Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/54/20241659

Enhancing Convolutional Neural Networks via separately trained kernels for digit recognition

Zhaochen Yang

Department of Computing, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong, 999077, China

yangzhaochen2017@163.com

Abstract. This paper introduces a novel methodology for pre-training neural networks. Instead of the traditional approach of running a single classifier for all objects, the method used in the research constructs individual classifiers for each object to extract unique features. These classifiers, in essence, act as a series of binary classifiers, each pre-training a distinct set of convolutional kernels. These individually trained kernels are then combined and processed by a comprehensive classifier. The methodology leverages the power of individual feature extraction and collaborative processing to enhance the overall performance of the classifier. The results demonstrate that this innovative pre-training approach leads to a significant improvement in classification performance, yielding higher accuracy compared to conventional methods. It underscores the benefits of incorporating individual object feature extraction and combined processing in the pre-training phase of neural network-based classification tasks. The study opens new paths for improving pre-training methods in neural networks, with potential applications in various fields that require high-accuracy object recognition. Future work will delve deeper into the potential challenges associated with model complexity and overfitting, and will also include a more comprehensive evaluation using extensive training, cross-validation, and independent test sets. This will further validate the effectiveness of our proposed pre-training methodology.

Keywords: Convolutional Neural Networks, Pre-trained Convolutional Kernels, Transfer Learning, Image Classification, Model Optimization

1. Introduction

Deep learning, a subset of machine learning, has revolutionized various fields across science and technology, including image recognition, natural language processing, medical diagnosis, and autonomous driving. Neural networks, the backbone of deep learning, have displayed remarkable capabilities in handling complex tasks. Despite their widespread adoption, enhancing the training efficiency and accuracy of these networks remains a significant challenge. Traditional methodologies typically employ a single classifier for all objects during the training phase, which may not fully capture the unique features of each object.

In light of this, we propose a novel pre-training methodology for neural networks that aims to address these challenges. Our approach diverges from conventional methods by constructing individual classifiers for each object, effectively creating a series of binary classifiers. Each of these classifiers is

DOI: 10.54254/2755-2721/54/20241659

responsible for pre-training a distinct set of convolutional kernels, which are fundamental components for feature extraction in neural networks. This individualized approach allows for more nuanced and detailed feature extraction, which is essential for efficient object recognition.

Once pre-trained, these individually trained kernels are then brought together and processed by a comprehensive classifier. The integration of individual feature extraction with combined processing capitalizes on the strengths of both strategies, leading to enhanced overall performance.

This paper is organized as follows: We first provide a detailed description of our proposed pre-training methodology, followed by a comparative analysis demonstrating the superior performance of our approach compared to traditional methodologies. We further discuss potential applications and implications of our work in various fields that require precise object recognition. Finally, we outline potential challenges related to model complexity and overfitting, and set the direction for future research, including a more comprehensive evaluation of our proposed methodology using extensive training, cross-validation, and independent test sets.

Understanding the limitations of the traditional approach, this study proposes a novel methodology that constructs individual classifiers for each object to extract distinctive features. These features are then amalgamated and processed by an overarching classifier. This approach aims to enhance the classification performance by leveraging the individual characteristics of each object.

This paper is organized as follows: the paper first reviews the related work on neural network classification tasks. Then, we introduce our method of building separate classifiers for each object, extracting unique features, and combining them for the overall classification. We present and discuss our experimental results, which demonstrate the superior performance of our proposed model. Finally, we conclude with a summary of our findings and suggestions for future work.

2. Literature Review

The role of pre-training in neural networks has been the subject of extensive research over the past decade. These studies have recognized the critical importance of pre-training in enhancing the performance of neural networks and have proposed various methods to improve its effectiveness.

A significant body of work focuses on the use of a single classifier for all objects during the pre-training stages. This approach, while widely adopted, has inherent limitations. For instance, Hinton et al. demonstrated that a one-size-fits-all classifier may not fully capture the unique features of each object [1]. This can lead to sub-optimal performance in tasks requiring precise object recognition.

To address this, some researchers have explored the use of multiple classifiers. However, these efforts have mostly concentrated on employing multiple classifiers during the final training phase, rather than the pre-training phase [2]. While such studies have shown promising results, they do not tap into the potential benefits of using individual classifiers for each object during pre-training.

The concept of convolutional kernels in neural networks has also received considerable attention. LeCun et al. highlighted the significance of convolutional kernels in feature extraction, a crucial process in neural network-based classification tasks [3]. However, the potential of pre-training distinct convolutional kernels for individual objects has been largely overlooked in the literature.

The landscape of image classification was reshaped with the advent of AlexNet, introduced by Krizhevsky et al. [4]. This deep convolutional neural network (CNN) utilized an unprecedented depth in its architectural design, which was pivotal in its success at the ImageNet competition. The essence of AlexNet was its ability to leverage a vast dataset for pre-training, allowing the network to develop robust feature detectors. This approach set a new benchmark for image recognition tasks, establishing deep learning as the method of choice for such challenges.

