Improvement of Performance Related to Cross Dataset Handwritten Recognition Based on Transfer Learning

Kaiyuan Chen

Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham, B15 2TT, United Kingdom
kxc263@student.bham.ac.uk

Abstract. Given the escalating diversity and intricacy of handwritten samples, it remains challenging to enhance the accuracy and robustness of recognition algorithms. This study proposed a solution to this problem by optimizing a pre-existing Convolutional Neural Network (CNN) model for handwritten recognition using a new dataset. Initially, the model was trained on the MNIST dataset. It was then fine-tuned using a self-collected dataset with diverse handwritten styles, employing transfer learning to capitalize on existing knowledge and alleviate the training load. This approach involved freezing certain layers in the pre-trained model to prevent overfitting and encourage more generalized feature extraction. The model exhibited high accuracy on the MNIST dataset, with a training accuracy of 99.07% and a testing accuracy of 99.29%. However, when tested on the self-collected dataset, the accuracy dropped to 11.89%. After applying transfer learning, the model achieved an improved testing accuracy of 38.46% on the self-collected dataset. Despite this improvement, the significant gap between training and testing accuracy indicated overfitting, suggesting the need for additional strategies to enhance model generalization. This study establishes a foundation for future work on improving CNN model performance on diverse handwritten digit datasets.

Keywords: Handwritten Recognition, Convolutional Neural Network, Transfer Learning

1 Introduction

Handwritten recognition stands as a pivotal challenge within the domains of computer vision, holding broad application prospects. Its evolution in commercial exploitation has significantly flourished. Despite the development of numerous efficacious algorithms over previous decades for handwritten recognition, the escalating diversity and intricacy of handwritten samples necessitate the enhancement in accuracy and robustness of these algorithms.

During the nascent stage of handwritten symbol recognition, techniques of digital image processing were employed, fundamentally based on rudimentary image knowledge and simplistic image processing techniques. These traditional techniques, however, demonstrate inherent limitations within the context of handwritten symbol

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_68

verification. For instance, they frequently encounter dilemmas while managing low-quality images. Studies have demonstrated a comparison between the performances of various algorithms such as K-nearest neighbor (KNN), Support Vector Machine (SVM), Backpropagation (BP) neural networks, and Convolutional Neural Network (CNN) in handwritten digit recognition applications, with the most commendable recognition rate exhibited by CNN [1].

A plethora of CNN models have demonstrated exceptional performance in this task. For instance, Liu et al. employed LeNet 5 for implementing the handwritten recognition [2]. LeNet-5 is an early-stage CNN model, explicitly designed for the task of handwritten digit recognition. LeNet-5 possesses a relatively small model complexity, enabling decent performance without demanding extensive computational resources, thus making it ideal for beginners and basic applications. Furthermore, Farahbakhsh et al. also considered the application of AlexNet in related tasks [3]. Though AlexNet was initially designed to address large-scale natural image classification tasks, its potent representational capability enables it to adeptly handle handwritten digit recognition tasks, thereby yielding high accuracy. ResNet is another typical model in this case [4]. By implementing residual connections, ResNet enables deeper network training, thus enhancing the model's representational capabilities. For intricate handwritten samples, ResNet may potentially demonstrate superior performance.

However, most previous models mainly focused on their application on the current dataset and did not consider their generalization ability on real-world or dissimilar data. To address this issue, the prime objective of this study is to adjust and optimize a pre-existing, trained model of a handwritten recognition neural network using a new dataset of handwritten samples. This novel dataset, comprising an array of different handwritten styles, will facilitate the model in better adapting to handwritten samples present in real-world scenarios. In this process, transfer learning will be employed to capitalize on existing knowledge and alleviate the training load for the new task. Finally, the methodology will be validated through a series of experiments aimed at assessing the performance of the model on the new dataset and comparing it with other handwritten recognition algorithms. The evaluation metrics such as accuracy will be employed to render a comprehensive assessment of the model's performance.

2 Methods

2.1 Dataset Description and Preprocessing

The source domain dataset used in this study is the MNIST dataset, which is widely employed in computer vision task. This dataset consists of 60,000 training images and 10,000 testing images, all of which are grayscale images of handwritten digits (0-9), making the number of categories equal to 10. The sample image is provided in the Fig. 1.

