Pattern Recognition 137 (2023) 109347

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

A Comprehensive Survey of Image Augmentation Techniques for Deep Learning

Mingle Xu[a], Sook Yoon[b,*], Alvaro Fuentes[c], Dong Sun Park[c,*]

[a] Department of Electronics Engineering, Jeonbuk National University, Jeonbuk 54896, South Korea
[b] Department of Computer Engineering, Mokpo National University, Jeonnam 58554, South Korea
[c] Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonbuk 54896, South Korea

ARTICLE INFO

Article history:
Received 28 April 2022
Revised 22 November 2022
Accepted 15 January 2023
Available online 18 January 2023

Keywords:
Image augmentation
Deep learning
Image variation
Vicinity distribution
Data augmentation
Computer vision

ABSTRACT

Although deep learning has achieved satisfactory performance in computer vision, a large volume of images is required. However, collecting images is often expensive and challenging. Many image augmentation algorithms have been proposed to alleviate this issue. Understanding existing algorithms is, therefore, essential for finding suitable and developing novel methods for a given task. In this study, we perform a comprehensive survey of image augmentation for deep learning using a novel informative taxonomy. To examine the basic objective of image augmentation, we introduce challenges in computer vision tasks and vicinity distribution. The algorithms are then classified among three categories: model-free, model-based, and optimizing policy-based. The model-free category employs the methods from image processing, whereas the model-based approach leverages image generation models to synthesize images. In contrast, the optimizing policy-based approach aims to find an optimal combination of operations. Based on this analysis, we believe that our survey enhances the understanding necessary for choosing suitable methods and designing novel algorithms.

© 2023 The Author(s). Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Over the recent years, deep learning has achieved significant improvements in computer vision based on three key elements, efficient computing devices, powerful algorithms, and large volumes of images. A main work over the last decade was designing a powerful model with numerous trainable parameters[1]. The training of such a model requires a large volume of images to achieve competitive performance. However, collecting images is frequently an expensive and challenging process. Obtaining satisfactory performance with a limited dataset is particularly challenging in practical applications, such as medical [1] and agricultural images [2].

To address this issue, image augmentation has been confirmed to be an effective and efficient strategy [3,4]. As listed in Table 1, many image augmentation methods have been utilized for image classification and object detection. Understanding existing image augmentation methods is, therefore, crucial in deploying suitable algorithms. Although similar surveys have been conducted previously [5–7], our study is characterized by several essential differences. First, we do not confine ourselves to a specific type of image, such as facial images [8]. Likewise, we consider many types of image augmentation algorithms, including generative adversarial networks [9] and image mixing [10]. Third, we do not focus on a specific application, such as object detection [5]. Conversely, we consider image classification and object detection as two primary applications, along with other image and video applications such as segmentation and tracking. Finally, unlike two related studies [6,7], our survey encompasses more recent yet effective image augmentation algorithms such as instance level multiple image mixing, as well as a comprehensive analysis of model-based methods. Consequently, this paper encompasses a wider range of algorithms that yield a novel informative taxonomy.

Specifically, we first explain why different image augmentation algorithms have been designed and leveraged across diverse applications. More specifically, challenges in computer vision and vicinity distribution are introduced to illustrate the necessity of image augmentation. By augmenting image data, the aforementioned challenges can be mitigated, and the vicinity distribution space can be dilated, thereby improving the trained model's generalizability. Based on this analysis, we argue that novel image augmentation

* Corresponding authors.
[1] https://spectrum.ieee.org/andrew-ng-data-centric-ai

https://doi.org/10.1016/j.patcog.2023.109347
0031-3203/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

Table 1
Image augmentation algorithms used studies pertaining to image classification (up) and object detection (bottom).

Paper	Image augmentation method
AlexNet [11]	Translate, Flip, Intensity Changing
ResNet [12]	Crop, Flip
DenseNet [13]	Flip, Crop, Translate
MobileNet [14]	Crop, Elastic distortion
NasNet [15]	Cutout, Crop, Flip
ResNeSt [16]	AutoAugment, Mixup, Crop
DeiT [17]	AutoAugmentat, RandAugment, Random Erasing, Mixup, CutMix
Swin Transformer [18]	RandAugment, Mixup, CutMix, Random Erasing
Faster R-CNN [19]	Flip
YOLO [20]	Scale, Translate, Color space
SSD [21]	Crop, Resize, Flip, Color Space, Distortion
YOLOv4 [22]	Mosaic, Distortion, Scale, Color space, Crop, Flip, Rotate, Random erase, Cutout, Hide-and-Seek, GridMask, Mixup, CutMix, StyleGAN

Fig. 1. Example of image variations from Class CS231n.

methods are promising when new challenges are recognized. Simultaneously, once a challenge is observed in an application, it can be mitigated using an appropriate augmentation method.

In summary, our study makes the following contributions.

- We examine *challenges* and *vicinity distribution* to demonstrate the necessity of image augmentation for deep learning.
- We present a comprehensive survey on image augmentation with a novel *informative taxonomy* that encompasses a wider range of algorithms.
- We discuss the current situation and future direction for image augmentation, along with three relevant topics: understanding image augmentation, new strategy to leverage image augmentation, and feature augmentation.

The remainder of this paper is organized as follows. The second section introduces the research taxonomy. We then present two basic inspirations of image augmentation in the third section: the challenges of computer vision tasks and the vicinity distribution. Model-free image augmentation is covered in the fourth section, whereas the model-based methods are discussed in the fifth section. The process of determining an optimal image augmentation is introduced in the sixth section, followed by a discussion section. Concluding remarks are presented in the final section.

2. Taxonomy

As shown in Table 2, we classify the image augmentation algorithms among three main categories. A model-free approach does not utilize a pre-trained model to perform image augmentation, and may use single or multiple images. Conversely, model-based algorithms require the image augmentation algorithms to generate images using trained models. The augmentation process may be unconditional, label-conditional, or image-conditional. Finally, optimizing policy-based algorithms determine the optimal operations with suitable parameters from a large parameter space. These algorithms can further be sub-categorized into reinforcement learning-based and adversarial learning-based methods. The former leverages a massive search space consisting of diverse operations and their magnitudes, along with an agent to find the optimal policy within the search space. In contrast, adversarial learning-based methods locate algorithms with the corresponding magnitude to allow the task model to have a sufficiently large loss.

3. Motivation to perform image augmentation

3.1. Challenges

Table 3 describes the four types of challenges faced in computer vision tasks. The first challenge is *image variation*, resulting from effects such as illumination and deformation. Fig. 1 illustrates im-

Table 2
Taxonomy with relevant methods.

Categories			Relevant methods
Model-free	Single-image	Geometrical transformation	translation, rotation, flip, scale, elastic distortion.
		Color image processing	jittering.
		Intensity transformation	blurring and adding noise, Hide-and-Seek [23], Cutout [24], Random Erasing [25], GridMask [26].
	Multiple-image	Non-instance-level	SamplePairing [27], Mixup [28], BC Learning [29], CutMix [30], Mosaic [22], AugMix [31], PuzzleMix [32], Co-Mixup [33], SuperMix [34], GridMix [35].
		Instance-level	CutPas [36], Scale and Blend [37], Context DA [38], Simple CutPas [39], Continuous CutPas [40].
Model-based	Unconditional		DCGAN [41], [42–44]
	Label-conditional		BDA [45], ImbCGAN [46], BAGAN [47], DAGAN [48], MFC-GAN [49], IDA-GAN [50].
	Image-conditional	Label-preserving	S+U Learning [51], AugGAN [52], Plant-CGAN [53], StyleAug [54], Shape bias [55].
		Label-changing	EmoGAN [56], δ-encoder [57], Debiased NN [58], StyleMix [59], GAN-MBD [60], SCIT [2].
Optimizing policy-based	Reinforcement learning-based		AutoAugment [61], Fast AA [62], PBA [63], Faster AA [64], RandAugment [65], MADAO [66], LDA [67], LSSP [68].
	Adversarial learning-based		ADA [69], CDST-DA [70], AdaTransform [71], Adversarial AA [72], IF-DA [73], SPA [74].

Table 3
Challenges in computer vision tasks from the perspectives of datasets and deep learning models.

