International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Implementing Convolutional Neural Networks for Simple Image Classification

Aditeya Nanda, Praveen Kumar, Seema Rawat

Abstract*— In recent years, huge amounts of data in form of images has been efficiently created and accumulated at extraordinary rates. This huge amount of data that has high volume and velocity has presented us with the problem of coming up with practical and effective ways to classify it for analysis. Existing classification systems can never fulfil the demand and the difficulties of accurately classifying such data. In this paper, we built a Convolutional Neural Network (CNN) which is one of the most powerful and popular machine learning tools used in image recognition systems for classifying images from one of the widely used image datasets CIFAR-10. This paper also gives a thorough overview of the working of our CNN architecture with its parameters and difficulties.

Keywords*— Convolutional Neural Network, image recognition, CIFAR-10, machine learning.

I. INTRODUCTION

With the advent of the social media age the amount of digital data being churned out every minute has increased many folds over the years. The estimated amount of this data has always been fluctuating. Conventional tools are unable to process or store these vast datasets that have turn out to be immeasurable. These data sets could be as large as maybe numerous terabytes. This brings into existence a critical problem of automatic classification of these images for further processing. While it is very easy for a human to identify images and distinguish them from one another, giving same abilities to a computer has proven to be a difficult task.

Artificial Intelligence is a field that aims to give computers similar processing capabilities as of humans. And Computer Vision is a sub field of Artificial Intelligence that aims at providing computers the tools to classify and analyze images. But classification of images has turned out to be a very demanding task, as humans can use their visual senses to quickly identify objects within images but for a computer an image consists of various pixels and it turns out to be a difficult task to extract all features from the image and identify it. Much research has been done in this area without much fruition. But in recent years neural networks have turned out to be a powerful machine learning tool for classification and identification of digital images.

The working of a neural network systems is inspired by the design of the brain. Human brains have innumerable neurons connected to each other in different stratum, sending information to and fro. Similarly, a neural network system constitutes of a multitude of neurons[6] connected to different layers processing data in an accurately manner. These networks are first exposed to a training data set from where the system performs feature extraction and 'learn' about the image and are then tested against a testing data set. The neural networks can be Artificial Neural Networks (ANN) or Convolutional Neural Networks (CNN). CNNs are used for the task of image recognition.

This project aims at creating a Convolutional Neural Network architecture for image recognition. The network will be trained and tested on the CIFAR-10 dataset.

A Convolutional Neural Network comprises of various layers each equipped with performing a particular task. These layers are Input Layer, Convolution Layer, Activation Function Layer, Pool Layer and finally the Fully-Connected Layer. Initially the raw image is input in the network. The convolution layer then computes a dot product of the filters and the image. This dot product is the output volume of the image. The output of this layer is input of activation function layer which apply activation function to this data. Sigmoid, Tanh, RELU etc. are some of the activation functions. RELU is one of the most popular and widely used activation function. The output volume of the data is not hindered after applying the activation function. As the output volume is still vast the computation speed remains very slow and memory consumption is also substantial, thus pooling layer is used. The pooling layers minimizes the output[2] volume and prevents the data from overfitting. Max pooling and average pooling are common types of pooling layers. Finally, the fully-connected layer computes the output in a one-dimensional array. The number of classes to be identified is equivalent to the size of this array.

The Canadian Institute for Advanced Research created the CIFAR-10 dataset. It is a universally used collection of images for machine learning and computer vision tasks. The CIFAR-10 dataset contains 60,000 color images. The size of the images is 32x32. There are total 10 different classes with 6,000 images each in which the images are divided. some of these classes are cars, cats, dogs, horses, ships, frogs etc. The CIFAR-10 dataset is widely used as it consists of many tiny images, thus various systems can run on this dataset quickly as the size of the mages decreases the computation size dramatically.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
Praveen Kumar*, Department of Computer Science & Engineering, Amity University Uttar Pradesh, Noida, India. Email: pkumar3@amity.edu
Aditeya Nanda, Department of Computer Science & Engineering, Amity University Uttar Pradesh, Noida, India, Email: adityananda@gmail.com
Seema Rawat, Department of Department of Information Technology, Amity University Uttar Pradesh, Noida, India. Email: srawat1@amity.edu

Image recognition has wide array of applications ranging from Facial Recognition System to Contextual image classification, reverse image search, machine vision etc.

II. Approach Used

A. Network Architecture

A convolutional neural network is an intricate network of multiple nodes with various layers in between that process the data before the final layer of nodes give the output. Every neural network has five variety of layers.

The input layer takes all the input from different classes and passes it on to the next layer for processing. The number of nodes in an input layer depend on the number of input parameters.

The convolutional layer takes the output from the input layer as input in the input channels parameter. The output of the convolutional layer is defined in the output channels parameter. Convolution function is defined derived from two given functions using integration and it tells how one function is affected by the other function. Input image, Feature detector and Feature map are the elements that are enter into the convolution operation. A 3x3 matrix is one of the widely used size for a feature detector. A feature detector can also be called a "kernel" or a "filter,".

The max pooling layer allows the network to detect features in an image when presented in any manner. Max pooling is one of the widely used pooling techniques which is also used in our network. Other popular pooling techniques are Mean pooling and Sum pooling.

The fully connected layer takes data from previous layers and turn the data into a one-dimensional array which can be entered into an artificial neural network.[9] The artificial neural network combines different features into a single attribute thus helping the convolutional neural network to better identify images. And the output layer finally gives the output of what class the image belongs to. The number of nodes in output layer depends on number of classes the image can be classified into.

The dataset used in this project contains 60,000 tiny images of size 3x32x32. The images are of size 32x32 pixels with 3 channels i.e. red, blue, green.

The 2nd layer in the network is a convolutional layer with 3 input channels, 32 output channels with kernel size of 3. This layer has padding set to 1 and also stride set to 1. This will convolve 32 filters each of size 3x3x3. As [(32−3 +2(1))/1]+1=32, thus the output size of this layer will be 32x32x32 with ((3×3×3)+1)×32=896 parameters.

The 3rd layer is a down sampling layer. This layer uses max pooling with a kernel size of 2x2. Thus, the output size from previous layer drops from 32x32x32 to 32x16x16.

The next layer is also a convolutional layer with 32 input channels, 64 output channels with kernel size of 3. This layer has padding set to 1 and also stride set to 1. This will convolve 64 filters each of size 32x3x3. As [(16−3 +2(1))/1]+1=16, thus the output size of this layer will be 64x16x16 with ((3×3×64)+1)×32=18464 parameters.

The 5th layer is again a down sampling layer. This layer also uses max pooling with a kernel size of 2x2. Thus, the output size from previous layer drops from 64x16x16 to 64x8x8.

The 6th layer is a convolutional layer with 64 input channels, 64 output channels with kernel size of 3. This layer has padding set to 1 and also stride set to 1. This will convolve 64 filters each of size 64x3x3. As [(8−3 +2(1))/1]+1=8, thus the output size of this layer will be 64x8x8 with ((3×3×64)+1)×64=36,928 parameters.

The next layer is the final max pooling layer. Here, the output size from previous convolutional layer drops from 64x8x8 to 64x4x4. The ouput of this layer needs to be flattened because the output of this layer will be used as input for a fully connected layer. The first fully connected layer has the ReLU activation function with 500 nodes. In total this layer needs ((64×4×4)+1)×500=512,500 parameters.

The last layer is a fully-connected layer with a total of ten nodes. Here each node represents a class from the CIFAR-10 dataset. This layer requires (500+1)×10=5010 parameters.

Thus, the convolutional neural network created for the project requires a total of 896+18464+36928+512500+5010=573798 parameters.

TABLE I. Neural Network Model

	Layers
Layer 1	Conv1[3x32x3]
Layer 2	Pool[2x2]
Layer 3	Conv1[32x64x3]
Layer 4	Pool[2x2]
Layer 5	Conv1[64x64x3]
Layer 6	Pool[2x2]
Layer 7	Fc[500]
Layer 8	Fc[10]

B. Dataset

In this project CIFAR-10 dataset has been used. It was created by the Canadian Institute for Advanced Research. The CIFAR-10 dataset contains 60,000 color images with a size of 32x32belonging to a total 10 different [10][11]classes with 6,000 images in each class. some of these classes are cars, cats, dogs, horses, ships, frogs etc. It is a universally used collection of images for machine learning and computer vision tasks.

MNIST is other universally used dataset that is widely used for computer vision systems. MNIST dataset can also be used with the network stated in this paper. This dataset contains images of dimension 28x28x1 in only black and white color.

C. Experiment

This paper indicates the working of a convolutional neural network. The main objective of the project is to create a network that can learn the different features of the images and can accurately[6] accurate images from the test set. The initial version of the program had an accuracy of 52%. Many steps like Data augmentation, increasing the number of layers, changing learning rate etc. were taken to increase the accuracy of the network.

The table below shows the number of layers of initial models and the final models and the accuracy we got with each layer.

Table II. Effect on accuracy by changing network parameters

	Network - 1	*Network - 2*	*Network - 3*
Layer 1	Conv1[3x6x5]	Conv1[3x16x3]	Conv1[3x32x3]
Layer 2	Pool[2x2]	Pool[2x2]	Pool[2x2]
Layer 3	Conv1[6x16x5]	Conv1[16x32x3]	Conv1[32x64x3]
Layer 4	Pool[2x2]	Pool[2x2]	Pool[2x2]
Layer 5	Fc[120]	Conv1[32x64x3]	Conv1[64x64x3]
Layer 6	Fc[84]	Pool[2x2]	Pool[2x2]
Layer 7	Fc[10]	Fc[500]	Fc[500]
Layer 8		Fc[10]	Fc[10]
Accuracy	52%	70%	75%

The algorithm below states all the general steps for image classification in neural networks used in training and testing of the CIFAR-10 data set:

1. Batch size = 4, classes = 10, number of epochs = 20
2. Size of the image is 32 ×32 pixel
3. Load the CIFAR-10 data set
4. Testing dataset has 10000 images of size 32x32 pixel and Training dataset has 50000 images of size 32x32 pixel
5. Create the Model
6. Compile the Model
7. Train the network.

For this project Pytorch library was used for creating the machine learning model. This library has been developed by the artificial intelligent research group at Facebook. It is an open source library with various applications like natural language processing, deep learning etc. After importing the libraries, the data set is downloaded[7] and loaded onto the system. But before loading the system some artificial changes are made to the dataset to increase the accuracy and the learning of the model. The total size of the data is exponentially increased by randomly flipping or changing the brightness of the dataset. This also prevents the over-fitting of the model. Then the images are converted from Python Image Library (PIL) format to PyTorch tensors. Then normalization of the data takes place.

The dataset is then divided into training and testing dataset and loaded onto the system. The data goes through all the different layers of the convolutional neural network for processing. In this model cross-entropy is used as loss function. A loss function is a very significant part of a neural network as it is used to measure the difference between the actual value from the dataset and the predicted value[8] from the model. Stochastic gradient descent is used in the model to optimize the network. The network is also employed with a momentum of 0.9 and a learning rate of 0.001.

The network goes over the complete training data in batches of 4 images. This process is repeated for 20 epochs. After every 2000 batches of the data, the running loss value and the current epoch is declared. After the training is completed the total accuracy of the network is checked against the testing dataset.

III. RESULT

The initial version of the program had an accuracy of 52%. Many steps like Data augmentation, increasing the number of layers, changing learning rate etc. were taken to increase the accuracy of the network. The final version of the model had total accuracy of 75% with individual accuracy of individual classes as mentioned below.

