Applied Convolutional Neural Network Algorithm in Handwritten Recognition for Digital Transformation

Nguyen Tuan[1], Duc-Hoang Chu[2], Nguyen Huu Phat[3]

[1]Management School of Culture, Sport and Tourism, 128A Ho Tung Mau, Caugiay, Hanoi, Vietnam

[2]Ministry of Science and Technology, 39 Tran Hung Dao, Hoan Kiem, Hanoi, Vietnam

[3]Hanoi University of Science and Technology, 01 Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam

Abstract: ***In the paper, we exploit a small branch of identification handwritten recognition problem. That is the recognition of letters, handwritten charactersfor digital transformation. The information is transmitted digitally using handwriting capture camera. Images will then be processed and put into identification for information by text. In the paper, we propose the handwriting recognition algorithms based on convolutional neural network (CNN). The solution method is to separate character from handwritten letters taken by camera, and then applying recognition algorithm to determine what is the letter. The goal of the paper is to reduce the parameter of CNN. Experimental resultsusing MATLAB show that the proposal algorithm improves up to 90 percent when selecting suitable learning factor of neural networks. The flexibility of this design allows it to extend to other languages easily.***

Keywords: Recognition algorithm, contours, neural networks, handwritten, image processing

1. Introduction

Currently, the strong development of identification algorithms has opened up applications using computer vision and machine learning. Handwriting recognition is one of interesting applications. However, handwriting has a disadvantage that computers cannot recognize. Therefore, recognition algorithm will help computer convert handwriting from image to text. Besides, handwriting recognition can be applied in a number of fields such as digitization, helping disabled people to control robots [1] [6].

Computer vision solves the problem of how computers can understand digital images and videos. It seeks to automate the tasks that human visual system can do. Computer vision includes methods for acquiring, processing, analyzing and understanding digital images, and extracting multi-dimensional data to produce digital or symbolic information. The development of this field for identification and understanding of digital image of computers is interesting. The subfields of computer vision include scene reconstruction, event detection, video tracking, object recognition, pattern recognition, three-dimensional gesture estimation, motion estimation, and image recovery [7] ÷ [12].

On the other hand, pattern recognition is machine learning industry. It is intended to classify data (samples) based on prior knowledge or statistical information extracted from available patterns. Handwriting recognition helps computer analyze and identify characters. This is development direction that can be applied in a number of areas such as digitizing documents, reading postal addresses, helping to order robots to support people based on converting information from digital images into text.

Digital Transformation initiatives often falter when trying to eliminate the processing of both physical and digital documents. Fortunately, handwritten data does not need to sound the death knell for a Digital Transformation initiative, as it does not preclude the digitization and automated processing of critical handwritten business information.

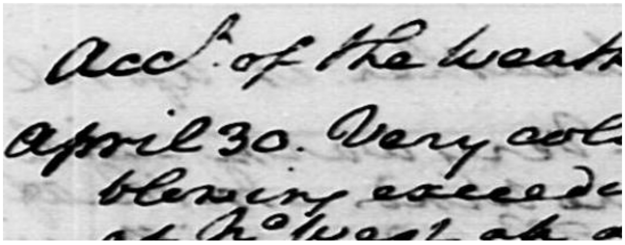


Figure 1: Examples of typical image degradations frombleed through of ink from the other side of the document

Therefore, we propose handwriting recognition algorithms in this paper. The algorithm uses images captured by camera, and then analyzing and calculating to be able to produce textual data that computer can understand based on neuralnetworks. The object to determine accuracy of algorithm is handwriting on white paper.

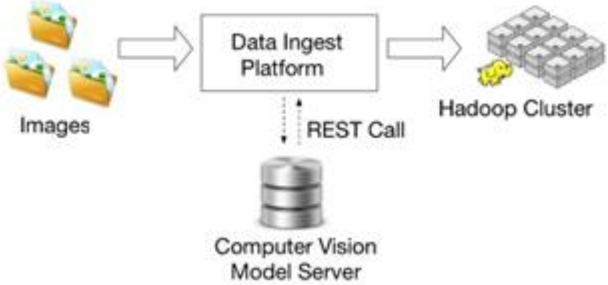


Figure 2: Diagram of working perceptions [11]

2. Neural network

2.1 Convolutional Neural network

The core idea of convolutional neural networks is to capture local features. For text, a local feature is a sliding window of several words, similar to an N-gram. The advantage of convolutional neural networks is that they can automatically combine and filter N-gram features to obtain semantic information of different abstract levels, and the training speed is faster due to the weight sharing mechanism.

