



Image enhancement effect on the performance of convolutional neural networks

Xiaoran Chen

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:

Author:
Xiaoran Chen
E-mail: xicb17@student.bth.se

University advisor:
Dr. Abbas Cheddad
Department of Computer Science

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

Context. Image enhancement algorithms can be used to enhance the visual effects of images in the field of human vision. So can image enhancement algorithms be used in the field of computer vision? The convolutional neural network, as the most powerful image classifier at present, has excellent performance in the field of image recognition. This paper explores whether image enhancement algorithms can be used to improve the performance of convolutional neural networks.

Objectives. The purpose of this paper is to explore the effect of image enhancement algorithms on the performance of CNN models in deep learning and transfer learning, respectively. The article selected five different image enhancement algorithms, they are the contrast limited adaptive histogram equalization (CLAHE), the successive means of the quantization transform (SMQT), the adaptive gamma correction, the wavelet transform, and the Laplace operator.

Methods. In this paper, experiments are used as research methods. Three groups of experiments are designed; they respectively explore whether the enhancement of grayscale images can improve the performance of CNN in deep learning, whether the enhancement of color images can improve the performance of CNN in deep learning and whether the enhancement of RGB images can improve the performance of CNN in transfer learning?

Results. In the experiment, in deep learning, when training a complete CNN model, using the Laplace operator to enhance the gray image can improve the recall rate of CNN. However, the remaining image enhancement algorithms cannot improve the performance of CNN in both grayscale image datasets and color image datasets. In addition, in transfer learning, when fine-tuning the pre-trained CNN model, using contrast limited adaptive histogram equalization (CLAHE), successive means quantization transform (SMQT), Wavelet transform, and Laplace operator will reduce the performance of CNN.

Conclusions. Experiments show that in deep learning, using image enhancement algorithms may improve CNN performance when training complete CNN models, but not all image enhancement algorithms can improve CNN performance; in transfer learning, when fine-tuning the pre-trained CNN model, image enhancement algorithms may reduce the performance of CNN.

Keywords: Image Enhancement, Convolutional Neural Networks, Deep Learning, Transfer Learning

Acknowledgment

I would like to thank my mentor, Dr. Abbas Cheddad, he patiently taught the experiment and gave me encouragement and help. At the same time, I also want to thank my parents and friends, they gave me care.

Contents

Abstract.....	3
Acknowledgment.....	4
Contents.....	5
List of Figures	7
List of Tables	8
List of Formula.....	9
Chapter 1	10
1.1 Problem background	10
1.2 Objectives	11
1.3 Research questions	11
Chapter 2	13
2.1 Transfer learning	13
2.2 CNN.....	13
2.2.1 Activation function.....	13
2.2.2 Optimization function.....	14
2.2.3 Pooling layer strategy	15
2.2.4 Classification	16
2.3 Model evaluation	16
2.3.1 Evaluation method.....	16
2.3.2 Performance metric	17
2.3.3 Hypothetical testing.....	18
Chapter 3.....	19
3.1 Related literatures	19
Chapter 4.....	20
4.1 Variable, Treatment, Object and Subject	20
4.2 Image data set.....	20
4.3 Image processing	21
4.3.1 Filter low-quality images.....	21

4.3.2 Image type conversion.....	21
4.3.3 Image channel conversion	21
4.3.4 Image enhancement	21
4.3.5 Summary.....	23
4.4 CNN model.....	24
4.4.1 Improved Lenet5	24
4.4.2 Pre-trained VGG16.....	25
4.5 Error reduction	26
4.6 Experiments.....	26
4.6.1 Experiment 1	27
4.6.2 Experiment 2	28
4.6.3 Experiment 3	28
Chapter 5.....	30
5.1 Statistical hypothesis testing	30
5.2 Result.....	33
5.3 Discussion	33
5.4 Threat analysis.....	34
Chapter 6.....	36
6.1 Conclusion.....	36
6.1.1 Answering RQs.....	36
6.1.2 Hypothesis testing.....	37
6.2 Future Work	37
References	38

List of Figures

Fig1. Four activation functions	14
Fig2. Performances of four activation functions.....	14
Fig3. Performances of different optimization functions	15
Fig4. Enhanced images in “Cat VS Dog” data set	22
Fig5. Enhanced images in “Chest X-ray” data set.....	22
Fig6. Image processing of gray images	23
Fig7. Image processing of RGB images	24
Fig8. Structure of improved Lenet5 model in the experiment	25
Fig9. Structure of VGG16 in the experiment	26

List of Tables

Table1. Accuracy in 10-fold cross-validation of experiment1	27
Table2. Recall rate in 10-fold cross-validation of experiment1.....	27
Table3. Accuracy in 10-fold cross-validation of experiment2.....	28
Table4. F1 score in 10-fold cross-validation of experiment2	28
Table5. Accuracy in 10-fold cross-validation of experiment3.....	29
Table6. F1 in 10-fold cross-validation of experiment3	29

List of Formula

(1) Accuracy	17
(2) Precision	17
(3) Recall	17
(4) F1 score	17
(5) paired T-test	30

1.1 Problem background

In recent years, deep learning performs well in the field of image recognition, and convolutional neural networks (CNN) are the most outstanding models. As a very classic CNN model, the Lenet5 model has been proven many times in handwritten digit recognition. And the 152-layer “Resnet” network can achieve more than 95% accuracy in the task of one thousand classifications [36], which makes the CNN model widely used in the field of image recognition in recent years. In addition, for the data sets with the small number of images, or to reduce training time, a pre-trained CNN network can be used for transfer learning. Among them, fine-tuning the pre-training network is the most common transfer learning method. Two important ways to train the CNN model are to train the complete CNN model and fine-tune the pre-trained CNN model. However, no matter what kind of training method, before training the CNN models, the image data sets are usually processed in various ways, such as image cropping, extending the number of images, data normalization, and image enhancement.

How to improve the recognition performance of the CNN model is an endless study, and the quality of the image dataset has a great impact on the performance of the model. Because in the image dataset, due to factors such as illumination or camera, the contrast of images will be insufficient, resulting in insufficient features in the images. One research has shown that image preprocessing is essential when training CNN models, which can effectively improve the recognition performance of CNN models. [38] And image enhancement is a very important part of preprocessing. So, it is important to research the relationship between image enhancement and the CNN model.

An image enhancement algorithm can emphasize the overall or local characteristics of the images, clear the unclear image, emphasize certain features of interest, suppress features that are not of interest, enlarge the difference between the features of different objects in the images. It can improve image quality, enrich information, enhance image interpretation and recognition, meet the needs of some special analysis, and make images more suitable for human visual systems and experience [11]. Image enhancement has a lot of strategies, and it can be divided into “Spatial Domain Methods” and “Frequency Domain Methods” according to the principle [11]. Among them, the Fourier transform is a typical “Frequency Domain Method”. Comparing with Fourier transform, although the wavelet transform has a larger computational cost, it can reduce the errors, so the wavelet transform is a good algorithm that belongs to the “Frequency Domain Methods” [39]. There are many methods of “Spatial Domain Methods”, which can be roughly divided into enhanced image contrast and enhanced local features of the images. There are many ways to enhance image contrast. The contrast limited adaptive histogram equalization (CLAHE), the successive means of transform (SMQT), and the gamma correction are some of them. They have different advantages and disadvantages. The CLAHE algorithm

has proven to perform well on medical image X-ray images [32]. SMQT can extract the structure in a robust manner which makes it insensitive to changes in bias and gain, and compare with histogram equalization; it can reduce problems with oversaturation and artifacts [12]. Gamma correction can enhance contrast and brightness on darker images. [6]. Edge enhancement is one of the ways to enhance the local features of the image, and the Laplace operator can effectively extract and enhance the edge information of the images. But machine vision and human vision are different. Whether image enhancement algorithms can improve the performance in image recognition of the Convolutional neural network? Few studies have shown the impact of image enhancement algorithms on training CNN models.

1.2 Objectives

In order to explore the influence of the image enhancement algorithms on the CNN model, the experimental method is adopted as the research method. In order to use image enhancement algorithms with different principles as much as possible, in the experiment, successive means quantization transform (SMQT), contrast limited adaptive histogram equalization (CLAHE), Adaptive gamma correction, Wavelet transform, and Laplace operator were used to enhance the original image. The original images and the enhanced images by different image enhancement algorithms will be compared on the CNN model by two different types of training: train the complete CNN model and fine-tune the pre-trained CNN model. The stakeholders of the thesis are those people who use the CNN model for image recognition, and the conclusions of this thesis can provide a reference for them whether using image enhancement algorithms when training the CNN models.

1.3 Research questions

RQ1: Whether increasing image intensity can improve the performance in image recognition of the Convolutional neural network when training a complete Convolutional neural network?

Motivation: RQ1 is used for exploring the influence of image intensity on training a complete convolutional neural network, the conclusions of this research question can provide a proposal and help for the image preprocessing strategy when training a complete CNN model.

RQ2: Whether increasing image intensity will influence the performance in image recognition of the Convolutional neural network when using transfer learning (fine-tuning)?

Motivation: RQ2 is used for exploring the influence of image intensity on fine-tuning a pre-trained CNN model (transfer learning), the conclusions of this research question can provide a proposal and help for the image preprocessing strategy when fine-tuning a pre-trained CNN model (transfer learning).

The null hypothesis 1: Increasing the image intensity can't improve the performance in image recognition of the Convolutional neural network when training a complete Convolutional neural network.

Alternative Hypothesis 1: Increasing image intensity can improve the performance in image recognition of the Convolutional neural network when training a complete Convolutional neural network.

The null hypothesis 2: Increasing the image intensity will not reduce the performance in image recognition of the Convolutional neural network when using transfer learning.

