2020 International Conference on Intelligent Engineering and Management (ICIEM)

Enhancing Performance of Deep Learning Models with different Data Augmentation Techniques: A Survey

*

Cherry Khosla
School of Computer Science and Engineering
Lovely Professional University
Phagwara, India
cherry.khosla86@gmail.com

Baljit Singh Saini
School of Computer Science and Engineering
Lovely Professional University
Phagwara, India
baljitsaini28@gmail.com

***Abstract*—Deep convolutional neural networks have shown impressive results on different computer vision tasks. Nowadays machines are fed by new artificial intelligence techniques which makes them intelligent enough to cognize the visual world better than humans. These computer vision algorithms rely heavily on large data sets. Having a large training data set plays a very crucial role in the performance of deep convolutional neural networks. We can enhance the performance of the model by augmenting the data of the image. Data augmentation is a set of techniques that are used to increase the size and quality of the image with label preserving transformations. This survey paper focuses on different data augmentation techniques based on data warping and oversampling. In addition to data augmentation techniques, this paper gives a brief discussion on different solutions of reducing the overfitting problem.**

***Index Terms*—data augmentation, deep learning models, over-fitting, data warping, synthetic oversampling**

I. Introduction

The machines nowadays are fed by new Artificial Intelligence techniques which makes them intelligent enough to apprehend and cognize the visual world better than humans. This leads to the field of computer vision which is taking a great leap forward today. The computer vision aspired the researchers to design algorithms for such visual perceptions [1]. These visual perceptions include object recognition for identifying a specific object in the given image data, object detection by doing the semantic analysis and labelling it under a class, understanding the scene to parse the image into a meaningful segment [2]. Deep convolutional neural networks have done exceptionally well in the field of computer vision. These neural networks are showing incredible results in computer vision tasks. The convolutional neural networks CNN preserves the features of the image data when complex images are taken into consideration by maintaining their spatial and temporal dependencies. However, when it comes to big data, many of the applications does not have much access to data like medical images [3]. Having large training data set plays a very crucial role in the performance of deep convolutional neural networks. We can enhance the performance of the model by augmenting the data of the image. Data augmentation is a set of techniques that are used to increase the size and quality of the image with label preserving transformations. It is the method of creating new data by having different data orientations as shown in Fig. 1. Data augmentation helps the researchers in two ways: first by generating more data from limited amount of data, secondly it restrains over fitting. It artificially generates new data by introducing new data samples [4] [5]. To build a better learning model, the validation error rate should decrease correspondingly with training error. Data augmentation plays a crucial role in achieving this by inflating data set thereby decreasing the distance among training, validation and test data sets.

In this paper we look into different image data augmentation techniques. We will discuss different technqiues based on two broad categories: Data Warping and Oversampling. The classification of different augmentation techniques based on data warping and oversampling is shown in Fig. 2.The paper also provides brief discussion on other functional solutions for increasing the size of the small data sets like transfer learning, pre-training, dropout regularization and batch normalization. Apart from data augmentation techniques there are many other techniques which are used for increasing the performance of the architectures relies on the model of the architecture. This has led to the development of complicated architectures like AlexNet [6], VGG-16 [7], ResNet [8], Inception-V3 [9] and DenseNet [10].

However, along with the techniques mentioned above, data augmentation inflates the original training dataset size by either oversampling or data warping [11].Different image datasets are used by many researchers on effectiveness of image data augmentation to benchmark the results. Some of the these datasets include: MNIST, CIFAR-10, CIFAR-100, ImageNet, SVHN, Caltech101, MIT places, MIT-Adobe 5K dataset etc. Commonly used datasets are CIFAR-10, CIFAR-

978-1-7281-4097-1/20/$31.00 ©2020 IEEE

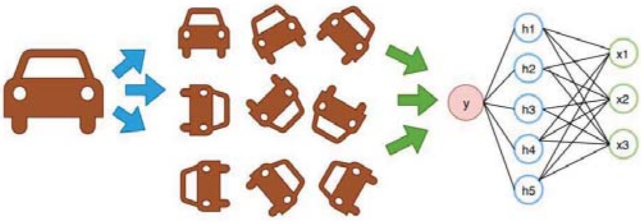

Fig. 1. Data Augmentation into picture

100 and ImageNet.

