DEEP LEARNING FOR IMAGE SEARCH

SEMINAR

By

Malvi Shrey
16BCE099

DEPARTMENT OF COMPUTER ENGINEERING
Ahmedabad 382481

DEEP LEARNING FOR IMAGE SEARCH

SEMINAR

Submitted in fulfillment of the requirements

For the degree of

Bachelor of Technology in Computer Engineering

By

Malvi Shrey
16BCE099

Guided By
Dr. Zunnun Narmawala
DEPARTMENT OF COMPUTER ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING
Ahmedabad 382481

CERTIFICATE

This is to certify that the Seminar entitled “Deep Learning for Image Search” submitted by Shrey Malvi (16BCE099), towards the partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Engineering of Nirma University is the record of work carried out by him/her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination.

Dr. Zunnun Narmawala
Associate Professor
Department of Computer Eng.,
Institute of Technology,
Nirma University,
Ahmedabad

Dr. Sanjay Garg
Dept. of Computer Eng.,
Institute of Technology,
Nirma University,
Ahmedabad

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Dr. Zunnun Narmawala for guidance and encouragement in carrying out research for my Seminar. I would also like to thank Computer Engineering Dept, Institute of Technology for considering me and giving me this opportunity to present this topic, which will be helpful in my future research and studies. At last, I would also like to thank my classmates for supporting and helping me throughout the semester.

ABSTRACT

This Seminar is based on how to search an image using Deep Learning algorithms. Deep learning extracts features of an image, learn them provided large dataset and then gives the output to the user according to the given input (image searched). Seminar also covers basic terminologies - Neural Network, Machine Learning, Gradient Descent, etc. Deep Learning Architectures have been discussed in detail with applications. Overview of Deep Learning Frameworks is being given. Lastly, three techniques for image search are been compared on basis of accuracy and time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Overview

Image Search or Image Retrieval system searches images from large dataset of digital images. Deep Learning is a subset of machine learning and machine learning is referred as machines that have ability to “learn”. In short, Deep Learning is implementation of Neural Network with numerous hidden layers. Deep Learning algorithms works very well when the input can be transformed in matrix which makes it easier to learn and adjust parameters for more accuracy. These algorithms first extract features from an image using an DL architecture and then to last layer (reduced dimension) neural network is applied for classification and required image of particular class is found. Deep learning machines are able to learn generic features rather than specific ones. [3]

1.2 Objective

This seminar covers the topic not related to how algorithms work but how they can be used for several applications. This seminar gives understanding of OCR (Optical Character Recognition), One-shot Classification, Auto Encoders which google uses for their search engine and other architectures and how they learn from relatively large dataset.

1.3 Scope of Work

Image Search is not only restricted to what google allow us to search but it has numerous applications. OCR can be used for retrieving suppose Aadhaar card no. form Aadhaar photo or address from specific document. One-Shot Classification can be used for Face Recognition and Signature Recognition by extracting features of an image in one go. Auto-encoders is currently being used by google. Deep learning makes uses of very deep (more no. of layers) Neural Network for learning such that there is hardly a chance of failure of model.

CHAPTER 2 LITERATURE REVIEW

2.1 General

Image Retrieval based on Deep Learning is implemented generally on Python using TensorFlow. A high-end CPU and GPU are needed for training the large dataset. These datasets include ImageNet, Caltech256, Oxford, Paris, Pubfig83LFW. For learning the concepts of Deep Learning, Machine Learning and statistical probability are pre-requisite. So, for further understanding of implementation of these algorithms I had taken the course “Machine Learning Andrew Ng” on Coursera. Also watched several videos how neural network works for CNN and why TensorFlow is utilized for implementing CNN. [4]

2.2 Terminologies

- Neural Networks are combination of several neurons in each layer. ANNs consists of input layer, output layer and hidden layers. These neurons get weight assigned (parameters) through logistic regression. Then, these parameters are optimized using Back Propagation (checks the change in output with respect to every parameter). [3]
- Gradient Descent is a process of optimizing these parameters by partial differentiation at every stage.
- Fully Connected layer is referred as the layer which is going to be fed to the neural network.