Alternative Hypothesis 2: Increasing the image intensity will reduce the performance in image recognition of the Convolutional neural network when using transfer learning (fine-tuning).

2.1 Transfer learning

Although the CNN model has excellent performance in the field of computer vision, training a complete CNN model usually requires a large number of data sets, otherwise, the model is prone to over-fitting, and the over-fitting will result in poor generalization ability of the model [19]. In many fields, such as the medical field, the number of image data sets is limited; training a complete CNN in such data sets is very hard. However, transfer learning can solve this problem. Fine-tuning is the most common strategy for transfer learning, which means change the parameters of a pre-trained CNN model [20]. There are many strategies for fine-tuning. The common methods are fine-tuning all convolutional and fully-connected layers, or to freeze some layers and fine-tune the remaining layers. This thesis will study the effect of image enhancement algorithms on transfer learning (fine-tuning).

2.2 CNN

In this topic, how to build a good convolutional neural network is one of the keys to the experiment. The CNN model needs to reduce the training time and have better performance. As the most classic CNN model, the Lenet5 model has been repeatedly proven to achieve good results in the field of image recognition, such as achieving over 98% accuracy in handwritten digit recognition [14]. Moreover, the structure of Lenet5 is relatively simple and the training speed of Lenet5 is very fast. However, the activation function used in the initial Lenet5 model is the “hyperbolic tangent” function, the pooling strategy is average-pooling, the loss function is mean square error (MSE), the optimization algorithm is gradient descent algorithm, and the classifier is Gaussian connections. These methods are only a good choice in 1998. Now, many studies have published better activation functions, pooling strategies, loss functions, optimization algorithms and classifiers. Therefore, in the experiment, the original Lenet5 model needs to be improved.

2.2.1 Activation function

Each layer of the neural network needs to choose one activation function, the value of each neuron needs to be calculated by the activation function to obtain a final value. In neural networks, the role of the activation function is to transform the neural network from linear to nonlinear, so that the neural network can better solve more complex problems. Common activation functions are the “sigmoid” function, the “hyperbolic tangent” function, the “Relu” function and the “Elu” function. There is a paper [18], the experiment compares four activation functions. The CNN model was also used in the experiment. The CNN model

used the structure of Lenet5. The experiment used the “MNIST” data set. The experiment iterated a total of 95,000 times. After 95,000 iterations, the “ELU” activation function reached the highest accuracy.

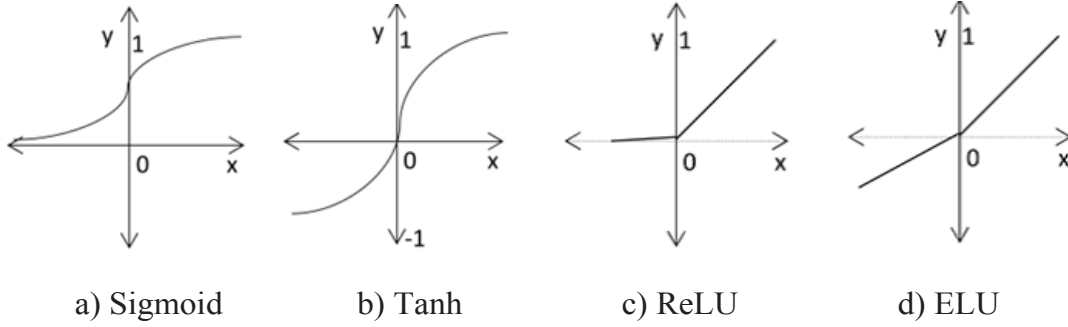


Fig1. Four activation functions [18]

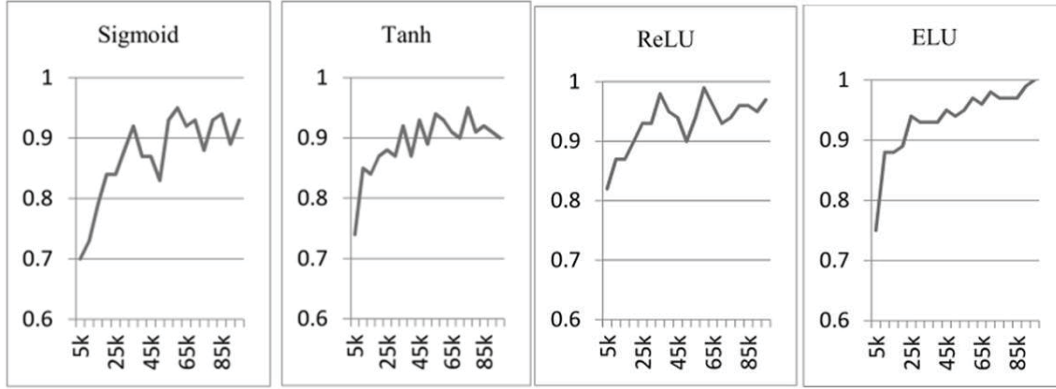


Fig2. Performances of four activation functions [18]

The conclusion of the paper is that it has been observed that “ReLU” the most popular activation function performs better than “sigmoid” and “tanh”. Although the accuracy of the “ELU” function is significantly better than the “ReLU” function after 95,000 iterations, the “ReLU” function only needs about 30,000 iterations to reach the peak, which means that the calculation of the “ELU” function is much larger than the “ReLU”, and the “ELU” adds a large amount of calculation to obtain stability; “ReLU” function is faster and “ELU” is more stable

2.2.2 Optimization function

When training a neural network, it is necessary to update the parameters in the neural network by an optimization algorithm to minimize the value of the loss function, such as weights and bias in each layer of the network. The most classical optimization function is the gradient descent function, but the gradient descent function has very obvious shortcomings: it is easy to reach the local optimal value but it is difficult to reach the global optimal value. Therefore, many optimization algorithms are updated based on the gradient descent algorithm.

A paper proposes an optimization algorithm called “Adam” [9]. The Adam

optimization algorithm is a ramification of the stochastic gradient descent algorithm. Recently, it is widely used in deep learning applications, especially tasks such as computer vision and natural language processing. The Adam algorithm differs from the traditional stochastic gradient descent. Stochastic gradient descent maintains a single learning rate (i.e. α) to update all weights, and the learning rate does not change during training. Adam calculates an independent adaptive learning rate for different parameters by calculating the first-moment estimation and second-moment estimation of the gradient. Overall, Adam is a good optimization function on the deep learning model. The following figure shows the performance of different optimization algorithms for neural networks on MNIST:

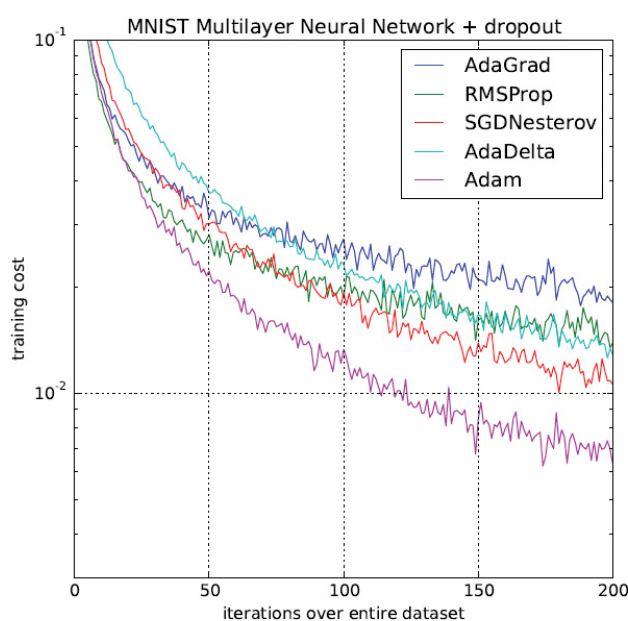


Fig3. Performances of different optimization functions [9]

2.2.3 Pooling layer strategy

In the CNN model, the role of the pooling layer is mainly to extract image features, reduce the input size of the next layer. That means the pooling layer can reduce the computation of the CNN, and prevent over-fitting. The common pooling strategies are max pooling and mean pooling. Max pooling selects the largest value in the pooling kernels, while mean pooling calculates the averages of pooling kernels. So, the calculation of max pooling is much smaller than the mean pooling.

Studies have shown that using the largest pooling layer can greatly reduce the amount of computation when training CNN models and shorten the training time [3]. In addition, there is a paper comparing the max pooling and the average pooling. Experiments have shown that the maximum pooling on the “MNIST” dataset can improve the accuracy around 1% to 1.5% compared to the average pooling. In face recognition, the max pooling can improve the recognition

accuracy around 4% to 5% than the average pooling [28]. Obviously, the max pooling has a better performance than the average pooling in CNN.

2.2.4 Classification

The original Lenet5 model used the Euclidean Radial Basis Function (RBF) based Gaussian connections classifier. The classifier has been completely replaced and it is hard to find any documents about it. Now, in the CNN models, the most common classifier is “softmax”. There is a paper show that the CNN model using softmax has a very good performance on the recognition of street view images [7]. And the cross-entropy cost function always uses with a softmax classifier. Some papers show that using cross-entropy cost function and softmax together has excellent performance on the CNN model [1] [10], so it is very reasonable to use softmax and cross-entropy cost functions instead of Gaussian connections and MSE.

