



Review

Image data augmentation techniques based on deep learning: A survey

Wu Zeng*

Engineering Training Center, Putian University, Putian 351100, China

* **Correspondence:** Email: wuzeng515@ptu.edu.cn.

Abstract: In recent years, deep learning (DL) techniques have achieved remarkable success in various fields of computer vision. This progress was attributed to the vast amounts of data utilized to train these models, as they facilitated the learning of more intricate and detailed feature information about target objects, leading to improved model performance. However, in most real-world tasks, it was challenging to gather sufficient data for model training. Insufficient datasets often resulted in models prone to overfitting. To address this issue and enhance model performance, generalization ability, and mitigate overfitting in data-limited scenarios, image data augmentation methods have been proposed. These methods generated synthetic samples to augment the original dataset, emerging as a preferred strategy to boost model performance when data was scarce. This review first introduced commonly used and highly effective image data augmentation techniques, along with a detailed analysis of their advantages and disadvantages. Second, this review presented several datasets frequently employed for evaluating the performance of image data augmentation methods and examined how advanced augmentation techniques can enhance model performance. Third, this review discussed the applications and performance of data augmentation techniques in various computer vision domains. Finally, this review provided an outlook on potential future research directions for image data augmentation methods.

Keywords: image data augmentation; generative adversarial networks; image mixing; sample augmentation; deep learning

1. Introduction

In recent years, deep learning (DL) has achieved exciting success in various fields of computer vision. For instance, it has enabled more accurate identification and localization in object detection [1], image analysis [2], improved the performance of image context restoration [3], image classification [4], sentiment analysis [5, 6] in natural language processing [7, 8], significantly enhanced the capabilities of speech recognition [9–11], speech synthesis [12], fuzzy image classification [13, 14], and led to the generation of more realistic images by generative adversarial networks (GANs) [15, 16]. The success

in these areas is largely attributed to the support of large-scale datasets [17–19], high-performance computing devices, and excellent network architectures [20–23]. Among these, the quality of datasets plays a decisive role in most tasks. Many detection models with strong performance are trained on large datasets. These extensive datasets encompass the majority of scenarios involving target objects, facilitating the model’s learning of feature information from multiple angles and poses.

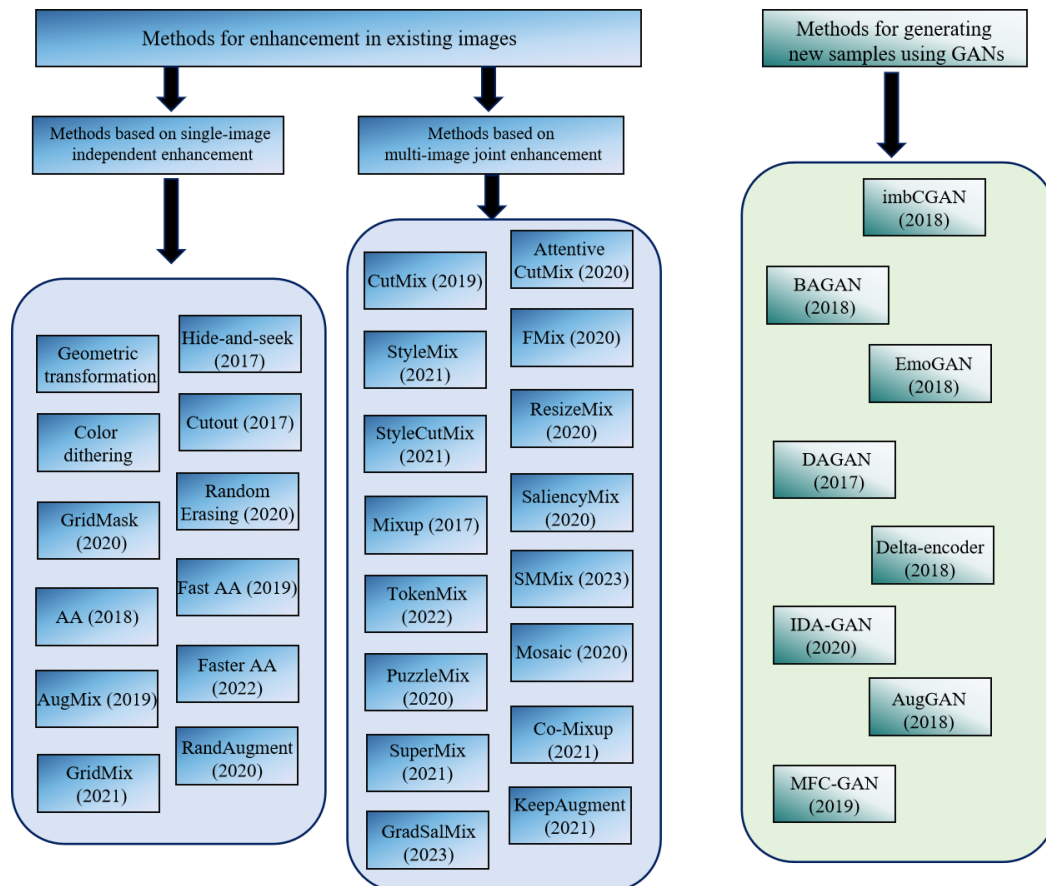


Figure 1. This review provides a rough classification of most image data augmentation methods.

Based on the above discussion, we recognize the importance of a large amount of data. Therefore, in most DL tasks, the larger the amount of data in the dataset, the higher the performance of the trained model. However, when data is limited or insufficient, models tend to suffer from issues such as low detection accuracy and poor generalization. Typically, it is challenging for individuals or small groups to gather an ideal number of images for model training in DL tasks. Even with an adequate number of images, annotating them for different tasks can be a time-consuming and labor-intensive endeavor. To address this, various strategies have been proposed to enhance model performance despite limited image quantities in datasets. These include few-shot learning (FSL) strategies [24–26] and data augmentation strategies [27–29]. Among them, image data augmentation is a relatively straightforward approach. Most data augmentation strategies can generate numerous synthetic images

based on existing samples, offering a quick and convenient solution.

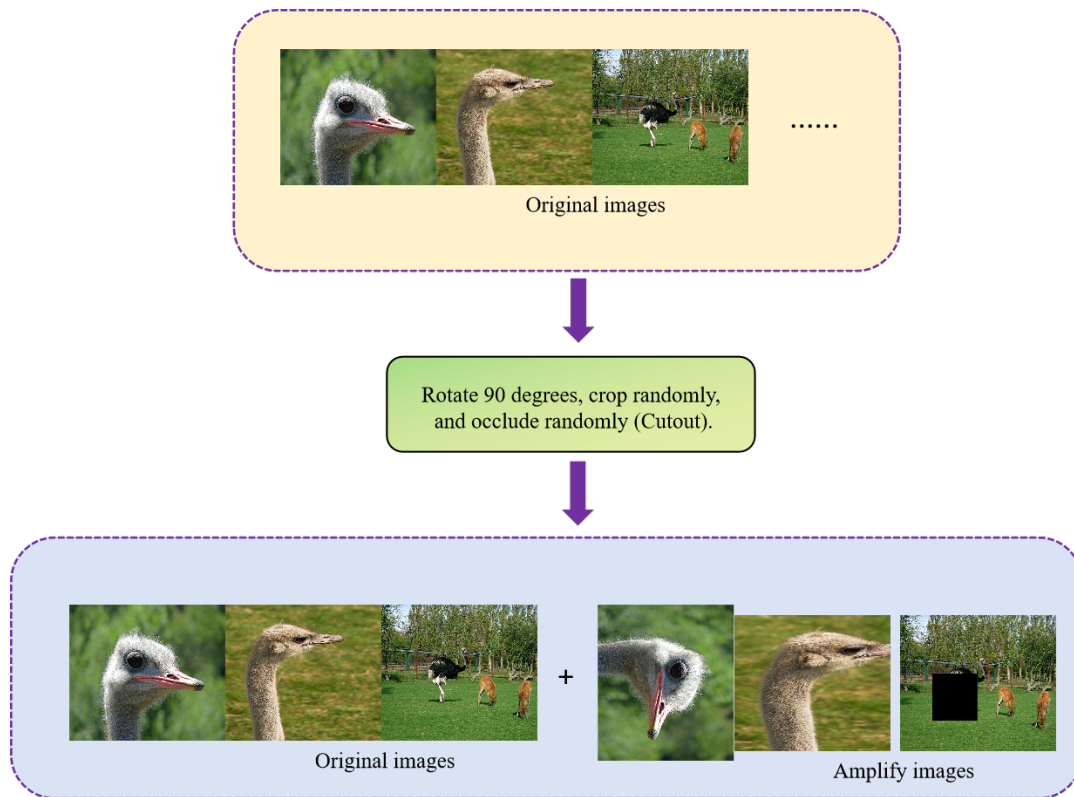


Figure 2. The method of enhancing based on the original image can amplify samples on the basis of the original image through a simple data enhancement strategy. By generating samples with certain differences from the original image, the model can be given different types of images, thereby promoting the improvement of model performance. However, the samples generated by this method will still have partially identical pixel content as the original image.

The iterative process of learning feature information from images through DL methods is essentially a process of learning these features. The greater the number of images, the more diverse they become, facilitating the model's ability to learn feature information about target objects in different scenarios. This in turn enhances the recognition ability and generalization performance of the model in various scenarios. Data augmentation methods can generate additional virtual samples (distinct from the original images in the dataset), enabling the model to learn feature information from a wider range of image styles and achieve optimal performance. Obviously, image data augmentation methods are effective in improving model performance. Therefore, they have also been widely studied by scholars. As shown in Figure 1, this review broadly categorizes data augmentation methods into two types: those that generate new image samples based on existing data (the method of directly enhancing the original image is shown in Figure 2) and those that utilize GANs to generate new image samples (the general strategy of this method is shown in Figure 3). Furthermore, based on different augmentation strategies,

the methods that generate new image samples from existing data are further subdivided into approaches that independently augment single samples and those that involve multiple images jointly.

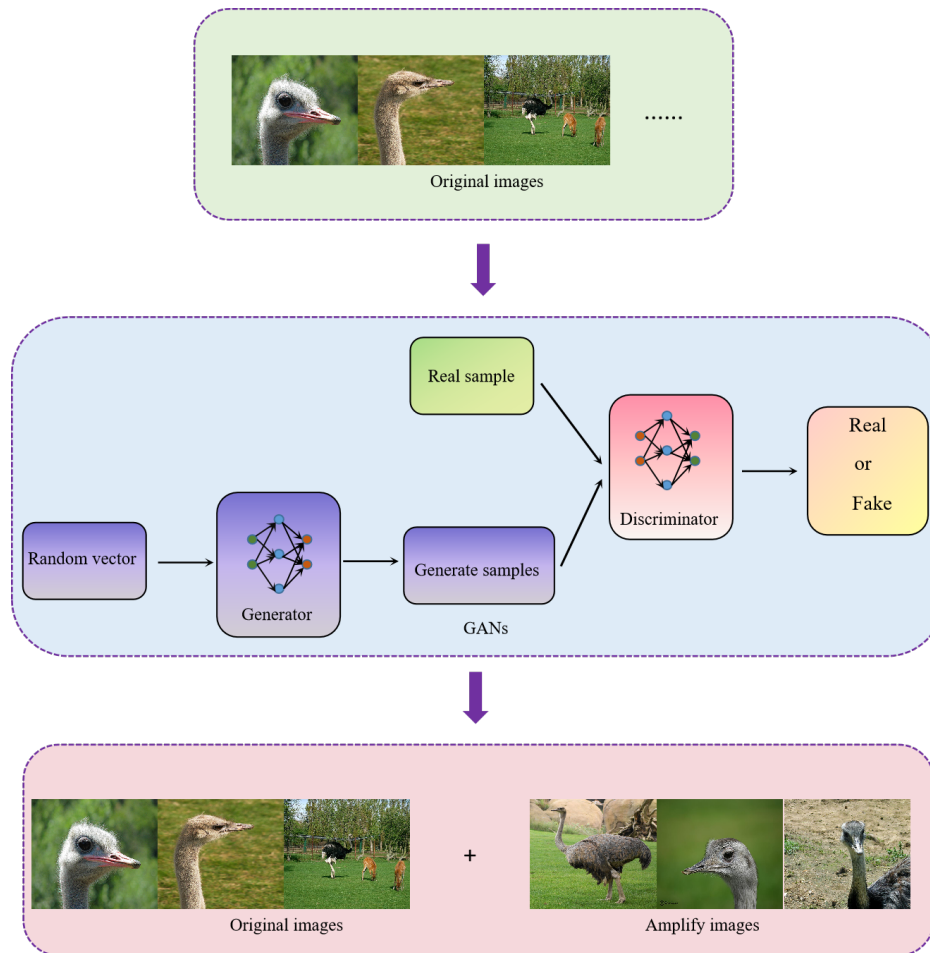


Figure 3. Methods for generating new samples using GANs can produce images with different pixel content from the original images (i.e., the generated images can be completely different from the original ones) through the GAN model and partial input samples and conditions. By adding conditions, some methods can also generate new image samples that meet specific requirements. (note: amplified samples are only examples and not generated by real GANs models)

The organization of the chapters in this review is roughly outlined in Figure 4. The second section introduces methods for generating new images based on existing data. The third section presents approaches for generating new sample images using GANs-based methods. The fourth section explores the application of some data augmentation techniques in popular DL domains such as image classification, object detection, and semantic segmentation. The fifth section summarizes the current achievements of data augmentation methods and provides a prospect for future research. Finally, the sixth section concludes the entire review.

