VISVESVARAYA TECHNOLOGICAL UNIVERSITY

JNANA SANGAMA, BELAGAVI – 590 018

An Internship Project Report on

"HANDWRITTEN TEXT RECOGNITION"

Submitted in partial fulfillment of the requirements as a part of the

AI/ML INTERNSHIP

(NASTECH)

For the award of degree of

Bachelor of Engineering
in
Information Science and Engineering

Submitted by

Rushwanth K

1RN19IS123

Under the Guidance of

Mr. Ravi Kumar S G
Assistant Professor
Department of ISE

Department of Information Science and Engineering

RNS Institute of Technology

Channasandra, Dr. Vishnuvardhan Road, RR Nagar Post,
Bengaluru – 560 098

2022 -2023

RNS Institute of Technology

Channasandra, Dr. Vishnuvardhan Road, RR Nagar Post,

Bengaluru – 560 098

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the internship project report entitled ***HANDWRITTEN TEXT RECOGNITION*** has been successfully completed by **RUSHWANTH K** bearing USN **1RN19IS123**, presently VIII semester student of **RNS Institute of Technology** in the partial fulfillment of requirements as a part of the ***AI/ML Internship (NASTECH)*** for the award of the degree of ***Bachelor of Engineering in Information Science and Engineering*** under ***Visvesvaraya Technological University, Belagavi*** during academic year **2022-2023**. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report and deposited in the department library. The Internship project report has been approved as it satisfies the academic requirements as a part of AI/ML Internship.

Mr. Ravi Kumar S G	**Dr. R Rajkumar / Mr. Pramoda R**	**Dr. Suresh L**
Internship Guide	Internship Coordinators	Professor & HoD
Assistant Professor	Associate/Assistant Professor	Department of ISE
Department of ISE	Department of ISE	RNSIT

External Viva

Name of the Examiners	**Signature with date**
1. _____	_____
2. _____	_____

NASTECH

Microsoft Partner Network

CERTIPORT
A PEARSON VUE BUSINESS
AUTHORIZED TESTING CENTER

CERTIFICATE

This is to certify that **Rushwanth K** bearing USN: **1RN19IS123** from RNSIT has taken part in internship training on the **Artificial Intelligence & Machine Learning and successfully completed** project work" **Handwritten text recognition** "in New Age Solutions Technologies (NASTECH) from March 2022 to June 2022.

Rushwanth K has shown great enthusiasm and interest in the project. The incumbents' conduct and performance were found satisfactory during all phases of the project.

Thank you very much.

Sincerely Yours,

Deepak Garg

Founder

Address: Block No.10, Kailashnagar, Highland Residency, Kolshet Road, Thane (W)- 400608, Mumbai, Maharashtra.
Mobile Number: 8591132820| E-mail: deeprgarg@nastech.in

DECLARATION

I, **Rushwanth K [USN: 1RN19IS123]**, student of VIII Semester BE, in Information Science and Engineering, RNS Institute of Technology hereby declare that the Internship project work entitled ***Handwritten Text Recognition*** has been carried out and submitted in partial fulfillment of the requirements for the *VII Semester degree of* ***Bachelor of Engineering in Information Science and Engineering*** *of Visvesvaraya Technological University, Belagavi* during academic year 2022-2023.

Place : Bengaluru

Date :

Rushwanth K
1RN19IS123

ABSTRACT

Character recognition is one in all the emerging fields within the computer vision. The most abilities of humans are they will recognize any object or thing. The hand transcription can be easily identified by humans. Different languages have different patterns to spot. Humans can identify the text accurately. The hand transcription cannot be identified by the machine. It's difficult to spot the text by the system.

During this text recognition, the input image is processed, extraction of features, and classification schema takes place, training of system to acknowledge the text. During this approach, the system is trained to seek out the similarities, and also the differences among various handwritten samples. This application takes the image of a hand transcription and converts it into a digital text.

ACKNOWLEDGMENT

The fulfillment and rapture that go with the fruitful finishing of any assignment would be inadequate without the specifying the people who made it conceivable, whose steady direction and support delegated the endeavors with success.

I would like to profoundly thank **Management** of **RNS Institute of Technology** for providing such a healthy environment to carry out this AI/ML Internship Project.

