Training Small Neural Networks on Limited Data: Regularization and Augmentation Methods

Research Work

December 12, 2025

Abstract

This work addresses the problem of training small neural networks with limited training data. The research focuses on applying various regularization and data augmentation methods to improve the generalization ability of models. Experiments were conducted using convolutional neural networks on datasets with limited number of examples. Methods such as dropout, batch normalization, weight decay, as well as various image augmentation techniques are considered. Results show that a combination of regularization and augmentation methods allows achieving significant improvement in classification accuracy even when working with small datasets. The average accuracy improvement was about 12-15% compared to baseline models without applying special techniques.

Keywords: neural networks, machine learning, regularization, data augmentation, overfitting, small datasets

1 Introduction

In recent years, deep learning has shown impressive results in various fields, from computer vision to natural language processing. However, most successful models require enormous amounts of data for training. In real-world tasks, there often arises a situation where limited amounts of labeled data are available, which creates serious problems for training deep neural networks.

This problem becomes especially relevant when working with small neural networks, which may be more suitable for devices with limited computational resources, such as mobile devices or embedded systems. In such cases, classical training approaches often lead to overfitting of the model on the training set, which significantly reduces its ability to generalize to new data.

In this work, we investigate various regularization and data augmentation methods that can help improve the training of small neural networks when working with limited datasets. We consider both classical methods, such as dropout and weight decay, as well as more modern approaches to data augmentation.

The goal of the research is to determine the most effective combinations of regularization and augmentation methods for training small neural networks on limited data. The practical significance of the work lies in the possibility of applying the obtained results to tasks where collecting large volumes of data is difficult or impossible.

2 Literature Review

The problem of training neural networks on limited data is actively studied in the scientific community. One of the classical approaches to solving this problem is the application of regularization methods. Srivastava et al. [1] proposed the dropout method, which randomly disables part of neurons during training, helping to prevent overfitting. This method has shown its effectiveness in various tasks.

Batch normalization, proposed by Ioffe and Szegedy [2], is also an important regularization method that normalizes the inputs of each layer and can improve training stability and model generalization ability.

Regarding data augmentation, there are many techniques for different types of data. For images, methods such as random rotations, reflections, brightness and contrast changes are widely used [3]. More modern approaches, such as Mixup [4] and Cutout [5], also show good results.

However, most studies focus on large models and datasets. The question of the effectiveness of these methods specifically for small networks and limited data is insufficiently studied, which motivates our research.

3 Methods

3.1 Model Architecture

For experiments, we used a small convolutional neural network consisting of three convolutional blocks. Each block includes a convolutional layer, a batch normalization layer, a ReLU activation function, and a max pooling layer. After the convolutional layers, there is a fully connected layer with dropout and a final classification layer. The total number of model parameters is approximately 50,000, making it compact enough for use on devices with limited resources.

3.2 Regularization Methods

The following regularization methods were investigated in the work:

Dropout: Applied with a neuron dropout probability of 0.3-0.5 after fully connected layers. This helps prevent neuron co-adaptation and improve generalization.

Batch Normalization: Used after each convolutional layer to normalize activations and stabilize the training process.

Weight Decay (L2 regularization): Applied with a coefficient of 10^{-4} to limit the magnitude of model weights and prevent overfitting.

Early Stopping: Training was stopped if validation accuracy did not improve for 10 epochs.

3.3 Data Augmentation Methods

The following techniques were used for image augmentation:

- Random rotations at angles from -15 to +15 degrees
- Random horizontal and vertical reflections

- Brightness changes in the range of $\pm 20\%$
- Contrast changes in the range of $\pm 15\%$
- Random scaling from 0.9 to 1.1
- Random horizontal and vertical shifts up to 10% of the image size

A combination of these methods with more modern approaches, such as random image region cutting (cutout) and image mixing (mixup), was also investigated.

