A

Summer Internship Report on

MNIST Handwritten Digit Recognition

Submitted to

Department of Information Technology

Submitted by

Akanksha Gautam (1773413003)

Guided by

Dr. Vibhash yadav

Mr. Umar Badr Shafeeque

RAJKIYA ENGINEERING COLLEGE, ATARRA, BANDA

(Affiliated to DR. A. P. J. Abdul kalam Technical University, Lucknow, UP)

ACKNOWLEDGEMENTS

I would also like to take this chance to thank module lecturers and faculty mentors for their enthusiastic support and attention during my BTECH study, especially DR. Vibhash Yadav and Mr. Umar Badr Shafeeque. My mentors gave me some advice and helped me to analyse the pros and cons very patiently when I faced problems with the time conflict between job and dissertation.

This work would not have been possible without the encouragement and assistance of my dear group members. They always provided me with unconditional help and valuable recommendations. I state my special gratitude to my mother in law because she came to Ireland to take care of my daughter to let me concentrate on the research. Last but not least, I would also like to thank all my friends and classmates for their endless patience, generous support and total criticism which assisted the whole group in improving my research.

SUMMARY:

In this paper, multiple learning techniques based on Optical character recognition (OCR) for the handwritten digit recognition are examined, and a new accuracy level for recognition of the MNIST dataset is reported. The proposed framework involves three primary parts, image pre-processing, feature extraction and classification. This study strives to improve the recognition accuracy by more than 99% in handwritten digit recognition. As will be seen, pre-processing and feature extraction play crucial roles in this experiment to reach the highest accuracy.

Firstly, it was found that forms of image pre-processing such as normalization, slant correction or elastic distortion have a significant effect on the feature selection of the sample. In particular, slant correction is the central focus of this work because it can solve the problem whereby different people's handwriting is more or less tilted. In the feature extraction stage, Principal Component Analysis (PCA) and Histogram of Oriented Gradients (HOG) feature descriptor are presented to reduce the dimension of data and extract the relevant information. The classification task is performed by a number of classifiers namely, Support Vector Machine (SVM), Convolutional Neural Network (CNN), K-Nearest Neighbours (K-NN) and Random Forest (RF) to determine which classifier has the highest accuracy rate in this experiment. The experimental results indicated that the entire performance of CNN and K-NN models is superior to SVM and RF in the field of handwritten number recognition. Two combinations can improve the recognition accuracy to over 99% in this study, respectively Pre-processing + CNN and Pre-processing + PCA + K-NN. Moreover, four experimental results are analysed and evaluated by a series of tools such as the confusion matrix, 10-fold cross-validation, error rates, and classification reports. An interesting finding is that the level of accuracy achieved by using the HOG feature descriptor based on K-NN and RF was lower than the raw data. Notably, the combination of pre-processing and CNN reached the highest recognition rate of 99.44% in the experiment.

TABLE OF CONTENTS:

1.INTRODUCTION

1.1 Background

The rapid growth of new documents and multimedia news has created new challenges in pattern recognition and machine learning (Cecotti, 2016). Handwriting character recognition has become a standard research area due to advances in technologies such as the handwriting capture devices and powerful mobile computers (Elleuch, Maalej & Kherallah 2016). However, since handwriting very much depends on the writer, building a high-reliability recognition system that recognizes any handwritten character input to an application, is challenging.

This work considers the problem of recognizing handwritten digits, i.e. numbers from 0 to 9. Typically, handwritten digit recognition is an essential function in a variety of practical applications, for example in administration and finance (Niu & Suen, 2012). These industries require an excellent recognition rate with the highest reliability. Unconstrained handwritten number recognition has been applied with excellent results, to the amounts in written form on checks, to forms filled by hand such as tax forms or postal zip codes for a postcard (Lauer, Suen & Bloch, 2007). Constraint recognition refers to the extent to which individuals believe that factors beyond their control limit their behaviour. By contrast, the unconstrained recognition system can be broken down into several parts: pre-processing, feature extraction, classification, evaluation and verification.

Optical character recognition (OCR) is one of a multitude of research fields in artificial intelligence and character recognition (Pramanik & Bag, 2018). OCR has developed many applications. For example, verification code images, automatic license plate recognition and text information extraction (Sarkhel et al., 2016). Besides, investigators working on the OCR systems have considered extensive features for handwritten digit recognition. While the majority of features are generic, several of them apply the particular attributes to enhance the function of the classifiers such as graph-theoretic methods, shadow-based characteristics, gradient-based characteristics, etc. (Biswas et al., 2017).

Although many researchers have discussed images of isolated handwritten digits, only a few people mentioned pre-processing the image. For example, Niu and Sune (2012) proposed a hybrid model of combining the two superior classifiers: Convolutional Neural Network (CNN)

and the Support Vector Machine (SVM), which have been conducted on the non-pre-processing MNIST database and achieved the recognition rate of 94.4% with 5.6% rejection. Image pre-processing which includes filtering, segmentation, normalization, thinning, slant correction, etc. may deliver the dramatic positive effects on the characters and the results of image analysis. Most image pre-processing can reduce noise and reconstruct images so that operations can be easily performed on the image and further improve OCR accuracy. Moreover, different people's writings are more or less sloped. To correct this, elastic distortion is employed in the process of rotating an image that provides a method to increase the similarity between two samples representing the same digit.

