Evaluating Image Quality through Latent Space Analysis of Autoencoders

Gofur Halmuratov* and Arnošt Veselý†
Czech University of Life Sciences, Prague, CZECH REPUBLIC

Abstract

In the domain of deep learning-driven image classification, the underpinning algorithms often grapple with reduced performance efficacy when working with constrained datasets. Such algorithms typically thrive in scenarios with a considerable volume of data at their disposal, but they falter when their operational spectrum narrows down to a handful of images. This necessitates data augmentation or the synthesis of new data instances via Generative Adversarial Networks (GANs). However, these methodologies do not always yield images that align with the desired quality criteria. As such, the onus of evaluating and sieving out low-quality images falls on the researchers, who must conduct a meticulous manual review of each image. This approach, albeit thorough, is riddled with the challenges of being highly time-consuming and resource-exhaustive.

In this study, we introduce a novel methodology that leverages the latent space of autoencoders for image quality assessment. This unique approach bypasses the need for manual review, allowing us to infer image quality by analyzing the latent space representation. We furnish empirical evidence of our methodology's efficacy through extensive experimentation, which unveils its superior performance over conventional image quality evaluation techniques.

Key Words: Autoencoder, latent space analysis, image quality, perceptual metrics, regularization, deep learning.

1 Introduction

Deep learning algorithms have shown remarkable performance in image classification and recognition tasks. However, their effectiveness is often hindered when working with a small number of images [2]. When the training dataset consists of only a few images, researchers usually augment them or synthesize new images using generative adversarial networks (GANs) [5]. Nevertheless, these methods may not always produce images of the required quality. As a result, manual review of individual images is necessary to remove low-quality ones from the dataset, which is a time-consuming and resource-intensive process [10]. One solution to this issue is to evaluate the quality of images through the analysis of the latent space of autoencoders. An autoencoder is a type of neural network that learns to encode and decode data, reducing the dimensions of the input data into a latent space [6]. The latent space is a compressed representation of the input data, where each dimension represents a learned feature of the input [1]. Analyzing the latent space of autoencoders can provide a more reliable measure of image quality than manual review [8]. Several studies have investigated the use of autoencoders and their latent space for image quality evaluation. In [7], a novel method was proposed for evaluating image quality using the variance of the latent representation. In [12], a perceptual loss function was introduced to improve the quality of the generated images using GANs. A regularization term was added to the loss function to encourage the generated images to be more similar to the real images in the latent space. In [4], the latent space was used to evaluate the quality of face images, and the results were compared with the human perceptual evaluation. Other studies have used autoencoders and their latent space for image generation tasks. In [3], an adversarial autoencoder was introduced, which combines the advantages of both autoencoders and GANs. In [9], a conditional autoencoder was proposed, where the input data and a condition vector are combined to generate high-quality images. In this paper, we propose a methodology for evaluating the quality of images by analyzing the autoencoder latent space. We demonstrate the effectiveness of our approach in comparison to conventional methods for determining image quality [11]. Our method can be used in various image-related tasks, including image restoration, synthesis, and recognition.

2 Methodology

We start with a small number of training datasets and our goal is to improve classification accuracy. First, we attempt to classify the existing datasets using deep learning algorithms. If the results are not satisfactory, we try image augmentation or generative adversary networks to improve the datasets.

However, if these methods do not produce the desired results, we turn to quality determination using the latent space of an autoencoder. Here are the steps involved:

* ORCiD ID: Ing. Gofur Halmuratov https://orcid.org/0000-0003-4403-8046. Email: halmuratov@pef.czu.cz
† ORCiD ID: doc. Ing. Arnošt Veselý, CSc https://orcid.org/0000-0001-8979-1336. Email: vesely@pef.czu.cz

1. We train an autoencoder for all the input training datasets and encode all available images. By decoding them, we obtain the latent space of the autoencoder.
2. We represent the first class of images in latent space.
3. Points of training datasets are generated in the latent space, and we hypothesize a sphere that filters out 95% of the points. We find the radius of this sphere.
4. We synthesize the required number of images using image augmentation or generative adversarial networks.
5. The newly synthesized images are imaged again in latent space. Images outside the sphere are considered low-quality and are excluded from the deep learning models. The ones inside are included in our training dataset. This sphere serves as a qualifier to divide the quality of images into good or bad.
6. Starting from step 2, we repeat the process for the rest of the classes.