I would like to express my thanks to my Principal **Dr. M K Venkatesha** for his support and for inspiring me towards the attainment of knowledge.

I wish to place on record my words of gratitude to **Dr. Suresh L,** Professor and Headof the Department, Information Science and Engineering, for having accepted to patronize me in the right direction with all his wisdom.

I like to express my profound and cordial gratitude to my Internship Project Guide**, Mr. Ravi Kumar S G,** Assistant Professor, Department of Information Science and Engineering for his valuable guidance, constructive comments, continuous encouragement throughout the Internship.

I like to express my profound and cordial gratitude to my Internship Project Coordinators**, Dr. R Rajkumar** Associate Professor and **Mr. Pramoda R,** Assistant Professor, Department of Information Science and Engineering for their valuable guidance, constructive comments, continuous encouragement throughout the Internship Work and guidance in preparing report.

I would like to thank all other teaching and non-teaching staff of Information Science& Engineering who have directly or indirectly helped me to carry out the Internship Project Work.

Also, I would like to acknowledge and thank my parents who are source of inspiration and instrumental in carrying out this Internship Project Work.

RUSHWANTH K
USN:1RN19IS123

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 ORGANIZATION/INDUSTRY

1.1.1 COMPANY PROFILE

NASTECH is formed with the purpose of bridging the gap between Academia and Industry. Nastech is one of the leading Global Certification and Training service providers for technicaland management programs for educational institutions. They collaborate with educational institutes to understand their requirements and form a strategy in consultation with all stakeholders to fulfill those by skilling, reskilling and upskilling the students and faculties on new age skills and technologies.

1.1.2 DOMAIN/TECHNOLOGY

The domain chosen for my project is AI/ML. Machine learning, the fundamental driver of AI,is possible through algorithms that can learn themselves from data and identify patterns to make predictions and achieve your predefined goals, rather than blindly following detailed programmed instructions, like in traditional computer programming. This technology allows the machine to perceive, learn, reason and communicate through observation of data, like a child that grows up and acquires knowledge from examples. Machines also have the advantage of not being limited by our inherent biological limitations. With machine learning, manufacturing companies have increased production capacity up to 20%, while lowering material consumption rates by 4%.

Nowadays, the revolutionary AI technology evolved from rule-based expert systems to machine learning and more advanced subcomponents such as deep learning (learning representations instead of tasks), artificial neural networks (inspired by animal brains) and reinforcement learning (virtual agents rewarded if they made good decisions).

The AI can master the complexity of the intertwining industrial processes to enhance the whole flow of production instead of isolated processes. This enormous cognitive capacity gives the AI the ability to consider the spatial organization of plants and the timing constraints of live production. Another key advantage is the capability of AI algorithms to think probabilistically with all the subtlety this allows in edge cases, instead of traditional rule-based methods that require rigid theories and a full comprehension of problems.

1.1.3 Department

R.N. Shetty Institute of Technology (RNSIT) established in the year 2001, is the brain-child ofthe Group Chairman, Dr. R. N. Shetty. The Murudeshwar Group of Companies headed by Sri.R. N. Shetty is a leading player in many industries viz construction, manufacturing, hotel, automobile, power & IT services and education. The group has contributed significantly to the field of education. A number of educational institutions are run by the R. N. Shetty Trust, RNSIT being one amongst them. With a continuous desire to provide quality education to the society, the group has established RNSIT, an institution to nourish andproduce the best of engineering talents in the country. RNSIT is one of the best and top accredited engineering colleges in Bengaluru.

1.2 PROBLEM STATEMENT

1.2.1 Existing System and their Limitations

The Current system is a bit flawed because it recognizes only the typed letters. But here I am recognizing the human handwritten texts and extract the words from it.

1.2.2 Proposed Solution

Handwritten Text Recognition (HTR) system implemented with TensorFlow (TF) and trained on the IAM off-line HTR dataset. This Neural Network (NN) model recognizes the text contained in the images of segmented words.

1.2.3 Problem formulation

Besides the two decoders shipped with TF, it is possible to use word beam search decoding. Using this decoder, words are constrained to those contained in a dictionary, but arbitrary non-word character strings (numbers, punctuation marks) can still be recognized.

