Article

Data Augmentation in Classification and Segmentation: A Survey and New Strategies

Khaled Alomar *, Halil Ibrahim Aysel and Xiaohao Cai

School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK
* Correspondence: k.a.alomar@soton.ac.uk

Abstract: In the past decade, deep neural networks, particularly convolutional neural networks, have revolutionised computer vision. However, all deep learning models may require a large amount of data so as to achieve satisfying results. Unfortunately, the availability of sufficient amounts of data for real-world problems is not always possible, and it is well recognised that a paucity of data easily results in overfitting. This issue may be addressed through several approaches, one of which is data augmentation. In this paper, we survey the existing data augmentation techniques in computer vision tasks, including segmentation and classification, and suggest new strategies. In particular, we introduce a way of implementing data augmentation by using local information in images. We propose a parameter-free and easy to implement strategy, the random local rotation strategy, which involves randomly selecting the location and size of circular regions in the image and rotating them with random angles. It can be used as an alternative to the traditional rotation strategy, which generally suffers from irregular image boundaries. It can also complement other techniques in data augmentation. Extensive experimental results and comparisons demonstrated that the new strategy consistently outperformed its traditional counterparts in, for example, image classification.

Keywords: data augmentation; deep learning; convolutional neural networks; image processing; segmentation; classification

1. Introduction

Deep neural networks, like convolutional neural networks (CNNs), have been used in computer vision with numerous research applications, such as action recognition [1,2], object detection and localisation [3,4], face recognition [5], and image characterisation [6]. They have achieved superior performance against conventional approaches in many challenging computer vision tasks [7]. Nevertheless, their shortcomings, such as large-scale data requirements, long training time, overfitting, and performance slumps upon data scarcity, may hinder their generalisation and effectiveness [8,9].

The fruitful results presented by the CNN models encourage researchers to pursue higher accuracy models. These results are generally achieved by building more complex architectures [10]. Note that model complexity is often described by the number of trainable parameters. The more trainable parameters a model has, the more complex it is. More specifically, model complexity may also be defined in terms of the number of layers (i.e., non-linearity) and the number of neurons (e.g., filters) in individual layers. On the other hand, in supervised learning, data complexity can be determined according to the inter-class multiplicity (i.e., different classes) in addition to the intra-class differences. In general, the complex of the data and the model needed is proportional. If the training data is insufficient, complex models may be susceptible to the issue of memorising the training data. It is also well known that deep neural networks prevail partly because of the availability of high volume data. The networks can easily memorise data points due to their complex structure. However, the increasing complexity of the model architectures with insufficient data could exacerbate the shortcomings of CNN models [11]. One of the most

apparent issues when adopting complex CNN models is the overfitting problem [12], which can be described as the performance difference between the training and validation/test stages, where the model loses its ability to generalise. Overfitting generally occurs when a model is either too complex for the data or the data itself is insufficient [13]. Figure 1 shows an example of the loss curve of an overfit model. Although the training accuracy and validation accuracy improved concurrently during the early stages of training, they diverged after a certain point, where the model started losing its generalisation ability [12]. Strategies like reducing the model complexity, applying regularisation, and/or acquiring more extensive data volumes have been considered to mitigate the overfitting issue in deep learning models, see Figure 2.
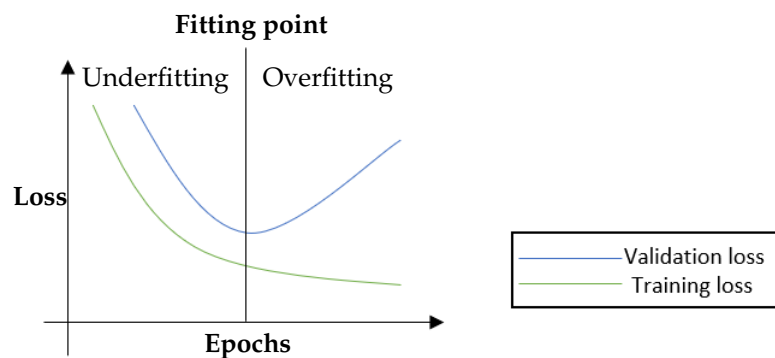

Figure 1. An illustration of the training and validation loss curves. Training and validation losses decrease simultaneously until the fitting point. After that, the validation loss begins to rise while the training loss is still decreasing, i.e., the so-called overfitting. Overfitting is associated with good performance on the training data but poor generalisation to the validation/test data (cf. underfitting is associated with poor performance on the training data and poor generalisation to the validation data) [12].

Regularisation techniques are implemented at the model architectural level [14,15], such as dropout [16], ridge regression (ℓ_2 regularisation) [17], and Lasso regression (ℓ_1 regularisation) [18]. The main objective of these techniques is to reduce the complexity of a neural network model during training, which is considered the main reason behind overfitting, especially when the model is trained on small datasets. Other techniques, like batch normalisation and transfer learning, may speed up the training process and also have an impact on preventing overfitting [19,20]. These techniques could be regarded as byproducts of the constant competition in the pursuit of higher performance by innovating new complex deep neural architectures, such as VGG-16 [21], ResNet [22], Inception-V3 [23] and DenseNet [24]. These models, in fact, aim to achieve higher accuracy on large datasets like Imagenet [25], which has over 14 million images [25]. However, when applying these models to small-scale applications with small datasets, they usually suffer from poor generalisation and overfitting, indicating the necessity of developing methods to reduce their complexity.
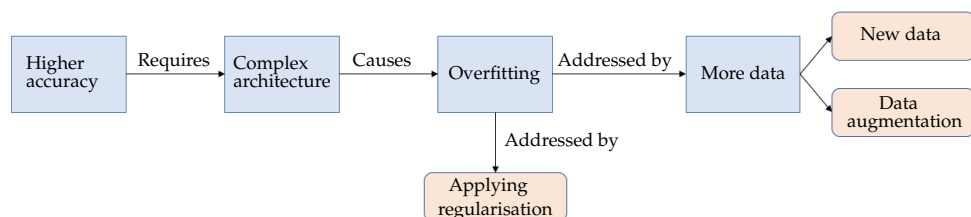

Figure 2. Diagram illustrating the overfitting problem and its well-known solutions.

Data augmentation methodology encompasses a broader range of techniques that function at the data level, rather than at the model architectural level. It can help deep

learning models perform better by artificially creating different and diverse samples with balanced classes for the training dataset. When the dataset is sufficient in terms of quantity and quality, a deep learning model performs better and more accurately. In other words, the training data must fulfil two requirements, i.e., adequate diversity and size, both of which can be achieved by data augmentation [26].

Data augmentation can be categorised based on the intended purpose of applying it (i.e., increasing training dataset size and/or diversity), or it can be categorised based on the problems. The following are examples of the latter: the random erasing technique was proposed to address the occlusion problem [8]; rotation and flipping were supposed to partially resolve the viewpoint problem [27–29]; brightness was used to address the change in lighting [30]; and cropping and zooming were used to address the scaling and background issues. In particular, the most popular categorisation of data augmentation divides it into deep learning-based data augmentation and traditional data augmentation [10], which is further divided into geometric, photometric, and noise data augmentation, see Figure 3. For reviews on the deep learning approaches for data augmentation, see e.g., [31,32].

Several studies evaluating the efficacy of data augmentation have utilised standard academic image datasets to assess results. For example, MNIST, CIFAR-10, CIFAR-100 and ImageNet are four commonly used datasets [10,33–35]. Note that some of these datasets, especially ImageNet, are considered "big data" [36] and may not require data augmentation techniques to further increase their size. To simulate data scarcity challenges, many experiments testing data augmentation techniques limit themselves to small subsets of the original large datasets [34]. It is worth emphasising that data augmentation techniques may also be used to improve the data diversity, except for the data quantity.

Figure 3. Data augmentation (DA) taxonomy.

This survey mainly focused on recent articles that used data augmentation techniques in image classification and segmentation, regardless of the data augmentation category, models, or datasets used in the studies. To the best of our knowledge, there are few surveys in the fields of data augmentation in image classification and segmentation. Another main contribution of this article is that we propose a new geometric data augmentation technique, which can complement the current data augmentation strategies. It is well known

that traditional rotation is one of the most commonly used geometric data augmentation techniques, see Figure 4. It, however, has drawbacks; for example, the loss of a significant amount of pixel information when rotating. It is noticeable that rotating a square-shaped image in a circular trajectory produces black patches at the boundaries, which do not accurately reflect the original data and may affect the final augmentation performance. Filling these black patches with modified pixel values via the wrap, constant, reflection, and/or nearest rotation techniques was a common solution to this issue (see Figure 5). In this study, we suggest exploiting local information in images and propose conducting rotation randomly and locally to address the limitations of the traditional rotation. We named our method "*random local rotation*" (RLR). RLR rotates an image's internal circular region by selecting random location, area and angle, which is easy to implement. Rotation performed in a local manner avoids forming black regions near image boundaries. Moreover, this method could also improve the data diversity. Extensive experiments demonstrated its superior performance compared to its counterpart, i.e., the traditional rotation technique.

The remainder of this article is organised as follows. Sections 2 and 3 recall the most common traditional data augmentation methods and the most common deep learning-based data augmentation methods, respectively. Section 4 reviews some recent research in image classification and segmentation utilising data augmentation for performance enhancement. Sections 5 and 6 present our proposed data augmentation method and the experimental results validating its promising performance. We conclude our study in Section 7.
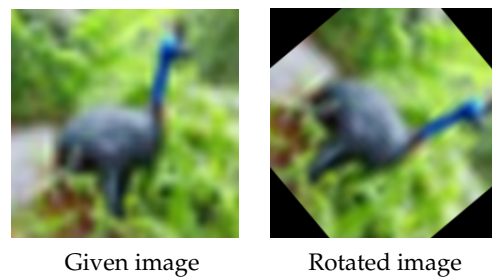

Given image   Rotated image

Figure 4. Traditional rotation. Left and right: the given image (from the CIFAR-10 dataset) and the rotated image with a randomly rotated angle. Black areas appear in the corners of the rotated image and the corners in the given image are cut off in the rotated image.

Figure 5. Data augmentation by different types of rotation techniques. The left three figures in the first row show the "constant" technique, i.e., the traditional rotation (TR), resulting in black areas around the boundary. The RNR technique is used in the right three figures of the first row. The first three figures of the second row give the results of filling up the black areas by using RRR, and the right three figures use RWR. For each technique, three random angles were selected for rotation.

2. Traditional Data Augmentation Techniques

This section briefly recalls the most commonly used traditional data augmentation approaches.

2.1. Geometric Transformations

Basic geometric operations. like flipping, cropping and random rotation, are still sought-after techniques to augment data. They generally increase the data size to improve data diversity, and are fairly easy to apply, see below for more detailed description.

