UDC 004.852
DOI: 10.25140/2411-5363-2021-2(24)-83-91

Oleksandr Chaikovskyi, Artem Volokyta, Artemii Kyrianov, Heorhii Loutskii

DATA AUGMENTATION METHOD USING GENERATIVE ADVERSARIAL NETWORKS

The article discusses a data augmentation method based on generative adversarial networks to improve the accuracy of image classification by convolutional neural networks. A comparative analysis of the proposed method with classical image augmentation methods was performed.

Keywords: *data augmentation; generative adversarial networks; convolutional neural networks.*
Fig.: 10. Table: 3. References: 11.

Urgency of the research. One of the most important areas of research and development of modern computer technologies is the field of machine learning, pattern recognition, and computer vision. A large number of different data is the basis for building a reliable machine learning model because it helps to more accurately summarize the information about the task. At the same time, there are areas where data collection is either impossible or requires a huge number of resources. For example, in the medical field, patient data is protected by general data protection regulations, which makes it difficult to collect, store, and use information. Modern machine learning systems solve this problem using methods of synthetic data generation. The article considers the development of a method of data augmentation based on generative adversarial networks.

Target setting. An important step in building an accurate machine learning model is the collection and annotation of data that will be used to train and test the accuracy of the neural network. The accuracy and stability of the neural network in real conditions depend on the amount of data collected. The article proposes a method for improving the accuracy of image classification by convolutional neural networks based on data augmentation using generative adversarial networks.

Actual scientific researches and issues analysis. One of the most common ways to increase the accuracy of classification in a limited dataset is to build and train ensembles of classifiers. This approach is based on the idea that combining independent classifiers into an ensemble can compensate for their individual shortcomings through collective voting, thus providing higher classification accuracy and greater resistance to accidental emissions in the processed data. Known methods of constructing ensembles of classifiers can be divided into dependent, which uses the classifiers obtained in the previous stages, to train new, more advanced classifiers, and independent, in which each classifier of the ensemble is trained independently of the others [1].

Increasing the size of the training sample due to the transformation of the original training data is most often used in image recognition problems, so these methods are focused mainly on image processing. The rotation at a random angle, compression, and stretching vertically and horizontally, incline, mirror reflection, offset, and many other transformations are used often when generating images [2].

Algorithms for introducing artificial realistic deformation are proposed in [3], based on which the initial training dataset of original images was increased and used to train the Viola-Jones algorithm (algorithm of the AdaBoost family). Using this algorithm, the size of the initial dataset can be reduced by 10 times (1000 original images instead of 10000) by reducing the recognition accuracy in the range of 2-4 %.

An algorithm that generates feature values for a synthetic image as independent random variables in the range between the minimum and maximum values of the corresponding feature from the input training dataset was described in [4]. This algorithm was used to synthetically generate the most difficulty recognized samples in iterations of a modified boosting algorithm

designed to solve the problem of imbalanced dataset. The resulting algorithm provided recognition accuracy comparable to the AdaBoost algorithm for balanced datasets and significantly higher accuracy for imbalanced datasets.

In [5], an algorithm for the artificial generation of synthetic data called SMOTE (Synthetic Minority Over-sampling Technique) was described. For each training sample of the minority class in the input samples, several nearest neighbors are sought. Then several of them are randomly selected, and the number of selected samples is determined depending on the required generation factor (if the sample size needs to be increased by 200 %, select 2 random nearest neighbors, if 300% - 3 and so on). Next, for each selected neighbor, the vector of distances between its feature vector and the feature vector of the reference sample is calculated and then multiplied by a random number in the range from 0 to 1. The obtained vector is summed with the vector of features of the reference sample. The authors test their algorithm on several test datasets and note that in most cases, its usage allows achieving better results compared to traditional generation with repetitions. In their subsequent works, the authors combined boosting and SMOTE algorithm, which allowed them to achieve even higher results.

Uninvestigated parts of general matters defining. The possibilities of using generative adversarial networks in the problem of data augmentation remain unconsidered.

The research objective. The aim of the research is to develop a method of data augmentation using generative competitive networks to improve the accuracy of image classification by convolutional neural networks.

The statement of basic materials. Generative adversarial networks (GAN) is a machine learning algorithm that belongs to a family of generative models and is built on a combination of two neural networks, one of which generates samples and the other tries to distinguish real samples from generated ones. Such networks were first introduced by Ian Goodfellow in 2014 [6].

