IMPLEMENT CNN NETWORK BASED ON FPGA FOR MNIST HANDWRITING RECOGNITION

Ninh Nguyen Thi Hai[1], Anh Dang Thi Ngoc[1]

[1]Thai Nguyen University of Technology, Thai Nguyen city, Viet Nam

ABSTRACT

This article presents a solution to recognize MNIST handwriting using a convolutional neural network (CNN). From the solution, the authors implemented a CNN network on FPGA technology to increase processing speed and reduce power consumption. Through the testing process, the solution has met the basic requirements of the MNIST handwriting recognition problem using the convolutional neural network method.

Keyword: *Recognize handwriting, Convolutional neural network, CNN, MNIST*

1. INTRODUCTION

Convolutional neural network (CNN) is a highly accurate artificial neural network model, deployed in many applications, especially in image recognition [1] [2]. To achieve the required accuracy, a conventional CNN network model must perform a large number of mathematical operations with floating point numbers. Because of that, CNN networks often require complex hardware with fast processing speed and large memory space to train and perform recognition tasks.

The method of implementing a CNN network on FPGA architecture has been proposed by many researchers around the world [6] [7] as well as domestically [9] to achieve two goals, both increasing calculation speed when processing images to meet real-time requirements while optimizing energy consumption.

To evaluate the effectiveness of the CNN network in the image recognition problem, thereby making recommendations for future research, we need to rely on large image databases recognized by the international community. One of them is the MNIST database (Modified National Institute of Standards and Technology database). This is a large dataset containing handwritten digits that is often used in training various image processing systems and is of great research interest to the scientific community [2], [8].

In this article, the authors focus on researching the MNIST handwriting recognition method using the CNN network model and implementing it on FPGA technology to increase processing speed and reduce power consumption.

The article is divided into 6 parts: (1) Introduction; (2) CNN network; (3) CNN network application solution for MNIST handwriting recognition; (4) Implement cnn network based on fpga technology; (5) Results; (6) Conclusion.

2. CNN NETWORK

2.1 Structure of artificial neural network

Neural networks (NN), also known as artificial neural networks (ANN) or simulated neural networks (SNN), are a subset of learning. machine, the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way biological neurons transmit signals to each other.

Artificial neurons have software modules, called nodes. An artificial neural network is a software program or algorithm that essentially uses a computer system to solve math problems.

Each artificial neural network is a multi-layer Perceptron, a Neural Network usually includes 3 specific types of layers as shown in Figure 1.

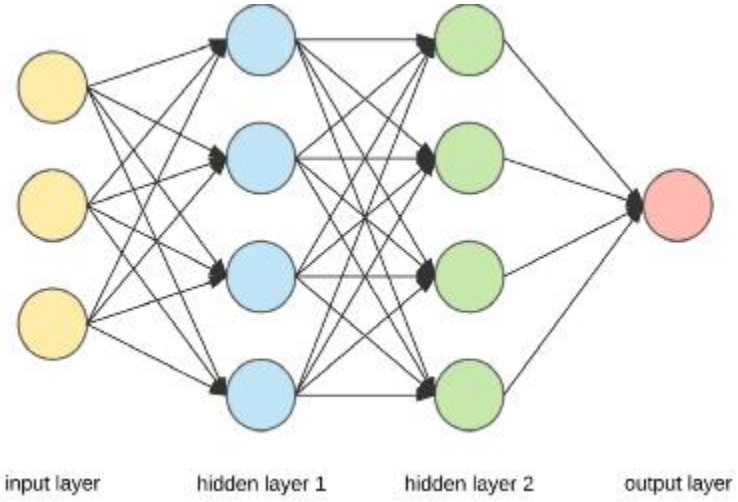

Fig - 1: Structure of artificial neural network

2.2 CNN network structure

A CNN network is a set of overlapping Convolution layers and uses nonlinear activation functions such as ReLU and Tanh to activate the weights in the nodes. Each layer, after passing activation functions, will create more abstract information for the next layers.

Each layer uses different filters, usually hundreds of thousands of such filters, and combines their results. In addition, there are some other layers such as pooling/subsampling layer used to filter more useful information (remove noise information).

During network training, CNN automatically learns values through filter layers based on how you do it. For example, in the image classification task, CNNs will try to find the optimal parameters for the corresponding filters in the order raw pixel > edges > shapes > facial > high-level features. The final layer is used to classify images.