Challenges	Descriptions	Strategies and related studies
Images variations	The following basic variations exist in many datasets and applications, including illumination, deformation, occlusion, background, viewpoint, and multiscale, as shown in Fig. 1.	Geometrical transformation and color image processing improve the majority of the variations. Occlusion: Hide-and-Seek [23], Cutout [24], Random Erasing [25], GridMask [26]. Background or context: CutMix [30], Mosaic [22], CutPas [36]. Multiscale: Scale and Blend [37], Simple CutPas [39].
Class imbalance and few images	Number of images vary between classes or some classes have only few images.	Reusing instance from minority class is one strategy by instance-level operation, Simple Copy-Paste [39]. Most studies attempt to generate images for the minority class: ImbCGAN [46], DAGAN [48], MFC-GAN [49], EmoGAN [56], δ-encoder [57], GAN-MBD [60], SCIT [2].
Domain shift	Training and testing datasets represent different domains, commonly referring to styles.	Changing styles for existing images is a main strategy, including S+U Learning [51], StyleAug [54], Shape bias [55], Debiased NN [58], StyleMix [59].
Data remembering	Larger models with many learnable parameters tend to remember specific data points, which may result in overfitting.	The mechanism is increasing dataset size within or between vicinity distributions. Within version assumes label-preserving while between version changes labels, such as Mixup [28], AugMix [31], Co-Mixup [33].

age variations[2]. *Class imbalance* is another challenge, wherein different objects are observed with different frequencies. In medical imaging, abnormal cases often occur with a low probability, which is further exacerbated by privacy. When trained with an imbalanced dataset, a model assigns a higher probability to the normal case. Besides, class imbalance becomes few images from multiple classes to one class. Furthermore, *domain shift* represents a challenge where the training and testing datasets exhibit different distributions. This is exemplified by the night and day domains in the context of automatic driving. Because it is more convenient to collect images during the daytime, we may desire to train our model with a daytime dataset but evaluate it at nighttime.

A new challenge introduced by deep learning is *data remembering*. In general, a larger set of learnable parameters requires more data for training, which is referred to as structural risk [75]. With an increase in parameters, a deep learning model may remember specific data points with an insufficient number of training images, which introduces a generalizability problem in the form of overfitting [76].

Fortunately, image augmentation methods can mitigate these challenges and improve model generalizability by increasing the number and variance of images in the training dataset. To utilize an image augmentation algorithm efficiently, it is crucial to understand the challenges of application and apply suitable methods. This study was conducted to provide a survey that enhances the understanding of a wide range of image augmentation algorithms.

3.2. Vicinity distribution

In a supervised learning paradigm, we expect to find a function $f \in \mathcal{F}$ that reflects the relationship between an input x and target y in a joint distribution $P(x, y)$. To learn f, a loss l is defined to reduce the discrepancy between the prediction $f(x)$ and actual target y for all examples in $P(x, y)$. We can then optimize f by minimizing l over $P(x, y)$, which is known as the expected risk [75] and can be formulated as $R(f) = \int l(f(x), y)dP(x, y)$. However, $P(x, y)$ is unknown in most applications [77]. Alternatively, we may use the empirical distribution $P_e(x, y)$ to approximate $P(x, y)$. In this case, the observed dataset $\mathcal{D} = (x_i, y_i)_{i=1}^n$ is considered to be the empirical distribution, where (x_i, y_i) is in $P_e(x, y)$ for a given i:

$$P_e(x, y) = \frac{1}{n}\sum_{i=1}^{n} \delta((x = x_i, y = y_i)], \tag{1}$$

where $\delta(x, y)$ is a Dirac mass function centered at point (x_i, y_i), assuming that all masses in the probability distribution cluster around a single point [78]. Another natural notion for approximating $P(x, y)$ is the vicinity distribution $P_v(x, y)$, which replaces the Dirac mass function with an estimate of the density in the vicinity of point (x_i, y_i) [79]:

$$P_v(x, y) = \frac{1}{n}\sum_{i=1}^{n} \delta_v(x = x_i, y = y_i), \tag{2}$$

where δ_v is the vicinity point set of (x_i, y_i) in $\mathcal{D}$. The vicinity distribution assumes that $P(x, y)$ is smooth around any point (x_i, y_i) [77]. In $P_v(x, y)$, models are less prone to memorizing all data points, and thus tend to yield higher performance in the testing process. One way to achieve vicinity distribution is to apply image augmentation, by which an original data point (x_i, y_i) can be moved within its vicinity. For example, the Gaussian vicinity distribution is equivalent to the addition of Gaussian noise to an image [79].

4. Model-free image augmentation

Image processing methods, such as geometric transformation and pixel-level manipulation, can be leveraged for augmentation purposes [6,7]. In this study, we refer to model-free image augmentation as contrasting model-based image augmentation. The model-free approach consists of single- and multi-image branches. As suggested by the names, the former produces augmented images from a single image, whereas the latter generates output from multiple images.

4.1. Single-image augmentation

From the vicinity distribution, single-image augmentation (SiA) aims to fluctuate the data points in the training dataset and increase distribution density. In general, SiA leverages traditional image processing, which is simple to understand and execute. SiA methods include geometric transformations, color image processing, and intensity transformations. Geometric transformation tries to modify the spatial relationship between pixels [80], including affine transformation and elastic deformation, while color image processing aims to vary the color of an input image. In contrast, the last one is advocated to change parts of the images and has recently received more attention.

4.1.1. Geometric transformation

Objects in naturally captured images can appear in many variations. Geometric transformations can be employed to increase this variability. For instance, *translation* provides a way to augment objects' position. Furthermore, an image can be *rotated*, changing the perspectives of objects. The angle of rotation should be carefully considered to ensure the preservation of appropriate labels. Likewise, a *flip* can be executed horizontally or vertically, according to the characteristics of the training and testing datasets. For instance,

[2] http://cs231n.stanford.edu/

Table 4
Studies focusing upon intensity transformations. Each study is highlighted with its corresponding figure if available.

Paper	Year	Highlight
Hide-and-Seek [23]	2017	Split an image into patches that are randomly blocked.
Cutout [24]	2017	Apply a fixed-size mask to a random location for each image.
Random Erasing [25]	2020	Randomly select a rectangular region and displace its pixels with random values. Fig. 3.
GridMask [26]	2020	Apply multiscale grid masks to an image to mimic occlusions. Fig. 4.

the Cityscapes [81] dataset can be augmented horizontally but not vertically. In addition, objects can be magnified or shrunk via *scaling* to mimic multiscale variation. Finally, *elastic distortion* can alter the shape or posture of an object. Among these methods, flips have been commonly utilized throughout many studies over the last decade for various computer vision tasks, such as image classification [11–13], object detection [82,83], and image translation [84,85]. Two factors must be considered when using these methods: the magnitude of the operation to preserve label identity and variations in the dataset.

4.1.2. Color image processing

Unlike greyscale images, color images consist of three channels. Color image processing for augmentation assumes that the training and testing dataset distributions fluctuate in terms of colors, such as contrast. Although color image processing yields superior performance, it is rarely used because the color variations between the training and testing datasets are small. However, one interesting point is the use of robust features for contrast learning [86] via color image processing, which represents a case of task-agnostic learning.

4.1.3. Intensity transformation

Unlike geometric transformations and color image processing, intensity transformations entail changes at the pixel or patch levels. Random noise, such as Gaussian noise, is one of the simplest intensity transformation algorithms [75]. The classical methods leverage random noise independently at the pixel level; however, the patch level has recently exhibited significant improvement for deep learning algorithms [23–26]. Studies pertaining to intensity transformations are listed in Table 4. The underlying concept is that the changes push the model to learn robust features by avoiding trivial solutions [76].

Cutout [24] randomly masks the most significant area with a finding mechanism to mimic occlusion. However, the most important aspect is cost. Hide-and-Seek [23] directly blocks part of the image with the objective of obscuring the most significant area through many iterations of a random process, which is simple and fast. Fig. 2 shows that images are divided into $s \times s$ patches, and each patch is randomly blocked. One disadvantage is that the identical size of each patch yields the same level of occlusion. To address this issue, Random Erasing [25] has been employed with three random values: the size of the occluded area, height-to-width ratio, and top-left corner of the area. Fig. 3 demonstrates some examples of Random Erasing for three computer vision tasks. Additionally, this method can be leveraged in image- and object-aware conditions, thereby simplifying object detection.