In the GAN system (Fig. 1) one of the networks (network G, from Generator, generative model) generates samples, and the other (network D, from Discriminator, discriminative model) tries to distinguish correct samples from incorrect ones. Using a set of latent space variables, the generative network tries to create a new sample by mixing several source samples. The discriminative network learns to distinguish between real and generated samples, and the results of the distinction are fed to the input of the generative network so that it can select the best set of latent parameters, and the discriminative network would no longer be able to distinguish real samples from generated ones. Thus, the goal of network G is to increase the error rate of network D, and the goal of network D is to increase the accuracy of recognition.

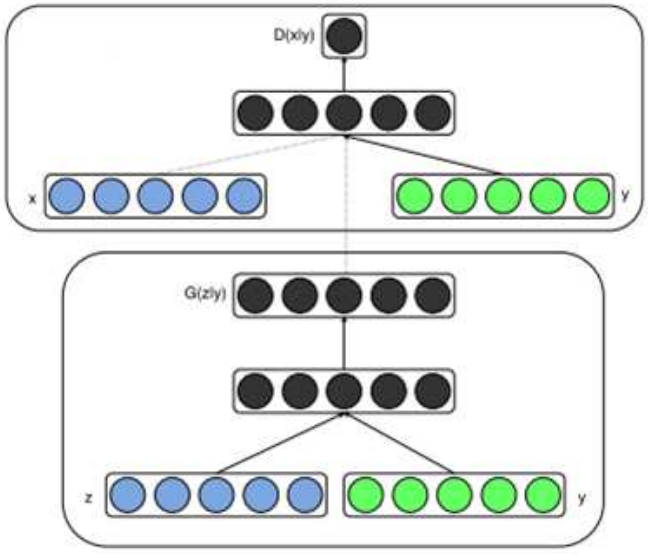

Fig. 1. Generative adversarial network architecture

Discriminative network D, analyzing samples from the original dataset and from samples generated by the generator, achieves a certain recognition accuracy. The generator G starts with random combinations of latent space parameters, and after evaluating the obtained samples from network D, the backpropagation method is used, which allows improving the quality of generation by adjusting the input set of latent parameters. Gradually, artificial images at the output of the generative network are becoming more qualitative. Network D is implemented as a convolutional neural network, while network G unfolds images based on latent parameters.

Generative adversarial network architecture for image augmentation. The transformation network can be represented as $f_W(x)$, which is built based on a deep convolutional neural network [7]. The input image x could be passed to the neural network and transformed into the output image y through the network $y = f_W(x)$. This network is trained by the method of stochastic gradient descent [8] to minimize the weighted combination of loss functions. The total loss can be defined as a linear combination of two content reconstruction losses and a style reconstruction loss, and the weights of the two losses can be modified and adjusted.

The transformation network consists of a generator and a discriminator (Fig. 2), which are parts of generative adversarial networks. Modeling GANs to generate high-resolution images is a very unstable process, so choosing the same architecture for both the generator and the discriminator for different datasets with different resolutions will not lead to the desired result. That is why the models were trained in such a way that the number of convolutional layers of the neural network in the generator and the discriminator depends on the resolution of the images in the dataset. The number of convolutional layers in the generator is calculated as:

$$n_G = log_2(h) - 2, \quad (1)$$

where n_G – the number of convolutional layers in the generator, h – image height.

The number of convolutional layers in the discriminator is calculated as:

$$n_D = n_G + 1 \quad (2)$$

where n_D – the number of convolutional layers in the discriminator.