Chapter 2

REQUIREMENT ANALYSIS, TOOLS &TECHNOLOGIES

2.1 Hardware and Software Requirements

2.1.1 Hardware Requirements:

- Processor : Any Processor above 500 MHz
- RAM : 2 GB and above
- Hard Disk : 4 GB
- Input device : Standard Keyboard and Mouse
- Output device : VGA and High Resolution Monitor

2.1.2 Software Requirements:

- Operating system : Windows 10
- IDE : Google collab (Any browser)
- Tools/Technologies : Python, tensorflow, cv2, streamlit

2.2 Tools/Languages/Platforms

- Python

Chapter 3

DESIGN AND IMPLEMENTATION

3.1 Problem Statement

- I use a NN for this task. It consists of a convolutional neural network (CNN) layer, recurrent neural network (RNN) layers, and a final Connectionist Temporal Classification (CTC) layer.
- In this project, I've taken 5 CNN (feature extraction) and a pair of RNN layers and a CTC layer (calculate the loss). First, preprocessing the pictures is done in order to reduce the noise.
- I can also view the NN in an exceedingly more formal way as a function (see Eq. 1) which maps a picture (or matrix) M of size W×H to a personality sequence (c1, c2, …) with a length between 0 and L. As you'll see, the text is recognized on character-level, therefore words or texts not contained within the training data is recognized too (as long because the individual characters get correctly classified).

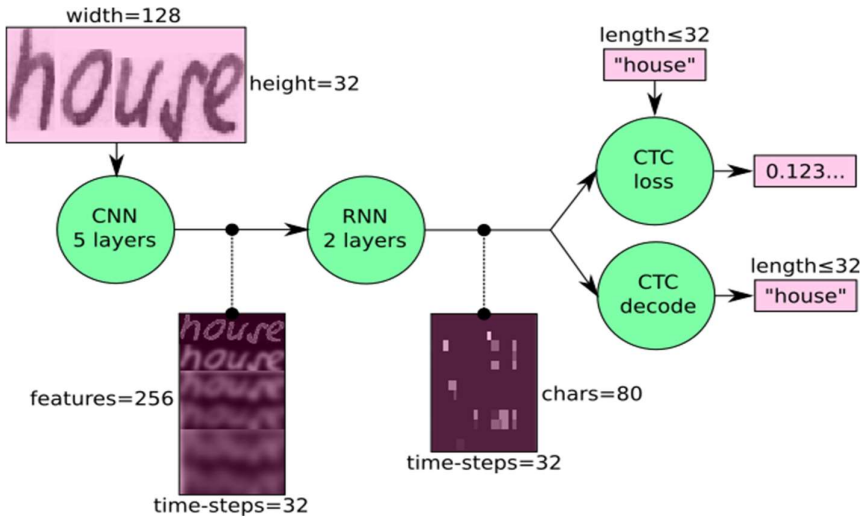

Fig 3.1 Overview of the NN operations (green) and the data flow through the NN (pink).

3.2 Operations

• **CNN**: the input image is fed into the CNN layers. These layers are trained to extract relevant features from the image. Each layer consists of three operations. First, the convolution operation, which applies a filter kernel of size 5×5 in the first two layers and 3×3 in the last three layers to the input. Then, the non-linear RELU function is applied. Finally, a pooling layer summarizes image regions and outputs a downsized version of the input. While the image height is downsized by 2 in each layer, feature maps (channels) are added, so that the output feature map (or sequence) has a size of 32×256.

• **RNN**: the feature sequence contains 256 features per time-step, the RNN propagates relevant information through this sequence. The popular Long Short-Term Memory (LSTM) implementation of RNNs is used, as it is able to propagate information through longer distances and provides more robust training-characteristics than vanilla RNN. The RNN output sequence is mapped to a matrix of size 32×80. The IAM dataset consists of 79 different characters, further one additional character is needed for the CTC operation (CTC blank label), therefore there are 80 entries for each of the 32 time-steps.

• **CTC**: while training the NN, the CTC is given the RNN output matrix and the ground truth text and it computes the **loss value**. While inferring, the CTC is only given the matrix and it decodes it into the **final text**. Both the ground truth text and the recognized text can be at most 32 characters long.

