International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Empirical Evaluation of the Effectiveness of Variational Autoencoders on Data Augmentation for the Image Classification Problem

Ozge Oztimur Karadag*[1], Ozlem Erdas Cicek[2]

Submitted: 01/04/2020 Accepted: 01/06/2020

Abstract: In the last decade, deep learning methods have become the key solution for various machine learning problems. One major drawback of deep learning methods is that they require large datasets to have a good generalization performance. Researchers propose data augmentation techniques for generating synthetic data to overcome this problem. Traditional methods, such as flipping, rotation etc., which are referred as transformation based methods in this study are commonly used for obtaining synthetic data in the literature. These methods take as input an image and process that image to obtain a new one. On the other hand, generative models such as generative adversarial networks, auto-encoders, after trained with a set of image learn to generate synthetic data. Recently generative models are commonly used for data augmentation in various domains. In this study, we evaluate the effectiveness of a generative model, variational autoencoders (VAE), on the image classification problem. For this purpose, we train a VAE using CIFAR-10 dataset and generate synthetic samples with this model. We evaluate the classification performance using various sized datasets and compare the classification performances on four datasets; dataset without augmentation, dataset augmented with VAE and two datasets augmented with transformation based methods. We observe that the contribution of data augmentation is sensitive to the size of the dataset and VAE augmentation is as effective as the transformation based augmentation methods.

Keywords: *data augmentation, generative models, image classification, variational autoencoders*

1. Introduction

In recent years, deep learning techniques have almost become the key solution for image classification problems. While promising results can be obtained by these methods, the main drawback of them is the requirement for large number of training samples. Unless a large dataset is available, deep learning systems cannot be trained enough to get sufficient generalization performance. However, obtaining a large dataset especially with labels is a challenging task which demands to much labor and time. For this reason, given a dataset researchers propose generating synthetic data. This process is referred as data augmentation and it is especially preferred in certain domains such as medical imaging [1], underwater image analysis [2], remote sensing [3], human tracking [4] where it is hard to obtain labeled large datasets.

Augmentation techniques can be grouped as transformation based methods and generative methods. Transformation based methods include techniques such as flipping, cropping, rotation, shifting, drop-out, and addition of noise while generative methods include models such as auto-encoders and generative adversarial networks (GAN). Among these methods, transformation based methods take as input an image and obtain a new image from that given image, while the generative models learn to generate synthetic data which possesses similar properties with a given set of image.

Recently, GANs are commonly used for augmentation. While training a GAN is more demanding than training an auto-encoder, auto-encoder is not thoroughly analyzed in the context of data augmentation. In this study, we examine the effect of data augmentation for various sized datasets. And we compare the classification performances for three conditions; dataset without augmentation, dataset augmented with transformation based method and dataset augmented with the generative method VAE.

[1] *Computer Eng., Alanya Alaaddin Keykubat University, Antalya, TURKEY*
ORCID ID: 0000-0001-7053-5383
[2] *Computer Eng., Alanya Alaaddin Keykubat University, Antalya, TURKEY*
ORCID ID: 0000-0003-4019-7744
** Corresponding Author Email: ozge.karadag@alanya.edu.tr*

2. Related Work

Shorten and Khoshgoftaar [5] present a taxonomy of augmentation methods, where they group those techniques into two as “Basis image manipulations” and “Deep Learning Approaches”. They conduct a survey on limited data problem and they discuss the use of augmentation techniques on various image datasets.

Shijie et. al. [6] reviewed the impact of various augmentation techniques; GANs, flipping, cropping, rotation, shifting etc. on image classification problem. They examined the effect of data augmentation as the size of the dataset changes and they reported that the contribution of data augmentation is more clearly observed as the size of available data shrinks.

Augmentation techniques are commonly used for medical image analysis. Frid-Adar *et al.* [1] propose employing GAN for augmenting dataset on liver lesion classification problem, while Ge *et al.* [7] propose employing GAN for brain image augmentation. Ornek and Ceylan [8] compare the augmentation performance of various traditional transformation methods for Medical Thermography.

A similar study to our study is conducted by Shijie *et al*. [6] which examines various augmentation methods for the image classification problem. They compare the classification performance over datasets which are augmented with the following methods; flipping, cropping, shifting, PCA jittering, Color

jittering, noise, rotation, GAN. As a generative method, GAN is employed, but VAE is not employed in that study.

