

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

A CNN based Handwritten Numeral Recognition Model for Four Arithmetic Operations

Chen ShanWei^{a,b}, Shir LiWang^{a*}, Ng Theam Foo^c, Dzati Athiar Ramli^d

^a*Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia.*

^b*Department of Education, Baoji University of Arts and Sciences, 721012 Baoji, China*

^c*Centre for Global Sustainability Studies, Universiti Sains Malaysia, 11800 Penang, Malaysia.*

^d*School of Electrical & Electronic Engineering, Universiti Sains Malaysia, Engineering Campus, Nibong Tebal 14300, Penang, Malaysia.*

Abstract

The pandemic of Covid-19 has caused a shift of paradigm of education, from face-to-face to e-learning. E-learning leads to an escalation in digitalization of handwritten documents because it requires submission of homework and assignments through online. To help teachers in checking digitalized handwritten homework, this paper proposes an automatic checking system based on a convolutional neural network (CNN) for handwritten numeral recognition. The CNN is used to recognize four arithmetic operations in mathematical questions consisting of addition, deduction, multiplication and division. The performance CNN in handwritten numeral recognition have been optimized in terms of activation function and gradient descent algorithm. The proposed CNN is also trained and tested with the MNIST handwritten data set. The experimental results show that the recognition accuracy the improved CNN improves to a certain extent as compared to before optimization.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of KES International.

Keywords: deep learning; CNN; handwritten numeral recognition; image processing

* Corresponding author. Tel.: +60-195588220; fax: +0-000-000-0000 .

E-mail address: shirli_wang@fskik.ups.edu.my

1. Introduction

The rapid development of artificial intelligence (AI) has led to technological changes and usages in various domains such as business, manufacturing, healthcare, education and social activities. The occurrence of pandemic coronavirus such as SARS, MERS and the recent Covid-19 has accelerated the development and implementation of digital and AI technology in the domains [1]. The Covid-19 pandemic also has forced more than 1.38 billion students to stay at home by March 2020 [2]. Thus, the pandemic causes a paradigm shift in education, from the traditional face-to-face learning to e-learning. The shift of paradigm in education has led to escalation in digitalization of handwritten documents because they are convenient and efficient. An example of digitalization of handwritten documents is the submission of homework through online by students. An automatic checking system for the digitalized handwritten homework will be helpful in reducing teacher's time in checking homework. As a result, teachers can spend more time and efforts in teaching and learning activities which benefit students.

In this paper, we propose an automatic checking system based on a convolutional neural network (CNN) in handwritten numeral recognition. The proposed system is used to recognize four arithmetic operations, which are addition, deduction, multiplication and division. The remainder of this paper is organized as follows: Section 2 describes the studies related to CNN in handwritten character and digit recognition. Section 3 presents the methodology of our proposed CNN model, and Section 4 describes the experimental setup. Section 5 discusses the experimental results through analysis. Section 6 presents the conclusions and future work.

2. Background

Handwritten numeral recognition has important applications in many fields such as banks, post, and education. At present, researchers have proposed many handwritten numeral recognition methods, such as multi-scale feature and neural network fusion method [3], a method based on prototype generation technology [4], a method based on affinity propagation clustering (AP) and back-propagation (BP) neural network [5], method based on probability measure support vector machine (SVM) [6], etc. However, the above methods have insufficient ability to express features and are easily affected by the external environment, which cannot meet the requirements of higher recognition rate.

Recently, CNNs have achieved good performance in handwritten numeral recognition. It has the ability of automatic feature extraction for image recognition and avoids the complex process of feature extraction and data reconstruction in traditional recognition methods [7].

In [8], a handwritten character classifier based on CNN and SVM was proposed. The model had produced a good classification result. A method of handwritten character recognition based on the deep neural network model of Siamese network (SN) was proposed in [9]. The recognition rate reached 98%, but the SN model did not learn well the different features of samples. Another CNN model known as binary convolutional neural network (B-CNN) was proposed in [10] for handwritten numeral recognition. Having the similar problem to SN, B-CNN achieved good recognition results but could not learn well the advanced features of samples.