By the end of this process, we have high-quality images to train deep learning models, and our training dataset is now larger. In summary, this methodology improves classification accuracy by using deep learning algorithms, image augmentation, generative adversary networks, and quality determination using the latent space of an autoencoder. This methodology aims to improve classification accuracy despite having a small number of training datasets. It does this by using deep learning algorithms and image augmentation or generative adversary networks to improve existing datasets. If these steps do not lead to significant improvement, quality determination is performed using the latent space of an autoencoder. Finally, the high-quality images obtained from this process are used to train deep learning models.

3 Results

We evaluated our proposed methodology on several open access datasets, including MNIST, CIFAR-10 and CIFAR-100. Our approach is designed to improve classification accuracy on datasets with very few images, so we intentionally used a small number of images from each class (5, 10, 20, and 100) in each of the test datasets. We then compared the classification results of these images with and without our proposed methodology to assess its effectiveness. All models used the same autoencoder, which is presented in Table 1: Model 1 and Table 2: Model 2.

Table 3 shows the datasets, the number of images in each

Table 1: Model 1. VAE encoder model

Layer(type)	**Output Shape**	**Param#**	**Connected to**
encoder_input(InputLayer)	[(None,28,28,1)]	0	[]
conv2d(Conv2D)	(None,14,14,512)	5120	['encoder_input[0][0]']
conv2d_1(Conv2D)	(None,7,7,1024)	4719616	['conv2d[0][0]']
flatten(Flatten)	(None,50176)	0	['conv2d_1[0][0]']
dense(Dense)	(None,400)	20070800	['flatten[0][0]']
z_mean(Dense)	(None,100)	40100	['dense[0][0]']
z_log_var(Dense)	(None,100)	40100	['dense[0][0]']
z(Lambda)	(None,100)	0	['z_mean[0][0]','z_log_var[0][0]']

Table 2: Model 2. VAE decoder model

Layer (type)	**Output Shape**	**Param #**
z_sampling(InputLayer)	[(None,100)]	0
dense_1(Dense)	(None,50176)	5067776
reshape(Reshape)	(None,7,7,1024)	0
conv2d_transpose(Conv2DTranspose)	(None,14,14,1024)	9438208
conv2d_transpose_1(Conv2DTranspose)l	(None,28,28,512)	4719104

Table 3: Results of experiments

Dataset	Number of images per category	Data augmentation	Number of qualified images	Accuracy before using proposed method	Accuracy after using proposed method
MNIST	5	500	425	20.6%	25.8 %
MNIST	10	1000	895	25.8%	30.2 %
MNIST	20	2000	1578	27.1%	30.9 %
MNIST	100	5000	3685	25.4%	24.8 %
CIFAR-10	5	500	458	18.9%	21.6 %
CIFAR-10	10	1000	901	23.4%	24.2 %
CIFAR-10	20	2000	1570	28.2%	31.2 %
CIFAR-10	100	5000	4001	34.3%	40.9%
CIFAR-100	5	2500	1978	14.6%	18.2 %
CIFAR-100	10	5000	3951	16.4%	29.5 %
CIFAR-100	20	10000	8012	18.2%	21.8 %
CIFAR-100	100	50000	38417	24.5%	27.6 %

dataset, the number of images after applying our methodology, the overall classification accuracy, and the classification accuracy of the qualified images.

The assessment of the proposed methodology was carried out across three standard benchmarks: MNIST, CIFAR-10, and CIFAR-100. The critical premise of this analysis was to gauge the effectiveness of the classification model when restricted to datasets with a limited number of images per category.

