DOI: https://doi.org/10.15276/aait.08.2025.17
UDC 004.93

Methods for addressing labeled data scarcity in applied computer vision systems

Anatolii O. Nikolenko[1)]
ORCID: https://orcid.org/0000-0002-9849-1797; nikolenko@op.edu.ua
Kyrylo B. Mikhalev[1)]
ORCID: https://orcid.org/0009-0007-8058-4569; kirill.mikhalev.9035@gmail.com
[1)] Odesa Polytechnic National University. 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

This article systematizes approaches to addressing the shortage of labeled images in datasets for deep learning models, which constitute the core of modern computer vision applications. It covers methods such as transfer learning using pre-trained networks, weakly supervised and semi-automatic labeling techniques, and various sample expansion strategies through data augmentation. Particular attention is given to data augmentation as a widely applicable and technologically accessible approach. The article presents an experimental analysis of the impact of different augmentation methods, including classical transformations such as rotation, reflection, noise addition, and modern techniques like MixUp, CutMix, and AugMix, on the performance of convolutional neural networks in image classification tasks. Results demonstrate that even basic augmentation methods can substantially improve model generalization when training data are limited, and combining multiple strategies can approach the accuracy achievable with fully labeled datasets. In addition, the study explores the application of pseudo-labeling, a semi-automatic labeling method, with a focus on the RAF-DB dataset for facial expression recognition. By generating confident pseudo-labels for unlabeled samples, the model's performance improved by up to two percent in overall accuracy and up to ten percent in individual class recognition. This highlights the potential of combining augmentation with pseudo-labeling to enhance model robustness in scenarios with scarce labeled data. An additional experiment using a basic CNN on the CIFAR-10 dataset confirmed the positive effect of augmentation on classification accuracy. The employed transformations included horizontal flipping, random cropping, and color jittering, resulting in a afive percent increase in overall accuracy compared to models trained without augmentation. The findings indicate that integrating classical and advanced augmentation techniques with semi-automatic pseudo-labeling provides a practical and effective strategy for improving deep learning performance on limited datasets. This combined approach is particularly valuable for real-world applications where labeled data are scarce or costly to obtain, demonstrating a promising direction for further research and deployment in domains such as facial emotion recognition and general object classification.

Keywords: Machine learning; augmentation; convolution neural networks; labeled images; data generation; pseudolabeling

***For citation*: Nikolenko A. O., Mikhalev K. B. "Methods for addressing labeled data scarcity in applied computer vision systems". *Applied Aspects of Information Technology*. 2025; Vol.8 No.3: 263–273. DOI: https://doi.org/10.15276/aait.08.2025.17**

INTRODUCTION

Computer vision systems (CV) is today an integral part of numerous intelligent solutions in a wide variety of industries – from medicine and security to agricultural industry and autonomous transport.

The success of such systems is largely based on the use of deep learning, in particular convolutional neural networks (CNN), which have demonstrated outstanding results in image classification, segmentation, object detection, pose estimation, and other tasks.

However, the effectiveness of deep models largely depends on a large amount of high-quality labeled data that allows the model to learn to highlight relevant patterns in visual information.

Despite the availability of large public datasets such as ImageNet, COCO (**C**ommon **O**bjects in **C**ontex) or Open Images, most real-world applications cannot use this data directly due to domain specifics, the need for precise markup, or legal and ethical constraints.

For example, in medical diagnostics, image labeling requires the participation of highly qualified specialists, and in video surveillance systems, compliance with confidentiality rules.

In automated control fields, rare or atypical scenarios are often encountered, for which there is simply a lack of examples in available datasets. Therefore, the problem of lack of labeled images is one of the key obstacles to creating reliable, generalizable, and interpretable computer vision systems in real-world applications.

In response to this challenge, the research community has developed a number of approaches that allow for efficient training of models even with a limited number of labeled images. Among them, the following deserve special attention: retraining of models based on pre-trained neural networks.

The goal of this study is to investigate practical strategies for mitigating the shortage of labeled

© Nikolenko A., Mikhalev K., 2025

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0 /deed.uk)

images in computer vision applications. In particular, we focus on two directions. Data augmentation, which artificially increases the variability of the training set through transformations and semi-supervised learning with pseudo-labeling, which enables the exploitation of large amounts of unlabeled data.

Based on this focus, we formulate the following research question – to what extent can data augmentation and pseudolabeling improve the performance of convolutional neural networks.

LITERATURE REVIEW

Deep learning methods, in particular convolutional neural networks, have become the basis of modern approaches to image recognition and computer vision tasks. One of the first works that laid the foundation for CNN is the study of LeCun et al., where the LeNet-5 architecture was presented and its effectiveness in the task of recognizing handwritten characters was demonstrated [1]. The direction was further developed thanks to the works of Krizhevsky, where the deep network AlexNet was created for the ImageNet competition [2], as well as a general review by Schmidhuber, who systematized the key achievements in the field of deep learning [3].

The problem of knowledge transfer between different data sets and subject areas is of considerable interest. The classic review by Pan and Yang [4] was the starting point for research in the field of transfer learning. Further generalizations are given in the works of Weiss [5] and Zhuang [6], where different strategies of transfer learning, from inductive to multitasking, were described and their practical application in various application systems.

Particular attention in modern research is paid to data augmentation methods. The work of Taylor and Nitschke [7] summarized basic image transformation techniques and showed their effectiveness for improving CNN training. Similar results are given in the works of Cubuk, who presented the AutoAugment approach for automatic search of augmentation policies [8], as well as DeVries and Taylor, who proposed the Cutout method for regularization of deep networks [9].

Another important direction concerns semi-automatic and weakly supervised learning. The review by Van Engelen and Hoos [10] systematizes modern methods of semi-automatic learning, and the works of Chapelle [11] and Zhu and Goldberg [12] provide the classical theoretical foundations of this approach. Additionally, Zhou [13] summarized modern research in the field of weakly supervised learning, emphasizing practical aspects of its application.