The work in [11] pointed out that disrupting the sample data in the training stage could speed up the learning ability of the handwritten character recognition network model. The method helps to improve the ability of model to learn advanced features of samples. When using CNN for image recognition [12] proposed that the convolution kernel should be set in the form of a weighted PCA matrix. After the mapping between hidden layer neurons was completed, the final feature vector was generated by codebook by making full use of the mapping results of each layer.

Traditional CNNs mostly adopt Softmax classifier for classification and recognition after feature extraction. However, with the continuous development of shallow classifiers such as SVM, sparse matrix, and manifold learning, their classification performances have also been greatly improved. Therefore, some researchers combined CNN models with the classifiers to improve the classification performance. Therefore, the work in [13] proposes a method combining CNN and SVM for handwritten digit recognition. Although the recognition rate of this method had been further improved, it required higher performance of computer hardware. Another example of hybrid of CNN and another classified can be found in [14], which proposed a CNN interlayer feature fusion method combined with a manifold classifier to solve the problem of character recognition.

3. Methodology

In this section, our proposed CNN is introduced to recognize four arithmetic operations. Figure 1 is the general flow chart of implementing CNN in handwritten numeral recognition. To complete the automatic check of the mathematical assignment, it is necessary to take photos first, and then correct the skew images from the photos. Then, CNN is used to recognize the characters in the images, and finally, the recognized results are compared with the right answers. Sections 3.1 to 3.6 describe the processes shown in Figure 1 in details, from skew image correction to algorithm improvement.

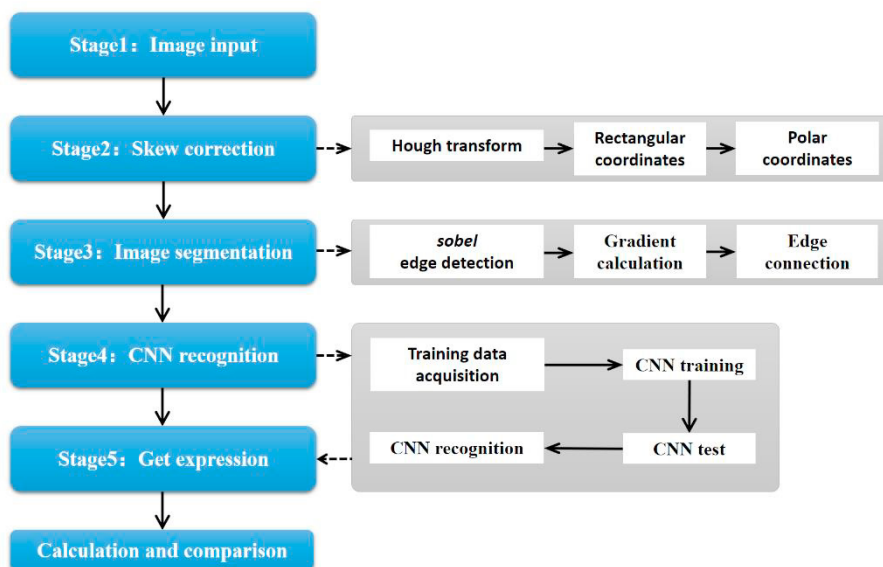


Fig. 1. The program general flow chart

3.1. Skew Image Correction

Handwritten numeral recognition starts with photos acquisition and the photos usually require skew image correction. The captured images are often tilted to some extent, which will not affect the reading and understanding of text information for human eyes. However, the tilted images will lead to recognition errors for computers, and thus, affecting the final character recognition accuracy [15]. There are many datum lines in the image, such as division line, table line, and horizontal grid line. For our case, we need to correct the image according to the direction of the reference line. For the pure character image involving only text or formula, we need to choose a reasonable text image skew correction algorithm. In the field of image processing and computer vision, Hough transform is generally used to recognize the geometric shape in the image. Therefore, the improved Hough transform and perspective transform [16] are adopted in our study. The method not only solves the problem of slant image, but also detects the line or circle in the image quickly and accurately.