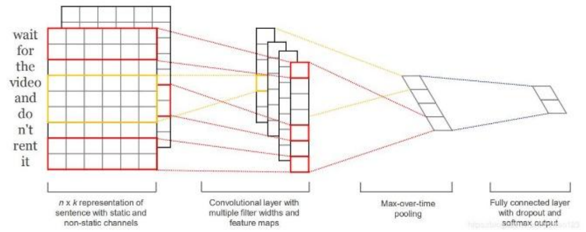


Figure 3: Diagram of working perceptions [11].

The input layer is a matrix of N × K dimensions, N corresponds to the number of articles, and K is the dimension of the word vector. The K-dimensional vector of each word can be trained in other corpora through word2vec in advance, or it can be trained as an unknown parameter by the network. The structure in the above figure indicates that the input has two channels, one-word embedding is static, the training process does not change, and the other is initialized in the same way, but it will change as a parameter with the network training process.

The second layer is a 1D convolution layer. The dimension of the convolution kernel is h×K. K is fixed, consistent with the word vector dimension, and h can be changed. Convolution kernels of different sizes can be defined to extract features.

The third layer is the pooling layer, which can be 1-1000 pooling, that is, the maximum value of each convolution kernel is extracted, or K-Max pooling can be used, that is, the top K largest features are selected.

Splicing the features into a one-dimensional vector, accessing a fully connected layer and classifying.

In the handwriting recognition problem, we use three-layer neural network, namely input, output, and hidden layer. Each pixel value will be used as an input. This value will be converted to a true value within [0,1]. Therefore, an image identified as a 28×28 image will provide 784 input samples. The English alphabet has 26 letters since there will be 26 corresponding output. The output value of output neuron at position i equals 1 represents network that recognizes letter i. Therefore, network will have 784 inputs and 26 outputs. The number of neurons in hidden layer will be changed to find the best solution that balances between accuracy and time of network training.

2.2 Training for the network

Neural networks are built based on understanding of human brain. Network consists of a series of different link units. It is a mapping between input and output set. Each of them is called a neuron. Two important types of artificial neurons are perceptron and sigmoid.

Perceptron was developed in 1950 by scientist Frank Rosenblatt based on inspiration from previous research by Warren McCulloch and Walter Pitts. The way perceptrons works is to use several binary inputs x1, x2, ... and create a binary output as shown in Fig. 2.

We suppose that perceptions have three inputs x1, x2, x3to calculate output, we assume that weights w1, w2, w3 … are real numbers expressing the importance of corresponding input. The output of 0 or 1 neuron is determined by comparing sigmoid of products wj∗xj with a threshold value in expression

$$output = \begin{cases} 0 \text{ if } \sum_j w_j x_j \le threshold \\ 1 \text{ if } \sum_j w_j x_j > threshold \end{cases} \quad (1)$$

We see that when we change weight or thresholds, we can create different outputs. In other words, we can make different decisions.

Given a preprocessed and tokenized sentence containing two entity types of interest (i.e. chemical and disease), our model first extracts the shortest dependency path (SDP) (on the dependency tree) between such two entities. The SDP contains tokens (together with dependency relations between them) that are important for understanding the semantic connection between two entities. For multi-channel embedding, instead of concatenating three partial embeddings of each token on SDP we maintain three independent embedding channels for them.

In this paper, handwriting recognition algorithms will be implemented on software using C ++ and python. To complete process of identification, three stages need to be addressed. First stage is pre-processing image. The second stage is to separate words in image into separate letters and standardize characters to a fixed size of 28×28. Third stage is recognition of words in standardized images using neural networks.

3. Proposed model

3.1 Pre-processing images

The handwriting will be written in black ink on white paper. We then use Samsung Galaxy Note 5 camera to record them. They are performed pre-processing. Each image has Pre-processing images by 5 steps: (1) Convert to gray image; (2) Using Gaussian Blur; (3) Filter; (4) Convert to bit using Otsu function; (5) Filter by using closing function.

Given that the added costs of improved image capture quality (over and above the minimum necessary for the project) are often small compared with the costs of repeating the project for more quality sometime in the future, some additional likely uses for the digital captures from the project were considered. These images can provide a permanent record of each item in the collection and its condition at the time of capture.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

3.2 Splitting and normalizing images

Identifying a word will be difficult if we put into the network a missing data. To make identification simpler, we input each letter. It reduces the number of inputs dimensions and limits the number of labels used. Networking will be performed more feasible [7]. Images obtained after the pre-processing step will be separated into separate letters using Opencv to find by find Contours function. The selected letters will be cut into separate images.