In some previous works, Shorten et al. [27] focused on designing practical applications of GANs-based augmentation methods. Khalifa et al. [28] emphasized traditional data augmentation techniques and some GANs-based methods. Alomar et al. [29] concentrated on the application of data augmentation technologies in image segmentation and classification. Unlike these reviews, my comprehensive review not only provides a detailed introduction to traditional and advanced data augmentation methods (including GANs and various image blending enhancements), but also offers a thorough and specific explanation of the application of some advanced methods in various visual fields, such as image segmentation, object detection, and image classification. This review aims to achieve a more comprehensive analysis and comparison.

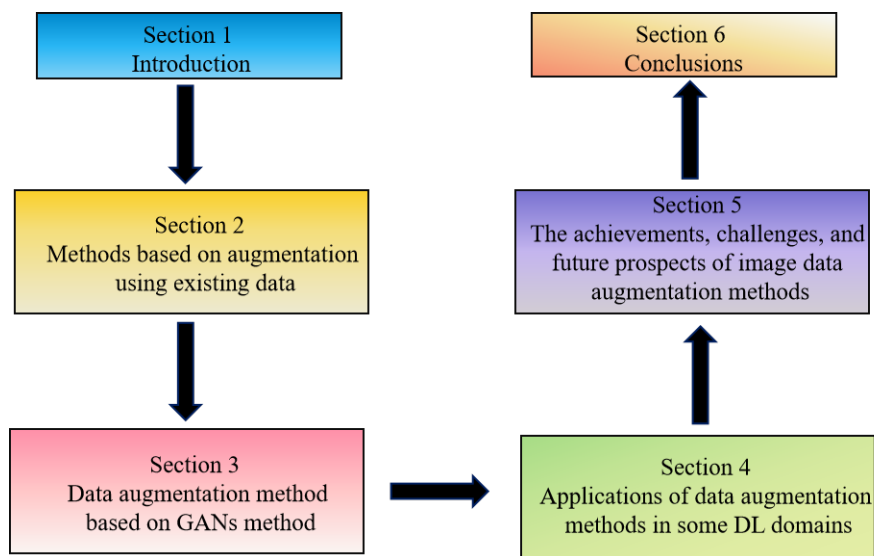


Figure 4. The specific architecture arrangement of this review.

2. Methods based on augmentation using existing data

To enhance the performance of models trained through DL, researchers in related fields have proposed numerous data augmentation methods based on generating new samples from existing images. These image data augmentation methods can broadly be classified into two categories: single-sample independent augmentation methods and multi-image combined augmentation methods. Section 2.1 introduces single-sample independent augmentation methods for image data, while multi-image combined augmentation methods will be presented in Section 2.2.

2.1. Methods based on single-sample independent augmentation

As the name suggests, methods based on single-sample independent augmentation involve enhancing an existing image within its own context. In practical DL vision tasks, image data augmentation methods based on geometric transformations are commonly used. These methods generate images as shown in Figure 5 (by applying various simple transformation strategies to a given

original image, numerous new samples can be generated). Typical examples include random cropping, translation cropping, rotations by 90, 180, and 270 degrees, horizontal flipping, vertical flipping, and random scaling. Random cropping involves randomly cutting away portions of the original image along its edges, with the amount of cropping manually adjustable. Generally, the amount of cropping should not be excessive, as this can lead to a loss of significant pixel information (i.e., when the cropped area includes the target object in the original image). Translation cropping, on the other hand, involves cropping the original image along a specific edge and can be used as a complementary method to random cropping. Rotations by 90, 180, and 270 degrees involve rotating the image to generate new samples. Additionally, there are methods that involve rotating the image by other angles to generate new images. By rotating the images, the model can learn feature information about the target object from different perspectives. Horizontal and vertical flipping involve flipping the image along the horizontal or vertical axis, respectively, and produce similar effects as rotating the image. Random scaling involves resizing the image randomly without changing its image content. The augmented images generated by these methods share a high degree of similarity with the original images and do not produce significantly different samples. Therefore, while these methods are easy to implement, their limited ability to generate diverse samples results in a limited improvement in model performance.

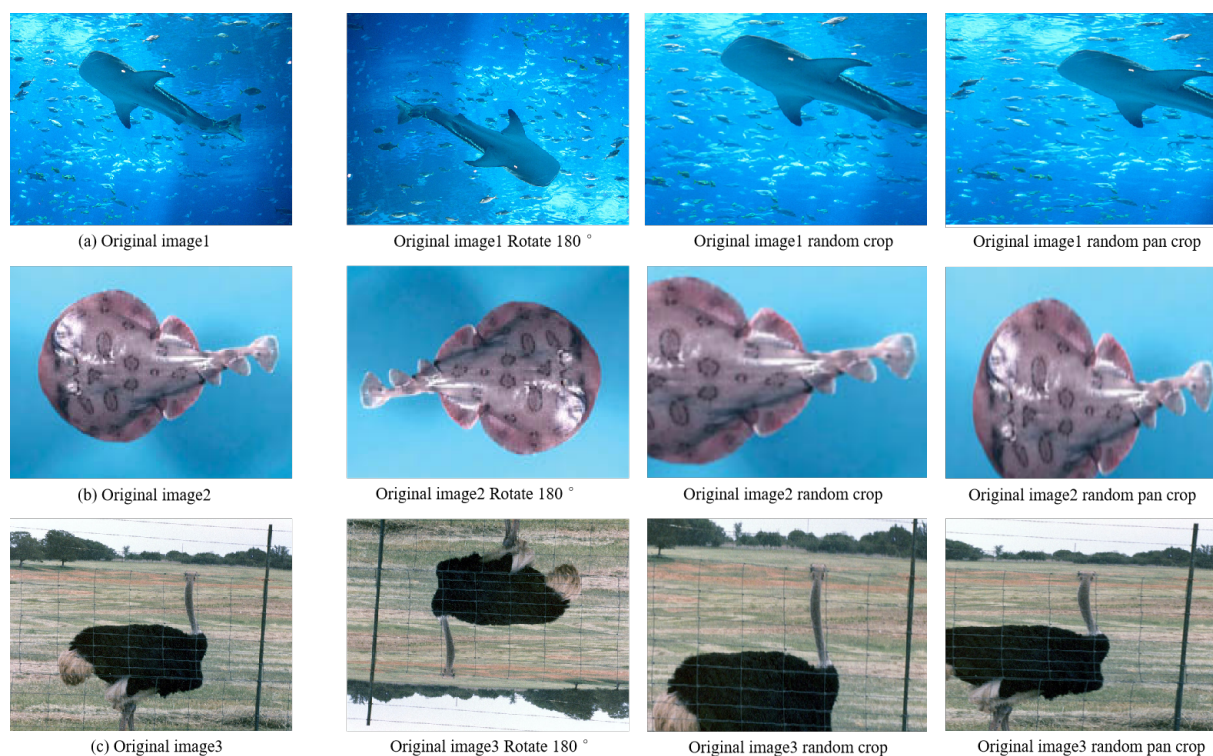


Figure 5. Some simple and classic image data augmentation methods generate sample displays.

In addition to the methods of enhancing image samples through geometric transformations

mentioned above, there are also techniques such as adding noise, applying blurring effects, and performing color changes. Commonly used noises for image enhancement include Gaussian noise and salt-and-pepper noise, as demonstrated in Figure 6. By incorporating these noise points into the image, new samples with some interfering pixels can be generated. Furthermore, color jittering involves increasing or decreasing certain color channel values in the image, or altering the order of color channels, resulting in a jittering effect that creates color variations in the original pixel regions, yielding new samples. Blurring, on the other hand, involves applying a blurring effect to certain regions of the original image, resulting in samples with partially blurred areas. Although these methods enrich the categories of data augmentation techniques, the generated image samples still exhibit a high degree of consistency with the original images, making it challenging to produce significantly different samples compared to the original images.

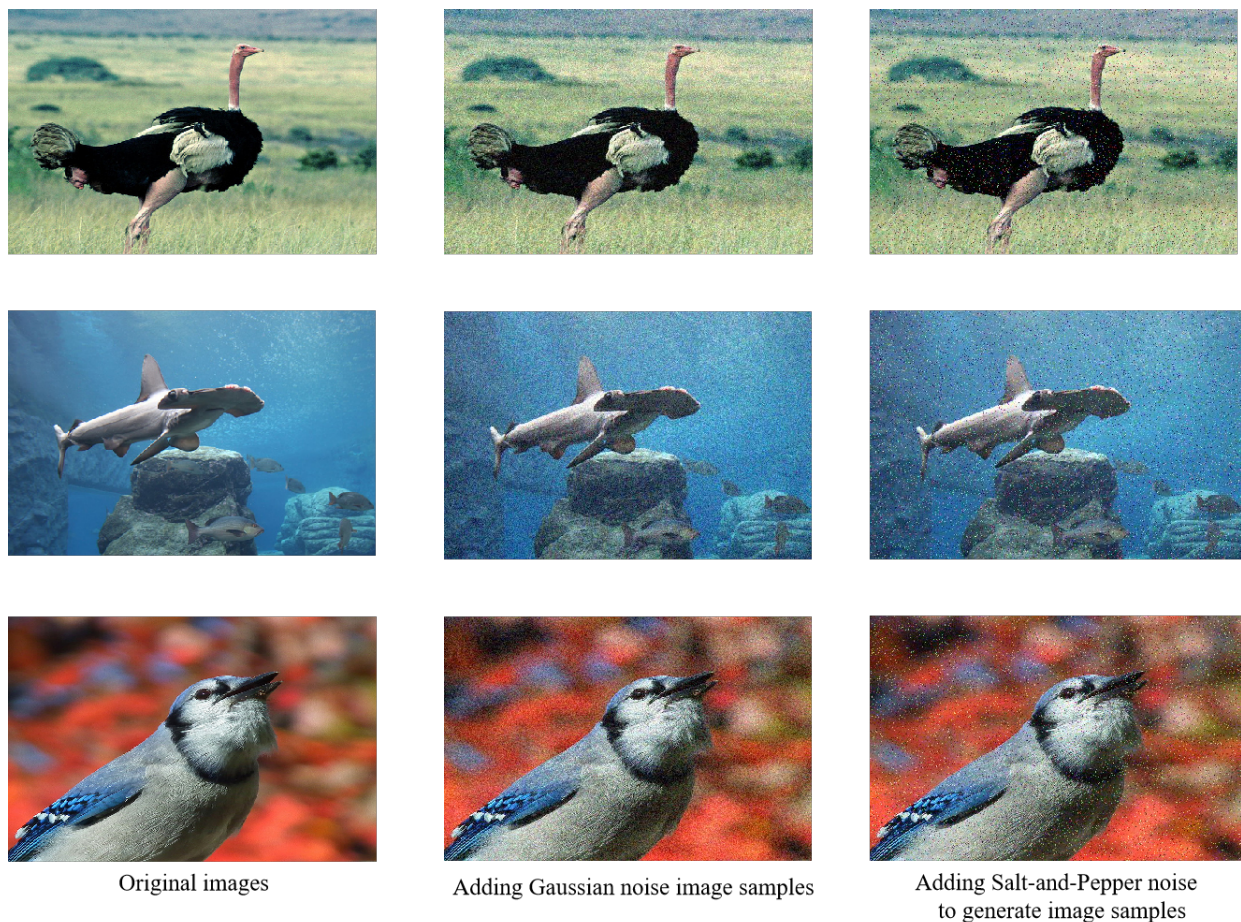


Figure 6. Some simple and classic image data augmentation methods generate sample displays.

Figure 7 illustrates the process of image processing using the Cutout [30] algorithm. The new image samples generated by Cutout are shown in Figure 8. Cutout achieves image augmentation by randomly removing some pixel regions from the image. During network training, the model tends

to focus on specific regions in the image, neglecting other areas. This can lead to a situation where the model's detection accuracy peaks and then stops improving with further iterations, potentially resulting in overfitting. The Cutout method addresses this by forcing the model to pay more attention to the remaining pixel regions, particularly those that may have been previously ignored but are crucial for the model's performance. However, this approach also has some limitations. First, the augmented regions consist of meaningless pixels, which can reduce the model's training efficiency. Second, the cropped regions in Cutout are randomly generated, and the size of these regions is also random. This means that important regions in the image may be excessively removed.

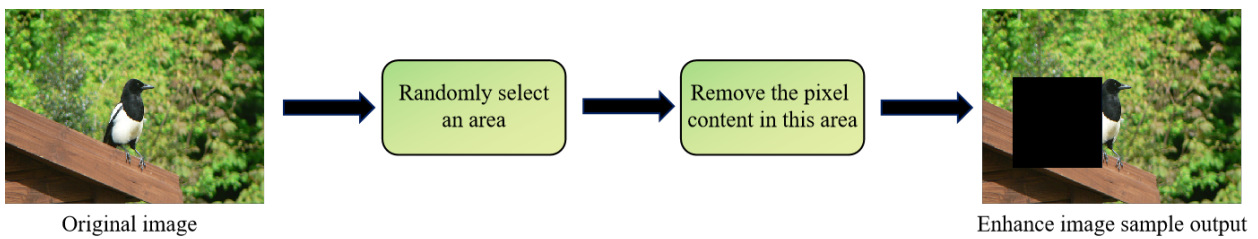


Figure 7. The general operation process of the Cutout algorithm.