3.2. Image Segmentation

The goal of image segmentation is to classify the pixels of the image according to the objects in the image and then extract the objects of interest. In this study, we first binarize and equalize the images. Then, we remove the noise by using Gaussian filter and median filter. An edge detection algorithm is used to get the text edge features in the image. Due to the sensitivity of the Laplacian edge detection algorithm based on the second-order derivatives to noise, we decided to use the Sobel algorithm [17] based on the first-order derivative to detect the edge of the image. By adjusting the parameters and size of dilation and erosion, we are able to get a complete picture of a formula, an entry

picture of an English word, or a picture of entries in ancient poems. However, the extracted results are affected by the conditions of the pictures being captured. In our study, the assignment pictures may be different owing to their image format, lighting and printing conditions. Therefore, it is necessary to optimize the capture method to obtain the ideal segmentation effect for the assignment pictures.

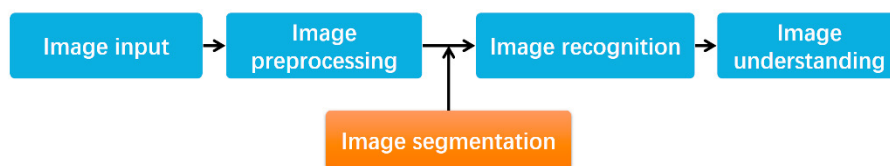


Fig. 2. General image recognition process of computer

3.3. Numeral Recognition

Numeral recognition refers to the process of using electronic equipment to determine the shape of paper handwriting by detecting the dark and bright patterns and then using the character recognition method to translate the shape into computer text [18]. The common used numeral recognition patterns mainly include structure recognition, artificial neural network (ANN) recognition, and the hybrid of the methods. The ANN is widely used in pattern recognition, computer vision, and other fields owing to its self-organizing and adaptive learning ability [19]. Recently, the use of CNN in pattern recognition has drawn attention [20]. The main strength of CNN over the traditional recognition methods is recognition accuracy and computation speed [21]. Therefore, we decided to use a CNN in the handwritten numeral recognition.

Firstly, the image of mathematical formula obtained by image segmentation is transformed into grayscale image and binarized. Then, the image is cut and separated into numbers and symbols. The images consisting of numbers and symbols are used to train the CNN to recognize them. The recognized symbols are input into the syntactic analysis machine according to the character sequence. The structure of the formula is obtained through syntactic analysis, including determining the spatial relationship between characters, structural analysis, and grammar analysis, etc. Then, an analysis tree is constructed to calculate the formula result.

3.4. Training Data Acquisition

For training data acquisition, we decided to use the MNIST data set owing to its good training results [23]. The MNIST handwritten numeral database consist of 60,000 training sample sets and 10,000 test sample sets. We have used translation, scaling, rotation, horizontal and vertical stretching to deform the data to increase the diversity of training data. The purpose of the procedure is to increase the diversity of data set with limited samples [23], and thus improve the recognition of CNN when it used to train CNN.

3.5. Assignments Check

The recognition samples of this program are mainly four arithmetic operations, which can also be extended to ancient poetry and English words. For the four operations of arithmetic, the identified character information needs to be converted into mathematical formulas. The program computes and produces the correct answers for the mathematical formulas. The program compares the correct with the identified answers. The standard answers are stored in the database and are relatively fixed, the recognized numeral information can be directly compared with the correct results in the database. If the identified answer is correct, the program can either produce the comparison results or tick the correct answers. If the identified answer is wrong, the program can either produce the correct answers or cross the incorrect answers. The users can determine the types of program outputs.

3.6. Algorithm Improvement

In this study, LeNet5 was used as the basic structure of handwriting recognition. Figure 3 is the LeNet5 classical

CNN structure proposed by LeCun et al. It consists of the input layer, the convolutional layer (C1, C3, C5), the pooling layer (S2, S4), the full connection layer and the output layer. Excluding the output layer, the structure has a total of 7 layers. The “convolution layer + pooling layer” structure connected alternately is the key component of CNN that automatically extracts image features. The specific parameter configuration of LeNet5 network model is shown in Table 1.