GridMask aims to balance deleting and reservation, with the objective of blocking certain important areas of an object while preserving others to mimic real occlusion. To achieve this, GridMask uses a set of predefined masks, as opposed to a single mask [23–25]. As illustrated in Fig. 4, the generated mask is obtained from four values, denoting the width and height of every grid and the vertical and horizontal distance of the neighboring grid mask. By adjusting these four values, grid masks of different sizes and heigh-width ratios can be obtained. Under these conditions, GridMask achieves a better balance between deleting and reservation,

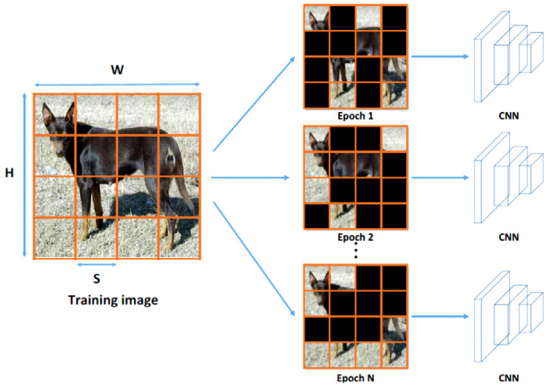

Fig. 2. Hide-and-Seek [23] carries out image augmentation where one image is split into several patches, and each patch is randomly blocked with a specified probability.

and a preliminary experiment suggests that it has a lower chance of producing failure cases than Cutout [24] and Hide-and-See [23].

4.2. Multiple-image augmentation

Multiple-image augmentation (MiA) algorithms are executed on more than one image. These methods can further be categorized as instance- and non-instance-level. Because one image may include more than one instance, we can mask instances and use them independently. Unlike SiA, MiA requires algorithms to merge multiple input instances.

4.2.1. Non-instance-level

In the context of MiA algorithms, the non-instance-level approach adopts and fuses the images. Studies pertaining to this concept are listed in Table 5. One of the simplest methods is to compute the average value of each pixel. In Pairing Samples [27], two images are fused to produce an augmented image with a label from one source image. This assumption is generalized in Mixup [28], where the labels are also fused. Fig. 5 illustrates the difference between Pairing Samples and Mixup. Mathematically, $\tilde{x} = \lambda x_i + (1-\lambda)x_j$ and $\tilde{y} = \lambda y_i + (1-\lambda)y_j$, where x_i and x_j are two images, y_i and y_j are the corresponding one-hot labels, and $\tilde{x}$ and $\tilde{y}$ denote the generated image and label, respectively. By adjusting $0 \leq \lambda \leq 1$, many images with different labels can be created, thereby smoothing out the gap between the two labels in the augmented images. Although Pairing Samples and Mixup produce satisfactory results, the fused images are not reasonable for humans. Accordingly, these fused images have been declared to make sense for machines from the perspective of a waveform [29]. In addition, vicinity distribution can also be utilized to understand this situation. To be more specific, changing image variations yet maintaining the label can be regarded a deviation in the vicinity distribution space of a specific label, whereas image fusion can be considered as an interpolation between the vicinity distribution of two labels [28].

In contrast to BC Learning [29], CutMix [30] spatially merges images to obtain results that are interpretable by humans. The last

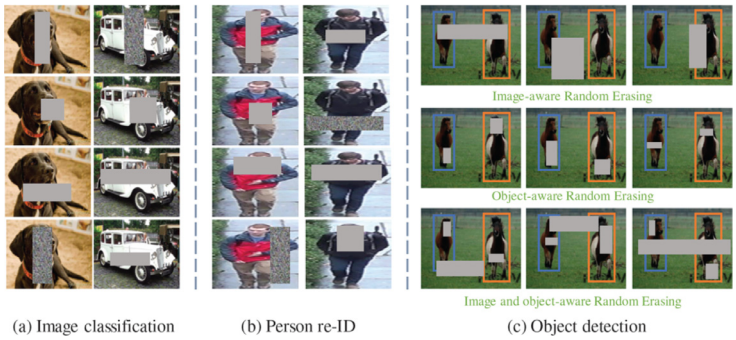

Fig. 3. Examples of Random Erasing [25].

Table 5
Studies related to multiple-image augmentation, divided into non-instance- (up) and instance-level (bottom).

Paper	Year	Highlight
SamplePairing [27]	2018	Combine two images with a single label.
Mixup [28]	2018	Linearly fuse images and their labels. Fig. 5.
BC Learning [29]	2018	Combine two images and their labels. Treat the image as a waveform, and declare that image mixing makes sense for machines.
CutMix [30]	2019	Spatially fuse two images and linearly fuse the labels. Fig. 5.
Mosaic [22]	2020	Spatially mix four images and their annotations, thereby enriching the context for each class.
AugMix [31]	2020	One image undergoes several basic augmentations, and the results are fused with the original image.
PuzzleMix [32]	2020	Optimize a mask for fusing two images to utilize the salient information and underlying statistics.
Co-Mixup [33]	2021	Maximize the salient signal of input images and diversity among the augmented images.
SuperMix [34]	2021	Optimize a mask for fusing two images to exploit the salient region with the Newton iterative method, 65x faster than gradient descent.
GridMix [35]	2021	Split two images into patches, spatially fuse the patches, and linearly merge the annotation.
Cut, Paste and Learn [36]	2017	Cut object instances and paste them onto random backgrounds. Fig. 6.
Scale and Blend [37]	2017	Cut and scale object instances, and blend them in meaningful locations.
Context DA [38]	2018	Combine object instances using context guidance to obtain meaningful images.
Simple Copy-Paste [39]	2021	Randomly paste object instances to images with large-scale jittering.
Continuous Copy-Paste [40]	2021	Deploy Cut, Paste and Learn to videos.

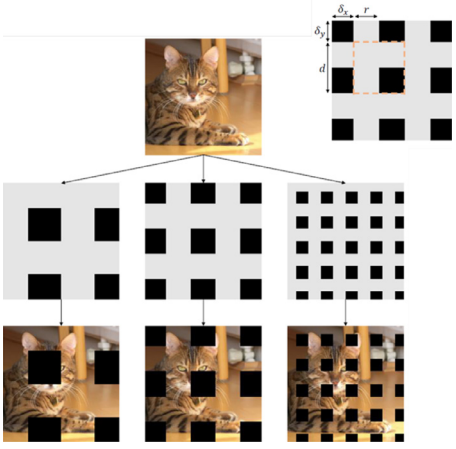

Fig. 4. GridMask [26] and its setting.

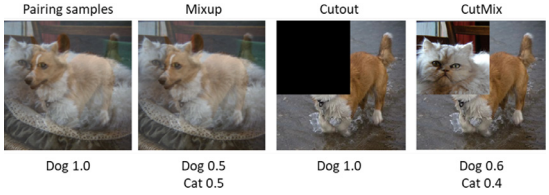

Fig. 5. Comparison of non-instance-level multiple-image algorithms [30].

picture in Fig. 5 illustrates its strategy, wherein the merged image consists of two source images spatially, and its label is obtained from the ratio of certain pixels between two images. Although multiple-image augmentation generally utilizes two images, more images can be used. For example, Mosaic [22] employs four images wherein the number of objects in one image is increased, thus significantly reducing the need for a large mini-batch size for dense prediction. AugMix [31] randomly applies basic multiple methods of image augmentation, and the results are adopted to merge with the original image.

Non-instance-level image augmentation has extensions similar to those of intensity transformations. To account for the most important area, PuzzleMix [32] discriminates the foreground from the background, and mixes important information within the foreground. Further, salient areas from multiple input images are maximized to synthesize each augmented image [33], simultaneously maximizing the diversity among the augmented images. To quickly locate dominant regions, SuperMix [34] employs a variant of the Newton iterative method. As in Hide-and-Seek [23], GridMix [35] divides images into fixed-size grids, and each patch of the output image is randomly taken from the corresponding patches of two input images. Through this analysis, we believe that GridMask [87] can be adapted to fuse image pairs with changeable sizes.

4.2.2. Instance-level

Whereas the non-instance-level approach employs images directly, the instance-level approach leverages instances masked from images. Related studies are listed in the second part of Table 5. The instance-level approach comprises two main steps. As shown in Fig. 6, the first step involves cutting instances from source images given a semantic mask, and obtaining clean background senses. Next, the obtained instances and background are merged. Cut, Paste and Learn [36] is an early instance-level method, wherein local artifacts are noticed after pasting instances to the background. Because local region-based features are important for object detection, various blending modes are employed to

Table 6
Studies relating to model-based image augmentation, label-conditional (top), label-preserving image-conditional (middle), and label-changing image-conditional (bottom).