2.3 Model evaluation

In this project, how to compare the performance of the same CNN model in different data sets is the key. But in fact, comparing performance is very complicated, it involves three factors. First of all, it needs to compare the generalization performance of the CNN model. However, one test or one experiment can only get the performance on the test set. Performance on the test set is different from generalization performance. Second, the performance on the test set will be influenced by the choice of the test set, the number of test sets and the sample in the test set. Third, the CNN model is not always stable, that is, using the same parameters to test multiple times on the same test set, the results will be different. So selecting good evaluation method, performance measurement, and hypothesis testing method is very important

2.3.1 Evaluation method

There are many evaluation methods in machine learning, including hold-out, K-fold cross-validation and bootstrapping. The hold-out method is the most basic method, which directly divides the data set into two mutually exclusive sets, one of the two sets is the training set and the other one is testing set. Since this method has a big contingency, different methods of division will produce different results, so the method is not stable. But this method only needs to train the model once to get the conclusion, so it is more suitable for those very large data sets, hold-out can reduce the training time. K-fold cross-validation is usually used with Leave-One-Out, which divides the data set into K parts and tests K times. Each subset is used as the testing set, which ensures that all data will be tested once. Although K-fold crosses-validation is highly reliable, if using it on a large data set, the calculated amount will be very large and the training time will be very long. Bootstrapping is to randomly sample one data from the original data set to testing set, repeat m times to get testing set, use the original data set as a training set. When m tends to be infinite, about 36.8% of the data in the original

data set will not appear in the testing set [34]. Studies have shown that due to the bias of bootstrapping, it may show poor performance in large sample cases. K-fold cross-validation is severely biased upward for the small sample case [8]. So bootstrapping performs better on small data sets, while K-fold cross-validation performs better when the number of data set is big enough. At the same time, there are also papers that show using K-fold cross-validation and Leave-One-Out has good performance in image recognition. [17].

2.3.2 Performance metric

Among all classification problems, whether it is a two-category problem or a multi-classification problem, accuracy and error rate are both the most common and important metric. Accuracy is the ratio of the number of correctly classified samples to the total number of samples, and the error rate is the proportion of the number of samples with the wrong classification. Accuracy is the ratio of the number of correctly classified samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

However, the error rate and accuracy do not meet all the task requirements, and different metrics are required in different tasks. For example, in a system for capturing deep-sea fish, in order to capture salmon, it is necessary to recognize images of the salmon. Since the system does not need to capture all the salmon, but it is important to ensure that as many salmon as possible are caught in the fish caught. At this time, the image recognition system is more concerned with "the proportion of the true salmon to all the fish which are considered as salmon", that is, the "Precision" value. Precision's calculation method is shown in the formula, where True Positive is the number of correctly identified salmon, and False Positive is the number of fish that are not salmon but are considered as salmon.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

In the medical field, on the identification system for detecting diseases, it is necessary to ensure that the disease is not missed as much as possible. At this time, the recognition system is more concerned with "The proportion of correctly identified diseases in all disease cases", that is, the recall rate. Recall's calculation method is shown in the formula, where "True Positive" means the number of samples that are correctly detected, and "False Negative" means the number of samples of diseases that have not been detected.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

For the unbiased two-class system, Precision and Recall are equivalent. In this case, the F1 score can be used as the standard for performance measurement. The calculation method of the F1 score is shown in the formula.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

2.3.3 Hypothetical testing

Hypothesis testing is to verify whether the results of the experiment can overturn the null hypothesis. There are many methods for hypothesis testing, such as T-test, Paired T-test, Mann-Whitney test, F-test, Wilcoxon test, chi-square test and Kruskal-Wallis test and so on. Different experiments and different evaluation methods require different methods. [23].

Chapter 3

Related work

3.1 Related literatures

One paper proposes to use adaptive region-based histogram equalization to improve the contrast of grayscale images [33]. The paper uses the “Yale B” database, the “Extended Yale B” database and the “CMU-PIE” dataset, which contains a large number of face images with single-sided illumination and low-intensity. The paper uses the “Euclidean distance nearest-neighbor” classifier to classify images. Experiments show that using an adaptive region-based histogram equalization algorithm can improve the accuracy of image classification to 100%. However, the “Euclidean distance nearest-neighbor” classifier has lower accuracy for image classification than the CNN model. This thesis also tried to use the datasets to experiment, but the CNN model performed very well on the original dataset and can achieve 100% accuracy, so it can't be proved that image enhancement can help CNN to perform better on these datasets.

Studies have shown that using bimodal image enhancement can greatly improve the accuracy of recognition when CNN is natural scene character recognition [21]. Different bimodal image enhancements will impact on the accuracy of image recognition. The proposed bimodal image enhancement algorithm performs better than the traditional bimodal image enhancement algorithm on ICDAR 2003 Robust OCR dataset [29]. The recognition accuracy rate from about 70% has increased to over 80%. Although the conclusion of the paper is obvious, the paper does not compare the recognition rate on the original data set. In addition, the ICDAR 2003 Robust OCR dataset contains a large number of low-quality images, while normal clear images account for only one-third of the data set. Under normal circumstances, the dataset of the CNN model is most the clear images. The accuracy of CNN with the bimodal image enhancement algorithm proposed in the paper is lower than the accuracy of CNN with traditional bimodal image enhancement. Therefore, the paper does not explain whether the image enhancement algorithm will affect the recognition rate of CNN on the normal image dataset.

In addition, one paper has shown that using image enhancement function with the Laplace operator can improve the performance of R-CNN and fast R-CNN in pedestrian detection task [16]. The experiment in this paper compares the original metro crowd image dataset with the enhanced image dataset with Laplace operator. The result is Laplace operator can increase the detection rate of 2% and 1% in the two R-CNN models respectively. The paper's experiments used transfer learning, fine-tuning pre-trained R-CNN model. However, the experiment did not use cross-validation and did not use hypothesis testing, which indicates that there is a high risk in the conclusion of the experiment. In addition, there is also a difference between target detection and image recognition.

4.1 Variable, Treatment, Object and Subject

In this project, the research method is the experiment. Because the project can be designed as a technology-oriented experiment, it is relatively easy to control, and all research questions can be concluded experimentally, so the experiment is the best research method. In this project, the experiment can be designed as one factor with more than two treatments. The subject of the experiment is the convolutional neural network models. The object of the experiment is the image data sets. The experimental factors (independent variable) and the treatments are the image enhancement methods. The dependent variable of the experiment is the performance of CNN model, this thesis use recognition accuracy and recall rate and f1 score to measure

4.2 Image data set

In order to research the performance of CNN models on RGB images and grayscale images, two data sets were selected in experiments. One of the data sets is an RGB image data set and the other is a grayscale image data set.

The RGB image dataset comes from the "Cat VS Dog" dataset in the Kaggle Challenge [25], which is typical of the nature image dataset. The training set contains 12,500 images of cats and 12,500 images of dogs, the number of images in the testing set is half of the training set, the label of the dog is 1, and the label of the cat is 0. Since the model chosen by the experiment is Lenet5, it is relatively simple and it is not easy to over-fit. In order to reduce the time spent on the experiment, this thesis only selected 2000 images for training the complete CNN model, in which the number of images of the cats is 1000 and the number of images for dogs is 1000. In addition, because the number of data sets required for transfer learning is small, this thesis used 1000 images for transfer learning, where the number of images for the cats is 500 and the number of images for the dogs is 500.

The grayscale image is from a dataset in Kaggle, which is a Chest X-ray image for detecting pneumonia [27]; it is a very typical medical image data set. Chest X-ray images were selected from pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. The data set contains thousands of Chest X-ray images diagnosed as normal and thousands of Chest X-ray images diagnosed as pneumonia X-ray images. Similarly, the experiment randomly selected 1000 normal images and 1000 pneumonia images for the experiment.

4.3 Image processing

4.3.1 Filter low-quality images

Because there are not many images in the data set, it is necessary to ensure that the images of the data set have high quality, so the experiment manually filter the original datasets and remove some images. In the "Cat VS Dog" dataset, some images have incorrect labels, such as images of dogs, are labeled as cat; and some images are cartoon images such as "hello kitty"; and some images have wrong subjects, for example, the subjects are humans rather than cats and dogs, they need to be removed. In the Chest X-ray data set, some images have 3 channels; it means these images are "RGB" images, so they need to be remove.

4.3.2 Image type conversion

The two selected data sets are all JPG images. Studies have shown that neural network image classifiers are vulnerable to adversarial images; JPG compression often reverses the drop in classification accuracy to a large extent [22]. In order to avoid the secondary compression of the original image in the image enhancement process affecting the CNN model, the JPG images need to be converted to PNG images before enhancing the original images.

4.3.3 Image channel conversion

Some studies have shown that the contrast limited adaptive histogram equalization (CLAHE) method, the successive means quantization transform (SMQT) method, the adaptive gamma correction method and the wavelet transform should use different strategies on grayscale images and RGB images. In the grayscale images, these image enhancement algorithms can act on the gray channel. But in the RGB images, it is wrong to strengthen the RGB channels, it will because it will destroy the original color distribution of the images. It is effective to convert the RGB image into hue-saturation-value. (HSV) channel and enhances the image by retaining the H channel, S channel and changing the V channel [4] [5]. In order to prevent the influence of the conversion from RGB images to HSV channels and from HSV channels to RGB channels, all the image data sets including the original images will be converted from RGB channels to HSV and convert back to RGB channels.

4.3.4 Image enhancement

The experiment performed image enhancement in the Matlab R2017a environment. The contrast limited adaptive histogram equalization method is used the function "adapthisteq()" in Matlab, which is the algorithm proposed by Zuiderveld, Karel [35]. The experiment uses successive means of the quantization transform method proposed by Julián Calderón González et al. [26]. The experiment uses adaptive gamma correction method proposed by Shih-Chia

Huang et al. [6]. The experiment uses wavelet transform method proposed by Ding X M et al. [24]. The image enhancement based on Laplace operator method uses the operator of $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ for edge extraction [30]. Figures compare the performance of five image enhancement algorithms in two data sets.