A separate group of methods is the technique of using generative models for augmentation. Antoniou [14] proposed GAN networks for generating synthetic images that improve CNN training. Similar approaches were used in the study of Frid-Adar [15] where GANs were applied to create synthetic medical images, as well as in Bowles [16] who tested the capabilities of generative models for augmenting training datasets.

The effectiveness of classical and modern augmentation methods was analyzed in detail by Wang and Perez [17], who showed an increase in CNN accuracy due to different transformation techniques. Similar results were confirmed by Mikołajczyk and Grochowski [18], who investigated the role of augmentation for small datasets.

Semi-automatic labeling approaches, such as pseudolabeling, have gained considerable popularity. Lee [19] was one of the first to use pseudolabels for semi-automatic learning. Subsequent studies have developed this idea: Iscen et al. used label propagation in combination with deep learning, and Arazo [20] investigated the effects of pseudolabeling and confirmation bias in semi-automatic scenarios.

PURPOSE AND OBJECTIVES OF THE RESEARCH

The aim of this work is to systematize approaches to overcome the shortage of labeled images in datasets for deep learning of CNN models in applied computer vision systems. Special emphasis is placed on experimental research of augmentation methods and semi-automatic (weakly supervised) labeling, which allow to significantly expand and improve the quality of training datasets and thereby increase the accuracy of image classification using CNN models that are being developed.

To achieve this goal, it is necessary to solve the following main tasks:

1) to systematize modern approaches to overcoming the shortage of labeled images in datasets for deep learning of CNN models in applied computer vision systems;

2) to investigate the impact of semi-automatic and weakly supervised labeling methods on the quality of data classification using CNN models;

3) to investigate the impact of different methods of augmenting training samples on the accuracy of image classification using CNN models using the CIFAR-10 dataset;

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

4) to summarize the results and formulate conclusions regarding the effectiveness of the considered methods in the context of overcoming the limitations of small data sets.

REVIEW OF CURRENT APPROACHES TO OVERCOMING THE DEFICIT OF LABELED IMAGES

To systematize modern approaches to overcoming the shortage of labeled images in datasets for deep learning of CNN models, we will consider in detail a number of methods and technologies that allow us to compensate for the shortage of labeled data or reduce the need for them.

Among the most effective approaches, the following groups of methods can be distinguished: model retraining (Transfer Learning), synthetic image generation, semi-automatic and weakly supervised markup (Weak Supervision), sample expansion (Data Augmentation) identifier.

Transfer learning is one of the most popular approaches to overcome the lack of labeled images. Its idea is to use the knowledge acquired by the model on a large source dataset (e.g. ImageNet) to solve the target problem on a smaller specific dataset. Typically, retraining is done in two stages.

First, pre-trained weights are used to initialize the model. Then fine-tuning on the target dataset – full or partial (e.g., freezing low-level layers) [21].

Advantages of this approach: reduced need for a large number of labeled examples. Significantly faster learning process. Possibility of using "common" features (edge detectors, corner detectors), which the models have already learned to distinguish.

Transfer learning has proven itself well in areas such as medical diagnostics, satellite monitoring, and security systems. Modern frameworks (PyTorch, TensorFlow) provide easy integration of pre-trained models into the development pipeline.

Semi-automatic labeling involves using pre-trained models to automatically assign labels to new, previously unannotated images. The process typically begins with a model trained on a limited sample being used to predict classes or other annotations on a larger amount of unlabeled data.

These predictions can be taken as is or they can be further validated by a human expert, which allows both to increase the labeled sample and to control the quality of the labels. In cases where the labeling is completely transferred to the model without human intervention, it is said to be weakly supervised markup.

Pseudolabeling is the process of adding data from an unlabeled dataset that the model is most confident in to the training data and predicting a label for the unlabeled dataset. This method belongs to semi-supervised learning. This process can be iterative: after each training cycle, the model improves its classification ability, and the dataset with pseudo-labels becomes better.

There is also an approach where weak labels are not given by a neural network, but by simple heuristics or templates, for example, based on metadata, context analysis, or even through interactive tools for rapid labeling. It is important to note that weakly supervised labeling does not always guarantee high annotation accuracy, but its advantage is that it allows to significantly reduce the cost of preparing large amounts of data for training the model.

The Fig.1 shows a conceptual diagram of how pseudo-labeling works.

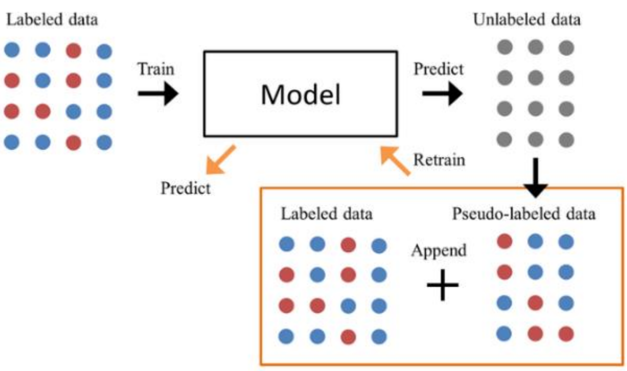

Fig. 1. **Pseudolabeling operation scheme**
***Source*: compiled by the [21]**

STUDYING THE IMPACT OF SEMI-AUTOMATIC AND WEAKLY SUPERVISED MARKING ON THE QUALITY OF DATA CLASSIFICATION IN CNN MODELS

Authors of the article [22] addressed a similar problem and attempted to address the lack of labeled images using a pseudolabeling approach. The challenge is that it is not enough to simply increase the amount of data, but also to increase the accuracy of the models that work with this data.

Let's conduct an experiment using the RAF-DB dataset. The Real-world Affective Faces Database (RAF-DB) is a large-scale facial expression database that contains about 30,000 different facial images downloaded from the Internet, including 29.672 real-world images. The images are divided into 7 classes (emotions) and a 7-dimensional expression distribution vector for each image (7 emotions in total). We will recognize emotions, which are:

anger, disgust, fear, happiness, sadness, surprise, neutrality (lack of emotion).