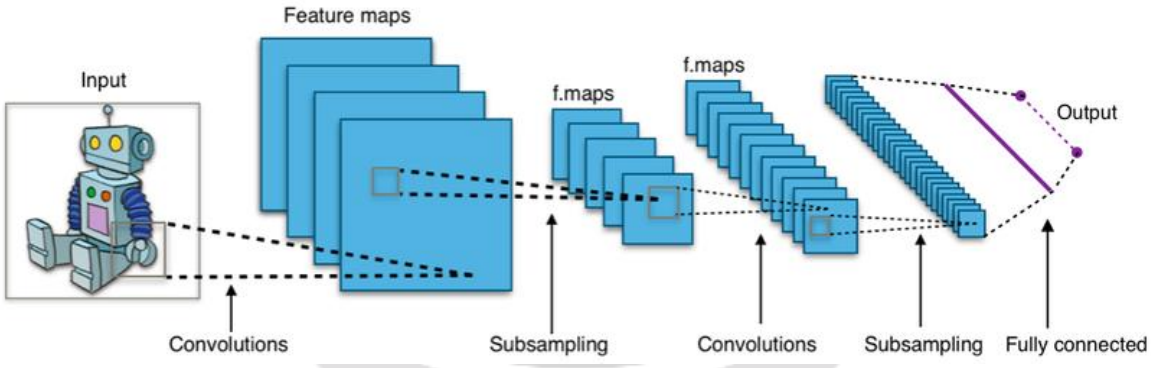

Fig - 2: CNN network structure

2.3 Implement CNN network on FPGA platform

In this part, we implement CNN network in NB2DSK01 Nanoboard. It is an Altium Desktop FPGA, microprocessor, and embedded system development toolkit capable of configuring hardware to suit design needs and purposes. It is a hardware platform integrated with many implementation functions. and interaction and debugging of digital designs on a single board. This is truly a low-cost programming and simulation tool that can perform many functions well and powerfully. For those reasons, the NanoBoard NB2DSK01 Altium has been widely used in laboratories, digital product development centers and as experimental modules for students in universities.

Fig - 3: Nanoboard NB2 board

3. CNN NETWORK APPLICATION FOR MNIST HANDWRITING RECOGNITION

Currently, handwriting recognition is still a major challenge for researchers. The biggest difficulty when studying the problem of handwriting recognition is the extreme variation in each person's writing style. The same person writes, but sometimes there are many differences in writing style depending on the context, a person's writing style can also change over time or according to habit.... This causes many difficulties. Obstacles in feature extraction as well as recognition model selection.

3.1 Handwriting recognition methods

There are various pattern recognition methods widely applied in handwriting recognition systems. These methods can be integrated in the following approaches: Pattern matching, statistics, structure, neural networks and SVM.

- Pattern matching: is the simplest word recognition technique based on comparing prototypes with each other to identify characters or words.

- Structural approach: This method poses a problem to solve the general word recognition problem. Currently, popular structural recognition is to extract all the features of the learning sample, partition the character table based on these features, then the image to be recognized will be extracted features, then compared on partition table to find characters with matching characteristics.

- Hidden Markov Model HMM: HMM is a finite state probability model in the form of process generation by defining link probabilities on observation sequences.

- SVM (Support Vector Machines): SVM is considered an advanced machine learning method that is widely applied in the fields of data mining and computer vision...

- Neural network: is defined as a computational structure consisting of many "neural" processors interconnected in parallel. Due to the parallel nature of neurons, it can perform calculations at higher speeds than other layering techniques.

3.2 Network architecture for handwriting recognition problem from MNIST sample set

MNIST is a handwriting dataset derived from the NIST dataset provided by the National Institute of Standards and Technology (NIST).

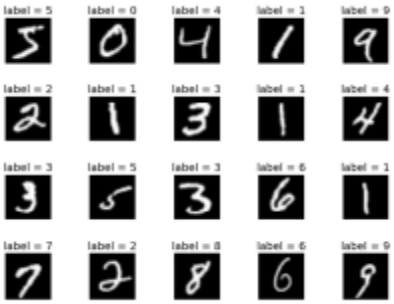

Fig - 4: Examples of some samples of the MNIST data set

In this article, the authors will build a model to solve the problem of handwritten digit recognition. The model will be a combination of CNN and MLP layers. In particular, layers of CNN are used to filter information in images to build feature vectors used to classify images. MLP acts as a classifier, receiving as input the feature vectors built by the layers of the CNN and the output are the classification results. Figure 5 is the general architecture of the solution.