Flipping. The term flipping refers to the process of flipping images horizontally or vertically or both, see Figure 6. The most commonly used flipping is horizonal flipping, since it is more realistic. For example, a cat versus dog dataset may include all the dog images heading to the left from the spectator view. Not surprisingly, the trained model may suffer from misclassifying dogs heading to the right. The best way to alleviate this problem is to collect more training images that include as many different views as possible. When collecting more images is difficult, flipping may directly solve this type of problem.

Flipping is one of the most intuitive strategies to increase data size or diversity. However, it may be inappropriate when the data has unique properties. For example, considering the concept of label safety, discussed in [10], asymmetric or direction sensitive data, such as letters or digit numbers, cannot use the flipping strategy since it results in inaccurate labels, or even opposite labels.
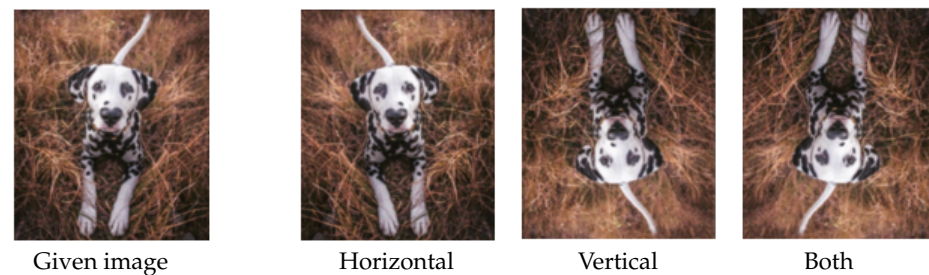

Given image   Horizontal   Vertical   Both

Figure 6. Data augmentation by flipping. Images from left to right represent the given image, horizontally flipped image, vertically flipped image and the image flipped horizontally and vertically, respectively.

Cropping. Cropping is a basic augmentation technique that randomly crops a part of the given image and then resizes the cropped part back to a certain size. As training data may include samples of different sizes, cropping images to a certain size is a widely used step before training [10,37].

It is worth mentioning that cropping may generate samples with incorrect labels. For example, images containing more than one object, which are labelled according to the object with dominant size, may experience a problem when using the cropping technique. In such a case, it is possible to crop an area of the given image that has more details of the accompanying object, rather than the dominant object, see Figure 7. The conventional strategy for training modern state-of-the-art architectures is to crop patches as small as 8% of the given image and label them the same as the given image [38]. This frequently results in incorrect labelling in the augmented data, as in the example shown in Figure 7.

Given image   Cropped patch

Figure 7. Data augmentation by cropping. Left and right: the given image (from ImageNet) labelled as "Dog" and the cropped patch. It is clear that the "Dog" is no longer visible in the cropped patch.

Rotation. Rotation is a simple geometric data augmentation technique. The images are rotated by a specified angle, and the newly created images are used alongside the originals

as training samples. The disadvantage of rotation is that it may result in information loss at the image boundary, see Figure 4 and the first row in Figure 5. There are several possible solutions, e.g., random nearest neighbor rotation (RNR), random reflect rotation (RRR) and random wrap rotation (RWR), to fix the boundary problem of the rotated images. In particular, the RNR technique repeats the nearest pixel values to fill in the black areas, while the RRR technique employs a mirror-based approach and the RWR technique uses the periodic boundary strategy to fill in the gaps; see Figure 5 for an example.

These geometric data augmentation techniques have been shown to be highly effective in improving diversity and increasing data quantity. For example, Mash et al. [39] used a fine-grained dataset of ten classes to test a variety of geometric augmentation methods for the task of aircraft classification. Cropping, rotating, rescaling, polygon occlusion, and a combination of these techniques were all tested. The cropping technique combined with occlusion achieved the highest improvement, i.e., increasing the task performance by 9% against the benchmark result. Their study, however, did not examine photometric data augmentation strategies (see below).

2.2. Photometric Transformations

A different type of traditional transformation is to change pixels' values rather than their positions. This approach includes different techniques, such as changing brightness, contrast, and/or colours.

Typically, a digital image is encoded as a tensor of three dimensions, i.e., height × width × colour channels. The difference between different colour representation schemes lies in the channel part of the tensor. For example, the RGB colour representation scheme uses a combination of three colour channels (i.e., red, green and blue) to represent individual pixels. Manipulating these individual colour channels is a very basic technique in colour augmentation [10]. For example, an image can be swiftly transformed into its representation in one colour channel if the others are set to black.

In addition to the RGB colour space, there are many other colour spaces. For example, the HSL colour representation scheme combines hue, saturation and lightness to represent individual pixels [40]. A hue is a single pigment that has no tint or shade. Saturation refers to colour intensity and lightness refers to how light a colour is. HSL is user-friendly since it is convenient to see how a particular colour appears using different values for these three attributes. Please refer to e.g., [40,41] for different colour spaces. Transferring from one colour space to another can be a useful technique for data augmentation.

Colour jittering is a photometric data augmentation technique that employs either random colour manipulation [42] or predetermined colour adjustments [43], such as randomly changing the brightness, contrast or colour properties of an image, see Figure 8.

(**a**) Given image　　　(**b**) Saturation　　　(**c**) Brightness　　　(**d**) Contrast

Figure 8. Data augmentation by colour jittering. (**a**–**d**) represent the given image, and the augmented images by manipulating the colour saturation, brightness and contrast, respectively.

Traditional photometric techniques in augmenting data may have limitations, e.g., high memory and computation requirements. In addition, they may result in crucial image information loss, particularly when the feature is a colorimetric feature capable of differentiating different dataset categories [44].
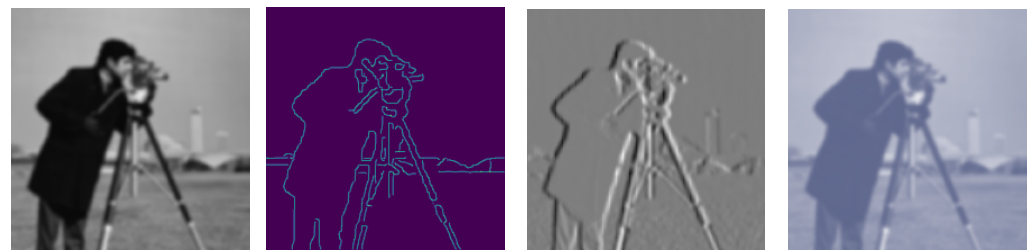
2.3. Kernel/Filter

Kernel plays an important role in deep learning. It can extract certain features from given images as a filter by sliding a window across the images. CNN models can learn

features from images by automatically updating their kernel values according to the back-propagation process. Similarly, kernels with distinct values can also be used to conduct data augmentation and generate specific images containing specific features [10].

In computer vision, filters can be used for edge detection (e.g., using the Sobel [45] or Canny [46] filters), sharpening (e.g., using high-contrast vertical or horizontal edge filters), and blurring (e.g., using the Gaussian filter). In particular, edge enhancement that improves object edges within images can be used for data augmentation. It is hypothesised that using training images with augmented edges could improve CNN performance, since the learned kernels in CNN could detect objects' shapes more easily [47]. Analogously, blurring images can also be utilised for data augmentation and could make models more resistant to blur or noise. Figure 9 shows an example of using different kernels/filers to augment images.

Using filters for data augmentation is a relatively unexplored field, even though the idea is straightforward. Its application in areas, such as action recognition, could be advantageous. For instance, edge detection filters may aid in recognising the human shape, thereby enabling the inference of its action. Motion blur may be used to augment data so as to improve models' resistance to blurring in action recognition [48,49].

(**a**) Given image　　　　(**b**) Canny filter　　　　(**c**) Sobel filter　　　　(**d**) Gaussian filter

Figure 9. Data augmentation by using kernels/filters.

2.4. Noise Transformations

Noise is commonly defined as a random variation in brightness or colour information [50]. It is frequently caused by technical limitations of the image capture sensor or poor environmental conditions. Unfortunately, these issues are often unavoidable in actual situations, making image noise a prevalent problem to address.

Noise in data may appear to be a problem for neural networks in particular. Real-world data is rarely perfect [51]. When neural networks are evaluated on real-world data, noise can impair their accuracy and cause them to perform poorly in generalisation. At the very least, the data used to test deep learning models may not be as clean as the data used to train them. This may account for why deep neural network models frequently perform poorly in tests. Their robustness could be improved by augmenting data with different types of noise. Gaussian, salt and pepper, and speckle noise are three well-known forms of noise that can be used to augment image data [52], e.g., see Figure 10.

Gaussian noise is statistical noise with a probability density function equal to the normal distribution. The distribution of Gaussian noise is uniform throughout the signal [53]. Since it is additive noise, the pixels in a noisy image are made up of the sum of their original pixel values plus random Gaussian noise values. It is also independent at each pixel, and independent of the signal magnitude. Salt-and-pepper noise is also known as "spike noise" or "impulsive noise". It causes white and black pixels to appear at random points in the image. This type of noise is mainly created by data transfer errors [54]. Speckle noise is multiplicative. It is generated by multiplying random values with different image pixels [53]. These different types of noise described above are generally dispersed over the image level. When they are used to augment data, deep learning models could be resistant to data that contains certain types of noise.
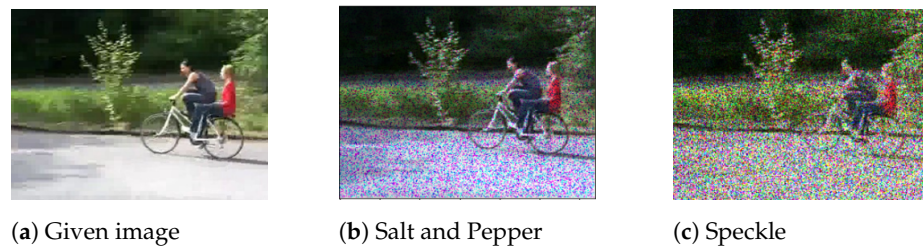
(**a**) Given image   (**b**) Salt and Pepper   (**c**) Speckle

Figure 10. Data augmentation by using noise transformation.

2.5. Random Erasing

Random erasing [8] is a data augmentation technique which does not attempt to change individual image pixel values in general. Instead, it replaces the values of the pixels within a random size rectangle in an image by a random value, see Figure 11 for example. We could regard random erasing as a kind of noise technique focusing on local areas rather than individual pixels. It intends to make the model resistant to occlusion of objects in images (e.g., the datasets CIFAR-10, CIFAR-100, and ImageNet) and, thus, to reduce the possibility of overfitting. It enhances the data diversity holistically without increasing the data size, which is different from the other aforementioned data augmentation methods.

Since the random erasing technique selects a rectangular area (i.e., occlusion region) randomly, it may entirely erase the object information to be classified in the image. Therefore, it may not be recommended in categorising sensitive data which cannot withstand the deletion of a randomly generated local area in images, such as the cases of categorising licence plate numbers and letters.