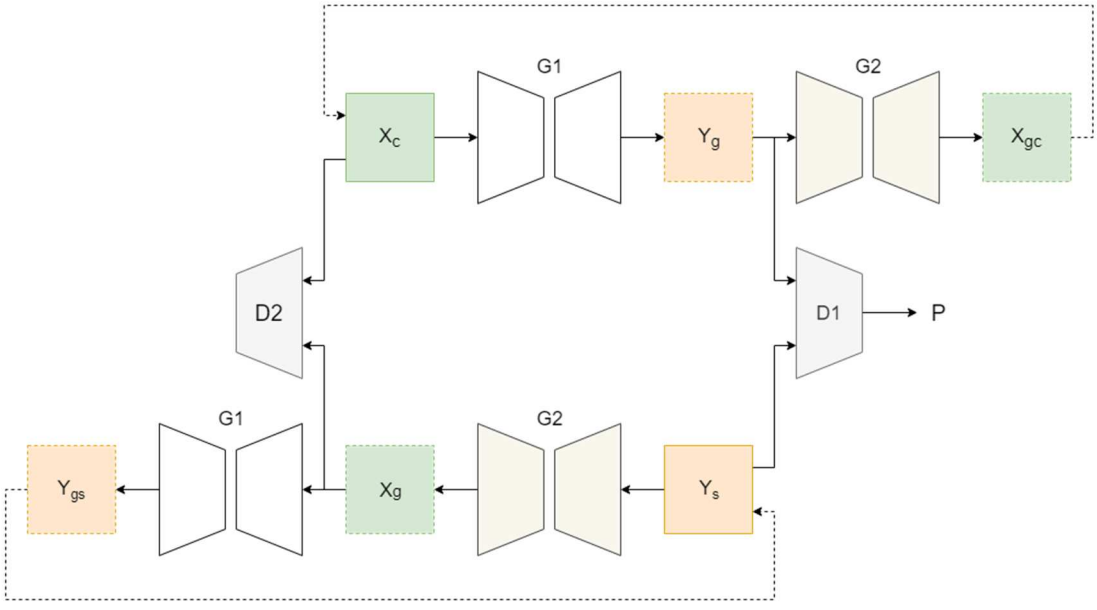

Fig. 2. Generative adversarial network architecture for image augmentation

When the transformation network is trained, it is applied to the training dataset for augmentation. The generated samples are used together with the original images to train convolutional neural networks for image classification. In research, a convolutional neural network with VGG16 architecture is used [9].

Datasets for image augmentation. CIFAR-10 [10] and MNIST [11] datasets were used to augment and test image classification accuracy.

The CIFAR-10 dataset consists of 60,000 color images (Fig. 3) with a size of *32x32* pixels with 6,000 images per class. The dataset is divided into samples for training and testing. There are 50,000 images in the training dataset, and the other 10,000 images are in the test dataset.

The MNIST dataset contains images of handwritten digits (Fig. 4). The MNIST dataset contains 70,000 handwritten digits. Images are *28×28* pixels in size, which belong to 10 different classes. The training dataset consists of 60,000 images, and the test dataset consists of 10,000 images.