CHAPTER 3 DEEP LEARNING ARCHITECTURES

There are mainly four Architectures:

-CNN

-RNN

-LSTM

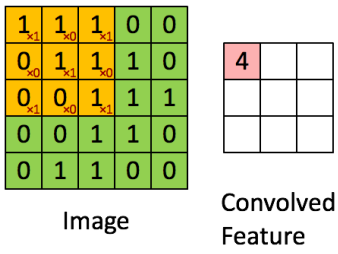
-DBN

3.1 Convolutional Neural Network

In deep learning, CNN are best for finding or extracting features in images to the next layer to form a hierarchy of non-linear features that grow in complexity. A CNN consists of multiple hidden layers such as convolutional layer, pooling layers, padding layers, fully connected layers, etc. [4]

3.1.1 Convolutional layer

In Convolutional layer, a filter (weight matrix) is multiplied to the input layer and we got reduced dimension matrix (3*3 in this case). This reduced matrix shows the contribution of that particular feature in the input layer by multiplication



3.1.2 Pooling and Padding

- Pooling-Used for reducing the number of parameters and prevent overfitting of model. It is present after each convolution layer.
- Padding-For getting the image of same dimension, zeroes are appended to matrix.

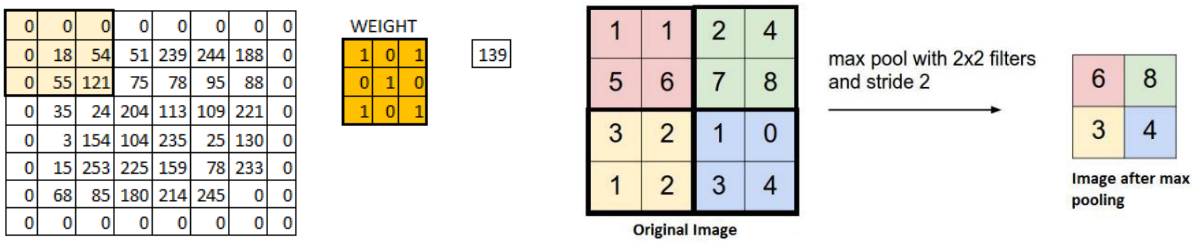


Fig 3.1.2 Pooling and Padding

3.1.3 CNN Diagram

So, CNN is composed of several convolutional and pooling layers.

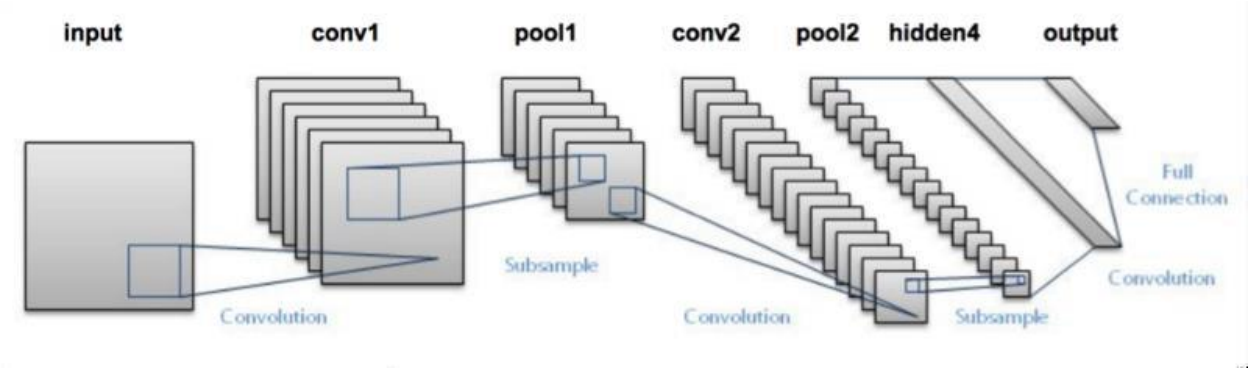
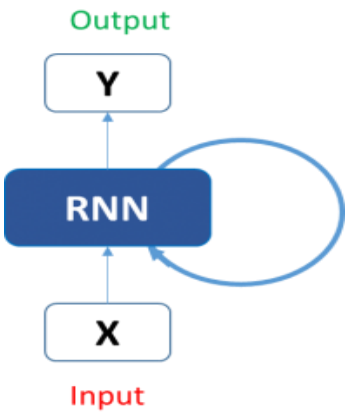


Fig 3.1.3 CNN

Fully-Connected layer which consists of less no. of parameters is then feed to Neural Network for Classification and gives the output which is then compared to the output layer for error generation.