Image normalization is an important step in identification to ensure that each input parameter has the same data distribution. In this case, it is the pixels. This makes convergence faster in network. The result of the above separation step is that the letters fit in images of different sizes. In this normalization step, we will shrink image to a size smaller than 20×20 and fix for every letter.

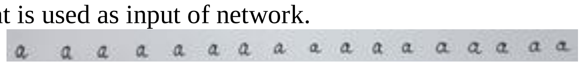
Another technique using to enhance performance of identification is the center of mass technique. The key of algorithm is that density distribution of pixel is balanced around a central point. It makes difference among letters. We set out a 28×28 frame. We determine center of 20×20 input image and then place this image in 28×28 frame where focus of two images overlaps.

Finally, we obtain an image 28×28 containing character that is placed in the center of image. In the next step, this image is uploaded to the network for teaching or identification.

3.3 Identifying using neural network

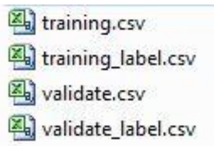
With neural network model in section 2, we build teaching and testing data set. The network is built by Python language with three classes of input, hidden, and output. The input consists of 784 neurons corresponding to 784 pixel values of each 28×28 image. The pixel value of 28×28 image will be spread out into one dimensional array of 784×1. The hidden layer consists of 20 neurons. The output layer consists of 26 neurons corresponding to 26 English letters of output. If the output have value of 1 in any position, the letter of that will be recognized.

After building program, we need to train in order for the network to work. A teaching and testing data sets are built. A sample table is printed on A4 paper, and then distributed to different people to collect for information. We use the program for pretreatment and separation for sampling. Samples are standardized into 28×28 image. A database of teaching and testing samples is created in excel ".csv" file that is used as input of network.



a)

b)



c)

Figure 4: Input data: (a) Sample, (b) Post-split sample, and (c) Database.

3.4 Digital transformation

The term "digital transformation" pervades the modern world. However, a generally valid definition for the concept of digital transformation does not yet exist. Some researchers focus on specific technologies to explain an "organizational shift to big data analytics, while others focus on technology in general as the driver of radical change. We want to underline, however, that DT does not merely refer to technological changes, but also to the impacts thereof on the organization itself. It leads to handwritten recognition for digital transformationas well as organizational structures and management concepts. The changes that come along with the digitalization affect people, society, communication and the whole business.

Many of the technologies that affect digital transformationare not new. The innovation is about "combinations of information, computing, communication, and connectivity technologies". The major technological areas which enable DT are very diverse and traditionally called "general purpose technologies". These include, for example, cyber-physical systems (CPS), (industrial) internet of things (I/IoT), cloud computing (CC), big data (BD), artificial intelligence but also augmented and virtual reality.

The digital transformation release to convolutional neural network algorithm in handwritten recognition including three significant changes: (1) digitally supported and cross-linked processes, (2) digitally enabled communication, and (3) new ways of value generation based on digital innovations or gained digital data. These major changes can be found worldwide and in all industries. Moreover, digital transformation has spawned new business areas such as e-government, e-banking, e-marketing, e-tourism and the highly innovative field of e-health where two research areas (medicine and information systems).

4. Result and Discussion

In this paper, we collected 14541 samples for teaching and 260 samples for testing. Test samples are randomly selected each with 10 letters in the database and certainly not in the group to teach. We will conduct to survey of network parameters to evaluate its influence on the results obtained.

4.1 Simulation results

In first scenario, we obtain an image 28×28 containing character that is placed in the center of image. In the next step, this image is uploaded to the network for teaching or identification. We will change the number of neurons of hidden layer. Other parameters have not changed. The results are shown in Tab. 1 and Fig. 5.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Table 1: Evaluating the parameters of hidden layer.

Name	Value
Coefficients	1
Size of an input batch (mini-batch)	10
Number of times of teaching review (epoch)	100
The number of hidden layer neurons (varies)	15, 20, 30, 40, 70, 100

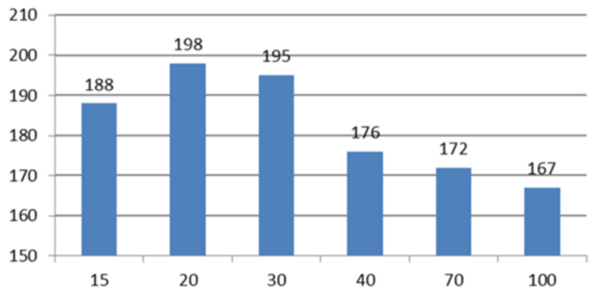


Figure 5: Evaluate number of hidden layer neurons

In second scenario, we choose coefficients of different size s from 0.001 to 100. Other parameters are fixed and the results are shown in Tab. 2 and Fig. 6.

