# Project Report

# on

# Automated Handwriting Recognition of English and Native Indian Languages with Deep Learning

(A dissertation submitted in partial fulfilment of the requirements of Bachelor of Technology in Computer Science and Engineering of the Maulana Abul Kalam Azad University of Technology, West Bengal)

Submitted by

Souporno Ghosh

Soumya Nasipuri

Rahul Roy

Sharanya Saha

Under the guidance of

Smt. Jaya Paul

Asst. Prof.,

Dept. of Computer Science and Engineering

# Government College of Engineering and Leather Technology

(Affiliated to MAKAUT, West Bengal)

Kolkata - 700106, WB

2020-2021

# Certificate of Approval

This is to certify that the project report on “Automated Handwriting Recognition of English and Native Indian Languages with Deep Learning” is a record of bona fide work, carried out by Shri Souporno Ghosh, Shri Soumya Nasipuri, Shri Rahul Roy and Smt Sharanya Saha under my guidance and supervision.

In my opinion, the report in its present form is in conformity as specified by Government College of Engineering and Leather Technology and as per regulations of the Maulana Abul Kalam Azad University of Technology, West Bengal. To the best of my knowledge the results presented here are original in nature and worthy of incorporation in project report for the B.Tech. Program in Computer Science and Engineering.

Signature of Signature of

Supervisor/ Guide Head, Dept. of CSE

# ACKNOWLEDGEMENT

With great pleasure, I would like to express my profound gratitude and indebtedness to Smt Jaya Paul, Department of Computer Science and Engineering, Government College of Engineering and Leather Technology, W.B. for his continuous guidance, valuable advice and constant encouragement throughout the project work. His valuable and constructive suggestions at many difficult situations are immensely acknowledged. I am in short of words to express his contribution to this thesis through criticism, suggestions and discussions.

I would like to take this opportunity to thank Dr. Santanu Halder, HOD, Department of Computer Science & Engineering, Government College of Engineering and Leather Technology.

I would like to express my gratitude to Sri S. Mondal for his valuable suggestions and help.

Signatures

1. Souporno Ghosh - 11200117028
2. Soumya Nasipuri - 11200117029
3. Rahul Roy - 11200117039
4. Sharanya Saha - 11200117033

# Dedicated to

# Alan Turing, Ada Lovelace and John von Neumann

# The pioneers on whose work we expand upon.

**ABSTRACT**

**Significant amount of research has been done in the field of handwriting recognition, particularly for characters in the Latin-based alphabets (English, French, Spanish, German, etc). However, there is a significant lack of literature and research on handwriting recognition for Devanagari based languages, such as Hindi, Bangla, Sanskrit, etc. In this project, we attempt to remedy that in an attempt to create an API that is able to recognize English, Hindi and Bangla handwriting. The primary goal of the API is to identify the author of a word, sentence or passage from the handwriting written in English, Hindi or Bangla. We also provide the methods used by us in this attempt in order to facilitate further study and replication of this API for future research.**

# CONTENTS

1. **INTRODUCTION 1-2**
   1. **Motivation 1**
   2. **Background 1**
   3. **Summary of present work 1**
   4. **Organization of the thesis 1**
   5. **Hardware/Software used 1**

1. **DATA COLLECTION AND PREPROCESSING 3-3**
   1. **Collection of Dataset 3**
   2. **Preparation of Dataset 3**
   3. **Construction of Model 3**
2. **METHODOLOGY 4-4**
   1. **About VGG 16 Model**
      1. **Architecture of VGG16**
      2. **Challenges of VGG16**
3. **PROGRAM WALKTHROUGH 5-5**
4. **OUTPUTS AND RESULTS 6-6**

**CONCLUSIONS 60**

**REFERENCES 62**

# CHAPTER 1. INTRODUCTION

## Motivation

Our great nation has produced many literary geniuses; Munshi Prem Chand, Rabindranath Tagore, Vikram Seth, Bankim Chandra Chatterjee, Sukumar Roy, etc. These pioneers have blessed us with a variety of literary masterpieces that provide not only an insight into their own minds but also insight on humanity and contemporary times. It is sufficed to say one learns a lot about mankind from their works. Moreover, their works and by extension, the manuscripts of said works, are a national treasure. Hence, it is imperative that their original works can be verified as their own. We can achieve this by analysing the handwritings of the writers.

