. This combination of TransDet effectively enhances small object detection performance. In addressing the limitations of UAV-based vehicle recognition under conditions of significant occlusion, researchers [205] have developed a novel approach combining meta-learning and transfer learning. This method initially trains a classifier to extract features, and then employs meta learning techniques to train both a "learner" and a "meta learner," thereby acquiring a set of initial parameters. These parameters are particularly effective for training the recognition model with minimal samples in new, uncharted tasks, offering a significant improvement over traditional Bayesian inference and component models in challenging scenarios.

Xu et al. [204] focuses on consistent object detection performance and [205] excelling in rapid adaptability under challenging conditions.

2.5.9 Feature-space image augmentation

The advancement in small object detection, particularly in aerial datasets (UAV imagery and remote sensing, etc), has seen significant contributions from various studies. For instance [206] employed various image interpolation methods, including bilinear, under the nomenclature UA-net, aiming to improve the performance of small image processing. Similarly, an improved YOLOv5-based method [207] utilizes bilinear interpolation upsampling for feature fusion, effectively reducing information loss and improving the detection of small and medium-sized targets in drone aerial images. The utilization of bilinear interpolation upsampling enhances feature fusion, mitigating information loss during upsampling. Additionally, the integration of a specialized small target detection layer (size 160) effectively locates and recognizes small targets using shallow feature information, thereby minimizing missed detection rates. The modification of the Faster R-CNN framework [208] represents another significant stride. This approach integrates top-down and skip connections to achieve high-resolution feature maps, which are crucial for detecting small and tiny objects in optical remote sensing images. The study also tailors anchor sizes to the dataset's object size distribution and employs 'random rotation' data augmentation and strategic sampling methods to address class distribution imbalance. In the context of minor insulator defects, [209] introduces an enhanced detection network incorporating a Batch Normalization Convolutional Block Attention Module (BN-CBAM). This optimizes channel information utilization and enhances feature map channel impacts. Additionally, a feature fusion module proposed in this study significantly boosts small object detection. To address the scarcity of aerial images, the researchers developed a unique data augmentation technique that blends target segments with backgrounds, effectively enriching the dataset and enhancing the model's robustness. This combination of BN-CBAM and feature fusion, along with the novel data augmentation strategy, marks a significant step forward for small object detection. Li and Zhang [210] tackles false detections in object recognition by integrating Mosaic and Mixup techniques into a novel method, MDAB, which enhances background sample data augmentation. This study also incorporates FaPN into the Rep-PAN structure and merges Transformer technology with CNN, achieving a 6% accuracy improvement on the Visdrone dataset. Another study [211] introduces an innovative attention-guided feature alignment fusion module, which effectively fuses features across scales to enhance spatial and contextual information, thus mitigating the problem of feature inconsistent deterioration in small target detection. Additionally, it proposes a shallow feature supplement module that employs attention mechanisms for augmenting small object information in low-level features directly at the base of the Neck, improving the detection efficiency for small objects without the need for an extra detection head. These advancements collectively enhance the detector's performance in identifying small-scale targets. Addressing challenges like limited features and sample imbalance, [212] proposes a method combining multiscale feature fusion with dilated convolution. This enhances the receptive field of feature maps and incorporates attention mechanisms to improve context awareness in object detection. In another study [213], UAV image mosaic speed and accuracy are enhanced by analyzing aerial image characteristics and addressing ghosting and blurring issues. This approach employs the Quick-SIFT operator and the As-Natural-As-Possible algorithm for effective image registration and transformation. This approach employs the frame difference method and region-growing algorithm for moving object segmentation and linear weighted fusion to mitigate ghosting. Comparative experiments with existing methods demonstrate a significant 78% improvement in matching time, showcasing the method's efficiency in rapid image mosaic processing by developing custom dataset for image mosaic. A notable contribution is the FS-SSD method, an extension of the FSSD framework, as detailed in [211] introduces a novel feature fusion and scaling approach. This includes an additional scaling branch with a deconvolution module and average pooling to form a feature pyramid, enhancing detection capabilities for smaller objects. The integration of two feature pyramids, from the deconvolution and feature fusion modules, allows for combined predictions, improving detection accuracy. Moreover, incorporating spatial context analysis, which considers interclass and intraclass distances, further refines multiclass small object detection, showcasing a significant advancement in the field. Complementing these advancements, [214] introduced a groundbreaking approach for tiny object detection. Their method, a feature pyramid composite neural network, integrates a context enhancement module (CEM) and a feature purification module (FPM). The CEM utilizes multi-scale dilated convolution features for enhanced context information, while the FPM employs channel and space dimension purification to highlight tiny objects amidst conflicting information. Additionally, [214] proposed a unique data-enhancement technique, termed copy-reduce-paste. This technique is designed to amplify the impact of tiny objects within the loss function, thereby addressing the issue of imbalanced training samples. This methodological innovation not only enhances the detection accuracy but also contributes to the overall robustness of the training process.

3 Discussion

In the past five years, basic approaches to augmentation for small object detection have predominantly favored geometric and region-wise methods. Despite being widely adopted in this domain, there has been a noticeable shift towards the utilization of Generative Adversarial Networks (GANs) for synthesizing training data, reflecting a growing trend in the exploration of synthetic data generation techniques. The transition towards using GANs for data synthesis in small object detection is driven by the ability of GAN-generated images to closely mirror real-world scenarios. This shift addresses challenges related to variations in images and enhances the complexities present in training models, ultimately contributing to improved model performance. In basic data augmentation approaches, photometric and sample-level methods are less frequently used, primarily due to their limited applicability and potential for introducing unrealistic modifications. These methods, which include adjustments to color, brightness, and texture, can sometimes distort the essential characteristics of objects in images. Moreover, in sample-level augmentation, the creation of entirely new samples may not accurately reflect real-world scenarios, particularly in specialized fields like medical imaging or aerial surveillance, where fidelity to actual conditions is crucial. Meta-augmentation, a state-of-the-art approach in computer vision tasks like few-shot learning, has not yet been widely employed for small object detection due to several potential reasons. First, the specific characteristics of small objects in images, such as their scale and sparse distribution, may not align well with the general principles of meta-augmentation. This approach typically requires a level of generalization that might not be effective for the detailed and precise detection of small objects. Additionally, the complex and varied backgrounds often present in images with small objects could further challenge the efficacy of meta-augmentation methods, which are designed for broader and more general applications in computer vision.