3.2 Recurrent Neural Network



RNN is typical multilayer neural network which have connections that feed back into the same layer. This allows RNNs to maintain memory of past inputs and model problems in time. RNNs can use their internal state for processing sequences of inputs. [3]

3.2.1 Long Short-Term Memory Units

LSTM is modified version of RNN. It consists of input gate, forget gate and output gate. [2]

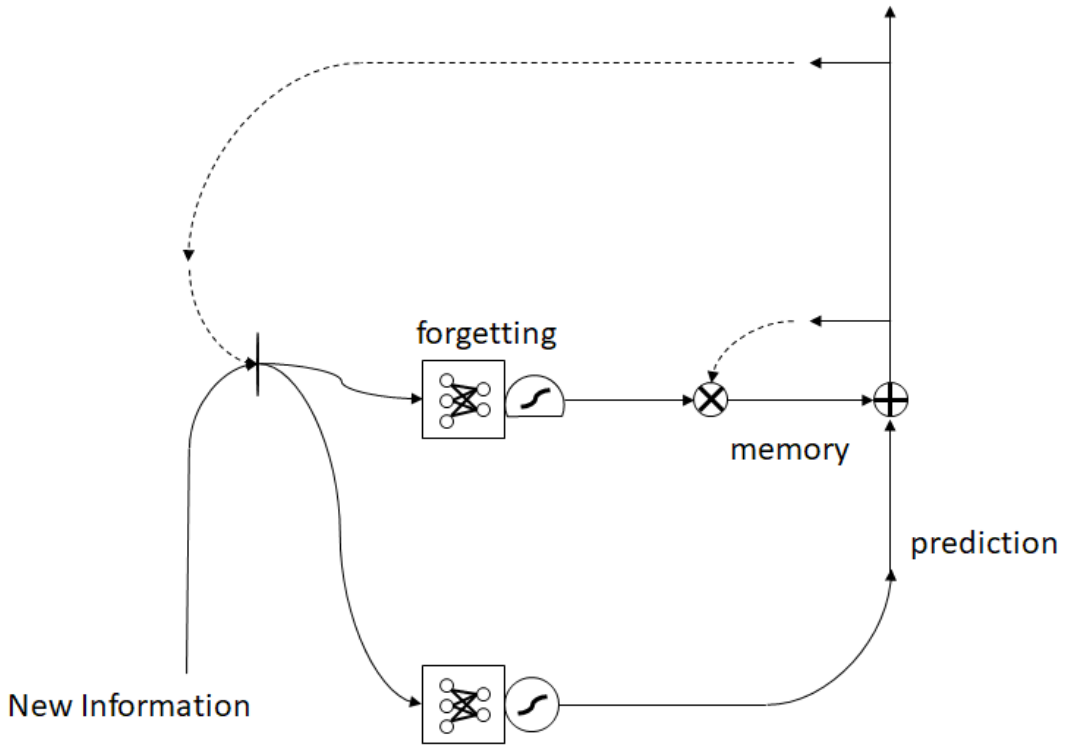


Fig 3.2.2 LSTM

3.3 RNN VS CNN

RNN	**CNN**
RNN is made up of one node. It is fed data then outputs a result back into itself and continues to do this.	CNNs essentially have three parts, convolution layers, pooling layers, and fully-connected layers.
RNN unlike feedforward neural networks - can use their internal memory to process arbitrary sequences of inputs.	CNN is a type of feed-forward artificial neural network - are variations of multilayer perceptron which are designed to use minimal amounts of preprocessing.
RNNs are ideal for text and speech analysis.	CNNs are ideal for images and video processing.

CHAPTER 4 FRAMEWORKS

Implementing these algorithms will be very time consuming if we start building from scratch and also need time to optimize. Hence many Deep Learning Frameworks are available for more accurate and optimized model. Every framework offers a unique range of features. There are mainly four frameworks which are commonly used for implementing these algorithms.

-TensorFlow

-Caffe

-Keras

-DeepLearning4j

4.1 TensorFlow

TensorFlow is a python-based library. Mathematically a Tensor can be considered as a N-dimensional vector.