Handwriting recognition will also help accelerate the field of forensic analysis, and in turn help the law enforcement authorities. Notes found on crime scenes and related to victims and suspects can be analysed and such analysis can help us identify perpetrators of a crime.

the field of education and academia, handwriting analysis can help us curb plagiarism. Plagiarism checking is a major field of research in academia and handwriting analysis can also help support those efforts.

These are only a few of the applications of handwriting recognition that we drove us to choose this topic for our project. Recognition and analysis of handwriting has applications in various fields such as archaeology, criminal detection, academia, education, etc. However, so far handwriting analysis has only been performed by human hands. In modern days, handwriting recognition has mostly only been attempted for languages based on Latin-based alphabet.

Literature related to handwriting recognition is limited for Devanagari related languages, such as Hindi and Bengali. Hence, our feeble attempt at remedying that.

## Background

<Literature Survey and related work and discuss in detail>

## Summary of present work

## Organisation of the Thesis

The thesis has been organised into primarily four chapters.

The Data Collection and Pre-processing chapter deals with the details of all the work that must be done before the actual programming part can be approached. This includes, but is not limited to data collection and preparation

The Methodology chapter deals with the necessary theory required to approach the problems. This includes any experiments we ran to test out our algorithms and any preliminary work which we referred to help with our project.

The Program Walkthrough chapter deals with all the actual program written for the image preparation and the creation of the model.

The Outputs and Results chapter deals with the output that we get from executing the program and how they are significant.

## Hardware/Software used

Primary language used for programming is Python 3.8.

Packages used in Python are TensorFlow v2.0 (a deep learning library by Google, Inc.), Keras (to help with integrating TensorFlow with Python), NumPy, Pandas, MatPlotLib, Pyplot and Seaborn. The primary algorithm used for image recognition is VGG16.

The primary hardware used comprise the personal laptop computers belonging to the team members; HP Pavilion with Intel i5 Quad-Core, Integrated Graphics Card, 8GB RAM; Dell G3 with 2.6GHz Hexa-Core Intel i7 Processor, Integrated Graphics Card, 8GB RAM and Asus Vivibook 2GHz Quad-Core AMD Ryzen 5 Processor, Integrated Graphics Card, 8GB RAM.

The processor and the RAM determine how fast the program will run and the dataset will be trained. The graphics card would also have accelerated the process, if GPU parallel processing was implemented in the project. A major issue with deep learning projects is that if the processing power is too less, then the model might take as long as a few days to train. However, our initial project does not take as much time and can be replicated with a machine with as low specifications as 1.6 GHz Dual Core Processor, Integrated Graphics Card and 4GB RAM. Anything lower might take too long to process. (Training the Dataset will require much better specifications, however, since the original VGG16 model was trained on 2-3 weeks on a Nvidia Titan GPU. In such a scenario, training the data on a cloud platform like GCP, AWS or Azure might yield better results).

# CHAPTER 2. INITIAL STAGES

## Collection of Dataset

A major criterion of success in working on problems whose solutions depend on machine or deep learning is the presence of large datasets. The larger the dataset, the better and more accurate the model becomes.

One of the main challenges of attempting handwriting recognition in a new language is the lack of sufficient datasets. Our supervisor for the project, Smt Jaya Paul, was generous enough to assist us with that. We used 30 volunteers of various native languages and asked them to write certain passages in their native and English languages. This provided us with the necessary dataset needed for proper model training.

There are three languages in the dataset are Bangla, Hindi and English.

## Preparation of Dataset

The handwritten passages were scanned into image. The scanned images passages were processed to eliminate noise and then segmented into word-sized images. The methods used for this was provided to us from a previous project attempted by Tapan Mondal, Suraj Ahmed Hossain, Spandan Mondal, Raihan Afroz and Anowar Hossain [1] These word sized images are then arranged into several folders in the following method.

* The images are distinguished according to pairs of authors and for each author, there are five (5) sets of data.
* Each of these 5 sets are divided into two subfolders; train containing three sets and test containing two sets.
* The images are in Tag Image File Format (TIFF). They are named in the following format: <Author Code>\_<Set Number>\_<Image Number>. For example, the first image of the first set of the first author (Author Code 0) is *0000\_01\_0.tiff*.
* There are five thusly organized datasets.

These images are now ready to be fed as inputs for the model.

## Construction of the Model

Currently, the base model that we will be using is called VGG16. VGG16 is one of the most popular models for image recognition. It was introduced in 2014 [2]. VGG16 takes the input in 224 x 224 images and the pretrained VGG 16 model will predict from the 1000 categories provided in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014.