Figure 11. Data augmentation by the random erasing technique. The first and second rows represent the given images (from CIFAR-10) and the images after random erasing, respectively.

3. Deep Learning-Based Data Augmentation Techniques

This section briefly recalls the most commonly used deep learning-based data augmentation approaches.

3.1. Texture Transfer

Texture transfer [55] aims to generate textures from source images while maintaining control over the semantic content of the source images, e.g., see Figure 12. It allows the generation of new images with given textures, while preserving the original images' visual characteristics, such as contours, shading, lines, strokes and areas. The study in [56] demonstrated that CNNs are biased towards objects' texture rather than shape, indicating that employing texture transfer may make a model more texture resistant.

Base image	Reference style image	Resultant image

Figure 12. Data augmentation by texture transfer. Content of the base image (**left**) is mixed with the style of the reference style image (**middle**) to obtain the resultant image (**right**).

The majority of traditional texture transfer methods resample textures into each particular content image [57]. For example, image quilting [55] creates a new image by stitching together small patches of other images. The work in [58] developed an image analogue technique, using pixel resampling to transfer textures from one image to another [59]. The newly generated images could be added into the training dataset to enlarge the data size and enhance its diversity.

3.2. Adversarial Training

Adversarial examples, also known as machine illusion, have attracted considerable attention in the deep learning community. Adversarial examples can also be seen as members of the noise injection data augmentation family. By injecting a systematic noise into a given image, the CNN model outputs a completely different prediction, even though the human eye cannot detect the difference, see Figure 13. For example, the work in [60] created adversarial examples by changing a single pixel per image. Adversarial training is where these examples are added to the training set to make the model robust against attacks. As adversarial examples can detect weak points in a trained model, this way of augmenting data can be seen as an effective data augmentation approach.
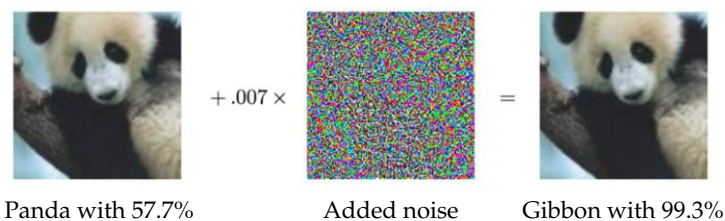

Panda with 57.7%	Added noise	Gibbon with 99.3%

Figure 13. An adversarial example taken from [61]. Even though the given image and the image after adversarial noise added look exactly the same to the human eye, the noise fools the model successfully, i.e., the model labels the two images as different classes.

3.3. Generative Adversarial Networks for Data Augmentation

Inspired by adversarial examples, the generative adversarial network (GAN), proposed in [62], has been widely used for data augmentation. Synthetic images created by GANs, which even humans find difficult to distinguish from the real images, help models significantly increase their robustness. GAN consists of two networks, i.e., a generator, which creates new images, and a discriminator, which tries to detect if the generated images are real or fake. For the variants of GANs, please refer to e.g., DCGAN [63], progressively growing GANs [64] and CycleGANs [65].

4. Data Augmentation in Image Classification and Segmentation

Data augmentations performed using traditional transformation techniques is still the most popular among academics, due to their simplicity [10,66]. Often, traditional and deep learning-based augmentation approaches are used either separately or in tandem. Image classification and image segmentation are two common, yet important, research areas in computer vision, which typically use data augmentation approaches. In this section, we

discuss recent research, mostly within the past five years, in these two areas that leverage data augmentation for performance enhancement.

4.1. Data Augmentation on Image Classification

Lots of works have used data augmentation in image classification tasks, and their results vary, depending on aspects such as, models, data and applications. See Table 1 for a brief survey in this respect.

Table 1. Survey of data augmentation techniques in recent image classification works.

Papers	Dataset	Aug. Techniques	Model	Task	Findings
Shijie et al., 2017 [34]	CIFAR10; ImageNet (10 categories)	GAN/WGAN, flipping, cropping, shifting, PCA jittering, colour jittering, noise, rotation.	AlexNet	Image classification	Four methods (i.e., cropping, flipping, WGAN and rotation) perform generally better than other augmentation methods, and some appropriate combination methods are slightly more effective than the individuals.
Perez et al., 2017 [67]	A small subset of ImageNet; MNIST.	Neural augmentation, CycleGAN, GANs, cropping, rotating, and flipping.	SmallNet	Image classification	GANs do not perform better than traditional techniques.
Hussain et al., 2017 [68]	Digital Database for Screening Mammography (DDSM).	Flipping, cropping, noise, Gaussian filters, principal component analysis (PCA).	VGG-16	Medical images classification	The flipping and Gaussian filter techniques are better than noise transformation.
Pawara et al., 2017 [69]	Folio, AgrilPlant, and the Swedish Leaf datasets	Rotation, blur, contrast, scaling, illumination, and projective transformation.	AlexNet; GoogleNet	Plant image classification	CNN models trained from scratch benefit significantly from data augmentation.
Inoue et al., 2018 [70]	ILSVRC2012; CIFAR-10	SamplePairing, flipping, distorting, noise, and cropping.	GoogLeNet	Image classification	Developed a new technique known as SamplePairing.
Li et al., 2018 [71]	Indian Pines and Salinas datasets	Pixel-block pair, flipping, rotation, and noise.	PBP-CNN	Hyperspectral imagery classification	A threefold increase in sample size is often sufficient to reach the upper bound.
Frid-Adar et al., 2018 [72]	Liver lesions dataset	Translation, rotation, scaling, flipping and shearing, and GAN-based synthetic images.	Customised small CNN architecture	Medical image classification	Combining traditional data augmentation with GAN-based synthetic images improves small datasets.
Pham et al., 2018 [73]	Skin lesion dataset (ISBI Challenge)	Geometric augmentation and colour augmentation.	InceptionV4	Skin cancer image classification	Skin cancer and medical image classifiers could benefit from data augmentation.
Motlagh et al., 2018 [74]	Tissue Micro Array; Breast Cancer Histopathological Images (BreaKHis)	Random resizing, rotating, cropping, and flipping.	ResNet50	Breast cancer image classification	Traditional data augmentation techniques are adequate for obtaining distinct samples of various types of cancer.
Zheng et al., 2019 [75]	Caltech 101; Caltech 256.	Neural style transfer, rotation, and flipping.	VGG16	Image classification	Neural style transfer can be utilised as a deep-learning data augmentation technique.
Ismael et al., 2020 [76]	Brain tumor dataset	Horizontal and vertical flips, rotating, shifting, zooming, shearing, and brightness alteration.	ResNet	MRI image classification (Brain Cancer)	The effectiveness of traditional augmentation methods varied among classes.
Gour et al., 2020 [77]	BreaKHis dataset	Stain normalisation, image patch generation, and affine transformation.	ResHist model	Breast cancer histopathological image classification	The model performance for classifying histopathology images is better with data augmentation than with pre-trained networks.
Nanni et al., 2021 [78]	Virus, a bark, a portrait, and a LIGO glitches datasets	Kernel filters, colour space transforms, geometric transformations, and random erasing.	ResNet50	Image classification	Introduced the discrete wavelet transform and the constant-Q Gbor transform as two new methods for data augmentation.
Anwar et al., 2021 [79]	Customised image based ECG signals	Flipping, cropping, contrast and Gamma distortion.	EfficientNet B3	ECG images classification	In the experiment with images of ECG signal, traditional data augmentation did not improve the performance of neural networks.
Kandel et al., 2021 [80]	MURA dataset	Horizontal flip, vertical flip, rotation, and zooming.	VGG19; ResNet50; InceptionV3; Xception; DenseNet121	X-ray images classification	Augmentation was found to significantly enhance classification performance.

In 2017, the work in [67] suggested that deep learning-based augmentation methods, like GANs, do not perform significantly better than traditional techniques, but consume nearly three times more computational cost. Moreover, in [67], a model called "SmallNet" was trained using traditional augmentation techniques and style transfer with Cycle-GAN [81]. It was observed that combining deep learning-based methods with traditional techniques could achieve better results. Hussain et al. [68] evaluated various augmentation strategies on a medical image dataset using VGG-16. They demonstrated that the flipping and Gaussian filter augmentation techniques yielded superior outcomes compared to the other ones, particularly when adding noise, which gave the lowest accuracy. Pawara et al. [69] applied data augmentation techniques, such as rotation, blur, contrast, scaling, illumination, projective transformation, and multiple combinations of these techniques to enhance plant image classification performance. In this challenge, pre-trained and un-trained AlexNet and GoogleNet models were used. It was observed that CNN models trained from scratch benefited significantly from data augmentation, whereas pre-trained CNN models did not. In addition, it was discovered that combinations of data augmentation techniques like rotation and varied illuminations could contribute most for CNN models trained from scratch in achieving excellent performance.

In 2018, Inoue et al. [70] developed a new technique, known as SamplePairing, in which a new sample was synthesised from one image by overlaying another image randomly selected from the training data, i.e., taking an average of two images. Li et al. [71] found that traditional data augmentation techniques were not cumulative, and that a threefold increase in sample size was often sufficient to reach the upper bound. In addition, the PBP technique proposed by the authors significantly increased the number of samples, and was proved to be effective for hyperspectral imagery classification. FridAdar et al. [72] classified liver lesions using a small customised CNN architecture. In order to accommodate small datasets and input sizes, they suggested that CNN designs should often contain fewer convolutional layers. By combining traditional data augmentation techniques with GAN-based synthetic images, more accurate results from a small dataset were obtained. Pham et al. [73] discussed how to solve the challenges of skin lesion classification and limited data in medical images by applying image data augmentation techniques, such as geometric augmentation and colour augmentation. The effects of a different number of augmented samples were evaluated on the performance of different classifiers, and it was concluded that the performance of skin cancer classifiers and medical image classifiers could be improved by utilising data augmentation. Motlagh et al. [74] classified several forms of cancer using 6402 tissue microarrays (TMAs) as training samples and utilising transfer learning and deep neural networks. Data augmentation techniques, such as random scaling, rotation, cropping, and flipping, were used to obtain sufficiently different samples, and the results showed that 99.8 percent of the four cancer types, including breast, bladder, lung and lymphoma, were correctly classified using the ResNet50 pre-trained model.

In 2019, Zheng et al. [75] assessed the efficacy of neural style transfer using VGG16 on the Caltech 101 and Caltech 256 datasets, and the results demonstrated a two-percent gain in accuracy. Recent research has demonstrated that neural style transfer algorithms can apply the artistic style of one image to another image without altering the latter's high-level semantic content, showing that neural style transfer can be used for data augmentation to add more variation to the training dataset.