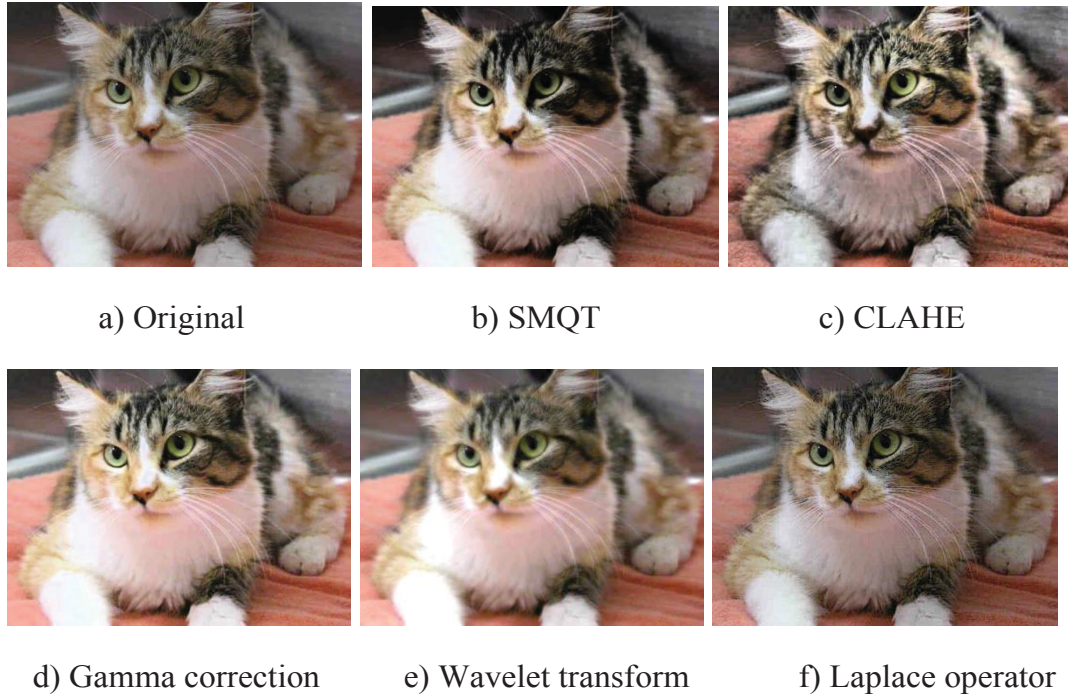


Fig4. Enhanced images in “Cat VS Dog” data set with image enhancement algorithms: a) Original image, b) Successive means quantization transform, c) Contrast limited adaptive histogram equalization, d) Adaptive gamma correction e) Wavelet transform, f) Laplace operator

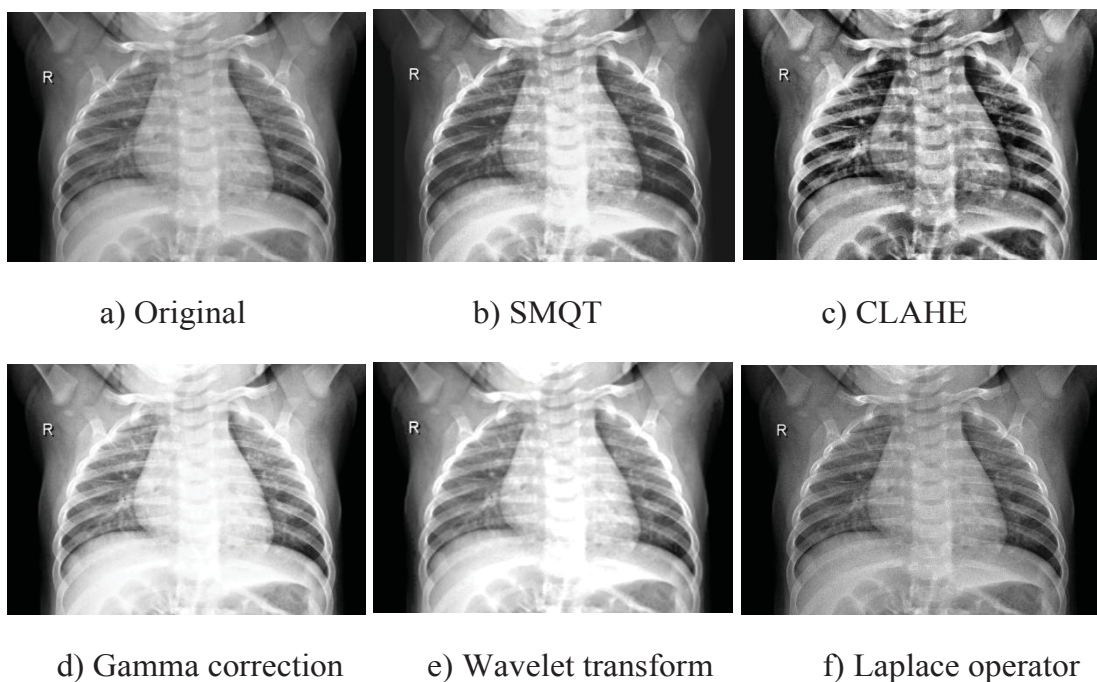


Fig5. Enhanced images in “Chest X-ray” data set with image enhancement algorithms: a) Original image, b) Successive means quantization transform, c) Contrast limited adaptive histogram equalization, d) Adaptive gamma correction e) Wavelet transform, f) Laplace operator

4.3.5 Summary

In the Chest X-ray data set, the original data set is called A1, the JPG images are first converted to PNG images, the data set is called A2, and this data set will serve as the blank control group in the experiment. The images in A2 are enhanced by CLAHE, SMQT, adaptive gamma correction, and Wavelet transform Laplace operator methods, the data set after enhancement are called A3, A4, A5, A6, and A7, respectively. The experiment will compare the performance of A2 and A3, A4, A5, A6, A7 in the CNN model.

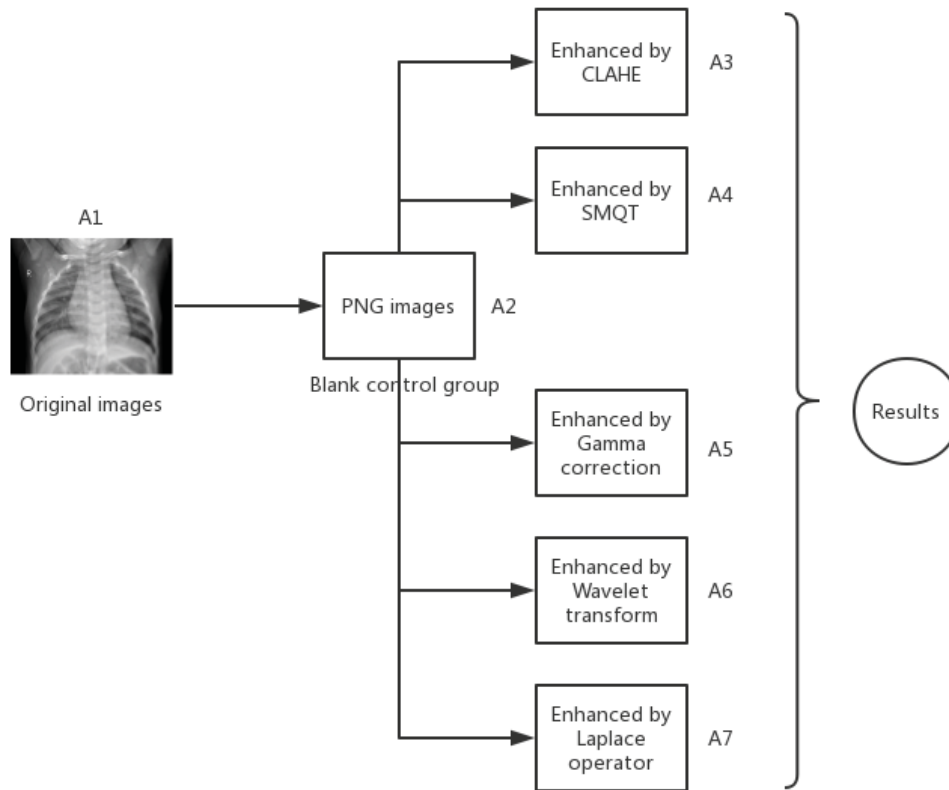


Fig6. Image processing of gray images

In the "Cat VS Dog" image data set, the original data set is called B1. The JPG images are first converted into PNG images, the data set called B2. All images in B2 were converted from RGB channels to HSV channels and converted back to RGB channels. This data set is called B3 and B3 will be used as the blank control group in the experiment. Convert B2 from RGB channel to HSV channel and enhance the “Value” channel by CLAHE, SMQT, adaptive gamma correction, Wavelet transform image enhancement respectively, and converts the enhanced

HSV channel to RGB channel, at this time, the data sets are called B5, B6, B7, B8 respectively. And enhance the images in B3 by Laplace operator, and the data set is called B4. The experiment will compare the performance of B3 and B4, B5, B6, B7, B8 in the CNN model.