Variational Autoencoders (VAE) were first introduced by Kingma and Welling [9] for optimizing the inference and learning processes which are based on probabilistic models with intractably distributed parameters. At the end of their study, they obtained a neural network model to be used in recognition, visualization and noise reduction tasks. The study of Jorge *et al.*, where a Convolutional VAE generating synthetic data is proposed, is an example usage of VAE in data augmentation [10]. They evaluate the contribution of this synthetic data by comparing the classification performance with and without the synthetic data. They conclude that the addition of synthetic data improves the classification performance using interpolation or extrapolation when transforming the data space.

Perez and Wang [11] propose a neural network based data augmentation technique to examine the effectiveness of data augmentation for image classification. They compare the classification performance of their neural augmentation method with traditional techniques such as shifting, zooming, distorting etc. The experiments show that the classification using both traditional and proposed augmentation methods outperform the classification without the synthetic data generation.

This study is a comparative research on data augmentation techniques for image classification. For image classification we employ a Convolutional Neural Network architecture. We examine the classification performance as the dataset expands. And we compare the performance of VAE with two transformation based methods; Dropout of Regions and Gaussian Blurring. In section 2 we introduce the data augmentation methods that we employ in this study. In section 3 we explain the experimental setup and give empirical results. Finally, we provide our conclusions and talk about possible future work in section 4.

3. Data Augmentation for Image Classification

3.1. Transformation Based Data Augmentation Techniques

In this study, two transformation based data augmentation methods are used in comparison with VAE. The first method is Gaussian Blurring which is commonly used for smoothing image. A Gaussian Kernel is applied to each pixel with its neighboring pixels for blurring effect. This method is specifically chosen as a baseline augmentation method as it is both easy and commonly used in the literature.

The second method is Dropout of Regions (DoR) which is inspired by the Random Erasing [12] and Cutout Regularization [13] methods. Both Random Erasing and Cutout Regularization remove certain parts of the image. Similarly, DoR randomly selects pixels and erase them from the original image to generate a new image. While doing this process, DoR selects pixels randomly from different color channels so that it both removes certain pixels or change their color. This technique is reported as producing the highest accuracy [5] for the CIFAR-10 dataset [14] which is also employed in this study.

3.2. Data Augmentation with VAE

Variational Autoencoder (VAE) consists of an encoder, a decoder and a loss function. The encoder is a neural network which takes as input a data point x and outputs a latent representation z. The goal of the encoder is to find an efficient compression of data. Let $q_\theta(z|x)$ denote the encoder which is assumed to be Gaussian Probability Density. The encoder outputs parameters to $q_\theta(z|x)$. Once the parameters are estimated, it is possible to sample from this distribution.

The decoder is another neural network, which takes as input the latent representation z and outputs the parameters to the probability distribution of the data. The decoder is represented by $p\phi(x|z)$, and it is expected that the decoder learns to reconstruct data given its latent representation. Information loss during decoding is measured by log-likelihood log $p\phi(x|z)$, which measures how effectively the decoder has learned to reconstruct input data x given its latent representation z.

The loss function for a datapoint x_i is given in Equation 1. The first term in this equation is the reconstruction loss which is defined as the expected negative log-likelihood of the i-th datapoint. The second term is the Kullback-Leibler divergence between the encoder's distribution $q_\theta(z|x)$ and $p(z)$, which measures information loss for using q to represent p.

$$l_i(\theta,\phi) = -\mathbb{E}_{Z\sim q_\theta(z|x_i)}\big[log p_\phi(x_i|z)\big] + \mathbb{KL}(q_\theta(z|x_i)\big|\big|p(z)\big) \quad (1)$$

Once the parameters of the probability distribution of data is learned, one can sample from the distribution and generate new data samples.

3.3. Image Classification

A Deep Convolutional Neural Network (CNN) architecture is used for classification of the images. The architecture of the classifier is inspired by the well-known classifiers AlexNet [14] and LeNet [15]. The architecture given in Figure 1 consists of two folds of convolution layers followed by max-pooling and dropout layers, which is then followed by a fatten layer before two dense layers. Three dropout layers are placed to avoid over-fitting.

