An Optimization of Image Super-Resolution Method Based on Residual Network

Yue Pan
School of Computer Science, University of Nottingham
akisoojin@hotmail.com

ABSTRACT: In this paper, an optimization with deep residual networks for the super-resolution method will be proposed. Meanwhile, the optimized method will be used into MNIST database to improve the image resolution, and the accuracy of the image identification of the improved database will be test by the residual network. In the optimization design, there are 4 convolutional layers, and 5 residual blocks will be used to replace the 3 convolutional layers of initial design, and there are 3 convolutional layers in each residual block. After the residual network testing, the experiment is positive and effective, and total number of errors is 34, and the accuracy is 99.43%.

1. Introduction
In the last ten years, accompanied with the prosperity of “ImagerNet competition”, deep learning neural network has faced a numerous of computer vision tasks, such as, the object-detection, segmentation, human pose estimation, video and image classification, object tracking, handwriting classification [1,2]. Meanwhile, He, Zhang, Ren, et al [3] stated that a series of breakthroughs for image classification have been achieved by the deep convolutional neural networks (CNN). Some iconic CNN models are created during the period of the deep learning neural network’s developments, which can not only reflect the computer vision tasks that be conquered by the CNN models, but also witness the progress of deep learning neural network. For an instance, VGGnet, residual network (ResNet), GoogLeNet, YOLOnet, etc. According to He et al [3], the relationship between the network depth and ImageNet dataset is a challenging for deep CNN, which means the implement for different deep CNN models for ImageNet dataset is imperative aim of related research.

Meanwhile, identifying, classifying, and translating human handwriting scripts to computer documents is one of main issues of deep learning field. Effectively and accurately identify human handwriting can not only preserve precious handwriting documents instead the to scan them or maintain the originals in museum, but also bring a superb convenient to the group who cannot type when they have to use computer.

Besides, the classical issue,” the low-resolution image”, in computer vision task is still researched by the field of deep learning neural network. Meanwhile, accompanied with the emergence of CNN models, the image size requirements to train for dataset are quite different in different CNN model. To overcome both issues, Dong, Loy, He, et al [4] proposed a deep learning solution for single image super-resolution. In the image super-resolution, the super-resolution image will be outputted after three convolutional processes from the original image. According to Dong et al [4] research, the super-resolution output is clearer than the traditional Bicubic output, which is the breakthrough for the deep learning neural network field.

However, during the experimental operation, there are some drawbacks has exposed in the original
super-resolution [4]. Moreover, the low-resolution image, and image misrecognition issues still exist in the original method. Attributing to the curiosities for the deep learning network and the desire of exploration for the super-resolution, in this project, an optimization with deep residual networks for the original super-resolution method will be proposed. Meanwhile, the optimized method will be used into MNIST database to improve the image resolution, and the accuracy of the image identification of the improved database will be test by ResNet.

2. Literature Review

2.1. Traditional convolutional neural network
The convolutional neural networks (CNN) are types of feedforward neural networks which has the convolutional calculations with deep structures. Meanwhile, CNN also is a representative algorithm of the deep learning field. Moreover, CNN has the ability of representation learning, which can do the shift-invariant classification for the output by the layers’ structures of CNN. The earliest research for CNN accompanied with two CNN models creation: the time delay neural network [5], and the LeNet [6], and the LeNet is the initial CNN model mainly created to solve the computer vision issues [6].

According to research of LeCun [6,7], and the analysis of Ketkar [8], the traditional convolutional neural networks have three layers structures: the input layer, the hidden layer, and the output layer. In the input layer, attributing to the gradient descent algorithm, the features of input need to be standardized to increase the learning efficiency and performance of CNN. Besides, there are three layers to constitute the hidden layer which are the convolutional layer, the pooling layer, and the fully-connected layer: The convolutional layer constituted by the convolutional kernel, the convolutional parameters, and the activation function, the pooling layer is mainly constituted by the pooling strategy function, and the function of the fully-connected layer can be regarded as the hidden layer of the feedforward neural networks. Moreover, the structure of LeNet-5 model is the input layer, the convolutional layer, the pooling layer, the fully-connected layer, and the output layer.