Paper	Year	Highlight
BDA [45]	2017	Use CGAN to generate images optimized by a Monte Carlo EM algorithm. Fig. 7.
ImbCGAN [46]	2018	Deploy CGAN as image augmentation for unbalanced classes.
BAGAN [47]	2018	Train an auto-encoder to initialize generator.
DAGAN [48]	2018	Image is taken as the class condition for generator and discriminator. Fig. 8.
MFC-GAN [49]	2019	Use multiple fake classes to obtain a fine-grained image for minority class.
IDA-GAN[50]	2021	Train a variational auto-encoder and CGAN simultaneously.
S + U Learning [51]	2017	Translate synthetic images from a graphic model to realistic images using CGAN.
AugGAN [52]	2018	Aim to semantically preserve object when changing its style.
Plant-CGAN [53]	2018	Translate semantic instance layout to real images using CGAN.
StyleAug [54]	2019	Change image style via style transfer.
Shape bias [55]	2019	Transfer image style from painted images to mitigate texture bias of CNN.
EmoGAN [56]	2018	Translate a neutral face with another emotion.
δ-encoder [57]	2018	Image is taken as a class condition to generate images for new or infrequent class.
Debiased NN [58]	2021	Merge style and content via style transfer and appropriate labels. Fig. 10.
StyleMix [59]	2021	Merge two images with style, content, and labels. Fig. 11.
GAN-MBD [60]	2021	Translate an image from one class to another while preserving semantics via multi-branch discriminator. Fig. 9.
SCIT [2]	2022	Translate healthy leaves to abnormal one while retaining its style.

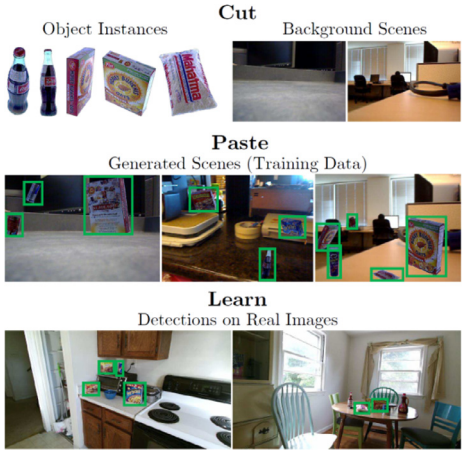

Fig. 6. Cut, Paste and Learn in training and testing process [36].

reduce local artifacts. With the exception of boundaries, the instance scale and position are not trivial, as objects may be multi-scale and recognizable with the help of their contexts, as addressed in [37].

Interestingly, instance-level image augmentation can mitigate the challenges posed by class imbalance. By repurposing rare instances, the number of images in the corresponding class increases. Simple Copy-Paste [39] indicates that the instance level enables strong image augmentation methods, for instance, segmentation. While it is based on Copy, Paste and Learn, Simple Copy-Paste differs in two characteristics. First, the background image is randomly selected from the dataset, and subjected to random scale jittering and horizontal flipping. Second, large-scale jittering is leveraged to obtain more significant performance. The copy-paste concept has also been utilized for time-series tasks [40] such as tracking.

5. Model-based image augmentation

A model must be pre-trained in model-based image augmentation to generate augmented images. The present study classifies this process among three categories, according to the conditions to generate images: unconditional, label-conditional, and image-conditional. Table 6 provides information regarding appropriate studies.

5.1. Unconditional image generation

An image synthesis model benefits image augmentation, which enables it to produce new images. Theoretically, the distribution of generated images is similar to that in the original dataset for a generative adversarial network (GAN) model after training [88]. However, the generated images are not the same as the original images and can be considered as points located in the vicinity distribution. In DCGAN [41], two random noises or latent vectors can be interpolated to generate intermediate images, which can be regarded as fluctuations between two original data points. Generally, a generative model with noise as input is deemed an unconditional model, and the corresponding image generation process is considered unconditional image generation. If the datasets encompass a single class, as in the case of medical images with one abnormal class [42], an unconditional image generation model can be directly applied to perform augmentation. Furthermore, a specific unconditional model can be leveraged for an individual class in the presence of multiple classes [43,44].

5.2. Label-conditional image generation

Although unconditional image generation has potential, the shared information of different classes cannot be utilized. In contrast, label-conditional image generation is expected to leverage the shared information and learn variations for minority classes using majority-class data. Label-conditional image generation requires one specific label as an extra input, and the generated image should align with the label condition.

The primary issue in label-conditional image generation is the use of label conditions. CGAN [89] uses the label for a generator, whereas the authenticator does not use the label. Consequently, the generator tends to ignore label information, as the authenticator cannot provide feedback regarding the condition. ACGAN [90] introduces an auxiliary classifier in the discriminator, which encourages the generator to produce images aligned with label conditions. With a more complex classifier, BDA [45] separates the classifier from the discriminator. Fig. 7 illustrates the differences between BDA and other label-conditional algorithms. In addition, MFC-GAN [49] adopts multiple fake classes in the classification loss to stabilize the training.

One of the main applications of label-conditional image generation is the class imbalance [49,46,50]. The generative model is expected to learn useful features from the majority class, and use them to generate images for the minority classes. The generated

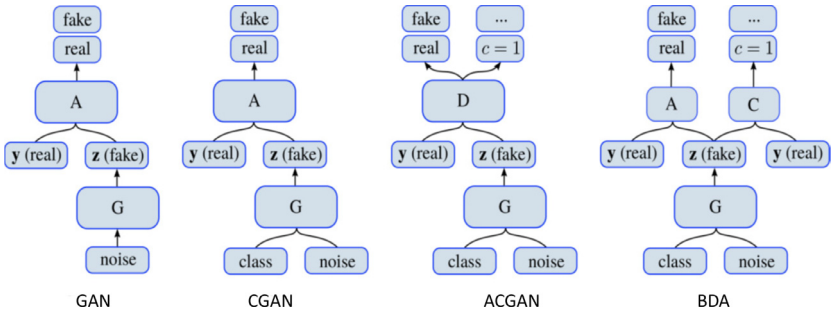

Fig. 7. GAN and variants of label-conditional GANs [45]. G: generator, A: authenticator, C: classifier, D: discriminator.

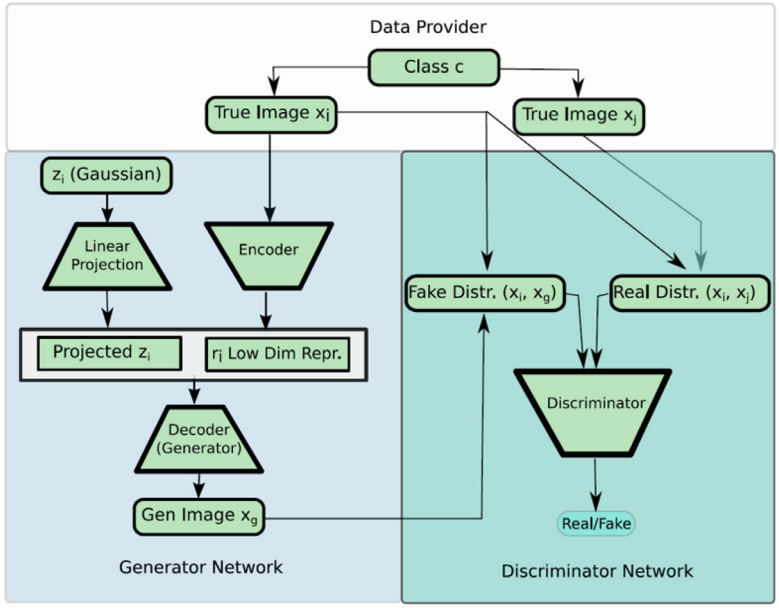

Fig. 8. Flowchart of DAGAN [48], where label information is obtained from an image via an encoder, rather than a label.

images are used to rebalance the original training dataset. However, it may be challenging to train a GAN model with an unbalanced dataset, as the majority class dominates the discriminator loss and the generator tends to produce images within the majority class. To address this challenge, a pretrained autoencoder with reconstruction loss has been employed to initialize a generator [47,50].

Although various discriminators and classifiers may be employed, the aforementioned algorithms utilize the class condition on a one-hot label. One resulting limitation is that the trained model can generate only known-class images. To overcome this limitation, DAGAN [48] utilizes an image encoder to extract the class, so that the generated image is assumed to have the same class as the original image. Fig. 8 illustrates the DAGAN algorithm.

5.3. Image-conditional image generation

In image generation, images can be employed as conditions, known as image translation. Generally, an image consists of content and style [91,92]. Content refers to class-dependent attributes, such as dogs and cats, whereas style denotes class-independent elements, such as color and illumination. Image-conditional image generation can be subcategorized into two types: label-preserving and label-changing. The former requires content to be retained, whereas the latter requires content to be changed.