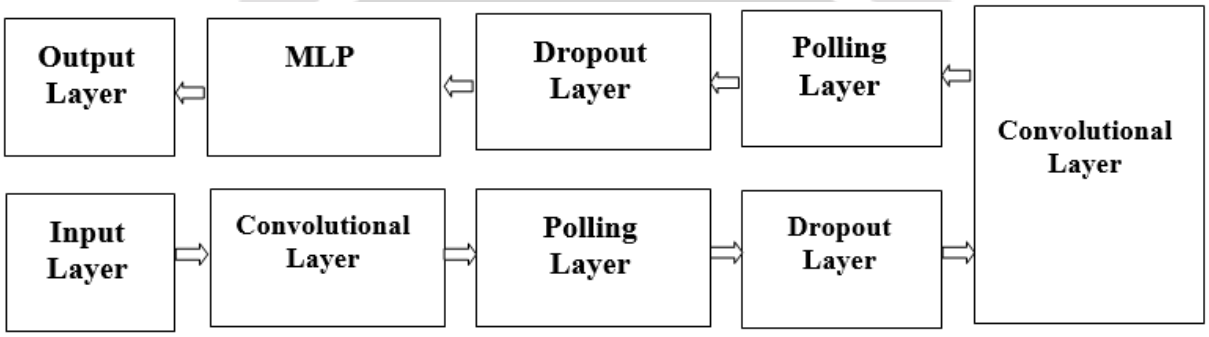

Fig - 5: CNN network model for handwriting recognition problem

4. IMPLEMENT CNN NETWORK BASED ON FPGA TECHNOLOGY

4.1 Describe the problem of implementing a CNN network for MNIST handwriting recognition

With the ability to handle large amounts of inputs and process them to infer hidden and complex relationships, CNN networks have played an important role in image processing, especially handwritten character recognition.

The main challenge arising from the problem of Hand Written Digits Recognition (HWDR) lies in the fact that handwritten digits (within the same digit) vary greatly in shape, line width and style, even if they are standardized in size and precisely focused.

One of the famous datasets used in research on the HWDR problem is the MNIST (Modified National Institute of Standards and Technology database). MNIST provides two separate datasets. The first dataset contains 60,000 training images and their corresponding digits from 0 to 9, and the second dataset contains 10,000 test images and their corresponding digits. Each image is an 8-bit grayscale image of size 28 28. Figure 6 depicts a sample image from MNIST representing the digit 5.

Fig - 6: Handwritten number "5" from the MNIST dataset

To build and evaluate the effectiveness of the experimental problem, we proceed with the following steps:

Step 1: Search for the optimal architecture and parameters of the CNN network for the handwriting recognition problem
Step 2: Implement CNN network using FPGA Development Kit NB2DSK01 and Altium Designer software
Step 3: Evaluate performance

4.2 Implement CNN network architecture on Kit NB2DSK01 and Altium Designer software

To evaluate the performance of the CNN network for the MNIST handwriting recognition problem on the FPGA platform. The authors plan to build an embedded system consisting of 10,000 image files to be tested that will be stored in *.jpg format on an SD memory card. The system will use these image files as input to the CNN network. The image will be displayed on the TFT screen and the image recognition results will be displayed on the 7-segment LED.

Implementing the above CNN network on Kit NB2DSK01 and Altium Designer software has the following results:

- Open bus architecture of the experimental model

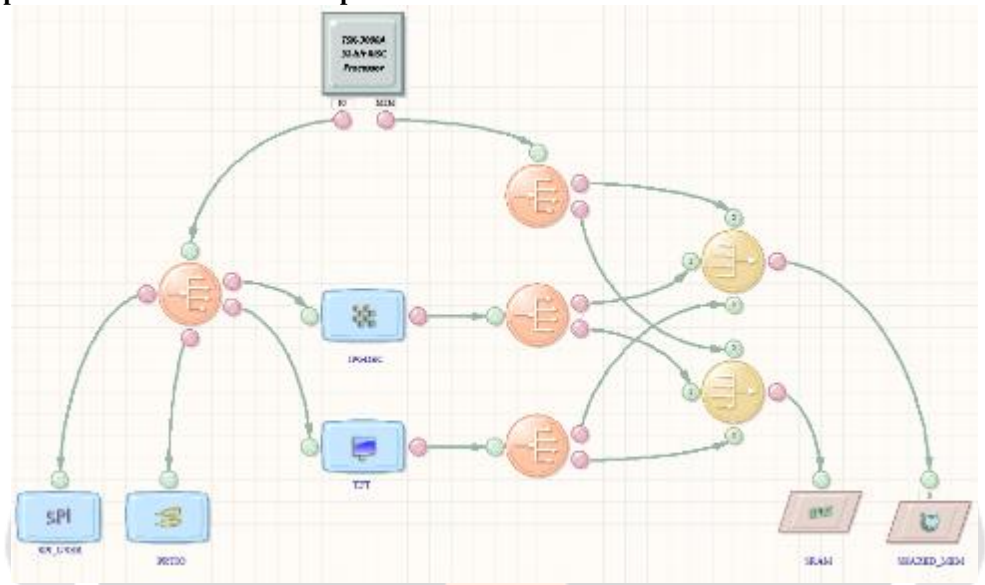

Fig - 7: Open bus architecture of the experimental model

- Principle diagram describing the device connection for the experimental model

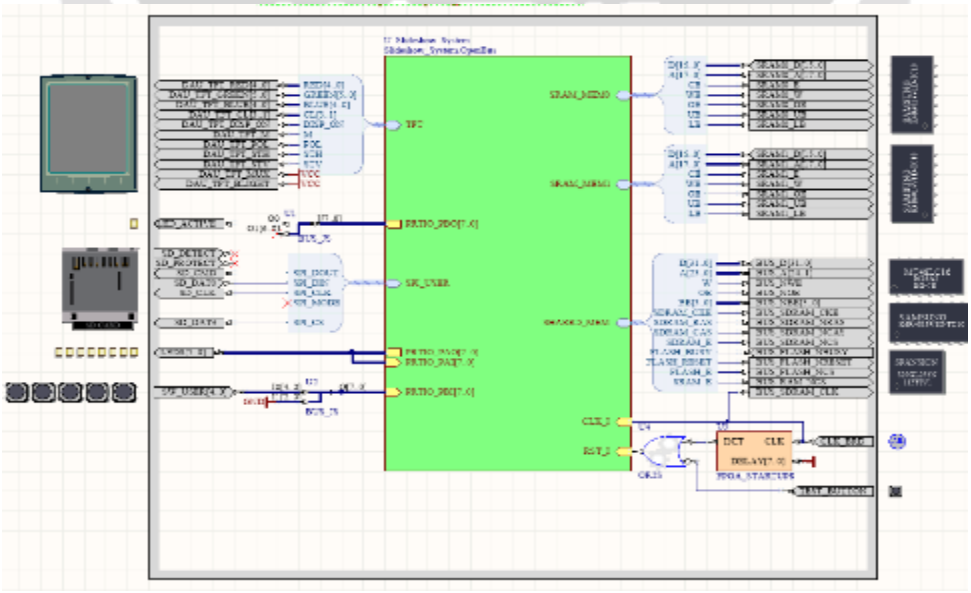

Fig - 8: Principle diagram describing the device connection for the experimental model