When developing a pseudolabeling method, the training part of the dataset, which contains 12271 images, should be divided into a training subset (Y_{train} = 8834), a validation subset (Y_{val} = 2209), and an "unlabeled" part ($Y_{unlabeled}$ = 1228).

During the pseudolabeling process, the class labels for images falling into $Y_{unlabeled}$ are simply ignored. This division is performed using the train_test_split function from the Scikit-learn library, using the stratify and random_state parameters.

The results of the RAF-DB data set partitioning for further pseudolabeling are illustrated in Fig. 2. It is important to note that the test part of the Test_labels set remains unchanged.

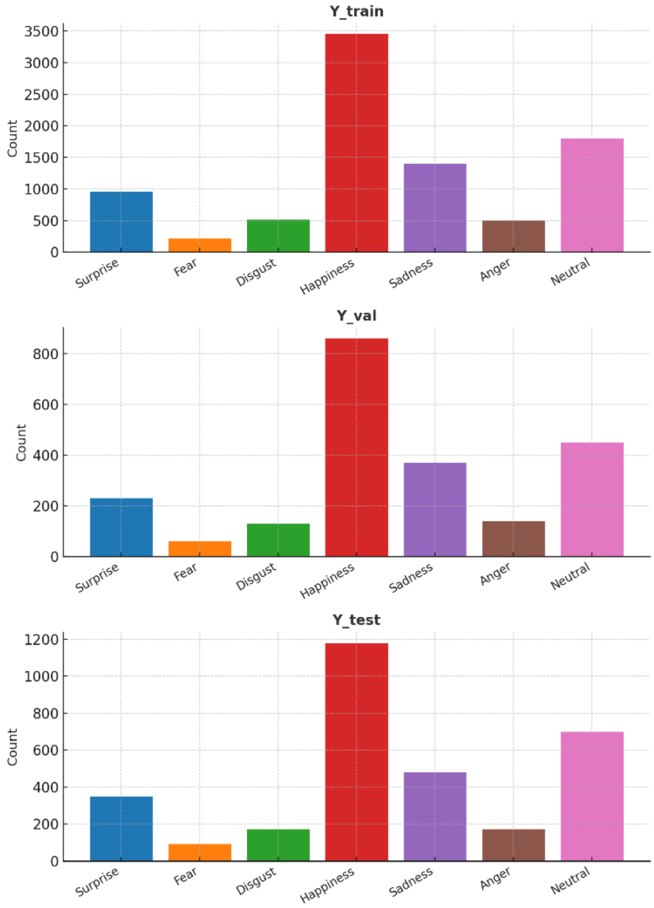

Fig. 2. **Distribution of images by classes and samples**
***Source:* compiled by the authors**

Pseudolabeling method is implemented using teacher (CNN-FER) and student (CNN (Retrain)-FER) models. It consists of four main steps:

1.Train the CNN-FER model on labeled images.

2.Use the trained CNN-FER model to create pseudo-labels on unlabeled images.

3. Add all confident predictions to the training data $Y_{unlabeled}$ with the prediction probability (P) for (y={0,1,2,3,4,5,6}) classes above a given threshold (t), P(y={0,1,2,3,4,5,6}|x) > t.

4.Train the CNN (Retrain)-FER model on a combination of ($Y_{train} + Y_{val} + Y_{unlabeled}$) samples.

The model is based on MobileNet V1 (pretrained on ImageNet). The base model is followed by GlobalAveragePooling2D, Dropout (0.3), and a dense output layer with softmax for seven classes. Training parameters: Adam optimizer with learning rate 0.0001, categorical cross-entropy loss, and metrics including accuracy and PR-AUC. Batch size is 64, and images are resized to 128×128×3. Each training stage runs for 10 epochs. The result of the program after the first and second iteration is shown in Fig. 3. As a result of the first execution of step 2 on images $Y_{unlabeled}$, 363 images ($Y_{pseudo\_labeled}$) were added to Y_{train}, the recognition accuracy of which is determined by the CNN-FER model with a probability of more than 0.99.According to step 4, the CNN (Retrain)-FER model has already been trained using Y_{train} + $Y_{pseudo\_labeled}$. After the second iteration of the pseudolabeling method, 95 images from $Y_{pseudo\_labeled}$ were added to Y_{train}. The training set is supplemented with those images in which the model is 90% confident. As can be seen from the results of the first iteration (Fig. 3), the CNN-FER model best identifies the emotions of happiness, anger, neutrality, and surprise. The overall classification accuracy is 76 %. Accuracy for particular classes are next: surprise – 71 %, fear – 64 %, disgust – 51%, happiness – 91 %, sadness – 72 %, anger – 64 %, neutrality – 66 %.

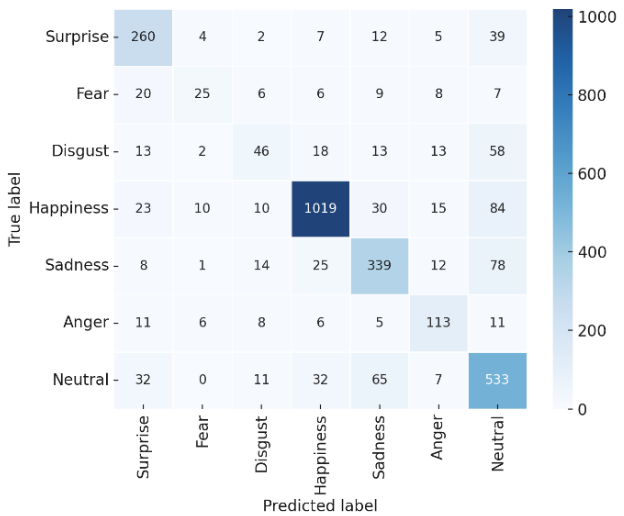

Fig. 3. **Confusion matrix before the first iteration**
***Source:* compiled by the authors**

In the second iteration (Fig. 4), the CNN (Retrain)-FER model recorded an improvement in accuracy for the emotions of surprise (from 71 to 77 %), fear (from 64 to 69 %), sadness (from 72 to 76 %), anger (from 64 to 74 %), and neutrality (from 66 to 71 %).

