HAND GESTURE COMMUNICATION USING CONVENTIONAL NEURAL NETWORKS

Mr.A.Vignesh, Asst.Professor, Faculty of Science and Humanities, Department of Computer Science,SRM Institute of Science and Technology, Ramapuram, Chennai – 600089.

Mr.U.Udayakumar, Asst.Professor, Faculty of Science and Humanities, Department of Computer Science,SRM Institute of Science and Technology, Ramapuram, Chennai – 600089.

Abstract

Hand gesture is one of the strategy utilized in communication through signing for non-verbal correspondence which is generally utilized by hard of hearing and unable to speak individuals to impart among themselves or with typical individuals. There is a requirement for an innovation that can perceive communication through signing motion so the correspondence is conceivable even with somebody who doesn't comprehend gesture based communication, connecting the correspondence hole. The current innovation for gesture based communication discovery is sensor based information glove which has surrenders like wearing gloves constantly and skin disturbance and so on Additionally, there are in excess of 170 communications via gestures all over the planet and examination were completed generally for just a portion of the communications via gestures like american communications via gestures, argentian gesture based communication etc.,and not adequate exploration is done for our Indian communication through signing. Thus, in this venture, Indian gesture based communication characters were arranged utilizing convolutional neural network (CNN) and the restrictions in the current innovation is settled by the PC vision. For this work to be completed the dataset for communication via gestures is to be ready, preprocessed and prepared with profound convolution neural organization which assists us with ordering the signs. Also utilizing that prepared model and the PC vision procedures, the hand motion can be anticipated.

Keywords : Convolutional Neural Network, gesture, Communication

Introduction

Movement of any body part like face, hand is a type of motion. Hand motion is one of the strategy utilized in communication via gestures for non-verbal correspondence which is ordinarily utilized by not too sharp individuals to impart among themselves or with ordinary individuals. There is a requirement for an innovation that can perceive gesture based communication motion so the correspondence is conceivable even with somebody who doesn't comprehend gesture based communication, crossing over the correspondence hole. Likewise, there are in excess of 170 communications through signing all over the planet and exploration were done broadly for just a portion of the gesture based communications like American communications through signing, Argentian communication through signing and so on, and not adequate examination is completed for our Indian communication through signing. Along these lines, In this venture, Indian communication through signing characters were ordered utilizing convolutional neural organization (CNN). For this work to be done the dataset for communication through signing is to be ready, preprocessed and prepared with profound convolution neural organization which assists us with grouping the signs. Also utilizing that prepared model and the raspberry pi vision framework, the hand signal can be anticipated.

Related Work

CONVOLUTIONAL NEURAL NETWORK (CNN)

Neural network is a machine learning technique which mimics our human brain and consist of a network of learning units known as neurons. The function of the neuron is to learn how to convert the input signals (picture, sound, video etc.,) into suitable output signal (i.e., with suitable labels for the input given) which is used for recognition.

A Convolutional Neural Network (CNN) is a type of feed forward artificial neural network and they have multiple hidden layers with numerous neurons in each layer based on the application. For convolutional neural network images are used as input layer. The Convolutional Neural Network has numerous layers like convolution layer, pooling layer and fully connected layer etc.,
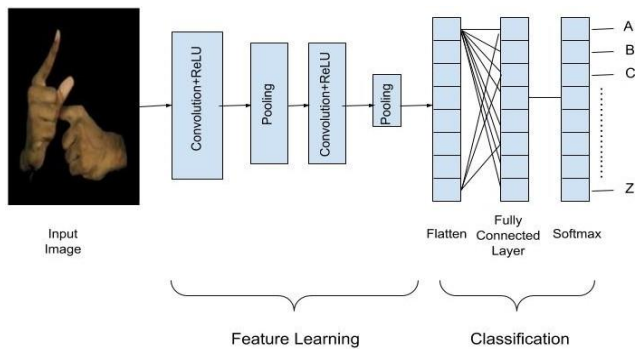

Figure 1. Convolutional Neural Network(CNN)

This is the first layer which receives input signal. It is a process where the network tries to label the input signal from the knowledge of what it had learned in the past. For example, if the input signal looked like a dog image it has seen before, the "dog" reference signal will be mixed or convolved with the input signal. The resulting output signal is then passed to the next layer which has numerous neurons. One of the nice property of Convolution is being translational invariant i.e., each convolution filter represents a feature of interest, and the CNN learns these features during the training process for different label of images given as input. The Convolution layer is the main building block of CNN and it comprises of numerous filters or feature maps. Each filter is convolved with the image and the shape of the convolved image varies depending on the kernel size of the feature

map or filter. The CNN may also consist of more than one convolution layer with different number of filters or feature maps depending upon the application. All the feature maps are initialized randomnly and during training they become our parameter.

STRIDE

Stride is the number of pixel shifts over the input matrix. For example, when the stride is 2 the filter is moves 2px over the input matrix. By default the stride is always 1px. The stride i of two types. They are,

i)   Stride along the width
ii)   Stride along the height

PADDING

Sometimes the feature detector does not fit perfectly with the input image, In that case padding can be applied. Padding is the amount of default data added to the borders of the image to maintain the output size. The padding usually has the value zero. There are four types of padding. They are,

✓ Padding height top
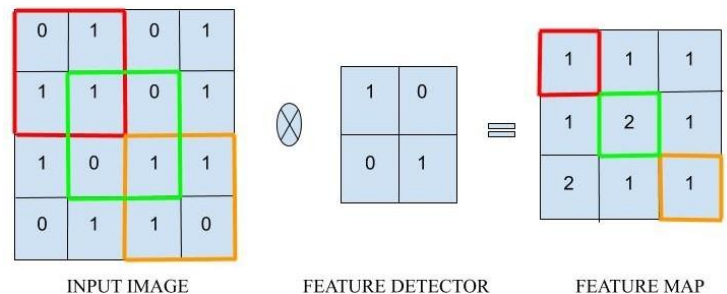✓ Padding height bottom
✓ Padding width left
✓ Padding width right

Fig 2. Convolving the image with filter/feature detector

CALCULATION OF CONVOLUTION OUTPUT SIZE

From the figure 2.3.5 we have an input image whose height is 4px and width is 4px making the shape of the image (4px * 4px). The input image is convolved with a filter or feature detector of shape (2px * 2px), resulting in the formation of feature map with the shape depending on the size of the input image, kernel size, stride height and stride width, padding height top and padding height bottom, padding width right and padding width left etc.,

The formula for finding the output size of the convolution is,

2.3.6.1 Output height =((Input height + padding height top +

padding height bottom - kernel height) / stride height) +1

2.3.6.2 Output width =((Input width + padding width right +

padding width left - kernel width) / stride width) +1

For the figure 2.3.5,the values for the parameters are,

Input height   = 4px

Input width   = 4px

Padding width right   = 0px

Padding width left   = 0px

Padding height top   = 0px

Padding height bottom   = 0px

Kernel height   = 2px

Kernel width   = 2px

Stride height   = 1px

Stride width   = 1px

Therefore,

i) Output height = ((4+0+0-2)/1)+1

= 2+1

=3

ii) Output width = ((4+0+0-2)/1)+1

= 2+1

=3

Figure 2.3.6 Convolving the image with edge filter

POOLING

The pooling layer is another important layer of CNN algorithm. The convolved images from output of the convolution layer can be "smoothened" by subsampling or Pooling technique to reduce the sensitivity of the filter subjected to noise and variation. This subsampling can be achieved by taking averages or taking the maximum over a sample of the input signal. Using subsampling technique the size of the image can be

reduced, or the colour contrast across red, green, and blue (RGB) channel can be reduced. The function of the pooling layer is to reduce the size of the image to reduce the amount of parameters and computational time in the network while training. Pooling layer acts on each feature maps or filters independently. There are many pooling techniques like mean pooling, max pooling etc.,. The most common approach used in the pooling technique is Maxpooling technique in which the maximum of a region is selected and others are neglected. In the below diagram a 2x2 region is replaced by the largest value inside that region. Also, the information loss is not more in pooling layer although it reduces the dimension of the images.

ACTIVATION LAYER

This layer controls how the signal goes from one layer to the subsequent layer by emulating the same way how neurons fire in our brain. CNN has many activation functions like Sigmoid activation, Softmax activation, Rectified Linear Activation (ReLU) function which model signal propagation. The most commonly used activation function is ReLU activation as it has faster training speed and better than the other two activation functions. The output of the ReLU activation function is,

$f(\text{x}) = \max(0,\text{x})$

CONCLUSION

To conclude the description about the project. The project developed using php and MySQL is based on the requirement specification of the user and the analysis of the existing system, with flexibility for future enhancement.Gestures are non-verbal form of communication and these gestures are understood with vision.This non-verbal communication of deaf and dumb(D&M) people is called sign language. There are more than 170 sign languages around the world and research were carried out widely for only some of the sign languages like American Sign Language, etc. Compared to other sign Languages, Sufficient research is not carried out properly for Indian sign Language and useful dataset is not available for researchers to work on it. In this project, the model is trained using convolutional Neural Network(CNN) algorithm which can recognize the ISL gestures. It is more users friendly and more GUI oriented.

REFERENCES

1. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," Proc. 2017 Int. Conf. Eng. Technol. ICET 2017, vol. 2018-Janua, no. April 2018, pp. 1–6, 2018, doi: 10.1109/ICEngTechnol.2017.8308186.

2. K. Dixit and A. S. Jalal, "Automatic Indian Sign Language recognition system," Proc. 2013 3rd IEEE Int. Adv. Comput. Conf. IACC 2013, pp. 883–887, 2013, doi: 10.1109/IAdCC.2013.6514343.

3. N. Dwina, F. Arnia, and K. Munadi, "Skin Segmentation based on Improved Thresholding Method," 2018 Int. ECTI North. Sect. Conf. Electr. Electron. Comput. Telecommun. Eng., pp. 95–99, 2018, doi: 10.1109/ECTI- NCON.2018.8378289.

4. S. Ghotkar, R. Khatal, S. Khupase, S. Asati, and M. Hadap, "Hand Gesture Recognition for Indian Sign Language," pp. 10–13, 2012.

5. H. H. S. Abdul-kareem, "Static hand gesture recognition using neural networks," no. January 2012, pp. 147–181, 2014, doi: 10.1007/s10462-011- 9303-1.

6. D. Scherer, M. Andreas, and S. Behnke, "Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition," no. September, 2010.