It is also noticed that traditional image augmentation and meta augmentation serve distinct purposes. Traditional methods aim to create more samples within a single task, whereas meta augmentation focuses on generating diverse tasks for a single example. The latter approach compels the model to rapidly learn new tasks from feedback, differentiating it from the objective of traditional augmentation methods. State-of-the-art (SOTA) image augmentation approaches exhibit comparable speed to traditional approaches for small object detection in aerial images. However, their efficiency differs, often requiring the integration of an additional network to extract significant information from input images. Throughout the analysis, it has been identified that augmentation methods have effectively addressed the following issues. During the review of the literature, it is noted that state-of-the-art (SoTA) augmentation approaches provide guidelines on selecting and combining techniques, they are constrained by a limited range of supported augmentation operations and tend to require more computational power and memory. Such limitations restrict the volume of data. Consequently, research in the field of data augmentation, particularly for the detection of small and tiny objects, remains a highly active area.

The main challenge of image augmentation approaches for small and tiny object detection in aerial images is the risk of overfitting. When augmenting the dataset by generating additional images with variations such as rotation, scaling, or flipping, there is a possibility that the model may learn features specific to the augmented images rather than generalize well to unseen data. To address overfitting in image augmentation for detecting small objects in aerial images, researchers should constrain augmentation parameters to realistic transformations, use adaptive policies like AutoAugment, and employ advanced techniques such as Mixup and CutMix. Targeted regularization methods, such as dropout, and rigorous evaluation on minimally augmented validation sets are essential. Additionally, region-level augmentations and meta-augmentation strategies can further improve model generalization and performance. Additionally, aggressive augmentation techniques might introduce unrealistic variations that do not accurately represent real-world scenarios, leading to a degradation in model performance when applied to actual aerial images. To mitigate the negative effects of aggressive augmentation, researchers can use regularization techniques like dropout or weight decay. Additionally, monitoring validation performance during training helps identify when augmentation becomes excessive, ensuring that the model generalizes well to real-world data. Moreover, extensive augmentation can significantly increase the computational complexity and training time. Researchers can optimize augmentation pipelines for efficiency by using GPU-accelerated libraries and parallelization. This approach helps manage the increased computational burden and ensures that the training process remains feasible in environments with limited resources. The paper primarily focuses on image augmentation for small and tiny object detection in aerial images, which may limit its applicability to other domains.

The following points outline the problems that the aforementioned augmentation approaches have addressed in the context of Small and Tiny Object Detection (STOD).

3.1 Problems solved with basic augmentation approaches

- The model faces challenges in performance when aerial images lack a sufficient number of small or tiny objects, making it difficult for the model to learn them accurately. To address this issue, basic augmentation approaches artificially generate samples for small and tiny objects.
- In aerial images, the size of objects varies significantly, with some appearing large and others too small to discern easily. This diversity in size poses a challenge for detection due to the relative scale of objects. Techniques such as resizing and scaling effectively address this challenge, providing a viable solution to ensure accurate and comprehensive detection of varying object sizes.
- During the processing or transformation of images, there is a risk of losing crucial information or details pertaining to objects. This concern is more pronounced for small objects, given their inherently lower pixel count in the image. This, in turn, results in poor performance. Augmentation approaches, such as flipping and rotation, serve as effective solutions to mitigate this problem.
- Dealing with a large-scale dataset containing numerous small and tiny objects and managing high detection accuracy can pose a challenge due to memory intensity, resulting in a slow inference rate. Efficient batch processing helps alleviate this challenge to some extent. Moreover, the combination of various augmentation techniques enhances both detection accuracy and inference rate.
- It is crucial for models to effectively handle inputs with diverse variations, such as changes in orientation, lighting, and more. This challenge becomes more complex in aerial images due to the varied conditions at high altitudes where the capturing devices are positioned. Approaches like rotation, flipping, scaling, translation, brightness/contrast adjustments, zooming, shearing, and color jittering play a key role in enhancing a model's robustness against a wide range of input variations during training.
- In aerial images, it is observed that small and tiny objects often lack detailed information due to occlusion, and environmental factors such as lighting conditions can impact both image and object quality, affecting the accuracy of models. Augmentation techniques, such as cropping and zooming to enlarge objects, help gather detailed information and address occlusion issues. Additionally, adjustments in brightness and color tackle challenges related to image quality. In conventional augmentation methods, researchers commonly employ a combination of approaches to effectively address issues with aerial images and performance concerns.

3.2 Problems solved with SoTA augmentation approaches

- Enhanced feature representation approaches improve the visibility of objects in aerial images, especially those that are often overlooked or poorly depicted using traditional methods. By employing feature augmentation, models can achieve higher accuracy in detecting small and tiny objects.
- Small and tiny objects tend to be underrepresented in training datasets, leading to biased models. SoTA augmentation approaches like copy-paste, mix-up, or custom data augmentation strategies increase the prevalence of these objects in training datasets, leading to more balanced and effective learning.
- One of the significant challenges in detecting small and tiny objects is the wide range of scales at which objects can appear. Advanced augmentation techniques involve scaling objects to various sizes during training, enabling the model to learn how to recognize objects across different scales.
- For tiny object detection, contextual information is crucial. Augmentation approaches can enhance the contextual surroundings of tiny objects in images, making it easier for models to discern them from the background or noise.
- By providing a more robust and varied training dataset, SoTA augmentation approaches help reduce false positives (incorrectly identifying objects where there are none) and false negatives (failing to identify existing objects), which are common issues in small object detection.
- With a more diverse and comprehensive training dataset, models become better at generalizing to new, unseen data, which is crucial for practical applications.
- In datasets where some classes of objects are much more common than others, augmentation can help to balance these classes by artificially increasing the presence of less common objects.

3.3 Challenges and future work

There are several challenges that need attention, and are elaborated in the following:

- Detecting small objects remains a significant challenge for researchers, particularly when working with aerial datasets. Despite efforts to address this issue through various data augmentation approaches, there is an ongoing need to identify the most accurate augmentation method specifically tailored to the challenges posed by aerial images. Adaptive augmentation approaches and improved annotation precision can be Potential solutions. Adaptive augmentation involves continuously modifying augmentation parameters based on feedback from the training process. It dynamically adjusts the type and intensity of approaches. On the other hand, improved annotation precision ensures that small objects in aerial images are labeled with high accuracy and detail. Precise annotations help the model learn the correct features associated with these objects, reducing confusion with background noise and improving detection performance.
- The application of traditional augmentation approaches in computer vision lacks a specific, universally applicable framework. Instead, the choice of techniques or their combination is largely determined by the challenges and characteristics of the dataset a researcher is working with. This tailored approach to augmentation is employed based on the specific needs identified during the experimental phase. This variability requires researchers to conduct extensive preliminary experiments to identify suitable augmentation approaches. Researchers can develop automated systems that can analyze

dataset characteristics and suggest optimal augmentation approaches. For future research, researchers can also develop a standardized guideline for common scenarios for aerial datasets that could streamline the augmentation process.