Table 2: Evaluating learning parameters

Name	Value
Coefficients	0.001, …, 100
Size of an input batch	10
Number of times of teaching review (epoch)	100
The number of hidden layer neurons (varies)	20

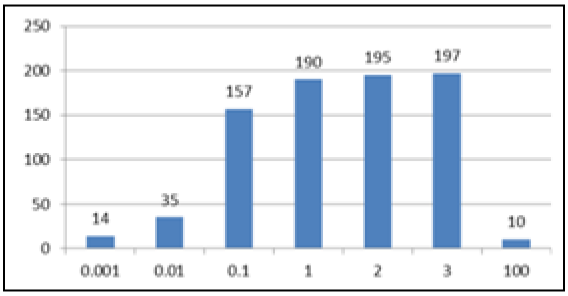


Figure 6: The results evaluate learning parameter

In third scenario, the key of algorithm is that density distribution of pixel is balanced around a central point. It makes difference among letters. We set out a 28 × 28 frame. We determine center of 20 × 20 input image and then place this image in 28 × 28 frame where focus of two images overlaps. We change input size and keep learning coefficients. The results are shown in Tab. 3 and Fig. 7.

Table 3: Evaluating learning parameters

Name	Value
Coefficients	3
Size of an input batch (mini-batch)	100, 10, 5, 3, 1
Number of times of teaching review (epoch)	100
The number of hidden layer neurons (varies)	20

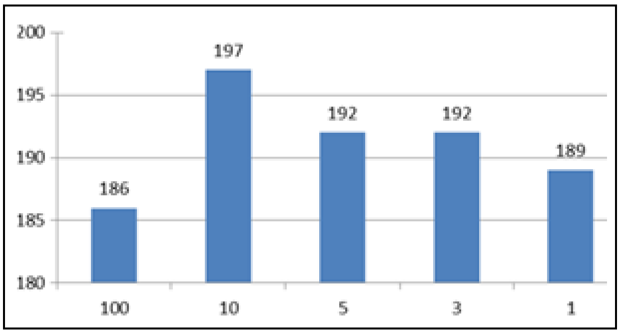


Figure 7: The results evaluate the size of teaching input.

3.2 Discussion

Image quality depends on the project's planning choices and implementation. Project designers need to consider what standard practices they will follow for input resolution and bit depth, layout and cropping, image capture metric (including color management), and the particular features of the capture device and its software. Benchmarking for any given type of source material can help one select appropriate image quality parameters that capture just the amount of information needed from the source material for eventual use and display. By maximizing the image quality of the digital master files, managers can ensure the on-going value of their efforts, and ease the process of derivative file production. Quality is necessarily limited by the size of the digital image file, which places an upper limit on the amount of information that is saved. The size of a digital image file depends on the size of the original and the resolution of capture (number of pixels per inch in both height and width that are sampled from the original to create the digital image), the number of channels (typically 3: Red, Green, and Blue: "RGB"), and the bit depth (the number of data bits used to store the image data for one pixel).

In Fig. 7, we found that when the number of hidden neurons increased from low to high (about 15), the accuracy increased. However, when the threshold reached certain level, accuracy of method will be decreased. The highest accuracy is when number of neurons is 20 and will decrease as number of neurons reaches 100. When number of neurons increases, the time to learn will increase. Therefore, we need to balance. In Fig. 6, we find that when coefficient is small at 0.001, the accuracy is low after 100 iterations. This shows that small coefficients are difficult to update weight and bias values that make network adaptive slowly.

When we increase the coefficients to 0.01, efficiency has increased. The effect is obvious when we increase coefficients to 0.1; 1; 2; 3. Results exceeded approximately 197/260 samples. However, when the coefficients are increased to 100, accuracy dropped marked slowly. The reason is that we need small coefficients to approximate formula for calculating in cost function C and expect this value to be negative. In this case, C is positive and we cannot find the smallest point (the bottom of the valley) of the cost function C. In the stochastic gradient descent algorithm, we need a sufficient number of samples to be able to approximate gradient of C. Therefore, we decided to choose the size of 10 samples for a first phase of teaching online.

Therefore, we need to find the most optimal set of parameters with influence of learning outcomes. This requires to perform many times. Finally, we selected the following set of parameters as follows: learning coefficient is 3, number of hidden class neurons is 20, size of input is 10, and number of reviews of the entire teaching sample is 100.

Digital transformation (DT) has become a buzzword, triggering different disciplines in research and influencing practice, which leads to independent research streams. Scholars investigate the antecedents, contingencies, and consequences of these disruptive technologies by examining the use of single technologies or of digitization, in general.