In our dataset, the size of the segmented images is random, and hence some pre- is required so that the input images can be converted to the required size of 224 x 224. We shall do that with the help of the in-built VGG 16 pre-processing tools provided in the Keras library. The Program Walkthrough mentions exactly how we do it.

# CHAPTER 3. METHODOLOGY

## About VGG 16 Model

VGG 16 was proposed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group Lab at the University of Oxford in 2014 in an article titled "Very Deep Networks for Large Scale Image Recognition." [2] This model achieves 92.7% of the top 5 accuracy tests on the ImageNet dataset of 14 million images belonging to 1000 categories. The model inputs 224 x 244 pixels size images in RBG channels.

### Architecture of VGG16:

The input to the network is image of dimensions (224, 224, 3). The first two layers have 64 channels of 3\*3 filter size and same padding. Then after a max pool layer of stride (2, 2), two layers which have convolution layers of 256 filter size and filter size (3, 3). This followed by a max pooling layer of stride (2, 2) which is same as previous layer. Then there are 2 convolution layers of filter size (3, 3) and 256 filter. After that there are 2 sets of 3 convolution layer and a max pool layer. Each have 512 filters of (3, 3) size with same padding. This image is then passed to the stack of two convolution layers. In these convolution and max pooling layers, the filters we use is of the size 3\*3 instead of 11\*11 in Alex Net and 7\*7 in ZF-Net. In some of the layers, it also uses 1\*1 pixel which is used to manipulate the number of input channels. There is a padding of 1-pixel (same padding) done after each convolution layer to prevent the spatial feature of the image.

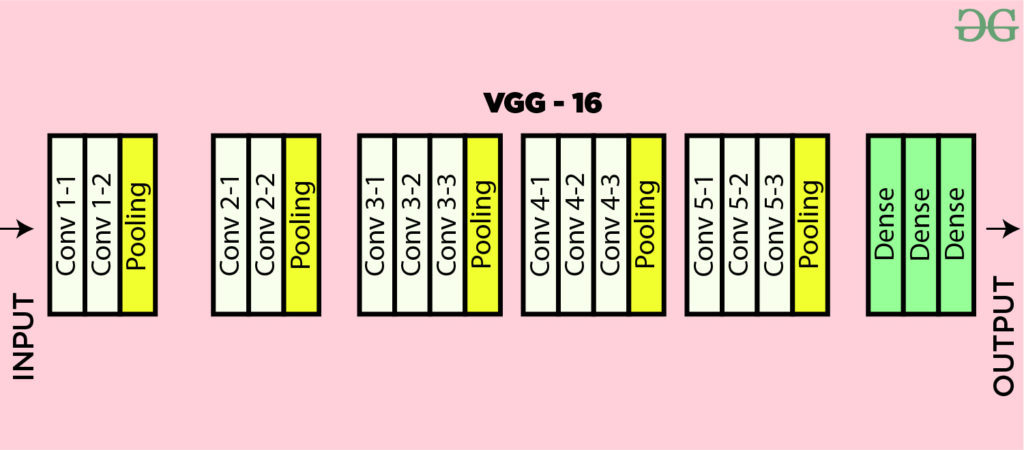


Fig 1.1. Architecture of VGG16 Model

After the stack of convolution and max-pooling layer, we got a (7, 7, 512) feature map. We flatten this output to make it a (1, 25088) feature vector. After this there are 3 fully connected layer, the first layer takes input from the last feature vector and outputs a (1, 4096) vector, second layer also outputs a vector of size (1, 4096) but the third layer output 1000 channels for 1000 classes of ILSVRC challenge, then after the output of 3rd fully connected layer is passed to SoftMax layer in order to normalize the classification vector. After the output of classification vector top-5 categories for evaluation. All the hidden layers use ReLU as its activation function. ReLU is more computationally efficient because it results in faster learning and it also decreases the likelihood of vanishing gradient problem.

This is how the pretrained VGG16 model works.

### Challenges of VGG16:

* It is very slow to train (the original VGG model was trained on NVidia Titan GPU for 2-3 weeks).
* The size of VGG16 trained ImageNet weights is 528 MB. So, it takes quite a lot of disk space and bandwidth that makes it inefficient.