TABLE III. Accuracy Of The Model

Class	*Accuracy*
Plane	80%
Car	90%
Bird	68%
Cat	57%
Deer	71%
Dog	61%
Frog	79%
Horse	79%
Ship	81%
Truck	84%
Total Accuracy	**75%**

As mentioned previously the dataset contains images belonging to a total 10 different classes. In each class the images are divided into training and testing dataset. The testing dataset of each class is compared with the predictions of each class by the model to find accuracy of individual classes.

IV. CONCLUSION

In this paper, we discussed the construction of a convolutional neural network with a brief account of working of the network for image classification. Cifar-10 dataset was used for the purpose of training and testing the machine learning model. A total accuracy of 75% was achieved with our model, with accuracy of individual classes as mentioned in the table above.

More accurate results can be obtained by increasing the number of layers in the model, increasing the amount of images in the training set, increasing epochs, using multiple GPU etc.

REFERENCES

1. Tapan Bhavsar, Bhavinkumar Gajjar "Image Classification using Convolution Neural Network", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-6 Issue-5, June 2017
2. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton "ImageNet Classification with Deep Convolutional Neural Networks", COMMUNICATIONS OF THE ACM, JUNE 2017
3. Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks."European conference on computer vision. Springer International Publishing, 2014.
4. Gu, Jiuxiang, et al. "Recent advances in convolutional neural networks."arXiv preprint arXiv:1512.07108 (2015).
5. Michael Blot, Matthieu Cord, Nicolas Thome. Max-min convolutional neural networks for image classification. ICIP 2016 - IEEE International Conference on Image Processing, Sep 2016, Phoenix, United States.
6. J. Redmon, A. Angelova, "Real-time grasp detection using convolutional neural networks", *IEEE International Conference on Robotics and Automation*, pp. 1316-1322, 2015.
7. X. Zhou, K. Yu, T. Zhang, T. Huang, "Image classification using super-vector coding of local image descriptors", *ECCV*, 2010.
8. K. E. A. Van de Sande, T. Gevers, C. G. M Snoek, "Evaluating color descriptors for object and scene recognition", *IEEE Transactions on Pattern Analysisand Machine Intelligence*, pp. 1582-1596, 2010.
9. T. Ahonen, A. Hadid, Pietikinen, "M. Face description with local binary patterns: Application to face recognition", *Pattern Analysis and Machine Intelligence*, pp. 2037-2041, 2016.
10. Chaturvedi, A., Kumar, P. and Rawat, S., 2016, October. Proposed noval security system based on passive infrared sensor. In 2016 International Conference on Information Technology (InCITe)-The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds (pp. 44-47). IEEE.
11. Kumar, P., Sanchita Kadambari, and S. rawat. "Prefetching web pages for improving user access latency using integrated Web Usage Mining." In 2015 Communication, Control and Intelligent Systems (CCIS), pp. 401-405. IEEE, 2015.

AUTHORS PROFILE

Mr. Aditeya Nanda, is a student at Amity School of Engineering Technology, Amity University, Uttar Pradesh. Leveraging technology for a better tomorrow is his area of interest. His area of research includes convolutional neural network and python deep learning.

Dr. Praveen Kumar, is working as Associate Professor at Amity School of Engineering & Technology. He is M.Tech in Computer Science & Engineering. He has a number of international and national publications to his credit. He is a lifetime member of IETE, ACM, and IET. His primary research area includes Big Data Analytics, Cloud Computing and Data mining.

.

Dr. Seema Rawat, is working as Assistant Professor at Amity School of Engineering & Technology. She is PhD and M.Tech in Computer Science and B.Tech in Information Technology. She has a number of international and national publications to her credit. She is a member of IEEE, IACSIT and IAENG. Her primary research area includes Cloud Computing, Data mining and Artifi cial Intelligence.