In 2020, Ismael et al. [76] employed data augmentation to solve the problem of insufficient training data and imbalanced classes in the MRI image classification task for brain cancer. Various augmentation techniques, including horizontal and vertical flipping, rotating, shifting, zooming, shearing and brightness alteration, were utilised. They observed that each augmentation technique had different effects on the performance of distinct classes. For instance, manipulation of brightness yielded 96 percent accuracy for class one, whereas the rotation technique yielded 98 percent accuracy for the same class. For class two, these two techniques achieved a score of 99 percent with brightness and 98 percent with rotation. By combining all of the previously mentioned augmentation techniques, they were able to

attain 99 percent overall accuracy, i.e., 4 percent improvement against the results obtained without data augmentation. Additionally, Gour et al. [77] developed ResHist, a 152-layer CNN based on residual learning, for breast cancer histopathological image classification. A data augmentation strategy was devised, based on stain normalisation, image patch generation and affine transformation, to improve the model performance. Experimental results demonstrated that with the help of data augmentation the model performance for classifying histopathology images was better than the pre-trained networks, including AlexNet, VGG16, VGG19, GoogleNet, Inception-v3, ResNet50 and ResNet152.

In 2021, Kandel et al. [80] examined the impact of test time augmentation (TTA) on X-ray images for bone fracture detection using the MURA dataset. It was observed that TTA could dramatically improve classification performance, especially for models with a low score, by comparing the performance of nine different augmentation techniques with five state-of-the-art CNN models. Nanni et al. [78] investigated the performance of over ten different kinds of data augmentation techniques, including kernel filters, colour space transforms, geometric transformations, random erasing/cutting and image mixing, and proposed two approaches: the discrete wavelet transform and the constant-Q Gbor transform. Using the aforementioned data augmentation techniques, the performance of several ResNet50 networks was evaluated on four benchmark image datasets (i.e., a virus dataset, a bark dataset, a portrait dataset, and a LIGO glitches dataset), representing diverse problems and different scales, indicating the efficacy of data augmentation techniques in enhancing model performance. In addition, the work in [79] investigated the impact of augmenting ECG images for COVID-19 and cardiac disease classification using deep learning. They argued that traditional data augmentation did not improve the performance of neural networks in their experiments with ECG signal images.

4.2. Data Augmentation on Image Segmentation

Image segmentation is also an important field in computer vision. It involves grouping an image into different parts where each part may share certain features and characteristics. It has a close relationship to image classification. For example, image segmentation, in some sense, could be achieved by classifying individual pixels in an image into different groups. A great deal of emphasis has been placed on data augmentation in order to achieve better segmentation results, particularly when working with small training datasets. For practical semantic segmentation applications, collecting and annotating sufficient training data for deep neural networks is notoriously difficult. Therefore, data augmentation techniques are of great importance. We, below, survey a number of studies that have involved data augmentation in image segmentation tasks. See Table 2 for a summary of relevant literature.

In 2018, the work in [82] used an encoder–decoder structure, adapted from an hourglass network, prevalent in the field of human-pose estimation [83], in order to classify and segment brain tumours in MRI scans for the BraTS 2018 challenge [84–87]. Two data augmentation techniques were utilised: vertical flipping, which matches up to the naturally symmetrical shape of the brain, and random intensity variation, used because the intensity between MRI scans varies significantly. The network was trained with, and without, data augmentation. It was discovered that data augmentation appeared to provide a small increase in accuracy for the Dice coefficient and a significant improvement in Hausdorff accuracy.

Table 2. Survey of data augmentation techniques in recent image segmentation works.

Papers	Dataset	Aug. Techniques	Model	Task	Findings
Benson et al., 2018 [82]	BraTS 2018	Vertical flipping and random intensity variation.	Hourglass Network [83]	Brain tumor segmentation	Data augmentation methods appear to have a different impact on the Dice coefficient and Hausdorff accuracy.

Table 2. *Cont.*

Papers	Dataset	Aug. Techniques	Model	Task	Findings
Casado et al., 2019 [88]	ISBI challenge [89].	Automated image augmentation tool.	U-Net architecture with four different models	Semantic segmentation	Introduced an approach that enables researchers to use image augmentation techniques automatically to the challenges of object classification, localisation, detection, semantic segmentation, and instance segmentation.
Ma et al., 2019 [90]	Sheep segmentation dataset (SSG)	Colour transformation, flipping, cropping, projection transformation, local copy, and JPEG compression.	DeepLabv3+ [91]	Semantic segmentation	A combination of augmentation methods could achieve the best performance, while excessive augmentation could degrade the performance.
Qiao et al., 2020 [92]	Cattle segmentation dataset	Random image cropping and patching.	Bonnet [93]	Semantic segmentation	The proposed method of randomly cropping and patching images to increase the number of training images improves segmentation performance.
khryashchev et al., 2020 [94]	Planet, and the Resurs datasets	Random chromatic distortion, rotation, and shifting.	U-Net with the ResNet34 as encoder	Semantic segmentation	The application of random chromatic distortion in HSV colour format improves the robustness of deep learning algorithms for images with noise, such as small clouds and glare from reflective surfaces.
Chen et al., 2020 [95]	Tongue image dataset	Cropping, rotation, flipping, and colour transformations.	U-Net with 15 different CNN models as encoders	Semantic segmentation	Geometric transformations can achieve higher performance than colour transformations, and segmentation accuracy can be increased by 5 to 20% compared to no augmentation.
Qin et al., 2020 [96]	Kidney Tumour dataset	An automatic deep reinforcement learning based augmentation method	An end-to-end augmentation segmentation architecture	Medical image segmentation	Conventional augmentation techniques (e.g., rotation, cropping, etc.) are random and sometimes damaging to the image segmentation task.
Cirillo et al., 2021 [97]	BraTS2020 Dataset	Flipping, rotation, scaling, brightness adjustment, and elastic deformation.	3D U-Net [98]	3D brain tumor segmentation	Conventional data augmentation significantly improves the validation performance of brain tumour segmentation.
Su et al., 2021 [99]	Narrabri and Bonn datasets	Random image cropping and patching (RICAP) method.	Bonnet	Semantic segmentation	The RICAP technique increases the mean accuracy and mean intersection over union (IOU) of the CNNs with the traditional data augmentation.
Zhang et al., 2021 [100]	PASCAL VOC 2012; Cityscapes; CRAG dataset	Object-level augmentation method.	MobileNet based DeepLab V3+	Semantic segmentation	ObjectAug can easily be integrated with existing image-level augmentation techniques to further improve the segmentation performance. ObjectAug supports category-aware augmentation that gives objects in each category a variety of options.
Mallios et al., 2021 [101]	VoxTox [102]	GAN-based synthetic images.	RS-FCN	Rectum segmentation	Demonstrated the viability of producing synthetic data and subsequently incorporating it into the training samples in order to get satisfactory outcomes.

In 2019, Casado et al. [88] presented a versatile method which was implemented in the open-source package CLoDSA, dedicated to classification, semantic segmentation, instance segmentation, localisation and detection. Three different datasets were used to demonstrate the benefits of applying data augmentation. Ma et al. [90] created the SSG dataset, i.e., a small-scale and open-source sheep segmentation dataset containing hundreds of images. To find the best technique for this small semantic segmentation dataset, they evaluated seven data augmentation methods, including colour transformation, flipping, cropping, projection transformation, local copy, a proposed technique named "JPEG compression" and their combinations. Experimental results showed that the combination of compression, cropping and local shift could achieve the best augmentation performance for their AI-Ranch application. However, they also found that excessive augmentation could degrade performance.

In 2020, Qiao et al. [92] introduced a data augmentation technique where images were randomly cropped into distinct regions and then patched together to form a new one. Experimental results on their acquired cattle dataset showed that this data augmentation

technique, together with an open-source semantic segmentation CNN architecture, "Bonnet" [93], achieved 99.5 percent mean accuracy and 97.3 percent mean intersection of unions. In [94], an U-Net neural network with the ResNet34 encoder was used for automated wildfire detection on high-resolution aerial photos using two small satellite RGB image datasets. To overcome the small data size challenge, data augmentation techniques, such as rotation, shifting and random chromatic distortion in HSV colour format, were used to increase the robustness of the deep learning algorithm for noisy images, such as small clouds and glare from reflective surfaces. The experimental results showed that data augmentation methods led to better results on test datasets for all metrics used in the experiments. Qin et al. [96] argued that the data generated by conventional augmentation techniques (e.g., rotation, cropping, etc.) was random and sometimes detrimental to the image segmentation process. In light of this, an automatic learning-based data augmentation technique was developed for CT kidney tumor segmentation.

The work in [95] focused on automatic tongue segmentation using 15 different pretrained network models (such as VGG, ResNet, ResNext, DenseNet, EfficientNet, inceptionV3, SE-ResNet, inception, ResNetV2, etc.). They utilised multiple label-preserving transformations to increase the size and diversity of the training dataset. Their findings indicated that geometric transformations could achieve greater performance than colour transformations, and that the segmentation accuracy could be improved by 5 to 20 percent compared to no augmentation.

In 2021, the work in [100] proposed a data augmentation technique, named ObjectAug, for image segmentation. The ObjectAug technique operates at the object level by first decoupling the image into individual objects and the background using semantic labelling and, then, each object is individually augmented using conventional augmentation techniques (e.g., scaling, shifting and rotation), followed by image inpainting, which is utilised to further restore the pixel artefacts introduced by object augmentation. The final step is integrating the augmented objects and background into an augmented image. Extensive experiments on both normal and medical image datasets demonstrated that the ObjectAug technique outperformed conventional augmentation techniques and improved segmentation performance. Cirillo et al. [97] examined how augmentation techniques, such as flipping, rotation, scaling, brightness adjustment and elastic deformation, affected the learning process when training a standard 3D U-Net [98] on the BraTS dataset [85,103,104]. In multiple cases, their findings indicated that data augmentation significantly improved validation performance. They presumed that the reason why data augmentation had not been thoroughly investigated for brain tumour segmentation was because the BraTS training set was quite large and several works [105,106] suggested that data augmentation would not be of much assistance.

Mallios et al. [101] investigated image-guided radiation therapy [102,107], which is one of the most prevalent methods for treating numerous types of cancer. Their study included the development of deep learning approaches for segmenting the organs-at-risk in CT images during radiation therapy. It was observed that the scarcity of annotated data, stemming from the difficulty and time consuming nature of manual annotation in this area, hindered research development for medical applications. In order to compensate for the shortage of labelled real-world data required to train very deep models, like FCN architecture [108], cGAN [109] was used to generate synthetic images. The experimental results illustrated the superior performance of the proposed segmentation methods for the rectum under the help of deep learning-based data augmentation. In [99], a framework for augmenting data for semantic segmentation, based on the random image cropping and patching (RICAP) method, was presented. Experiments on two datasets using Bonnet architecture [93] showed that the developed framework improved segmentation performance in terms of accuracy and mean intersection over union.

