A Project Report on

Image Classification using Convolutional Neural Networks

Computer Engineering

by

Mrunal S. Jadhav(16102030),Rohan P. Dhere(15102057)

Under the Guidance of

Prof. Sofiya Mujawar

Department of Computer Engineering
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2019-2020

Approval Sheet

This Project Report entitled Image Classification using Convolutional Neural Networks Submitted by Mrunal S. Jadhav(16102030),Rohan P. Dhere(15102057) is approved for the partial fulfillment of the requirenment for Semester VI in Computer Engineering, University of Mumbai.

(Prof. Sofiya Mujawar)
Guide

Prof. Sachin Malave
Head Department of Computer Engineering

Place:A.P.Shah Institute of Technology, Thane
Date: April 11, 2019

CERTIFICATE

This is to certify that the project entitled "Image Classification using Convolutional Neural Networks" submitted by Mrunal S Jadhav(16102030),Rohan Dhere(15102057) for the fulfillment of the requirement for Semester VI in Computer Engineering, University of Mumbai,is a bonafide work carried out during academic year 2019-2020.

(Prof. Sofiya Mujawar)
Guide

Prof. Sachin Malave
Head Department of Computer Science

Dr. Uttam D. Kolekar
Principal

External Examiner(s)

1.

2.

Place:A.P.Shah Institute of Technology, Thane
Date: April 11, 2019

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

——————————————————

(Signature)

——————————————————

(Mrunal Jadhav and 16102030)
(Rohan Dhere and 15102057)

Date: April 11,2019

Abstract

Image classification has become one of the key pilot use-cases for demonstrating machine learning.Image classification refers to the task of extracting information classes from a multiband raster image. The resulting raster from image classification can be used to create thematic maps.The objective of the project is to use “flavor” of logistic regression to tackle the multi-class classification problem. Embedding is a way to map discrete objects (images, words, etc.) to high dimensional vectors. Thus taking advantage of overall patterns of location and distance between vectors , a classifier can be trained on the same.Further, making effective use of multiple features of data, a neural network can be adapted as a baseline model ,which is a data-driven,visually aware feature extractor.Also evaluating the network under various conditions that affect the accuracy.Also ,to exploit the idea that the local understanding of image is good enough using CNN as fewer parameters may greatly improve the time it taken to learn as well as reduces the amount of data required to train the model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computer vision is moving from predicting discrete,categorical labels to generating rich descriptions of visual data, in particular, in the form of natural language.We are witnessing a surge of interest in tasks that involve cross-modal learning from image widely viewed as the “next frontier” of scene understanding[3].

In recent years, CNN has been used for various purposes. Especially, it has been widely used for image recognition and it shows good performance when applied to two-dimensional data such as images. [4]Image Classification in classic problem in computer vision. Given a image ,the machine has to identify the class of the image. The model predicts the class of the matrix from multiple classes. Image classification is classic problem in machine learning due to its various applications and uses in industry. In addition, modern social media and photo sharing/storage applications like Facebook and Google Photos use image classification to improve and personalize user experience on their products.[1]

During the rise of deep learning, feature extraction and classifier has been integrated to a learning framework which overcomes the traditional method of feature selection difficulties. The idea of deep learning is to discover multiple levels of representation, with the hope that high-level features represent more abstract semantics of the data. One key ingredient of deep learning in image classification is the use of Convolutional architectures.

The image classification problem is also representative of many common challenges in computer vision such as intraclass variation, occlusion, deformation, scale variation, viewpoint variation, and illumination. Methods that work well for image classification are likely to translate to methods that will improve other key computer vision tasks as well, such as detection, localization, and segmentation. [2]

A prime example of this is image captioning. The image captioning problem is to, given an image, output a sentence description of the image. The image captioning problem is similar to the image classification problem, but more detail is expected and the universe of possibilities is larger. Image classification is also very useful as a subcomponent for image search engines. Finally, image classification systems could become a crucial component of accessibility software in the future, helping humans with vision problems draw meaning from their surroundings.[2]

Chapter 2

Literature Review

Sr No	Title	Author	Abstract
1	Convolutional Neural Networks for Image Classification and Captioning	Sachin Padmanabhan	The paper explores the use of convolutional neu- ral networks (CNNs) for the image classification and image captioning problems.
2	Simple Convolutional Neural Network on Image Classification	Tianmei Guo, Jiwen Dong ,Henjian Li 'Yunxing Gao	On the basis of the Convolutional neural network, the paper analyzes different methods of learning rate set and different optimization algorithm of solving the optimal parameters of the influence on image classification.
3	Learning Two-Branch Neural Networks for Image-Text Matching Tasks	Liwei Wang, Yin Li, Jing Huang, Svetlana Lazebnik	The paper investigates two-branch neural networks for learning the similarity between these two data modalities.