Figure 8. Examples of image samples that can be generated by the Cutout algorithm. Cropping regions of random size can generate richer image samples.

Similar to Cutout, Zhong et al. [31] introduced another method called “Random Erasing.” This technique generates samples as shown in Figure 9, where a rectangular region is randomly selected in the image, and the pixels within that region are occluded. Unlike Cutout, Random Erasing allows for various ways to fill the occluded region, such as using random mosaics or meaningless pixel blocks. While this method shares some similarities with Cutout in terms of its strategy and core idea, it also faces similar challenges and limitations.

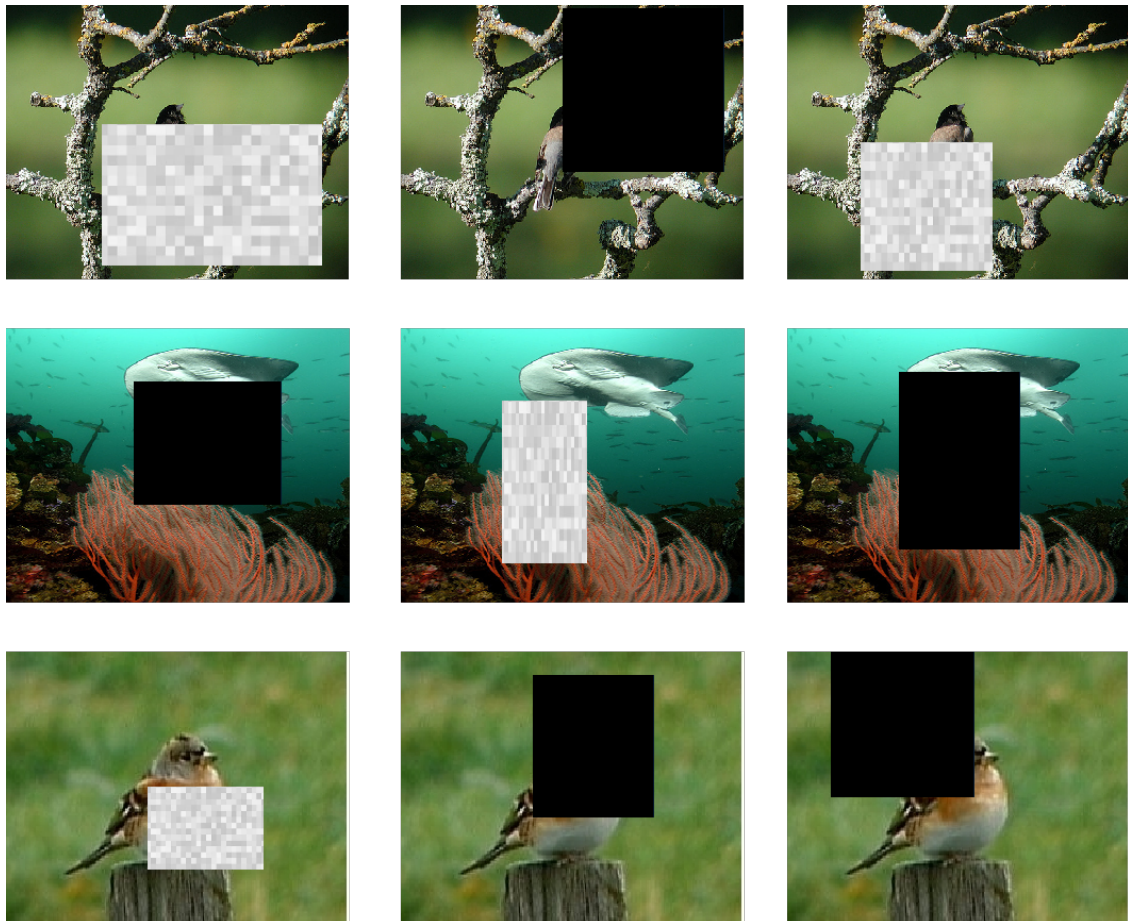


Figure 9. The random erasing method can generate sample examples. Compared to the Cutout method, this method has more diverse forms in filling the cropped area.

Singh et al. [32] proposed the hide-and-seek method based on Cutout, introducing improvements in the selection strategy of the cropped regions. As shown in Figure 10, this approach no longer removes the entire cropped region as a whole but instead divides the image into equal parts and randomly occludes some of these regions. This provides more flexibility in handling occluded areas. However, there is still a risk of removing important regions in the image. Chen et al. [33] further improved upon Cutout by introducing GridMask, which generates new samples as shown in Figure 11. This method considers that removing an entire rectangular region of pixels simultaneously may increase the probability of losing important regions in the original image. Therefore, GridMask disperses the

cropped rectangular blocks according to a selected strategy, breaking a single rectangular region into four, nine, or a different number of smaller blocks (the approximate strategy for generating samples using this method is shown in Figure 11). This helps mitigate some of the risks associated with Cutout.

In addition to the above methods, Cubuk et al. [34] introduced a data augmentation strategy called “AutoAugment” (AA), which presupposes numerous sub-strategies for data augmentation within the algorithm. Briefly, the model initially searches for the two most suitable sub-strategies to augment a single input sample and then applies them sequentially. Based on AA, Lim et al. [35] made improvements and optimizations. To reduce the search time among sub-strategies in the AA approach, Fast AutoAugment (Fast AA) was proposed, which significantly decreases the time required to generate samples using data augmentation methods. Faster AA introduced by Hataya et al. [36], optimizes the search strategy for data augmentation through approximate gradient information, achieving a faster search speed. It’s worth noting that Faster AA can perform image data augmentation not only on individual images, but also without relying on preset sub-strategies for augmentation. Additionally, this method integrates the backpropagation algorithm for model optimization.

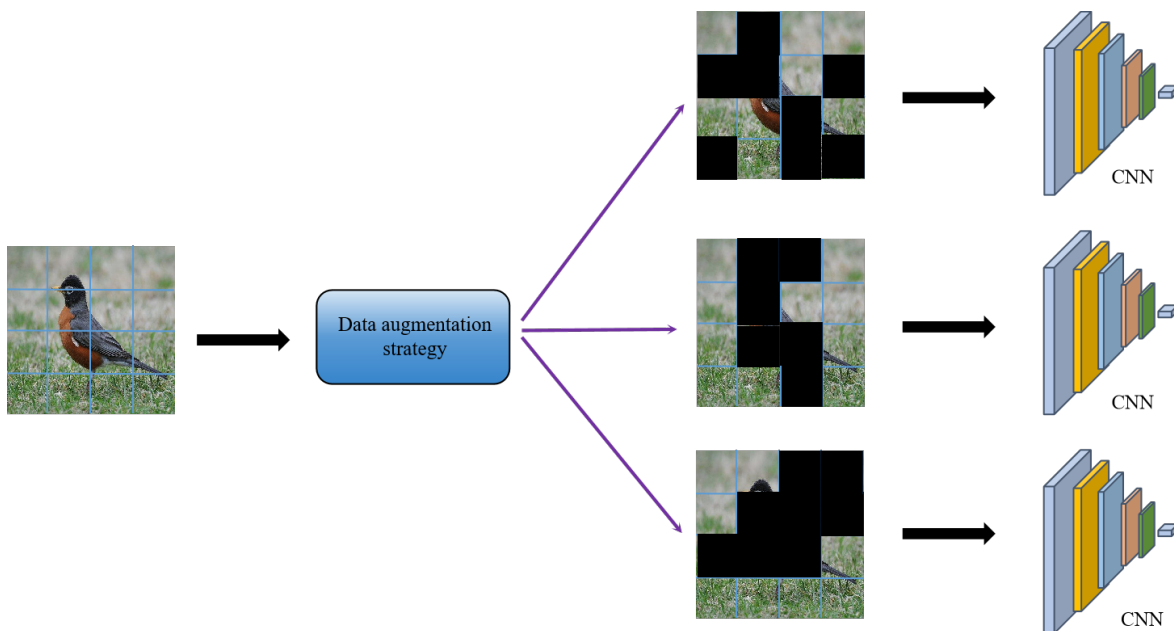


Figure 10. The rough strategy for generating image samples using the Hide-and-Seek method.

Cubuk et al. [37] further enhanced AA by introducing the RandAugment method. Its main strategy is to narrow down the search space and avoid excessive time consumption in searching for data augmentation strategies. This method randomly explores suitable strategies for image sample augmentation during the iteration process, making it suitable for most datasets. Hendrycks et al. [38] proposed the AugMix method for augmenting individual images. Specifically, this approach introduces various transformation techniques, such as rotation and color jittering, into a single image. Then, it blends variants generated by applying some data augmentation methods to the original image with

certain weight proportions, creating a completely new sample. Baek et al. [39] performed data augmentation through context information mapping and random sampling strategies within a single image, naming this method GridMix. Briefly, this approach first sets the context information for a single image sample and then maps each pixel of the image in a one-dimensional space. Subsequently, simple data augmentation methods are applied to process and synthesize these pixels (e.g., random sampling and Gaussian noise). Finally, the augmented information is mapped back to the original space to complete the augmentation process.

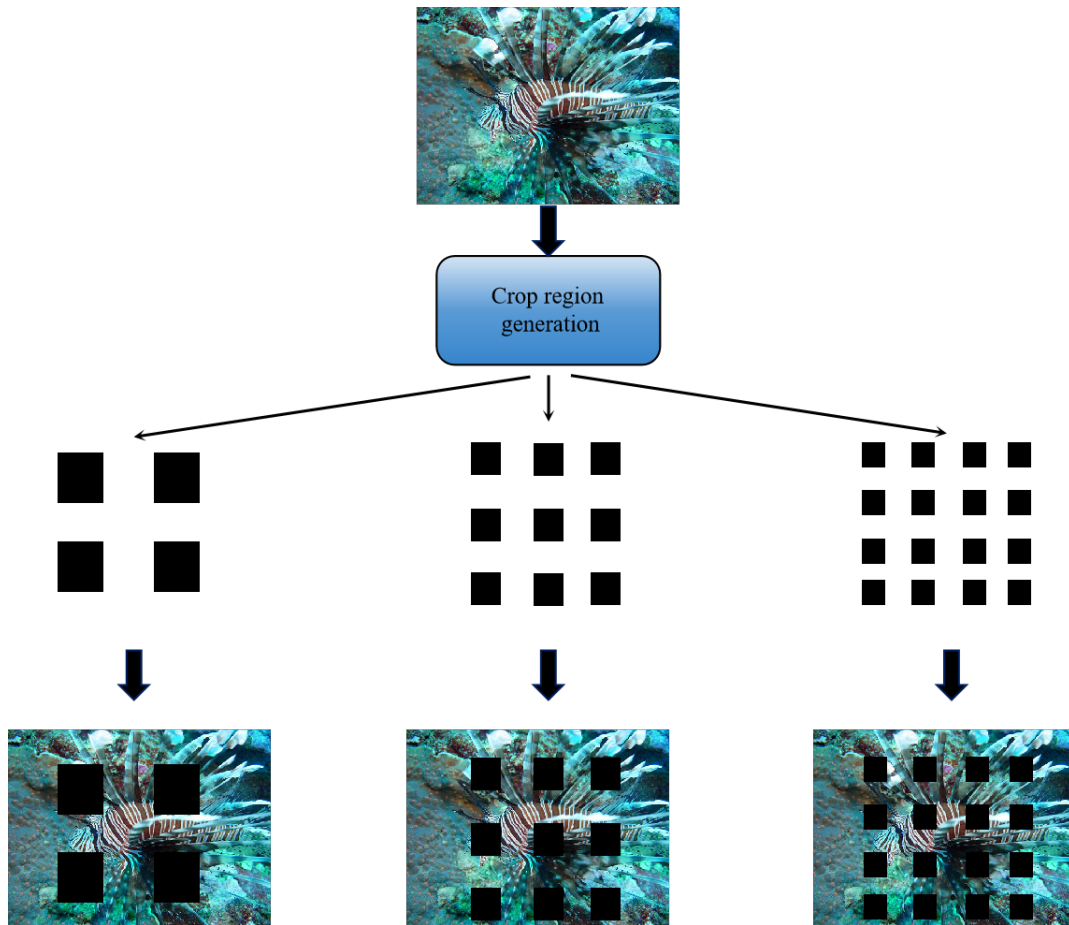


Figure 11. The approximate strategy for generating samples using the GridMask method.

2.2. Methods based on multi-image combined augmentation

Methods such as image geometric deformation, adding noise to images, and Cutout data augmentation all belong to the category of “single-image independent augmentation”, which involves enhancing data within a single image. Although these methods can achieve certain effects, their impact on improving model performance remains insignificant. Techniques like adding noise, Cutout, Random Erasing, and GridMask introduce noise into images, essentially creating meaningless pixels. Excessive learning of such meaningless pixels by the model may adversely affect its recognition capabilities.