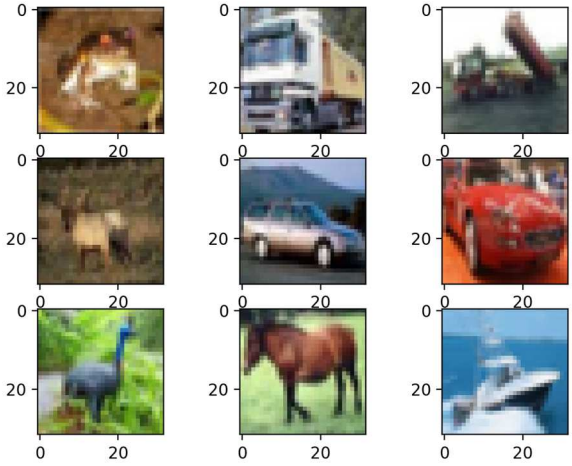

Fig. 3. CIFAR-10 image samples

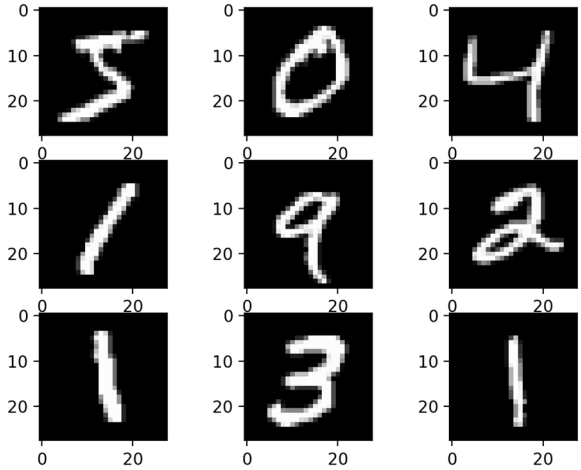

Fig. 4. MNIST image samples

Analysis of results. The convolutional neural network used to classify images is trained in three different configurations. The first configuration uses images from the original dataset without augmentation. This will give basic results for comparison. In the second configuration, additional data is generated using existing data augmentation methods (horizontal and vertical flipping, 90- and -90-degree rotation). This artificial data is added to the original training dataset used to train convolutional neural networks. In the third configuration, images generated by generative adversarial networks are added to the original training dataset. When comparing the results, the same training parameters are used in all experiments. The network is trained for 20,000 iterations with a batch size of 32 images and the stochastic gradient descent optimization algorithm, resulting in some instability of the results. For this reason, training is repeated 10 times with different random initial coefficients.

Table 1 – Datasets characteristics

Name	Publication date	Number of categories	Number of samples
MNIST	1998	10	70000
CIFAR-10	2010	10	60000

The recognition accuracy on the CIFAR-10 training set is 0.7081, and the recognition accuracy on the test set - 0.37. Note that the accuracy on the test set is extremely low compared to the accuracy on the training set. This is because only 5,000 images are used to optimize the network with hundreds of thousands of parameters that results in heavily overtraining.

On the other hand, on the MNIST dataset recognition accuracy achieved 0.9989 and 0.9513 on the training and test sets, respectively. The images in MNIST have fewer low-level features than the images in the CIFAR-10 dataset. This is the reason that a very massive convolutional neural network, such as VGG, has a high accuracy of classification on test data.

Table 2 – Classification accuracy on the CIFAR-10 dataset

Dataset	Training accuracy	Validation accuracy
Original images	0.7081	0.3021
Classical augmentation	0.7800	0.2467
GAN-based augmentation	**0.7913**	**0.3700**

Table 3 – Classification accuracy on the MNIST dataset

Dataset	Training accuracy	Validation accuracy
Original images	0.9689	0.8172
Classical augmentation	0.9089	0.2709
GAN-based augmentation	**0.9989**	**0.9513**

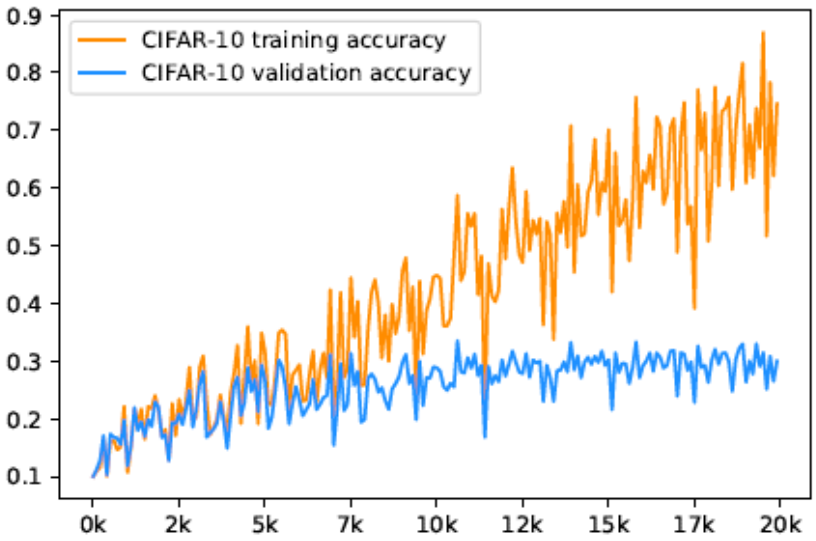

Fig. 5. Classification accuracy of the 5000 images from the CIFAR-10 dataset without augmentation