5.3.1. Label-preserving image generation

Label-preserving assumes that the label of a generated image is the same as that of the input image. One active field to deploy this approach is the domain shift, where the style of the source domain is different from that of the target domain. To address this challenge, original images can be translated from the source domain to the target domain. To preserve the object during image translation, AugGAN employs a segmentation module that extracts context-aware features to share parameters with a generator [52]. For practical applications, synthetic images generated by a graphical model are translated into natural images [51], and the leaf layout is translated as a real leaf image [53]. In addition, image translation can be utilized for semantic segmentation with a domain shift [93]. Furthermore, label-preserving can be leveraged to improve the robustness of a trained model. Inspired by the observation that CNNs exhibit bias on texture toward shape, original images are translated to have different textures, which allows the CNN to allocate more attention to shape [55].

It is often challenging to obtain the desired style during the image generation process. Most algorithms utilize an encoder to extract style from an image, as in the case of DRIT++ [94] and SPADE [95]. This approach to perform image translation can be regarded as image fusion. In contrast, Jackson et al. [54] proposed style augmentation, where the style is generated from a multivariate normal distribution. Another challenge is that the *one* model

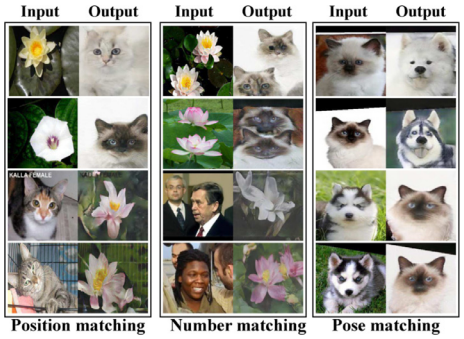

Fig. 9. Semantic level matching by GAN-MBD [60] for label-changing image augmentation, including position, number, and pose.

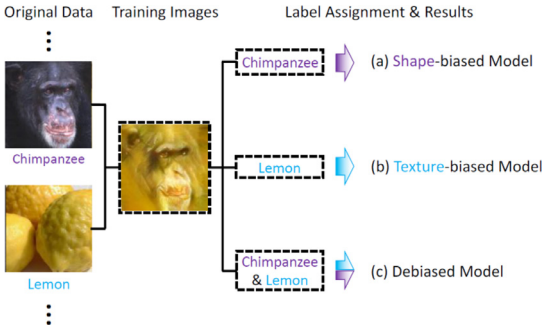

Fig. 10. Label assignment for the biased and unbiased model with respect to shape and texture [58].

can be adopted to generate images for *multiple* domains with fewer trained images. To address this, MetalGAN leverages domain loss and meta-learning strategies [96].

5.3.2. Label-changing image generation

In contrast to label-preserving, label-changing changes the label-dependent. For example, a neutral face can be transformed into a different emotion [56]. Although the generated images have poor fidelity, the approach improves the classification of emotions. In addition to changing label dependence, the preservation of label independence has recently received attention as a way to improve variability within the target class, thereby mitigating class imbalance. To take variation from one to another class, a style loss is leveraged to retain the style when translating an image [2]. Similarly, a multi-branch discriminator with fewer channels is introduced to achieve semantic consistency such as the number of objects [60]. Fig. 9 shows several satisfactory translated images. To address severe class imbalance, a δ-encoder has been proposed to extract label-independent features from one label to another [57]. As in the case of DAGAN [48], class information is provided by an image. The δ-encoder and decoder aim to reconstruct the given image in the training phase, whereas the decoder is provided a new label image and required to generate the same label in the testing phase.

Compared to label-preserving, label-changing yields more significant improvements in model robustness by changing the label and style simultaneously. As illustrated in Fig. 10, traditional image augmentation does not change the label after altering the color of the chimpanzee to that of a lemon, which incurs shape bias. By contrast, when a texture-biased model is trained, the translated image is labeled as a lemon. To balance the bias, the translated image by style transfer is taken with two labels [58] – chimpanzee and lemon – which eliminates bias. Inspired by Mixup [28], Hong et al. developed StyleMix [59], which merges the two inputs to obtain content and style labels, as shown in Fig. 11. These labels are then fused to obtain the final label for the generated images.

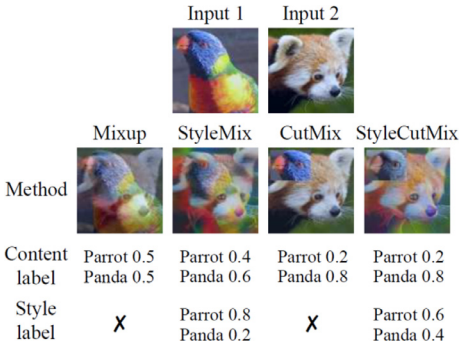

Fig. 11. Examples of label assignment with different algorithms [59].

6. Optimizing policy-based image augmentation

All algorithms mentioned in the previous two sections represent specific schemes, wherein domain knowledge is required to achieve better performance. In general, individual operations with the desired magnitude are utilized to perform image augmentation for specific datasets according to their characteristics. However, hyperparameter optimization is challenging and time-consuming. One way to mitigate this is to design algorithms that determine optimal augmentation strategies. These algorithms, termed policy-based optimization, encompass two categories: reinforcement learning-based, and adversarial learning-based. The former category employs reinforcement learning (RL) to determine the optimal strategy, whereas the latter category adopts augmented operations and their magnitudes that generates a large training loss and small validation loss. As generative adversarial networks (GANs) can be utilized for both model-based and optimizing policy-based image augmentation, the objective to adopt GANs is the primary difference. Model-based category aims to *directly* generate images, instead of other goals such as finding optimal transformations [69]. Studies pertaining to policy-based optimization are listed in Table 7.

6.1. Reinforcement learning-based

AutoAugment [61] is a seminal approach that employs reinforcement learning. As shown in Fig. 12, iterative steps are used to find the optimal policy. The controller samples a strategy from a search space with the operation type and its corresponding probability and magnitude, and a task network subsequently obtains the validation accuracy as feedback to update the controller. Because the search space is very large, lighter child networks are leveraged. After training, the controller is used to train the original task model and can be finetuned in other datasets.

Although AutoAugment achieves satisfactory classification performance across several datasets, it requires a long training time. To address this issue, several studies have been conducted from different perspectives. For instance, RandAugment [65] replaces several probabilities in AutoAugment with a uniform probability. Conversely, Fast AA [62] and Faster AA [64] leverage density matching, aligning the densities of the training and augmented training datasets, instead of Proximal Policy Optimization [97], to optimize the controller in AutoAugment. Furthermore, PBA [63] attempts to learn a policy schedule from population-based training, rather than a single policy.

Except for the long training phase, AutoAugment utilizes child models, by which the learned policy may not be optimal for the fi-

Table 7
Studies relating to optimizing policy of image augmentation. The upper and the bottom suggest reinforcement learning- and adversarial learning-based image augmentation.

Paper	Year	Highlight
AutoAugment [61]	2019	Use reinforcement learning to determine the optimal augmentation strategies. Fig. 12.
Fast AA [62]	2019	Use efficient density matching for augmentation policy search.
PBA [63]	2019	Adopt non-stationary augmentation policy schedules via population-based training.
Faster AA [64]	2019	Use a differentiable policy search pipeline via approximate gradients.
RandAugment [65]	2020	Reduce the search space of AutoAug via probability adjustment.
MADAO [66]	2020	Train task model and optimize the search space simultaneously by implicit gradient with Neumann series approximation.
LDA [67]	2020	Take policy search as a discrete optimization for object detection.
LSSP [68]	2021	Learn a sample-specific policy for sequential image augmentation.
ADA [69]	2016	Seek a small transformation that yields maximal classification loss on the transformed sample.
CDST-DA [70]	2017	Optimize a generative sequence using GAN in which the transformed image is pushed to be within the same class distribution.
AdaTransform [71]	2019	Use a competitive task to obtain augmented images with a high task loss in the training stage, and a cooperative task to obtain augmented images with a low task loss in the testing stage. Fig. 13.
Adversarial AA [72]	2020	Optimize a policy to increase task loss while allowing task model to minimize the loss.
IF-DA [73]	2020	Use influence function to predict how validation loss is affected by image augmentation, and minimize the approximated validation loss.
SPA [74]	2021	Select suitable samples to perform image augmentation.