- Creating effective augmentation strategies for object detection is notably more complex than for other computer vision tasks, receiving less research attention due to perceived lesser impact on performance and limited applicability across tasks. Future research may focus on developing task-specific augmentation approaches that enhance model sensitivity to small and contextually diverse objects. This involves integrating advanced geometric transformations and contextual augmentations that closely imitate real-world scenarios.
- Identifying the optimal number of samples for augmentation remains an unresolved question in this context. Determining the ideal quantity of data for effective augmentation approaches poses an ongoing challenge for researchers. In the future, more investigation on reinforcement learning approaches can be helpful in determining the most effective sample sizes for effective augmentation approaches.
- Machine learning and deep learning models demand substantial data for effective training, yet generating additional datasets proves challenging, time-consuming, and expensive. While data augmentation partially addresses this issue, the demand for extensive aerial datasets is crucial, particularly in the detection of small objects. Researchers can explore collaborative data-sharing platforms and advanced synthetic data generation approaches to cost-effectively expand and improve aerial datasets.
- Maintaining scalability, efficiency, and high-quality annotations for small objects in augmented datasets proves challenging. This complexity poses a significant hurdle in advancing the performance of detection models, requiring thoughtful solutions for real-world applicability. Developing automated annotation tools and scalable annotation platforms can enhance data handling while implementing quality control mechanisms ensures annotation accuracy.
- The effectiveness of using contextual information in small object detection is inconsistent, highlighting the need for future research to identify and refine strategies that better manage the use of such information. This involves exploring how contextual data can be optimally utilized or controlled to enhance detection accuracy in small object scenarios. Researchers can focus on developing sophisticated contextual-aware augmentation approaches that respect object boundaries and contextual relationships in aerial images. Investigating adaptive and multi-scale augmentation methods could help manage the complexities of high object density and varied sizes. Additionally, enhancing machine learning models to better process and learn from contextually enriched synthetic datasets could significantly improve detection accuracy.
- Image augmentation techniques become increasingly computationally intensive as they improve the performance of detection tasks during both training and testing. This added complexity poses a challenge, making it less feasible for real-time applications due to inefficient resource management. Using GPU-accelerated libraries and parallelization can be helpful in this scenario.
- Further research is needed in the area of object detection through meta-learning in aerial datasets. This approach is helpful when dealing with small or limited datasets, rare objects, or when there is a lack of labeled data for training.

4 Conclusion

Quality data is pivotal for the performance of deep learning systems, irrespective of the field. Augmentation methods provide a solution to address the voracious appetite of machine or deep learning models, compensating for the challenge of obtaining high-quality data. Image Augmentation approaches, gain much more importance in addressing the challenges posed by small and tiny objects in aerial images. In this paper, a comprehensive review of image augmentation approaches for detecting small and tiny objects in aerial images has been conducted. The review includes a summary of the libraries commonly utilized by programmers and researchers to augment datasets, offering developers and researchers a centralized resource. It also provided a brief description of both traditional and State-of-the-art image augmentation approaches, highlighting their respective advantages and disadvantages, that contribute to researchers in selecting optimal augmentation approaches. Furthermore, explored how these image augmentation techniques enhance the detection performance of small and tiny objects in the aerial datasets. Finally, analyzed the effectiveness of these approaches specifically for small and tiny object detection, discussing the challenges and future prospects in this field. Although traditional augmentation approaches improve performance in small and tiny object detection, challenges arise in determining the optimal combination of approaches in advance and the availability of existing training examples. State-of-the-Art approaches attempt to address this issue but encounter difficulties in predicting performance. These approaches aim to simplify this process, yet their narrow range of operations requires time-consuming trial-and-error experimentation with actual data. Future work in the field of small and tiny object detection for aerial images should prioritize refining augmentation approaches shaped specifically for aerial images, exploring meta-learning and few-shot learning for efficient training with limited data, and enhancing computational efficiency to facilitate real-time applications. Additionally, it is crucial to develop augmentation approaches that effectively utilize contextual information and ensure high-quality, scalable annotations to advance the accuracy and practicality of detection models in real-world scenarios. Hence, the research for developing and optimal use of augmentation approaches for small and tiny object detection in aerial images is still an open area to answer.

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

I declare that the content of this manuscript represents original work and has not been previously published. Furthermore, it is not currently under consideration for publication elsewhere. I consent to its submission to Multimedia Tools and Applications.

Conflict of interests/Competing interests The authors declare that they have no conflict of interest.

References

1. Kaur P, Khehra B.S, Mavi E.B.S (2021) Data augmentation for object detection: A review. In: 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS) 537–543. IEEE
2. Hasanpour SH, Rouhani M, Fayyaz M, Sabokrou M (2016) Lets keep it simple, using simple architectures to outperform deeper and more complex architectures. arXiv:1608.06037