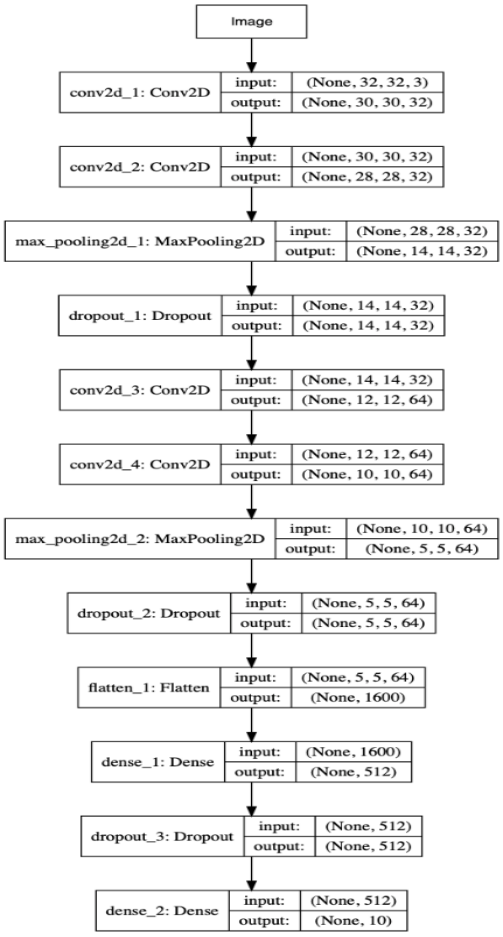


Fig. 1 Architecture of the Classifier.

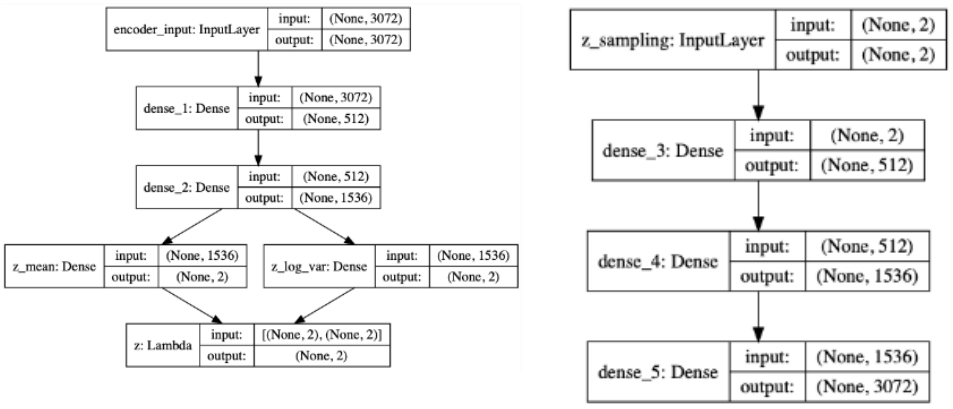


Fig. 3: Architecture of (a) encoder and (b) decoder.

4. Experiments

4.1. Dataset

In our experiments we use the CIFAR-10 dataset [14], which is commonly used in augmentation studies [16], [17], [18], [19], [20]. The dataset consists of 60000 colour images from 10 classes. The size of the images is 32x32 and there are 6000 images from each class. The dataset is separated into 50000 train and 10000 test images. Sample images from the dataset are given in Figure 2.

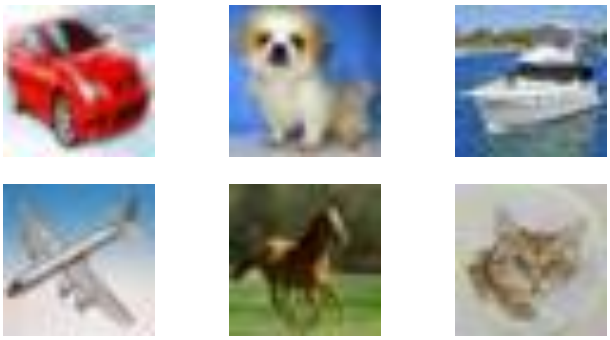


Fig. 2 Sample images from CIFAR-10.