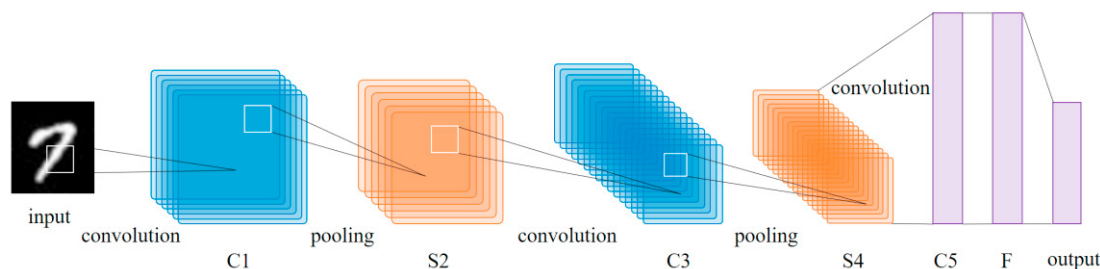


Fig. 3. LeNet5 Logical structure

Table 1. Parameter configuration of the CNN

Layer	Type	Output map size/number	Number of neurons	Connections	Step	Training parameters
C1	Convolution	28*28/6	4704	122304	1	156
S2	Pooling	14*14/6	1176	5880	2	12
C3	Convolution	10*10/16	1600	151600	1	1516
S4	Pooling	5*5/16	400	2000	2	32
C5	Convolution	1*1/120	120	48120	Null	48120
F6	Full connection	84*1	84	10164	Null	10164
7	Output	10*1	Null	Null	Null	10

In this study, we improve the performance of CNN in recognizing the handwritten characters on the basis on two aspects, which are active function and gradient descent algorithm. A comparative analysis before and after the improvement are carried out for the CNNs.

4. Experimental Setup

Since the key part of the program is to use CNN to recognize handwritten characters automatically, we improve the CNN model to improve and optimize the whole program. The optimization mainly starts with the active function of CNN forward propagation and the gradient descent of CNN back propagation.

4.1. Improvement of activation function

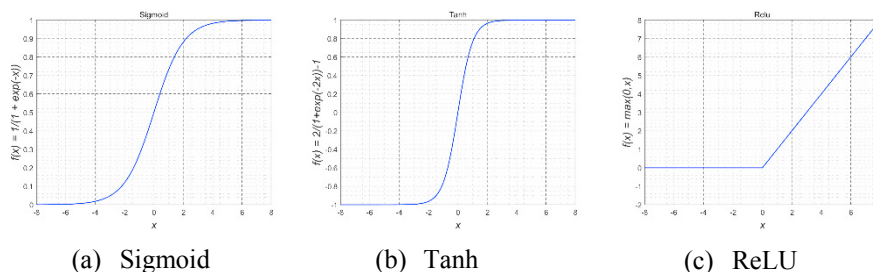


Fig. 4. Different active functions in CNN

Firstly, the active function of the convolution layer used in this program is changed from sigmoid and Tanh to rectified linear units (ReLU). Based on Figures 4(a) and 4(b), the functions sigmoid and Tanh approach the saturated

region at both ends, the transformation is very slow and the derivative approaches 0. In the back propagation, the gradient is easy to disappear, resulting in the loss of information [24]. Since functions sigmoid and Tanh are exponential operations, both functions require large amount of calculation than ReLU when calculating the error gradient of back propagation. Another strength of the ReLU function is it makes the output of some neurons to be 0. Therefore, the use of ReLU improves the network's sparsity, reduce the dependence of parameters, and avoid the occurrence of over-fitting [25]. The function of ReLU is shown in equation (1).

$$\sigma(x) = \max(0, x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (1)$$

4.2. Improvement of stochastic gradient descent algorithm

In the process of data training, the common stochastic gradient descent (SGD) algorithm does not result with good performance and it affects the parameter adjustment. Gradient descent optimizer can be used to update the weight and bias b according to the cost function obtained, but if the scale of the data itself is very different, it will cause a large difference in w . The use of inappropriate algorithm renders different training dynamics on batches.

Adaptive moment estimation (ADAM) optimizer can eliminate this phenomenon. Adam optimizer can be considered as the combination of momentum and root mean square propagation (RMSProp). Since SGD is prone to fall into an oscillation when it encounters a gully, momentum can accelerate the decline of SGD in the right direction and suppress the oscillation [26]. The next momentum is as follow:

$$m_t = \gamma m_{t-1} + \eta g_t \quad (2)$$

where γ is the weighted hyperparameter, η is the learning rate, and g_t is the gradient of the objective function with respect to the parameter.