The recognition accuracy of disgust remained unchanged, while for the emotion of happiness, which is the most common in the training sample, it decreased (from 91 to 86 %).

As a result, general accuracy of the CNN (Retrain)-FER model increased by 2 percent compared to CNN-FER, reaching 78 %.

Accuracy for particular classes are next: surprise – 77 %, fear – 69 %, disgust – 51 %, happiness – 86 %, sadness – 76 %, anger – 74 %, neutrality – 71 %.

Thus, the general conclusion is as follows: the use of the developed data pseudolabeling method gives good results in overcoming such shortcomings of datasets for deep learning of convolutional neural networks as the lack of data of a certain type, class imbalance, insufficient data volume for deep learning, etc.

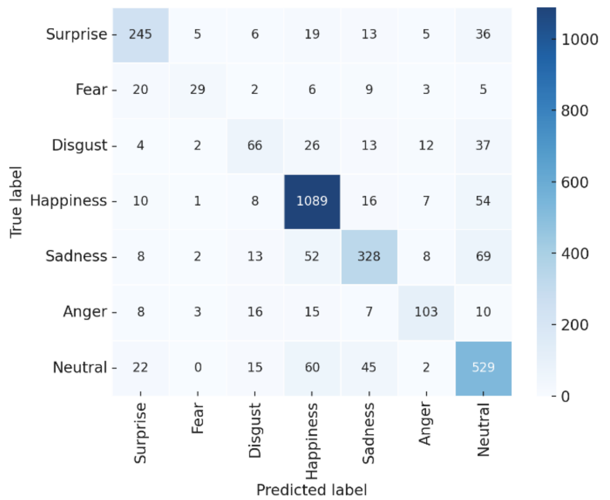

Fig. 4. **Confusion matrix after the first iteration**
Source: **compiled by the authors**

RESEARCH OF THE IMPACT OF DATA AUGMENTATION METHODS ON THE ACCURACY OF IMAGE CLASSIFICATION USING THE CIFAR-10 DATASET

Data Augmentation is one of the most fundamental and effective methods for dealing with the lack of labeled data in computer vision systems. This approach involves artificially increasing the size of the training sample by modifying existing images that preserve the main characteristics of the object, but at the same time introduce additional variable representation of the input data.

The special importance of data augmentation is useful in application areas where collecting new labeled examples is difficult or impossible, as well as in tasks where the model must remain robust to various changes in the external environment, in particular, noise, lighting variations, changes in perspective, etc. [23]. Augmentation is of particular importance in application areas where collecting new labeled data is difficult or impossible (e.g., medical imaging).

The basic idea is to modify the image x via a transformation function T(·), while preserving the corresponding label y:

$$(x,\ y) \to (T(x),\ y),\ T \in \mathrm{A}, \quad (1)$$

where A is a set permissible transformations .

Classical methods of augmentation: geometric transformations, rotation, scaling, shifts, horizontal or vertical reflection, color and light transformations: changing brightness, contrast, saturation of color shades, adding noise, blurring [24].

For example, the application of random rotation can be described as

$$T_{rot(x)} = R_\theta \cdot x, \quad (2)$$

where R_θ is the rotation matrix with angle θ.

Adding noise can be described as follows:

$$x' = x + \varepsilon,\ \ \varepsilon \sim N(0,\ \sigma^2), \quad (3)$$

where x is original image (pixel matrix), x' is augmented image, ε is random noise that is added, $N(0, \sigma^2)$ is normal (Gaussian) distribution with mean 0 and variance σ^2.

The color or brightness of a color can be described by the following formula:

$$x' = \alpha \cdot x + \beta, \quad (4)$$

where x is the original image (pixel value), x' is the new image after transformation, α is the contrast ratio, β is the brightness change (offset). For color images, this formula is applied to each channel (R, G, B) separately [25], [26].

Augmentation is an important step in training machine learning models. Data augmentation refers to the process of increasing the training data set by modifying existing data. Data augmentation has proven to be effective in image classification. However, there has been little research into the impact of augmentation on the accuracy of recognition models.

Given the resources available for image labeling, augmentation can sometimes be even more useful for object recognition tasks. Researchers have found that data augmentation methods for image classification can be useful for recognition. However, such methods provide limited gains in model accuracy.

There are some of the advanced Augmentation Methods:

- MixUp: creating new images by linearly combining a pair of images and mixing the labels accordingly;
- CutMix: pasting a fragment of one image into another with recalculation of labels based on the area of the pasted region;
- AugMix: stochastic combination of multiple base augmentations with regularization due to distortion invariance;
- MixUp can be described by the following formula

$$\tilde{x} = \lambda \cdot x_i + (1 - \lambda) \cdot x_j$$
$$\tilde{y} = \lambda \cdot y_i + (1 - \lambda) \cdot y_j, \quad (5)$$

where x_i, x_j are images from the sample, y_i, $y_j$are their labels in the format of one-hot vectors, $\lambda \in$ [0, 1] is a coefficient that determines the "weight" of each image, λ usually represents the interpolation weight between two images and α > 0, is a hyperparameter of the distribution that determines the variability of λ.

Where α>0 is a hyperparameter that controls how much the data is mixed. If α→0, then λ is close to 0 or 1, and the examples are almost not mixed. If $\alpha \to 1$, then the examples are mixed approximately equally.

CutMix: replacing a random region of one image with another is described as follows (formula for the image itself and the new label):

$$\tilde{x} = M \cdot x_i + (1 - M) \cdot x_j$$
$$\tilde{y} = \lambda \cdot y_i + (1 - \lambda) \cdot y_j, \quad (6)$$

where x_i is the base image, x_j is the second image that is partially inserted, M is a binary mask (a matrix of 0 and 1) that determines where to leave pixels from x_i, and where to replace them with x_j. $y_{i,}$ and $y_j$are the class labels for the original and second im ages, respectively. λ is the ratio of the saved area of the image $x_i$to the total area of the image.