For the MNIST dataset, the proposed methodology showcased marked improvements. With only 5 images per category, the model's accuracy experienced an augmentation of 25.24%, increasing from 20.6% to 25.8%. When the model was supplied with 10 images per category, the observed accuracy escalated from 25.8% to 30.2%, marking an improvement of approximately 17.05%. Furthermore, with 20 images per category, the model's accuracy ascended from 27.1% to 30.9%, signifying a growth of about 14.02%. Intriguingly, a marginal degradation of 2.36% in accuracy, from 25.4% to 24.8%, was observed when the model was tested with 100 images per category.

The application of the methodology on the CIFAR-10 dataset also yielded enhanced accuracy. With 5 images per category, the accuracy was amplified from 18.9% to 21.6%, denoting a relative improvement of 14.29%. For 10 images per category, the accuracy ascended from 23.4% to 24.2%, marking a modest growth of 3.42%. When provided with 20 images per category, the model accuracy improved from 28.2% to 31.2%, constituting a relative gain of 10.64%. Moreover, the accuracy exhibited a substantial surge of 19.24%, improving from 34.3% to 40.9% when tested with 100 images per category.

The CIFAR-100 dataset, although more complex, still witnessed improvements in accuracy with our methodology. For 5 images per category, accuracy augmented from 14.6% to 18.2%, marking an improvement of 24.66%. When 10 images per category were used, a significant leap in accuracy was observed, from 16.4% to 29.5%, translating to an outstanding improvement of 79.88%. For 20 images per category, accuracy rose from 18.2% to 21.8%, implying a relative growth of 19.78%. Lastly, for 100 images per category, the model's accuracy improved from 24.5% to 27.6%, denoting an enhancement of approximately 12.65%.

The results emanating from this study underscore the effectiveness of our proposed methodology. The results reveal consistent improvements in classification accuracy across MNIST and CIFAR-10 datasets. Even in the more complex CIFAR-100 dataset, our methodology continued to show efficacy, indicating its robustness in diverse and challenging contexts. The observed anomalous reduction in accuracy for MNIST with 100 images per category, however, suggests the need for an in-depth investigation into the intricate dynamics at play and potential refinement of the method for better performance.

In summary, our methodology exhibits a commendable potential to enhance the performance of image classification models, particularly when dealing with datasets that offer limited images per category. Future directions of this research will focus on dissecting the unexpected result observed with MNIST (100 images per category), and broadening the applicability of our methodology across other datasets of varying complexity.

4 Discussion and Conclusions

The proposed methodology, conceived to address the challenge of improving classification accuracy with a restricted number of training datasets, was rigorously evaluated on a series of established open-access datasets, namely MNIST, CIFAR-10, and CIFAR-100. Our methodology employs an amalgamation of deep learning algorithms, image augmentation or Generative Adversarial Networks (GANs), and if the results from these steps are insufficient, quality determination is performed via the latent space of an autoencoder.

An inherent limitation of this approach is its dependence on the hypothesis that the autoencoder can accurately capture the true distribution of images within the dataset. This assumption may not stand in scenarios involving complex datasets or datasets containing outliers not well represented by the autoencoder. Additionally, the quality determination phase operates under a predetermined threshold to segregate high-quality from low-quality images, which may not universally hold for all datasets.

The empirical results from our methodology, although tested on smaller datasets, have been promising. The most substantial improvements were observed for datasets with smaller quantities of training images. For instance, with just five images per category in the CIFAR-100 dataset, we observed an improvement of approximately 24.66%, increasing the accuracy from 14.6% to 18.2%. When the model was trained with 10 images per category, the accuracy escalated from 16.4% to a notable 29.5%, translating to an outstanding improvement of nearly 80%. This highlights the efficacy of our methodology in scenarios with limited training data.

However, the observed anomalous reduction in accuracy for the MNIST dataset with 100 images per category, dropping by 2.36% from 25.4% to 24.8%, prompts further investigation. This indicates that while our methodology is potent in scenarios with limited data, its performance may vary in cases where a larger pool of training images is available.