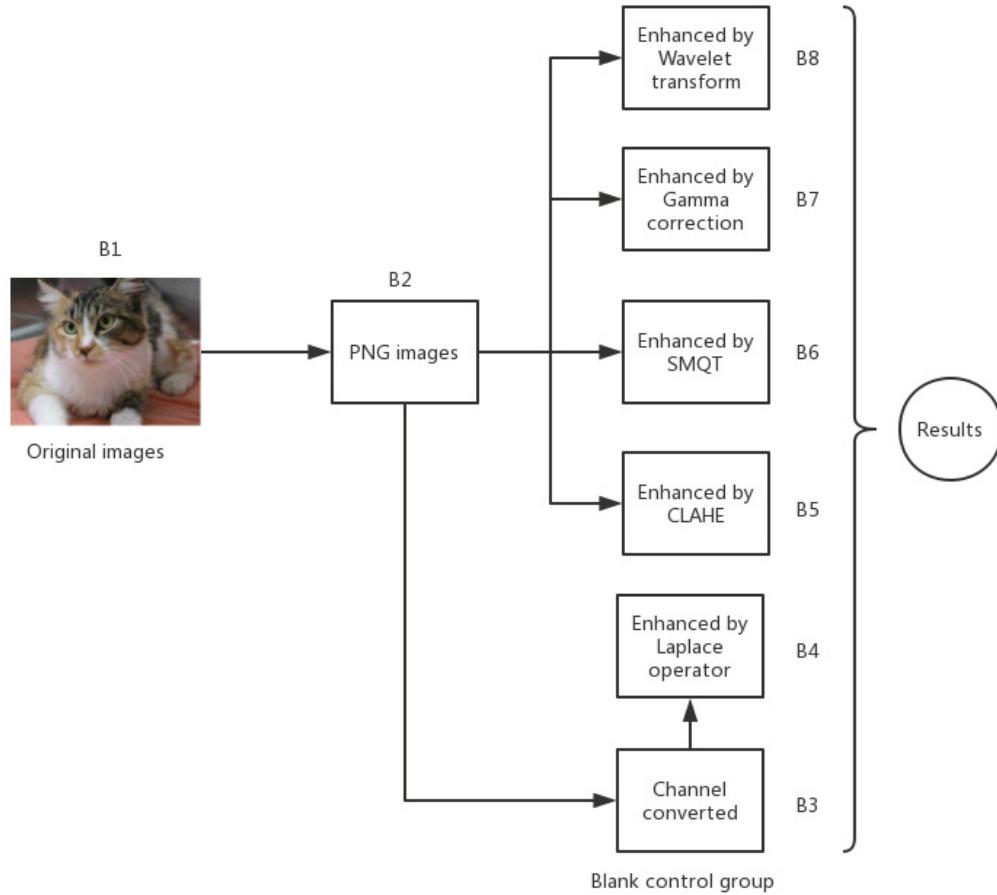


Fig7. Image processing of RGB images

4.4 CNN model

4.4.1 Improved Lenet5

There are two convolutional neural network models in the experiment. The first model is used to train the complete CNN model, and the second model is used for transfer learning. In the first model, the experiment chose a simpler convolutional neural network model. It is a modified LeNet5. The input layer is the matrix with a $250 \times 250 \times 3$ size. It has 5 layers, not including the input layer and the pooling layer, which are the “convolution layer--pooling layer--convolution layer--pooling layer--fully connected layer--fully connected layer--output layer

(softmax)”. The size of the convolution kernel is 3×3 and the step size is 1, and the number of convolution boxes is 16 in each convolution layer. The pooling layers use max pooling and the kernel size is 2×2 , the step size is 2. In fully connected layers there are 128 neurons in each layer, and the output layer uses the softmax as a classifier. The size of all parameters in the model is approximately 300M. In addition, because the model performance not stable during training, in order to improve the stability of the model, the model uses the “ELU” function as the activation function. In addition, in order to improve the training speed of the model, the model uses “Adam” as the optimization algorithm. The figure below is a simplified schematic of the model.

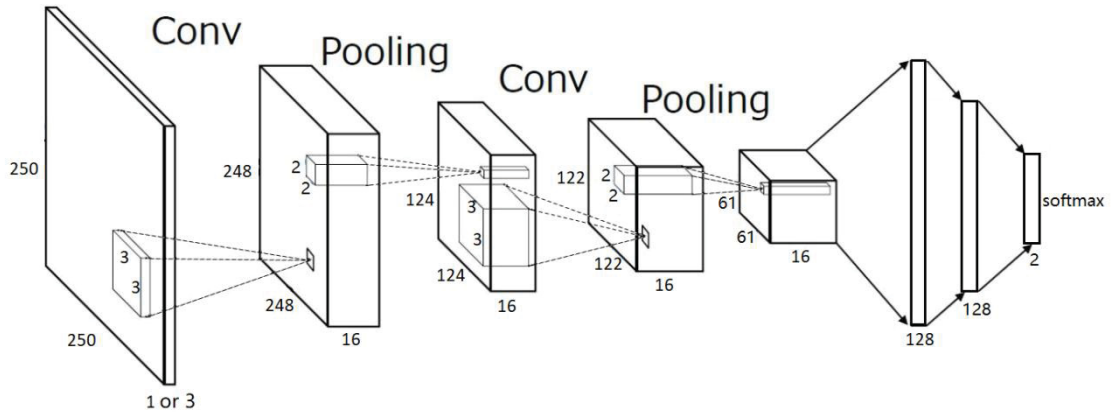


Fig8. Structure of improved Lenet5 model in the experiment [14]

4.4.2 Pre-trained VGG16

The experiment selected the VGG16 model in transfer learning. VGG16 has 16 layers in total, not including the input layer and the pooling layer, 13 of which are convolutional layers, 2 of which are fully connected layers and one of which is the output layer (softmax). The input layer is the matrix with a $224 \times 224 \times 3$ size, The size of convolution kernel is 3×3 and the step size is 1; the pooling layer uses max pooling, kernel size is 2×2 , the step size is 2. There are 4096 neurons in each fully connected layer, and softmax is used as the classifier in the output layer. All parameters of the model exceed 500M. In addition, since training the model requires a long training time, in order to make the model train faster, the model uses the “ReLU” function as the activation function and the model still uses the Adam optimization algorithm. The reason for choosing VGG16 is that the structure of the model is very easy to understand, the role of each layer is very clear. Like the Lenet5 model, the convolutional layers and the pooled layers of VGG16 are equivalent to do feature extraction on the images. Before entering the fully connected layer, the shape of the matrix is $7 \times 7 \times 512$, which is much smaller than $224 \times 224 \times 3$. The fully connected layers and softmax are used to classify the extracted features. VGG16 can achieve high accuracy, its parameters are less than VGG19 but the performance is not much different from VGG19 [15]. In the experiment, the parameters of all convolutional layers of VGG16 are frozen, and all the parameters of the fully connected layers and the softmax layer are

the X-ray grayscale images to train the complete CNN model in the improved Lenet5 model. Experiment 2 uses the "Cat VS Dog" RGB images to train the complete CNN model in the improved Lenet5 model. Experiment 3 uses the "Cat VS Dog" RGB images to do transfer learning in the pre-trained VGG16 model, freeze the convolution layers and fine-tune full connection layers and output layer.

Experiments were done on a computer with an i7-8750H CPU @2.20GHz, with 16.0GB of RAM (not all of it was used during runtime), and an NVIDIA GeForce GTX 1070 GPU. Our CNN classification implementation relies on the usage of Python (3.5) along with the TensorFlow (1.8.0) frame.

4.6.1 Experiment 1

Using Chest X-ray images as data sets, experiments were performed on the improved LeNet5 model using five different image enhancement algorithms. Since the sizes of the original images are large and different, the length and width of the images are basically between 1000 pixels and 2000 pixels. In order to reduce the training time of the CNN model and unify the size of the input layer, the experiment resizes the size of images to 250×250, the experimental results show that the image has a very high accuracy after resize. The experiment uses a learning rate of 0.0001, the number of iterations of 300, and the “Batchsize” of 200. Since in this experiment, all pneumonia images need to be detected as much as possible, the recall rate is more important than the precision rate. In many medical fields, the recall rate is often used as the metric to improve the rate of disease recognition or detection. For example, in this paper on breast cancer detection, recall rate is used as the metric instead of precision rate and f-measure [40]. So, the experiment uses accuracy and recall rate as metrics of performance. The table below shows the results of the experiment.

Accuracy (%)	1	2	3	4	5	6	7	8	9	10	Mean	Median
Original	96.9	98.2	95.5	95.9	97.3	98.6	99.2	98.4	96.2	97.7	97.39	97.5
SMQT	97.5	98.2	94.3	97.0	96.4	98.2	99.1	97.7	94.5	98.1	97.1	97.6
CLAHE	97.9	98.2	95.4	97.7	98	99	99.5	98.5	94.5	97.7	97.64	97.95
Gamma	96.7	97.9	95.0	96.7	97	99.1	98.6	97	96	97.5	97.15	97
Wavelet	97.5	98.3	95.9	96.0	97.6	98.6	99.1	98.5	95.9	97.4	97.48	97.55
Laplace	98.0	98.2	97.0	98.0	99.0	100	99.3	99.0	93.3	98.6	98.04	98.4

Table1. Accuracy in 10-fold cross-validation of experiment1

Recall (%)	1	2	3	4	5	6	7	8	9	10	Mean	Median
Original	96.3	97.5	92.6	95	97.5	98.6	98.5	99.2	99.5	95.7	97.04	97.5
SMQT	96.9	97.3	90.5	99	96.7	97.1	98.9	97.9	96.6	96.5	96.74	97
CLAHE	97.8	97.4	92.1	98	98	98	99	98	97.3	96.4	97.2	97.9
Gamma	96.1	97	91.8	98.2	97.5	98.7	98.1	99.1	99.2	95.6	97.13	97.8
Wavelet	96.1	97.5	93.3	96.5	97.9	98.8	98.2	100	98.4	95.7	97.24	97.7
Laplace	97.0	96.6	95.0	97.0	99.0	100	98.6	98.2	100	99.0	98.04	98.4

Table2. Recall rate in 10-fold cross-validation of experiment1

4.6.2 Experiment 2

Using the "Cat VS Dog" images as the data set, experiments were performed on the improved LeNet5 model using five different image enhancement algorithms. The length and width of the original dataset image range from 100 pixels to 500 pixels and they are resized to 250×250. The experiment uses a learning rate of 0.0001, the number of iterations of 500, and the "Batchsize" of 200. Since the two types of data set do not have a bias, the accuracy and F1 score are used as metrics to measure performance. The following table shows the results of the experiment.