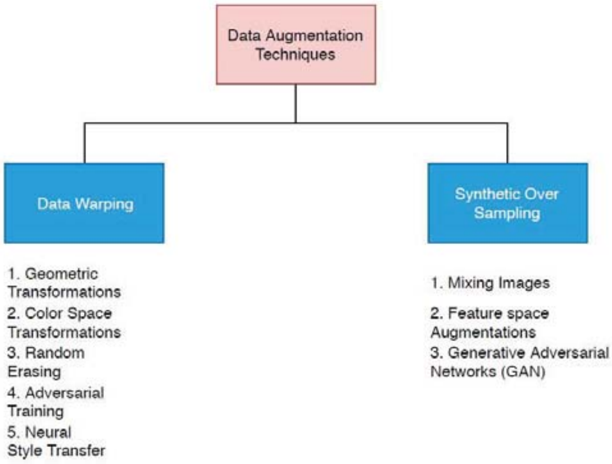

Fig. 2. Classification of Data Augmentation Techniques

Label Preserving Transformations

Before any image data augmentation technique is applied, it is better to endow some time in selecting an appropriate set of transformations. Some of these set of transformations also have parameters which can be tuned, and decent set of values for these parameters must be settled. It depends on the dataset that what is included as "decent". While augmenting the image, one must make sure that its meaning(label) must be preserved. It is called as label-preserving.

The rest of the paper is organized as follows: In Section II we have described brief Related Work on data augmentation techniues. Different Data Augmentation Techniques based on data warping and oversampling are discussed in Section III. Section IV compares different data augmention technqiues and then paper concluds with Conclusion and Future work in Section IV

II. Related Work

Many researchers have taken steps in this field and have identified various techniques for increasing the data sample size and increasing the generalization of the neural network used. The authors in the [12] have discussed and compared the different augmentation techniques over AlexNet model of CNN architecture. The dataset used by the authors were ImageNet and CIFAR10. The authors have compared the performance of different augmentation techniques like Flipping, Rotation, Noise, Shifting, Cropping, PCA jittering GAN and WGAN. As per authors, only rotations and WGAN have shown better results than others.The survey was conducted by [2] for semantic segmentation of image and video using deep learning techniques. The authors have focused on focused on the review of deep learning methods for semantic segmentation which are applied to various applications. Sematic segmentation in computer vision is one of the key problems in 2D images, videos, or even 3D images and volumetric data. The problem of semantic segmentation is very well defined by the authors and reviewed on various deep learning models using different datasets. An augmentation strategy with Perlin noise for image classification was presented by [13].This augmentation is applied to pixel-by-pixel and different image patterns. They have used images of 106 patients. They have considered 100 regions of interest for each of class of image patterns were selected for deep learning classification. The main problem faced by many researchers in the field of deep learning to maximize its generalization is the availability of large dataset. There are so many techniques used to increase the size of data set like image augmentation, dropout, transfer learning etc.In [14] the authors have presented new method called smart augmentation to increase the precision and to solve the problem of overfitting of the target network. The smart augmentation creates a network which automatically to generate the augmented data during training process thereby reducing network loss.A method is proposed by [15] used generative adversarial network to train the synthetic MRI images with brain tumors. They have used two public data sets of brain MRI. They have presented two benefits of using this synthetic data: 1) efficient performance of tumor segmentation by resisting the synthetic images as a form if image augmentation; 2) they have demonstrated the generative models as a anonymous based tool.

III. Data Augmentation Techniques

Despite availability of data, fetching the correct information which best possible matches with our research and experiment is a challenging task. In addition, the data should be diverse enough to be presented in varying sizes, poses, colors, lighting condition for better performance of the model working over this data. To overcome this problem of limited amount of data, data augmentation techniques are applied. These techniques will help to generate our own data from the existing data [16]. In follwing section different image data augmentation techniques are covered that are categorised under data warping and oversampling.

A. *Data Augmentation based on Data Warping*

Geometric Transformations

The following section describes different image augmentation solutions on the basis of geometric transformations along with other image processing techniques. The image is transformed on the basis of its similarity, Euclidean, affine, projectile, etc [17]. Different techniques discussed in this section are: Flipping, Color Space, Cropping, Rotation, Translation, and Noise injection as shown in Fig 3. Different geometric augmentations are also described in the context of its probability of preserving the label after transformation.

Flipping

The image can be flipped either horizontally or vertically. It produces images by rotating the image at the multiple of 90 degrees. Though there are frameworks which do not support vertical flipping. Vertical flipping can be done by rotating the image by 180 degrees and then perform a horizontal flip [12]. The flipping of image is shown in Fig. 3(b). This augmentation technique is label preserving on datasets like CIFAR-10 and Imagenet. However the datasets involving recognition of digits, it is not prooved as a label preserving transformation.

Color Space

The color space transformations are also referred to as photo metric transformations. In this technique, 3 stacked matrix of the image is created where each matrix is of size height X width. The matrix represent the pixel values for the individual RGB color value. The color distributions of the image can be altered to overcome the problem of lighting challenges [3].