In the domain of OCR handwriting digit recognition has been intensively researched for ten years in many systems and classification algorithms. These include, for example, the SVM, CNN and Random Forest (RF) algorithms. However, the recognition accuracy of the experiments is mostly around 95%. Since lots of classifiers cannot adequately handle the original images or data, feature extraction is one of the pretreatment steps that has the purpose of decreasing the dimension of data and abstracting the valid information (Lauer, Suen & Bloch, 2007).

Traditional manual design feature selection is a cumbersome and time-consuming mission that cannot process the original image, while an automatic extraction method by CNN can detect features directly from the original image (Bernard, Adam & Heutte, 2007). Lauer, Suen, and Bloch (2007) replaced the last layer of the LeNet4 network with the K-Nearest Neighbours (KNN) classifier to process the abstracted features. A CNN is a feed-forward network that extracts topological attributes from images. It collects features from the original image in the first layer and uses its last layer to classify the pattern. At the classification stage, the SVM constructs the best separation hyperplane in the high dimensional characteristic space. Also, the k-NN algorithm is one of the most straightforward machine learning algorithms, and the input consists of the k nearest training samples in the feature space. RF build various decision trees and associate them to receive more accurate and stable predictions.

The Histogram of Oriented Gradient (HOG) is accessible for object detection that feature extraction needs to invert black/white pixels. In the HOG feature descriptor, the distribution of directions of gradients is used as a feature. The gradient image removes a lot of nonessential information but highlights the outline. Moreover, Principal component analysis (PCA)

is an eigenvector-based multivariate analysis technique that usually extracts the best data variance. The other main benefit of PCA is that once the patterns are detected in the data, then the data is decreased without much loss of information.

1.2 Project

This research aims to recognize the handwritten digits by using tools from Machine Learning to train the classifiers, so it produces a high recognition performance. Furthermore, the use of tools from Computer Vision is explored to investigate the effect of the selection of classifiers, features, and image pre-processing on the entire error rate. The dataset used for the application is a MNIST dataset containing 60,000 training and 10,000 testing images originally, which are 28 x 28 grayscale (0 255), labelled and bitmap format. It is an excellent database for machine learning and pattern recognition methods while needing minimal efforts in pre-processing and formatting.

There are many features in this data, so it has many dimensions. PCA is a dimension-reduction tool that is applied to reduce the elements into a small but informative kind of set of characteristics before using the data in the machine learning models.

The OCR technique transforms the input graphics into a flexible format in the computer (Phangtriastu, Harefa & Tanoto, 2017). In OCR applications, the performance accuracy and speed of digital recognition is critical to the overall performance. In a handwriting recognition system, feature extraction is one of the vital factors for success. A good group of features ought to represent traits that are specific to one class (Lauer, Suen & Bloch, 2007). The commonly applied functions in character recognition are crossing points, structures, directions, intersections and contours (Niu & Suen, 2012). However, many classifiers such as SVM and RF cannot process raw images or data efficiently, because extracting appropriate structural features from complex shapes is a considerable challenge (Pramanik & Bag, 2018). While, the automatic extraction method by CNN can extract elements directly from the raw image (Bernard, Adam & Heutte, 2007), as well as the HOG feature vector is also very useful for tasks like image recognition and object detection, as when it fesds into the classification algorithms like SVM or RF it produces good results. Besides, PCA can project digital images onto a low-dimensional interspace composed of few primary images for further feature extraction.

Some problems occur during the development of the OCR system. Firstly, raw image data may have a variety of issues such as blurring or skewing, hence it may not generate the most excellent computer vision results. That is why image pre-processing is considered in depth. Another problem is how to extract features with background noise. One clear example is the contrast between fonts and paper (Phangtriastu, Harefa & Tanoto, 2017). Furthermore, the performance of the classifier can depend on the feature quality of the classifier itself (Elleuch, Maalej & Kherallah 2016). Additionally, a common problem in the digital classification is considered. The similarity between numbers such as 1 and 7, 5 and 6, 3 and 8, 9, and 8 etc., makes recognition a difficult task. Because people write the same number in various ways, the uniqueness and variety in handwriting affect the structure and appearance of the digits. Therefore, how to use the combination of image pre-processing and classifier is the main problem of OCR in handwritten digit recognition.

1.3 Objectives

The objectives of this research are:

To find out what opinions of the various image pre-processing techniques can be applied to this study

To identify whether can these image pre-processing methods have a significant impact on the error rate of selected classification models;

To examine the potential of the HOG feature descriptor to predict the model in image recognition and object detection;

To recommend which algorithms can improve the accuracy of handwritten digital recognition to up to 99% based on the evaluated findings;

Multiple handwritten digital images are applied to the model with the highest performance for verification.