## Initial Experimentation

Before we proceeded to handwriting recognition, we experimented with a cat and dog classifier. The dataset we found was from Kaggle. We attempted to create our own VGG16 model in order to better understand the VGG16 model. We shall once walk through the program to understand the VGG 16 model better.

First, we import Keras and OS libraries. Keras is the package that helps us use the TensorFlow 2.0 Library by Google, Inc.

import keras

import os

Then we import the necessary modules from keras. This include but are not limited various algorithms that will constitute each layer of the model (namely Dense, Convolution, Max Pool and Flattening), image pre-processing tools and various keras utilities.

from keras.models import Sequential

from keras.layers import Dense, Conv2D, MaxPool2D, Flatten

from keras.preprocessing.image import ImageDataGenerator

from keras.callbacks import ModelCheckpoint, EarlyStopping

from keras.preprocessing import image

from keras.models import load\_model

from keras.optimizers import Adam

We will also import NumPy and Mat Plot Library. NumPy assists us with handling matrices and Mat Plot Library helps us to view graphs and plots.

import numpy as np

import matplotlib.pyplot as plt

Here we will load the datasets into the program. The data is divided into two subsets; test and train.

trdata = ImageDataGenerator()

traindata = trdata.flow\_from\_directory(

directory="./train", target\_size=(224, 224))

tsdata = ImageDataGenerator()

testdata = tsdata.flow\_from\_directory(

directory="./test", target\_size=(224, 224))

From here we will add will add all the layers relevant to the VGG16 model in sequentially.

model = Sequential()

The following lines of code will add the first set of layers. This includes two Convolutional networks. The first one takes in the image of size 224 x 224 in RGB channels. The next one has 64 filters. ReLU is often used as the activation function. The output after these two layers is pooled with max pooling algorithm.

model.add(Conv2D(input\_shape=(224, 224, 3), filters=64,

kernel\_size=(3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=64, kernel\_size=(3, 3),

padding="same", activation="relu"))

model.add(MaxPool2D(pool\_size=(2, 2), strides=(2, 2)))

Then we will add two more convolutional networks with 128 filters, activation function ReLU and the same padding. The output after these two layers is pooled with max pooling algorithm.

model.add(Conv2D(filters=128, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=128, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(MaxPool2D(pool\_size=(2, 2), strides=(2, 2)))

Then we will add three more convolutional networks with 256 filters, activation function ReLU and the same padding. The output after these two layers is pooled with max pooling algorithm.

model.add(Conv2D(filters=256, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=256, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=256, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(MaxPool2D(pool\_size=(2, 2), strides=(2, 2)))

Then we will add three more convolutional networks with 512 filters, activation function ReLU and the same padding. The output after these two layers is pooled with max pooling algorithm.

model.add(Conv2D(filters=512, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=512, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=512, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(MaxPool2D(pool\_size=(2, 2), strides=(2, 2)))

Then we will add another set of three more convolutional networks with 512 filters, activation function ReLU and the same padding. The output after these two layers is pooled with max pooling algorithm.

model.add(Conv2D(filters=512, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=512, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(Conv2D(filters=512, kernel\_size=(

3, 3), padding="same", activation="relu"))

model.add(MaxPool2D(pool\_size=(2, 2), strides=(2, 2)))

We use Rectified Linear Unit (ReLU) for activation so that the negative values are not passed to the next layers. After this, we will add two Dense layers with ReLU activation with 4096 units followed by a SoftMax Dense layer with two units.

model.add(Flatten())

model.add(Dense(units=4096, activation="relu"))

model.add(Dense(units=4096, activation="relu"))

model.add(Dense(units=2, activation="softmax"))

Here, we will be using Adam optimiser to reach to the global minima while training out model. If the algorithm stuck in local minima while training then the Adam optimiser will help us to get out of local minima and reach global minima. We will also specify the learning rate of the optimiser, here in this case it is set at 0.001. If our training is bouncing a lot on epochs then we need to decrease the learning rate so that we can reach global minima.

opt = Adam(lr=0.001)

model.compile(optimizer=opt,

loss=keras.losses.categorical\_crossentropy, metrics=['accuracy'])

We will then print the summary of the model.

model.summary()

We shall use the ModelCheckpoint Utility from Keras library to save the model only when the validation accuracy of the model in current epoch is greater than what it was in the last epoch. The EarlyStopping Utility from Keras library to stop the learning in case there is no increase in validation accuracy for 20 epochs.