Cropping

Randomly sampling a particular section of the original image is called as cropping. The section is then resized to the size of the original image. It is a part of the original image and then resizing the image to some scale if necessary. This method is also called as random cropping. The random cropping and translation are different in the sense that cropping will decrease the size of the image while translations preserve its spacial dimensions [12] [18]. The Fig. 3(c) shows how cropping is applied to the image.

Rotation

Depending upon the need, the image can be rotated either at the 90 degrees angles or it can orient at minute angles. When the image is rotated at 90 degrees angles, no background noise is added in the image after orientation. But this is not true in case when rotation is done at minute angles. Also, if the background of the image is of black or white color then newly added color in the form of noise will blend in the image. But if there are different colors present in the background of the image then it will not blend and the network will learn it in the form of a feature of the image [3] [12]. The rotation degree parameter will determine the safety of the augmentations. It is useful in digit recognition for slight rotations. When we increase the degree of the rotation, the label of the data is not preserved after transformation. Fig. 3(d) shows how rotation is applied to an image.

Translation

In order to identify the object in any part of the image, then translation concept is applied to image. It involves moving of the image along the X or Y direction or both. Left, right, up and down shifting of images is useful in avoiding positional bias in the data. This augmentation helps the network to look everywhere inside the image which may result in the addition of background noise in the image [16].

Noise Injection

Noise injection is an important technique of data augmentation as it helps in preventing overfitting of the neural network model. Generally, salt and pepper noise are added to the image. This refers to the addition of white and black dots in the image as shown in Fig. 3(e). Addition of noise to the image can help the model to learn properly [19].

Geometric transformations are considered to be a good solution for images in training data where there are positional biases. However there are some disadvantages of the using geometric transformations which may include additional memory and training time, transformation cost etc. Scope of usage of geometric transformations is very less.

Color Space Transformations

This technique involves transformation of the first image with first multi-dimension color space RGB to second image with second multi-dimensional color space CMYK. The image is transformed into 3 stacked matrices. The size of each matrix is represented as height X width. The matrices demonstrate the pixel value for an individual RGB color value. The color distributions of the image can be altered to overcome the problem of lighting challenges. The color space transformations can also be attained by using different image editing applications. A histogram is formed by aggregating the RGB color channel of an image's pixel value. Then this histogram is manipulated by applying filters which helps in changing the color space characteristics of an image [3] [20]. Just like printers, images in computer systems are represented in color spaces like RGB and CMYK. Just like geometric transformations, color transformations increases the memory space, transformation cost and training time. Moreover, the color transformations can result in lossy transformations. There are different color tranformations that can be performed on an image like RGB, LAB, YCrCb, HSV etc, some of which are shown in Fig. 4 where Fig. 4(b) shows LAB color transformation, Fig. 4(c) depicts YCrCb transformation and Fig. 4(d) shows HSV transformation.

Random Erasing

Occlusion is one of the factors which is critically influencing the generalization ability of Convolutional Neural Networks. In the cases where, where training samples are clearly visible, i.e. no occlusion is present in the images, the CNN model works well on the testing images. However, it might fail to recognize the object consisting of partial occlusion in them. To solve this problem, and to improve the generalization

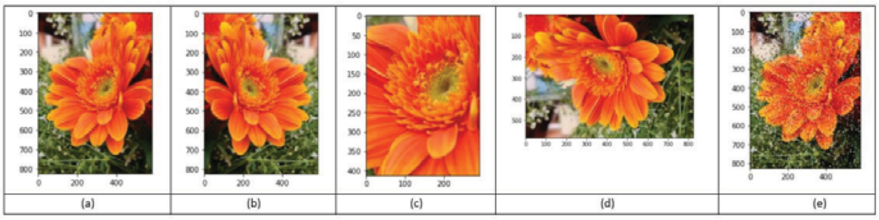

Fig. 3. (a) Original Image (b) Flipping (c) Cropping (d) Rotation (e) Noise Injection

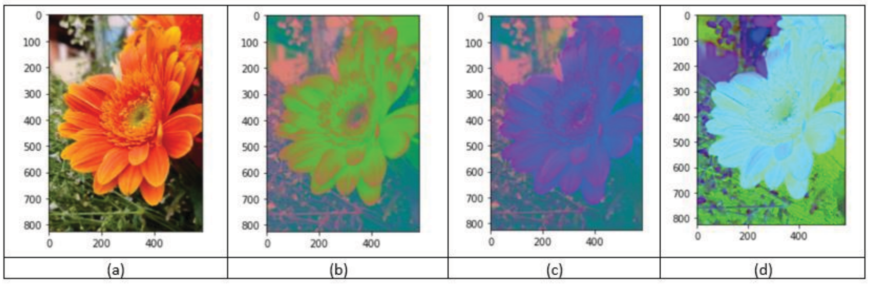

Fig. 4. Color Space Transformations (a) Original Image (b) LAB color Transformation (c) YCrCb Transformation (d) HSV Transformation

ability of the CNN model, a new data augmentation technique is presented called as random erasing [21]. In random erasing the image is occluded in a random position and with a arbitrary-sized patch or a rectangular-region.