5. Proposed Strategy for Data Augmentation

In this section, we propose a new data augmentation technique, belonging to the traditional data augmentation category. It was inspired by techniques focusing on local areas in images, e.g., the random erasing technique.

Let $\mathcal{D}$ be the training dataset. Let $C_{x,y,r}$ be a circular area in an image $I \in \mathcal{D}$, with centre location (x,y) and radius r. Let $\theta \in [0, 2\pi]$ be an angle for rotation.

The main procedure of the proposed augmentation technique is given below. Firstly, $\forall I \in \mathcal{D}$, we select a circular area $C_{x,y,r}$ within image I, with a randomly generated centre (x,y) and radius r. Then, the image content within the circular area $C_{x,y,r}$ is rotated with a randomly generated angle $\theta \in [0, 2\pi]$, while the image content outside the circular area $C_{x,y,r}$ is kept, and we call this newly generated image $\tilde{I}$. Finally, image $\tilde{I}$ is used to augment the original training dateset $\mathcal{D}$. Here we suggest two ways. The first one is to use the generated image $\tilde{I}$ to replace the original image $I \in \mathcal{D}$. This way does not change the size of the dataset $\mathcal{D}$, but may change the data diversity. The other way is to add image $\tilde{I}$ into the dataset $\mathcal{D}$, which increases the dataset size and enhances the data diversity. We call the above technique random local rotation (RLR), see Figure 14 for the diagram showing the conducting of the RLR data augmentation strategy and Algorithm 1 for the summary of RLR.
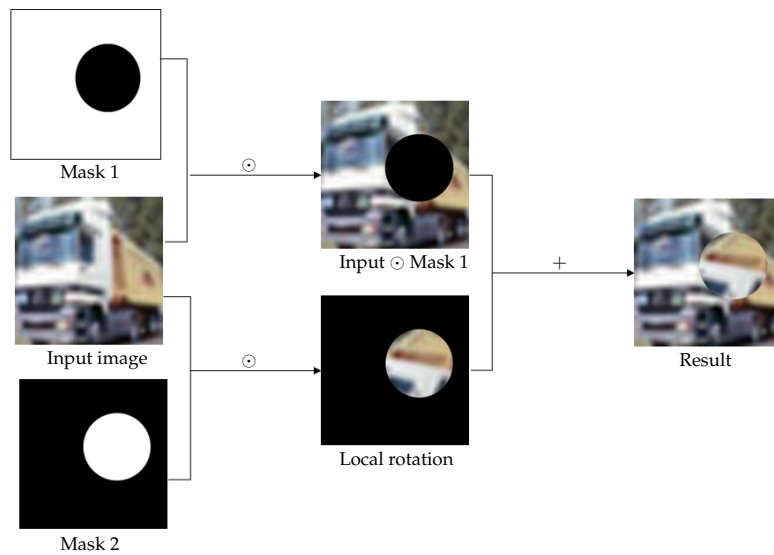

Figure 14. The proposed random local rotation data augmentation strategy. Symbol $\odot$ represents pointwise multiplication.

Algorithm 1 Random Local Rotation

1: **Input:** The training dataset $\mathcal{D}$
2: **Output:** The augmented training dataset $\tilde{\mathcal{D}}$
3: Create a subset say $\mathcal{D}' \subseteq \mathcal{D}$ by randomly selecting N images from $\mathcal{D}$
4: Create an empty set say $\mathcal{D}^*$
5: **for** $\forall I \in \mathcal{D}'$ **do**
6:     Randomly generate centre (x,y) and radius r within image I
7:     Form a circular area $C_{x,y,r}$ within image I with centre (x,y) and radius r
8:     Randomly generate angle $\theta \in [0, 2\pi]$
9:     Form image $\tilde{I}$ by rotating the area within $C_{x,y,r}$ in image I with angle θ
10:     Add $\tilde{I}$ into $\mathcal{D}^*$
11: **end for**
12: Way 1: $\tilde{\mathcal{D}} \leftarrow \mathcal{D}^* \cup \mathcal{D} \setminus \mathcal{D}'$
13: Way 2: $\tilde{\mathcal{D}} \leftarrow \mathcal{D}^* \cup \mathcal{D}$

A special case of RLR uses the largest possible circular rotation area in the image centre, see Figure 15. In the rest of this article, we call this special case random centre rotation (RCR). RCR could be applied as a direct replacement of the traditional rotation technique for data augmentation.

An obvious advantage of RLR against the traditional rotation is that it avoids the black boundary caused by traditional rotation, as shown in Figure 4. Moreover, the local area information distortion brought by RLR could improve the data diversity, without removing much information from the given images, like other augmentation techniques, e.g., image cropping, random erasing, etc. Detailed validation of RLR is presented in the next section.
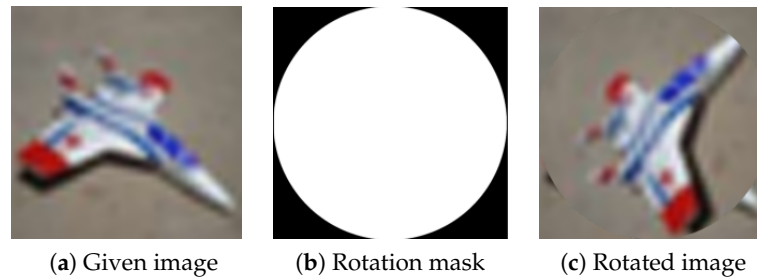

      (**a**) Given image   (**b**) Rotation mask   (**c**) Rotated image

Figure 15. Random local rotation data augmentation technique using the largest possible circular rotation area in the image centre.

6. Experiments

To validate the proposed RLR agumentation technique, we employed three state-of-the-art CNN models, i.e., ResNet50 [22], MobileNet [110] and InceptionV3 [23], which were all trained from scratch. We conducted experiments in both classification and segmentation tasks and mainly compared with the traditional rotation technique (shortened to TR) with randomly selected rotation angles. The quantitative results reported below with standard deviation were obtained by repeating the experiments five times.

6.1. Classification Experiment

The CIFAR-10 dataset was selected for conducting experiments regarding the classification task. It contained 60,000 coloured images, where every image was of size 32×32, A total of 50,000 images were for training and 10,000 images were for testing. CIFAR-10 consisted of ten classes, each with 6000 images. To simulate the scenario of data scarcity, we reduced the original training data size to 2%, 4% and 6%, forming three subsets with numbers of samples of 1000, 2000 and 3000, respectively, and used the original test set for testing.

For each subset, three extra copies were created by the TR, RCR and RLR data augmentation techniques. Each augmented copy was twice as large as its corresponding original subset. The data balance between the classes was also taken into consideration when constructing these subsets. Additionally, the image resolution was adjusted to fit the default input shape of each CNN model used in the experiments, i.e., 299×299 for InceptionV3, and 244×244 for MobileNet and ResNet50. According to the constructed datasets, each model was subjected to a total of 12 tests, (i.e., number of subsets $\times$ number of techniques).

For fair comparison, same hyperparameters were kept for each model. Models were trained for 50 epochs with the Adam optimiser and categorical cross entropy loss function. Test accuracy was selected as the monitoring metric. The Spyder platform was utilised to train and evaluate the models.

6.1.1. Classification Results

Table 3 gives the classification accuracy of the CNN models (i.e., ResNet50, MobileNet and InceptionV3) with the TR, RCR and RLR data augmentation techniques on the

three subsets, including the comparison with the baseline results (i.e., the ones obtained on the subsets without using data augmentation).

The results in Table 3 show that our proposed data augmentation technique RLR constantly achieved the best classification accuracy with all the three CNN models on all the subsets, indicating its excellent performance. In contrast, the traditional rotation technique did not improve performance, and, in many cases, degraded the results, compared with the baseline results. This might be because of the aforementioned limitations of the traditional rotation technique, i.e., the black and irregular boundary it introduces. As for the performance of the RCR technique, its results were slightly better than the TR results and were comparable to the baseline results. This was what we expected, since RCR is quite similar to TR. Yet, the images generated by RCR did not suffer from the black and irregular boundaries, and, therefore, performed slightly better than TR.

For further performance evaluation, we also reported the comparison of the RLR method with the mostly used traditional data augmentation techniques, see Table 4. The smallest subset of CIFAR-10 (i.e., the one with 1000 samples) was employed with data augmentation techniques, including RLR, RNR, RWR, RRR, flipping, shifting, zooming, and brightness. The results in Table 4 show that, generally, data augmentation techniques could indeed enhance the performance of different models. It again demonstrated the great performance of the proposed RLR method; for example, RLR achieved the best accuracy when the ResNet model was used. The results in Table 4 also show that the performance of the augmentation techniques might differ for different models, which is worth investigating further in the future.

Table 3. Classification accuracy comparison between the TR, RCR and RLR data augmentation techniques. CNN models, i.e., MobileNet, ResNet and InceptionV3, with the data augmentation techniques, were applied on three different CIFAR-10 subsets, with numbers of samples of 1000, 2000 and 3000, respectively. The results indicated the superior performance of the proposed RLR technique.

Model	Subset	Baseline	RLR	TR	RCR
MobileNet	1000	41.69 ± 0.29	$\mathbf{42.24} \pm 0.44$	40.57 ± 0.22	39.51 ± 0.52
	2000	50.62 ± 0.43	$\mathbf{51.76} \pm 0.56$	48.77 ± 0.51	50.6 ± 0.64
	3000	56.95 ± 0.62	$\mathbf{60.96} \pm 0.54$	55.30 ± 0.82	58.18 ± 0.66
ResNet	1000	39.73 ± 0.64	$\mathbf{41.47} \pm 0.39$	38.11 ± 0.52	38.28 ± 0.51
	2000	50.16 ± 0.49	$\mathbf{51.06} \pm 0.54$	47.95 ± 0.84	48.84 ± 0.59
	3000	53.78 ± 0.49	$\mathbf{56.15} \pm 0.76$	53.38 ± 0.56	53.31 ± 0.56
InceptionV3	1000	42.65 ± 0.63	$\mathbf{45.41} \pm 0.55$	43.32 ± 0.47	43.15 ± 0.58
	2000	54.63 ± 0.45	$\mathbf{55.71} \pm 0.48$	54.85 ± 0.57	53.78 ± 0.28
	3000	61.24 ± 0.37	$\mathbf{62.45} \pm 0.86$	59.72 ± 0.42	60.18 ± 0.66

Table 4. Classification accuracy comparison between RLR and other common data augmentation techniques. CNN models, i.e., MobileNet, ResNet and InceptionV3, with the data augmentation techniques, were applied on the smallest subset of CIFAR-10 (i.e., the one with 1000 samples).