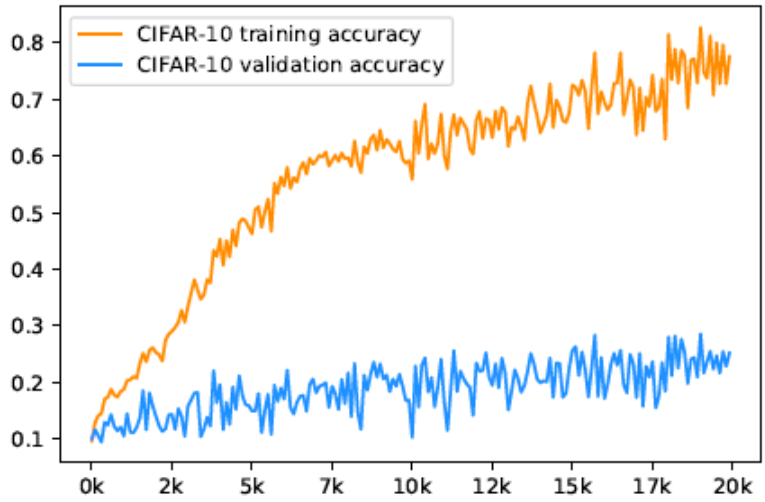

Fig. 6. Classification accuracy of the 20000 images generated based on the CIFAR-10 dataset using classical augmentation methods

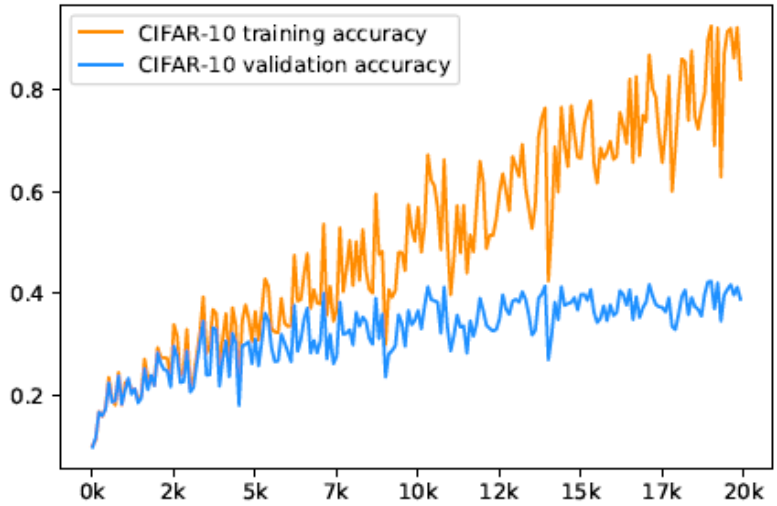

Fig. 7. Classification accuracy of the 20000 images generated based on the CIFAR-10 dataset using GAN-based augmentation

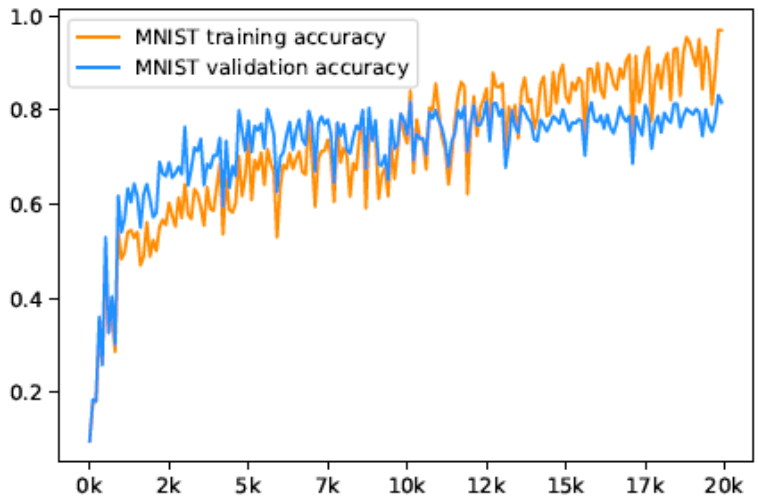

Fig. 8. Classification accuracy of the 5000 images from the MNIST dataset without augmentation