Therefore, when the number of images in a dataset is limited, it is crucial to make more efficient use of the available images to generate a higher quality of augmented samples. In this review, methods that combine multiple existing samples from images to generate new samples are referred to as “data augmentation methods based on multi-image combined.”

When discussing “data augmentation methods based on multi-image combined”, it is worth mentioning the CutMix approach proposed by Yun et al. [40]. This method considers the potential impact of meaningless pixel regions in the Cutout technique on the model. Therefore, it replaces these regions with patches from the same location in another image. The final sample label is represented using interpolation. Specifically, as illustrated in Figure 12, CutMix enhances images by combining existing images from the dataset. Based on Cutout, CutMix first selects two images of the same pixel size. It then generates crop regions at the same position and size in both images for augmentation. Figure 13 provides a schematic representation of samples generated using CutMix. Notably, CutMix-generated images do not contain meaningless pixels. The crop region in the original image is replaced with a patch from the corresponding area in another image from the dataset. The absence of meaningless pixel regions in the new samples contributes to improving model training efficiency. The labels of the newly generated samples are primarily determined by two factors: 1) the proportion of the remaining area after data augmentation and 2) the ratio of the patch area to the whole image area. Compared to Cutout, CutMix generates higher-quality samples without meaningless pixels, thereby enhancing the learning efficiency and performance of the model. However, since the crop region and size in CutMix are randomly selected, inappropriate positioning or sizing can degrade the quality of the generated samples and potentially mislead the model.

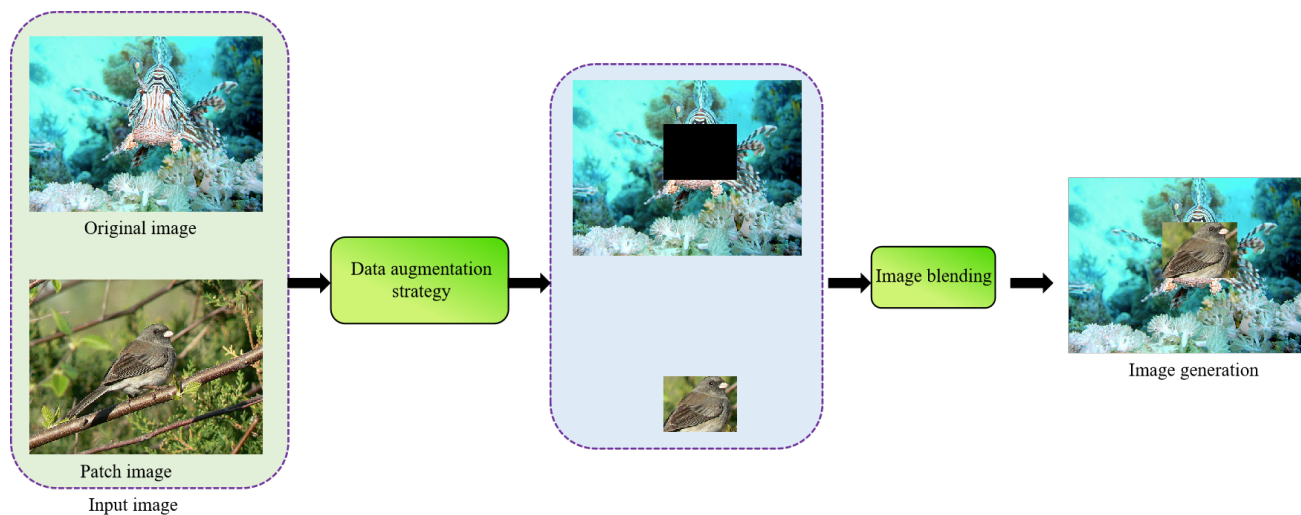


Figure 12. The approximate strategy for generating samples using the CutMix method.

Later, Hong et al. [41] introduced a GAN-based approach for sample generation, proposing the StyleMix method. This technique generates samples with diverse styles by incorporating styles from other images. However, the pure StyleMix method does not significantly improve model performance. Subsequently, the authors combined StyleMix with CutMix, resulting in the StyleCutMix method. This

approach inherits the advantages of CutMix while introducing more style variations, allowing for the generation of images with a wider range of styles and significantly improving the performance of the baseline model. Furthermore, Walawalkar et al. [42] proposed the Attentive CutMix method based on CutMix. This technique considers the importance of different image regions by incorporating an attention mechanism to search for patches within the image. The identified patches are considered critical regions in the augmented image and are pasted onto the corresponding areas in the original image. Experimental results across multiple datasets and networks demonstrate that this method significantly improves model detection performance and outperforms advanced techniques such as CutMix and Mixup [43] (the generated samples are shown in Figure 14). Harris et al. [44] introduced the FMix method for image augmentation. This approach strategically processes high-frequency and low-frequency regions of the image. Subsequently, it enhances the low-frequency image through Fourier sampling variations and performs weighted blending of pixel content using binarization to generate entirely new samples. Additionally, FMix can produce higher-quality samples for model learning without increasing computational complexity.



Figure 13. The CutMix method generates display of image samples.

Qin et al. [45] proposed the ResizeMix method to acquire more information from images. In simple terms, the main strategy of this method is to remove part of the original image. However, instead of selecting patches from the same region in the original image, the entire image is chosen as the patch, compressed to the same size as the crop region, and pasted into it to generate a new sample. Additionally, methods that leverage salient regions in images have garnered attention. Uddin

et al. [46] employed saliency detection techniques to select crop regions in patch images, ensuring that the newly generated samples contain salient regions from multiple images. By increasing the saliency information in the generated samples, the model can acquire rich feature information, thereby improving its performance.



Figure 14. Visualization of new sample images generated by the Mixup method. Through pixel-level blending across the entire region, this approach can produce completely new and diverse samples. This process not only combines visual features from two source images, but also generates intermediate representations that can enhance the generalization capabilities of the model.

Based on CutMix, Bochkovskiy et al. [47] introduced the Mosaic data augmentation technique, primarily used in object detection. This method folds and compresses multiple image samples before stitching them together to create new samples. Figure 15 provides a general representation of the generated samples. The Mosaic approach has gained significant popularity in various YOLO (you only look once) models, and experimental results from most models demonstrate its effectiveness in object detection. Liu et al. [48] proposed the TokenMix method, which utilizes self-attention mechanisms and image tokens to generate new samples. Specifically, the method segments an image into a series of patches using tokens. To capture important contextual information from these patches, a self-attention mechanism is employed. After obtaining relevance scores among the tokens, patch mixing and replacement are performed based on these scores. Similarly, Chen et al. [49] proposed SMMix (self-motivated image mixing), which also aims to enhance the model's ability to capture crucial contextual information in images through the introduction of a self-attention mechanism. This

mechanism searches for scores in the image to determine the correlation between different positions and generates images with important feature information from each input image. Yang et al. [50] presented a new image augmentation strategy called “RecursiveMix.” The main approach of this method involves scaling down previously iterated images and pasting them into a new batch of images. Subsequently, the model’s performance is further improved by adjusting the loss function of the labels. Experimental results indicate that this method enables the model to achieve superior performance by learning information from different scales.



Figure 15. The Mosaic method can generate schematic diagrams of image.

In addition, pixel-level blended image data augmentation methods have been proposed successively. The data augmentation areas for Cutout and CutMix methods are partial areas within the image, while the Mixup method introduced by Zhang et al. [43] performs data augmentation on the entire image area. This algorithm first selects two images of the same pixel size for pixel-level blending across the entire region, generating new samples in this process. The newly generated samples, as shown in Figure 14, demonstrate that multiple sample feature information can be combined. The central idea of the Mixup method is to generate diverse samples through globally random pixel proportion blending. However, its overly random blending strategy may lead to locally chaotic and blurry phenomena, which can hinder the model’s ability to locate target objects. Therefore, improvements are needed in its random blending ratio strategy. Nonetheless, its pixel-level blending concept is worth considering. Subsequently, Verma

et al. [51] proposed the Manifold Mixup method based on this idea. This approach first selects two random samples and propagates them through the forward network architecture to obtain hidden feature representations of these two samples. Interpolation is then performed on the hidden feature vectors to synthesize a completely new vector. Finally, this new vector is passed through the backward propagation network to generate the final prediction. Overall, the image samples generated by the Manifold Mixup method appear smoother and more natural visually. Its effectiveness has also been demonstrated in multiple experimental datasets.

Kim et al. [52] introduced a data augmentation method called “Co-Mixup”, which combines saliency guidance with supermodular diversity. This method aims to use saliency detection techniques to guide the synthesis of new samples, resulting in better-quality samples. Briefly, the method first generates saliency values for the input samples using saliency detection techniques, selects regions with strong saliency from these images, and then blends these regions. Additionally, the method incorporates supermodular diversity techniques to enhance the rationality of the feature space distribution of the generated samples. Overall, this approach improves the generalization ability and performance of DL models. Kim et al. [53] proposed a more flexible method called “PuzzleMix.” This method not only incorporates saliency detection techniques, but also enhances images by dividing them into smaller patches. It selects patches with important feature regions by calculating the saliency strength within each patch. In terms of blending between patches, PuzzleMix employs a flexible strategy that allows for both pixel-level blending and feature-level patch replacement, achieving local optimal solutions. Dabouei et al. [54] introduced the SuperMix method, which utilizes prior knowledge to obtain saliency regions (also known as key regions) in the input images and then blends pixels within these key regions. This method also performs smoothing at the edges of pixel-level blends to ensure more natural-looking generated samples. Additionally, the authors developed a Newton iteration method to expedite image generation, which is reportedly 65 times faster than traditional gradient descent methods.

Gong et al. [55] proposed the KeepAugment method, which suggests avoiding critical regions in the original image during data augmentation. Experimental results combining this strategy with CutMix have demonstrated its effectiveness. Subsequently, Kang et al. [56] introduced saliency detection techniques to generate better-quality samples and presented the GuideMixup method. Briefly, this method generates saliency maps for the input images, equally divides them into matrices, and calculates the important region matrix blocks for two samples. During the final image blending weight assignment, higher weight coefficients are given to these important patches, ensuring that the generated images contain more saliency feature content. Hong et al. [57] proposed the GradSalMix method by combining image gradient information with saliency information. This method involves cropping patches near the maximum saliency response from one image and directly transferring them to another image to complete the generation of new samples.

3. Data augmentation method based on GANs method

In addition to the above methods of generating entirely new samples using existing images, there are also approaches that leverage GANs for this purpose. Classic GAN methods include GAN, CGAN (conditional GAN) [58] (as shown in Figure 16), and ACGAN (auxiliary classifier GAN) [59] (as shown in Figure 17). GAN-based methods offer unique advantages as they can effortlessly generate

entirely new samples without the need for extensive augmentation operations on the original images, thereby significantly expanding the diversity and possibilities of the dataset.

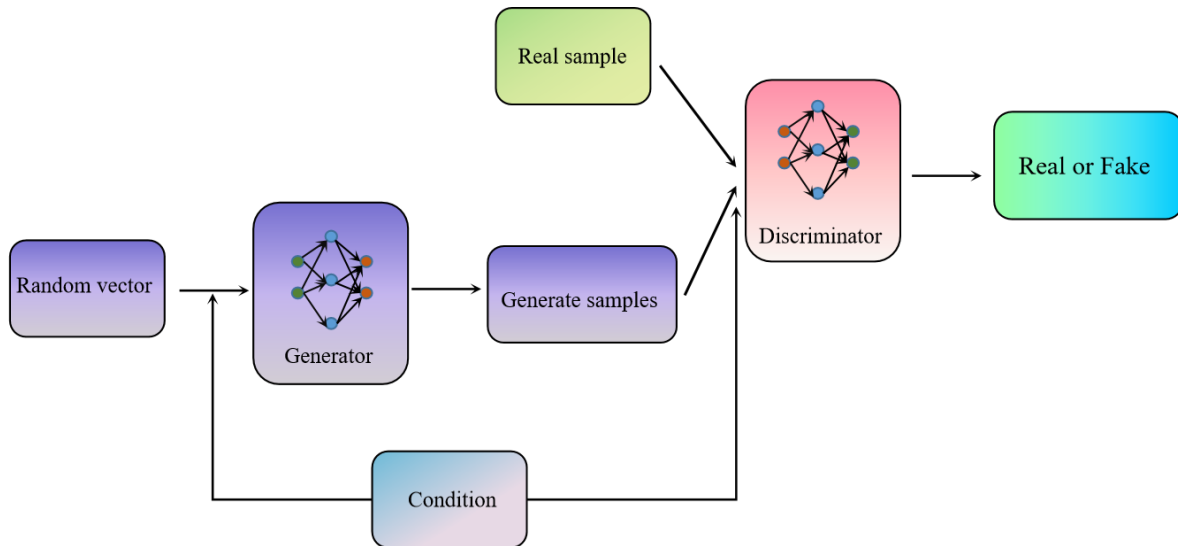


Figure 16. The general architecture of CGAN method.