4.2. Experimental Setup

In this study the data augmentation with VAE is compared with two transformation based augmentation methods; Gaussian Blur(GB) and Dropout of regions (DoR). First, Gaussian Blurring is applied on a set of images, the standard deviation of the Gaussian Kernel is randomly selected for each image in the range [0:1; 3:0]. Then, for dropout of regions, augmenter which sets a certain fraction of pixels in an image to zero is applied. The pixels to be set are selected randomly and three color channels are considered separately so that both color variation and omitting of certain pixels are obtained. We use the "imaug" library [21] for transformation based methods. For data augmentation with VAE, a VAE is trained for each class in the dataset. The model of the VAE employed for data augmentation is a three layered encoder-decoder architecture which is depicted in Figure 3. For each class a VAE is trained for 750 epochs and synthetic images for each class are generated using corresponding VAEs.

Sample images obtained by Gaussian Blurring, Dropout of Regions methods and sample images which are generated by VAE are provided in Figure 4. In this figure, columns correspond to the three augmentation techniques; Gaussian Blurring, Dropout of Regions and VAE respectively. At each row of the figure, samples images from different classes are given. Classes from the first row to the last row are; horse, frog, car, ship, truck and cat respectively.



Fig. 4 Sample images that are obtained by Gaussian Blurring, Dropout of Regions and VAE.

4.3. Results

We evaluate the effectiveness of data augmentation with three different techniques on various sized datasets. The experiments are run in four folds for the number of original images {5000,10000,25000,40000}. At each fold of the experiment, we select an N number of images randomly for training. And then we select N*0.25 number of images randomly for testing.

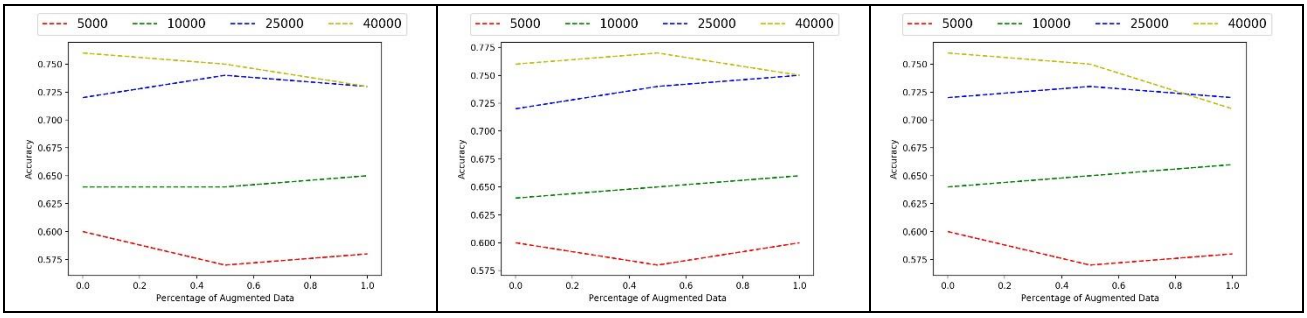


Fig. 5 Classification Accuracies obtained by augmentation with Gaussian Blurring, Dropout of Region and VAE respectively.

Table 1. Classification performance for various sized datasets with and without augmentation

Num. of Or. Im.	Num. of Gen. Im.	Accuracy			
		Baseline	GB	DoR	VAE
5000	-	0.60	-	-	-
5000	2500	-	0.57	0.58	0.57
5000	5000	-	0.58	0.60	0.57
10000	-	0.64	-	-	-
10000	5000	-	0.64	0.65	0.65
10000	1000	-	0.65	**0.66**	**0.66**
25000	-	0.72	-	-	-
25000	12500	-	0.74	0.74	0.73
25000	25000	-	0.73	0.75	0.72
40000	-	0.76	-	-	-
40000	20000	-	0.75	0.77	0.75
40000	40000	-	0.73	0.75	0.71

In the first run of the fold, only the original images are used for training. In the second run of the fold, N*0.5 number of images are generated by three techniques separately and those images together with the original images are used for training. In the third run of the fold, N number of images are generated by three techniques separately and they are used together with the original images for training. In the third run, there are as many synthetic samples as the number of original images. Classification performance is measured as the accuracy which is the percentage of correct classifications.

The results for four fold experiments are provided in Table 1. In this table, each fold is separated by bold lines. At each fold, the first row is the classification without augmentation which is referred as the Baseline method.

Analysis of Table 1, points out that VAE is as much effective as DoR in the second fold where the number of original images is set as 10000.