One of the main advantages of augmentation is its versatility and ease of implementation. Most modern deep learning frameworks (such as PyTorch or TensorFlow) have ready-made libraries for implementing basic augmentation methods , and specialized tools such as Albumentations or imgaug allow you to easily add even the most complex types of transformations to the training pipeline [27].

In addition, augmentation is an effective tool in combating overfitting. Due to the constant variation of input data, the model is not able to simply "memorize" training examples, but is forced to learn to detect deep patterns that are resistant to change.

However, it is important to remember that augmentation is not a one-size-fits-all solution. In tasks where every detail of an image is critical (e.g., medical segmentation), applying certain transformations can cause distortion of important features.

In cases of a sharp change in the subject domain, when the training data is radically different from the one on which the model will operate in real conditions, simple augmentations may not be enough. In such situations, it is necessary to combine Data Augmentation with other approaches, including weakly supervised markup or synthetic image generation [28], [29].

The main goal of the experiment is to practically evaluate the effectiveness of various data augmentation methods in image classification tasks using convolutional neural networks. In particular, it is investigated how the application of basic and advanced augmentation techniques affects the accuracy of the model with a limited amount of labeled data. As the experimental dataset, which contains 60.000 color images of 32x32 pixels, divided into 10 classes (plane, car, bird, cat, deer, dog, frog, horse, ship, truck). To simulate the data shortage conditions, only a part of the full set was used in training (e.g., 10-20 % of the training samples).

As a basic architecture, a classical convolutional neural network with three convolutional layers and two fully connected layers in the output part was chosen. In this experiment, such augmentation approaches as Random Horizontal Flip, RandomCrop (random cropping), ColorJitter (Color distortion) were used.

An experimental study of the impact of data augmentation on the performance of a convolutional neural network showed a number of characteristic patterns that correspond to theoretical expectations regarding the behavior of models when using different sample expansion strategies.

Fig. 5 shows the loss graphs of the model (blue curve – model with augmentation, yellow – without).

In the initial stages of training (the first 3-4 epochs), the model trained without augmentation demonstrated higher accuracy in classifying images

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

on the test set. This is explained by the fact that in the absence of augmentation, the model receives "clean" and predictable training examples that it can relatively quickly "remember" and begin to correctly classify similar images on the test set.

In conditions of limited data volume, this behavior is a typical sign of initial overfitting – the model adapts to the specific details of the training set, but its ability to generalize remains low. The accuracy graphs are shown in Fig. 6.

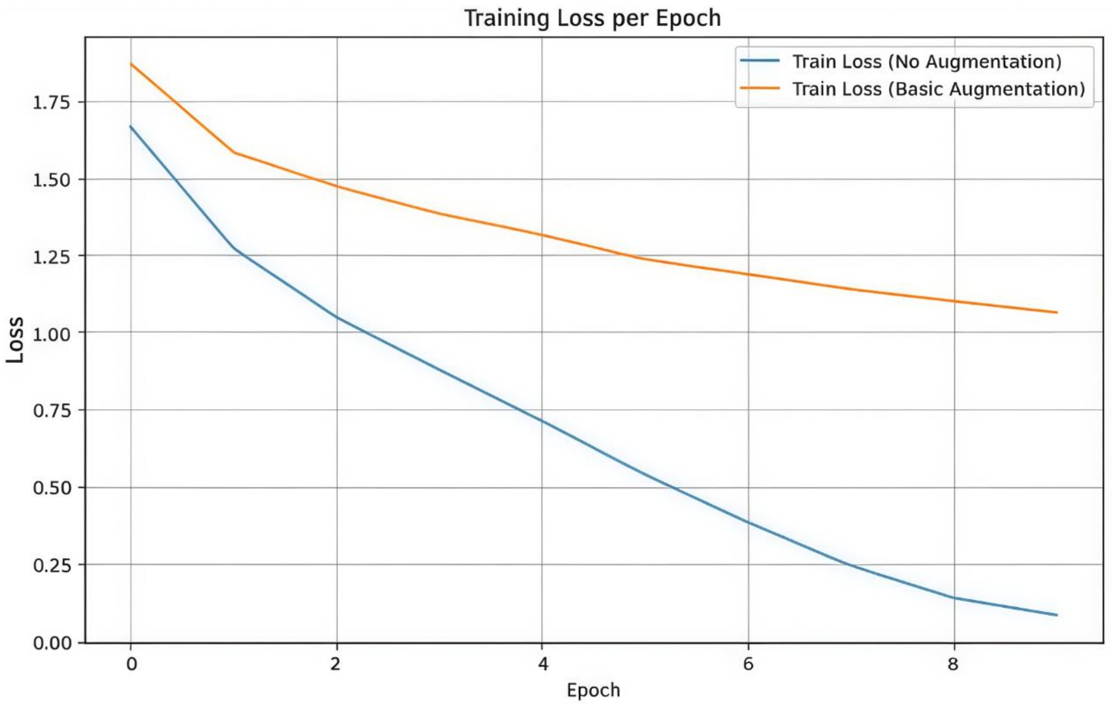

Fig. 5. **Graph of losses for both models**
***Source:* compiled by the authors**