1.4 Methodologies

Hypothesis: The null hypothesis (H0) of this research is that the accuracy of handwritten digit recognition using the combination of image pre-processing and classifiers based on the OCR will be less than 95%, while the alternative hypothesis (H1) is that accuracy will be no less than 99%.

The objective of the research is to show the particularly template-based model, such as CNN and RF. However, the error rate increases when the number cantered on the bounding box rather than the centre of mass. So, feature extraction is one of the pre-treatment steps, aimed at reducing the dimension of data and extracting the relevant information. On the other hand, image pre-processing such as sharpening, slant correction or elastic distortion is necessary because the oblique numbers and blurred images will affect the accuracy of feature extraction. Traditional manual design feature selection is a cumbersome and time-consuming mission that cannot process the original image, while an automatic extraction method by the LeNet5 CNN architecture can retrieve features directly from the original image and HOG is another feature descriptor.

The research methods used in this paper are quantitative. Specifically, quantitative methods emphasize objective measurements and manipulate pre-existing statistical data using software tools. The MNIST database files are online freely available train.csv of 60,000 examples and test.csv of 10,000 samples that contain images of handwritten English numerals. Also, the RF, CNN, K-NN and SVM in Python were used to study and build prediction models. The four pre-processed models will be evaluated and compared, and then the results will clearly show the difference in performance for the classifiers. Moreover, the accuracy of the classifiers will be assessed to determine whether to reject or accept the null hypothesis.

2. DISCRIPTION OF THE INTERNSHIP

2.1 Introduction

How the image pre-processing, feature selection and the relevant classification techniques contribute to handwritten digit recognition. Also, it provides an in-depth and detailed overview of the recent literature, corresponding to this study. Firstly, the part of this section presents an overview with references to the approaches to OCR and the template matching Machine Learning (ML) techniques. In the second part, an analysis of the factors which affect the recognition error rate is expressed. Furthermore, the applied classification techniques in ML and the evaluation of design will be reviewed in the third part. The final section will provide a summary of the next stage in the study and what will happen next and what will come in the design of the experiment.

2.2 The Importance of Handwritten Digit Recognition

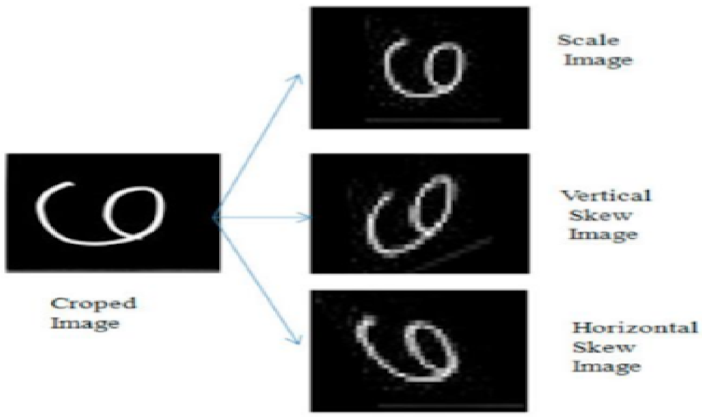
More and more people are focusing on the use of the personal computer rather than acquiring excellent handwriting skills. The one reason is that the internet and applications are becoming more intelligent than before. Additionally, the poor quality or illegible handwriting is the main reason for inaccurate handwritten character recognition.

OCR refers to the recognition of characters on optical scanning and

digital text pages by computer. Although many systems are available for identifying printed text, identifying handwritten characters is still a challenge in the field of pattern recognition. Despite its problems, it widely contributes to the progress of improving the interface between man and machine in a lot of applications. Due to a variety of potential applications such as the reading of postal codes, medical prescription reading, interpreting handwritten addresses, processing bank checks, credit authentication, social welfare, forensic analysis of crime evidence which includes a handwritten note, etc., handwritten digital recognition is still an active area of research (Winkler, 1980). In recent years, the availability of devices has further broadened the range of applications for handwritten digital recognition for multiple personal uses such as note taking and extracting data from filling out forms, etc

2.3: OCR

OCR is a technique that recognizes printed text in scanned documents. But it serves many other purposes as well. For instance, the Google Translate application contains an OCR technique that works with the device's camera. It captures text from magazines, documents, and other handwritten characters and converts it to another language. OCR is a complex process which involves many steps. The steps involved in OCR are pre-processing, feature selection and classification. First capture the image of the digit is categorized in a standard image format such as JPEG, PNG or bitmap. Image formats are broadly categorized as lossy or non-lossy image formats which are used depending on the application. For example, it is usually a requirement that non-lossy image formats are used in medicine. Next, the image is pre-processed to standardize features such as size and resolution. From this, we extract features such as an



edge outline or a chain code depending on the algorithm being used. Finally, these features will be passed on to the classification engine.