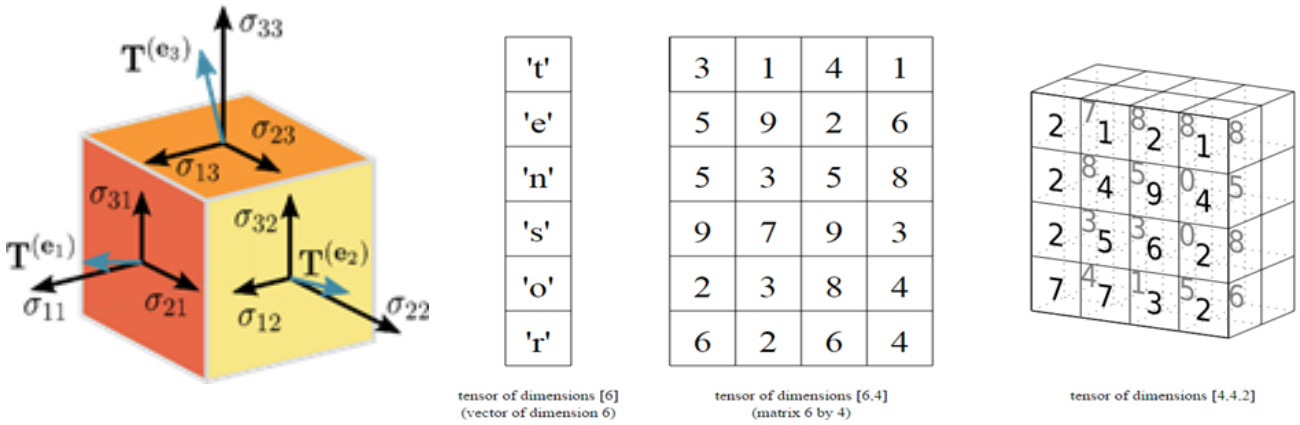


Fig 4.1 Tensors of different dimensions

4.1.1 Characteristics of TensorFlow

- TensorFlow provides Computational Graph in which calculations are performed more efficiently than directly performing in Python.
- TensorFlow can calculate gradients that are needed to be optimize (minimizing the cost function) the variables of the graph for more efficient model.
- For calculating derivatives, Computational Graph uses chain rule which will be time efficient.

4.2 Caffe

Caffe is a C++ based library and finds its primary application in building Convolutional Neural Networks.

4.3 Keras

TensorFlow is somewhat hard to implement so, a more simplified interface library namely Keras is used for building efficient neural networks.

4.4 DeepLearning4j

DeepLearning4j is a Java based library.

CHAPTER 5 IMAGE RETRIEVAL TECHNIQUES

Three different approaches of retrieving the image are been discussed in briefly. These are as follows: [5]

5.1 Optical Character Recognition

OCR is a process in which handwritten images, printed or typed text are converted into machine-encoded text. Automated recognition of credit cards, documents, car plates significantly simplifies the way we collect and process data.

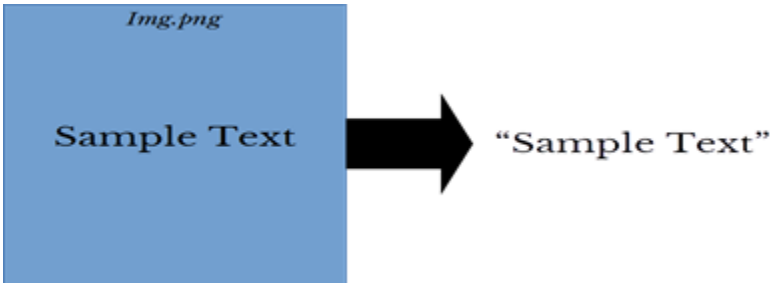


Fig 5.1.1 OCR

OCR techniques are multistage processes. First the image is divided to smaller regions that contain the individual characters and then character is recognized and finally the output is pieced back together.