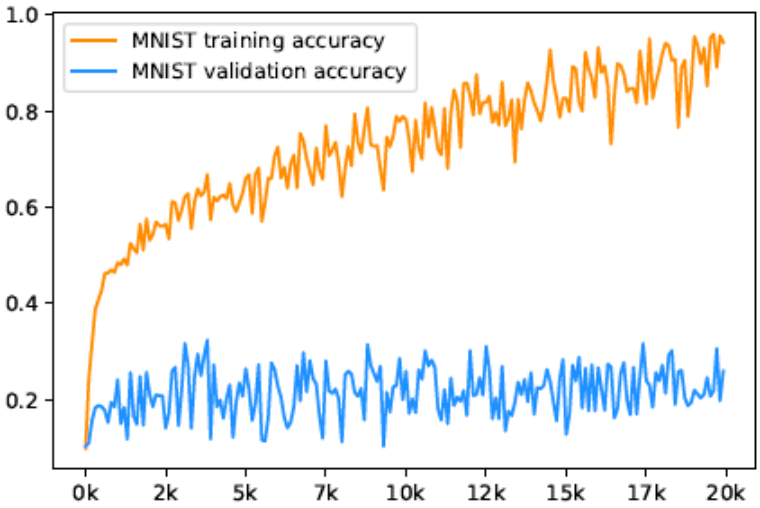

Fig. 9. Classification accuracy of the 20000 images generated based on the MNIST dataset using classical augmentation methods

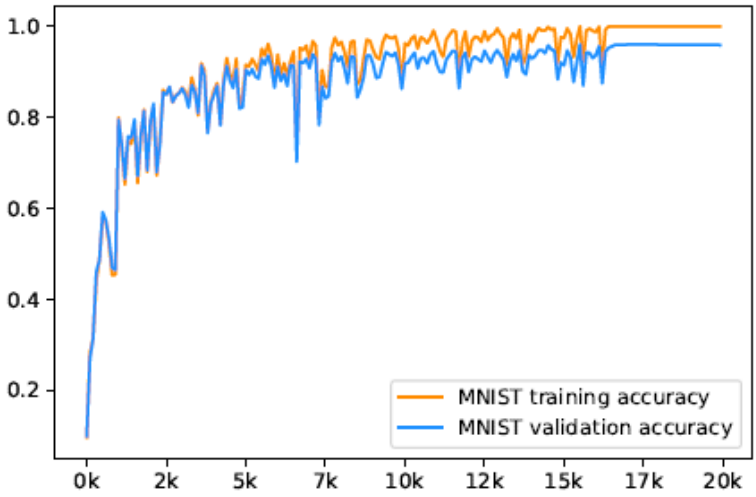

Fig. 10. Classification accuracy of the 20000 images generated based on the MNIST dataset using GAN-based augmentation

Conclusions. In this paper the method of data augmentation with the use of generative adversarial networks was proposed, the comparative analysis of classical image augmentation methods with the offered one was performed.

The advantage of the proposed method is increasing the accuracy of image classification problems with the use of convolutional neural networks. The classification accuracy of the VGG16 network was increased by 13.41% and 6.79% for the MNIST and CIFAR-10 datasets, respectively.

However, the main disadvantage of the proposed method is increasing the training time of the convolutional neural network relative to the number of augmented images and the training time of the generative adversarial network itself.

As part of the further improvements of the data augmentation method to improve the classification accuracy by convolutional neural networks, I consider it most appropriate to research the methods of reconstruction based on the input data of multidimensional density distribution probability of the feature vector. This approach will reduce the dataset size by reducing the total training time of the model.

It is also necessary to analyze the results of the proposed method on a larger number of datasets, such as Caltech101, Caltech256, ImageNet, and consider upgrading the architecture of the generative adversarial network to work with tabular data.

References

1. Rokach, L. Ensemble-based Classifiers. *Artificial Intelligence Review*. 2010. Vol. 33. Pp. 1–39.
2. Ciresan, D. C., Meier, U., Gambardella, L. M., Schmidhuber, J. (2010). Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition. *Neural Computation*. Vol. 22(12).
3. Akimov, O. V., Syrota, O. O. (2016). Models and algorithms of artificial dataset augmentation for training algorithms persons recognition using Viola-Jones method. *Computer optics*. Vol. 40, № 6. Pp. 911-918.
4. Guo, H., Viktor, H. L. (2014). Learning from Imbalanced Data Sets with Boosting and Data Generation: The DataBoost IM Approach. *ACM SIGKDD Explorations Newsletter*. Vol. 6(1). Pp. 30–39.
5. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2017). SMOTE: Synthetic Minority Oversampling Technique. *J. Artificial Intelligence Research*. Vol. 16. Pp. 321–357.
6. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks. arXiv:1406.2661 [cond-mat] (2014) (available at https://arxiv.org/abs/1406.2661).
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
8. Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010* (pp. 177-186).
9. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
10. Krizhevsky, A., Nair, V., & Hinton, G. (2014). The CIFAR-10 dataset. http://www.cs.toronto.edu/kriz/cifar.html.
11. Mery, D., Rio, V., Zscherpel, U., Mondragn, G., Lillo, I., Zuccar, I., ... Carrasco, M. (2015). GDXray: The database of X-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, 34(4), 42.