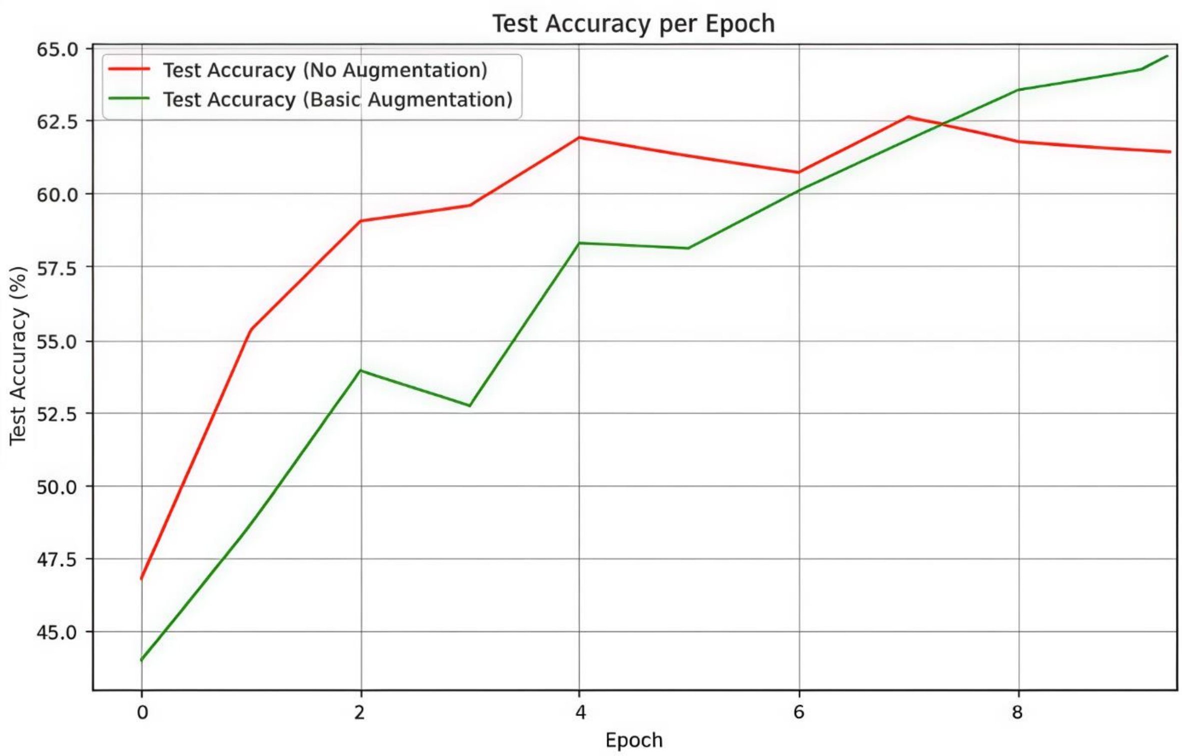

Fig. 6. **Graph of accuracy for both models**
***Source:* compiled by the authors**

To simulate a limited-data scenario, only 10,000 images were randomly selected from the original 50,000 training examples, while the full 10,000-image test set was used for evaluation.

A lightweight convolutional neural network (CNN) was implemented, consisting of two convolutional layers (3×3 kernels, ReLU activations, batch normalization, and max-pooling), followed by a fully connected layer with 512 units and dropout (p = 0.3), and a final classification layer with 10 outputs.

Training was carried out for 10 epochs using the Adam optimizer with an initial learning rate of 0.001. The batch size was set to 64, and cross-entropy loss was used as the training objective.

In contrast, in the baseline augmentation scenario (random reflection, cropping, color shifts), the model faced a much more complex training environment. The introduction of noise, geometric, and color distortions into the training samples complicates the task for the model, as it can no longer rely on fine-grained features of the images and is forced to look for more robust and generalized features for classification. As a result, in the first epochs, the accuracy of the augmented model was lower than that of the baseline.

After 4-5 epochs of training, the accuracy curve of the augmented model began to catch up, and in the final epochs, even surpass the result of the model without augmentation. This indicates the gradual formation of more stable and generalized features, which allowed the model to more correctly classify previously unseen images on the test sample.

This dynamic is direct evidence of the positive impact of augmentation on the generalization ability of the model. The final accuracy of the model trained with augmentation methods is 5% higher than the accuracy of the model without augmentation.

It is also worth paying attention to the shape of the learning loss curves. The model without augmentation showed a sharp drop in losses in the first epochs, but later its losses stabilized at a level indicating the beginning of retraining. At the same time, the model with augmentation had a smoother decrease in losses, and the fluctuations in loss were less pronounced, which also indicates better resistance to retraining.

In the original technical report introducing the CIFAR-10 and CIFAR-100 datasets, Krizhevsky, Hinton, and Nair [30] presented benchmark results obtained with a variety of classical algorithms. Specifically, they evaluated k-nearest neighbors (kNN), support vector machines (SVM), decision trees, and shallow convolutional neural networks (CNNs). Their findings showed that even for a relatively simple task – classifying 10 categories of 32×32 color images – achieving high accuracy was challenging. Reported results were generally in the range of 60 %, which highlighted the limitations of traditional machine learning approaches as well as the restricted representational power of early neural network architectures.

It is important to note that at that time, techniques such as data augmentation or semi-automatic dataset expansion were rarely used. Models were trained directly on the raw data, which significantly reduced their generalization capability and made them prone to overfitting. In comparison, our experiment demonstrates that even a relatively simple CNN architecture, when combined with basic data augmentation techniques such as random cropping, horizontal flipping, brightness and contrast adjustments, can achieve notable improvements. Despite training on a deliberately limited subset of CIFAR-10 (10,000 images instead of the full training set), the augmented model displayed higher stability during training and achieved test accuracy that surpassed the baseline numbers reported by Krizhevsky.

The methods presented in this work are most effective when applied to moderately sized labeled where classes that are reasonably balanced. Simple CNNs or ResNet18 architectures show good performance under these conditions. Performance may degrade on datasets with highly imbalanced classes. The experiments do not cover very deep transformer models or high-resolution real-world datasets, so the findings should be interpreted within these boundaries.

Practical Application and Future Extensions. Augmentation strategies discussed in this paper can be combined with semi-supervised learning techniques such as pseudo-labeling, FixMatch, or Mean Teacher to further improve performance on limited labeled datasets. Additionally, generative models like GANs can be used to synthetically expand the dataset, complementing classical augmentations. This combined approach allows more robust training while minimizing manual labeling effort.

Computational considerations should guide the choice of architecture and augmentation complexity: start with lightweight CNNs or ResNet18, and scale to larger models or more extensive augmentations after initial experiments confirm benefits.

Overall, the experimental results confirmed that the use of data augmentation and pseudolabeling is

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

an effective strategy for improving model quality in the context of a limited sample of labeled images.