5.1.1 OCR Working

Let's take an example of an image containing "apple" word. This image is feed to CNN for extracting features and we can reduced dimension (required features) matrix of size 4*8*4. [5]

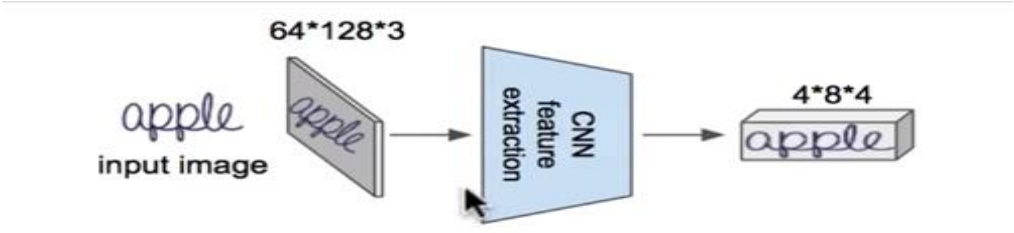


Fig 5.1.2 CNN and reshaping

Next, we reshape the obtain tensor so that we get the sequence of 8 vectors of 16 elements. These 8 vectors are then feed to LSTM network and gives the output which is also a vector of 16 elements. Followed by applying fully connected layer and we get the vector of 6 elements. This vector (6 elements) contains probability distribution of each alphabet symbols at each LSTM step.

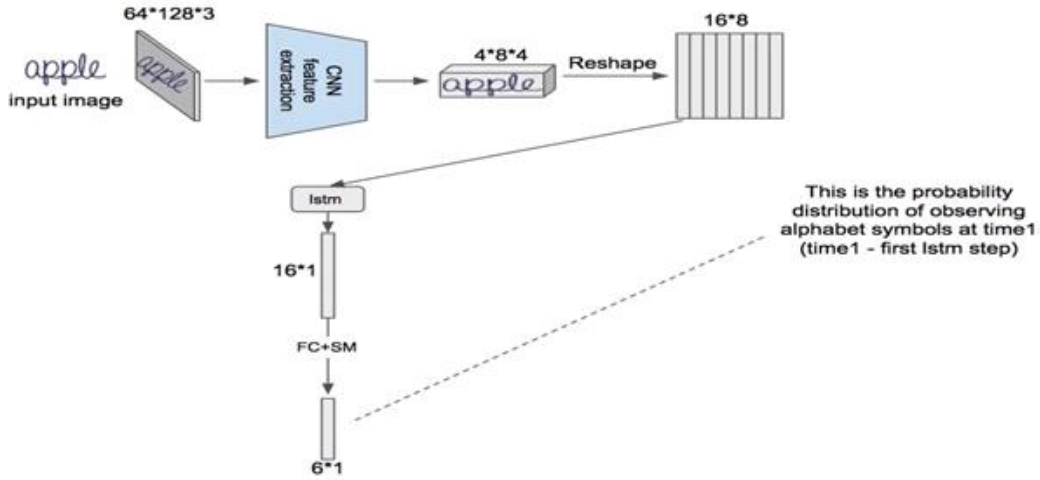


Fig 5.1.3 LSTM and FC+SM

Every vector obtained from LSTM gives the most probable alphabet symbol. The "best path decoding" algorithm is used. Two consecutive repeating characters are glued into one ("e" in our example). Special character "-" allows us to split symbols which are repeated. Then we remove all blank symbols. [5]