Further exploration into the effectiveness of our methodology on larger, more complex datasets is necessary. As the greatest improvements were observed for datasets with smaller numbers of training images, it remains to be seen how the methodology would perform when the number of training images increases significantly. This highlights the need for further research into the methodology's scalability and its performance with larger, more complex datasets.

In conclusion, the proposed methodology provides a robust and efficacious strategy for bolstering classification accuracy in scenarios where the number of training images is limited. Despite its reliance on certain assumptions and pre-set thresholds, it shows a consistent trend of enhancing classification accuracy across various datasets. Further studies are warranted to explore its effectiveness with larger datasets and more complex image distributions, which could potentially unlock its full potential and extend its applicability.

References

[1] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798-1828, 2013.

[2] Y. Bengio and Y. LeCun, *Scaling Learning Algorithms Towards AI,* Large-scale Kernel Machines, MIT Press, Cambridge, ;MA 02142, USA, 1(1):1-41, (2007).

[3] Y. Blau and T. Michaeli, "The Perception-Distortion Tradeoff," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6228-6237, 2018.

[4] A. Dosovitskiy and T. Brox, "Inverting Visual Representations with Convolutional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4829-4837, 2016.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S., Ozair, .and Y. Bengio, "Generative Adversarial Nets," Advances in Neural Information Processing Systems, pp. 2672-2680, 2014.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, 313(5786):504-507, 2006.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708, 2017.

[8] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," arXiv preprint arXiv:1511.06434, 2015.

[9] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative Adversarial Text to Image Synthesis," *Proceedings of the International Conference on Machine Learning*, pp. 1060-1069, 2016.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, 13(4):600-612, 2004.

[11] Y. Zhang, X. Song, H. Qi, J. Liu, and H. Zhang, "An End-to-End Framework for Image Quality Assessment based on Conditional Generative Adversarial Networks," *IEEE Transactions on Multimedia*, 20(10):2678-2690, 2018.

[12] J. Y. Zhu, T. Park, P, Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223-2232, 2017.

Gofur Halmuratov (photo not available) emerges as a dynamic force at the heart of the Czech University of Life Sciences (ČZU), distinguishing himself as an Engineer (Ing.) and a dedicated Ph.D. student. Situated within the Provozně Ekonomická Fakulta (Faculty of Economics and Management) and the Katedra Informačního Inženýrství (Department of Information Engineering), Gofur is a pivotal contributor to the academic landscape.

In his capacity as a Ph.D. student, Gofur is actively engaged in pushing the frontiers of knowledge, embodying a spirit of curiosity and innovation. His role transcends mere research, as he generously offers his time for consultations and official hours, illustrating a steadfast commitment to the academic community and the development of future information engineers.

Communication with Engineer Gofur Halmuratov is facilitated through email, providing a direct conduit to his profound insights and collaborative potential. Within the vibrant academic milieu of ČZU, Gofur's contributions are instrumental in shaping the future of information engineering, rendering him an invaluable asset within the Provozně Ekonomická Fakulta.

Arnošt Veselý (photo not available)stands as a luminary figure within the academic community of the Czech University of Life Sciences (ČZU). Possessing the esteemed title of Associate Professor and having attained a Doctorate in Sciences (CSc.), he is an eminent presence within the Provozně Ekonomická Fakulta (Faculty of Economics and Management). His academic journey is intricately intertwined with the Katedra Informačního Inženýrství (Department of Information Engineering), where he imparts profound expertise in the realm of information engineering.

As a teacher Associate Professor Veselý's dedication to education is discernible, manifesting in his pivotal role as a guiding force for students. His influence extends beyond the confines of the classroom, encompassing consultations and administrative responsibilities, with his office at PEF/E550 serving as a focal point for academic discourse. Beyond his educational endeavors, Associate Professor Arnošt Veselý is a prolific author, leaving an indelible mark on the scholarly landscape associated with ČZU.

For those seeking his erudition or collaboration, Associate Professor Veselý is easily accessible through email or the ČZU hotline. His multifaceted role as an educator, consultant, and researcher underscores his commitment to the comprehensive development of students and the advancement of knowledge in the dynamic field of information engineering.