УДК 004.852

Олександр Чайковський, Артем Волокита, Кир'янов Артемій, Георгій Луцький

МЕТОД АУГМЕНТАЦІЇ ДАНИХ ІЗ ВИКОРИСТАННЯМ ГЕНЕРАТИВНИХ ЗМАГАЛЬНИХ МЕРЕЖ

Велика кількість різноманітних даних є основою для побудови надійної моделі машинного навчання, адже це допомагає точніше узагальнити інформацію про поставлену задачу. Водночас існують галузі, де збір даних або неможливий, або потребує величезної кількості ресурсів. Наприклад, у медичній галузі дані пацієнтів захищені законами про конфіденційність та приватність інформації, через які їх пошук, зберігання та використання викликає великі проблеми. Сучасні системи машинного навчання вирішують цю проблему методами генерації синтетичних даних. У статті розглянуто розробку методу аугментації даних на базі генеративних змагальних мереж.

Важливим етапом побудови точної моделі машинного навчання є пошук та анотація даних, які будуть використовуватися для навчання та тестування точності роботи нейронної мережі. Від кількості зібраних даних залежить точність та стабільність роботи мережі в реальних умовах. У статті пропонується метод для підвищення точності класифікації зображень згортковими нейронними мережами на базі аугментації даних із використанням генеративних змагальних мереж.

Нині добре описано та проаналізовано такі методи аугментації зображень, як поворот на деякий випадковий кут, стиснення та розтягнення по вертикалі й горизонталі, зміщення, дзеркальне відображення. Також наявні роботи, що розглядають внесення реалістичної деформації в зображення та генерації нових векторів ознак на базі декількох сусідніх зразків.

Нерозглянутими на даний момент залишаються можливості використання генеративних змагальних мереж у задачі аугментації даних.

Описано метод аугментації зображень із використанням генеративних змагальних мереж для підвищення точності роботи згорткових нейронних мереж, проведено порівняльний аналіз запропонованого методу з класичними методами аугментації зображень. Виділено основні переваги та недоліки запропонованого методу аугментації даних, висунуто плани щодо подальших досліджень.

Ключові слова: *аугментація даних; генеративні змагальні мережі; згорткові нейронні мережі.*

Рис.: 10. Табл.: 3. Бібл.: 11.

Chaikovskyi Oleksandr – PhD Student, Department of Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (37 Pobedy Av., 03056 Kyiv, Ukraine).
Чайковський Олександр Ігорович – аспірант, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (просп. Перемоги, 37, м. Київ, 03056, Україна).
E-mail: alex.programmr@gmail.com
ORCID: http://orcid.org/0000-0001-7451-1127

Volokyta Artem – PhD, Associate Professor, Department of Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (37 Pobedy Av., 03056 Kyiv, Ukraine).
Волокита Артем Миколайович – кандидат технічних наук, доцент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (просп. Перемоги, 37, м. Київ, 03056, Україна).
E-mail: artem.volokita@kpi.ua
ORCID: http://orcid.org/0000-0001-9069-5544
Scopus Author ID: 54421406500

Kyrianov Artemii – PhD Student, Department of Computer Engineering**,** National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (37 Pobedy Av., 03056 Kyiv, Ukraine).
Кир'янов Артемій Юрійович – аспірант, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (просп. Перемоги, 37, м. Київ, 03056, Україна).
E-mail: hunter953214@gmail.com
ORCID: http://orcid.org/0000-0003-3116-0122

Loutskii Heorhii – Doctor of Technical Sciences, Professor, Department of Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (37 Pobedy Av., 03056 Kyiv, Ukraine).
Луцький Георгій Михайлович – доктор технічних наук, професор, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (просп. Перемоги, 37, м. Київ, 03056, Україна).
E-mail: georgijluckij80@gmail.com
ORCID: http://orcid.org/0000-0002-3155-8301
Scopus Author ID: 16473143100

Chaikovskyi, O., Volokyta, A., Kyrianov, A., Loutskii, H. (2021). Data augmentation method using generative adversarial networks. *Technical sciences and technologies,* (2(24)), 83-91.