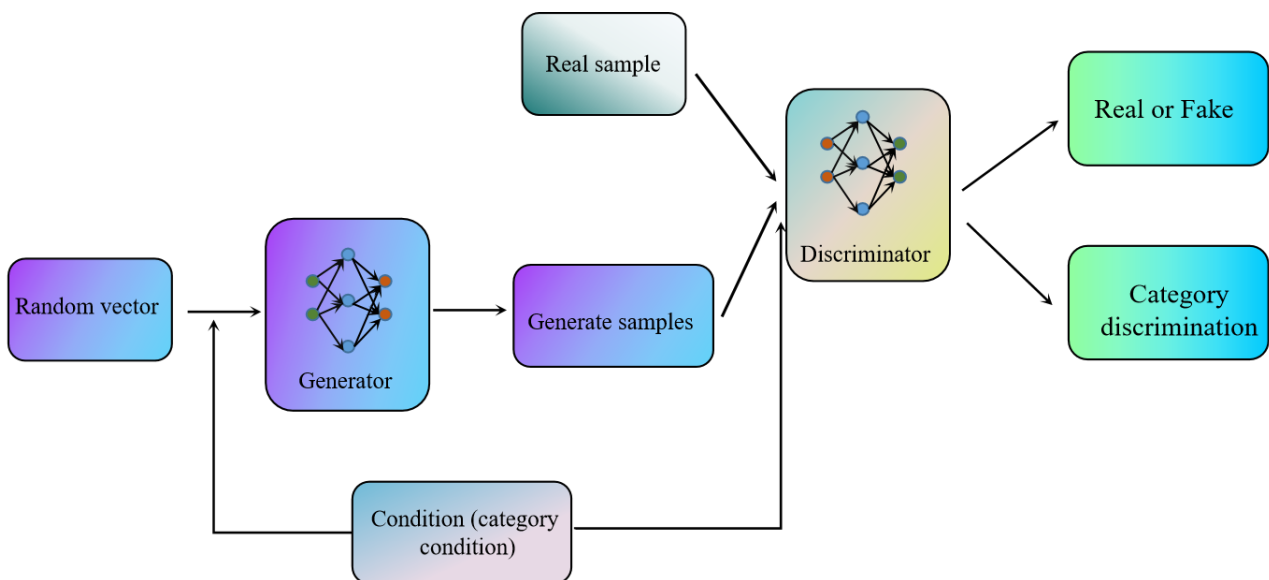


Figure 17. The general architecture of ACGAN method.

To more effectively address the issue of insufficient minority class samples in imbalanced datasets, Douzas et al. [60] proposed a method called “imbCGAN” (imbalance conditional GAN). This method integrates the strategies of CGANs to not only generate more realistic samples that align with the

distribution patterns of the minority class, but also effectively mitigate the scarcity of minority class samples. By emphasizing the authenticity of the generated samples and their fit with the original data distribution, imbCGAN demonstrates superior performance in imbalanced datasets.

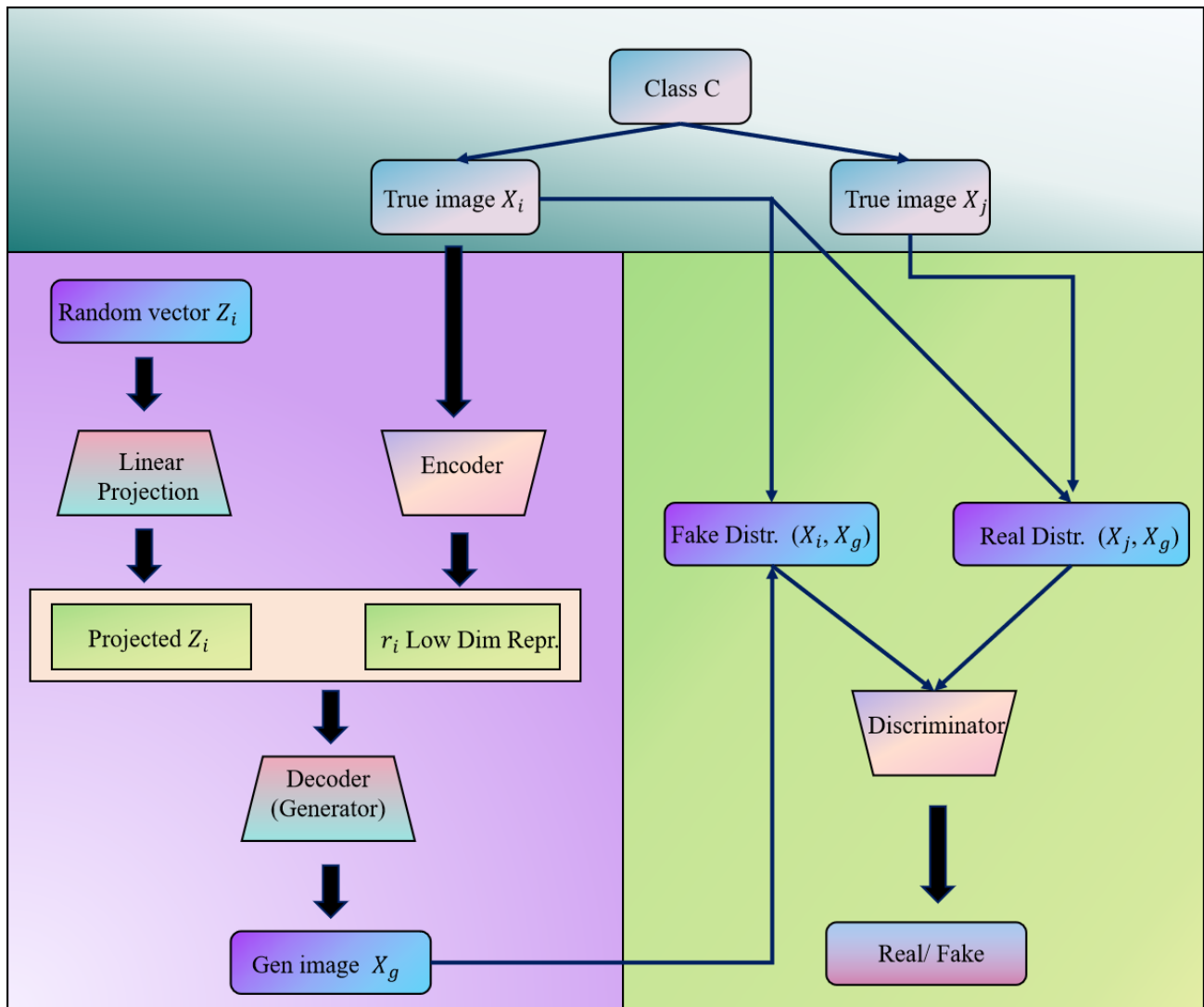


Figure 18. The general architecture of DAGAN method.

Subsequently, Antoniou et al. [61] introduced the DAGAN (data augmentation GAN) method. This approach leverages the flexibility of conditional GANs to learn intra-class features from source domains without restricting categories. The general process of this method is shown in Figure 18. This allows DAGAN to generate new samples that are highly similar to the original ones while also enabling it to generate entirely new categories. Compared to traditional data augmentation methods, DAGAN breaks some limitations and provides a new perspective for handling imbalanced datasets. Mariani et al. [62] proposed the BAGAN (balancing GAN) method, which also focuses on augmenting minority class samples in imbalanced datasets. Considering that standard GAN strategies often require

a large number of training samples to achieve good classification results, which minority class samples may not satisfy, BAGAN adopts a new training strategy. This strategy fully utilizes images from both minority and majority classes for training and successfully generates more images related to the minority class by introducing category conditions in the latent space. The effectiveness of this approach has been validated in multiple experiments.

Huang et al. [63] addressed the challenge of maintaining consistent image structure and objects across domains and introduced the AugGAN method. This method incorporates a novel structure-aware network that integrates components such as an encoder, generator, and discriminator to achieve more realistic sample generation. AugGAN exhibits excellent performance when handling complex datasets, offering a new solution for cross-domain data augmentation. Zhu et al. [64] proposed EmoGAN (Emotion GAN), which leverages CycleGAN technology to tackle the issue of unbalanced category labels in emotional image datasets. This method searches for optimal decision boundaries between adjacent categories to generate new samples. Additionally, to alleviate the common problem of vanishing gradients, EmoGAN employs the least squares loss function as the loss function for the GAN network, further improving training stability and the quality of generated samples. Schwartz et al. [65] proposed the Delta-encoder method to address the challenge of insufficient samples in FSL scenarios. This approach utilizes an improved encoder that can synthesize new samples relying only on a limited number of samples. Among other recent methods, MFC-GAN [66] (multiple fake class GAN) stands out for its novel approach to handling sample scarcity in imbalanced datasets. This method generates more diverse minority class samples by constructing multiple fake classes, leading to better-quality samples. Yang et al. [67] introduced the IDA-GAN (imbalanced data augmentation GAN) method. This approach effectively addresses the issue of class imbalance by skillfully combining GANs with variational autoencoders to generate samples. In this method, the variational autoencoder is used to learn the spatial distribution of majority and minority classes, capturing patterns that aid in generating more realistic minority class samples. Furthermore, IDA-GAN can capture subtle feature differences between different categories, providing richer feature information for subsequent tasks. Through extensive experimental validation, IDA-GAN has demonstrated superior performance on multiple imbalanced datasets. The introduction of these methods has expanded the local methods for generating new samples.

From the detailed discussion above, we can observe that data augmentation methods based on GANs demonstrate significant advantages when dealing with imbalanced datasets. Compared to traditional data augmentation techniques, these methods exhibit notably less dependence on existing data samples. Without the need for direct transformations or enhancements on the images of existing samples, they can generate completely new samples. This advantage makes GANs particularly suitable for handling imbalanced datasets, effectively supplementing minority class samples and thus mitigating issues arising from class imbalance. However, it is important to note that while data augmentation methods based on GANs are powerful, their performance heavily relies on the design of the GANs model. Every aspect, from the choice of network architecture and the definition of loss functions to the formulation of training strategies, can profoundly impact the quality of the final generated samples. Furthermore, due to the extensive matrix operations and backpropagation involved in the training process of GANs, the requirements for hardware performance are relatively high. Therefore, there is still some room for improvement.

4. Applications of data augmentation methods in some DL domains

Image data augmentation methods can enhance the performance of models in practical applications. In this review, this review exemplify the applications of data augmentation methods in image classification, object detection, and semantic segmentation.

4.1. Application in image classification

In general, the data augmentation methods mentioned in Section 2 are widely favored due to their intuitiveness and ease of operation. These methods do not rely on specific generative models, allowing researchers to directly generate augmented samples based on existing datasets. As a result, they have found widespread application and promotion in various scenarios. Furthermore, researchers typically evaluate the performance of newly proposed image data augmentation methods on three popular image classification datasets: CIFAR-10 [68], CIFAR-100 [68], and ImageNet [69]. As shown in Table 1, the basic information for these datasets is as follows: The CIFAR-10 dataset comprises 10 categories with a total of 50,000 training samples, distributed evenly with 5000 images per category, and 1000 images per category in the test set. The CIFAR-100 dataset contains 100 categories with a total of 50,000 training samples, evenly distributed with 500 images per category, and 100 images per category in the test set. The ImageNet dataset has 1000 categories, with approximately 1.3 million training samples, averaging 1300 images per category, and 50 images per category in the test set.

To better assess the performance improvement of data augmentation methods compared to baseline methods, Tables 2, 3, and 4 present the performance of various data augmentation techniques on these three datasets. Notably, “Top-1 Error” refers to the proportion of instances where the model’s highest predicted probability category does not match the actual label when predicting a single image. “Top-5 Error” represents the proportion of cases where the actual label is not among the model’s top five predicted categories. In other words, if a model generates five most likely category predictions for an image but the true label is not included in these predictions, it counts as one Top-5 Error. This metric effectively evaluates the model’s ability to include the correct answer among multiple candidate responses.

Analyzing these tables reveals several key findings: First, most methods demonstrate some degree of performance improvement compared to the baseline model. Second, the effectiveness of different data augmentation techniques in enhancing baseline model performance varies. Specifically, in the CIFAR-100 dataset, using the CutMix method with the PyramidNet200 [70] network achieves a significant performance gain of 2.64% compared to the baseline. Conversely, the Cutout method does not improve performance and actually leads to a decrease in performance compared to the baseline. Finally, these findings underscore the need for continuous exploration of effective data augmentation techniques to further enhance model performance.

Mikolajczyk et al. [71] proposed an image data augmentation method based on style transfer, which generates new samples that not only retain the pixel content of the original images, but also exhibit various styles. They conducted extensive image classification experiments on three medical datasets, and the results verified the effectiveness of this method. Bird et al. [75] incorporated transfer learning techniques and image data augmentation based on generative models into the training process to improve the classification accuracy of lemon images. By adding data augmentation techniques, the classification accuracy of lemon images increased significantly from 83.77% to 88.75%. Gao et al. [76]

Table 1. A classification dataset partially used for enhancing the performance of image data enhancement strategies. Among them, the image pixels in the CIFAR-10 and CIFAR-100 datasets are lower and the number of images is relatively small. The ImageNet dataset has a larger number of images and higher pixels.

Dataset	Number of categories	Number of train set	Number of test set
CIFAR-10	10	50000	10000
CIFAR-100	100	50000	10000
ImageNet	1000	Approximately 1.3 million	50000

Table 2. Performance demonstration of some classic and advanced data augmentation methods in the CIFAR-100 dataset. From the data in the table, it can be seen that most data augmentation methods can improve the benchmark model, among which CutMix, a method based on multigraph simultaneous augmentation, can make significant improvements compared to the benchmark model. Data from [40] and [41].