2.2. The deep residual networks
The ResNet is an optimization for VGGNet-16 model inspired from the “Highway Network” thought form Srivastava, Greff, and Schmidhuber [9] that build the direct path in the traditional deep CNN model, which can solve the main deep CNN issues: lost information problem during information transmission, gradient disappear and explode problem, and untraining problem in deeper layer [3]. According to He et al [3], they created the residual block to inherit the “Highway Network” thought, and the residual block can guarantee the better predictive effect with the more image features obtained. In other words, the deeper network structure the better predictive effect in ResNet.

2.3. Image Super-Resolution by deep convolutional networks
The low-resolution image is the classical issue in computer vision task which is still researched by the field of deep learning neural network. Meanwhile, accompanied with the emergence of CNN models, the image size requirements to train for dataset are quite different in different CNN model.

To overcome both issues, Dong, Loy, He, et al [4] proposed a deep learning solution for single image super-resolution. In the image super-resolution, the super-resolution image will be outputted after three convolutional processes from the original image. According to Dong et al research [4], the super-resolution output is clearer than the traditional Bicubic output, which is the breakthrough for the deep learning neural network field.

2.4. MNIST dataset
MNIST dataset is an imageNet dataset established by National Institute of Standards and Technology (MNIST), which constituted by different human’s handwritings, such as high school students and handwriting records form the Census Bureau [10].
3. Methodology

3.1. Introduction
Experimental environment description: Python version: 3.7.1, Pytorch version: 1.7, CUDA version: 11.2, Anaconda version: 3, and Graphics core: RTX3070.

3.2. Implement of ResNet
In this part, the first step is to do the design of the residual block. Each residual block is constituted by 2 convolutional layers, and after each convolutional layer a ReLU function and a BatchNorm2d function will be used. The BatchNorm2d function is a normalization function for the image feature. Hence, the structure of residual block can be summarized as: a convolutional layer, a BatchNorm2d function, a ReLU function, another convolutional layer, a BatchNorm2d function, and a ReLU function.

During the period of forward propagation of ResNet, the input and output of the residual block should be gathered as the Figure 1 shows.

The whole ResNet structure can be stated as, there are 3 residual units constituted the whole ResNet, and there are 2 residual blocks constituted each residual unit. Moreover, a Averagepool layer and a fully-connected layer linked the last residual block. Meanwhile, the pooling kernel size is 8*8, and the image features will get 64 after the Averagepool. Moreover, the feature will be decrease from 64 into 10 to fit the MNIST dataset. In the convolution layer, during the whole training process, the convolutional kernel size set up as 3, the stride set up as 1, the padding set up as 1, the batch size for each iteration set up as 64, the number of iterations set up as 20, and the learning rate set up as 0.001.

3.3. Implement of super-resolution
In the part of implement of super-resolution, an optimization for the initial design of Dong et al [4] will be implement in this project: there are 4 convolutional layers, and 5 residual blocks will be used to replace the 3 convolutional layers of initial design to obtain the more super-resolution image.

The structure of the optimized super-resolution can be stated as, there are three convolutional layers in each residual block, in each residual block, the convolutional kernel size set up as 1 in first convolutional layer, the convolutional kernel size set up 3, the stride set up as 1, the padding set up as 1 in second convolutional layer, and the convolutional kernel size set up as 1 in third convolutional layer.

In the 4 convolutional layers, the convolutional kernel size set up 3, the stride set up as 1, the padding set up as 1 in first convolutional layer, the convolutional kernel size set up as 1 in second convolutional layer, the convolutional kernel size set up 3, the stride set up as 1, the padding set up as 1 in third
convolutional layer, and the convolutional kernel size set up as 1 in forth convolutional layer. Moreover, the batch size for each iteration set up as 128, the number of iterations set up as 30, and the learning rate set up as 0.001.

4. Finding

4.1. Introduction

In this experiment, the division ratio for MNIST dataset as,

\[ \text{Training Set: Test Set: Validation Set} = 8:1:1 \]

Hence, the Training set has 47,999 images, the Test Set has 6,001 images, and the Validation Set has 6000 images.

