How the Convolutional Neural Network Work to Identify Numbers?

1st Guanlin Jiang
Canada British Columbia International Schools- Chengdu
Chengdu, China
jiangguanlin2003@outlook.com

Abstract—This document is a model and introduction for Convolutional Neural Network. Introduce the Convolutional Neural Network, Processes of Convolutional Neural Network, and application on identify objects.

Index Terms—introduce, processes, application

I. INTRODUCTION

Convolutional Neural Network is a kind of Deep Neural Networks that work in Machine Learning. The most common fields to be used like computer vision, natural language processing, and analysis the visual image information. Nowadays, the Convolutional Neural Network is the most popular Deep Learning Neural Network in this world. The reason we choose the Convolutional Neural Network is that I am interested in it, because this is the so powerful, and Convolutional Neural Network is using most popular Neural Network in Machine Learning. Another reason is I want to use this Convolutional Neural Network to make a identify system use in life. This paper will include the introduction of Convolutional Neural Network, How the Convolutional Neural Network process, and Convolutional Neural Network application on identifying objects.

II. MACHINE LEARNING AND CONVOLUTIONAL NEURAL NETWORK

A. What Is Machine Learning

Machine learning is a part of Artificial intelligence which is “a study of computer algorithms that use data and neural network to let machine learning and improve their accuracy” [1]. The dataset and algorithms are the important elements for Machine Learning. They use algorithms to build the neural network which means build the structures. Then the dataset will be label and classify and put into the neural network to train the model which means the algorithms use data to learn, and against their own, practicing the learning and get the max score. Then, researchers can use the model on the raspberry pi, or other electronic devices, or they can add more datasets to training, and improve the accuracy of the models.

The application of machine learning includes computer vision, natural language processing, and information analysis. For example, when a company needs to install a face scanner to help manage employees, they need to collect the employee’s face photos to be the datasets. Then they are loading those face data to the computer, and they section off the data. Then programming the algorithms for the Neural Network and use that to identify patterns which are to train the model. The program of Neural Network can assess and output the results, so people can use the resulting model to run in electronic device. The machine learning is like people to think the things, but computer vision is like to envision the things can shows in people’s brain.

B. What Is Computer Vision

When a photo scan or take a photo by computer, the display will be showing the photo, and perform image classification, those processes are a part of computer vision. The image or vision will receive by the computer as the input data. The photo data will be translated to the binary language, which is a language computer can understand, “They first detect pixels, then edges and contours, then whole objects, before eventually producing a final guess about what they’re looking at.” [2]. After scanning it, the computer will guess what it is. The computer vision converts the image into a two-dimensional matrix format, which display to computer is the binary language. Also, the computer vision based on the Neural Network to accurately identify and classify images, which relies on the models trained by the computer neural network. Like Figure 1, the image will be divided into many blocks, and each small block will correspond to many two-dimensional matrices. So, this is what computer see.

Fig. 1. Example of Computer Vision.

Nowadays, computer vision application is in facial recognition, and image analysis software, including catching criminals and take photos. For example, when people open their computer which can use face to unlock, people put the computer
in front of their face, then the camera will be scanning the face and transfer it to binary language, the computer will be analyzing the data by the model and comparing the owner’s face data to check and identify is or not owner’s face.

C. What Is Convolutional Neural Network

“The Convolutional Neural Network is a kind of deep learning neural network which designed for processing structured arrays of data such as images, video, and neural text” [3], which is researched in 1980 by Yann LeCun. The object recognition of the Convolutional Neural Network model imitates the visual processing mechanism of the human brain, each neuron on every layer is connect to next layers neurons, which means the whole network use the layered features. Comparing with “the Back Propagation Model find in 1960, which is a method of fine-tuning the weights of a neural net based on the error rate obtained in the previous epoch” [4]. Convolutional Neural Network comparing with the Back Propagation Model way, which is reduce the number of processes for computers to run, and more computing can use in add the accuracy. “After coming out Convolutional Neural Network, the domain of Computer Vision which make amazing things happen, the training model’s accuracy rate is greatly increased, and the error rate is greatly reduced” [4]. The application of Convolutional Neural Network use on the training model which is a process on Machine Learning. AI scientists will build a Neural Network to train their model more accurately and easy to be building and reading the code. The Convolutional Neural Network is the most common for scientists professional people use because of the high performance and easy to build.

III. PROCESSES OF CONVOLUTIONAL NEURAL NETWORK

A. Convolutional Layer

The Convolutional Layer is the first layer of a Convolutional Neural Network. “The convolution layer (CONV) uses filters that perform convolution operations” [5]. This means the convolution layer use the filters to get the useful information for whole network. “Input data like photos, Convolutional Layer will output some images of size 28x28 transfer to the next layer, which normally identifies basic features such as straight edges and corners” [3]. This layer includes input data, a filter to filter out some noise like the in a animal photo there have a gun in there, also get the important data, and a map of connection the layers. “When an image which made up of a matrix of pixels in 3D, and input to the Convolutional Neural Network, computer understanding the RGB, and the image build by three dimensions, which include a height, width, and depth” [1]. Like Figure 2, the image will be divided into many blocks, and each small block will correspond to many two-dimensional matrices. After, the Convolutional Layer will be cut the image to many pieces, and transfer to RGB (which is the pixel of Red, Green, and Blue in computer) of each small image to the binary language, which is display 0 and 1 in the layer, like Figure 1. “While they can vary in size, the filter size is typically a 3x3 matrix. The filter is then applied to an area of the image” [1]. This means the output matrix is always 3 × 3 matrices, the computer will be calculating every point by the filter and pixels, and those matrices will be transferred to the next layer. The Convolutional Layer is a layer to analyze the input data and convert the image to the number which the computer can understand, help other layers to use the data to training and building a model.

\[
\begin{pmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 1
\end{pmatrix}
\]

Fig. 2. Example of Matrix.

B. Pooling Layer

The Pooling Layer are dividing into two layers, one is Max Pooling, another is Average Pooling. “The functions of Pooling Layer are conducted dimensionality reduction, reducing the number of parameters in the input” [1]. This shows the Pooling layer reducing the size of the image, which means the Neural Network can reduce some difficulties of calculating. Also, this layer helps the Machine Learning to summarize all the data from Convolutional Layer, and collect features of the image to processing, which help the model more accuracy when using in identify objects [6]. The Pooling layer is the second layer, also like a summarize layer to decrease the difficulties of computing and make the Convolutional Neural Network clearer.

1) Max Pooling: Max Pooling is choosing the most special data like the most feature point in an image from summarize information and output a map which has the most characteristic parameter from every matrix [6]. Like Figure 3, Matrix 1 after Max Pooling become Matrix 2. The Max Pooling is to optimization the data and chooses the high feature parameters to train, and the model output by training use in Machine, the process will be to use the test images to identify will focus on the most feature or special part to identify.

2) Average Pooling: The Average Pooling is a process to calculate every matrix average parameter and output a new matrix transfer to the next layer. The point is the Max Pooling just choose the highest feature one, but the “Average Pooling blends them in” which is using all the parameters which are 2 × 2 matrixes to get the average value [7]. Like Figure 4, Matrix 3 after Average Pooling, the output will be Matrix 4. The Average Pooling is just using the average value of each matrix to identify the objects, if they can have similar value in the calculating.

\[
\begin{pmatrix}
6 & 2 & 4 \\
9 & 6 & 5 \\
4 & 2 & 7
\end{pmatrix}
\quad \rightarrow \quad
\begin{pmatrix}
9 & 6 \\
9 & 7
\end{pmatrix}
\]

Fig. 3. Example of Max Pooling.

Fig. 4. Example of Average Pooling.
C. Fully-connected (FC) Layer

“Fully-connected Layer is multiplying the input parameters by a matrix with the weight, then adds a bias vector” which is the addition of the neurons’ weights in the neural network that do not need the input parameters [8]. In the Fully-connected Layer, input is connected to all neurons. The functions of neurons are to help the data calculating parameters to connect each point and transfer the calculating result. “The full connected layers which compile the data extracted by last layers to form the final output” [9]. The Fully-connected Layer has three sub-layers—Input Layer, Hidden Layer, and Output Layer. “In those layers, there have many units, like the feature units, activation units, and bias units” [9]. The Fully-connected Layer will be processing the weight and class recognition by those layers and units. The input layer will get the parameters, and the hidden layers calculate the parameters with weight and bias vector, after the calculation, the result will be transferred to the output layer to be the result. The work of class recognition are the neurons in this layer will catch a certain feature, then the Network needs to storage the value of each feature, and give the class to the values, also will check it is right or not [10]. After output layer, the output model file will storage all the network, when people put model file into other device, they do not need to be training again, just use is ok. The Fully-connected Layer basically is to classify the value and add the weight to each value to make sure the output model can identify accurately.

IV. USING CONVOLUTIONAL NEURAL NETWORK TO IDENTIFY OBJECTS

A. Loading Data

Nowadays, the use of Convolutional Neural Network to get the Machine Learning model is common thing, because of the clear reading and easy building, data researchers and academic people always use this framework to train the data and build the model. I use Convolutional Neural Network to identify numbers to be the example to explain how. In here, we use MNIST dataset which is a dataset have many numbers write by other people but offer by Keras which is a Deep Learning framework to be the dataset as the input data. We use Python3 to be the computer programming language, and Jupyter Notebook to be the Python3 Editor based on the web. Importing the Keras deep learning framework, Keras MNIST dataset, and the Numpy which is a calculating part for Python. The size of the image needs to be set, 28 x 28 x 1 is the input shape we need. The x-train, y-train, x-test, and y-test need to be setting when we start to load the MNIST dataset. “The x-train and x-test are the features you are using as input for the model of train data and test data, y-train and y-test are the expected labels of train data and test data” [11].