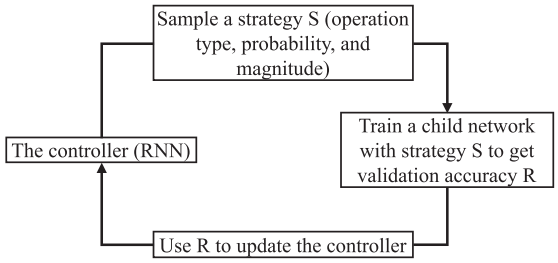

Fig. 12. Overview of AutoAugment [61], a reinforcement learning-based image augmentation method.

nal task model. To address this issue, Hataya et al. [66] trained the target model and image augmentation policy simultaneously using the same differentiable image augmentation pipeline in Faster AA. In contrast, Adversarial AA [72] leverages adversarial loss simultaneously with reinforcement learning.

One limitation of the algorithms mentioned above is that the learned image augmentation policy is at the dataset level. Conversely, class- and sample-level image augmentation methods were considered in [98] and [68], respectively, wherein each class or sample utilizes a specific policy. Furthermore, instance-level image augmentation was considered in [67] for object detection, where operations were performed only inside the bounding box.

6.2. Adversarial learning-based

The primary objective of image augmentation is to train a task model with a training dataset to achieve sufficient generalizability on a testing dataset. One assumption is that hard samples are more useful, and the input images that cause a larger training loss are considered hard samples. Adversarial learning-based image augmentation aims to learn an image augmentation policy to generate hard samples based on the original training samples.

An early method [69] attempts to find a small transformation that maximizes training loss on the augmented samples, wherein learning optimization finds an optimal magnitude given an operation. One of the main limitations is the label-preserving assumption that the augmented image retains the same label as the original image. To meet this assumption, a common strategy is to design the type of operation and range of corresponding magnitude using human knowledge. To weaken this assumption, Ratner et al. [70] introduced generative adversarial loss to learn a transformation sequence in which the discriminator pushes the generated images to one of the original classes, instead of an unseen or null class.

Interestingly, SPA [74] attempts to select suitable samples, and image augmentation is leveraged only on those samples in which the augmented image incurs a larger training loss than the original image. Although SPA trains the image augmentation policy and task model simultaneously at the sample level, the impact of the learned policy in the validation dataset is unknown. To address this challenge, an influence function was adopted for approximating the change in validation loss without actually comparing performance [73]. Another interesting concept is the use of image augmentation in the *testing stage*. To achieve this, AdaTransform [71] learns two tasks – competitive and cooperative – as illustrated in Fig. 13. In a competitive task, the transformer learns to increase the input variance by increasing the loss of the target network, while the discriminator attempts to push the augmented image realistically. Conversely, the transformer learns to decrease the variance of the augmented image in the cooperative task by reducing the loss of the target network. After training, the transformer

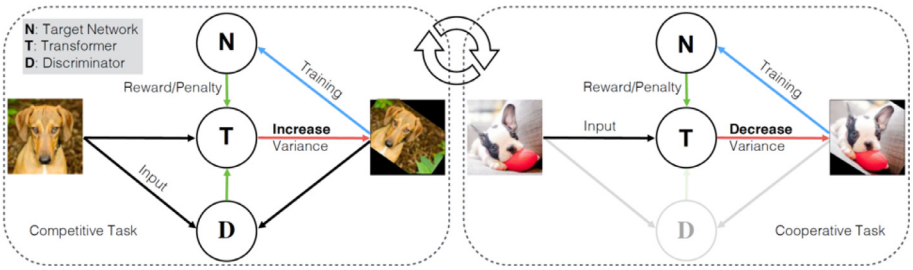

Fig. 13. Overview of AdaTransform [71]. AdaTransform encompasses two tasks – competitive training and cooperative testing – and three components: transformer T, discriminator D, and target network N. The transformer increases the variance of training data by competing with both D and N. It also cooperates with N in the testing phase to reduce data variance.

is utilized to reduce the variance of the input image, thereby simplifying the testing process.

7. Discussions

In this section, the usage of the mentioned strategies to perform image augmentation are first discussed. Several future directions are then illustrated. Furthermore, three related topics are discussed: understanding image augmentation from theory perspective, adopting image augmentation with other strategy, and augmenting features instead of images.

Current situation. Datasets are assumed to be essential to obtain satisfactory performance. One way to generate an appropriate dataset is through image augmentation algorithms, which have demonstrated impressive results across multiple datasets and heterogeneous models. For instance, Mixup [28] increases the validation accuracy in ImageNet-2012 by 1.5 and 1.2 percent with ResNet-50 and ResNet-101. Non-trivially, GAN-MBD [60] achieves 84.28 classification accuracy with an unbalance dataset setting in 102Flowers, 33.11, 31.44, and 14.05 higher than non-image augmentation, geometrical transformation, and focal loss, respectively. Currently, model-free and optimizing policies are widely leveraged, whereas the model-based approach is an active research topic for specific challenges, such as class imbalance and domain adaptation. In addition, although most algorithms are label-preserving, label-changing algorithms have recently received attention.

Future direction. Although many image augmentation algorithms exist, developing novel algorithms remains crucial to improve the performance of deep learning. We argue that recognizing new challenges or variations may inspire novel methods if they can be mimicked using image augmentation. Further, most algorithms of image augmentation are designed for classification and hence extending them to other applications is one of the most applicable directions by incorporating application-based knowledge, such as time-series in video [40]. Another interesting direction is distinguishing specific applications from general computer vision tasks such as ImageNet [99] and COCO [100] and then finding new motivations to design image augmentation. For example, most variations in plant healthy and diseased leaves are shared and thus can be converted from one to another [2]. Finally, considering image augmentation from a systematic perspective is appealing. For example, the effects of image augmentation schedules on optimization such as learning rate and batch size, are analyzed in [101].

Understanding image augmentation. This study was conducted to understand the objectives of image augmentation in the context of deep learning, from the perspectives of challenges and vicinity distribution. Although it was also verified that image augmentation is similar to regularization [79], most of the evidences are empirically from experiments. Understanding them in theory is therefore appealing. Recently, kernel theory [102] and group theory [103] have been used to analyze the effects of image augmentation. In addition, the improvement yielded by image augmentation in the context of model generalizability has been quantified using affinity and diversity [104].

New strategy to leverage image augmentation. Although image augmentation is commonly used in a supervised manner, this must not necessarily be the case. First, a pretext task can be created via image augmentation, such as predicting the degrees of rotation [105] and relative positions of image patches [106]. Second, image augmentation can be leveraged to generate positive samples for contrast learning under the assumption that an augmented image is similar to the corresponding original image [107–109]. Furthermore, semi-supervised learning benefits from image augmentation [79,110,111].

Feature augmentation attempts to perform augmentation in feature space instead of image space in image augmentation, and thus reduces the computation cost but without visual evidences. A feature space generally has dense information in semantic level than an image space. Consequently, operation in feature space is more efficient [112], such as domain knowledge [113]. Simultaneously, we believe that most of the techniques in image augmentation can be extended to feature augmentation, such as Manifold Mixup [114] from Mixup [28] and occluded feature [115].

8. Conclusion

This study surveyed a wide range of image augmentation algorithms with a novel taxonomy encompassing three categories: model-free, model-based, and optimizing policy-based. To understand the objectives of image augmentation, we analyzed the challenges of deploying a deep learning model for computer vision tasks, and adopted the concept of vicinity distribution. We found that image augmentation significantly improves task performance, and many algorithms have been designed for specific challenges, such as intensity transformations for occlusion, and model-based algorithms for class imbalance and domain shift. Based on this analysis, we argue that novel methods can be inspired by new challenges. Conversely, appropriate methods can be selected after recognizing the challenges posed by a dataset. Furthermore, we discussed the current situation and possible directions of image augmentation with three relevant interesting topics. We hope that our study will provide an enhanced understanding of image augmentation and encourage the community to prioritize dataset characteristics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This research was partly supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No.2019R1A6A1A09031717), supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) (No. 2020R1A2C2013060), and supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) and Korea Smart Farm R&D Foundation (KosFarm) through the Smart Farm Innovation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA), Ministry of Science and ICT (MSIT), and Rural Development Administration (RDA) (No. 421027-04).

References

[1] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[2] M. Xu, S. Yoon, A. Fuentes, J. Yang, D. Park, Style-consistent image translation: a novel data augmentation paradigm to improve plant disease recognition, Front. Plant Sci. 12: 773142. doi: 10.3389/fpls (2022).
[3] S.C. Wong, A. Gatt, V. Stamatescu, M.D. McDonnell, Understanding data augmentation for classification: when to warp? in: 2016 international conference on digital image computing: techniques and applications (DICTA), IEEE, 2016, pp. 1–6.
[4] L. Taylor, G. Nitschke, Improving deep learning with generic data augmentation, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2018, pp. 1542–1547.