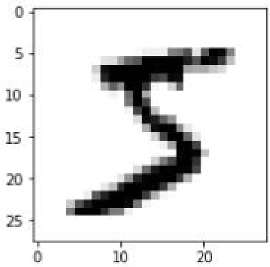

Fig. 1. The sample image of the MNIST dataset (Photo/Picture credit: Original).

The target domain dataset for this study was collected from the internet, which includes 286 different images but with the same categories as the MNIST dataset.

The preprocessing consisted of three parts. Firstly, the dataset underwent a reshaping procedure to make it compatible with the CNN model, and the pixel values were normalized by dividing by 255 to scale them between 0 and 1. Secondly, one-hot encoding was applied to the labels of the dataset. Lastly, reading the new image data from directories, resizing them to match the dimensions of the MNIST dataset (28×28). The sample image is provided in the Fig. 2.

Fig. 2. The sample image of the collected handwritten digit dataset (Photo/Picture credit: Original).

2.2 Transfer Learning-based Convolutional Neural Network

The main model used in this study is CNN. A CNN is a type of deep learning model which is particularly good at recognizing patterns present in images [5-7]. Transfer learning was applied in this study to utilize the knowledge learned from the MNIST dataset to the new target dataset [8-10]. This is done by training the model on the source dataset, and then "fine-tuning" it on the target dataset, with some layers' weights frozen to prevent overfitting.

The structure of the CNN used here consists of two convolutional layers each followed by a max-pooling layer, then a dropout layer to prevent overfitting, a flattening layer, and two dense layers, the last of which uses the softmax activation function to output probabilities for each of the 10 categories. The details structure of the model can be found in Fig. 3.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 28, 28, 16)        416

 max_pooling2d (MaxPooling2D  (None, 14, 14, 16)       0
 )

 conv2d_1 (Conv2D)           (None, 14, 14, 32)        12832

 max_pooling2d_1 (MaxPooling  (None, 7, 7, 32)         0
 2D)

 dropout (Dropout)           (None, 7, 7, 32)          0

 flatten (Flatten)           (None, 1568)              0

 dense (Dense)               (None, 128)               200832

 dropout_1 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 10)                1290

=================================================================
Total params: 215,370
Trainable params: 215,370
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3. The structure of the constructed CNN (Photo/Picture credit: Original).

The construction and fine-tuning of CNN models are pivotal to the implementation of the study, with judicious adjustments playing a vital role in enhancing recognition accuracy.

2.3 Adaptation of a CNN model

In terms of transfer learning, the tactic of freezing certain layers in a pre-trained model, bestows multiple merits, enhancing the overall efficacy of the model. It brings computational efficiency by avoiding the update of initial layers' weights, consequently speeding up the training process. This approach preserves the generic feature extraction capacity of the CNNs, which tends to occur in the initial layers, identifying common visual components such as edges and shapes. Furthermore, overfitting, a common problem when dealing with a small new dataset, is mitigated by adopting a pre-trained model and limiting the training to the final layers. Lastly, improved performance on the new task is achieved as the final, task-specific layers are fine-tuned to adapt to the new dataset's specific features. Collectively, these benefits ensure a robust model performance with optimized utilization of computational resources and data when transitioning from recognizing MNIST digits to classifying hand-written digits in a distinct set of images.

2.4 Implementation Details

Training on source dataset. For the training on the MNIST dataset, the model used the categorical cross-entropy as the loss function, Adam as the optimizer, and accuracy as the evaluation metric. The training was implemented with a batch size of 200 and for 10 epochs.

Training on target dataset. For the training on the target dataset, the same loss function, optimizer, and evaluation metrics were used. However, the training was carried out with half of the data for validation and for 100 epochs. Before this training, all layers except the last two were frozen to ensure the fine-tuning process.

This effectively means that only the last two layers of the model will learn from the new dataset during the training process. This strategy is particularly useful when the new dataset is small, as it helps to avoid overfitting by not allowing the majority of the parameters to adjust to the new data. Instead, the model is forced to extract features that are relevant to both the original dataset and the new one, thus promoting the creation of more generalized representations.

3 Results and Discussion

3.1 The Performance of the Model Based on Various Configurations

The experiment evaluated three configurations of the developed CNN. The results are summarized in the Table 1.

Table 1. The results of models after training and testing.