Accuracy (%)	1	2	3	4	5	6	7	8	9	10	Mean	Median
Original	70.3	69.7	68.9	70.5	74.1	67.9	67.5	69.8	70.3	70.9	69.99	70.05
SMQT	72.9	70.7	70.5	70.5	69.6	68	71.2	71.1	69.5	66.1	70.01	70.5
CLAHE	72.8	68.5	67.1	71	71	70.7	68.1	67.7	72.4	69.7	69.9	70.2
GAMMA	78.2	67.4	64.5	70.8	74.3	70.1	71.5	68.2	69.5	67	70.15	69.8
Wavelet	71.8	67.2	66.2	68.1	75.1	68.8	70.6	67.5	71.6	67.2	69.41	68.45
Laplace	73.3	70.0	67.7	72.1	72.0	67	71.5	67.6	73.3	72.7	70.72	71.75

Table3. Accuracy in 10-fold cross-validation of experiment2

F1 score (%)	1	2	3	4	5	6	7	8	9	10	Mean	Median
Original	71.2	69.1	70.1	71.5	73.6	66.7	66.6	70.2	68.8	70.7	69.85	70.15
SMQT	70.9	71.3	70.3	70.5	68.3	66.7	72	71.1	66.5	65.7	69.33	70.4
CLAHE	72.2	67.6	66.7	71.5	73	69.8	67.7	67.1	69.7	70.7	69.6	69.75
GAMMA	77.7	66.8	66.9	70.5	74.3	67.7	70.1	67.1	67.5	67.7	69.63	67.7
Wavelet	68.3	67.0	65.6	68.7	74.5	67.1	70.1	66.6	71.1	67.6	68.66	67.95
Laplace	71.9	68.5	67.4	73.5	72.5	64.8	71.4	68.2	71.3	72.7	70.22	71.35

Table4. F1 score in 10-fold cross-validation of experiment2

4.6.3 Experiment 3

Using the "Cat VS Dog" images as the data set, experiments were performed transfer learning on the pre-trained VGG16 model using five different image enhancement algorithms. And the image was resized to 224×224 before training. The experiment used the parameters in the VGG16 model convolutional layers and frozen all convolutional layers. Remove all fully connected layers and output layer and build new fully connected layers and output layer for training. The experiment uses a learning rate of 0.00005, the number of iterations of 1500, and the "Batchsize" of 100. Accuracy and F1 score were used as performance metrics. The following table is the experimental results.

Accuracy (%)	1	2	3	4	5	6	7	8	9	10	Mean	Median
Original	96.7	96.0	91.2	85.3	92.3	88.9	94.7	91.4	93.2	91.6	92.13	91.95
SMQT	96.2	94.6	89.2	88.4	88.3	89.5	86.9	92.2	85.6	88.5	89.94	88.85
CLAHE	94.8	95.6	90.5	85.1	92.4	90.1	94.3	90.3	87.9	89.2	91.02	90.4
Gamma	94.9	93.1	90.5	89.0	92.2	91.0	91.6	91.4	87.9	89.2	91.08	91.2
Wavelet	91.0	93.3	88.5	86.7	90.2	88.4	91.1	91.9	87.2	88.0	89.63	89.35
Laplace	92.0	93.0	85.5	89.0	92.3	90	88	90	88.4	90.3	89.85	90

Table5. Accuracy in 10-fold cross-validation of experiment3

F1 score (%)	1	2	3	4	5	6	7	8	9	10	Mean	Median
Original	96.4	95.8	91.0	83.5	92.5	88.0	94.5	90.7	93.0	91.0	91.64	91.75
SMQT	96.2	94.5	89.1	87.8	87.9	88.8	85.9	91.5	84.7	87.6	89.4	88.35
CLAHE	94.7	95.5	90.2	83.5	92.2	89.5	93.5	89.7	87.2	88.5	90.45	89.95
Gamma	94.9	92.6	89.5	88.4	92.0	90.3	90.7	91.0	87.0	88.5	90.49	90.5
Wavelet	90.0	93.0	87.2	85.4	89.7	87.8	90.3	91.3	85.7	87.2	88.76	88.75
Laplace	91.7	92.6	85.1	88.4	91.8	89	86.9	89.8	87.7	89.6	89.26	89.3

Table6. F1 in 10-fold cross-validation of experiment3

Chapter 5

Result and Discussion

5.1 Statistical hypothesis testing

Since the experiments use ten-fold cross-validation, ten results are produced for each experiment. So the thesis used the paired T-test for hypothesis testing. And there is a report shows the variance of k-folds cross-validation cannot be unbiasedly estimated [31]. That means the hypothesis testing should use the population variance and population standard deviation rather than the sample variance and sample standard deviation. In the paired T-test, it is necessary to compare the relationship between T_t and $t_{\alpha,k-1}$ in one-sided test, and reject the null hypothesis when T_t is larger than (or equal with) $t_{\alpha,k-1}$, where α is the confidence interval. When the α is 0.05, the $t_{0.05,9}$ is 1.83 [41]. The formula for T_t is as follows [34], where μ is the mean value of the differences and σ is the population standard deviation of the differences:

$$T_t = \left| \frac{\sqrt{K}\mu}{\sigma} \right| \quad (4)$$

In “Experiment 1”, comparing the accuracy between original images and the images enhanced by SMQT, the $T_t \approx \left| \frac{0.29 \times \sqrt{10}}{0.82} \right| \approx 1.12$, because 1.12 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by CLAHE, the $T_t \approx \left| \frac{0.25 \times \sqrt{10}}{0.85} \right| \approx 0.93$, because 0.93 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by adaptive gamma correction, the $T_t \approx \left| \frac{0.24 \times \sqrt{10}}{0.564} \right| \approx 1.35$, because 1.35 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by wavelet transform, the $T_t \approx \left| \frac{0.09 \times \sqrt{10}}{0.27} \right| \approx 1.05$, because 1.05 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by Laplace operator, the $T_t \approx \left| \frac{0.65 \times \sqrt{10}}{1.345} \right| \approx 1.53$, because 1.53 is smaller than 1.83, so there is no significant difference. Comparing the recall rate between original images and the images enhanced by SMQT, the $T_t \approx \left| \frac{0.3 \times \sqrt{10}}{1.84} \right| \approx 0.52$, because 0.52 is smaller than 1.83, so there is no significant difference. Comparing the recall rate between original images and the images enhanced by CLAHE, the $T_t \approx \left| \frac{0.16 \times \sqrt{10}}{1.374} \right| \approx 0.37$, because 0.37 is smaller than 1.83, so there is no significant difference. Comparing the recall rate between original images and the images enhanced by adaptive gamma correction, the $T_t \approx \left| \frac{0.09 \times \sqrt{10}}{1.066} \right| \approx 0.27$, because 0.27 is smaller than 1.83, so there is no significant difference. Comparing the recall rate between original images and the images enhanced by wavelet transform, the $T_t \approx \left| \frac{0.2 \times \sqrt{10}}{0.67} \right| \approx 0.94$, because 0.94 is smaller than 1.83, so there is no significant difference. Comparing the recall rate between original images and the images enhanced by Laplace operator, the

$T_t \approx |\frac{1 \times \sqrt{10}}{1.32}| \approx 2.40$, because 2.40 is bigger than 1.83, so there is a significant difference between the recall of original images and the recall of the images enhanced by Laplace operator. Because the recall rate between original images and the images enhanced by Laplace operator has a significant difference, this thesis also compares the accuracy and recall rate between Laplace operator and other algorithms. Comparing the accuracy between the Laplace operator and the SMQT, the $T_t \approx |\frac{0.94 \times \sqrt{10}}{1.15}| \approx 2.58$, because 2.58 is larger than 1.83, so there is a significant difference. Comparing the accuracy between the Laplace operator and the CLAHE, the $T_t \approx |\frac{0.4 \times \sqrt{10}}{0.75}| \approx 1.69$, because 1.69 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between Laplace operator and the adapt gamma correction, the $T_t \approx |\frac{0.89 \times \sqrt{10}}{1.32}| \approx 2.13$, because 2.13 is larger than 1.83, so there is a significant difference. Comparing the accuracy between the Laplace operator and the wavelet transform, the $T_t \approx |\frac{0.56 \times \sqrt{10}}{1.21}| \approx 1.46$, because 1.46 is smaller than 1.83, so there is no significant difference. Comparing the recall rate between the Laplace operator and the SMQT, the $T_t \approx |\frac{1.3 \times \sqrt{10}}{1.90}| \approx 2.16$, because 2.16 is larger than 1.83, so there is a significant difference. Comparing the recall rate between the Laplace operator and the CLAHE, the $T_t \approx |\frac{0.84 \times \sqrt{10}}{1.44}| \approx 1.84$, because 1.84 is larger than 1.83, so there is a significant difference. Comparing the recall rate between Laplace operator and the adapt gamma correction, the $T_t \approx |\frac{0.91 \times \sqrt{10}}{1.40}| \approx 2.06$, because 2.06 is larger than 1.83, so there is a significant difference. Comparing the recall rate between Laplace operator and the SMQT, the $T_t \approx |\frac{0.8 \times \sqrt{10}}{1.27}| \approx 1.99$, because 1.99 is larger than 1.83, so there is a significant difference.