Model	Baseline	RLR	RNR	Flip	Shift	Zoom	Bright	RWR	RRR
MobileNet	41.89	42.28	41.37	42.52	45.83	45.75	47.84	45.53	45.10
ResNet	39.40	41.70	40.07	40.18	41.57	40.19	39.06	41.22	41.18
InceptionV3	42.85	45.61	44.84	45.86	43.03	45.81	46.11	46.67	45.41

6.1.2. Qualitative Comparison via Saliency Maps

To further evaluate the effectiveness of the proposed RLR technique against the traditional rotation, we employed GradCAM [111], one of the well-known methods illustrating the decision made by CNNs, to show the saliency maps regarding the TR and RLR techniques.

Figure 16 shows the saliency maps of the TR and RLR techniques on images randomly selected from the test datasets. The truck image (first row in Figure 16) was classified as truck with 94% (here the percentage was the probability produced by the Softmax activation function in the CNN architectures) via MobileNet model trained with the TR augmentation technique, and nearly 100% with the proposed RLR technique. The bird image (second row in Figure 16) was classified as bird with 95% via ResNet50 model trained with the TR augmentation technique, and nearly 100% with the proposed RLR technique. The saliency maps shown in Figure 16 for the TR and RLR techniques indicated that the proposed RLR technique was, indeed, more effective in terms of assisting the CNN architectures to make decisions based on more reasonable areas within the test images.
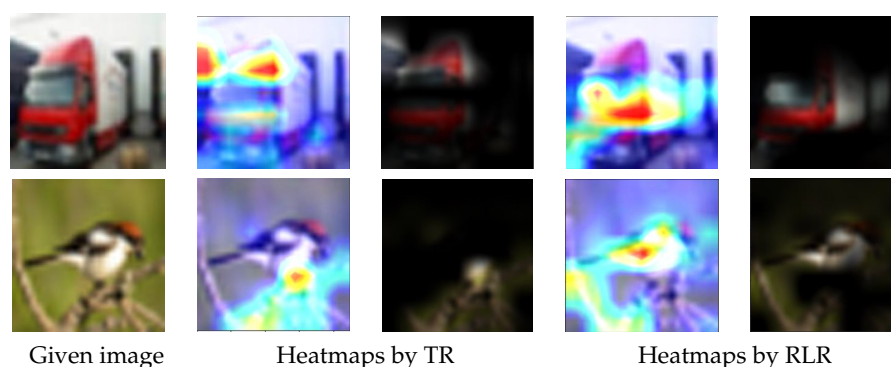

Given image   Heatmaps by TR   Heatmaps by RLR

Figure 16. Data augmentation techniques evaluation by saliency map. Column 1: given images; columns 2 and 3: two types of saliency maps for TR; columns 4 and 5: two types of saliency maps for RLR. In particular, for the two types of saliency maps evaluating each data augmentation technique, the first saliency map highlights the activated area in the given image, and the second highlights the activated area using the content of the given image. The CNNs used for the test images in the first and second rows are MobileNet and ResNet50, respectively. The saliency maps created for the models, which were trained with the dataset augmented with RLR, clearly focus on the wider part of the object while for the other cases where augmentation is achieved with TR, the models focus on a smaller area of the object. The models trained with RLR output more reliable results, together with the wider focus on the target object shown in the above saliency maps, demonstrating the superior performance of RLR compared to TR.

6.2. Segmentation Experiment

Two publicly available datasets were selected for conducting experiments regarding the segmentation task. The first dataset was the Supervisely Person [112], which contained 5711 images and 6884 high-quality annotated human instances for human semantic segmentation, see e.g., Figure 17. The second dataset was the Nuclei images dataset [113], which contained 670 microscopic images with their corresponding segmentation masks, see e.g., Figure 18. Each augmented copy was twice as large as its corresponding original dataset. Then, each dataset copy was divided into training (90% of the data) and validation (10% of the data) subsets. Note that, in this experiment, we also considered the concept of equivariance. Equivariance implies that the output changes in proportion to the input. The concept of equivariance is important in segmentation, where the location of the object and the location of the segmented object shift proportionally, e.g., see Figures 17 and 18. In contrast, invariance refers to a change in the location of an object while the output remains unchanged, which is considered in Section 6.1 for the classification task.
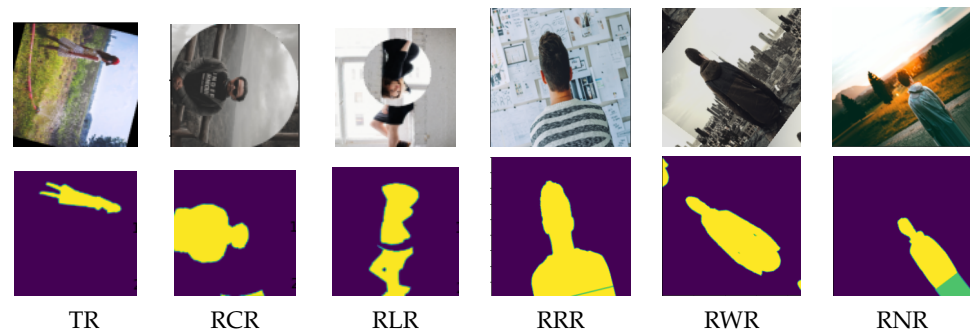
| TR | RCR | RLR | RRR | RWR | RNR |

Figure 17. Samples of the Supervisely Person dataset by applying the RLR, TR, RCR, RRR, RWR and RNR augmentation techniques. Rows one and two are the augmented samples with their corresponding human body segmentation, respectively.
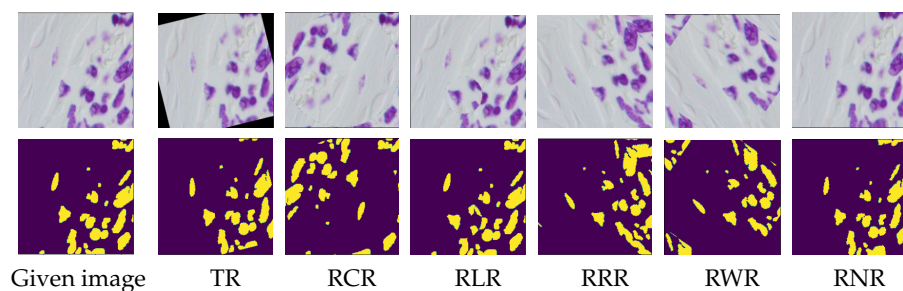

| Given image | TR | RCR | RLR | RRR | RWR | RNR |

Figure 18. Samples of the Nuclei images dataset by applying the RLR, TR, RCR, RRR, RWR and RNR augmentation techniques. Rows one and two are the augmented samples with their corresponding segmentation, respectively.

Two autoencoders were used to conduct the semantic segmentation task. These two autoencoders were constructed based on two models (i.e., MobileNet and VGG16), each with a customised decoder, see Table 5 for the detailed architectures. Each autoencoder was subjected to a total of seven tests. They were trained for 200 epochs with the Adam optimiser and binary cross entropy loss function.

Segmentation Results

Table 6 gives the segmentation accuracy of the autoencoders (i.e., MobileNet-based, and VGG16-based) with the TR, RCR, RLR, RNR, RWR and RRR data augmentation techniques on the Supervisely Person dataset, including the comparison with the baseline results (i.e., the one obtained on the original samples without using data augmentation). In contrast to the classification results, the results in Table 6 show that all the augmentation techniques tested did not improve the segmentation performance. This might imply that using rotation alone to augment data might not be a good method for the segmentation task, particularly if the shape feature was the most important one in the dataset, as in the Supervisely Person dataset. For further investigation into the influence of different features on the performance of the augmentation techniques, we conducted an experiment using the Nuclei images dataset [113].

Differing from the results obtained in Table 6 on the Supervisely Person dataset, the results in Table 7 on the Nuclei images dataset demonstrated that the rotation augmentation methods could improve the segmentation performance. This performance gain might be due to the fact that the augmentation techniques did not degrade the image qualities much in the Nuclei dataset, since the shape feature was not that critical, compared to the Supervisely Person dataset. In the Nuclei Images dataset, colours and textures are likely to be more essential than the shape features. In particular, the segmentation results in Table 7 on the Nuclei dataset showed that RLR achieved the best performance among the rotation augmentation methods. This might be due to the information preservation ability that RLR

provided, whereas RRR, RWR and RNR either lost parts of the information in the image's periphery or repeated some parts of the image, see Figure 19.
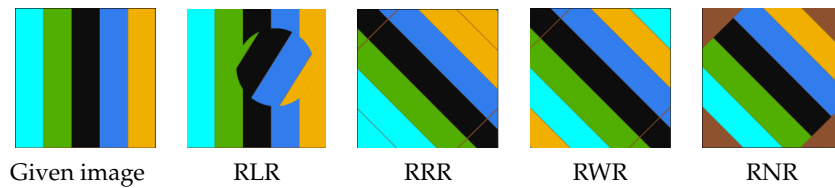

| Given image | RLR | RRR | RWR | RNR |

Figure 19. The effect of different rotation methods on the rotated image. The RRR and RWR expanded the central region (i.e., the black stripe) by repeating parts of it. RNR resulted in the loss of image content at the image's periphery and the creation of artificial pixel values to fill the gap. In contrast, RLR manipulated the content of the image while preserving the information around the image's periphery well.

Table 5. The architectures of the decoders of the MobileNet-based and VGG16-based auto-encoders. Conv2D and Conv2DT represent the 2D convolutional layer and the transposed 2D convolutional layer, respectively.

MobileNet Decoder				VGG16 Decoder			
Layer	Kernel Size	Filters Number	Activation Function	Layer	Kernel Size	Filters Number	Activation Function
Conv2DT	(3,3)	1024	Relu	Conv2DT	(3,3)	1024	Relu
Batch Normalisation				Batch Normalisation			
Conv2D	(3,3)	1024	Relu	Conv2D	(3,3)	1024	Relu
Batch Normalisation				Batch Normalisation			
Conv2DT	(3,3)	512	Relu	Conv2DT	(3,3)	512	Relu
Batch Normalisation				Batch Normalisation			
Conv2D	(3,3)	512	Relu	Conv2D	(3,3)	512	Relu
Batch Normalisation				Batch Normalisation			
Conv2DT	(3,3)	256	Relu	Conv2DT	(3,3)	256	Relu
Batch Normalisation				Batch Normalisation			
Conv2D	(3,3)	256	Relu	Conv2D	(3,3)	256	Relu
Batch Normalisation				Batch Normalisation			
Conv2DT	(3,3)	128	Relu	Conv2DT	(3,3)	128	Relu
Batch Normalisation				Batch Normalisation			
Conv2D	(3,3)	128	Relu	Conv2D	(3,3)	128	Relu
Batch Normalisation				Batch Normalisation			
Conv2DT	(3,3)	64	Relu	Conv2DT	(3,3)	64	Relu
Batch Normalisation				Batch Normalisation			
Conv2D	(3,3)	64	Relu	Conv2D	(3,3)	64	Relu
Batch Normalisation				Batch Normalisation			
Conv2DT	(3,3)	32	Relu	Conv2DT	(3,3)	32	Relu
Batch Normalisation				Batch Normalisation			
Conv2D	(3,3)	32	Relu	Conv2D	(2,2)	32	Relu
Batch Normalisation				Batch Normalisation			
Conv2D	(3,3)	1	Sigmoid	Conv2D	(3,3)	1	Sigmoid

Table 6. Segmentation accuracy comparison between different data augmentation techniques (i.e., TR, RCR, RLR, RNR, RWR and RRR). MobileNet-based and VGG16-based autoencoders were applied on the Supervisely Person dataset. The results indicated that using rotation solely to augment data might not be a good for the segmentation task in this case.