- Software platform library architecture of the experimental model

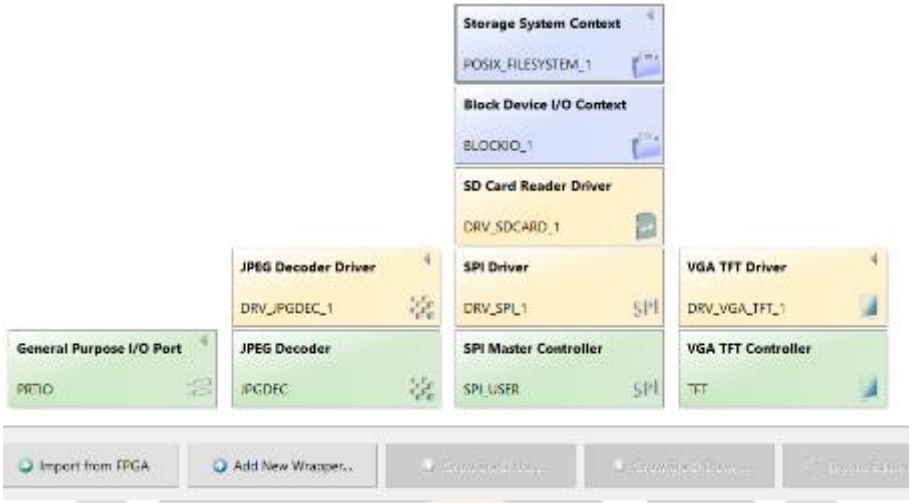

Fig - 9: Software platform library architecture of the experimental model

5. RESULTS

Figure 10 shows the correct performance results of the experimental model when the input is an image of the number 3. To accurately evaluate the performance of the CNN network, the authors also wrote software for the evaluation system. Evaluation of the recognition results of 10,000 test images in the MNIST data set. These images are saved to the SD memory card and are sequentially fed through the CNN network for identification. The system's performance results reached 98.08% after 76 seconds.

Fig - 10: Execution results on Kit NB2DSK01

6. CONCLUSION

Handwriting recognition is an attractive research field because it can be applied in many practical problems. This is also a complex problem but it will be solved if we know how to apply research achievements in fields such as digital signal processing, artificial intelligence... In particular, deploying CNN networks on the FPGA hardware will shorten operating time and is very useful for real-time image recognition applications such as cameras and videos. In the near future, the authors will conduct experiments on some similar problems of human face recognition, fruit classification, etc. via camera or video. Research and deploy on the FPGA platform other more effective DeepLearning network architectures for recognition problems such as Long short-term memory (LSTM) networks, Deep Belief Networks (DBNs), Autoencoders (AEs).

ACKNOWLEDGEMENT

The authors thanks for the financial support of Thai Nguyen University of Technology for the research team to complete this study.

7. REFERENCES

[1] Byerly, A., Kalganova, T., Ott, R. (2022). The Current State of the Art in Deep Learning for Image Classification: A Review. In: Arai, K. (eds) Intelligent Computing. SAI 2022. Lecture Notes in Networks and Systems, vol 507. Springer, Cham. https://doi.org/10.1007/978-3-031-10464-0_7

[2] Chen, Leiyu, Shaobo Li, Qiang Bai, Jing Yang, Sanlong Jiang, and Yanming Miao. 2021. "Review of Image Classification Algorithms Based on Convolutional Neural Networks" Remote Sensing 13, no. 22: 4712. https://doi.org/10.3390/rs13224712

[3] Huỳnh Việt Thắng (2018), Nghiên cứu thiết kế lõi IP mạng nơ-ron nhân tạo cho nhận dạng mẫu trên phần cứng FPGA, Đề tài NCKH cấp bộ, ĐH Đà Nẵng

[4] LeCun Y, et al. Deep learning. Nature. 2015;521(7553):436–444. PMID: 26017442. Available from: https://doi.org/10.1038/nature14539.

[5] He K, et al. Deep Residual Learning for Image Recognition. arXiv:1512.03385, 2015; PMID: 26180094. Available from: https://doi.org/10.1109/CVPR.2016.90

[6] Chen J, Ran X. Deep Learning with Edge Computing: A Review. Proceedings of the IEEE. 2019;107(8). Available from: https://doi.org/10.1109/JPROC.2019.2921977

[7] Mittal, S. A survey of FPGA-based accelerators for convolutional neural networks. Neural Comput & Applic 32, 1109–1139 (2020). https://doi.org/10.1007/s00521-018-3761-1

[8] Yan, Fei, Zhuangzhuang Zhang, Yinping Liu, and Jia Liu. 2022. "Design of Convolutional Neural Network Processor Based on FPGA Resource Multiplexing Architecture" Sensors 22, no. 16: 5967. https://doi.org/10.3390/s22165967

[9] Shen, Guobin & Li, Jindong & Zhou, Zhi & Chen, Xiang. (2021). FPGA-Based Neural Network Acceleration for Handwritten Digit Recognition. 10.1007/978-3-030-67514-1_15.