CONCLUSION

The paper considered modern methods for overcoming the lack of labeled data in computer vision tasks, in particular, methods of augmentation and semi-automatic labeling. It is shown that data augmentation remains one of the simplest and most effective ways to artificially increase the volume of the training sample, ensuring the stability of models to changes in the external environment and reducing the risk of overtraining. The use of such approaches as geometric transformations, changing color characteristics, adding noise or more advanced techniques (MixUp, CutMix, AugMix) allows the model to learn on more variable data, which increases its generalization ability.

Special attention is paid to semi-automatic labeling methods, among which pseudolabeling occupies an important place. This approach involves using a pre-trained model to automatically generate labels for unknown examples, after which the resulting pseudolabels are used for further training. This can significantly reduce the need for manual labeling of large volumes of data, which is especially relevant in areas with high cost or complexity of obtaining labeled examples, such as medical imaging or remote sensing.

An experiment using CNN on the CIFAR-10 dataset demonstrated the positive impact of augmentation on classification quality. The results confirmed that even basic image transformation methods contribute to improving model accuracy compared to training on “clean” data.

When datasets are very small, consider GANs or other generative approaches to synthetically enlarge the dataset alongside classical augmentations, improving model robustness.

This suggests that combining augmentation with semi-automatic labeling (in particular, pseudolabeling) may be a promising direction for further research and practical application in conditions of limited resources for data collection.

Thus, the combined use of different approaches – from classical Data Augmentation methods to modern semi-automatic markup strategies – opens up new opportunities for building more accurate, robust, and efficient computer vision systems.

REFERENCES

1. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. “Gradient-Based learning applied to document recognition”. *Proceedings of the IEEE*. 1998; 86 (11): 2278–2324. DOI: https://doi.org/10.1109/5.726791.

2. Krizhevsky, A., Sutskever, I. & Hinton, G. E. “ImageNet classification with deep convolutional neural networks”. *Advances in Neural Information Processing Systems*. 2012; 25: 1097–1105. DOI: https://doi.org/10.1145/3065386.

3. Schmidhuber, J. “Deep learning in neural networks: An overview”. *Neural Networks*. 2015; 61: 85–117. DOI: https://doi.org/10.1016/j.neunet.2014.09.003.

4. Pan, S. J. & Yang, Q. “A survey on transfer learning”. *IEEE Transactions on Knowledge and Data Engineering*. 2010; 22 (10): 1345–1359. DOI: https://doi.org/10.1109/TKDE.2009.191.

5. Weiss, K., Khoshgoftaar, T. M. & Wang, D. “A survey of transfer learning”. *Journal of Big Data*. 2016; 3 (1): 9. DOI: https://doi.org/10.1186/s40537-016-0043-6.

6. Zhuang, F., Qi, Z., Duan, K., et al. “A comprehensive survey on transfer learning”. *Proceedings of the IEEE*. 2021; 109 (1): 43–76. DOI: https://doi.org/10.1109/JPROC.2020.3004555.

7. Taylor, L. & Nitschke, G. “Improving deep learning with generic data augmentation”. *arXiv preprint*. 2017. DOI: https://doi.org/10.48550/arXiv.1708.06020.

8. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V. & Le, Q. V. “AutoAugment: Learning augmentation policies from data”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019. p. 113–123. DOI: https://doi.org/10.1109/CVPR.2019.00020.

9. DeVries, T. & Taylor, G. W. “Improved regularization of convolutional neural networks with cutout”. *arXiv preprint*. 2017. DOI: https://doi.org/10.48550/arXiv.1708.04552.

10. Van Engelen, J. E. & Hoos, H. H. “A survey on semi-supervised learning”. *Machine Learning*. 2020; 109: 373–440. DOI: https://doi.org/10.1007/s10994-019-05855-6.

11. Chapelle, O., Scholkopf, B. & Zien, A. “Semi-Supervised learning”. *MIT Press*. 2006. DOI: https://doi.org/10.7551/mitpress/9780262033589.001.0001

12. Zhu, X. & Goldberg, A. B. "Introduction to semi-supervised learning". *Synthesis Lectures on Artificial Intelligence and Machine Learning*. 2009; 3 (1): 1–130. DOI: https://doi.org/10.2200/S00196ED1V01Y200906AIM006.

13. Zhou, Z.-H. "A brief survey on weakly supervised learning". *National Science Review*. 2018; 5 (1): 44–53. DOI: https://doi.org/10.1093/nsr/nwx105.

14. Antoniou, A., Storkey, A. & Edwards, H. "Data augmentation generative adversarial networks". *arXiv preprint*. 2017. DOI: https://doi.org/10.48550/arXiv.1711.04340.

15. Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J. & Greenspan, H. "Synthetic data augmentation using GAN for improved liver lesion classification". *IEEE Transactions on Medical Imaging*. 2018; 38 (3): 675–685. DOI: https://doi.org/10.1109/TMI.2018.2864810.

16. Bowles, C., Chen, L., Guerrero, R., et al. "GAN augmentation: Augmenting training data using generative adversarial networks". *arXiv preprint*. 2018. DOI: https://doi.org/10.48550/arXiv.1810.10863.

17. Wang, J. & Perez, L. "The effectiveness of data augmentation in image classification using deep learning". *arXiv preprint*. 2017. DOI: https://doi.org/10.48550/arXiv.1712.04621.

18. Mikołajczyk, A. & Grochowski, M. "Data augmentation for improving deep learning in image classification problem". *2018 International Interdisciplinary PhD Workshop (IIPhDW)*. IEEE. 2018. p. 117–122. DOI: https://doi.org/10.1109/IIPHDW.2018.8388338.

19. Lee, D.-H. "Pseudo-Label: The simple and efficient semi-supervised learning method for deep neural networks". *International Conference on Machine Learning Workshop*. 2013. DOI: https://doi.org/10.1103/PhysRevLett.110.262002.

20. Iscen, A., Tolias, G., Avrithis, Y. & Chum, O. "Label propagation for deep semi-supervised learning". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019. p. 5070–5079. DOI: https://doi.org/10.1109/CVPR.2019.00520.

21. Arazo, E., Ortego, D., O'Connor, N. E. & McGuinness, K. "Pseudo-Labeling and confirmation bias in deep semi-supervised learning". *International Joint Conference on Neural Networks (IJCNN)*. 2020. p. 1–8. DOI: https://doi.org/10.1109/IJCNN48605.2020.9207304.