Method	Top-1 Error (%)	Top-5 Error (%)
PyramidNet200	16.45	3.69
PyramidNet200 + Label smoothing [72]	16.73	3.37
PyramidNet200 + Cutout	16.53	3.65
PyramidNet200 + Cutout + Label smoothing	15.61	3.88
PyramidNet200 + DropBlock [73]	15.73	3.26
PyramidNet200 + Mixup	15.63	3.99
PyramidNet200 + Manifold Mixup	16.14	4.07
PyramidNet200 + CutMix	14.47	2.97
PyramidNet200 + StyleCutMix	14.61	2.95

Table 3. Performance demonstration of some classic and advanced data augmentation methods in the ImageNet dataset. The data in the table also indicates that most advanced methods can make improvements to the benchmark model. The degree of improvement in model performance varies with different methods. Data from [40].

Method	Top-1 Error (%)	Top-5 Error (%)
ResNet50	23.68	7.05
ResNet50 + StochDepth [74]	22.46	6.27
ResNet50 + Cutout	22.93	6.66
ResNet50 + DropBlock	21.87	5.98
ResNet50 + Mixup	22.58	6.40
ResNet50 + Manifold Mixup	22.50	6.21
ResNet50 + CutMix	21.40	5.92

Table 4. Performance demonstration of some classic and advanced data augmentation methods in the CIFAR-10 dataset. Due to the small categories of the CIFAR-10 dataset and the large number of samples in each category, most methods can recognize images with high accuracy. However, most methods can still improve the performance of the model. Data from [41].

Method	Top-1 Error (%)
PyraMidNet200	3.85
PyramidNet200 + Cutout	3.10
PyramidNet200 + Mixup	3.09
PyramidNet200 + Manifold Mixup	3.15
PyramidNet200 + CutMix	2.88
PyramidNet200 + StyleMix	3.56
PyramidNet200 + StyleCutMix	2.79
PyramidNet200 + StyleCutMix (auto)	2.55

enhanced image diversity by incorporating data augmentation methods and used the augmented samples to improve model performance in small-sample hyperspectral datasets, facilitating better application of hyperspectral data in real-world scenarios. Shawky et al. [77] addressed the challenge of limited remote sensing images leading to poor generalization ability in trained models. They used data augmentation methods to increase the number of images and conducted extensive tests on three publicly available VHR (very high-resolution) remote sensing datasets. The experimental results demonstrated that incorporating data augmentation techniques can enhance model performance in image classification tasks. Abayomi-Alli et al. [78] proposed an image data augmentation method that modifies color distribution in images to improve the model's ability to recognize cassava leaf diseases, especially when the number of images in the dataset is insufficient. This method generates low-quality image datasets to simulate issues encountered by devices in real-world scenarios. The experimental results showed that data augmentation techniques significantly improve model performance. Cap et al. [79] aimed to enhance model performance in plant disease diagnosis by proposing LeafGAN, an excellent method for generating new samples to expand the training dataset. This approach improved the model's diagnostic performance by 7.4%.

In practical industrial applications, defect detection remains a critical and complex task. Training efficient and accurate defect detection models typically requires a large number of defective images as training data. However, collecting sufficient and diverse defective images is often challenging due to the rarity of defects and various factors influencing their occurrence during production. Additionally, existing simple data augmentation methods like rotation and cropping can increase the amount of data to some extent but often struggle to generate realistic and detailed defective images, thus failing to meet the demands of high-precision defect detection. To address this issue, Li et al. [80] proposed the DLS-GAN (defect location sensitive data augmentation GAN) method. This approach involves two "encoder-decoder" structures for extracting and fusing features of defective and non-defective regions. The first encoder-decoder focuses on capturing key features of defective images, including shape, size, location, and interaction with the surrounding environment. The second encoder-decoder extracts complete features of non-defective components to ensure that the generated images align with

real components in all aspects except for the defects. Finally, through feature transfer techniques, DLS-GAN skillfully fuses these two sets of features, generating images that contain both realistic defective features and a high degree of realism. This method effectively alleviates the problem of insufficient defective images in practical applications and provides richer data resources for training defect detection models. Training with these generated images significantly improves detection model performance. Similarly, Jain et al. [81] tackled the challenge of collecting defective data by proposing the use of GAN models to generate realistic samples with defects, addressing the issue of data scarcity. Furthermore, the newly generated samples can be used in image classification tasks. Multiple experimental results demonstrate that the data augmentation method using these augmented samples significantly improves the performance of convolutional neural network (CNN) models. Wang et al. [82] introduced a novel semantic data augmentation approach called “ISDA” (implicit semantic data augmentation). This method performs semantic transformations in feature space to enhance samples, and experimental results on multiple public classification datasets show that ISDA significantly improves the generalization performance of baseline models.

4.2. Application in object detection

As a popular research direction in the field of DL vision, object detection has always attracted much attention. Many researchers are dedicated to improving model architectures and algorithms to enhance model performance. However, besides these efforts, there is also a method that directly acts on the underlying data, which can significantly improve model performance, known as the data augmentation strategy. In the field of object detection, a widely popular data augmentation method is “Mosaic”. By skillfully combining and transforming image data, the Mosaic method can effectively expand the dataset, improve the generalization ability and robustness of the model, and further enhance the performance of object detection. Zoph et al. [83] conducted in-depth research on the improvement of image data augmentation techniques in object detection performance. They adopted the Auto Augment method, leveraging its advantage of automatically searching for appropriate data augmentation strategies, to optimize the detection model. Through extensive experiments on the COCO (common objects in context) dataset, especially in models based on the ResNet-50 (residual network) network, they achieved significant performance improvements, increasing the model’s mAP (mean average precision) by 2.3 points, validating the effectiveness of data augmentation techniques in the field of object detection. Tang et al. [84] proposed a method called “AutoPedestrian” to address the issues of crowding and occlusion in pedestrian detection. This method first augments existing samples using advanced data augmentation strategies and then employs an optimized loss function for training. The model trained using this method achieved excellent performance on multiple pedestrian detection datasets. Wang et al. [85] used GAN to enhance images for detecting cracks and defects on the surface of lychees, aiming to expand the data volume of minority class samples and enrich the sample diversity in the dataset. The authors tested their strategy using three object detection models, and the experimental results showed that this method could significantly improve model performance. Moreover, the data augmentation strategy could better distinguish key regions and boundaries of target objects in images.

In the field of 3D (three-dimensional) object detection, Zhang et al. [86] proposed using transformation reversal and replay to augment existing data, aiming to bridge the gap between multimodal and unimodal data augmentation and improve the model’s detection ability in multimodal

scenarios. Additionally, to avoid potential occlusion interference that could affect model performance, they introduced multimodal cutting and pasting, achieving promising results. Similarly, in the field of 3D object detection, Wang et al. [87] tackled the problem of 3D object detection in autonomous driving by proposing a method called “PointAugmenting”. This method combines a cross-modal data augmentation approach that involves pasting virtual samples into point clouds and images. Experimental results demonstrated that this method effectively enhances the model’s generalization ability and robustness. The performance of the model trained with this method improved by 6.5% in mAP compared to the baseline method. In the same autonomous driving domain, Li et al. [88] introduced the InverseAug data augmentation method, which utilizes geometric deformation strategies such as rotation to enhance input images. This strategy aligns image pixels with radar points, leading to stronger model performance. Cheng et al. [89] incorporated a progressive population-based data augmentation method into 3D point cloud object detection. This approach searches for optimal parameters specific to each sample in the search space, enhancing model performance. Experimental results showed that this method requires fewer labeled samples to achieve similar performance compared to baseline methods without data augmentation.

Data augmentation techniques have also been widely applied in YOLO series object detection. Zhu et al. [90] aimed to enhance the performance of object detection tasks in drone scenarios by proposing the TPH-YOLOv5 (transformer prediction heads YOLOv5) model. They introduced several effective modules to boost YOLOv5’s performance. To further improve model performance, they incorporated Mosaic, Mixup, and several traditional data augmentation strategies to augment the original dataset. These strategies contributed to their method achieving competitive results in the 2021 VisDrone Challenge. Sun et al. [91] applied the YOLOv5 method to defect detection on the inner wall of combustion boilers. However, due to the difficulty of collecting defect images, the initial number of images was limited. To address this issue, they used five data augmentation methods to expand the initial image set, effectively enhancing sample diversity and improving model performance. Liu et al. [92] considered that traditional image data augmentation methods might have limitations and proposed a novel adaptive framework called “IA-YOLO” (image adaptive YOLO). This framework searches for the optimal augmentation strategy for each input image. By combining this method with YOLOv3, they significantly improved the detection performance of the model in foggy and low-light conditions. Chung et al. [93] evaluated various data augmentation strategies in a vehicle detection dataset. Their experimental results showed that even simple data augmentation methods could significantly enhance the performance of the YOLOv3 detection model.

4.3. Application in semantic segmentation

Su et al. [94] aimed to achieve better results in the semantic segmentation of crop and weed images. They adopted random cropping and image stitching to enhance existing images, conducting extensive experiments on datasets from two farms. The experiments demonstrated that their strategy significantly improved model performance, particularly in the Narrabri dataset, where the average accuracy and mean intersection over union (IOU) increased by 3.01% and 7.18%, respectively. Recognizing the heavy reliance of semantic segmentation tasks on labeled data, which is often difficult to collect and annotate in real-world scenarios, Choi et al. [95] proposed a solution. They first used GANs to synthesize new samples similar to existing ones and then employed a self-ensembling strategy to enhance the performance of the segmentation network, achieving better segmentation results.

Similarly, addressing the challenge of semantic segmentation's heavy reliance on labeled datasets, which are typically time-consuming and labor-intensive to annotate, Yuan et al. [96] presented a simple semi-supervised learning framework. This method involved initially augmenting the original dataset using data enhancement techniques and then processing the enhanced samples through a batch normalization module. Liu et al. [97] tackled the issue of imbalanced semantic label distribution in semantic segmentation tasks. They introduced a data augmentation approach using GANs to increase the number of samples in minority classes. Experimental results showed that this strategy improved the overall average performance of the model by enhancing the performance of the minority classes.

Exploring large-scale video data augmentation for semantic segmentation in driving scenarios, Budvytis et al. [98] proposed an occlusion-aware and uncertainty-enabled label propagation algorithm to generate additional labeled data. This approach aimed to increase the scale and diversity of training data. They augmented two commonly used semantic datasets for driving scenes, providing numerous high-quality new samples and significantly improving the average classification accuracy and IOU of the benchmark model. Olsson et al. [99] addressed the limitations of traditional data augmentation methods in semi-supervised classification. They introduced a data augmentation technique called "ClassMix", which generates entirely new samples by mixing unlabeled ones. To achieve optimal results, this method carefully considered the boundaries of various sample objects during the augmentation process, enhancing the plausibility of the newly generated samples. Zhang et al. [100] noted that traditional data augmentation techniques primarily focus on enhancing entire images, making them image-level approaches that struggle to augment specific parts of an image. To better enhance boundaries between objects and backgrounds in images, they proposed ObjectAug. This method decouples and separates target objects from their backgrounds, allowing for independent augmentation of each object and subsequent recombination with various backgrounds. Their extensive experiments on several semantic segmentation tasks demonstrated the superiority of their approach.

4.4. *Brief summary*

In this section, the present review delves deeply into the classic applications of data augmentation methods within the domain of computer vision, encompassing image classification, object detection, and semantic segmentation. The widespread application and significance of data augmentation techniques in these tasks are demonstrated.

In the context of image classification, particularly when addressing challenges associated with insufficient datasets or imbalanced classes, a range of methods such as style transfer, generative models, and simple transformations (e.g., rotations, croppings) have been effectively employed to enhance model performance. Furthermore, data augmentation has also exhibited its practical utility in specialized domains such as medicine, agriculture, and industrial defect detection. Regarding object detection tasks, various data augmentation methods have also found widespread application. Within this field, image data augmentation techniques can be applied to tasks such as 3D object detection and pedestrian detection. Experimental results indicate that these methods significantly improve the detection performance of models. In the realm of semantic segmentation, data augmentation techniques can be utilized to augment the number of samples in a dataset, addressing issues related to sample scarcity in this task and thereby reducing reliance on labeled data.

5. The achievements, challenges, and future prospects of image data augmentation methods

Image data augmentation techniques have achieved notable accomplishments in various fields of computer vision, but they also present certain limitations and challenges. This review will delve into the future development directions of these techniques, aiming to provide a beneficial outlook for their improvement.

5.1. The achievements attained by image data augmentation methods

In recent years, data augmentation methods have achieved significant accomplishments in various fields of computer vision, which can be mainly summarized as follows:

(1) Improving baseline model performance: By utilizing data augmentation techniques to increase the number of training images, models can receive images with different pixel content, thereby enhancing their performance in various tasks such as image classification, object detection, and semantic segmentation.

(2) Enhancing model robustness and generalization ability: Essentially, data augmentation strategies involve transforming images or adding other pixel content (essentially a form of interference or noise). Models can learn more robust feature representations from these samples, enabling them to exhibit better robustness and generalization capabilities when confronted with input images containing interference or noise.

(3) Reducing annotation costs: The collection and annotation of initial datasets can be time-consuming, labor-intensive, and costly. Therefore, in certain semi-supervised learning tasks, the adoption of data augmentation techniques can effectively improve the performance of training models without heavily relying on annotation information, thereby achieving a significant reduction in annotation costs.