• **CNN output**: Fig. 3.2.1 shows the output of the CNN layers which is a sequence of length 32. Each entry contains 256 features. Of course, these features are further processed by the RNN layers, however, some features already show a high correlation with certain high-level properties of the input image: there are features which have a high correlation with characters (e.g. “e”), or with duplicate characters (e.g. “tt”), or with character-properties such as loops (as contained in handwritten “l”s or “e”s).

• **RNN output:** Fig. 3.2.2 shows a visualization of the RNN output matrix for an image containing the text “little”. The matrix shown in the top-most graph contains the scores for the characters including the CTC blank label as its last (80th) entry. The other matrix-entries, from top to bottom, correspond to the following characters: “ !”#&’()*+,-./0123456789:;?ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz”. It can be seen that most of the time, the characters are predicted exactly at the position they appear in the image (e.g. compare the position of the “i” in the image and in the graph). Only the last character “e” is not aligned. From the bottom-most graph showing the scores

for the characters “l”, “i”, “t”, “e” and the CTC blank label, the text can easily be decoded: I just take the most probable character from each time-step, this forms the so-called best path, then we throw away repeated characters and finally all blanks: “l---ii--t-t--l-…-e” → “l---i--t-t--l-…-e” → “little”.

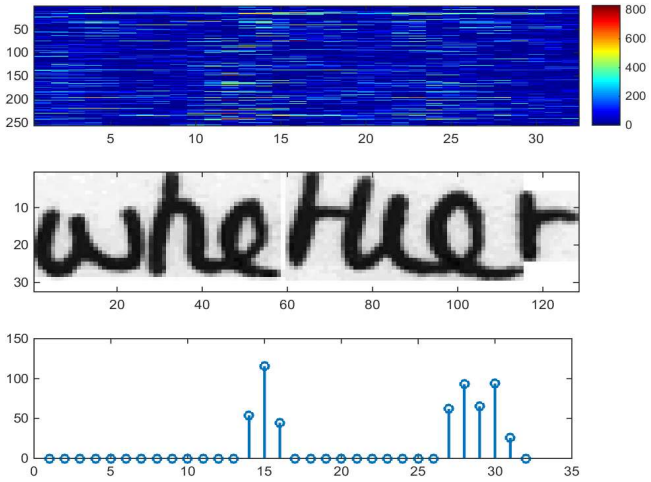

Figure 3.2.1 Top: 256 feature per time-step are computed by the CNN layers. Middle: input image. Bottom: plot of the 32nd feature, which has a high correlation with the occurrence of the character “e” in the image.

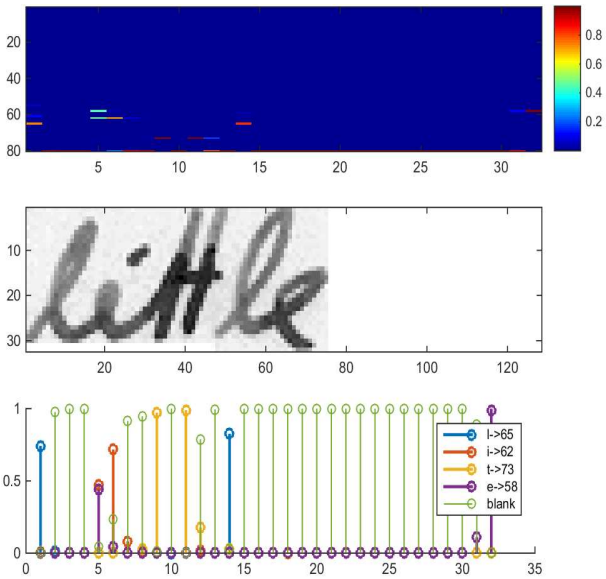

Figure 3.2.1 Top: output matrix of the RNN layers. Middle: input image. Bottom: Probabilities for the characters “l”, “i”, “t”, “e” and the CTC blank label.