In 2016, He et al. [5] built upon this concept with the development of ResNet, a deeper neural network that incorporated residual connections. These connections enabled the training of networks with significantly more layers, by effectively addressing the vanishing gradient problem. The ResNet architecture demonstrated that increasing network depth, when combined with an effective pre-training regime, could lead to substantial gains in performance, as evidenced by its victory in the ILSVRC 2015.

The work of He et al. validated the hypothesis that depth is a crucial factor for the performance of CNNs, provided that the architecture is designed to support efficient training and learning [5].

Further research by Simonyan and Zisserman introduced VGG, a network that emphasized the depth aspect by using a series of small convolutional filters, which provided improved feature extraction capabilities [6]. Similarly, Szegedy et al.'s inception modules, used in GoogLeNet, demonstrated that a combination of kernels of different sizes in a CNN architecture could lead to enhanced performance in image classification tasks [7]. These studies underscore the evolution and refinement of pre-training strategies, suggesting that more nuanced architectures can lead to better generalization and feature extraction.

In summary, while existing research has contributed substantially to our understanding of pre-training in neural networks, there remains a gap regarding the use of individual classifiers for each object during the pre-training phase. This literature review underscores the need for innovative methodologies that can fill this gap and potentially enhance the accuracy and efficiency of neural network-based classification tasks.

3. Methodology

The methodology used in this research involves a distinct approach to handwritten digit recognition using convolutional neural networks (CNNs). The key distinguishing feature of this method is the individual creation of a binary classifier for each digit, which is then combined into a single comprehensive classifier.

3.1. Data Collection and Pre-processing

The MNIST dataset, a large database of handwritten digits commonly used for training various image processing systems, serves as the source of data for this study. The dataset consists of 60,000 training images and an additional 10,000 images for testing. Before the implementation of the model, the images were reshaped to conform to the model's input requirements.

3.2. Model Development and Training

The initial phase of model development involved the creation of a binary classifier for each digit in the MNIST dataset. Each classifier model is comprised of a single convolutional layer, a layer to flatten the input, and a dense layer. The convolutional layer consisted of three kernels and utilized a sigmoid activation function. The model was trained over five epochs with the Adam optimization algorithm and the categorical cross-entropy loss function.

After training, the weights (or convolutional kernels) from the convolutional layer of each model were extracted and saved. These kernels were then aggregated and used as the initial weights for the comprehensive classifier.

3.3. Comprehensive Classifier

With the initial weights from the individual classifiers, a comprehensive classifier was constructed. This comprehensive model was then trained over two epochs using the Adam optimizer and the sparse categorical cross-entropy loss function. Simultaneously, a conventional model with the same architecture was created and trained without any predefined initial weights for comparison purposes.

3.4. Model Evaluation

Finally, both the comprehensive and conventional models were evaluated. The models were tested using the testing portion of the MNIST dataset. Their accuracy was determined and compared to assess the effectiveness of the comprehensive classifier model when contrasted with the conventional model.

The entire research process was conducted using TensorFlow, a well-known deep-learning framework.

In this section, this paper presents the results obtained from the experiments conducted using the methodology described in the previous section. The primary focus of the analysis is the comparison of

DOI: 10.54254/2755-2721/54/20241659

the performance of our comprehensively trained model with pre-trained convolutional kernels and a conventional model with randomly initialized convolutional kernels of the same architecture.

4. Conclusion

The experimental findings provide compelling evidence supporting the efficacy of the proposed approach, which incorporates pre-trained convolutional kernels into a comprehensive classifier for digit recognition using the MNIST dataset. The model's performance, in terms of accuracy on the test set, was found to be significantly better compared to a conventional CNN model with randomly initialized kernels. The enhancement in performance, an increase of approximately 2% in accuracy, is non-trivial considering the complexity of the task at hand and the already high baseline performance of conventional CNNs on the MNIST dataset.

This improvement can be mainly attributed to the initial weights provided by the pre-training phase, which is a form of transfer learning. The pre-training phase allows each binary classifier to capture unique, digit-specific features. When these pre-trained classifiers are integrated into the comprehensive classifier, the model could leverage these learned features to achieve improved performance in recognizing digits. As such, the results suggest that the pre-trained convolutional kernels offer a good starting point for the training process and lead to a more effective model for digit recognition.

The promising results obtained in this study provide a strong foundation for further exploration. Future research will focus on more complex recognition tasks beyond digit recognition. For instance, the selection of activation functions used during the pre-training phase could significantly impact performance. In this study, various activation functions yielded different results, with the sigmoid function proving to be the most effective. However, this approach may result in diminished performance for specific class recognition.

To mitigate this, future work will consider optimizing this method by selecting the optimal pre-trained convolutional kernels for different classes. This direction is expected to further refine the proposed method and potentially lead to a more robust and versatile model for image recognition tasks. The insights gained from this research, therefore, pave the way for advancing the field of image recognition and deep learning, particularly in the realm of transfer learning and domain-specific feature extraction.

References

[1] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, vol. 18, no. 7, pp. 1527–1554, 2006.
[2] C. Szegedy et al., "Going deeper with convolutions," in Proc. of the IEEE conference on computer vision and pattern recognition, pp. 1–9, 2015.
[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.
[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, 2012.
[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in International Conference on Learning Representations, 2015.
[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.