(4) Positive impacts on imbalanced and few-sample datasets: In the real world, many datasets suffer from long-tailed distributions (where some categories have abundant samples while most have scarce samples) and few-sample issues (where it is difficult to obtain sufficient data for model training). Excellent data augmentation methods can effectively generate high-quality samples to supplement the quantity of minority categories and enrich few-sample datasets. This approach can significantly improve the overall performance of models, enabling them to better adapt to various practical scenarios.

5.2. The challenges faced by data augmentation methods

Although image data augmentation methods have achieved significant accomplishments in various fields of computer vision, they nonetheless present some limitations and challenges, which can be roughly summarized as follows:

(1) Computational resource consumption: Generally speaking, some simple data augmentation methods (such as rotation, cropping, etc.) may struggle to produce satisfactory results in certain tasks. On the other hand, more complex data augmentation methods (e.g., those based on GANs) can generate high-quality samples to improve model performance. However, these methods often require substantial computational resources and time, posing a challenge in scenarios with limited computational capabilities.

(2) Discrepancies with real-world scenarios: Currently, most data augmentation methods are based

on geometric transformations of the image itself and color adjustments of pixels, making it difficult to simulate complex variations encountered in real-world situations. Such variations include changes in camera perspective, occlusion angles of objects, and variations in lighting conditions.

(3) Dependence on labeled data: While data augmentation methods can alleviate the issues caused by insufficient labeled data in certain scenarios (such as semi-supervised learning tasks), high-quality and abundant labeled information remains crucial for improving model performance in more complex tasks (such as fine-grained classification and semantic segmentation).

5.3. Prospects for the future

In light of the limitations and adjustments present in the aforementioned data augmentation techniques, the potential future development directions can be summarized as follows:

(1) Adaptive augmentation strategies: The effectiveness of a single data augmentation method can vary significantly across different tasks. Therefore, it is crucial to adaptively select appropriate data augmentation strategies tailored to specific tasks. Future work should focus on researching and developing adaptive augmentation strategies to achieve optimal data augmentation results across various tasks.

(2) Simulating more realistic augmentation strategies: Currently, there are still discrepancies between the images generated by data augmentation methods and those encountered in real-world scenarios. To enhance the model's generalization ability in practical settings, we need to explore augmentation strategies that better align with real-world conditions. Future efforts should leverage advanced techniques such as GANs to generate more realistic image samples. This can draw on the current achievements in image synthesis and virtual reality.

(3) Exploring lightweight augmentation methods: While some complex data augmentation methods can produce high-quality samples, they often come with high computational costs and large model parameters. In future research, we aim to investigate lightweight augmentation techniques that combine methods like model pruning and distillation to reduce computational requirements and model size, enabling more efficient data augmentation. By combining knowledge distillation and quantification, the demand for storage can be reduced.

(4) Cross-modal data augmentation: In recent years, cross-modal learning techniques have gained increasing attention. Exploring ways to combine ideas from different modalities for augmentation and generating new image samples represents a promising research direction (e.g., combining speech with images, text with images, etc.). By combining these technologies, it is expected to achieve cross-modal data augmentation, enabling the model to achieve better performance in more complex tasks.

6. Conclusions

In recent years, DL has made significant progress in various computer vision fields. However, this success has been largely achieved through the support of large-scale training datasets. Unfortunately, in practical applications, many DL tasks struggle to obtain a sufficient number of training samples for model training. To address this issue, various image data augmentation methods have been widely used. First, this review provides a detailed introduction to image data augmentation methods that directly enhance existing samples and those that generate new samples using GANs models. These methods effectively improve model performance in situations with limited samples. Second, this

review delves into the applications of some data augmentation methods in popular fields such as image classification, object detection, and semantic segmentation, as well as their impact on model performance. Finally, this review discusses the achievements, limitations, and challenges of current data augmentation methods and provides an outlook on potential future improvement directions.

Use of AI tools declaration

The author declares they have not used artificial intelligence (AI) tools in the creation of this article.

Conflict of interest

The author declares there is no conflict of interest.

References

1. P. Li, Y. Zhang, L. Yuan, H. X. Xiao, B. B. Lin, X. H. Xu, Efficient long-short temporal attention network for unsupervised video object segmentation, *Pattern Recogn.*, **146** (2024), 110078. <https://doi.org/10.1016/j.patcog.2023.110078>
2. E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, D. Van Valen, Deep learning for cellular image analysis, *Nat. Methods*, **16** (2019), 1233–1246. <https://doi.org/10.1038/s41592-019-0403-1>
3. L. Chena, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, D. Rueckert, Self-supervised learning for medical image analysis using image context restoration, *Med. Image Anal.*, **58** (2019). <https://doi.org/10.1016/j.media.2019.101539>
4. Y. A. Nanekaran, D. F. Zhang, J. D. Chen, Y. Tian, N. Al-Nabhan, Recognition of plant leaf diseases based on computer vision, *J. Ambient Intell. Human. Comput.*, (2020), 1–18. <https://doi.org/10.1007/s12652-020-02505-x>
5. M. Wankhade, A. C. S. Rao, C. Kulkarni, A survey on sentiment analysis methods, applications, and challenges, *Artif. Intell. Rev.*, **55** (2022), 5731–5780. <https://doi.org/10.1007/s10462-022-10144-1>
6. D. M. E. D. M. Hussein, A survey on sentiment analysis challenges, *J. King Saud Univ. Eng. Sci.*, **30** (2018), 330–338. <https://doi.org/10.1016/j.jksues.2016.04.002>
7. K. R. Chowdhary, Natural language processing, in *Fundamentals of Artificial Intelligence*, Springer, (2020), 603–649. https://doi.org/10.1007/978-81-322-3972-7_19
8. V. Raina, S. Krishnamurthy, Natural language processing, in *Building an Effective Data Science Practice*, Springer, (2022), 63–73. https://doi.org/10.1007/978-1-4842-7419-4_6
9. M. Malik, M. K. Malik, K. Mehmood, I. Makhdoom, Automatic speech recognition: A survey, *Multimed. Tools Appl.*, **80** (2021), 9411–9457. <https://doi.org/10.1007/s11042-020-10073-7>
10. D. Wang, X. D. Wang, S. H. Lv, An overview of end-to-end automatic speech recognition, *Symmetry*, **11** (2019), 1018. <https://doi.org/10.3390/sym11081018>
11. L. Deng, X. Li, Machine learning paradigms for speech recognition: An overview, *IEEE Trans. Audio*, **21** (2013), 1060–1089. <https://doi.org/10.1109/TASL.2013.2244083>

12. X. Tan, T. Qin, F. Soong, T. Y. Liu, A survey on neural speech syntheses, preprint, arXiv:2106.15561.
13. V. Mario, G. Angiulli, P. Crucitti, D. D. Carlo, F. Laganà, D. Pellicanò, et al., A fuzzy similarity-based approach to classify numerically simulated and experimentally detected carbon fiber-reinforced polymer plate defects, *Sensors*, **22** (2022), 4232. <https://doi.org/10.3390/s22114232>
14. M. Versaci, G. Angiulli, P. D. Barba, F. C. Morabito, Joint use of eddy current imaging and fuzzy similarities to assess the integrity of steel plates, *Open Phys.*, **18** (1) (2020), 230–240. <https://doi.org/10.1515/phys-2020-0159>
15. W. Zeng, H. L. Zhu, C. Lin, Z. Y. Xiao, A survey of generative adversarial networks and their application in text-to-image synthesis, *Elect. Res. Arch.*, **31** (2023), 7142–7181. <https://doi.org/10.3934/era.2023362>
16. I. Goodfellow, P. A. Jean, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial nets, in *2014 Advances in Neural Information Processing Systems (NIPS)*, **27** (2014), 1–9.
17. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft COCO: Common objects in context, in *2014 European conference computer vision (ECCV)*, (2014), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48
18. J. Zou, M. Huss, A. Abid, P. Mohammadi, A. Torkamani, A. Telenti, A primer on deep learning in genomics, *Nat. Genet.*, **51** (2019), 12–18. <https://doi.org/10.1038/s41588-018-0295-5>
19. A. Borji, S. Izadi, L. Itti, iLab-20M: A large-scale controlled object dataset to investigate deep learning, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 2221–2230. <https://doi.org/10.1109/CVPR.2016.244>
20. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis.*, **115** (2015), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
21. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
22. A. G. Howard, M. L. Zhu, B. Chen, D. Kalenichenko, W. J. Wang, T. Weyand, et al., MobileNets: Efficient convolutional neural networks for mobile vision applications, preprint, arXiv:1704.04861.
23. X. Y. Zhang, X. Y. Zhou, M. X. Lin, J. Sun, ShuffleNet: An extremely efficient convolutional neural network for mobile devices, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2018), 6848–6856. <https://doi.org/10.1109/CVPR.2018.00716>
24. W. Zeng, Z. Y. Xiao, Few-shot learning based on deep learning: A survey, *Math. Biosci. Eng.*, **21** (2024), 679–711. <https://doi.org/10.3934/mbe.2024029>
25. J. Yang, X. M. Wang, Z. P. Luo, Few-shot remaining useful life prediction based on meta-learning with deep sparse kernel network, *Inform. Sci.*, **653** (2024), 119795. <https://doi.org/10.1016/j.ins.2023.119795>
26. Y. Q. Wang, Q. M. Yao, J. T. Kwok, L. M. Ni, Generalizing from a few examples: A survey on few-shot learning, *ACM Comput. Surveys*, **53** (2020), 1–34. <https://doi.org/10.1145/3386252>

27. C. Shorten, T. M. Khoshgoftaar, A survey on Image Data Augmentation for Deep Learning, *J. Big Data*, **6** (2019), 60. <https://doi.org/10.1186/s40537-019-0197-0>
28. N. E. Khalifa, M. Loey, S. Mirjalili, A comprehensive survey of recent trends in deep learning for digital images augmentation, *Artif. Intell. Rev.*, **55** (2022), 2351–2377. <https://doi.org/10.1007/s10462-021-10066-4>
29. K. Alomar, H. I. Aysel, X. H. Cai, Data augmentation in classification and segmentation: A survey and new strategies, *J. Imaging*, **9** (2023), 46. <https://doi.org/10.3390/jimaging9020046>
30. T. DeVries, G. W. Taylor, Improved regularization of convolutional neural networks with cutout, preprint, arXiv:1708.04552.
31. N. H. Li, S. J. Liu, Y. Q. Liu, S. Zhao, M. Liu, Random erasing data augmentation, in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, **34** (2020), 13001–13008. <https://doi.org/10.1609/aaai.v34i07.7000>
32. K. K. Singh, Y. J. Lee, Hide-and-Seek: Forcing a network to be meticulous for weakly-supervised object and action localization, in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, (2017), 3544–3553. <https://doi.org/10.1109/ICCV.2017.381>
33. P. G. Chen, S. Liu, H. S. Zhao, X. G. Wang, J. Y. Jia, GridMask data augmentation, preprint, arXiv:2001.04086.
34. E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q. V. Le, AutoAugment: Learning augmentation policies from data, preprint, arXiv:1805.09501.
35. S. Lim, I. Kim, T. Kim, C. Kim, S. Kim, Fast autoaugment, in *2019 Advances in Neural Information Processing Systems (NIPS)*, (2019).
36. R. Hataya, J. Zdenek, K. Yoshizoe, H. Nakayama, Faster autoaugment: Learning augmentation strategies using backpropagation, in *2020 European conference computer vision (ECCV)*, (2022), 1–16. https://doi.org/10.1007/978-3-030-58595-2_1
37. E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, Faster autoaugment: Learning augmentation strategies using backpropagation, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2020), 3008–3017. <https://doi.org/10.1109/CVPRW50498.2020.00359>
38. D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, B. Lakshminarayanan, Augmix: A simple data processing method to improve robustness and uncertainty, preprint, arXiv:1912.02781.
39. K. Baek, D. Bang, H. Shim, GridMix: Strong regularization through local context mapping, *Pattern Recogn.*, **109** (2021), 107594. <https://doi.org/10.1016/j.patcog.2020.107594>
40. S. Yun, D. Han, S. Chun, S. J. Oh, S. Chun, J. Choe, et al., CutMix: Regularization strategy to train strong classifiers with localizable features, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, (2019), 6022–6031. <https://doi.org/10.1109/ICCV.2019.00612>
41. M. Hong, J. Choi, G. Kim, StyleMix: Separating content and style for enhanced data augmentation, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2021), 14857–14865. <https://doi.org/10.1109/CVPR46437.2021.01462>
42. D. Walawalkar, Z. Q. Shen, Z. C. Liu, M. Savvides, Attentive cutmix: An enhanced data augmentation approach for deep learning based image classification, preprint, arXiv:2003.13048.