checkpoint = ModelCheckpoint("vgg16\_1.h5", monitor='val\_acc', verbose=1,

save\_best\_only=True, save\_weights\_only=False, mode='auto', period=1)

early = EarlyStopping(monitor='val\_acc', min\_delta=0,

patience=20, verbose=1, mode='auto')

hist = model.fit\_generator(steps\_per\_epoch=100, generator=traindata, validation\_data=testdata,

validation\_steps=10, epochs=100, callbacks=[checkpoint, early])

The fit generator method in the model class helps us pass the data to the model for training and validation. The method handles everything necessary for passing data to the model and training the model. After this we will print the information about the model training with the help of the MatPlot Library.

plt.plot(hist.history["acc"])

plt.plot(hist.history['val\_acc'])

plt.plot(hist.history['loss'])

plt.plot(hist.history['val\_loss'])

plt.title("model accuracy")

plt.ylabel("Accuracy")

plt.xlabel("Epoch")

plt.legend(["Accuracy", "Validation Accuracy", "loss", "Validation Loss"])

plt.show()

Now to predict using the model that we created, we will import an image and use the Keras library to preprocess it for the VGG16 model we created.

img = image.load\_img("image.jpeg", target\_size=(224, 224))

img = np.asarray(img)

plt.imshow(img)

img = np.expand\_dims(img, axis=0)

saved\_model = load\_model("vgg16\_1.h5")

output = saved\_model.predict(img)

if output[0][0] > output[0][1]:

print("cat")

else:

print('dog')

This above experimental program gives us in depth understanding of the VGG16 algorithm and its architecture.

Google: data augmentation (what is); ImageNet; vgg16 pre-processing; ImageNet dataset; Alex net, capsule net

## Extracting Features

Before we could start training the actual model, we needed to pre-process the input. This is primarily because the input for the VGG16 model requires the images to be of size 224 x 224 pixels and RGB channel. Additionally, before we train the model, we will extract the necessary feature required for prediction.

To obtain the necessary features, we take the pretrained VGG16 model available in the Keras library and remove the top layer, where the prediction (with the 1000 categories from ILSVRC challenge) actually happens. Then we pass the data through it to get a “prediction” to get a feature matrix as an output. The below program does that for one image in the dataset.

First, we will import all the necessary packages.

import keras

from PIL import Image

from keras.preprocessing import image

from keras.applications.vgg16 import VGG16

from keras.applications.vgg16 import preprocess\_input

from keras.preprocessing.image import ImageDataGenerator

import numpy as np

#data representation, data store, data handling

import pandas as pd

import matplotlib.pyplot as plt

#beautifies graphs based on seaborn

import seaborn as sns

# initialises seaborn

sns.set()

# top is not included as we are only interested in feature extraction

model = VGG16(weights='imagenet', include\_top=False)

# dealing with only feature extraction

model.summary()

# prepares images for loading into the model into the program

def image\_loader(img\_path):

# uses the keras preprocessing library to prepare the images for loading

# here the image path is loaded and the target size is made 224 224 acc to vgg 16 specs

img = image.load\_img(img\_path, target\_size=(224, 224))

# print(img.shape)

# gives error as the object is 'Image', not array

# shape: numpy method, input array, no of rows and column

# shows the image

plt.imshow(img)

# converts the image to an array

img\_data = image.img\_to\_array(img)

# prints the shape

print(img\_data.shape)

# converts the image array into a vector from a single dimensional array

img\_data = np.expand\_dims(img\_data, axis=0) # google axis = 0

# prints the shape

print(img\_data.shape)

# google

img\_data = preprocess\_input(img\_data)

print(img\_data.shape)

return img\_data

img\_path = '../Datasets/writerPair1/0-10/train/0/0000\_01\_0.tif'

img\_data = image\_loader(img\_path)

#using the pretrained vgg16 model to output the feature matrix of one single image for testing

model\_feature = model.predict(img\_data)

print(model\_feature.shape)

# CHAPTER 4. PROGRAM WALKTRHOUGH

For each set, we will create one csv file with 2d array with feature matrices of multiple images

# CHAPTER 5. OUTPUTS AND RESULTS

# Reference

1. Tapan Mondal, Suraj Ahmed Hossain, Spandan Mondal, Raihan Afroz and Anowar Hossain, “Preprocess the Handwritten Document Image for Preparing Writer Recognition” as Project Work
2. Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition” in *International Conference on Learning Representations (ICLR 2015)*, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)