3. Wang J, Yang W, Guo H, Zhang R, Xia G.-S (2021) Tiny object detection in aerial images. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp 3791–3798.IEEE
4. Liu Y, Sun P, Wergeles N, Shang Y (2021) A survey and performance evaluation of deep learning methods for small object detection. Expert Syst Appl 172:114602
5. Duong HT, Nguyen-Thi TA (2021) A review: preprocessing techniques and data augmentation for sentiment analysis. Computational Social Networks 8(1):1–16
6. Tong K, Wu Y (2022) Deep learning-based detection from the perspective of small or tiny objects: A survey. Image and Vision Computing 104471
7. Chen G, Wang H, Chen K, Li Z, Song Z, Liu Y, Chen W, Knoll A (2020) A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. IEEE Transactions on systems, man, and cybernetics: systems 52(2):936–953
8. Wu X, Li W, Hong D, Tao R, Du Q (2021) Deep learning for unmanned aerial vehicle-based object detection and tracking: A survey. IEEE Geoscience and Remote Sensing Magazine 10(1):91–124
9. Muzammul M, Li X (2021) A survey on deep domain adaptation and tiny object detection challenges, techniques and datasets. arXiv:2107.07927
10. Nguyen ND, Do T, Ngo TD, Le DD (2020) An evaluation of deep learning methods for small object detection. Journal of electrical and computer engineering 2020:1–18
11. Tong K, Wu Y, Zhou F (2020) Recent advances in small object detection based on deep learning: A review. Image Vis Comput 97:103910
12. Muzammul M, Li X (2021) A survey on deep domain adaptation and tiny object detection challenges, techniques and datasets. arXiv:2107.07927
13. Zhou B, Buyya R (2018) Augmentation techniques for mobile cloud computing: A taxonomy, survey, and future directions. ACM Computing Surveys (CSUR) 51(1):1–38
14. Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW), pp 117–122. IEEE
15. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. Journal of big data 6(1):1–48
16. Nalepa J, Marcinkiewicz M, Kawulok M (2019) Data augmentation for brain-tumor segmentation: a review. Front Comput Neurosci 13:83
17. Bissoto A, Valle E, Avila S (2021) Gan-based data augmentation and anonymization for skin-lesion analysis: A critical review. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 1847–1856
18. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A (2021) A review of medical image data augmentation techniques for deep learning applications. J Med Imaging Radiat Oncol 65(5):545–563
19. Chen Y, Yang XH, Wei Z, Heidari AA, Zheng N, Li Z, Chen H, Hu H, Zhou Q, Guan Q (2022) Generative adversarial networks in medical image augmentation: a review. Computers in Biology and Medicine 105382
20. Khosla C, Saini BS (2020) Enhancing performance of deep learning models with different data augmentation techniques: A survey. In: 2020 International Conference on Intelligent Engineering and Management (ICIEM), pp 79–85. IEEE
21. Wang X, Wang K, Lian S (2020) A survey on face data augmentation for the training of deep neural networks. Neural Comput Appl 32(19):15503–15531
22. Liu P, Wang X, Xiang C, Meng W (2020) A survey of text data augmentation. In: 2020 International Conference on Computer Communication and Network Security (CCNS), pp 191–195. IEEE
23. Naveed H, Anwar S, Hayat M, Javed K, Mian A (2021) Survey: Image mixing and deleting for data augmentation. arXiv:2106.07085
24. Shorten C, Khoshgoftaar TM, Furht B (2021) Text data augmentation for deep learning. Journal of big Data 8:1–34
25. Khalifa NE, Loey M, Mirjalili S (2022) A comprehensive survey of recent trends in deep learning for digital images augmentation. Artificial Intelligence Review 1–27
26. Yang S, Xiao W, Zhang M, Guo S, Zhao J, Shen F (2022) Image data augmentation for deep learning: A survey. arXiv:2204.08610
27. Mumuni A, Mumuni F (2022) Data augmentation: A comprehensive survey of modern approaches. Array 100258
28. Oubara A, Wu F, Amamra A, Yang G (2022) Survey on remote sensing data augmentation: Advances, challenges, and future perspectives. In: Advances in Computing Systems and Applications: Proceedings of the 5th Conference on Computing Systems and Applications, pp 95–104. Springer
29. Lewy D, Mańdziuk J (2023) An overview of mixing augmentation methods and augmentation strategies. Artif Intell Rev 56(3):2111–2169

30. Li Z, Wang Y, Zhang N, Zhang Y, Zhao Z, Xu D, Ben G, Gao Y (2022) Deep learning-based object detection techniques for remote sensing images: A survey. Remote Sensing 14(10):2385
31. Connor S, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. Journal of big data 6(1):1–48
32. Naveed H, Anwar S, Hayat M, Javed K, Mian A (2023) Survey: Image Mixing and Deleting for Data Augmentation
33. Maharana K, Mondal S, Nemade B (2022) A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings
34. Farahanipad F, Rezaei M, Nasr MS, Kamangar F, Athitsos V (2022) A survey on gan-based data augmentation for hand pose estimation problem. Technologies 10(2):43
35. Lalitha V, Latha B (2022) A review on remote sensing imagery augmentation using deep learning. Materials Today: Proceedings
36. Lewy D, Mańdziuk J (2022) An overview of mixing augmentation methods and augmentation strategies. Artif Intell Rev 56(3):2111–2169. https://doi.org/10.1007/s10462-022-10227-z
37. Alomar K, Aysel HI, Cai X (2023) Data augmentation in classification and segmentation: A survey and new strategies. Journal of Imaging 9(2):46
38. Kumar T, Turab M, Raj K, Mileo A, Brennan R, Bendechache M (2023) Advanced data augmentation approaches: A comprehensive survey and future directions. arXiv:2301.02830
39. Bloice MD, Stocker C, Holzinger A (2017) Augmentor: an image augmentation library for machine learning. arXiv:1708.04680
40. Jung A (2019) Imgaug documentation. Readthedocs. io, Jun 25
41. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA (2020) Albumentations: fast and flexible image augmentations. Information 11(2):125
42. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M et al (2016) Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467
43. Chen T, Li M, Li Y, Lin M, Wang N, Wang M, Xiao T, Xu B, Zhang C, Zhang Z (2015) Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. arXiv:1512.01274
44. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia: pp 675–678
45. Chollet FK (2015) Available online: https://keras.io (accessed on 14 august 2019). 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/)
46. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L et al (2019) Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems 32
47. Kovesi PD (2000) MATLAB and Octave functions for computer vision and image processing
48. McAuliffe M.J, Lalonde F.M, McGarry D, Gandler W, Csaky K, Trus BL (2001) Medical image processing, analysis and visualization in clinical research. In: Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001, pp 381–386. IEEE
49. Klinger T (2003) Image Processing with LabVIEW and IMAQ Vision.Prentice Hall Professional
50. Demirkaya O, Asyali MH, Sahoo PK (2008) Image Processing with MATLAB: Applications in Medicine and Biology. CRC Press
51. McCaslin S, Kesireddy A (2015) Metallographic image processing tools using mathematica manipulate. In: Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering, pp 357–363. Springer
52. Geosystems L (2004) Erdas imagine. Atlanta, Georgia 7(12):3209–3241
53. DeVries T, Taylor GW (2017) Dataset augmentation in feature space. arXiv:1702.05538
54. Singh KK, Yu H, Sarmasi A, Pradeep G, Lee YJ (2018) Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond. arXiv:1811.02545
55. Yang Z, Wang Z, Xu W, He X, Wang Z, Yin Z (2019) Region-aware random erasing. In: 2019 IEEE 19th International Conference on Communication Technology (ICCT), pp 1699–1703. IEEE
56. Huang J, Zhu Z, Huang G, Du D (2020) Aid: Pushing the performance boundary of human pose estimation with information dropping augmentation. arXiv:2008.07139
57. Zhong Z, Zheng L, Kang G, Li S, Yang Y (2020) Random erasing data augmentation. Proceedings of the AAAI Conference on Artificial Intelligence 34:13001–13008
58. Chen P, Liu S, Zhao H, Jia J (2020) Gridmask data augmentation. arXiv:2001.04086