43. H. Y. Zhang, M. Cisse, Y. N. Dauphin, D. Lopez-Paz, Mixup: Beyond empirical risk minimization, preprint, arXiv:1710.09412.
44. E. Harris, A. Marcu, M. Painter, M. Niranjana, A. Prügell-Bennett, J. Hare, Fmix: Enhancing mixed sample data augmentation, preprint, arXiv:2002.12047.
45. J. Qin, J. M. Fang, Q. Zhang, W. Y. Liu, X. G. Wang, X. G. Wang, Resizemix: Mixing data with preserved object information and true labels, preprint, arXiv:2012.11101.
46. A. F. M. S. Uddin, M. S Monira, W. Shin, T. C. Chung, S. H. Bae, Saliencymix: A saliency guided data augmentation strategy for better regularization, preprint, arXiv:2006.01791.
47. A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, Yolov4: Optimal speed and accuracy of object detection, preprint, arXiv:2004.10934.
48. J. H. Liu, B. X. Liu, H. Zhou, H. S. Li, Y. Liu, Tokenmix: Rethinking image mixing for data augmentation in vision transformers, in *2022 European conference computer vision (ECCV)*, (2022), 455–471. https://doi.org/10.1007/978-3-031-19809-0_26
49. M. Z. Chen, M. B. Lin, Z. H. Lin, Y. X. Zhang, F. Chao, R. R. Ji, SMMix: Self-Motivated Image Mixing for Vision Transformers, in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2023), 17214–17224. <https://doi.org/10.1109/ICCV51070.2023.01583>
50. L. F. Yang, X. Li, B. Zhao, R. J. Song, J. Yang, RecursiveMix: Mixed Learning with History, in *2020 Advances in Neural Information Processing Systems (NIPS)*, (2022), 8427–8440.
51. V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, D. Lopez-Paz, et al., Manifold mixup: Better representations by interpolating hidden states., in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, **97** (2019), 6438–6447.
52. J. H. Kim, W. Choo, H. Jeong, H. O. Song, Co-mixup: Saliency guided joint mixup with supermodular diversity, preprint, arXiv:2102.03065.
53. J. H. Kim, W. Choo, H. O. Song, Puzzle mix: Exploiting saliency and local statistics for optimal mixup, in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, **119** (2020), 5275–5285.
54. A. Dabouei, S. Soleymani, F. Taherkhani, N. M. Nasrabadi, SuperMix: Supervising the mixing data augmentation, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2021), 13789–13798. <https://doi.org/10.1109/CVPR46437.2021.01358>
55. C. Y. Gong, D. L. Wang, M. Li, V. Chandra, Q. Liu, KeepAugment: A simple information-preserving data augmentation approach, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2021), 1055–1064. <https://doi.org/10.1109/CVPR46437.2021.00111>
56. M. Kang, S. Kim, GuidedMixup: An efficient mixup strategy guided by saliency maps, in *2023 Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, (2023), 1096–1104. <https://doi.org/10.1609/aaai.v37i1.25191>

57. T. Hong, Y. Wang, X. W. Sun, F. Z. Lian, Z. H. Kang, J. W. Ma, GradSalMix: Gradient saliency-based mix for image data augmentation, in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, (2023), 1799–1804. <https://doi.org/10.1109/ICME55011.2023.00309>
58. M. Mirza, S. Osindero, Conditional generative adversarial nets, preprint, arXiv:1411.1784v1.
59. A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier GANs, in *2017 Proceedings of the 34th International Conference on International Conference on Machine Learning*, PMLR, (2017), 2642–2651.
60. G. Douzas, F. Bacao, Effective data generation for imbalanced learning using conditional generative adversarial networks, *Expert Syst. Appl.*, **91**, (2018), 464–471. <https://doi.org/10.1016/j.eswa.2017.09.030>
61. A. Antoniou, A. Storkey, H. Edwards, Data augmentation generative adversarial networks, preprint, arXiv:1711.04340.
62. G. Mariani, F. Scheidegger, R. Istrate, C. Bekas, C. Malossi, Bagan: Data augmentation with balancing gan, preprint, arXiv:1803.09655.
63. S. W. Huang, C. T. Lin, S. P. Chen, Y. Y. Wu, P. H. Hsu, S. H. Lai, Auggan: Cross domain adaptation with gan-based data augmentation, in *2018 Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 731–744. https://doi.org/10.1007/978-3-030-01240-3_44
64. X. Y. Zhu, Y. F. Liu, J. H. Li, T. Wan, Z. H. Qin, Emotion classification with data augmentation using generative adversarial networks, in *2018 Advances in Knowledge Discovery and Data Mining (PAKDD)*, **10939** (2018), 349–360. https://doi.org/10.1007/978-3-319-93040-4_28
65. E. Schwartz, L. Karlinsky, J. Shtok, S. Harary, M. Marder, A. Kumar, et al., Delta-encoder: An effective sample synthesis method for few-shot object recognition, in *2018 Advances in Neural Information Processing Systems (NIPS)*, **31** (2018).
66. A. Ali-Gombe, E. Elyan, MFC-GAN: Class-imbalanced dataset classification using multiple fake class generative adversarial network, *Neurocomputing*, **361** (2019), 212–221. <https://doi.org/10.1016/j.neucom.2019.06.043>
67. H. Yang, Y. Zhou, Ida-gan: A novel imbalanced data augmentation gan, in *2020 International Conference on Pattern Recognition (ICPR)*, IEEE, (2020), 8299–8305. <https://doi.org/10.1109/ICPR48806.2021.9411996>
68. A. Krizhevsky, *Learning Multiple Layers of Features from Tiny Images*, Master’s thesis, University of Toronto, 2009.
69. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, ImageNet: A large-scale hierarchical image database, in *2009 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2009), 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
70. D. Han, J. Kim, J. Kim, Deep pyramidal residual networks, in *2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2017), 6307–6315. <https://doi.org/10.1109/CVPR.2017.668>

71. A. Mikołajczyk, M. Grochowski, Data augmentation for improving deep learning in image classification problem, in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, IEEE, (2018), 117–122. <https://doi.org/10.1109/IIPHDW.2018.8388338>
72. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2016), 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
73. G. Ghiasi, T. Y. Lin, Q. V. Le, Dropblock: A regularization method for convolutional networks, in *2018 Advances in Neural Information Processing Systems (NIPS)*, **31** (2018).
74. G. Huang, Y. Sun, Z. Liu, D. Sedra, K. Q. Weinberger, Deep networks with stochastic depth, in *2016 European Conference Computer Vision (ECCV)*, (2016), 646–661. https://doi.org/10.1007/978-3-319-46493-0_39
75. J. J. Bird, C. M. Barnes, L. J. Manso, A. Ekárt, D. R. Faria, Fruit quality and defect image classification with conditional GAN data augmentation, *Sci. Hortic.*, **293** (2022), 110684. <https://doi.org/10.1016/j.scienta.2021.110684>
76. H. M. Gao, J. P. Zhang, X. Y. Cao, Z. H. Chen, Y. Y. Zhang, C. M. Li, Dynamic data augmentation method for hyperspectral image classification based on Siamese structure, *J. Sel. Top. Appl. Earth Observ. Remote Sens.*, **14** (2021), 8063–8076. <https://doi.org/10.1109/JSTARS.2021.3102610>
77. O. A. Shawky, A. Hagag, E. S. A. El-Dahshan, M. A. Ismail, Remote sensing image scene classification using CNN-MLP with data augmentation, *Optik*, **221** (2020), 165356. <https://doi.org/10.1016/j.ijleo.2020.165356>
78. O. O. Abayomi-Alli, R. Damaševičius, S. Misra, R. Maskeliūnas, Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning, *Expert Syst.*, **38** (2021), e12746. <https://doi.org/10.1111/exsy.12746>
79. Q. H. Cap, H. Uga, S. Kagiwada, H. Iyatomi, Leafgan: An effective data augmentation method for practical plant disease diagnosis, *IEEE Trans. Autom. Sci. Eng.*, **19** (2022), 1258–1267. <https://doi.org/10.1109/TASE.2020.3041499>
80. W. Li, C. C. Gu, J. L. Chen, C. Ma, X. W. Zhang, B. Chen, et al., DLS-GAN: Generative adversarial nets for defect location sensitive data augmentation, *IEEE Trans. Autom. Sci. Eng.*, (2023), 1–17. <https://doi.org/10.1109/TASE.2023.3309629>
81. S. Jain, G. Seth, A. Paruthi, U. Soni, G. Kumar, Synthetic data augmentation for surface defect detection and classification using deep learning, *J. Intell. Manuf.*, **33** (2022), 1007–1020. <https://doi.org/10.1007/s10845-020-01710-x>
82. Y. L. Wang, G. Huang, S. J. Song, X. R. Pan, Y. T. Xia, C. Wu, Regularizing deep networks with semantic data augmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2021), 3733–3748. <https://doi.org/10.1109/TPAMI.2021.3052951>
83. B. Zoph, E. D. Cubuk, G. Ghiasi, T. Y. Lin, J. Shlens, Q. V. Le, Learning data augmentation strategies for object detection, in *2020 Proceedings of the European Conference on Computer Vision (ECCV)*, **12372** (2020), 566–583. https://doi.org/10.1007/978-3-030-58583-9_34

84. Y. Tang, B. P. Li, M. Liu, B. Y. Chen, Y. N. Wang, W. L. Ouyang, Autopedestrian: An automatic data augmentation and loss function search scheme for pedestrian detection, *IEEE Trans. Image Process.*, **30** (2021), 8483–8496. <https://doi.org/10.1109/TIP.2021.3115672>
85. C. L. Wang, Z. F. Xiao, Lychee surface defect detection based on deep convolutional neural networks with gan-based data augmentation, *Agronomy*, **11** (2021), 1500. <https://doi.org/10.3390/agronomy11081500>
86. W. W. Zhang, Z. Wang, C. C. Loy, Exploring data augmentation for multi-modality 3D object detection, preprint, arXiv:2012.12741.
87. C. W. Wang, C. Ma, M. Zhu, X. K. Yang, Pointaugmenting: Cross-modal augmentation for 3D object detection, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 11789–11798. <https://doi.org/10.1109/CVPR46437.2021.01162>
88. Y. W. Li, A. W. Yu, T. J. Meng, B. Caine, J. Ngiam, D. Y. Peng, et al., Deepfusion: Lidar-camera deep fusion for multi-modal 3d object detection, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), 17161–17170. <https://doi.org/10.1109/CVPR52688.2022.01667>
89. S. Y. Cheng, Z. Q. Leng, E. D. Cubuk, B. Zoph, C. Y. Bai, J. Ngiam, et al., Improving 3d object detection through progressive population based augmentation, in *2020 Proceedings of the European Conference on Computer Vision (ECCV)*, **12366** (2020), 279–294. https://doi.org/10.1007/978-3-030-58589-1_17
90. X. K. Zhu, S. C. Lyu, X. Wang, Q. Zhao, TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios, in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, (2021), 2778–2788. <https://doi.org/10.1109/ICCVW54120.2021.00312>
91. X. M. Sun, X. C. Jia, Y. Q. Liang, M. G. Wang, X. B. Chi, A defect detection method for a boiler inner wall based on an improved YOLO-v5 network and data augmentation technologies, *IEEE Access*, **10** (2022), 93845–93853. <https://doi.org/10.1109/ACCESS.2022.3204683>
92. W. Y. Liu, G. F. Ren, R. S. Yu, S. Guo, J. K. Zhu, L. Zhang, Image-adaptive YOLO for object detection in adverse weather conditions, in *2022 Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, **36** (2022), 1792–1800. <https://doi.org/10.1609/aaai.v36i2.20072>
93. Q. M. Chung, T. D. Le, T. V. Dang, N. D. Vo, T. V. Nguyen, K. Nguyen, Data augmentation analysis in vehicle detection from aerial videos, in *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, (2022), 1–3. <https://doi.org/10.1109/RIVF48685.2020.9140740>
94. D. Su, H. Kong, Y. L. Qiao, S. Sukkarieh, Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics, *Comput. Electron. Agric.*, **190** (2021), 106418. <https://doi.org/10.1016/j.compag.2021.106418>
95. J. Choi, T. Kim, C. Kim, Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, (2019), 6829–6839. <https://doi.org/10.1109/ICCV.2019.00693>

96. J. L. Yuan, Y. F. Liu, C. H. Shen, Z. B. Wang, H. Li, A simple baseline for semi-supervised semantic segmentation with strong data augmentation, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, (2021), 8209–8218. <https://doi.org/10.1109/ICCV48922.2021.00812>
97. S. T. Liu, J. Q. Zhang, Y. X. Chen, Y. F. Liu, Z. C. Qin, T. Wan, Pixel level data augmentation for semantic image segmentation using generative adversarial networks, in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, (2019), 1902–1906. <https://doi.org/10.1109/ICASSP.2019.8683590>
98. I. Budvytis, P. Sauer, T. Roddick, K. Breen, R. Cipolla, Large scale labelled video data augmentation for semantic segmentation in driving scenarios, in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, IEEE, (2017), 230–237. <https://doi.org/10.1109/ICCVW.2017.36>
99. V. Olsson, W. Tranheden, J. Pinto, L. Svensson, Classmix: Segmentation-based data augmentation for semi-supervised learning, in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, (2021), 1368–1377. <https://doi.org/10.1109/WACV48630.2021.00141>
100. J. W. Zhang, Y. C. Zhang, X. W. Xu, Objectaug: Object-level data augmentation for semantic image segmentation, in *2021 International Joint Conference on Neural Networks (IJCNN)*, IEEE, (2021), 1–8. <https://doi.org/10.1109/IJCNN52387.2021.9534020>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).