Figure 2.1: An example of the transformed images

There is no standard large dataset available for handwritten Marathi numerals, Mane and Kulkarni (2018) have performed various transformations to add the size of the data set**.**

Horizontal and Vertical skewing: each image is tilted vertically and horizontally by a factor of 0.5, which is represented in Figure 2.1. These conversions have increased the dataset fourfold.

In other words, the task of recognizing handwritten digits has been broken down into the following steps which are depicted in Fig.2.2. For handwritten characters, one of the first variances in ways of writing ways is caused by slope, which is defined as the slant of the writing trend relative to the vertical line. Besides that, slant correction must precede other pre-processing tasks, which is that correction operations usually create a rough outline of the character and smoothing tends to change the image topology. Bilinear interpolation is useful for generating other distorted character images at the selected resolution (29x29). Furthermore, one method of achieving non-uniform thickness invariance includes determining a constant thickness "middle line" for each letter for identification purposes. Therefore, the process is called Skeletonization.

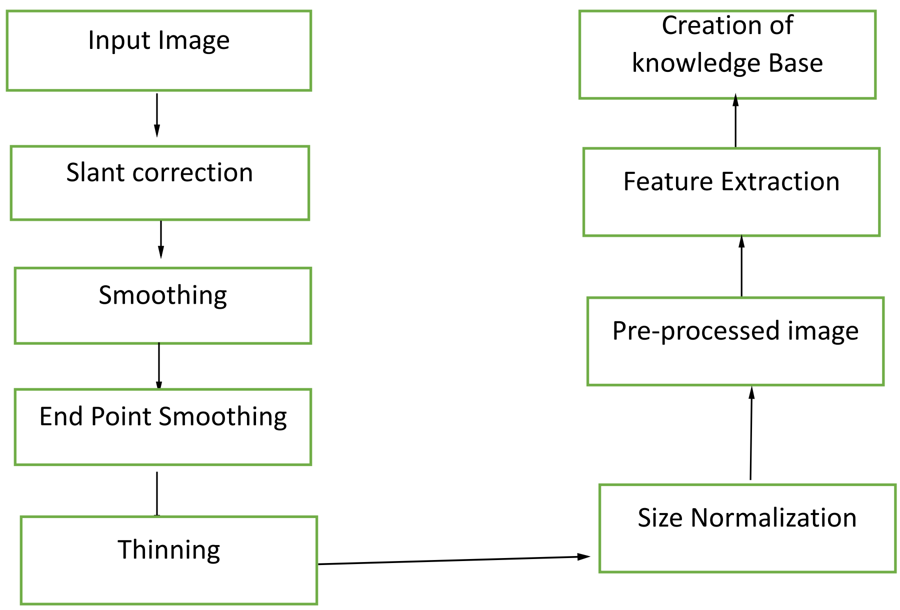


Figure 2.2: Block diagram describing system implementation

2.4 DESIGN AND METHODOLOGY

Discuss the structure of the research including the data collection methods, sampling size, the principles and design of template-based models such as CNN, SVM, K-NN, and RF, the comparison of the technical details and the methodology adopted for designing and evaluating the solution. In particular, to achieve higher accuracy, the sharpening, grayscale normalization, slant correction and elastic distortion techniques will be applied to the image pre-processing stage. The PCA will be used to extract the best data variance, and the HOG feature vector will be applied to image recognition and object detection. Finally, the four previously mentioned classification techniques will be evaluated using K-fold cross-validation, error rates, accuracy, classification reports and confusion matrix.

Since previous researchers have well verified the four classifiers mentioned above, CNN, SVM, K-NN and RF will be applied and compared in this experiment to determine which classifier delivers the highest performance. However, this research will pay more attention to pre-processing and feature extraction steps than in previous studies to reach the highest accuracy.

Firstly, raw image data may have a variety of issues such as blurring or skewing and thus are less likely to produce optimal computer vision results. That is why careful consideration of image pre-processing is fundamental. In particular, grayscale normalization will decrease the effect of an illumination's differences. Slant correction will be used to solve the problem that different people's handwritings are more or less skewed writing.

In handwriting recognition systems, feature extraction is one of the critical factors for success. However, extracting appropriate structural features from complex shapes is a considerable challenge. CNN will use the LeNet5 automatic extraction method to extract elements directly from the original image. Moreover, the HOG feature vector will be adopted in other classifiers, since it is useful for tasks such as image recognition and object detection. Besides, PCA can project digital images onto low-dimensional space composed of a small number of elemental images for further feature extraction. Therefore, HOG and PCA are the nuclear technologies in the feature extraction phase.

Finally, the handwritten numbers never seen by the systems will be applied to finalise the model. The resulting model with an accuracy of more than 99% will identify the handwritten digits and display the predicted numbers. Figure 3.1 presents some handwritten digits never seen by the systems.