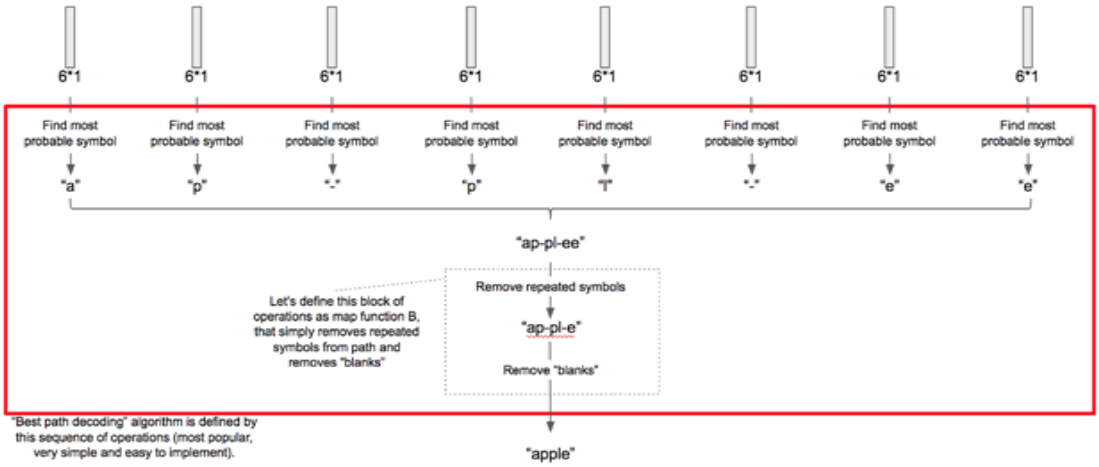


Fig 5.1.4 Best path algorithm

5.2 Siamese Architecture

In Siamese Networks, instead of a model learning to classify its inputs, the neural network learns to differentiate between two inputs. Siamese Networks learns the similarity between two sets of input. [1]

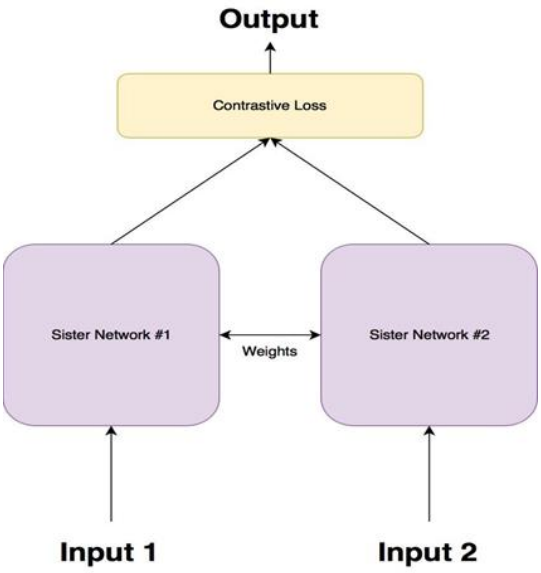


Fig 5.2.1 Siamese Network

This architecture makes use of pre-extracted features from CNN and learns a distance representation based on the user's relevance feedback.

Example: Suppose two are given with two images x1 and x2. Siamese network compute distance d between their encoding f(x1) and f(x2). If it is less than a threshold(hyperparameter), it means that two images are of the same person else they are of different persons.

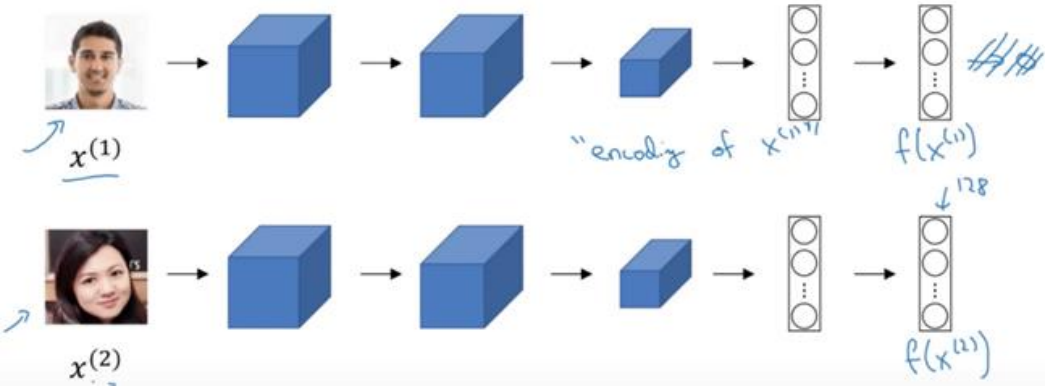


Fig 5.2.2 Comparing two images

5.3 Autoencoders

Autoencoders are neural networks composed of an encoder and a decoder. The main aim is to compress the input data with encoder and decompressing this encoded data with the decoder such that output is a good reconstruction of the input image.

The intermediate layer of autoencoder which contains encoded or compressed data are feed to neural network for training because if we use original image with such huge and hard-to-work-with space of RGB pixels, it would be very time inefficient. Compressed data contains only meaningful dimensions such as “image brightness”, “head shape”, “location of eyes”, etc.