22. Arsirii, O. O. & Petrosiuk, D. V. "Pseudo-Labeling of transfer learning convolutional neural network data for human facial emotion recognition". *Herald of Advanced Information Technology*. Odesa, Ukraine. 2023; 6 (3): 203–214. DOI: https://doi.org/10.15276/hait.06.2023.13.

23. Shorten, C. & Khoshgoftaar, T. M. "A survey on image data augmentation for deep learning". *Journal of Big Data*. 2019; 6 (1): 60. DOI: https://doi.org/10.1186/s40537-019-0197-0.

24. Feng, S. & Wang, X. "A review of data augmentation approaches for computer vision". *Journal of Visual Communication and Image Representation*. 2021; 74: 103021. DOI: https://doi.org/10.1016/j.jvcir.2021.103021.

25. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L. & Haworth, A. "A review of medical image data augmentation techniques for deep learning applications". *Journal of Medical Imaging and Radiation Oncology*. 2021; 65 (5): 545–563. DOI: https://doi.org/10.1111/1754-9485.13261.

26. Kumar, T. "Image data augmentation approaches". *arXiv preprint*. 2023. DOI: https://doi.org/10.48550/arXiv.2301.02830.

27. Shorten, C. & Khoshgoftaar, T. M. "Deep learning image augmentation techniques: A Practical overview". *Journal of Big Data*. 2020; 7 (1): 1–21. DOI: https://doi.org/10.1186/s40537-020-00385-0.

Cubuk, E. D., Zoph, B., Shlens, J. & Le, Q. V. "RandAugment: Practical automated data augmentation with a reduced search space". *arXiv preprint*. 2019. DOI: https://doi.org/10.48550/arXiv.1909.13719.

28. Lim, S., Kim, I., Kim, T., Kim, C. & Kim, S. "Fast AutoAugment". *Advances in Neural Information Processing Systems (NeurIPS)*. 2019. p. 6665–6675. DOI: https://doi.org/10.5555/3454287.3454885.

29. Krizhevsky, A., Nair, V. & Hinton, G. "CIFAR-10 and CIFAR-100 Datasets". *The Toronto University Technical Report*. 2009. DOI: https://www.cs.toronto.edu/~kriz/cifar.html.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 16.07.2025
Received after revision 10.09.2025
Accepted 23.09.2025

DOI: https://doi.org/10.15276/aait.08.2025.17
УДК 004.93

Методи подолання нестачі маркованих зображень при побудові прикладних систем комп'ютерного зору

Ніколенко Анатолій Олександрович[1)]
ORCID: https://orcid.org/0000-0002-9849-1797; anatolyn@ukr.net
Міхалев Кирило Богданович[1)]
ORCID: https://orcid.org/0009-0007-8058-4569; kirill.mikhalev.9035@gmail.com
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

У статті систематизуються підходи до вирішення проблеми нестачі маркованих зображень у наборах даних для моделей глибокого навчання, які складають основу сучасних програм комп'ютерного зору. Вона охоплює такі методи, як трансферне навчання з використанням попередньо навчених мереж, слабо контрольовані та напівавтоматичні методи маркування, а також різні стратегії розширення вибірки шляхом доповнення даних. Особлива увага приділяється доповненню даних як широко застосовному та технологічно доступному підходу. У статті представлено експериментальний аналіз впливу різних методів доповнення, включаючи класичні перетворення, такі як обертання, відбиття, додавання шуму, та сучасні методи, такі як MixUp, CutMix та AugMix, на продуктивність згорткових нейронних мереж у завданнях класифікації зображень. Результати показують, що навіть базові методи доповнення можуть суттєво покращити узагальнення моделі, коли навчальні дані обмежені, а поєднання кількох стратегій може наблизитися до точності, досяжної з повністю маркованими наборами даних. Крім того, дослідження досліджує застосування псевдомаркування, напівавтоматичного методу маркування, з акцентом на наборі даних RAF-DB для розпізнавання виразів обличчя. Завдяки генерації впевнених псевдоміток для немаркованих зразків, продуктивність моделі покращилася на два відсотка за загальною точністю та до десяти відсотків за розпізнаванням окремих класів. Це підкреслює потенціал поєднання доповнення з псевдомаркуванням для підвищення стійкості моделі в сценаріях з обмеженими маркованими даними. Додатковий експеримент з використанням базової CNN на наборі даних CIFAR-10 підтвердив позитивний вплив доповнення на точність класифікації. Використані перетворення включали горизонтальне перевертання, випадкове кадрування та кольорове тремтіння, що призвело до збільшення загальної точності на п'ять відсотків порівняно з моделями, навченими без доповнення. Результати показують, що інтеграція класичних та передових методів доповнення з напівавтоматичним псевдомаркуванням забезпечує практичну та ефективну стратегію для покращення продуктивності глибокого навчання на обмежених наборах даних. Цей комбінований підхід особливо цінний для реальних застосувань, де марковані дані є обмеженими або дорогими для отримання, демонструючи перспективний напрямок для подальших досліджень та впровадження в таких областях, як розпізнавання емоцій обличчя та загальна класифікація об'єктів.

Ключові слова: Машинне навчання; аугментація; згорткові нейронні мережі; марковані зображення; генерація даних; псевдомаркування

ABOUT THE AUTHORS

Anatolii O. Nikolenko - PhD, Associate Professor, Department of information systems. Odesa Polytechnic National University. 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-9849-1797; nikolenko@op.edu.ua
***Research field*:** Information technology; artificial intelligence; decision support systems; machine learning; neural networks

Ніколенко Анатолій Олександрович - кандидат технічних наук, доцент кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Kyrylo B. Mikhalev - Master, Department of Information Systems. Odesa Polytechnic National University. 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0009-0007-8058-4569; kirill.mikhalev.9035@gmail.com
***Research field*:** data science, image recognition systems, artificial intelligence

Міхалев Кирило Богданович - магістр кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна