Hand Gesture Recognition using Convolutional Neural Networks

Shengchang Lan
Dept. of Microwave Engineering
Harbin Institute of Technology
Harbin, Heilongjiang, China
lansc1015@hit.edu.cn

Zonglong He
Photonics Systems Research Lab
KAIST
Daejeon, Korea
hezonglong@kaist.ac.kr

Weichu Chen, Lijia Chen
Dept. of Microwave Engineering
Harbin Institute of Technology
Harbin, Heilongjiang, China
1150550106@stu.hit.edu.cn; ljchen@hit.edu.cn

***Abstract*—This paper introduced a hand gesture recognition method based on convolutional neural networks (CNNs). The recognition scenario consisted in a three dimensional radar array to transmit and receive 24GHz continuous electromagnetic (EM) wave, and convert the scattered EM wave to the intermediate frequency (IF) signals. This paper used the the processed frequency spectrum as the input to the CNN. Then the CNN feature detection layer learned through data training, avoiding supervised feature extraction while learning implicitly from training data. It highlighted these features through convolution operating, pooling and a softmax function. Results showed that this system could achieve a high recognition accuracy rate higher than 96%.**

I. Introduction

With its success in the fields of image recognition and classification, convolutional neural networks (CNNs) has attracted great attention in many fields [1][2]. In 1960s when Huber and Wiesel studied neurons for locality sensitivity and direction selection in the cortex of the cat, they discovered its unique network structure can effectively reduce the complexity of feedback neural network[3]. Currently, CNNs have become one of the research hotspots in many fields, especially in the field of pattern classification. The network has been widely used because it avoids the complicated pre-processing of images and can directly input original images. In recent years, developing more efficient human-computer interaction becomes increasingly attractive in the computer science society, aiming to digitally interpreting the hand gesture information to the computers based on visual images and wearable inertial sensors. However, microwave engineering has involved in this new realm as an emerging solution. Featured for simple structure, high sensitivity and narrow spectrum occupancy, Frequency Modulated Continuous Wave (FMCW) radar is able to detect the frequency shift when the transmitted electromagnetic wave is scattered by the moving targets according to Doppler Effect[4][5]. In this paper, we investigated the feasibility of designing a hand gesture recognition system using convolutional neural networks to classify basic human hand gestures including hand motion, finger motion and fist motions. The experiment result showed a satisfactory accuracy rate higher than 96% in the recognition and verified the proposed method with a promising vision in controlling the modern electronics.

II. Radar array design

Hand gestures are nature physical movements of human fingers, palm, wrist and forearm with the intention of conveying the meaningful information from brain to others. However, human hands are complex system with many joints. So we established a radar array with three commercial FMCW radar sensors perpendicularly mounted to each other to measure the gestures. The radar sensors from RfBeam Switzerland, K-LC2, operate at 24.125GHz with a bandwidth of 50MHz. The EM wave scattered by hand was heterodyned by the local oscillator and converted to the intermediate frequency ready for being digitalized by Analog-digital converters. The system structure was shown in Fig. 1. These I/Q signals contained the velocity and range information of the target.

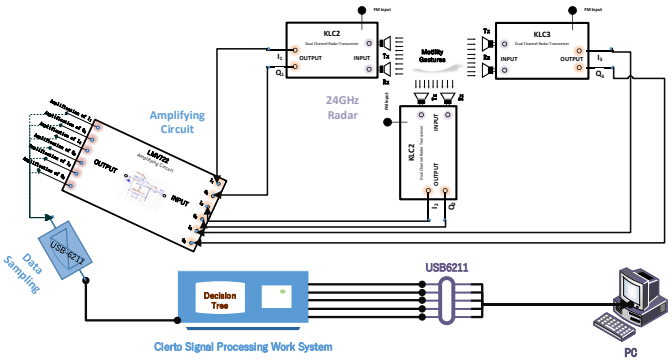

Fig. 1. The 24GHz radar array for hand gesture recognition.

III. Classification based on convolutional neural networks

Convolutional neural networks are similar to ordinary Neural Networks, making up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a nonlinearity. The convolutional neural networks relies less on the feature extraction, enabling the raw input of the data on one end to class scores at the other. And they still have a loss

function such as softmax on the fully-connected layer to fulfill the classfication.

A. network architecture

The entire structure of the proposed convolutional neural networks included the convolutional layer, pooling layer, fully connected layer and softmax layer, as shown in Fig. 2. Firstly, convolutional layer contained a convolution operation to the input, passing the result to the pooling layer. And the pooling layer combined the outputs of neuron clusters at convolutional layer into a single neuron in the next layer. In this system, we used max pooling algorithm which only takes the maximum value from each of a cluster of neurons at the convolutional layer. Then, the data was processed by the double convolutional layer and the double pooling layer. And the fully connected layers connected every neuron in one layer to every neuron in another layer, the same as the traditional multilayer neural network. Finally, after convolution operation through convolution kernel and training samples, the network yielded softmax function to highlight the largest values and suppress values below the maximum value so that the could can be classified eligibly.

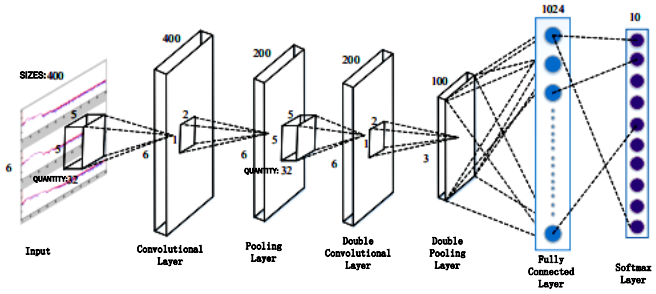

Fig. 2. The convolutional neural network system for hand gesture recognition.

B. dataset

The dataset from I/Q signal spectrum was constructed by 10 types of spectrograms, including hand lifting, taping, opening, clenching, left, right, up, down, front, back. Firstly, FFT (Fast Fourier Transformation) was performed on the original signal of the same group of 10 groups to obtain a spectrum, and logarithm of the spectrum value was used to make the spectrum curve more intuitive. Then, based on the six channels orthogonal I/Q signals and their spectrum ranging from 0 to 400Hz , a 400×6 dataset matrix was constructed as the input to the network. The dataset matrix contained the training set and the test set. To prevent confusion about the specific location of the characteristics, the max pooling algorithm should be used after convolution.

IV. EXPERIMENTS

We used the NI DAQ USB 6211 card to sample the I/Q signals of the three radar sensors with a sampling rate of 5000Hz and sampling length of 5000 points. We invited four volunteers with different genders and ages participating the experiment, providing 4000 groups of data for testing and two more volunteers, male and female, providing 2000 groups of data for constructing the TensorFlow environment and datasets. Fig. 3 showed the experiment result with convolutional neural networks. All the six hand motions included were recognized with an average accuracy of 100%, higher than finger and fist motions, which implied that the movement characteristics of these six types of hand motions was easier to be recognized. Two finger motions had the lowest accuracy in the test, only reaching to 96% (Lifting) and 96.5% (Taping), almost 1.5% lower than two fist motions which reaching to 97.5% (Opening) and 98.25% (Clenching). The reason of the difference was that the flexion and extensions of the fist or finger have the varying velocity in the gesture generation.

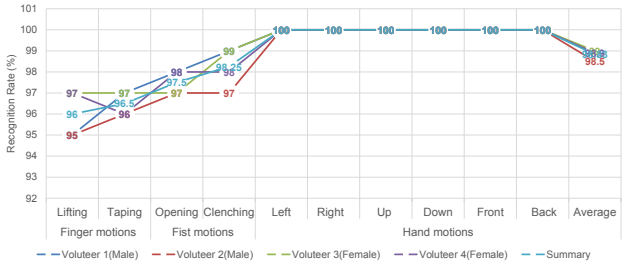

Fig. 3. The recognition result using CNN.

V. CONCLUSION

This paper has demonstrated a three dimensional hand gesture recognition system using convolutional neural networks. Temporal and frequency signatures of I/Q signals were extracted and the convolutional neural networks was used to perform recognition algorithm. This system has a better performance in recognizing the in-plane motion than flexion and extensions. In the future works, our main goal is to improve the recognition accuracy and computational speed of convolutional neural networks model.

ACKNOWLEDGEMENT

We appreciate the support from “China National Natural Science Foundation Project (Grant No. 61601141)”, “the Fundamental Research Funds for the Central Universities (Grant No. HIT.NSRIF.201817).

REFERENCES

[1] J. Schmidhuber, “Deep learning in neural networks: An overvie”, *Neural Netw.*, vol. 61, pp. 85-117, Jan. 2015.
[2] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising”, *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142-3155, July 2017.
[3] Y. LeCun, Y. Bengio, G. Hinton, “Deep learning”, *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015.
[4] J. Lien, N. Gillia , “Soli: Ubiquitous gesture sensing with millimeter wave radar”, *ACM Transactions on Graphics*, vol. 35, pp. 142, April, 2016.
[5] S. Lan, Z.He, H. Tang, K. Yao, W, Yuan, “A hand gesture recognition system based on 24GHz radars”, *International Symposium on Antennas and Propagation*, Phuket, October, 2017.