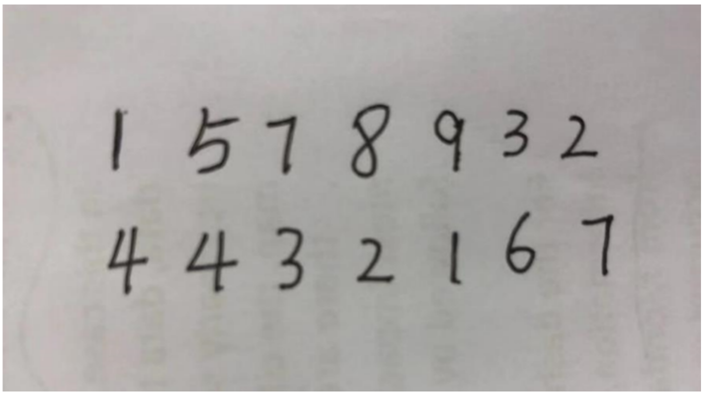


Figure3.1: the several of handwritten digits never seen by the system

2.4.1 DATA COLLECTION:

A commonly applied dataset for handwritten digit recognition called MNIST can be searched on the Y. Lecun website. It is a gathering of 70,000 digits written by the different 750 Census Bureau employees and high school students. This dataset is a widely known benchmark that includes a training set with 60,000 images and testing set with 10,000 images. Numerals were size-normalised, cantered and stored sequentially as 28 × 28-pixel images in grey-level bitmaps. The resulting datasets are provided with the labels, and each image includes a single digit. This ready-to-use database is the data that was applied in the experiments below. Figure 3.2 displays some sample numbers in the training set.

Figure 3.2: MNIST database: the sample number in the training set

The MNIST dataset consists of NIST's unique database 3 and unique database 1, which involves binary images of handwritten numbers. In particular, the MNIST training set that contained samples from approximately 250 writers is constructed from 30,000 patterns in SD3 and 30,000 patterns in SD-1. Similarly, the test set consists of 5,000 patterns in SD-3 and 5,000 patterns in SD-1. NIST formerly appointed SD-3 as the training set and SD-1 as the test set. However, compared with SD-1, SD-3 is more transparent and more accessible to identification. The reason for this is that SD-3 was collected from Census Bureau staff, while SD-1 was collected from high school students. Since it is essential to ensure that the writers of the training set and the test set are disjoint, a new database MNIST was built by mixing the NIST data sets.

Some researchers have used the database for analysis and have achieved beautiful results. For instance, Bernard and his team (2007) reached an average of 94.93% for handwritten digit recognition accuracy, and Cecotti (2016) gained the highest accuracy of 98.54%. In some experiments, to add the training set, the artificial distortions are applied to each sample to extend new samples.

2.4.2 CNN- A Convolutional Neural Network (CNN) is a multi-layer neural feed-forward network with deep supervised learning architecture, which can be regarded as a two-part combination: automatic feature extractor and trainable classifier. The classifier and weights of the backpropagation algorithm in the feature extractor are applied. Besides, CNN can also extract topology attributes from images. It abstracts features from the primary image in the first layer and classifies the pattern with the last layer.

2.4.3 K-NN- K-NN is one of the most straightforward and well-known non-parametric algorithms which is suitable for large numbers of data. The K-NN algorithm has been applied for the statistical calculation, scene identification and also writer recognition systems. In some previous research studies, K-NN has been used for handwritten character recognition, and a high recognition accuracy was obtained. In 2014, Babu et al. proposed four feature extraction techniques that are composed of water Reservoir principle-based features, the number of loops in the image, maximum profile distances and fill hole density feature, and then experimented on the MNIST dataset. The recognition accuracy with this method is 96.94%.

2.4.4 SVM- The Support Vector Machine (SVM) algorithm invented by Vapnik and Cortes (1998) is a powerful discriminant classifier that has been effectively applied to many pattern recognition or classification problems and has obtained positive results. Besides, due to its simplicity, flexibility, prediction capability and global optimality, it is considered to be the most advanced tool for solving linear and nonlinear classification problems. They are based on structural risk minimization instead of the empirical risk minimization that is traditionally used for artificial neural networks.

2.4.5 RF- Random Forest (RF) is a collective term for a combination of classifiers using the L-tree classifier $\{h (x, \Theta k), k = 1,... L\}$, where Θk is an independent random vector of the same distribution and x is Input. It can be said that random forest is a series of methods which include several algorithms based on this definition. The concept of the random forest was introduced based on the bagging principle in 2001 by Bierman. A decision forest is a collection of some decision trees that act in a parallel pattern. It is distinct from the way that each tree attempts to classify the data completely independent of other trees

2.5: The key techniques in this Experiment

Normalization and Reshape Data- The MNIST handwritten digits have been size-normalized, centered and stored sequentially as 28 × 28 pixel images in grey-level bitmaps. The pixel values are grayscale between 0 and 255. The background is mostly close to 0 and those close to 255 represents the digit. Besides, the pixel values can be quickly normalized to the range of 0 and 1 by dividing each value by a maximum of 255. Some grayscale samples from the MNIST dataset are displayed in Fig. 3.3.