Root mean square prop (RMSProp) is an adaptive learning rate method proposed by Geoff Hinton, which can avoid the continuous accumulation of second-order momentum and improve the training speed with a larger learning rate [27]. The next training speed of RMS prop is as follow:

$$v_t = \gamma v_{t-1} + (1 - \gamma) \cdot \text{diag}(g_t^2) \quad (3)$$

where γ is the weighted hyperparameter, g_t is the descending gradient in the latest time window, $g_t^2 = g_t \odot g_t$.

5. Result and Analysis

Table 2 shows the comparison of the training process after 3000 times of execution of the two algorithms at the frequency of printing every 150 training sessions.

Table 2. The probability of various accuracy rates after 3000 times.

Algorithm	Learn rate	Normal value/times	Max value/times	Stable value	Average
Adam	0.01	0.98/2	1.00/2	1.00	98.2
Stochastic gradient descent	0.01	0.96/2	0.98/2	0.98	96.2

Based on the experimental results, it is found that the CNNs can achieve accurate segmentation of printed font and handwritten font, and combine the recognized results into a formula after the segmentation. It can recognize the four fundamental operations, decimal operation, etc., as shown in Figure 4. Instead of recognizing the handwritten numbers manually, the proposed CNN model can accurately recognize especially the basic four arithmetic operations. The CNN model has a relatively stable performance in checking the mathematical questions.

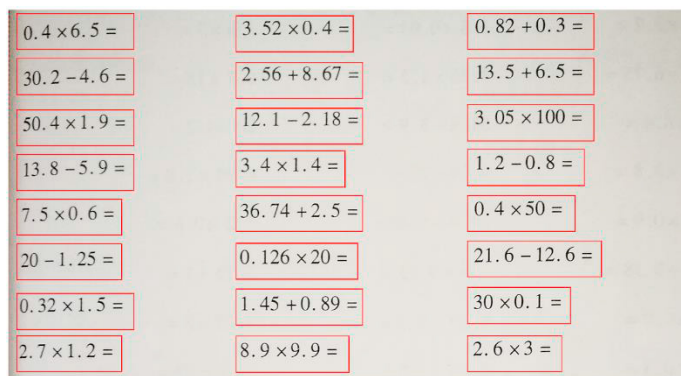


Fig. 5. Mathematical question cards.

The CNN is tested on the MNIST handwritten data set consisting of contains 60000 training sample sets and 10000 test sample sets respectively. The performances of the CNNs on the data set before and after optimization are shown in Table 3. Based on the results shown in Table 3, the CNNs before and after optimization have the same training settings except their activation function and gradient descent algorithm. The results show that the recognition rate of CNN after optimization has an increment of 7.3% to 91.2% as compared to before optimization. Through the improvement of active function and gradient descent algorithm, the convergence speed of the CNN handwritten recognition model reduces from 250 to 200. This means that the recognition effectiveness and convergence speed of the model have improved.

Table 3. Comparison of CNNs in handwritten numeral recognition before and after optimization.

CNN	Training samples	Training batches	Test samples	Test batches	Performance index			
					Optimal learning rate	Weight hyperparameter	Convergence speed (batch)	Recognition Rate (%)
Before optimization	60000	400	10000	67	0.01	0.95	250	83.9
After optimization	60000	400	10000	67	0.01	0.95	200	91.2

Figure 6 and 7 show the comparison of cost function and accuracy of the CNNs during the training process before and after optimization. Based on Figures 6 and 7, the optimized network structure is much better than that before optimization both in terms of convergence speed and recognition accuracy. On the basis of accurate segmentation of printed and handwritten fonts, the recognition rate of the program is improved by optimizing the handwritten recognition network. The improved handwritten recognition network can effectively and efficiently recognize the four operations, fractional operations and decimal operations, which are commonly done in manual.