Chapter 3

Building the Foundation: Logistic Regression

Scope

The scope of our project is to construct a Convolutional Neural Network(CNN) which provides high class accuracy, as compared to premium quality CNN like VGGNet and AlexNet models. This CNN will be built upon concrete facts and pure facts based research.

3.1 Logistic Regression, 1st building block

3.1.1 Getting Familiar

In statistics, the **logistic model** (or **logit model**) is a widely used statistical model that, in its basic form, uses a logistic function to model a binary dependent variable.In regression analysis, **logistic regression** (or **logit regression**) is estimating the parameters of a logistic model; it is a form of binomial regression. Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail, win/lose or healthy/sick.

Logistic regression was developed by statistician David Cox in 1958. The binary logistic regression model has extensions to more than two levels of the dependent variable: categorical outputs with more than two values are modeled by multinomial logistic regression, and if the multiple categories are ordered, by ordinal logistic regression.

In our project we are going to be dealing with multinomial logistic regression which is a classification method that generalizes logistic regression to multiclass problems, i.e. with more than two possible discrete outcomes.That is, it is a model that is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables.

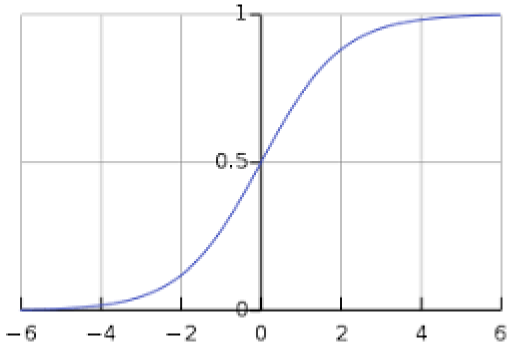
3.1.2 Use of Cost Function: Sigmoid Function

The cost function represents optimization objective i.e. we create a cost function and minimize it so that we can develop an accurate model with minimum error.

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2.$$

Cost Function for Linear regression

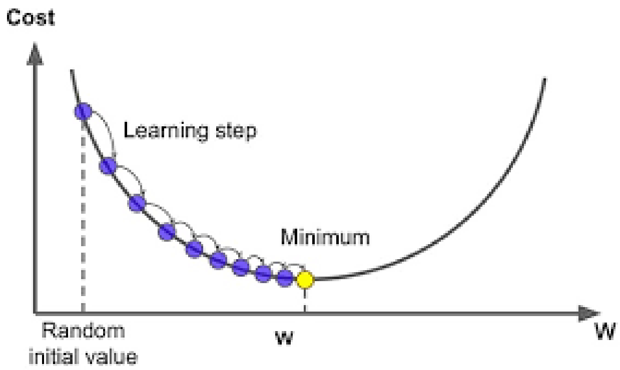
In order to map predicted values to probabilities, we use the Sigmoid function. The function maps any real value into another value between 0 and 1. In machine learning, we use sigmoid to map predictions to probabilities.



Sigmoid Function

3.1.3 Gradient Descent

The most basic and vastly used optimisation technique to minimise the above function is **Gradient Descent.** It is an iterative optimisation algorithm to find the minimum of a function. To find the local minimum using gradient descent, steps proportional to the negative of the gradient of the function at the current point are taken. If taken in the positive direction, the algorithm finds local maximum and this process is called as **Gradient Ascent.**



Gradient Descent

3.2 Working with Artificial Neural Networks, 2nd Building Block

3.2.1 Getting familiar

A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes. Thus a neural network is either a biological neural network, made up of real biological neurons, or an artificial neural network, for solving artificial intelligence (AI) problems.

The concept of a neural network appears to have first been proposed by Alan Turing in his 1948 paper Intelligent Machinery.

Neural networks take a different approach to problem solving than that of conventional computers. Conventional computers use an algorithmic approach i.e. the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem. That restricts the problem solving capability of conventional computers to problems that we already understand and know how to solve.

3.2.2 Neurons

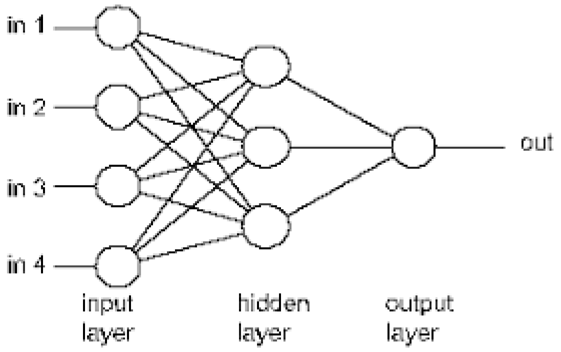
An artificial neuron is a device with many inputs and one output. The neuron has two modes of operation; the training mode and the using mode. In the training mode, the neuron can be trained to fire (or not), for particular input patterns. In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not.

Graphical representation of Neuron

3.2.3 Architecture of Artificial Neural Network

Artificial neural networks are composed of a set of neurons, joined together by synapses. Neurons perform a simple computational task, generally a basic yes/no decision. Synapses link neurons together by linking their inputs and outputs.

In programming terms, a synapse is an object which links one neuron connected to its input to another connected to its output. A neuron is a slightly more complex object which can be connected to one or more input synapses and one or more output synapses. The structure of any neural network is therefore defined by the way in which various neurons and synapses are linked together.



Basic Architecture of ANN

Chapter 4

Building Convolutional Neural Network

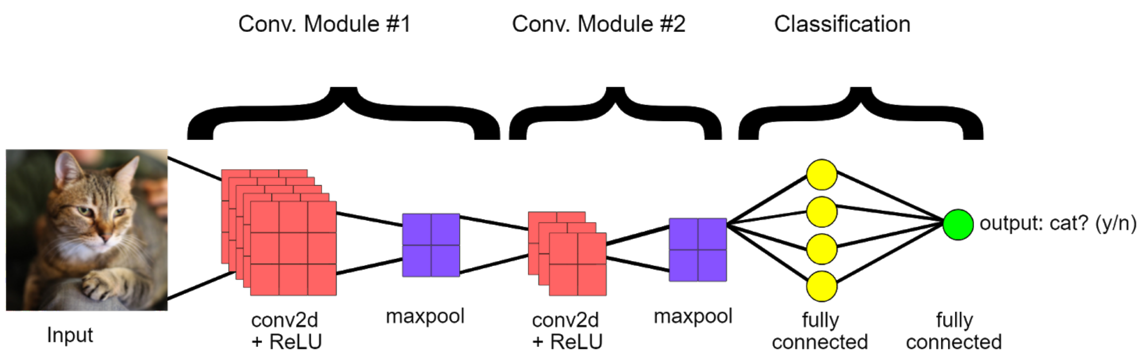
4.1 Getting Familiar

4.1.1 Convolutional Neural Network(CNN) vs Artificial Neural Network(ANN)

Convolutional Neural Networks are very similar to ordinary Neural Networks from the previous chapter: they are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. So what is the difference between CNN and ANN?

Generally speaking, an ANN is a collection of connected and tunable units (a.k.a. nodes, neurons, and artificial neurons) which can pass a signal (usually a real-valued number) from a unit to another. Where as a CNN, in specific, has one or more layers of ***convolution*** units. A convolution unit receives its input from multiple units from the previous layer which together create a proximity. Therefore, the input units (that form a small neighborhood) share their weights.

4.2 Architecture of CNN



Architecture of CNN

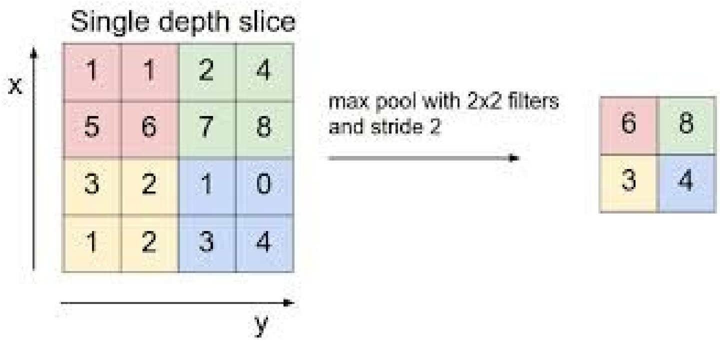
4.2.1 Convolution Layer

Each convolutional neuron processes data only for its receptive field. Although fully connected feedforward neural networks can be used to learn features as well as classify data, it is not practical to apply this architecture to images.