59. Feng S, Yang S, Niu Z, Xie J, Wei M, Li P (2021) Grid cut and mix: flexible and efficient data augmentation. In: Twelfth International Conference on Graphics and Image Processing (ICGIP 2020), vol. 11720, pp 656–662. SPIE
60. Li C.-L, Sohn K, Yoon J, Pfister T (2021) Cutpaste: Self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 9664–9674
61. Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y (2019) Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 6023–6032
62. Yoo J, Ahn N, Sohn KA (2020) Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8375–8384
63. Kang G, Dong X, Zheng L, Yang Y (2017) Patchshuffle regularization. arXiv:1707.07103
64. Xie T, Cheng X, Wang X, Liu M, Deng J, Zhou T, Liu M (2021) Cut-thumbnail: A novel data augmentation for convolutional neural network. In: Proceedings of the 29th ACM International Conference on Multimedia, pp 1627–1635
65. Kim J, Shin IH, Lee JR, Lee YJ (2020) Where to cut and paste: Data regularization with selective features. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp 1219–1221. IEEE
66. Walawalkar D, Shen Z, Liu Z, Savvides M (2020) Attentive cutmix: An enhanced data augmentation approach for deep learning based image classification. arXiv:2003.13048
67. Li H, Zhang X, Tian Q, Xiong H (2020) Attribute mix: Semantic data augmentation for fine grained recognition. In: 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP), pp 243–246. IEEE
68. Takahashi R, Matsubara T, Uehara K (2018) Ricap: Random image cropping and patching data augmentation for deep cnns. In: Asian Conference on Machine Learning, pp 786–798. PMLR
69. Hong S, Kang S, Cho D (2019) Patch-level augmentation for object detection in aerial images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp 0–0
70. Uddin A, Monira M, Shin W, Chung T, Bae SH et al (2020) Saliencymix: A saliency guided data augmentation strategy for better regularization. arXiv:2006.01791
71. Kim JH, Choo W, Song HO (2020) Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In: International Conference on Machine Learning, pp 5275–5285. PMLR
72. Dabouei A, Soleymani S, Taherkhani F, Nasrabadi NM (2021) Supermix: Supervising the mixing data augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 13794–13803
73. Liu J, Liu B, Zhou H, Li H, Liu Y (2022) Tokenmix: Rethinking image mixing for data augmentation in vision transformers. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVI, pp 455–471. Springer
74. Zhao H, Wang J, Chen Z, Lin S, Bao P (2022) Srk-augment: A self-replacement and discriminative region keeping augmentation scheme for better classification. Neural Processing Letters, 1–17
75. Zhang H, Cisse M, Dauphin Y.N, Lopez-Paz D (2017) mixup: Beyond empirical risk minimization. arXiv:1710.09412
76. Kim JH, Choo W, Jeong H, Song HO (2021) Co-mixup: Saliency guided joint mixup with supermodular diversity. arXiv:2102.03065
77. Inoue H (2018) Data augmentation by pairing samples for images classification. arXiv:1801.02929
78. Lee JH, Zaheer MZ, Astrid M, Lee SI (2020) Smoothmix: a simple yet effective data augmentation to train robust classifiers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp 756–757
79. Guo H, Mao Y, Zhang R (2019) Mixup as locally linear out-of-manifold regularization. Proceedings of the AAAI Conference on Artificial Intelligence 33:3714–3722
80. Yu H, Wang H, Wu J (2021) Mixup without hesitation. In: Image and Graphics: 11th International Conference, ICIG 2021, Haikou, China, August 6–8, 2021, Proceedings, Part II 11, pp 143–154. Springer
81. Lin WH, Zhong JX, Liu S, Li T, Li G (2020) Roimix: proposal-fusion among multiple images for underwater object detection. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 2588–2592. IEEE
82. Hendrycks D, Mu N, Cubuk ED, Zoph B, Gilmer J, Lakshminarayanan B (2019) Augmix: A simple data processing method to improve robustness and uncertainty. arXiv:1912.02781
83. Summers C, Dinneen MJ (2019) Improved mixed-example data augmentation. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp 1262–1270. IEEE