Model	Baseline	RLR	TR	RCR	RNR	RWR	RRR
MobileNet	75.13 ± 0.25	72.12 ± 0.21	73.75 ± 0.15	72.72 ± 0.18	71.12 ± 0.17	71.27 ± 0.10	71.09 ± 0.12
VGG16	75.42 ± 0.16	72.56 ± 0.19	73.16 ± 0.22	72.66 ± 0.24	71.31 ± 0.08	71.22 ± 0.14	71.06 ± 0.15

Table 7. Segmentation accuracy comparison between different data augmentation techniques (i.e., TR, RCR, RLR, RNR, RWR and RRR). MobileNet-based and VGG16-based autoencoders were applied on the Nuclei images dataset. The results indicated that using rotation to augment data could enhance the segmentation performance in this case.

Model	Baseline	RLR	TR	RCR	RNR	RWR	RRR
MobileNet	94.25 ± 0.05	97.6 ± 0.08	95.37 ± 0.11	94.50 ± 0.28	95.28 ± 0.19	95.24 ± 0.14	95.43 ± 0.13
VGG16	94.24 ± 0.12	97.81 ± 0.09	95.36 ± 0.21	94.74 ± 0.17	95.08 ± 0.18	94.91 ± 0.15	95.32 ± 0.16

6.3. Discussion

The vast majority of researchers combine many data augmentation techniques to obtain a final result. This makes it difficult to acquire an accurate evaluation for these techniques individually. In this study, we chose the random rotation technique and examined it in more detail, along with its impact on two significant tasks (i.e., classification and segmentation), in order to make a contribution to the data augmentation regime in general. Segmentation and classification are two distinct tasks. The notion that both rely on the same features to attain their desired outcomes may not be accurate. Our results in the previous section showed that the rotation augmentation techniques could enhance methods' performance for the classification task, but not the segmentation task. It was observed that the segmentation task naturally relied on shape features [114]. Geirhos et al. [56] conducted a quantitative experiment demonstrating that CNNs trained with ImageNet had a strong inclination to classify texture over shape. This feature distinction might account for the disparity between classification and segmentation results when the rotation augmentation techniques were applied. In particular, in the segmentation experiment, the RLR method distorted the shape of the human body the most, yielding a slightly poorer result than that of the TR method, which did not distort the shape of the human body. The distortion of the shape feature might explain the deterioration of the segmentation results when applying the rotation augmentation techniques. In contrast, for the classification task, the rotation augmentation techniques altered the object shape but not the overall texture, which benefited the performance enhancement for the classification task.

7. Conclusions

Deep learning models, like CNNs, are susceptible to overfitting. In this work we surveyed the data augmentation techniques, particularly recent research in image classification and segmentation employing data augmentation techniques, which are critical for deep learning models to overcome the overfitting issue and achieve better performance. In addition, we proposed a geometric augmentation technique, i.e., RLR (random local rotation), focusing on manipulating local information within images without adding non-original pixel values. Quantitative and qualitative experimental results demonstrated that RLR could be more effective than the traditional rotation technique in classification and some segmentation tasks, and, therefore, complemented the existing data augmentation techniques well.

Abbreviations

The following abbreviations were used in this manuscript:

TR	Traditional rotation
RLR	Random local rotation
RCR	Random center rotation

References

1. Yudistira, N.; Kurita, T. Gated spatio and temporal convolutional neural network for activity recognition: Towards gated multimodal deep learning. *Eurasip J. Image Video Process.* **2017**, *2017*, 1–12. [CrossRef]
2. Papakostas, M.; Giannakopoulos, T.; Makedon, F.; Karkaletsis, V. Short-term recognition of human activities using convolutional neural networks. In Proceedings of the 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Naples, Italy, 28 November–1 December 2016; pp. 302–307.
3. Milyaev, S.; Laptev, I. Towards reliable object detection in noisy images. *Pattern Recognit. Image Anal.* **2017**, *27*, 713–722. [CrossRef]
4. Zhou, X.; Gong, W.; Fu, W.; Du, F. Application of deep learning in object detection. In Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24–26 May 2017; pp. 631–634.
5. Ranjan, R.; Sankaranarayanan, S.; Bansal, A.; Bodla, N.; Chen, J.C.; Patel, V.M.; Castillo, C.D.; Chellappa, R. Deep learning for understanding faces: Machines may be just as good, or better, than humans. *IEEE Signal Process. Mag.* **2018**, *35*, 66–83. [CrossRef]
6. Druzhkov, P.; Kustikova, V. A survey of deep learning methods and software tools for image classification and object detection. *Pattern Recognit. Image Anal.* **2016**, *26*, 9–15. [CrossRef]
7. Rajnoha, M.; Burget, R.; Povoda, L. Image background noise impact on convolutional neural network training. In Proceedings of the 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Moscow, Russia, 5–9 November 2018; pp. 1–4.
8. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13001–13008. [CrossRef]
9. Joshi, S.; Verma, D.K.; Saxena, G.; Paraye, A. Issues in training a convolutional neural network model for image classification. In Proceedings of the International Conference on Advances in Computing and Data Sciences, Ghazibad, India, 12–13 April 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 282–293.
10. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48. [CrossRef]
11. Zoph, B.; Cubuk, E.D.; Ghiasi, G.; Lin, T.Y.; Shlens, J.; Le, Q.V. Learning data augmentation strategies for object detection. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 566–583.
12. Brownlee, J. Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions. Machine Learning Mastery. 2018. Available online: https://machinelearningmastery.com/better-deep-learning/ (accessed on 15 December 2022).
13. Guo, Y.; Zhang, J.; Cai, J.; Jiang, B.; Zheng, J. Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1294–1307. [CrossRef]
14. Wan, L.; Zeiler, M.; Zhang, S.; Le Cun, Y.; Fergus, R. Regularization of neural networks using dropconnect. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1058–1066.
15. Kang, G.; Dong, X.; Zheng, L.; Yang, Y. Patchshuffle regularization. *arXiv* **2017**, arXiv:1707.07103
16. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
17. Farebrother, R. Further results on the mean square error of ridge regression. *J. R. Stat. Soc. Ser. (Methodol.)* **1976**, *38*, 248–250. [CrossRef]

18. Ranstam, J.; Cook, J. LASSO regression. *J. Br. Surg.* **2018**, *105*, 1348. [CrossRef]

19. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

20. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]

21. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556

22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

23. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

24. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.

25. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

26. Yang, S.; Xiao, W.; Zhang, M.; Guo, S.; Zhao, J.; Shen, F. Image Data Augmentation for Deep Learning: A Survey. *arXiv* **2022**, arXiv:2204.08610

27. Divon, G.; Tal, A. Viewpoint Estimation—Insights & Model. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 252–268.

28. Ning, X.; Nan, F.; Xu, S.; Yu, L.; Zhang, L. Multi-view frontal face image generation: A survey. *Concurr. Comput. Pract. Exp.* **2020**, e6147. [CrossRef]

29. Massa, F.; Marlet, R.; Aubry, M. Crafting a multi-task CNN for viewpoint estimation. *arXiv* **2016**, arXiv:1609.03894

30. Liu, H.; Wang, C.; Peng, Y. Data Augmentation with Illumination Correction in Sematic Segmentation. In Proceedings of the 2021 3rd International Conference on Artificial Intelligence and Computer Science (AICS) 2021, Beijing, China, 29–31 July 2021; IOP Publishing: Bristol, UK, 2021; Volume 2025, p. 012009.

31. Chlap, P.; Min, H.; Vandenberg, N.; Dowling, J.; Holloway, L.; Haworth, A. A review of medical image data augmentation techniques for deep learning applications. *J. Med. Imaging Radiat. Oncol.* **2021**, *65*, 545–563. [CrossRef]

32. Lindner, L.; Narnhofer, D.; Weber, M.; Gsaxner, C.; Kolodziej, M.; Egger, J. Using synthetic training data for deep learning-based GBM segmentation. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 6724–6729.

33. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Virtual, 14–19 June 2020; pp. 702–703.

34. Shijie, J.; Ping, W.; Peiyi, J.; Siping, H. Research on data augmentation for image classification based on convolution neural networks. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 4165–4170.

35. Lee, H.; Hwang, S.J.; Shin, J. Rethinking data augmentation: Self-supervision and self-distillation. *arXiv* **2019**, arXiv:1910.05872;

36. Denton, E.; Hanna, A.; Amironesei, R.; Smart, A.; Nicole, H. On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data Soc.* **2021**, *8*, 20539517211035955. [CrossRef]

37. Lu, W.; Xing, X.; Cai, B.; Xu, X. Listwise view ranking for image cropping. *IEEE Access* **2019**, *7*, 91904–91911. [CrossRef]

38. Bagherinezhad, H.; Horton, M.; Rastegari, M.; Farhadi, A. Label refinery: Improving imagenet classification through label progression. *arXiv* **2018**, arXiv:1805.02641

39. Masi, I.; Tran, A.T.; Hassner, T.; Leksut, J.T.; Medioni, G. Do we really need to collect millions of faces for effective face recognition? In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 579–596.

40. Ibraheem, N.A.; Hasan, M.M.; Khan, R.Z.; Mishra, P.K. Understanding color models: A review. *ARPN J. Sci. Technol.* **2012**, *2*, 265–275.

41. Cai, X.; Chan, R.; Nikolova, M.; Zeng, T. A three-stage approach for segmenting degraded color images: Smoothing, lifting and thresholding (SLaT). *J. Sci. Comput.* **2017**, *72*, 1313–1332. [CrossRef]

42. Wu, R.; Yan, S.; Shan, Y.; Dang, Q.; Sun, G. Deep image: Scaling up image recognition. *arXiv* **2015**, arXiv:1501.02876.

43. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 806–813.

44. Khalifa, N.E.; Loey, M.; Mirjalili, S. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artif. Intell. Rev.* **2021**, *55*, 2351–2377 . [CrossRef] [PubMed]

45. Kanopoulos, Nick and Vasanthavada, Nagesh and Baker, Robert L Design of an image edge detection filter using the Sobel operator. *IEEE J. Solid-State Circuits* **1988**, *23*, 358–367. [CrossRef]

46. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *6*, 679–698. [CrossRef]

47. Taylor, L.; Nitschke, G. Improving deep learning with generic data augmentation. In Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bengaluru, India, 18–21 November 2018; pp. 1542–1547.