B. Build the Model Training Data

In this process, Keras be the Convolutional Neural Network builder framework. For this MNIST Convolutional Neural Network, I choose to build two groups of layers in this training, which means the data will be run two times. Building the two-dimensional Convolutional Layers’ filters, parameters set are 32 hidden units and 64 hidden units, also using “relu” which is a non-linear activation function can calculate, it to be the activation units and using two times two-dimensional Max Pooling to choose the most feature data to summarize. After, choosing the “softmax” which is a function that can output of a probability distribution, it to be the activation units, and output it. In this process, we need to build the Model of the Convolutional Neural Network and decide the layers want to set. After the Convolutional Neural Network model built, the input data can be training. Setting the epoch which is how many times people want the Network training and setting the accuracy and loss for training and test. When the program running, after training, the model file will be output. “The trained model needs to be evaluated in terms of performance” [11]. If the performance reach to more than 99 percentages, which means the model is trained well for prediction. So, this file people can put loading in the programs, and use it in any electric devices.

V. CONCLUSION

The Convolutional Neural Network is a part of the Machine Learning framework, and its application in Computer Visions and Neural Language Processing. This Network includes 3 layers—Convolutional Layer, Pooling Layer, and Fully-connected Layer which is also following the processes. Using Convolutional Neural Network to identify numbers to be an example of how to build algorithms and train Machine Learning model. This research paper which is basic information for my Personal Interesting Project offers the basic knowledge of my project—Identify the Numbers using Convolutional Neural Network to training, which can help me know the basic knowledge for Convolutional Neural Network method. Also, this paper can help me to organize my Convolutional Neural Network, not miss anything and steps. Like, first need to find
dataset and load it, second to build the layer and let data transfer to computer language, last is training it. After, we can use the model to test or in the real-life applications.

REFERENCES

[1] By: IBM Cloud Education, “What is Machine Learning?,” IBM. [Online]. Available: https://www.ibm.com/cloud/learn/machine-learning. [Accessed: 25-Mar-2021].

[2] K. Hartnett and substantive Quanta Magazine moderates comments to facilitate an informed, "Computers and Humans 'See' Differently. Does It Matter?,” Quanta Magazine. [Online]. Available: https://www.quantamagazine.org/computers-and-humans-see-differently-does-it-matter-20190917. [Accessed: 25-Mar-2021].

[3] T. Wood, “Convolutional Neural Network,” DeepAI, 17-May-2019. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network. [Accessed: 25-Mar-2021].

[4] K. D. Foote, “A Brief History of Deep Learning,” DATAVERSITY, 31-Jan-2017. [Online]. Available: https://www.dataversity.net/brief-history-deep-learning/. [Accessed: 25-Mar-2021].

[5] A. Amidi and S. Amidi, “Convolutional Neural Networks cheatsheet Star,” CS 230 - Convolutional Neural Networks Cheatsheet. [Online]. Available: https://stanford.edu/ shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks. [Accessed: 25-Mar-2021].

[6] S. Hosla, “CNN: Introduction to Pooling Layer,” GeeksforGeeks, 26-Aug-2019. [Online]. Available: https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/. [Accessed: 25-Mar-2021].

[7] Christian, “What are Max Pooling, Average Pooling, Global Max Pooling and Global Average Pooling?” MachineCurve, 30-Jan-2020. [Online]. Available: https://www.machinecurve.com/index.php/2020/01/30/what-are-max-pooling-average-pooling-global-max-pooling-and-global-average-pooling/. [Accessed: 25-Mar-2021].

[8] Mathwork, “fullyConnectedLayer,” Fully connected layer MATLAB, 2021. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.fullyconnectedlayer.html. [Accessed: 25-Mar-2021].

[9] S. P. Singh, “Fully Connected Layer: The brute force layer of a Machine Learning model,” OpenGenus IQ: Learn Computer Science, 02-Mar-2019. [Online]. Available: https://iq.opengenus.org/fully-connected-layer/. [Accessed: 25-Mar-2021].

[10] SuperDataScience Team, SuperDataScience, 18-Aug-2018. [Online]. Available: https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-4-full-connection. [Accessed: 25-Mar-2021].

[11] S. Mahapatra, “A simple 2D CNN for MNIST digit recognition,” 22-May-2018. [Online]. Available: https://towardsdatascience.com/a-simple-2d-cnn-for-mnist-digit-recognition-a998db1e579a. [Accessed: 09-Apr-2021].