In “Experiment 2”, comparing the accuracy between original images and the images enhanced by SMQT, the $T_t \approx |\frac{0.02 \times \sqrt{10}}{2.64}| \approx 0.02$, because 0.02 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by CLAHE, the $T_t \approx |\frac{0.09 \times \sqrt{10}}{1.98}| \approx 0.14$, because 0.14 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by adaptive gamma correction, the $T_t \approx |\frac{0.16 \times \sqrt{10}}{3.55}| \approx 0.41$, because 0.41 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by wavelet transform, the $T_t \approx |\frac{0.58 \times \sqrt{10}}{2.24}| \approx 0.82$, because 0.82 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by Laplace operator, the $T_t \approx |\frac{0.73 \times \sqrt{10}}{2.15}| \approx 1.08$, because 1.08 is smaller than 1.83, so there is no significant difference. Comparing the f1 score between original images and the images enhanced by SMQT, the $T_t \approx |\frac{0.52 \times \sqrt{10}}{3.03}| \approx 0.54$, because 0.54 is smaller than 1.83, so there is no significant difference. Comparing the f1 score between original images and the images enhanced by CLAHE, the $T_t \approx |\frac{0.25 \times \sqrt{10}}{1.89}| \approx 0.42$, because 0.42 is smaller

than 1.83, so there is no significant difference. Comparing the f1 score between original images and the images enhanced by gamma correction, the $T_t \approx |\frac{0.22 \times \sqrt{10}}{3.04}| \approx 0.23$, because 0.23 is smaller than 1.83, so there is no significant difference. Comparing the f1 score between original images and the images enhanced by wavelet transform, the $T_t \approx |\frac{1.19 \times \sqrt{10}}{2.60}| \approx 1.45$, because 1.45 is smaller than 1.83, so there is no significant difference. Comparing the f1 score between original images and the images enhanced by Laplace operator, the $T_t \approx |\frac{0.37 \times \sqrt{10}}{2.30}| \approx 0.51$, because 0.51 is smaller than 1.83, so there is no significant difference.

In “Experiment 3”, comparing the accuracy between original images and the images enhanced by SMQT, the $T_t \approx |\frac{2.19 \times \sqrt{10}}{3.36}| \approx 2.06$, because 2.06 is bigger than 1.83, so the accuracy of the original images is significantly better than the accuracy of images enhanced by SMQT. Comparing the accuracy between original images and the images enhanced by CLAHE, the $T_t \approx |\frac{1.11 \times \sqrt{10}}{1.70}| \approx 2.06$, because 2.06 is bigger than 1.83, so the accuracy of original images is significantly better than the accuracy of images enhanced by CLAHE. Comparing the accuracy between original images and the images enhanced by adaptive gamma correction, the $T_t \approx |\frac{1.05 \times \sqrt{10}}{2.50}| \approx 1.33$, because 1.33 is smaller than 1.83, so there is no significant difference. Comparing the accuracy between original images and the images enhanced by wavelet transform, the $T_t \approx |\frac{2.5 \times \sqrt{10}}{2.31}| \approx 3.42$, because 3.42 is bigger than 1.83, so the accuracy of original images is significantly better than the accuracy of images enhanced by wavelet transform. Comparing the accuracy between original images and the images enhanced by Laplace operator, the $T_t \approx |\frac{2.28 \times \sqrt{10}}{3.13}| \approx 2.30$, because 2.30 is bigger than 1.83, so the accuracy of original images is significantly better than the accuracy of images enhanced by Laplace operator. Comparing the f1 score between original images and the images enhanced by SMQT, the $T_t \approx |\frac{2.24 \times \sqrt{10}}{3.88}| \approx 1.83$, because 1.83 is equal with 1.83, so the accuracy of original images is significantly better than the accuracy of images enhanced by SMQT. Comparing the f1 score between original images and the images enhanced by CLAHE, the $T_t \approx |\frac{1.19 \times \sqrt{10}}{1.84}| \approx 2.05$, because 2.05 is bigger than 1.83, so the accuracy of original images is significantly better than the accuracy of images enhanced by CLAHE. Comparing the f1 score between original images and the images enhanced by adaptive gamma correction, the $T_t \approx |\frac{1.15 \times \sqrt{10}}{2.96}| \approx 1.23$, because 1.23 is smaller than 1.83, so there is no significant difference. Comparing the f1 score between original images and the images enhanced by wavelet transform, the $T_t \approx |\frac{2.88 \times \sqrt{10}}{2.78}| \approx 3.28$, because 3.28 is bigger than 1.83, so the accuracy of original images is significantly better than the accuracy of images enhanced by wavelet transform. Comparing the f1 score between original images and the images enhanced by Laplace operator, the $T_t \approx |\frac{2.38 \times \sqrt{10}}{3.53}| \approx 2.13$, because 2.13 is bigger than 1.83, so the accuracy of original images is significantly better than the accuracy of images enhanced by Laplace operator.

5.2 Result

In “Experiment 1”, although CLAHE and wavelet transform and Laplace operator image enhancement algorithms can improve the mean accuracy and recall rates, according to paired T-test, only Laplace operator image enhancement has a significant difference on recall rate with the original data set. There are no significant differences in all other accuracy and recall rates. And comparing enhanced images, in accuracy, the Laplace operator has a significant difference from SMQT and gamma correction but has no significant difference from CLAHE and wavelet transform. In recall rate, the Laplace operator is significantly larger than the other four image enhancement algorithms. And the mean accuracy, mean recall, median accuracy and median recall of Laplace operator are larger than others. So in this data set, using the Laplace operator to enhance the images can improve the performance of the CNN on the Recall rate, and the Laplace operator is the best algorithm in these five image enhancement algorithms.

In “Experiment 2”, the mean F1 score of the data set enhanced by Laplace operator is the highest, the original data set was higher than the mean F1 scores of data sets enhanced by the other four algorithms. The median F1 score of the data sets using SMQT and Laplace operator are higher than the original data set, and the data sets using CLAHE, gamma correction and wavelet transform are lower than original data set. On mean accuracy, the data sets using SMQT, adaptive Gamma correction, and Laplace operator are higher than the original dataset; data sets using CLAHE and wavelet transform are lower than the original dataset. On median accuracy, the data sets using SMQT, CLAHE and Laplace operator are higher than the original data set, and data sets using gamma correction and wavelet transform are lower than the original data set. However, all data sets have no significant difference in accuracy and F1 score.

In “Experiment 3”, in the mean accuracy, median accuracy, mean F1 score and median F1 score, all data sets processed by the image enhancement algorithms are lower than the original data set. And after paired T-test, it was found that only the accuracy and F1 score of the data set using adaptive Gamma-corrected have no significant difference from the accuracy and f1 score of the original data set. And the accuracy and F1 scores of the data sets processed by the other four image enhancement algorithms are significantly lower than the accuracy and the F1 score of the original data set.

5.3 Discussion

From the results of “Experiment 1”, we can know that the image enhancement algorithm based on Laplace operator can improve the Recall rate of CNN on X-ray images when training the complete CNN model. From this, we can reject the null hypothesis 1 and accept the alternative hypothesis 1. However, not all image enhancement algorithms can improve the performance of CNN. Only some suitable image enhancement algorithms can improve the performance of CNN. Compare Laplace operator with other algorithms, the results show that Laplace

operator is the best algorithm in these five image enhancement algorithms, which means on one data set, different image enhancement algorithms have different performance. After further experiments, it is found that using both the original data set and the data set enhanced by the Laplace operator has bad performance; the mean accuracy and the recall rate are both 97.21%, which is worse than using only the original image data set or the enhanced image data. Among the five image enhancement algorithms, only the Laplace operator image enhancement algorithm enhances the edge information of the X-ray image. In Experiment 1, edge enhancement was positive for recall rates. In practical applications, it is unrealistic to try to perform multiple image enhancement algorithms on a data set. Therefore, if there is no research to prove that an image enhancement algorithm is effective in a certain data set; it is better to use the original data set.

In “Experiment 2”, after paired T-test, there was no significant difference between the image enhanced by the Laplace operator and the original image in accuracy and F1 score. Comparing with the results of Experiment 1, it can be concluded that an image enhancement algorithm has different effects on the CNN model in different data sets. However, in the mean and median accuracy, and the mean and median of the F1 score, it can be found that the Laplace operator is still the highest group in Experiment 2.

From the results of “Experiment 3”, we know that SMQT, CLAHE, Wavelet transform and Laplace Operator will reduce the performance of VGG16 when fine-tuning. The image data set enhanced by Adapt Gamma Correction has no significant difference from the original image data set. But the mean accuracy, median accuracy, mean F1 score and median F1 score of original image data set are all higher than image data set enhanced by Adapt Gamma Correction. The possible reason is that the VGG16 model was trained on the ImageNet dataset [2], the images in the data set are all natural images. Therefore, the parameters in the convolutional layers are more suitable for extracting the features of the natural images, rather than extracting the features of the enhanced images. Therefore, in practical applications, when fine-tuning on natural images, it is better not to use image enhancement algorithms for image preprocessing. Image enhancement algorithms are likely to reduce the performance of pre-trained CNN models.

5.4 Threat analysis

In this thesis, experiments use the appropriate metrics and paired T-test to ensure the conclusion validity. In addition, in order to reduce the threat of internal validity, the experiments carried out 10-fold cross validation test which can reduce the contingency of the experiment.

The threat of the experiment mainly comes from the external validity of the experiment. Since there are too many strategies on one image enhancement algorithm, such as the adaptive gamma correction method, there are many papers publish different kinds of adaptive gamma correction methods. Laplace operator also has many kinds of operators. Since the experiment uses one of the Laplace operators, the conclusion can't be applied to all image enhancement algorithms

based on Laplace operators. And because there are many strategies to transfer learning, the conclusions of the experiment only apply to the fine-tuning method, not all the transfer learning methods. In addition, since the experiment only selected two data sets, and two CNN models, the conclusions are not necessarily applicable to all X-ray data sets and all CNN models.