48. Guo, G.; Lai, A. A survey on still image based human action recognition. *Pattern Recognit.* **2014**, *47*, 3343–3361. [CrossRef]

49. Wu, B.; Yuan, C.; Hu, W. Human action recognition based on context-dependent graph kernels. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2609–2616.
50. Ravishankar, A.; Anusha, S.; Akshatha, H.; Raj, A.; Jahnavi, S.; Madhura, J. A survey on noise reduction techniques in medical images. In Proceedings of the 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 20–22 April 2017; Volume 1, pp. 385–389.
51. Nazaré, T.S.; Costa, G.B.; Contato, W.A.; Ponti, M. Deep convolutional neural networks and noisy images. In Proceedings of the Iberoamerican Congress on Pattern Recognition, Valparaíso, Chile, 7–10 November 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 416–424.
52. Boonprong, S.; Cao, C.; Chen, W.; Ni, X.; Xu, M.; Acharya, B.K. The classification of noise-afflicted remotely sensed data using three machine-learning techniques: Effect of different levels and types of noise on accuracy. *Isprs Int. J.-Geo-Inf.* **2018**, *7*, 274. [CrossRef]
53. Boyat, A.K.; Joshi, B.K. A review paper: Noise models in digital image processing. *arXiv* **2015**, arXiv:1505.03489
54. Chen, S.; Yang, X.; Cao, G. Impulse noise suppression with an augmentation of ordered difference noise detector and an adaptive variational method. *Pattern Recognit. Lett.* **2009**, *30*, 460–467. [CrossRef]
55. Efros, A.A.; Freeman, W.T. Image quilting for texture synthesis and transfer. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, New York, NY, USA, 12–17 August 2001; pp. 341–346.
56. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.A.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv* **2018**, arXiv:1811.12231
57. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
58. Hertzmann, A.; Jacobs, C.E.; Oliver, N.; Curless, B.; Salesin, D.H. Image analogies. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, New York, NY, USA, 12–17 August 2001; pp. 327–340.
59. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujscie, Poland, 9–12 May 2018; pp. 117–122.
60. Su, J.; Vargas, D.V.; Sakurai, K. One pixel attack for fooling deep neural networks. *IEEE Trans. Evol. Comput.* **2019**, *23*, 828–841. [CrossRef]
61. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. *arXiv* **2014**, arXiv:1412.6572
62. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 139–144. .
63. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434
64. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196
65. Zhu, X.; Liu, Y.; Li, J.; Wan, T.; Qin, Z. Emotion classification with data augmentation using generative adversarial networks. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Melbourne, VIC, Australia, 3–6 June 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 349–360.
66. Antoniou, A.; Storkey, A.; Edwards, H. Data augmentation generative adversarial networks. *arXiv* **2017**, arXiv:1711.04340
67. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621
68. Hussain, Z.; Gimenez, F.; Yi, D.; Rubin, D. Differential data augmentation techniques for medical imaging classification tasks. In Proceedings of the AMIA Annual Symposium Proceedings, Washington, DC, USA, 4–8 November 2017; Volume 2017, p. 979.
69. Pawara, P.; Okafor, E.; Schomaker, L.; Wiering, M. Data augmentation for plant classification. In Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems, Antwerp, Belgium, 18–21 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 615–626.
70. Inoue, H. Data augmentation by pairing samples for images classification. *arXiv* **2018**, arXiv:1801.02929
71. Li, W.; Chen, C.; Zhang, M.; Li, H.; Du, Q. Data augmentation for hyperspectral image classification with deep CNN. *IEEE Geosci. Remote. Sens. Lett.* **2018**, *16*, 593–597. [CrossRef]
72. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **2018**, *321*, 321–331. [CrossRef]
73. Pham, T.C.; Luong, C.M.; Visani, M.; Hoang, V.D. Deep CNN and data augmentation for skin lesion classification. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Dong Hoi City, Vietnam, 19–21 March 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 573–582.
74. Motlagh, M.H.; Jannesari, M.; Aboulkheyr, H.; Khosravi, P.; Elemento, O.; Totonchi, M.; Hajirasouliha, I. Breast cancer histopathological image classification: A deep learning approach. *bioRxiv* **2018**. [CrossRef]
75. Zheng, X.; Chalasani, T.; Ghosal, K.; Lutz, S.; Smolic, A. Stada: Style transfer as data augmentation. *arXiv* **2019**, arXiv:1909.01056
76. Ismael, S.A.A.; Mohammed, A.; Hefny, H. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif. Intell. Med.* **2020**, *102*, 101779. [CrossRef] [PubMed]
77. Gour, M.; Jain, S.; Sunil Kumar, T. Residual learning based CNN for breast cancer histopathological image classification. *Int. J. Imaging Syst. Technol.* **2020**, *30*, 621–635. [CrossRef]

78. Nanni, L.; Paci, M.; Brahnam, S.; Lumini, A. Comparison of Different Image Data Augmentation Approaches. *J. Imaging* **2021**, *7*, 254. [CrossRef]

79. Anwar, T.; Zakir, S. Effect of image augmentation on ECG image classification using deep learning. In Proceedings of the 2021 International Conference on Artificial Intelligence (ICAI), Lucknow, India, 22–23 May 2021; pp. 182–186.

80. Kandel, I.; Castelli, M. Improving convolutional neural networks performance for image classification using test time augmentation: A case study using MURA dataset. *Health Inf. Sci. Syst.* **2021**, *9*, 1–22. [CrossRef]

81. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.

82. Benson, E.; Pound, M.P.; French, A.P.; Jackson, A.S.; Pridmore, T.P. Deep hourglass for brain tumor segmentation. In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 419–428.

83. Newell, A.; Yang, K.; Deng, J. Stacked hourglass networks for human pose estimation. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 483–499.

84. Menze, B.; Jakab, A.S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; Leemput, K.V. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [CrossRef]

85. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 1–13. [CrossRef]

86. Simpson, A.L.; Antonelli, M.; Bakas, S.; Bilello, M.; Farahani, K.; Van Ginneken, B.; Kopp-Schneider, A.; Landman, B.A.; Litjens, G.; Menze, B.; et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv* **2019**, arXiv:1902.09063.

87. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M.; et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv* **2018**, arXiv:1811.02629.

88. Casado-García, Á.; Domínguez, C.; García-Domínguez, M.; Heras, J.; Inés, A.; Mata, E.; Pascual, V. CLoDSA: A tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks. *BMC Bioinform.* **2019**, *20*, 323. [CrossRef]

89. Arganda-Carreras, I.; Turaga, S.C.; Berger, D.R.; Cireşan, D.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J.; Laptev, D.; Dwivedi, S.; Buhmann, J.M.; et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Front. Neuroanat.* **2015**, *9*, 142. [CrossRef]

90. Ma, R.; Tao, P.; Tang, H. Optimizing data augmentation for semantic segmentation on small-scale dataset. In Proceedings of the 2nd International Conference on Control and Computer Vision, Jeju Island, Republic of Korea, 15–18 June 2019; pp. 77–81.

91. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.

92. Qiao, Y.; Su, D.; Kong, H.; Sukkarieh, S.; Lomax, S.; Clark, C. Data augmentation for deep learning based cattle segmentation in precision livestock farming. In Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Virtual, 20–21 August 2020; pp. 979–984.

93. Milioto, A.; Stachniss, C. Bonnet: An open-source training and deployment framework for semantic segmentation in robotics using cnns. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 7094–7100.

94. Khryashchev, V.; Larionov, R. Wildfire segmentation on satellite images using deep learning. In Proceedings of the 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), Moscow, Russia, 11–13 March 2020; pp. 1–5.

95. Chen, H.; Jung, S.T. Enhancement of Tongue Segmentation by Using Data Augmentation. *J. Korea Inst. Information, Electron. Commun. Technol.* **2020**, *13*, 313–322.

96. Qin, T.; Wang, Z.; He, K.; Shi, Y.; Gao, Y.; Shen, D. Automatic data augmentation via deep reinforcement learning for effective kidney tumor segmentation. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Virtual, 4–8 May 2020; pp. 1419–1423.

97. Cirillo, M.D.; Abramian, D.; Eklund, A. What is the best data augmentation for 3D brain tumor segmentation? In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 36–40.

98. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.

99. Su, D.; Kong, H.; Qiao, Y.; Sukkarieh, S. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Comput. Electron. Agric.* **2021**, *190*, 106418. [CrossRef]

100. Zhang, J.; Zhang, Y.; Xu, X. Objectaug: Object-level data augmentation for semantic image segmentation. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8.

101. Mallios, D.; Cai, X. Deep Rectum Segmentation for Image Guided Radiation Therapy with Synthetic Data. In Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO), Virtual, 23–27 August 2021; pp. 975–979.

102. Burnet, N.G.; Scaife, J.E.; Romanchikova, M.; Thomas, S.J.; Bates, A.M.; Wong, E.; Noble, D.J.; Shelley, L.E.; Bond, S.J.; Forman, J.R.; et al. Applying physical science techniques and CERN technology to an unsolved problem in radiation treatment for cancer: The multidisciplinary 'VoxTox'research programme. *Cern Ideasquare J. Exp. Innov.* **2017**, *1*, 3.
103. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef]
104. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.; Freymann, J.; Farahani, K.; Davatzikos, C. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *Cancer Imaging Arch.* **2017**, *286*. [CrossRef]
105. Lyksborg, M.; Puonti, O.; Agn, M.; Larsen, R. An ensemble of 2D convolutional neural networks for tumor segmentation. In Proceedings of the Scandinavian Conference on Image Analysis, Copenhagen, Denmark, 15–17 June 2015; pp. 201–211.
106. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.M.; Larochelle, H. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **2017**, *35*, 18–31. [CrossRef] [PubMed]
107. Delaney, G.; Jacob, S.; Featherstone, C.; Barton, M. The role of radiotherapy in cancer treatment: Estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer Interdiscip. Int. J. Am. Cancer Soc.* **2005**, *104*, 1129–1137. [CrossRef]
108. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
109. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
110. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861
111. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
112. Supervise.ly. Supervisely Person Dataset. 2018. Available online: https://supervise.ly (accessed on 1 February 2022).
113. Hamilton, B.A. Find the Nuclei in Divergent Images to Advance Medical Discovery. 2018. Available online: https://www.kaggle.com/c/data-science-bowl-2018/data (accessed on 10 February 2023).
114. Bajcsy, R.; Solina, F.; Gupta, A. Segmentation versus object representation—Are they separable? In *Analysis and Interpretation of Range Images*; Springer: Berlin/Heidelberg, Germany, 1990; pp. 207–223.