Adversarial Training

It is a technique where we train the neural networks on the basis of adversarial examples. The adversarial training is an algorithm that perceive a highly resembled image to cheat the classifier. The researchers have found that the machine learning models can be easily mis leaded by slightly perturbated images [22]. These perturbations are determined when we optimize the input. This will maximize the prediction error. We call these perturbated examples as "adversarial examples".

Neural Style Transfer

It is ability of the neural network to transform images in style of another image thereby forming an aesthetic imagery by separating and recombining image content and style [23]. It is an image optimization technique which takes three images, first the content image, second the type of style that we want to incorporate as an image and the input image. Then these three images are combine together in such a way that the input image is modified so that it looks like the content image, but "painted" in the style of the style image [24]

B. Data Augmentation based on Oversampling

Mixing Images

Mixing images also called as Sample Pairing, in which a new sample is synthesized by overlaying one image with another image which is randomly chosen from the training data. The resultant image after mixing is used to train a classification model. Using these two randomly selected images from the training set, we can generate N^2 new samples from N training samples [25]. This technique is further improved by [26] by creating more generalized form of mixing images. They have used a non-linear way to integrate images into new training instances. Mixing of images can also be done through random cropping and patching.

Feature-Space Augmentations

The neural networks are remarkably good at mapping images which are high-dimensional in nature into their low-dimensional representations. The mapping of the images by networks is done either to binary classes or to $n *$ 1 vectors. These lower-level representations in the higher-level layers are called as feature space. As per DeVries and Taylor,

noise, interpolation, and extrapolation are types of feature space augmentations [27].

Generative Adversarial Network

Another form of data augmentation technique is Generative Adversarial Network (GAN). Generative adversarial networks are used for developing artificial samples from the dataset and these samples have similar characteristics as that of original set. [28] described GANs as the way which unlocks the information from the original dataset by generating synthetic samples. This is an extension to adversarial training networks. They belong to the set of generative models i.e. they have the capability to generate or produce the new data. There are two models used in GAN; Generator model and Discriminator model. The generator model will generate new samples and the discriminator model will take the input from the domain(real or generated) and predicts the binary class labels generated (real or fake).

C. Other Overfitting Solutions

Apart from data augmentation, there are other solutions as well which are used to solve the problem of overfitting. These solutions also increase the generalization performance of the model. Some of these solutions to overfitting are discussed in this section.

Transfer Learning

Conventional machine learning and data mining techniques make the predictions on the data the are trained on collected labelled and unlabelled data. The training and test data have same distributions and feature space. However, with data deluge from wide variety of sources, we find heterogeneity in the collected data. The concept of transfer learning is proposed to tackle this issue. Transfer Learning makes use of knowledge acquired while solving one problem to solve another similar kind of problem. It is considered as an important paradigm to reduce the problem of overfitting.A model is trained on a large data suing some weights. In Transfer learning these weights are resued as initial weights for trainig new model.

Drop-out

Dropout is a regularization technique in which the hidden units are dropped randomly during training time. It is considered as solution to reduce overfitting by using a single model to simulate different network architechtures by randomly removing some features at each iteration during training process .By doing this, the network will become more robust by learning feature [29].

Batch Normalization

Batch normalization is also a regularization technique. It is used to normalize the activations in a layer. The batch mean is subtarcted from the each activation and then dividing batch standard deviation.it standardizes the inputs given to the layer for each batch. This technique is in addition with standardization is also used preprocessing the pixel values.

Pretraining

Pretraining is similar to transfer learning. In transfer learning we tranfer both weights as well as the network, however in pretraining, it allows the initialization of weights over large datasets.