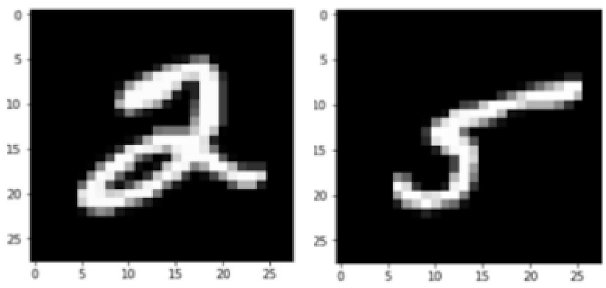


Figure. 3.3: Some grayscale samples from the MNIST dataset

HOG: The HOG descriptor was first proposed by Dalal and Triggs (2005) for human body detection in an image. Recently, it is one of the most commonly and successfully used descriptors for computer vision and image recognition for object detection. The principle of the HOG descriptor is that the appearance and shape of the local object within an image can be explained by the arrangement of intensity gradients or edge directions. This technique divides the image into small connected regions and then calculates a histogram of the gradient direction or edge direction according to the mean differences. Moreover, HOG vectors are computed by taking direction histograms of edge intensity in a local area.

PCA: PCA is a powerful and extensive technology applied for data exploration and compression in neural networks and machine learning. It involves linearly converting a set of related variables into alternative representations that emphasize the variance between observations. Effectively, it reduces the dimensions of the observed data by eliminating redundancy. PCA can provide a lower-dimensional representation if a multivariate data are visualized as a series of coordinates in the high-dimensional data space.

Flow Chart

This experiment will be completed according to the following flow chart is depicted in Fig.3.4.

1. Pre-Processing: Slant correction, Sharpening, and elastic distortion
2. Feature Extraction using PCA or HOG
3. Classification using the CNN, K-NN, RF and SVM

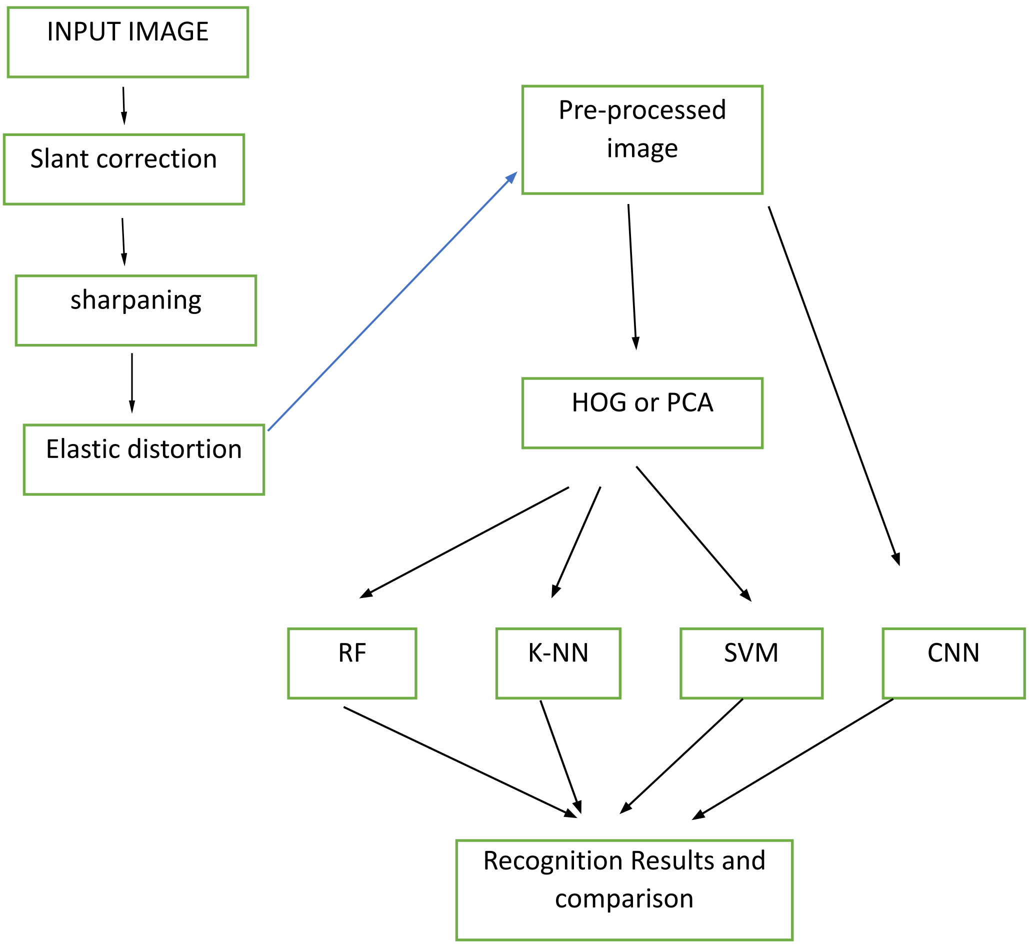


Figure3.4: The Flow Chart for this experiment

The process of this experiment has been displayed in detail in chapter four. The dataset adopted was revisited at the beginning to ensure the accuracy of the subsequent analyses. Then, the foremost techniques such as slant correction and elastic distortion in the pre-processing stage were illustrated in depth. Since an automatic extraction method LeNet5 by CNN can detect features directly from the original image, PCA and HOG technologies were not explored based on the CNN model. According to the previous research on different types of pre-processing, feature extraction and classifier technology,

five combinations were the focal points for further exploration:

1. Pre-processing +CNN
2. Pre-processing + PCA + K-NN
3. Pre-processing + PCA+ SVM

4. Pre-processing + HOG+ K-NN

5. Pre-processing + HOG+ RF

Moreover, the implementation details of the five combinations such as the selection of the package, and the adjustment of the parameters have been analysed.

3.REFLECTION OF EXPERIMENT

Initial Experiment

Firstly, the performance of the four classification models such as CNN, K-NN, RF and SVM was evaluated on the original MNIST data. Table 5.1 indicates a comparison of the four classifiers regarding error rates (ER) and training time (TT). It can be seen from this table that the ER of CNN in this experiment is the lowest at 1.25% compared with the other three classifiers, and the ER of SVM is up to 9%. On the other hand, the training time spent by CNN is 6,000 times higher than RF by 3.5 hours. The reason for this may be that CNN is well-suited for extensively used digital databases and images since they can recognize patterns with numerous features, namely pixels in 2D and characters in 1D. In contrast, RF, KNN and SVM demonstrate superiority in other kinds of challenges: mainly in the space of relatively few various features such as tens or hundreds. They will defeat CNN's easily there.

Future Experiment

In this report although the method of addressing the research question was found by training on the MNIST database, there are still some problems that need to be explored and solved in the future. For example, the accuracy of the KNN, SVM and RF models based on the combination of pre-processing and HOG are smaller than the initial experiment. Nevertheless, Ebrahim Zadeh et al. (2014) employed the linear SVM as the classifier, and the HOG feature descriptor on the MNIST database and a 97.25% accuracy rate was obtained. So, the causes of these problems mentioned above should be analysed and found to be resolved in the future. There are also some natural expansions to this research that would assist extend and reinforcing the results. The benchmark database of MNIST was developed for this work, and it is an excellent database for machine learning and pattern recognition methods while making minimal efforts in pre-processing and formatting.

4.CONCLUSION

The result of our work shows that the maximum accuracy 99.2% was obtained in MNIST dataset using deep learning technique and an appropriate learning rate at 15000 iterations. It is observed that accuracy slowly starts decreasing or remains constant after 15000 iterations. The performance ratio of GPU: CPU is found to be 30:1.It is concluded that computation time in GPU exponentially decreases as compared to CPU. Future works is focused o

This study attempted to recognize the handwritten digits by using tools from Machine Learning to train the classifier. Also, the use of techniques in Computer Vision was explored to investigate the effect of selection image pre-processing, feature extraction and classifiers on the overall accuracy. The dataset used for the experiment is MNIST dataset originally constituted of 60,000 training, and 10,000 testing images which are 28 x 28 grayscale (0 255) labeled and bitmap format. It is a brilliant database for machine learning and characters recognition methods while taking minimal efforts in pre-processing and formatting.

Compared with other research, this study focused on exploring which image pre-processing and feature extraction techniques based on OCR can work for improving the accuracy of classification models by more than 99%. In the initial experiment, the CNN algorithm won with a recognition accuracy of 98.75%, followed by K-NN with 96.68%. The performance of RF and SVM in this experiment is not outstanding because they are not good at pattern recognition, while they demonstrate superiority in other kinds of challenges: mainly in the space of relatively few different features such as tens or hundreds. After that, image pre-processing techniques (slant correction, sharpening and elastic deformation) and feature selection techniques (PCA and HOG) were applied to the experiment. Finally, CNN based on image pre-processing, and K-NN based on the combination of image pre-processing and PCA achieved the goal of successfully improving the accuracy to over 99%. In particular, slant correction played a significant role at the image pre-processing stage and the HOG feature descriptors did not perform well in image recognition and object detection.

Four experimental results were analysed and evaluated by a series of tools such as confusion matrices, k-fold cross-validation, error rates, and classification reports. Each modification produced changes in the results mostly improved accuracy and widely varying performance times. The original objective of this study was could the handwritten digit recognition accuracy is improved by image pre-processing and feature extraction, and it was confirmed to say yes. At the end of the experiment, to verify the performance of the model again, some handwritten digits never seen by the systems were forecasted using the classifier and achieved satisfactory results.