A very high number of neurons would be necessary, even in a shallow (opposite of deep) architecture, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10000 weights for *each* neuron in the second layer. The convolution operation brings a solution to this problem as it reduces the number of free parameters, allowing the network to be deeper with fewer parameters.

4.2.2 Pooling Layer

Convolutional networks may include local or global pooling layers. Pooling layers reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, typically 2 x 2. Global pooling acts on all the neurons of the convolutional layer. In addition, pooling may compute a max or an average. M*ax pooling* uses the maximum value from each of a cluster of neurons at the prior layer. A*verage pooling* uses the average value from each of a cluster of neurons at the prior layer.



Representation of Pooling layer

4.2.3 Fully Connected Layer

Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks. Their activations can hence be computed with a matrix multiplication followed by a bias offset. Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images.

Chapter 5

Result

The image classification used RGB-3 channel images with a size of 28x28 as input images.For training, NVIDIA Tesla K80 was used. The results for two datasets : Fashion Mnist and Build Dataset when applied on Logistic Regression, Neural Networks and CNN model build in Keras is as follows :

	Logistic Regression	**Neural Networks**	**Convolution Neural Networks**
Fashion Mnist	82.54 in 3 epochs	84.80 in 3 epochs	87.97 in 3 epochs
Build Dataset	70.05 in 5 epochs	71.36 in 5 epochs	89.45 in 5 epochs

Chapter 6

Conclusions

In this project, we were able to successfully construct a Convolutional Neural Network whose results were upto the mark. We were able to prepare our own dataset for training and testing. We understood the working of keras preprocessing libraries and were able to produce similar preprocessing results. The process of Image Augmentation was clearly understood and we were able to compute suitable matrices needed for image augmenting. We learnt about the function of filters present in Convolutional Layer, and we were able to replicate a filter which is utilized in edge detection. We were able to understand the significance of Pooling layer and how it sample-downs the entire image without losing critical features. We got a clear idea about the significance of optimizers and activation functions. We were able to compute confusion matrix and f1_score to determine the accuracy of our model.

Based on our results, we were able to prove that Convolutional Neural Network is definitely a huge upgrade in the area of image classification over the traditionally used methods. And we would hence like to conclude by saying Convolutional Neural Network is the future in the field of image classification techniques.

Future Scope

Most online shopping search engines are still largely depend on knowledge base and use keyword matching as their search strategy to find the most likely product that consumers want to buy. This is inefficient in a way that the description of products can vary a lot from the seller’s side to the buyer’s side. The explosive growth in the amount of available digital information has created the challenge of information overload for online shoppers, which inhibits timely access to items of interest on the Internet. This has increased the demand for recommendation systems.

Though almost every e-commerce company nowadays has its own recommendation system that can be used to provide all sorts of suggestions. With the rapid development of neural network these recent years, we can now change the traditional search paradigms from text description to visual discovery. A snapshot of a product tells a detailed story of its appearance, usage, brand and so on. While a few pioneering works about image-based search have been applied, the application of image matching using artificial intelligence in the online shopping field remains largely unexplored.

Bibliography

[1] Sachin Padmanabhan,”Convolutional Neural Networks for Image Classification and Captioning”

[2] Tianmei Guo, Jiwen Dong ,Henjian Li'Yunxing Gao,”Simple Convolutional Neural Network on Image Classification”
https://ieeexplore.ieee.org/abstract/document/807873

[3] Liwei Wang, Yin Li, Jing Huang, Svetlana Lazebnik,”Learning Two-Branch Neural Networks for Image-Text Matching Tasks”

[4]Seong-Hyeon Han, Kwang-Yeob Lee,”Implemetation of Image Classification CNN using Multi#thread GPU”

[5] Jouppi, Norman P., et al. "In-datacenter performance analysis of a tensor processing unit.", the 44th International Symposium on Computer Architecture (ISCA), 2017

Acknowledgement

We have great pleasure in presenting the report on **Image Classification using Convolutional Neural Networks.** We take this opportunity to express our sincere thanks towards our guide **Prof.Sofiya Mujawar** Department of Computer Engineering, APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Prof. Sachin Malave** Head of Department,IT, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Prof. Amol Kalugade T**E project co-ordinator, Department of Computer Engineering, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Student1 : Mrunal S Jadhav

ID1 : 16102030

Student2 : Rohan Dhere

ID1 : 15102057