Table 1 also reveals that, as the size of the dataset increases, the contribution of augmentation is almost not required at all. Moreover, at a size of 40000 images, generating another 40000 images causes a decrease in the performance for all three methods. Accuracy vs. percentage of synthetic data graphs are given in Figure 5. In this figure, the x axis refers to the amount of augmented data. 0:0 refers to no augmentation while 1:0 refers to N number of augmented images for N number of original images. This figure shows that; it is possible to obtain a slightly better performance than the baseline method which uses no augmentation. This increase is most evident with a dataset size of 10000 images. Results in Table 1 reveal that VAE is not more effective than the transformation based methods; Gaussian Blur and Dropout of Regions.

5. Conclusion

In this study, we empirically evaluate the effectiveness of VAEs on data augmentation for the image classification problem. Besides augmentation with VAE, we employ two widely used transformation methods; Gaussian Blurring and Dropout of Regions. Our experiments reveal that VAE augmentation is nearly as effective as the transformation based methods. However, VAE does not come out to be more effective than the two traditional methods.

Experiments show that the contribution of data augmentation becomes more evident as the size of the initial dataset shrinks. Nevertheless, if the dataset is too small, augmentation does not improve the classification performance. Moreover, classification after augmentation with very small data degrades the classification performance.

As a future work, variation of VAE, such as conditional VAE can be evaluated for data augmentation. Another interesting study would be combining various augmentation methods. Both the transformation based methods and generative models can be applied to augment data and its effectiveness on the classification performance can be observed.

References

[1] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington.

[2] Y. Xu, Y. Zhang, H.Wang and X. Liu, "Underwater image classification using deep convolutional neural networks and data augmentation," 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xiamen, 2017, pp. 1-5.

[3] M. Zhang, Z. Cui, X. Wang and Z. Cao, "Data Augmentation Method of SAR Image Dataset," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, 2018, pp. 5292-5295.

[4] *K. Chen, X. Zhou, Q. Zhou and H. Xu, "Adversarial Learning-based Data Augmentation for Rotation-robust Human Tracking,"* ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), *Brighton, United Kingdom, 2019, pp. 1942-1946*

[5] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[6] J. Shijie, W. Ping, J. Peiyi and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 4165-4170.

[7] C. Ge, I. Y. Gu, A. Store Jakola and J. Yang, "Cross-Modality Augmentation of Brain Mr Images Using a Novel Pairwise Generative Adversarial Network for Enhanced Glioma Classification," *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 2019, pp. 559-563.

[8] A. H. Ornek and M. Ceylan, "Comparison of Traditional Transformations for Data Augmentation in Deep Learning of Medical Thermography," *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, 2019, pp. 191-194.

[9] Kingma, D.P., Welling, M. (2014). "Auto-Encoding Variational Bayes", CoRR, abs/1312.6114.

[10] Jorge, J., Vieco, J., Paredes, R., Snchez, J., Bened, J. (2018). Empirical evaluation of variational autoencoders for data augmentation. VISIGRAPP.

[11] Perez, L., Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. CoRR, abs/1712.04621.

[12] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, 'Random Erasing Data Augmentation'. ArXiv e-prints 2017.

[13] Terrance V, Graham WT. Improved regularization of convolutional neural networks with cutout. arXiv preprint. 2017.

[14] Alex Krizhevsky, "Learning Multiple Layers of Features from Tiny Images", Technical Report, 2009.

[15] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geofrey E. (2017-05-24). "ImageNet classification with deep convolutional neural networks" (PDF). Communications of the ACM. 60 (6): 8490.

[16] Y. Lecun, L. Bottou, Y. Bengio and P. Ha_ner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[17] J. Shijie, W. Ping, J. Peiyi and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," *2017 Chinese Automation Congress (CAC)*, Jinan, 2017, pp. 4165-4170.

[18] A. Merchant, T. Syed, B. Khan and R. Munir, "Appearance-based data augmentation for image datasets using contrast preserving sampling," *2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, 2018, pp. 1235-1240.

[19] X. Liu *et al.*, "Data Augmentation via Latent Space Interpolation for Image Classification," *2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, 2018, pp. 728-733.

[20] E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan and Q. V. Le, "AutoAugment: Learning Augmentation Strategies From Data," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 113-123.

[21] R. Takahashi, T. Matsubara and K. Uehara, "Data Augmentation using Random Image Cropping and Patching for Deep CNNs," in *IEEE Transactions on Circuits and Systems for Video Technology*.

[22] https://imgaug.readthedocs.io/, last accessed in March 2020.