84. Negassi M, Wagner D, Reiterer A (2022) Smart (sampling) augment: Optimal and efficient data augmentation for semantic segmentation. Algorithms 15(5):165
85. Bello I, Zoph B, Vaswani A, Shlens J, Le Q (2019) Proceedings of the ieee/cvf international conference on computer vision. Unknown
86. Upchurch P, Gardner J, Pleiss G, Pless R, Snavely N, Bala K, Weinberger K (2017) Deep feature interpolation for image content changes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 7064–7073
87. Zhen W, Yao S, Lin J (2018) Learning adaptive receptive fields for deep image parsing networks. Computational Visual Media 4(3):1–14
88. Li X, Dai Y, Ge Y, Liu J, Shan Y, Duan LY () Uncertainty modeling for out-of-distribution generalization. arXiv:2202.03958
89. Goodfellow I, Warde-Farley D, Mirza M, Courville A, Bengio Y (2013) Maxout networks. In: International Conference on Machine Learning, pp 1319–1327. PMLR
90. Bouthillier X, Konda K, Vincent P, Memisevic R (2015) Dropout as data augmentation. arXiv:1506.08700
91. Kang G, Li J, Tao D (2016) Shakeout: A new regularized deep neural network training scheme. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30
92. Krueger D, Maharaj T, Kramár J, Pezeshki M, Ballas N, Ke NR, Goyal A, Bengio Y, Courville A, Pal C (2016) Zoneout: Regularizing rnns by randomly preserving hidden activations. arXiv:1606.01305
93. Chang AX, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, Savarese S, Savva M, Song S, Su H et al (2015) Shapenet: An information-rich 3d model repository. arXiv:1512.03012
94. McCormac J, Handa A, Leutenegger S, Davison AJ (2017) Scenenet rgb-d: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation? In: Proceedings of the IEEE International Conference on Computer Vision, pp 2678–2687
95. Barbosa IB, Cristani M, Caputo B, Rognhaugen A, Theoharis T (2018) Looking beyond appearances: Synthetic training data for deep cnns in re-identification. Comput Vis Image Underst 167:50–62
96. Roberts M, Ramapuram J, Ranjan A, Kumar A, Bautista MA, Paczan N, Webb R, Susskind JM (2021) Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 10912–10922
97. Gaidon A, Wang Q, Cabon Y, Vig E (2016) Virtual worlds as proxy for multi-object tracking analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4340–4349
98. Ros G, Sellart L, Materzynska J, Vazquez D, Lopez AM (2016) The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3234–3243
99. Cheung E, Wong TK, Bera A, Wang X, Manocha D (2016) Lcrowdv: Generating labeled videos for simulation-based crowd behavior learning. In: Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14, pp 709–727. Springer
100. Jain V, Aggarwal S, Mehta S, Hebbalaguppe R (2019) Synthetic video generation for robust hand gesture recognition in augmented reality applications. arXiv:1911.01320
101. Dosovitskiy A, Ros G, Codevilla F, Lopez A, Koltun V (2017) Carla: An open urban driving simulator. In: Conference on Robot Learning, pp 1–16. PMLR
102. Niemeyer M, Mescheder L, Oechsle M, Geiger A (2020) Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 3504–3515
103. Tremblay J, Meshry M, Evans A, Kautz J, Keller A, Khamis S, Loop C, Morrical N, Nagano K, Takikawa T et al (2022) Rtmv: A ray-traced multi-view synthetic dataset for novel view synthesis. arXiv:2205.07058
104. Wrenninge M, Unger J (2018) Synscapes: A photorealistic synthetic dataset for street scene parsing. arXiv:1810.08705
105. Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R (2021) Nerf: Representing scenes as neural radiance fields for view synthesis. Commun ACM 65(1):99–106
106. Tancik M, Casser V, Yan X, Pradhan S, Mildenhall B, Srinivasan PP, Barron JT, Kretzschmar H (2022) Block-nerf: Scalable large scene neural view synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8248–8258
107. Turki H, Ramanan D, Satyanarayanan M (2022) Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 12922–12931
108. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative Adversarial Networks
109. Wang Z, She Q, Ward TE (2021) Generative adversarial networks in computer vision: A survey and taxonomy. ACM Computing Surveys (CSUR) 54(2):1–38

110. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv:1411.1784
111. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434
112. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN
113. Jiang Y, Chang S, Wang Z (2021) Transgan: Two pure transformers can make one strong gan, and that can scale up. Adv Neural Inf Process Syst 34:14745–14758
114. Zhu JY, Park T, Isola P, Efros AA (2020) Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks
115. Karras T, Laine S, Aila T (2019) A Style-Based Generator Architecture for Generative Adversarial Networks
116. Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B (2015) Adversarial autoencoders. arXiv:1511.05644
117. Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network
118. Zhang H, Goodfellow I, Metaxas D, Odena A (2019) Self-Attention Generative Adversarial Networks
119. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J (2018) StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation
120. Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P (2016) InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets
121. Donahue J, Simonyan K (2019) Large Scale Adversarial Representation Learning
122. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Loy CC, Qiao Y, Tang X (2018) ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks
123. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S (2018) Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2794–2802
124. Esser P, Rombach R, Ommer B (2021) Taming Transformers for High-Resolution Image Synthesis
125. Gatys LA, Ecker AS, Bethge M (2015) A neural algorithm of artistic style. arXiv:1508.06576
126. Li Y, Fang C, Yang J, Wang Z, Lu X, Yang MH (2017) Universal style transfer via feature transforms. Advances in neural information processing systems **30**
127. Luan F, Paris S, Shechtman E, Bala K (2017) Deep photo style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4990–4998
128. Li Y, Liu MY, Li X, Yang MH, Kautz J (2018) A closed-form solution to photorealistic image stylization. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 453–468
129. Chen Z, Wang W, Xie E, Lu T, Luo P (2022) Towards ultra-resolution neural style transfer via thumbnail instance normalization. Proceedings of the AAAI Conference on Artificial Intelligence 36:393–400
130. Chun S, Park S (2021) Styleaugment: Learning texture de-biased representations by style augmentation without pre-defined textures. arXiv:2108.10549
131. Hong M, Choi J, Kim G (2021) Stylemix: Separating content and style for enhanced data augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 14862–14870
132. Rajendran J, Irpan A, Jang E (2020) Meta-learning requires meta-augmentation. Adv Neural Inf Process Syst 33:5705–5715
133. Liu J, Chao F, Lin CM (2020) Task augmentation by rotating for meta-learning. arXiv:2003.00804
134. Gong C, Ren T, Ye M, Liu Q (2020) Maxup: A simple way to improve generalization of neural network training. arXiv:2002.09024
135. Ni R, Shu M, Souri H, Goldblum M, Goldstein T (2021) The close relationship between contrastive learning and meta-learning. In: International Conference on Learning Representations
136. Yao H, Huang LK, Zhang L, Wei Y, Tian L, Zou J, Huang J et al (2021) Improving generalization in meta-learning via task augmentation. In: International Conference on Machine Learning, pp 11887–11897. PMLR
137. Li L, Han J, Yao X, Cheng G, Guo L (2020) Dla-matchnet for few-shot remote sensing image scene classification. IEEE Trans Geosci Remote Sens 59(9):7844–7853
138. Lim S, Kim I, Kim T, Kim C, Kim S (2019) Fast autoaugment. Advances in Neural Information Processing Systems **32**
139. Zhao A, Balakrishnan G, Durand F, Guttag JV, Dalca AV (2019) Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8543–8553
140. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV (2019) Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 113–123