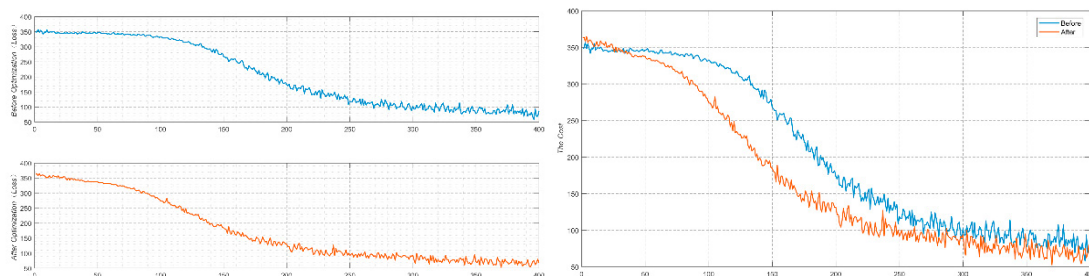


Fig. 6. Comparison of loss rate before and after optimization

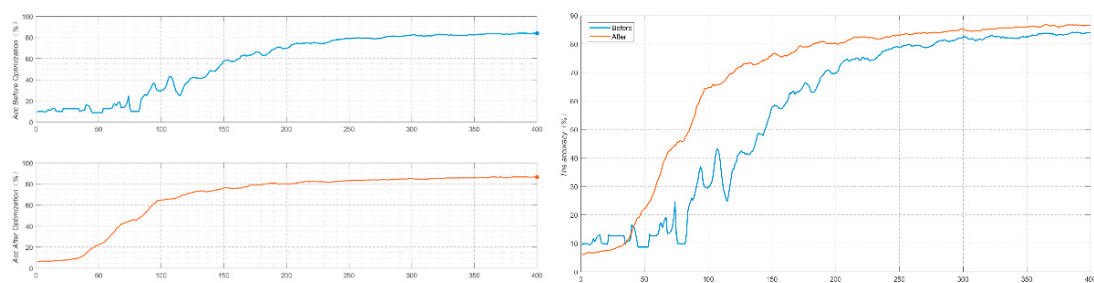


Fig. 7. Comparison of accuracy before and after optimization

6. Conclusion and Future Work

In this study, an improved CNN algorithm is proposed by replacing its activation function and gradient decent algorithm with ReLU and ADAM. The CNN's performance is trained and evaluated on the basis of the MNIST handwritten numeral data set. The improved CNN is evaluated in handwritten numeral recognition, whereby the CNN is used to automatically check four arithmetic operations consisting of addition, deduction, multiplication and division. The CNN based handwritten recognition model has achieved a reduction from 250 to 200 in convergence speed, and an increment from 83.9% to 91.2% in recognition accuracy. For future work, we can extend the potential of CNN in recognizing handwritten English letters and Chinese characters, so that the model can automatically check digitalized and handwritten assignments for other subjects. The CNN based handwritten recognition model can potentially reduce teachers' time in checking assignments so that they can spend more time and efforts to improve teaching and learning activities that benefit students.

Acknowledgment

The authors would like to acknowledge and thank the Universiti Sains Malaysia and the Ministry of Higher Education, Malaysia for supporting this research through the Fundamental Research Grant Scheme (FRGS) with account number 203.PELECT.6071478.