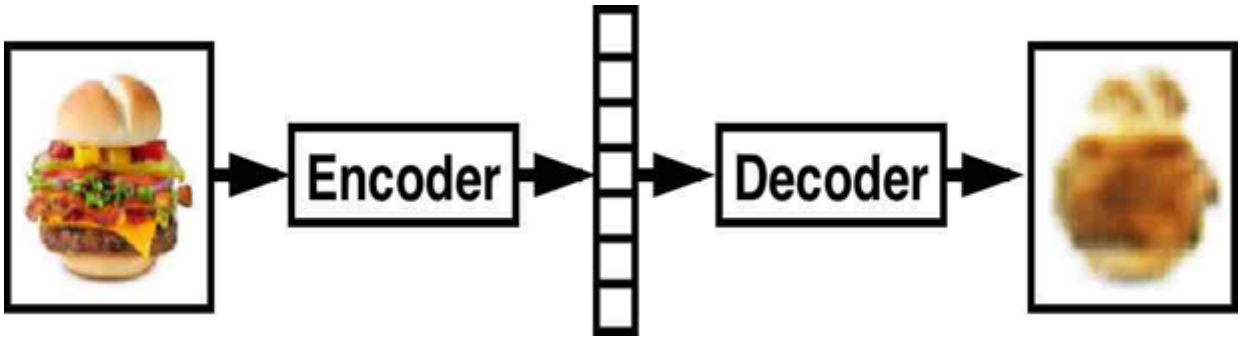


Fig 5.3.1 Dimensionality reduction

The practical applications of autoencoder are data denoising and dimensionality reduction for data visualization.