REFRENCES

[1] http://yann.lecun.com/exdb/mnist/

[2] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278-2324, November 1998

[3] https://www.kaggle.com/c/digit-recognizer

[4] Dan Cireşan, Ueli Meier, Juergen Schmidhuber: Multi-column Deep Neural Networks for Image Classification,CVPR,2012

[5] Liu,Qing,Tang Xianlun, Zhang,Na: Structure Optimized Convolutional Neural Network Based on Unsupervised Pre-training, Rongcheng Kxoe Yu Jishu/Advanced Engineering Science , 2017, Vol.49, p.210-215

[6] Li Wan, Matthew Zeiler, Sixin Zhang, Yann LeCun, Rob Fergus: Regularization of Neural Networks using DropConnect, ICML,2013

[7] Jia-Ren Chang, Yong-Sheng Chen: Batch-normalized Maxout Network in Network,arXiv 2015

[8] https://www.researchgate.net/publication/334957576_A_Survey_of_Handwritten_Character_Recognition_with_MNIST_and_EMNIST

[9] http://blog.csdn.net/weixin_38776853/article/details/72721077

[10] http://www.jeyzhang.com/cnn-learning-notes-1.html

[11] https://www.tensorflow.org/get_started/mnist/beginners

[12] http://neuralnetworksanddeeplearning.com/chap3.html#softmax

[13] https://github.com/kkweon/mnist-competition

[14] https://colah.github.io/posts/2014-10-Visualizing-MNIST/

APPENDIX-

The Combination of Pre-processing and CNN

CNN is a multi-layered neural feed-forward network with deep supervised learning architecture, which can be regarded as a two-part combination: automatic feature extractor and trainable classifier. In this section, the mixture of pre-processing and CNN has been applied to obtain a satisfactory result.

The Combination of Pre-processing, PCA and K-NN

PCA is a powerful and extensive technology applied for data exploration and compression in neural networks and machine learning. K-NN is one of the simplest and best well-known nonparametric algorithms which is suitable for large numbers of data. In this part, the combination of preprocessing, PCA and K-NN has been used for an experiment in handwritten digit recognition experiment

The Combination of Pre-processing, PCA and SVM

The Support Vector Machine (SVM) algorithm is a powerful discriminant classifier that has been effectively applied to many pattern recognition or classification problems and has obtained favorable results. Shamim et al. (2018) adopted the SVM with RBF and the ten cross-validation method to select the parameter to gain the highest recognition rates, reaching more than 93%. This section introduces a multivariate analysis framework for feature detection in a recognition system, while the PCA and SVM based supervision scheme can determine patterns in the recognition system. The data set for this experiment was pre-processed, which is a combination of pre-processing, PCA and SVM

The Combination of Pre-processing, HOG and K-NN

In recent times, the HOG descriptor has become one of the most common and successfully used descriptors for computer vision and image recognition for object detection. In the HOG feature descriptor, the distribution of directions of gradients is used as features. Moreover, HOG vectors are computed by taking direction histograms of edge intensity in a local area. A 2017 paper by Phangtriastu, Harefa, and Tanoto compared the most commonly classifiers SVM and ANN, while this experiment achieved the highest accuracy 94.43% using the SVM classifier with the combination of feature extraction algorithms which are a projection histogram and HOG.

The Combination of Pre-processing, HOG and RF

In the field of pattern recognition, researchers have paid more attention to multi-classifier systems in recent years, especially Bagging, Boosting. The concept of RF was introduced based on the bagging principle by Breiman in 2001. In the Forest-RI algorithm, not all the features contribute to the recognition rate. In fact, some features may degrade the results. Bernard, Adam and Heutte (2007) researched a conventional feature extraction technique based on a greyscale multi-resolution pyramid to find out the effect of the parameter values on the performance of the RF. They have experimented on the MNIST handwritten digital database and reached an accuracy level greater than 93%.

	Raw Data	Pre-processing	Pre+ PCA	Pre+ HOG
CNN	98.75%	99.44%	-----------	-----------
K-NN	96.68%	98.14%	99.17%	95.39%
RF	93.56%	95.60%	90.89%	92.6%
SVM	91.17%	94.54%	94.90%	93.02%

The summary of handwritten digit RR based on four classifier models

She combined performances of the CNN and K-NN models are higher than SVM and RF in the field of handwritten digit recognition. Furthermore, the performances of two combinations have successfully answered the challenge of this study and improved the accuracy to over 99%, respectively Preprocessing + CNN and Preprocessing + PCA + K-NN. Notably, the combination of pre-processing and CNN reached the highest efficiency of 99.44% throughout the experiment. However, an automatic extraction method LeNet5 by CNN can detect features directly from the original image, PCA and HOG technologies were not explored based on the CNN model. In contrast, most application SVM and RF models had recognition rates below 95% in general. An interesting finding is that the accuracy achieved from using the HOG feature descriptor based on K-NN and RF was lower than the raw data. One of the reasons may be that the two classification algorithms are not sensitive to the alignment of the intensity gradient of the image, and that will be the future research direction.