[5] P. Kaur, B.S. Khehra, E.B.S. Mavi, Data augmentation for object detection: A review, in: 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), IEEE, 2021, pp. 537–543.
[6] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.
[7] N.E. Khalifa, M. Loey, S. Mirjalili, A comprehensive survey of recent trends in deep learning for digital images augmentation, Artif. Intell. Rev. (2021) 1–27.
[8] X. Wang, K. Wang, S. Lian, A survey on face data augmentation for the training of deep neural networks, Neural Comput. Appl. 32 (19) (2020) 15503–15531.
[9] F.H.K.d.S. Tanaka, C. Aranha, Data augmentation using gans, arXiv preprint arXiv:1904.09135 (2019).
[10] H. Naveed, Survey: Image mixing and deleting for data augmentation, arXiv preprint arXiv:2106.07085 (2021).
[11] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[13] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
[14] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).
[15] B. Zoph, V. Vasudevan, J. Shlens, Q.V. Le, Learning transferable architectures for scalable image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8697–8710.
[16] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, et al., Resnest: Split-attention networks, arXiv preprint arXiv:2004.08955 (2020).
[17] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, Training data-efficient image transformers & distillation through attention, in: International Conference on Machine Learning, PMLR, 2021, pp. 10347–10357.
[18] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
[19] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, Adv. Inf. Process. Syst. 28 (2015).
[20] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
[21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: European conference on computer vision, Springer, 2016, pp. 21–37.
[22] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, Yolov4: optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934 (2020).
[23] 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, pp. 3544–3553.
[24] T. DeVries, G.W. Taylor, Improved regularization of convolutional neural networks with cutout, arXiv preprint arXiv:1708.04552 (2017).
[25] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, in: Proceedings of the AAAI conference on artificial intelligence, volume 34, 2020, pp. 13001–13008.
[26] P. Chen, S. Liu, H. Zhao, J. Jia, Gridmask data augmentation, arXiv preprint arXiv:2001.04086 (2020).
[27] H. Inoue, Data augmentation by pairing samples for images classification, arXiv preprint arXiv:1801.02929 (2018).
[28] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: beyond empirical risk minimization, arXiv preprint arXiv:1710.09412 (2017).
[29] Y. Tokozume, Y. Ushiku, T. Harada, Between-class learning for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5486–5494.
[30] S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: Regularization strategy to train strong classifiers with localizable features, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6023–6032.
[31] D. Hendrycks, N. Mu, E.D. Cubuk, B. Zoph, J. Gilmer, B. Lakshminarayanan, Augmix: A simple method to improve robustness and uncertainty under data shift, in: International conference on learning representations, volume 1, 2020, p. 6.
[32] J.-H. Kim, W. Choo, H.O. Song, Puzzle mix: Exploiting saliency and local statistics for optimal mixup, in: International Conference on Machine Learning, PMLR, 2020, pp. 5275–5285.
[33] J. Kim, W. Choo, H. Jeong, H.O. Song, Co-mixup: Saliency guided joint mixup with supermodular diversity, in: International Conference on Learning Representations, 2021.
[34] A. Dabouei, S. Soleymani, F. Taherkhani, N.M. Nasrabadi, Supermix: Supervising the mixing data augmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13794–13803.
[35] K. Baek, D. Bang, H. Shim, Gridmix: strong regularization through local context mapping, Pattern Recognit. 109 (2021) 107594.
[36] D. Dwibedi, I. Misra, M. Hebert, Cut, paste and learn: Surprisingly easy synthesis for instance detection, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 1301–1310.
[37] G. Georgakis, A. Mousavian, A.C. Berg, J. Kosecka, Synthesizing training data for object detection in indoor scenes, arXiv preprint arXiv:1702.07836 (2017).
[38] N. Dvornik, J. Mairal, C. Schmid, Modeling visual context is key to augmenting object detection datasets, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 364–380.
[39] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E.D. Cubuk, Q.V. Le, B. Zoph, Simple copy-paste is a strong data augmentation method for instance segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2918–2928.
[40] Z. Xu, A. Meng, Z. Shi, W. Yang, Z. Chen, L. Huang, Continuous copy–paste for one-stage multi-object tracking and segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15323–15332.
[41] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434 (2015).
[42] A. Madani, M. Moradi, A. Karargyris, T. Syeda-Mahmood, Chest x-ray generation and data augmentation for cardiovascular abnormality classification, in: Medical Imaging 2018: Image Processing, volume 10574, International Society for Optics and Photonics, 2018, p. 105741M.
[43] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Synthetic data augmentation using gan for improved liver lesion classification, in: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), IEEE, 2018, pp. 289–293.
[44] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification, Neurocomputing 321 (2018) 321–331.
[45] T. Tran, T. Pham, G. Carneiro, L. Palmer, I. Reid, A bayesian data augmentation approach for learning deep models, Adv. Neural Inf. Process. Syst. 30 (2017).
[46] G. Douzas, F. Bacao, Effective data generation for imbalanced learning using conditional generative adversarial networks, Expert Syst. Appl. 91 (2018) 464–471.
[47] G. Mariani, F. Scheidegger, R. Istrate, C. Bekas, C. Malossi, Bagan: data augmentation with balancing gan, arXiv preprint arXiv:1803.09655 (2018).
[48] A. Antoniou, A. Storkey, H. Edwards, Data augmentation generative adversarial networks, arXiv preprint arXiv:1711.04340 (2017).
[49] A. Ali-Gombe, E. Elyan, Mfc-gan: class-imbalanced dataset classification using multiple fake class generative adversarial network, Neurocomputing 361 (2019) 212–221.
[50] H. Yang, Y. Zhou, Ida-gan: A novel imbalanced data augmentation gan, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 8299–8305.
[51] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, R. Webb, Learning from simulated and unsupervised images through adversarial training, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2107–2116.
[52] 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: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 718–731.
[53] Y. Zhu, M. Aoun, M. Krijn, J. Vanschoren, H.T. Campus, Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants, in: BMVC, 2018, p. 324.
[54] P.T. Jackson, A.A. Abarghouei, S. Bonner, T.P. Breckon, B. Obara, Style augmentation: data augmentation via style randomization, in: CVPR Workshops, volume 6, 2019, pp. 10–11.
[55] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F.A. Wichmann, W. Brendel, Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness, arXiv preprint arXiv:1811.12231 (2018).
[56] X. Zhu, Y. Liu, J. Li, T. Wan, Z. Qin, Emotion classification with data augmentation using generative adversarial networks, in: Pacific-Asia conference on knowledge discovery and data mining, Springer, 2018, pp. 349–360.
[57] E. Schwartz, L. Karlinsky, J. Shtok, S. Harary, M. Marder, A. Kumar, R. Feris, R. Giryes, A. Bronstein, Delta-encoder: an effective sample synthesis method for few-shot object recognition, Adv. Neural Inf. Process. Syst. 31 (2018).
[58] Y. Li, Q. Yu, M. Tan, J. Mei, P. Tang, W. Shen, A. Yuille, et al., Shape-texture debiased neural network training, in: International Conference on Learning Representations, 2020.
[59] M. Hong, J. Choi, G. Kim, Stylemix: Separating content and style for enhanced data augmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14862–14870.
[60] Z. Zheng, Z. Yu, Y. Wu, H. Zheng, B. Zheng, M. Lee, Generative adversarial network with multi-branch discriminator for imbalanced cross-species image–to-image translation, Neural Netw. 141 (2021) 355–371.
[61] E.D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q.V. Le, Autoaugment: Learning augmentation strategies from data, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 113–123.
[62] S. Lim, I. Kim, T. Kim, C. Kim, S. Kim, Fast autoaugment, Adv. Neural Inf. Process. Syst. 32 (2019).
[63] D. Ho, E. Liang, X. Chen, I. Stoica, P. Abbeel, Population based augmentation: Efficient learning of augmentation policy schedules, in: International Conference on Machine Learning, PMLR, 2019, pp. 2731–2741.
[64] R. Hataya, J. Zdenek, K. Yoshizoe, H. Nakayama, Faster autoaugment: Learning augmentation strategies using backpropagation, in: European Conference on Computer Vision, Springer, 2020, pp. 1–16.
[65] E.D. Cubuk, B. Zoph, J. Shlens, Q.V. Le, Randaugment: Practical automated data augmentation with a reduced search space, in: Proceedings of the IEEE/CVF

Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 702–703.
[66] R. Hataya, J. Zdenek, K. Yoshizoe, H. Nakayama, Meta approach to data augmentation optimization, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 2574–2583.
[67] B. Zoph, E.D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, Q.V. Le, Learning data augmentation strategies for object detection, in: European conference on computer vision, Springer, 2020, pp. 566–583.
[68] P. Li, X. Liu, X. Xie, Learning sample-specific policies for sequential image augmentation, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 4491–4500.
[69] A. Fawzi, H. Samulowitz, D. Turaga, P. Frossard, Adaptive data augmentation for image classification, in: 2016 IEEE international conference on image processing (ICIP), Ieee, 2016, pp. 3688–3692.
[70] A.J. Ratner, H. Ehrenberg, Z. Hussain, J. Dunnmon, C. Ré, Learning to compose domain-specific transformations for data augmentation, Adv. Neural Inf. Process. Syst. 30 (2017).
[71] Z. Tang, X. Peng, T. Li, Y. Zhu, D.N. Metaxas, Adatransform: Adaptive data transformation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2998–3006.
[72] X. Zhang, Q. Wang, J. Zhang, Z. Zhong, Adversarial autoaugment, arXiv preprint arXiv:1912.11188 (2019).
[73] D. Lee, H. Park, T. Pham, C.D. Yoo, Learning augmentation network via influence functions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10961–10970.
[74] T. Takase, R. Karakida, H. Asoh, Self-paced data augmentation for training neural networks, Neurocomputing 442 (2021) 296–306.
[75] V. Vapnik, Principles of risk minimization for learning theory, Adv. Neural Inf. Process. Syst. 4 (1991).
[76] C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals, Understanding deep learning requires rethinking generalization (2016), arXiv preprint arXiv:1611.03530 (2017).
[77] V. Vapnik, The nature of statistical learning theory, Springer science & business media, 1999.
[78] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.
[79] O. Chapelle, J. Weston, L. Bottou, V. Vapnik, Vicinal risk minimization, Adv. Neural Inf. Process. Syst. 13 (2000).
[80] M.P. Ekstrom, Digital image processing techniques, volume 2, Academic Press, 2012.
[81] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic urban scene understanding, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.
[82] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao, Scaled-yolov4: Scaling cross stage partial network, in: Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, 2021, pp. 13029–13038.
[83] W. Ma, Y. Wu, F. Cen, G. Wang, Mdfn: multi-scale deep feature learning network for object detection, Pattern Recognit. 100 (2020) 107149.
[84] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.
[85] W. Xu, K. Shawn, G. Wang, Toward learning a unified many-to-many mapping for diverse image translation, Pattern Recognit. 93 (2019) 570–580.
[86] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International conference on machine learning, PMLR, 2020, pp. 1597–1607.
[87] J. Yang, Gridmask based data augmentation for bengali handwritten grapheme classification, in: Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing, 2020, pp. 98–102.
[88] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[89] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784 (2014).
[90] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier gans, in: International conference on machine learning, PMLR, 2017, pp. 2642–2651.
[91] X. Huang, S. Belongie, Arbitrary style transfer in real-time with adaptive instance normalization, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 1501–1510.
[92] V. Dumoulin, J. Shlens, M. Kudlur, A learned representation for artistic style, arXiv preprint arXiv:1610.07629 (2016).
[93] R. Li, W. Cao, Q. Jiao, S. Wu, H.-S. Wong, Simplified unsupervised image translation for semantic segmentation adaptation, Pattern Recognit. 105 (2020) 107343.
[94] H.-Y. Lee, H.-Y. Tseng, Q. Mao, J.-B. Huang, Y.-D. Lu, M. Singh, M.-H. Yang, Drit++: diverse image-to-image translation via disentangled representations, Int. J. Comput. Vis. 128 (10) (2020) 2402–2417.
[95] T. Park, M.-Y. Liu, T.-C. Wang, J.-Y. Zhu, Semantic image synthesis with spatially-adaptive normalization, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 2337–2346.
[96] T. Fontanini, E. Iotti, L. Donati, A. Prati, Metalgan: multi-domain label-less image synthesis using cgans and meta-learning, Neural Netw. 131 (2020) 185–200.
[97] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347 (2017).
[98] J. Lemley, S. Bazrafkan, P. Corcoran, Smart augmentation learning an optimal data augmentation strategy, IEEE Access 5 (2017) 5858–5869.
[99] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.
[100] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: European conference on computer vision, Springer, 2014, pp. 740–755.
[101] B. Hanin, Y. Sun, How data augmentation affects optimization for linear regression, Adv. Neural Inf. Process. Syst. 34 (2021).
[102] T. Dao, A. Gu, A. Ratner, V. Smith, C. De Sa, C. Ré, A kernel theory of modern data augmentation, in: International Conference on Machine Learning, PMLR, 2019, pp. 1528–1537.
[103] S. Chen, E. Dobriban, J. Lee, A group-theoretic framework for data augmentation, Adv. Neural Inf. Process. Syst. 33 (2020) 21321–21333.
[104] R. Gontijo-Lopes, S. Smullin, E.D. Cubuk, E. Dyer, Tradeoffs in data augmentation: An empirical study, in: International Conference on Learning Representations, 2020.
[105] N. Komodakis, S. Gidaris, Unsupervised representation learning by predicting image rotations, in: International Conference on Learning Representations (ICLR), 2018.
[106] C. Doersch, A. Gupta, A.A. Efros, Unsupervised visual representation learning by context prediction, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1422–1430.
[107] M. Ye, X. Zhang, P.C. Yuen, S.-F. Chang, Unsupervised embedding learning via invariant and spreading instance feature, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6210–6219.
[108] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al., Bootstrap your own latent-a new approach to self-supervised learning, Adv. Neural Inf. Process. Syst. 33 (2020) 21271–21284.
[109] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, A. Joulin, Emerging properties in self-supervised vision transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9650–9660.
[110] D. Berthelot, N. Carlini, E.D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, C. Raffel, Remixmatch: semi-supervised learning with distribution alignment and augmentation anchoring, arXiv preprint arXiv:1911.09785 (2019).
[111] Q. Xie, Z. Dai, E. Hovy, T. Luong, Q. Le, Unsupervised data augmentation for consistency training, Adv. Neural Inf. Process. Syst. 33 (2020) 6256–6268.
[112] Y. Wang, G. Huang, S. Song, X. Pan, Y. Xia, C. Wu, Regularizing deep networks with semantic data augmentation, IEEE Trans. Pattern Anal. Mach. Intell. (2021).
[113] Q. Wang, F. Meng, T.P. Breckon, Data augmentation with norm-vae for unsupervised domain adaptation, arXiv preprint arXiv:2012.00848 (2020).
[114] V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, D. Lopez-Paz, Y. Bengio, Manifold mixup: Better representations by interpolating hidden states, in: International Conference on Machine Learning, PMLR, 2019, pp. 6438–6447.
[115] F. Cen, X. Zhao, W. Li, G. Wang, Deep feature augmentation for occluded image classification, Pattern Recognit. 111 (2021) 107737.

Mingle Xu received the B.S. degree from Jiangxi Agricultural University, in 2015, and the M.S. degree from the Shanghai University of Engineering Science, in 2018. He is currently pursuing a Ph.D. degree majoring in electronic engineering with Jeonbuk National University. His research interests include artificial intelligence and machine learning, computer vision, and image understanding.

Sook Yoon received a Ph.D. degree in electronics engineering from Jeonbuk National University, South Korea, in 2003. She is currently a Professor at the Department of Computer Engineering, Mokpo National University, South Korea. She was a Researcher in electrical engineering and computer sciences with the University of California at Berkeley, Berkeley, USA, until June 2006. She joined Mokpo National University, in September 2006. She was a Visiting Scholar with the Utah Center of Advanced Imaging Research, University of Utah, USA, from 2013 to 2015. Her research interests include computer vision, object recognition, machine learning, and biometrics.

Alvaro Fuentes received the B.S. degree in mechatronics engineering from the Technical University of the North, Ecuador, in 2012, and the M.S. and Ph.D. degrees in electronics engineering majoring in arti_cial intelligence and computer vision from JeonBuk National University, South Korea, in 2016 and 2019, respectively. He is currently a Postdoctoral Researcher with the Department of Electronics Engineering, Jeonbuk National University. His research interests includemachine learning, deep learning, computer vision, and robotics.

Dong Sun Park received the B.S. degree from Korea University, Republic of Korea, in 1979, and the M.S. and Ph.D. degrees from the University of Missouri, USA, in 1984 and 1990, respectively. He is currently a Professor at Jeonbuk National University, Republic of Korea. He has published many papers in international conferences and journals. His research interests include computer vision and artificial neural networks, especially deep learning.