3.4 Experimental Setup

Experiments were conducted on the CIFAR-10 dataset, but with limited data. To simulate a situation with limited data, we used only 10% of the original training set, which is approximately 5,000 images. The test set remained full (10,000 images).

The model was trained using the Adam optimizer with an initial learning rate of 0.001 and a 2-fold reduction every 20 epochs. The batch size was 32. All experiments were conducted using the PyTorch framework.

4 Results

The conducted experiments showed a significant influence of regularization and augmentation methods on the quality of training small neural networks on limited data.

4.1 Baseline Results

The baseline model without applying special regularization and augmentation methods showed an accuracy on the test set of about 62%. At the same time, clear overfitting was observed: accuracy on the training set reached 95%, while on the test set it remained significantly lower.

4.2 Influence of Individual Methods

Applying only dropout with a probability of 0.4 improved accuracy to 68%. Batch normalization separately gave an improvement to 65%. Weight decay showed a less pronounced effect, improving accuracy to 64%.

Data augmentation without other regularization methods allowed achieving an accuracy of 71%, which is a significant improvement compared to the baseline model.

4.3 Combined Approaches

The most effective was the combination of all regularization methods together with data augmentation. This approach allowed achieving an accuracy of 74% on the test set, which is 12% higher than the baseline model. At the same time, the difference between accuracy on training and test sets decreased to 8%, indicating a significant reduction in overfitting.

Additional application of mixup and cutout techniques allowed improving the result by another 1-2%, achieving a final accuracy of about 75-76%.

4.4 Analysis of Data Volume Influence

An experiment was also conducted with different volumes of training data (5%, 10%, 20% of the original dataset). Results showed that the effectiveness of regularization and augmentation methods increases with decreasing data volume. When using only 5% of data (2,500 images), the combined approach allowed achieving an accuracy of 68%, while the baseline model showed only 52%.

5 Discussion

The obtained results demonstrate the importance of applying regularization and augmentation methods when training small neural networks on limited data. Particularly interesting is the fact that data augmentation turned out to be more effective than individual regularization methods, which may be related to the fact that it not only prevents overfitting but also actually increases the diversity of training data.

The combination of various methods showed a synergistic effect, indicating that these approaches complement each other. Dropout helps prevent neuron co-adaptation, batch normalization stabilizes training, and augmentation increases data diversity.

However, it should be noted that there are some limitations to the study. Experiments were conducted only on one type of data (images) and one network architecture. For more general conclusions, it is necessary to conduct studies on various types of data and architectures. In addition, optimal parameters of regularization methods may depend on the specific task and data volume.

It should also be noted that applying all methods simultaneously increases training time and computational costs, although for small networks this increase is not critical.

6 Conclusion

This work conducted a study of the effectiveness of regularization and data augmentation methods for training small neural networks on limited data. Results show that the combined application of these methods allows significantly improving the generalization ability of models and reducing overfitting.

Main conclusions of the work:

1. Data augmentation is the most effective individual method for improving training on limited data
2. The combination of regularization methods (dropout, batch normalization, weight decay) with data augmentation gives a synergistic effect
3. The effectiveness of methods increases with decreasing training data volume
4. Applying the proposed methods allows achieving an accuracy improvement of 12-15% compared to baseline approaches

The obtained results can be useful in solving practical tasks where collecting large volumes of data is difficult, but effective operation of neural networks is required. In the future, it is planned to expand the research to other types of data and network architectures, as well as to study the influence of more modern augmentation methods, such as AutoAugment and RandAugment.

7 References

References

[1] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[2] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, PMLR, 2015, pp. 448–456.

[3] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[4] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[5] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[9] U. Musaev and A. Khan, "Optimizing ML Training with Perturbed Equations," in *2025 6th International Conference on Problems of Cybernetics and Informatics (PCI)*, Baku, Azerbaijan, 2025, pp. 1-6, doi: 10.1109/PCI66488.2025.11219819.