Reference

- [1] Brem A, Viardot E, and Nylund P. A, "Implications of the coronavirus (COVID-19) outbreak for innovation: Which technologies will improve our lives?", Technological forecasting and social change, 2021, 163, 120451.
- [2] Li C., and Lalani F, "The COVID-19 pandemic has changed education forever", This is how. Retrieved September 22, 2020.
- [3] ZHAO Yuan-qing, and WU hua, "Handwritten Numeral Recognition Based on Multi-Scale Features and Neural Network", Computer Science, 2013, 40, (8), pp. 316-318.
- [4] REN Mei-li, and MENG Liang, "Handwriting digit recognition based on prototype generation technique", Computer Engineering and Design, 2015, (8), pp. 2211-2216.
- [5] Hosseiniasl E, and Guha A, "Similarity-based text recognition by deeply supervised Siamese network", Proceedings of Future Technologies Conference, USA: IEEE Press, 2015, pp. 1-7.
- [6] Ahmed E, Jones M, and Marks T K, "An improved deep learning architecture for person re-identification", Computer Vision and Pattern Recognition, USA: IEEE Press, 2015, pp. 3908-3916.
- [7] Shopon M, Mohammed N, and Abedin M A, "Image augmentation by blocky artifact in deep convolutional neural network for handwritten digit recognition", IEEE International Conference on Imaging, Vision & Pattern Recognition, IEEE, 2017, pp. 1-6.
- [8] DAS N, SARKAR R, and BASU S, et al, "A Genetic Algorithm Based Region Sampling for Selection of Local Features in Handwritten Digit Recognition Application", Applied Soft Computing, 2012, 12, (5), pp. 1592-1606.
- [9] HOSSEINI-ASL E and GUHA A, "Similarity-based Text Recognition by Deeply Supervised Siamese Network", Proceedings of Future Technologies Conference. Washington D.C., USA: IEEE Press, 2015, pp. 1-7.
- [10] AHM ED E, JONES M, and MARKS T K, "An Improved Deep Learning Architecture for Person Re-identification", Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition, Washington D. C., USA: IEEE Press, 2015, pp. 3908-3916.
- [11] LECUN Y, BOTTOU L, and BENGIO Y, et al, "Gradient-based Learning Applied to Document Recognition", Proceedings of the IEEE, 1998, 86, (11), pp. 2278-2324.
- [12] Wang Y, and Quan C, "Asymmetric optical image encryption based on an improved amplitude-phase retrieval algorithm", Optics and Lasers

- in Engineering, 2016,78,(10), pp.8-16.
- [13] Liu Xiaoyong, Cao Yiping, and Lu Pei, “Research on Optical Image Encryption Technique with Compressed Sensing”, *Acta Optica Sinica*, 2014, 34, (3), pp.91-99.
 - [14] Liu X Y, Cao Y P, Lu P. “Optical image encryption technique based on compressed sensing and Arnold transformation”, *Optik-International Journal for Light and Electron Optics*, 2013, 124, (24), pp. 6590-6593.
 - [15] L Gang, “Recognition of Multi-Font style Characters Based on Convolutional Neural Network”, *Computational Intelligence and Design (ISCID)*, 2011 Fourth International Symposium on IEEE,2011, pp. 223-225.
 - [16] Sun Y, Wang X and Tang X, “Deep Convolutional Network Cascade for Facial Point Detection”, *Conference on Computer Vision and Pattern Recognition*, IEEE, 2013, pp.3476-3483.
 - [17] Claudia I, Gonzalez , Patricia Melin, Juan R, Castro, Olivia Mendoza, and Oscar Castillo, “An improved sobel edge detection method based on generalized type-2 fuzzy logic, *Soft Computing*, 2016.
 - [18] Zeiler M D, and Fergus R, “Stochastic Pooling for Regularization of Deep Convolutional Neural Networks”, *Eprint Arxiv*, 2013.
 - [19] Timothy Dozat, “Incorporating Nesterov Momentum into Adam”, *ICLR Workshop*,(1), 2016.
 - [20] Reddi S J, Kale S, and Kumar S, “On the Convergence of ADAM and Beyond”, *arXiv*, 2019, arXiv:1904.09237.
 - [21] WEI Feng, and SHAN Lei, “A Study on Handwritten Digital Recognition Technology Based on CNN Optimization”, *Journal of Lianyungang Technical College*,2020.
 - [22] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, and Rabinovich A, “Going deeper with convolutions”, In *Computer Vision and Pattern Recognition (CVPR)*, 2015.
 - [23] Talathi S S, “Hyper-parameter optimization of deep convolutional networks for object recognition”, In *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3982–3986.
 - [24] Albelwi S, and Mahmood A, “Analysis of instance selection algorithms on large datasets with deep convolutional neural networks”, In *2016 IEEE Long Island Systems, Applications and Technology Conference(LISAT)*, 1-5,2016.
 - [25] Albelwi S, and Mahmood A, “A framework for designing the architectures of deep convolutional neural networks”, *Entropy*, 19(6), 2017.
 - [26] Andonie R, “Hyperparameter optimization in learning systems”, *Journal of Membrane Computing*,2019.
 - [27] L Gang, “Recognition of Multi-Font style Characters Based on Convolutional Neural Network”, *Computational Intelligence and Design (ISCID)*, 2011 Fourth International Symposium on. IEEE,2011, pp. 223-225.