141. Hataya R, Zdenek J, Yoshizoe K, Nakayama H (2020) Faster autoaugment: Learning augmentation strategies using backpropagation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16, pp 1–16. Springer
142. Liu A, Huang Z, Huang Z, Wang N (2021) Direct differentiable augmentation search. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 12219–12228
143. Gao Y, Tang Z, Zhou M, Metaxas D (2021) Enabling data diversity: efficient automatic augmentation via regularized adversarial training. In: Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings, pp 85–97. Springer
144. Miao H, Rahman LT (2020) Multi-class traffic sign classification using autoaugment and spatial transformer. Unknown
145. He X, Zhao K, Chu X (2021) Automl: A survey of the state-of-the-art. Knowl-Based Syst 212:106622
146. Ni R, Goldblum M, Sharaf A, Kong K, Goldstein T (2021) Data augmentation for meta-learning. In: International Conference on Machine Learning, pp 8152–8161. PMLR
147. Liu K, Mattyus G (2015) Fast multiclass vehicle detection on aerial images. IEEE Geosci Remote Sens Lett 12(9):1938–1942
148. Robicquet A, Sadeghian A, Alahi A, Savarese S (2016) Learning social etiquette: Human trajectory understanding in crowded scenes. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14, pp 549–565. Springer
149. Xia GS, Bai X, Ding J, Zhu Z, Belongie S, Luo J, Datcu M, Pelillo M, Zhang L (2018) 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
150. Li C, Yang T, Zhu S, Chen C, Guan S (2020) Density map guided object detection in aerial images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp 190–191
151. Li K, Wan G, Cheng G, Meng L, Han J (2020) Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS J Photogramm Remote Sens 159:296–307
152. Kisantal M, Wojna Z, Murawski J, Naruniec J, Cho K (2019) Augmentation for small object detection. arXiv:1902.07296
153. Li D, Rodriguez C, Yu X, Li H (2020) Proceedings of the ieee/cvf winter conference on applications of computer vision (wacv). Unknown
154. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp 21–37. Springer
155. Zhao H, Zhang H, Zhao Y (2023) 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
156. Sharma M, Dhanaraj M, Karnam S, Chachlakis DG, Ptucha R, Markopoulos PP, Saber E (2020) Yolors: Object detection in multimodal remote sensing imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14:1497–1508
157. Wang K, Fang B, Qian J, Yang S, Zhou X, Zhou J (2020) Perspective transformation data augmentation for object detection. IEEE Access 8:4935–4943. https://doi.org/10.1109/ACCESS.2019.2962572
158. Meethal A, Granger E, Pedersoli M (2023) Cascaded zoom-in detector for high resolution aerial images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 2045–2054
159. Gao C, Tang W, Jin L, Jun Y (2020) 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
160. Cunha E, Macêdo D, Zanchettin C (2023) Improving small object detection with detraug. In: 2023 International Joint Conference on Neural Networks (IJCNN), pp 1–8. https://doi.org/10.1109/IJCNN54540.2023.10191541
161. Wang S (2020) 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. https://doi.org/10.1109/ICISCE50968.2020.00054
162. Marvasti-Zadeh SM, Khaghani J, Ghanei-Yakhdan H, Kasaei S, Cheng L (2020) Comet: Context-aware iou-guided network for small object tracking. In: Proceedings of the Asian Conference on Computer Vision
163. Chen Y, Zhang P, Li Z, Li Y, Zhang X, Meng G, Xiang S, Sun J, Jia J (2020) Stitcher: Feedback-driven data provider for object detection 2(7): 12. arXiv:2004.12432
164. Ye Y, Ren X, Zhu B, Tang T, Tan X, Gui Y, Yao Q (2022) An adaptive attention fusion mechanism convolutional network for object detection in remote sensing images. Remote Sensing 14(3):516

165. Yim S, Cho M, Lee S (2023) Object-oriented cutout data augmentation for tiny object detection. In: 2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC), pp 1–4. IEEE
166. Lee G, Hong S, Cho D (2021) Self-supervised feature enhancement networks for small object detection in noisy images. IEEE Signal Process Lett 28:1026–1030. https://doi.org/10.1109/LSP.2021.3081041
167. Li Z (2022) Site safety target detection with improved yolov5. In: Proceedings of the 2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence, pp 738–742
168. Wang Y, Qin Y, Cui J (2021) Occlusion robust wheat ear counting algorithm based on deep learning. Front Plant Sci 12:645899
169. Kim J, Lee YJ, Moon YH (2021) Enhancing the data regularization effect with randomly combined features for object detection. In: 2021 International Conference on Information and Communication Technology Convergence (ICTC), pp 1065–1068. IEEE
170. Fan J, Liu H, Yang W, See J, Zhang A, Lin W (2022) Speed up object detection on gigapixel-level images with patch arrangement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 4653–4661
171. Hao Y, Luo W, Li Y, Zhang B, Bei J (2022) Copy and restricted paste: data augmentation for small object detection in specific scenes. In: Other Conferences. https://api.semanticscholar.org/CorpusID:255084336
172. Ma X, Lu J, Liu X, Kuang H (2023) An acne detector for skin image based on attention enhanced feature pyramid networks. In: 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), vol. 3: pp 151–155. https://doi.org/10.1109/ICIBA56860.2023.10165620
173. Niu T, Zhang Q (2022) Research on small object pattern recognition technology based on computer deep learning: Improved yolox model as an example. In: 2022 2nd International Conference on Social Sciences and Intelligence Management (SSIM), pp 93–98. https://doi.org/10.1109/SSIM55504.2022.10047938
174. Zhang H, Qin L, Li J, Guo Y, Zhou Y, Zhang J, Xu Z (2020) Real-time detection method for small traffic signs based on yolov3. IEEE Access 8:64145–64156. https://doi.org/10.1109/ACCESS.2020.2984554
175. Zhou LQ, Sun P, Li D, Piao JC (2022) A novel object detection method in city aerial image based on deformable convolutional networks. IEEE Access 10:31455–31465. https://doi.org/10.1109/ACCESS.2022.3156953
176. Niu Z, Jiang B, Xu H, Zhang Y (2023) Balance loss for multiattention-based yolov4. In: 2023 5th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP), pp 946–954. https://doi.org/10.1109/ICMSP58539.2023.10171084
177. Fang F, Hoang NM, Xu Q, Lim JH (2023) Data augmentation using corner cutmix and an auxiliary self-supervised loss. In: 2023 IEEE International Conference on Image Processing (ICIP), pp 830–834. https://doi.org/10.1109/ICIP49359.2023.10222009
178. Zhou H, Ma A, Niu Y, Ma Z (2022) Small-object detection for uav-based images using a distance metric method. Drones 6(10):308
179. Hu B, Liu Y, Chu P, Tong M, Kong Q (2022) Small object detection via pixel level balancing with applications to blood cell detection. Front Physiol 13:911297
180. Fujii S, Akita K, Ukita N (2021) Distant bird detection for safe drone flight and its dataset. In: 2021 17th International Conference on Machine Vision and Applications (MVA), pp 1–5. IEEE
181. Mucientes M, Seidenari L, Del Bimbo A (2023) Downsampling gan for small object data augmentation. neural networks (CNNs) 7:6
182. Mu J, Li S, Liu Z, Zhou Y (2021) Integration of gradient guidance and edge enhancement into super-resolution for small object detection in aerial images. IET Image Proc 15(13):3037–3052
183. Rabbi J (2020) Tiny object detection in remote sensing images: End-to-end super-resolution and object detection with deep learning. Unknown
184. Rabbi J, Ray N, Schubert M, Chowdhury S, Chao D (2020) Small-object detection in remote sensing images with end-to-end edge-enhanced gan and object detector network. Remote Sensing 12(9):1432
185. Wang H, Wang J, Bai K, Sun Y (2021) Centered multi-task generative adversarial network for small object detection. Sensors 21(15):5194
186. Stachoń M, Pietroń M (2023) Chosen methods of improving small object recognition with weak recognizable features. In: Future of Information and Communication Conference, pp 270–285. Springer
187. Courtrai L, Pham MT, Friguet C, Lefèvre S (2020) Small object detection from remote sensing images with the help of object-focused super-resolution using wasserstein gans. In: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, pp 260–263. https://doi.org/10.1109/IGARSS39084.2020.9323236