Chapter 6

Conclusion and Future work

6.1 Conclusion

In this paper, experiments have shown that the image enhancement algorithm based on Laplace operator get a positive conclusion on the "X-ray" data set when training the complete CNN models. It is proved that when training a complete CNN model, there is the possibility of using image enhancement algorithms to improve the performance of CNN models. In addition, experiments found that CLAHE, SMQT, adaptive gamma correction, and wavelet transform, four image enhancement algorithms did not perform positively on the "X-ray" data set when training the complete CNN models. In addition, in the X-ray dataset, the Laplace operator is significantly better than the other four image enhancement algorithms. It is proved that different image enhancement algorithms performance differently on the same CNN model, and not all image enhancement algorithms can improve the performance of CNN models when training a complete CNN model. In the RGB data set, all the enhanced image data set by five image enhancement algorithms including Laplace operator have no significant difference from the original data set. It is proved that one image enhancement algorithm in the different data set will influence the performance of CNN model differently.

In addition, experiments have shown that using "fine-tuning" to do the transfer learning in CNN models, CLAHE, SMQT, wavelet transform, Laplace operator, four image enhancement algorithms have the negative performance on the natural image data set, which will reduce the performance of the CNN models. Although the image data set using adapt gamma correction has no significant difference from the original data set, all of mean accuracy median accuracy, mean F1 score and median F1 score are lower than the original data set. It is proved that when using fine-tuning to do the transfer learning in CNN models, there is the possibility of using image enhancement algorithms to reduce the performance of pre-trained CNN models.

6.1.1 Answering RQs

RQ1

Increasing image intensity can improve the performance in image recognition of Convolutional neural network when training a complete Convolutional neural network, but not all image enhancement algorithms can improve the performance of CNN, and it is necessary to select an appropriate image enhancement algorithm according to the data set.

RQ2

Increasing image intensity can reduce the performance in image recognition of the Convolutional neural network when using transfer learning (fine-tuning). Although not all image enhancement algorithms will reduce the performance of CNN, it is better not to use image enhancement algorithms on transfer learning.

6.1.2 Hypothesis testing

According to the paired T-test, it was found that when training a complete CNN model, using the Laplace operator on the “X-ray” dataset can improve the performance of the CNN model. Therefore, this thesis can reject the “null hypothesis 1” and accept the “alternative hypothesis 1”: Increasing image intensity can improve the performance in image recognition of the Convolutional neural network when training a complete Convolutional neural network.

According to the paired T-test, it was found that in the transfer learning, using the SMQT, CLAHE, Wavelet transform and Laplace Operator on the "Cats VS Dogs" data set will reduce the performance of the CNN model. Therefore, this thesis can reject the “null hypothesis 2” and accept the “alternative hypothesis 2”: Increasing the image intensity will reduce the performance in image recognition of the Convolutional neural network when using transfer learning (fine-tuning).

6.2 Future Work

In order to reduce the external validity threat of the experiment, more data sets and more image enhancement algorithms will be used in the experiment in the future work. In addition, the image enhancement algorithms will be chosen based on the characteristics of the data sets. Moreover, using more CNN models and more strategies of transfer learning can also reduce external effectiveness threats in experiments. Moreover, more experiments can be designed to specifically research the impact of Laplace operators on the performance of CNN models.

References

- [1] J. Cao, Z. Su, L. Yu, D. Chang, X. Li, and Z. Ma, “Softmax Cross Entropy Loss with Unbiased Decision Boundary for Image Classification,” in 2018 Chinese Automation Congress (CAC), Xi’an, China, 2018, pp. 2028–2032.
- [2] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 248–255.
- [3] A. Giusti, D. C. Ciresan, J. Masci, L. M. Gambardella, and J. Schmidhuber, “Fast image scanning with deep max-pooling convolutional neural networks,” in 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 2013, pp. 4034–4038.
- [4] M. Hanmandlu and D. Jha, “An Optimal Fuzzy System for Color Image Enhancement,” *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2956–2966, Oct. 2006.
- [5] M. Hanmandlu, O. P. Verma, N. K. Kumar, and M. Kulkarni, “A Novel Optimal Fuzzy System for Color Image Enhancement Using Bacterial Foraging,” *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 8, pp. 2867–2879, Aug. 2009.
- [6] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, “Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution,” *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
- [7] L. Jian-Min and Y. Min-Hua, “The signal processing and recognition of street view images by CNNs and softmax,” in 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), Guangzhou, 2017, pp. 875–879.
- [8] J.-H. Kim, “Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap,” *Computational Statistics & Data Analysis*, vol. 53, no. 11, pp. 3735–3745, Sep. 2009.
- [9] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Dec. 2014.
- [10] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-Margin Softmax Loss for Convolutional Neural Networks,” *arXiv:1612.02295 [cs, stat]*, Dec. 2016.
- [11] R. Maini and H. Aggarwal, “A Comprehensive Review of Image Enhancement Techniques,” vol. 2, no. 3, p. 6, 2010.
- [12] Mikael Nilsson, Mattias Dahl, and Ingvar Claesson, “The Successive Mean Quantization Transform,” in *Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, Philadelphia, Pennsylvania, USA, 2005, vol. 4, pp. 429–432.

- [13]P. M. Radiuk, "Impact of Training Set Batch Size on the Performance of Convolutional Neural Networks for Diverse Datasets," *Information Technology and Management Science*, vol. 20, no. 1, Jan. 2017.
- [14]F. Rayar, M. Goto, and S. Uchida, "CNN Training with Graph-Based Sample Preselection: Application to Handwritten Character Recognition," in 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), Vienna, 2018, pp. 19–24.
- [15]K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556 [cs], Sep. 2014.
- [16]Q. Tian, G. Xie, Y. Wang, and Y. Zhang, "Pedestrian Detection Based on Laplace Operator Image Enhancement Algorithm and Faster R-CNN," in 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2018, pp. 1–5.
- [17]T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015.
- [18]R. Zaheer and H. Shaziya, "GPU-based empirical evaluation of activation functions in convolutional neural networks," in 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, 2018, pp. 769–773.
- [19]Q. Zheng, M. Yang, J. Yang, Q. Zhang, and X. Zhang, "Improvement of Generalization Ability of Deep CNN via Implicit Regularization in Two-Stage Training Process," *IEEE Access*, vol. 6, pp. 15844–15869, 2018.
- [20]Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, "Fine-Tuning Convolutional Neural Networks for Biomedical Image Analysis: Actively and Incrementally," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 4761–4772.
- [21]Y. Zhu, J. Sun, and S. Naoi, "Recognizing Natural Scene Characters by Convolutional Neural Network and Bimodal Image Enhancement," in *Camera-Based Document Analysis and Recognition*, vol. 7139, M. Iwamura and F. Shafait, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 69–82
- [22]GK Dziugaite, Z. Ghahramani and DM Roy, "A study of the effect of JPG compression on adversarial images," 2016, arXiv preprint arXiv:1608.00853.
- [23]C. Wohlin et al, *Experimentation in Software Engineering*. (2012th ed.) Berlin: Springer, 20129783642290442. P134-P150
- [24]DING X M. "The study of image enhancement algorithm based on wavelet transform". Anhui University, 2010. (in Chinese)
- [25]<https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/overview>
- [26]J. Calderón González and Ò. D. Carmona Salazar, "Image Enhancement with Matlab Algorithms," 2015.

- [27]Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-ray Images for Classification", Mendeley Data, v2
- [28]Liu Wanjun, Liang Xuejian, Qu Haicheng, "Learning performance of convolutional neural networks with different pooling models" National Natural Science Foundation Of China, 1006-8961 (2016) 09-1178-13
- [29]Lucas, SM, Panaretos, A., Sosa , L., Tang, A., Wong, S., Young, R.: ICDAR 2003 robust reading competitions. In: 7th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, vol. 2, pp. 682–687 (2003)
- [30]M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis, and Machine Vision. Cengage Learning, 2014. P137
- [31]Grandvalet Y, Bengio Y. Hypothesis testing for cross-validation[J]. Montreal Universite de Montreal, Operationnelle DdIeR, 2006, 1285.
- [32]Reza A M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement[J]. Journal of VLSI signal processing systems for signal, image and video technology, 2004, 38(1): 35-44.
- [33]S. Du and RK Ward, "Adaptive Region-Based Image Enhancement Method for Robust Face Recognition Under Variable Illumination Conditions," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, (9), pp. 1165-1175, 2010.
- [34]ZhiHua Zhou "Machine learning" Tsinghua University Press, 2016, P23-P47.
- [35]Zuiderveld, Karel. "Contrast Limited Adaptive Histogram Equalization." Graphic Gems IV. San Diego: Academic Press Professional, 1994. 474–485.
- [36]K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385 [cs], Dec. 2015.
- [37]Ding Han “VGG16 learning notes” [<http://deanhan.com/2018/07/26/vgg16/>]
- [38]K. Koo and E. Cha, "Image recognition performance enhancements using image normalization," Human-Centric Computing and Information Sciences, vol. 7, (1), pp. 1-11, 2017.
- [39]Z. Zhang et al, "Comparison of Fourier transform, windowed Fourier transform, and wavelet transform methods for phase calculation at discontinuities in fringe projection profilometry," Optics and Lasers in Engineering, vol. 50, (8), pp. 1152-1160, 2012.
- [40]Z. Z. Ang *et al*, "Classification of normal screening mammograms is strongly influenced by perceived mammographic breast density," *Journal of Medical Imaging and Radiation Oncology*, vol. 61, (4), pp. 461-469, 2017.
- [41]Milton, J.S and Arnold, J. (2002). Introduction to probability and statistics, McGraw-Hill, Appendix: Statistical Tables.