3.3 Libraries

- Pandas
- Numpy
- Seaborn

- Matplotlib
- Sklearn
- Tensorflow
- Streamlit

3.4 Functional Modules

Implementation using TF

The implementation consists of 4 modules:

- **SamplePreprocessor.py**: prepares the images from the IAM dataset for the NN
- **DataLoader.py**: reads samples, puts them into batches and provides an iterator-interface to go through the data
- **Model.py**: creates the model as described above, loads and saves models, manages the TF sessions and provides an interface for training and inference
- **main.py**: puts all previously mentioned modules together

I only look at Model.py, as the other source files are concerned with basic file IO (DataLoader.py) and image processing (SamplePreprocessor.py).

Chapter 4

OBSERVATION AND RESULTS

4.1 Training

Train Test split

The mean of the loss values of the batch elements is used to train the NN: it is fed into an optimizer such as RMSProp. To train the data the following code is used:

```
optimizer = tf.train.RMSPropOptimizer(0.001).minimize(loss)
```

4.2 Results & Snapshots

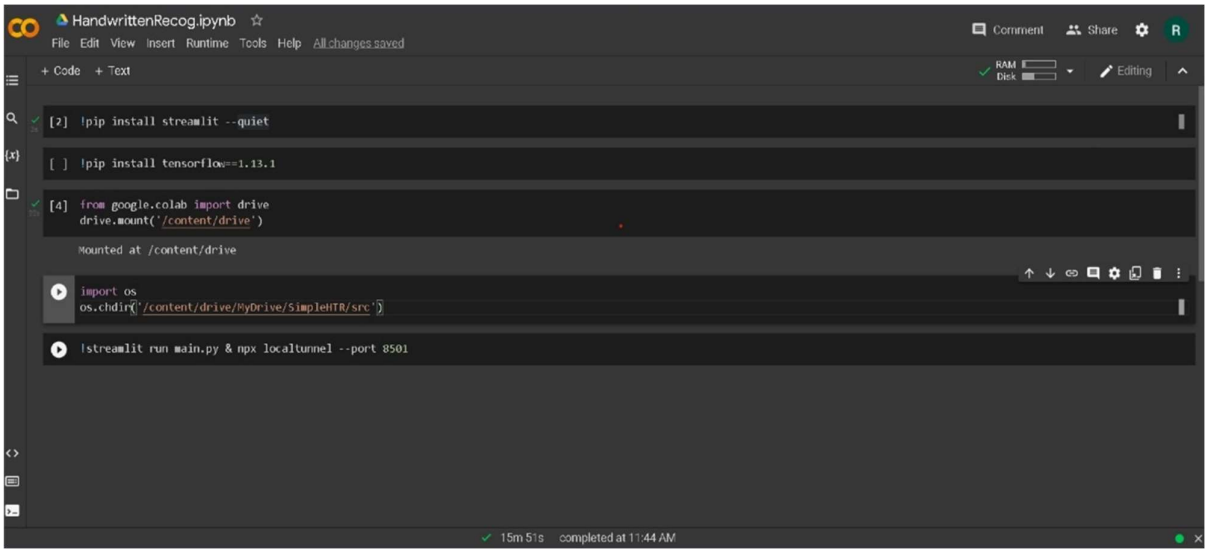

Figure 4.1 Commands

In the above diagram the commands do the following:

- First command will install the streamlit package in to the google colab ipynb file.
- Second command installs the tensorflow package of version 1.13.1 to the google colab ipynb file.
- Third and fourth line will link the google drive to the google colab.
- Sixth line will be used to open the particular directory in the google drive linked.
- Seventh line is used to run the local server to execute the program.