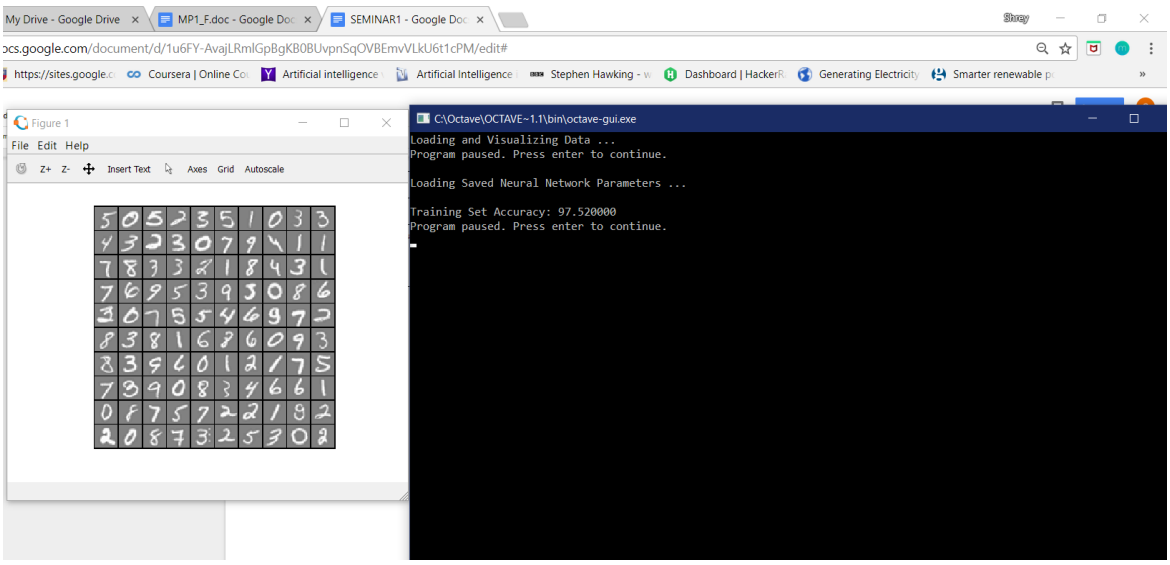
Fig 5.3.2 Denoising

5.4 Siamese Network vs Normal Classifier

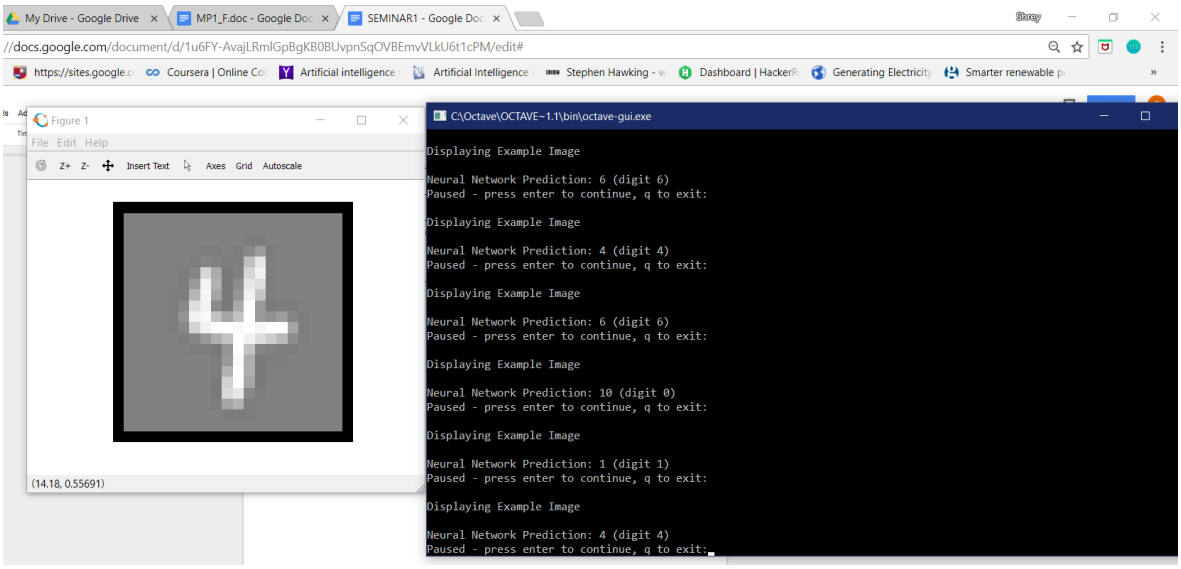
SAIMESE	**NORMAL CLASSIFIER**
Time Efficient	Requires more time to train
Requires less data for training	Relatively large dataset is required
User's feedback is needed	No feedback needed
Mainly used for image and signature verification	Mainly used for feature Extraction and OCR

CHAPTER 6 IMPLEMENTATION

I have implemented the Neural Network and Logistic Regression for recognizing handwritten digits from large dataset in higher level language Octave. Octave provides in-built libraries for calculating gradient descent which makes it easier to implement. The Neural Network I had implemented were having pre-assigned parameters. [1]



6.1 Dataset



6.2 Prediction

CONCLUSION

This Seminar focuses on algorithms and techniques used for searching an image. Google Photos uses deep learning for creating album based on face, landscape, portraits, etc. Every album groups the photo of respective class. Now, a more advanced feature “Google Lens” has also been introduced by Google which scans the photo and provides the necessary information for example: suppose if we scan a book then “Google Lens” displays book name, author name, publish date, etc. Apart from Image Search, deep learning applications includes Speech, failure prediction and many more.

APPENDIX

1. https://www.analyticsvidhya.com/blog/2017/05/25-must-know-terms-concepts-for-beginners-in-deep-learning/
2. https://www.kdnuggets.com/2017/08/deep-learning-neural-networks-primer-basic-concepts-beginners.html
3. https://devblogs.nvidia.com/deep-learning-nutshell-core-concepts/
4. https://towardsdatascience.com/a-beginner-introduction-to-tensorflow-part-1-6d139e038278
5. https://www.ibm.com/developerworks/library/cc-machine-learning-deep-learning-architectures/index.html
6. https://datahub.packtpub.com/deep-learning/top-10-deep-learning-frameworks/
7. http://www.ritchieng.com/machine-learning-photo-ocr/
8. https://dzone.com/articles/using-ocr-for-receipt-recognition

REFERENCES

- Machine Learning Course, Taught by: Andrew Ng: Linear Regression, Logistic Regression, Neural Network
- RNN and LSTM, Deep Neural Networks video by Brandon Rohrer
- Essentials of Deep Learning: Introduction to Long Short-Term Memory by anlayticsvidhya
- Architecture of Convolutional Neural Networks (CNNs) demystified by anlayticsvidhya
- Introduction to OCR - CVISION Technologies