Model Configuration	Training Loss	Training Accuracy	Testing Loss	Testing Accuracy
CNN (trained on MNIST, tested on MNIST)	0.0303	99.07%	0.0204	99.29%
CNN (trained on MNIST, tested on self-collected dataset)	-	-	9.0279	11.89%
Transfer learning-based model (trained on self-collected dataset, tested on self-collected dataset)	0.0085	99.30%	14.6861	38.46%

In the experiment, three different configurations of CNN were evaluated for handwritten digit recognition. The configurations included a CNN trained and tested on the MNIST dataset, a CNN trained on the MNIST dataset but tested on a self-collected dataset, and a transfer learning-based model trained and tested on the self-collected dataset. The first configuration achieved a training accuracy of 99.07% and a testing accuracy of 99.29% on the MNIST dataset. The second configuration, when tested on the self-collected dataset, obtained a testing accuracy of only 11.89%. The transfer learning-based model attained a training accuracy of 99.30% on the self-collected dataset but a testing accuracy of 38.46%.

3.2 Discussion

It is observed that the model performs exceptionally well on the MNIST dataset. However, a significant drop in testing accuracy is noted when the model trained on the MNIST dataset is tested on a different, self-collected dataset. This indicates that the model does not generalize well to new data that has a different distribution or characteristics compared to the MNIST dataset.

The transfer learning-based model exhibits an improved testing accuracy compared to the model trained only on the MNIST dataset. Despite potential disparities in background, image quality, and other aspects between the source dataset (MNIST) and the target dataset (self-collected dataset), they both encompass handwritten digits, implying a similarity in key features. Consequently, knowledge about digit shapes and structures learned from the MNIST dataset can be transferred to the new dataset via transfer learning. During the retraining of the model in the target domain, modifications need only be made to the features of the background. However, there is still a considerable difference between training and testing accuracy, suggesting overfitting to the training data. Further strategies such as data augmentation, regularization, or employing a more complex model architecture may be required to improve generalization.

4 Conclusion

In this study, a CNN was designed and implemented for the purpose of handwritten digit recognition. The initial training was conducted on the MNIST dataset, a well-established dataset comprising grayscale images of handwritten digits. Subsequently, the model was fine-tuned using a self-collected dataset through a transfer learning approach. The experimental results reveal that the CNN performs exceptionally well on the MNIST dataset, with high training and testing accuracy. However, a significant decline in performance is observed when the model is tested on a different dataset, indicating that the model does not generalize well to datasets with different characteristics. By employing a transfer learning approach and fine-tuning the model on a self-collected dataset, a better generalization to the new dataset was achieved compared to the model trained only on MNIST. However, overfitting to the training data was observed, suggesting that further strategies are needed to improve the model's generalization capabilities. Future work could involve experimenting with different model architectures, data augmentation techniques, and regularization methods to further improve model performance on diverse datasets.

References

1. Liu, Z.: Handwritten Character Recognition Based on Convolution Neural Network Models. In: CIBDA 2022; 3rd International Conference on Computer Information and Big Data Applications. pp. 1–5 (2022).
2. Liu, W., et al.: Comparisions on KNN, SVM, BP and the CNN for Handwritten Digit Recognition. In: 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA). pp. 587–590 (2020).
3. Farahbakhsh, E., et al.: Improving persian digit recognition by combining data augmentation and AlexNet. In: 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP). pp. 265–270 (2017).

4. Chi, X. et al.: Handwriting Recognition Based on Resnet-18. In: 2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). pp. 456–459 (2021).
5. Yu, Q., Yang, Y., Lin, Z., et al.: Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV. China Communications, 17(3): 46-57 (2020).
6. Kayalibay, B., Jensen, G., van der S., P.: CNN-based segmentation of medical imaging data. arXiv preprint arXiv:1701.03056, (2017).
7. Mortazi, A., Bagci, U.: Automatically designing CNN architectures for medical image segmentation, Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9. Springer International Publishing, 98-106 (2018).
8. Weiss, K., Khoshgoftaar, T. M., Wang, D. D.: A survey of transfer learning. Journal of Big data, 3(1): 1-40 (2016).
9. Ying, W., Zhang, Y., Huang, J., et al.: Transfer learning via learning to transfer, International Conference on Machine Learning. PMLR, 5085-5094 (2018).
10. Zamir, A. R., Sax, A., Shen, W., et al.: Taskonomy: Disentangling task transfer learning, Proceedings of the IEEE conference on computer vision and pattern recognition. 3712-3722 (2018).