IV. Comparison

As far as comparison of different augmentations is concerned, there are not many comparative studies which compares these image augmentation techniques on different parameters. The study of comparison depends on data sizes from small, medium to large data sets. Also it depends on different levels of augmentations applied. One study of comparison was conducted by [12] which compared GANs, WGANs, cropping, rotation, flipping, translations, jittering (PCA and color), adding noise on CIFAR-10 and ImageNet datasets. They found that flipping, cropping, rotation and WGAN performed better than others. Another comparison done by [30] used ResNet model over a subset of CIFAR-10 dataset and compared six different data augmentation techniques: skew, shear, random erasing, gaussian distortion and random distortion. They found that careful augmentation can improve the accuracy by +2.83 percent to 95.85 percent. Also their results indictated that injecting augmentation after initial learning phase is more effective than applying standard augmentation. Moreover the preformance of different data augmentation techniques can be measured with respect to different design decisions. These de- sign considerations include test-time augmentation, curriculum learning, reolution impact and final data set size.

V.Conclusion and Future Work

This paper presents the survey on the different image data augmentation solutions which are used to reduce the problem of overfitting in deep learning models. To give better results, deep learning models rely on large data sets.Different data augmentation techniques are discussed under data warping and oversampling. These solution of increasing data set size increases the generalization degree of the learning model

.Data augmentation is one of the prominent ways to increase the accuracy of classification tasks.The survey paper mainly discuss the techniques for image data augmentation. However many of these solutions can also be applied to other domains like videos data augmentation domains.

I. Conclusion and Future Work

This paper presents the survey on the different image data augmentation solutions which are used to reduce the problem of overfitting in deep learning models. To give better results, deep learning models rely on large data sets.Different data augmentation techniques are discussed under data warping and oversampling. These solution of increasing data set size increases the generalization degree of the learning model .Data augmentation is one of the prominent ways to increase the accuracy of classification tasks.The survey paper mainly discuss the techniques for image data augmentation. However many of these solutions can also be applied to other domains like videos data augmentation domains.

References

[1] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, “Computer vision algorithms and hardware implementations: A survey,” *Integration*, 2019.

[2] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, “A survey on deep learning techniques for image and video semantic segmentation,” *Applied Soft Computing*, vol. 70, pp. 41–65, 2018.

[3] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[4] M. D. Bloice, C. Stocker, and A. Holzinger, “Augmentor: an image augmentation library for machine learning,” *arXiv preprint arXiv:1708.04680*, 2017.

[5] L. Perez and J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” *arXiv preprint arXiv:1712.04621*, 2017.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[11] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, “Understanding data augmentation for classification: when to warp?” in *2016 international conference on digital image computing: techniques and applications (DICTA)*. IEEE, 2016, pp. 1–6.

[12] J. Shijie, W. Ping, J. Peiyi, and H. Siping, “Research on data augmentation for image classification based on convolution neural networks,” in *2017 Chinese automation congress (CAC)*. IEEE, 2017, pp. 4165–4170.

[13] H.-J. Bae, C.-W. Kim, N. Kim, B. Park, N. Kim, J. B. Seo, and S. M. Lee, "A perlin noise-based augmentation strategy for deep learning with small data samples of hrct images," *Scientific reports*, vol. 8, no. 1, p. 17687, 2018.

[14] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart augmentation learning an optimal data augmentation strategy," *Ieee Access*, vol. 5, pp. 5858–5869, 2017.

[15] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2018, pp. 1–11.

[16] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, 2018, pp. 117–122.

[17] M. Paschali, W. Simson, A. G. Roy, R. Göbl, C. Wachinger, and N. Navab, "Manifold exploring data augmentation with geometric transformations for increased performance and robustness," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 517–529.

[18] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep cnns," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[19] H. Noh, T. You, J. Mun, and B. Han, "Regularizing deep neural networks by noise: its interpretation and optimization," in *Advances in Neural Information Processing Systems*, 2017, pp. 5109–5118.

[20] B. Delean, "Method for performing a color space transformation," Oct. 9 2001, uS Patent 6,301,025.

[21] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," *arXiv preprint arXiv:1708.04896*, 2017.

[22] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[23] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015.

[24] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," *IEEE transactions on visualization and computer graphics*, 2019.

[25] H. Inoue, "Data augmentation by pairing samples for images classification," *arXiv preprint arXiv:1801.02929*, 2018.

[26] C. Summers and M. J. Dinneen, "Improved mixed-example data augmentation," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1262–1270.

[27] T. DeVries and G. W. Taylor, "Dataset augmentation in feature space," *arXiv preprint arXiv:1702.05538*, 2017.

[28] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, and D. Rueckert, "Gan augmentation: augmenting training data using generative adversarial networks," *arXiv preprint arXiv:1810.10863*, 2018.

[29] S. Wager, S. Wang, and P. S. Liang, "Dropout training as adaptive regularization," in *Advances in neural information processing systems*, 2013, pp. 351–359.

[30] K. McGuinness and S. O'Gara, "Comparing data augmentation strategies for deep image classification," 2019.