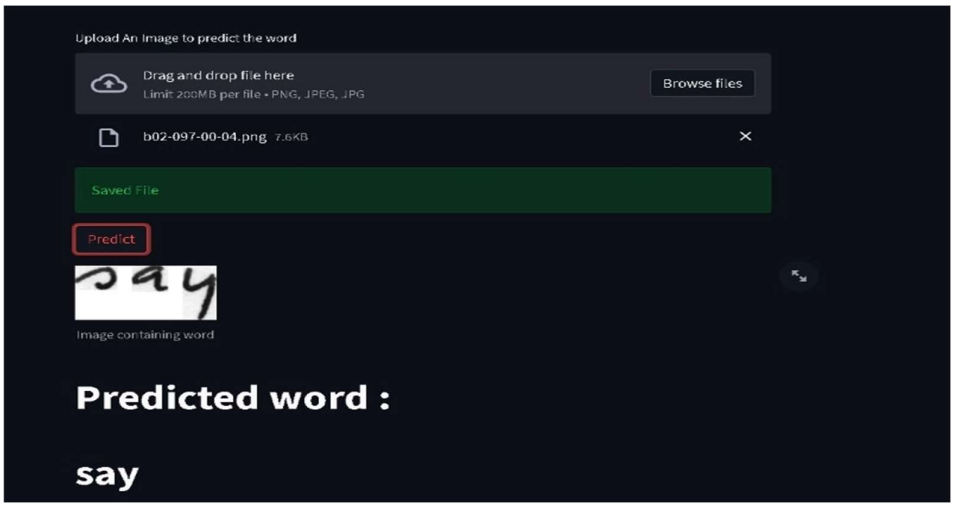

Figure 4.2 'Say' word prediction

In the above diagram the following steps takes place:

- First, drag and drop the word containing image or by browsing files.
- Once the image is uploaded it gets saved and then click on the predict button.
- The program will process the image and predict the word.

Chapter 5

CONCLUSION AND FUTURE ENHANCEMENT

5.1 Conclusion

In this project classification of characters takes place. The project is achieved through the conventional neural network. The accuracy obtained in this is above 90.3%. This algorithm will provide both the efficiency and effective result for the recognition. The project gives best accuracy for the text which has less noise. The accuracy completely depending on the dataset if the data is increased, I can get more accuracy. If I try to avoid cursive writing then also its best results.

In future I am planning to extend this study to a larger extent where different embedding models can be considered on large variety of the datasets. The future is completely based on technology no one will use the paper and pen for writing. In that scenario they used to write on touch pads so the inbuilt software which can automatically detects text which they are writing and convert into digital text so that the searching and understanding very much simplified.

5.2 Future Enhancement

- In case you want to feed complete text-lines as shown in Fig. 6 instead of word-images, you have to increase the input size of the NN.

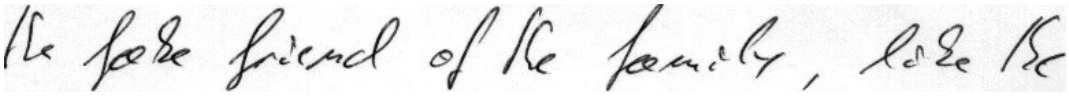

Fig.5.2: A complete text-line can be fed into the NN if its input size is increased (image taken from IAM).

If you want to improve the recognition accuracy, you can follow one of these hints:

- Data augmentation: increase dataset-size by applying further (random) transformations to the input images
- Remove cursive writing style in the input images (see DeslantImg)
- Increase input size (if input of NN is large enough, complete text-lines can be used)

  Decoder: use token beam search decoding to constrain the output to dictionary words
- Text correction: if the recognized word is not contained in a dictionary, search for the most similar one.

REFERENCES

[1] Fischer, A., Frinken, V., Bunke, H.: Hidden Markov models for off-line cursive handwriting recognition, in C.R. Rao (ed.): Handbook of Statistics 31, 421 – 442, Elsevier, 2013

[2] Frinken, V., Bunke, H.: Continuous handwritten script recognition, in Doermann, D., Tombre, K. (eds.): Handbook of Document Image Processing and Recognition, Springer Verlag, 2014

[3] S. Günter and H. Bunke. A new combination scheme for HMM-based classifiers and its application to handwriting recognition. In Proc. 16th Int. Conf. on Pattern Recognition, volume 2, pages 332–337. IEEE, 2002.

[4] U.-V. Marti and H. Bunke. Text line segmentation and word recognition in a system for general writer

[5] M. Liwicki and H. Bunke, "Iam-ondb - an on-line English sentence database acquired from the handwritten text on a whiteboard," in ICDAR, 2005

[6] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in Advances in neural information processing systems, 2009, pp. 545–552.

[7] Nafiz Arica, and Fatos T. Yarman-Vural, ―Optical Character Recognition for Cursive Handwriting,‖ IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.6, pp. 801-113, June 2002.