188. Courtrai L, Pham MT, Lefèvre S (2020) Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks. Remote Sensing 12(19):3152
189. Ji H, Gao Z, Liu X, Zhang Y, Mei T (2021) Small object detection leveraging on simultaneous super-resolution. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp 803–810. IEEE
190. Fu Y, Zheng C, Yuan L, Chen H, Nie J (2021) Small object detection in complex large scale spatial image by concatenating srgan and multi-task wgan. In: 2021 7th International Conference on Big Data Computing and Communications (BigCom), pp 196–203. IEEE
191. Chen Y, Li J, Niu Y, He J (2019) Small object detection networks based on classification-oriented super-resolution gan for uav aerial imagery. In: 2019 Chinese Control And Decision Conference (CCDC), pp 4610–4615. IEEE
192. Deshmukh AA, Kumari M, Krishnaiah VVJR, Bandhekar S, Dharani R (2023) Object detection and recognition in remote sensing images by employing a hybrid generative adversarial networks and convolutional neural networks. International Journal of Advanced Computer Science and Applications **14**(9). https://doi.org/10.14569/IJACSA.2023.0140965
193. Ahmad T, Chen X, Saqlain AS, Ma Y (2021) Fpn-gan: multi-class small object detection in remote sensing images. In: 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), pp 478–482. IEEE
194. Du S, Zhang B, Zhang P, Xiang P, Xue H (2021) Fa-yolo: An improved yolo model for infrared occlusion object detection under confusing background. Wirel Commun Mob Comput 2021:1–10
195. Wang S, Peng H, Zhao C, Wu Y (2022) Remote sensing image recognition network design based on filter pruning and lightweight gan. In: Second International Symposium on Computer Technology and Information Science (ISCTIS 2022), vol. 12474: pp 531–536. SPIE
196. Kim JH, Hwang Y () Gan-based synthetic data augmentation for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing 60: 1–12
197. Deng Z, Dong Z, Yang F, Xia L (2020) Data augmentation method of remote sensing image based on transfer learning and vgg model. In: AOPC 2020: Display Technology; Photonic MEMS, THz MEMS, and Metamaterials; and AI in Optics and Photonics, vol. 11565, pp 172–179. SPIE
198. Zhao J, Yang D, Li Y, Xiao P, Yang J (2022) Intelligent matching method for heterogeneous remote sensing images based on style transfer. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15:6723–6731
199. Tang H, Miao X, Shi J, Hua Z, Yang D (2022) An intelligent registration method of heterogeneous remote sensing images based on style transfer. In: Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition, pp 637–642
200. Dolgaia L, Illarionova S, Nesteruk S, Krivolapov I, Baldycheva A, Somov A, Shadrin D (2023) Apple tree health recognition through the application of transfer learning for uav imagery. In: 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA), pp 1–8. IEEE
201. Jiang Y, Zhu B, Xie B (2020) Remote sensing images data augmentation based on style transfer under the condition of few samples. In: Journal of Physics: Conference Series, vol. 1653: p. 012039. IOP Publishing
202. Tetila EC, Machado BB, Menezes GK, Oliveira AdS, Alvarez M, Amorim WP, Belete NADS, Da Silva GG, Pistori H (2019) Automatic recognition of soybean leaf diseases using uav images and deep convolutional neural networks. IEEE Geosci Remote Sens Lett 17(5):903–907
203. Pandian JA, Geetharamani G, Annette B (2019) Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques. In: 2019 IEEE 9th International Conference on Advanced Computing (IACC), pp 199–204. IEEE
204. Xu X, Zhang H, Ma Y, Liu K, Bao H, Qian X (2023) Transdet: Toward effective transfer learning for small-object detection. Remote Sensing 15(14):3525
205. Li H, Wu L, Niu Y, Wang C, Liu T (2019) Small sample meta-leaming towards object recognition through uav observations. In: 2019 IEEE International Conference on Unmanned Systems (ICUS), pp 860–865. IEEE
206. Chen L, Liu G, Tan Y, Sun Z, Ge H, Duan F, Zhu C (2021) A ua-net based salient object detection method for uav. In: 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp 1662–1667. IEEE
207. Yue H, Yan C, Mi T, Yan S, He X (2023) A small target detection method for uav aerial images based on improved yolov5. In: International Conference on Internet of Things and Machine Learning (IoTML 2023), vol. 12937: pp 88–93. SPIE
208. Ren Y, Zhu C, Xiao S (2018) Small object detection in optical remote sensing images via modified faster r-cnn. Appl Sci 8(5):813
209. Gao Z, Yang G, Li E, Liang Z (2021) Novel feature fusion module-based detector for small insulator defect detection. IEEE Sens J 21(15):16807–16814

210. Li Y, Zhang X (2023) Object detection for uav images based on improved yolov6. IAENG International Journal of Computer Science **50**(2)
211. Liang X, Zhang J, Zhuo L, Li Y, Tian Q (2019) Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis. IEEE Trans Circuits Syst Video Technol 30(6):1758–1770
212. Liu B, Du S, Li J, Wang J, Liu W (2022) Small object detection using multi-scale feature fusion and attention. In: 2022 41st Chinese Control Conference (CCC), pp 7246–7251. IEEE
213. Rui T, Hu Y, Yang C, Wang D, Liu X (2021) Research on fast natural aerial image mosaic. Computers & Electrical Engineering 90:107007
214. Xiao J, Guo H, Zhou J, Zhao T, Yu Q, Chen Y, Wang Z (2023) Tiny object detection with context enhancement and feature purification. Expert Syst Appl 211:118665