4.2. Finding of super-resolution

As Figure 2 shows, the first image is the original image from MNIST dataset, the second image is the original image after the initial super-resolution, the third image is a failed sample about the original image after the optimized super-resolution, and the fourth image is a successful sample about the original image after the optimized super-resolution.

4.3. The accuracy of ResNet after super-resolution

These two curves represent the training set (train) and validation set (val), and the result of accuracy between training set and validation set in tradition model and, the result of loss value between training set and validation set in tradition model as Figure 3 shows.

After the experiment, the experiment is positive and effective from the trend of the two curves, this ResNet model can predict MNIST dataset successfully. Meanwhile, the test set with 6,001 images is used to predict by this model, the result will output as .csv format, and the error condition is calculated as: total number of errors is 34, and the accuracy is 99.43%.
5. Conclusion

5.1. Conclusion for super-resolution
After the whole experiment, image super-resolution by deep convolutional networks can not only solve the low-resolution image issue, but also can be regarded as an advanced up-sampling process method. From Figure 2, the differences between the optimization super-resolution and initial super-resolution can be easily seen, which means the performance of the original super-resolution design has an enormous space to improve. Moreover, the optimization design for the super-resolution in this experiment is a primary optimization which motivated by the curiosities for the deep learning network and the desire of exploration for the function of CNN models, and it can be regarded as an inspiration for the one who has the same curiosity for the super-resolution by deep convolutional networks.

For a further discussion, from the initial super-resolution to the optimization super-resolution in this experiment, the quality of data has a great improvement, and this improvement brought a better result in the ResNet. The relationship between the quality of data and the performance of network model also need to pay attention by related researcher. To provide high quality data set, to avoid the noise of data set, and to search a better up-sampling process method also need to get enough attention by related researcher.

5.2. Conclusion for the accuracy experiment
After the experiment, from the trends of the two curves, the ResNet model implemented in the experiment can predict optimized MNIST dataset successfully. Meanwhile, after prediction for test set, the accuracy rate reached into 99.43%. Hence, accompanied with optimization for CNN models, the performances of the CNN models are constantly improving from this experiment. Moreover, the residual block in the ResNet can not only guarantee the better predictive effect with the more image features obtained, but also solve the lost information problem during information transmission, gradient disappear and explode problem, and untraining problem in deeper layer.

For a further discussion, from the experiment, an evolutionary process from traditional CNN to VGGNet to ResNet can be deeply experienced. Meanwhile, two evolution method can be summarized as to deepen the structure of neural network to obtain the better performance of deep neural network, and to add the residual block to strengthen the performance of deep neural network, which can correspond the two imperative thoughts of developments for deep neural network that the deeper thought, and residual thought. According to the residual thought that propagations of network is not forward propagation, and the propagations can be stacked, series of excellent network structures are created such as the GoogLeNet. Moreover, the performance of network models can be strengthened by stack of sensing vision and network models of different depths.

Reference
[1] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016) Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp.2818-2826). IEEE.
[2] Szegedy, C., Wei, L., Jia, Y., Sermanet, P., Rabinovich, A. (2015) Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE
[3] He, K., Zhang, X., Ren, S., Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.
[4] Dong, C., Loy, C. C., He, K., Tang, X. (2016) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell, 38(2), 295-307.
[5] Waibel, A. (1987) Phoneme recognition using time-delay neural networks. [6] Lecun, Y. (1989) Generalization and Network Design Strategies. Connectionism in Perspective. Elsevier.
[7] Lecun, Y., Bengio, Y. (1995) Pattern Recognition and Neural Networks. Handbook of Brain Theory & Neural Networks.
[8] Ketkar, N. (2017) Convolutional neural networks. Springer International Publishing.
[9] Srivastava, R. K., Greff, K., Schmidhuber, J. (2015) Highway networks. Computer Science.
[10] Blum, C. W. (2017) On the Effectiveness of Neural Networks Classifying the MNIST Dataset.