5. Conclusion

Within scope of this paper, we have performed to study and develop handwriting recognition algorithms. The simulation results give positive results. The text is well separated, clear and has little noise. The recognition rate is more than 80 percent. We realize that the limitation of the algorithm is raw input sample. The algorithm requirement needs to write letters separately. In fact, there are many types of writing that are more challenging. In addition, number of collected samples is not enough that affects the accuracy.

For future work, the filter will be expanded by AI software accelerometers in the processing update. The implementation is similar to the machine learning feedback. Also, this long-term correction will help handwritten recognition for digital transformation.

References

[1] J. Cai and Z. Q. Liu, Pattern recognition using markov random field models, Pattern Recognition, 35(3):725733, 2002.
[2] G. Chen, D. B. Tien, and A. Krzyzak, Invariant pattern recognition using radon, dualtree complex wavelet and fourier transforms, Pattern Recognition, 42(9), pp. 20132019, 2009.
[3] G. Chen and W. F. Xie, Pattern recognition with svm and dual-tree complex wavelets, Image Vision Comput., 25(6), pp. 960966, 2007.
[4] K. Feng, Z. Jiang, W. He, and B. Ma, A recognition and novelty detection approach based on curvelet transform, nonlinear pca and svm with application to indicator diagram diagnosis, Expert Syst. Appl., 38(10), pp.1272112729, 2011.
[5] L. L. SHEN and Z. JI, Gabor wavelet selection and svm classification for object recognition, Acta AutomaticaSinica, 35(4), pp.350 355, 2009.
[6] S. L. Y. Tian, C. Zhao, and X. Guo, Multiple classifier combi nation for recognition of wheat leaf diseases, IntellAutom Soft Co, 17(5), pp.519529, 2011.
[7] Handwritten Character Recogniton [Online], http://tcts.fpms.ac.be/rdf/hcrinuk.htm.
[8] R. K. Thampi, ”Classification of handwritten characters: definition, iden-tification and utilisation of regional characteristics for lower case english characters,” Aug. 1994.
[9] Using contours to extract characters in OpenCV, [Online],http://blog.ayoungprogrammer.com/2013/01/equation-ocr-part-1-using-contours-to.html/.
[10] W. H. Abdulla, A .O. M. Saleh, A. H. Morad, ”A preprocessi ng algorithm for hand-written character recognition”, Pattern Recognition Letters, vol. 7, no. 1, pp 13-18, 1998.
[11] ”Neural Networks and Deep learning”, [Online], http://neuralnetworksanddeeplearning.com/index.html.
[12] Y. Nesterov, A method for unconstrained convex minimization problem with the rate of convergence o(1/k2), Doklady ANSSSR, vol. 269, pp. 543 547, 1983.

Author Profile

Phat Nguyen Huureceived his B.E. (2003), M.S.(2005) degrees in Electronics and Telecommunications at Hanoi University of Science and Technology (HUST), Vietnam, and Ph.D. degree (2012) in Computer Science at Shibaura Institute of Technology, Japan. Currently, he is lecturer at School of Electronics and Telecommunications, HUST Vietnam. His research interests include digital image and video processing, wireless networks, ad-hoc and sensor network, and intelligent traffic system (ITS) and internet of things (IoT). He received the bestconference paper award in SoftCOM (2011), best student grant award in APNOMS (2011), Hisayoshi yanai honorary award by Shibaura Institute of Technology (SIT).

Hoang Chu Duc Head of R&D Funding Department, National Technology Innovation Fund (NATIF); Counselor of Vingroup Innovation Foundation (VINIF); Founder and Chairman of ZINMED (A Startup in management and treatment diabetics); Dr Duc-Hoang Chu get PhD in Biomedical Engineering at Hanoi University of Science and Technology and Washington University at St. Louis, He also former Researcher in Hanoi University of Science and Technology (2003 - 2016). Dr Duc-Hoang Chu has over 30 research papers, 20 conferences, 03 Books and 5 projects in Biomedical, Healthcare, ICT, Technology Management and Innovation. He has also consulted and assisted over 100 scientists and businesses on innovation and currently manages and sponsors over 30 science and technology projects.Main research interests: Biomedical engineering (Arrhythmia, Diabetes, Medical information system, Medical software); Information, Communication and Technology; Science Technology and Innovation Leaderships.

Tuan Nguyen received his B.E. (2004), M.S. (2010) degrees in Electronics and Computer Science at Posts and Telecommunications Institute of Technology, Vietnam. Currently, he is lecturer at Management School of Culture, Sport and Tourism, Hanoi, Vietnam. His research interests include digital image and video processing, tourist information system, wireless networks, ad-hoc and sensor network.