JNNCE Journal of Engineering & Management – A Peer Reviewed Bi-annual Journal
Available online @ https://jjem.jnnce.ac.in
ISSN: 2582-0079 (O)
https:// www.doi.org/10.37312/JJEM.2019.030206
Indexed in International Scientific Indexing (ISI)
Impact Factor Value: 1.025 for 2018-19
Volume: 3, Issue: 2
July - December 2019
Date of Publication: 12-30-2019

A Novel Approach Towards Implementation Of Optical Character Recognition Using LSTM And Adaptive Classifier

Ujwala B S
Assistant Professor,
Electronics and Communication
JNNCE,Shimoga-577204
ujwalaravi2004@jnnce.ac.in

Sumathi K
Assistant Professor,
Electronics and Communication
JNNCE, Shimoga – 577204
sumathik@jnnce.ac.in

Abstract

Optical character recognition (OCR) approach has been utilized in changing printed text into editable textual content.OCR using a neural network is a popular method in various applications and it is very useful .Text preprocessing and segmentation algorithm decides the accuracy of OCR. The recognition of optical characters is thought to be one of the earliest well known applications of Artificial Neural Networks. Fetching the text content from the image is difficult as a result of different length, fashion, orientation, complex heritage of photograph etc. The proposed work develops an OCR system using the artificial neural networks for recognition of characters and MNIST datasets. It presents a deep neural network for the character recognition issues. The design implemented presents the supportive achievement with higher certainty and with a smooth structure that can be hooked up on a ordinary PC. The proposed system is deployed using python and tesseract OCR tool adaptation pytesseract. It is now behind the main commercial engines in terms of its accuracy. Comparative analysis of image on different entropy showed that one can achieve the output near to perfection and the output will vary from one epoch to epoch. Thus the output of every level ought to be located and forced to OCR and the exceptional text output need to be taken into deliberation as the result.

1. Introduction

An ancient human invented the characters more than hundred decades ago to show symbols and objects in everyday lifestyles. These symbols and Characters were used due to the fact the support of the language to speak among community. There are many forms of languages and characters inside the global because specific communities of human beings have created one-of-a-kind characters and formed extraordinary languages. Characters are used to for phrases and sentences and it’s far critical to understand them to get a more awareness of the linguistic details. Characters seen in numerous types of item flare ups, e.g., files, propaganda forums, street signals and symptoms and product containers. To communicate with each other well, all these characters place an important role.

Figure (1) and (2) shows the printable and handwritten characters respectively containing different font size and orientation. The factors which affect the quality of images are firstly due to the more usage of smart phones usually the character images which are captured by cameras can introduce motion blur, geometry distortion. Light conditions also affect the quality of images. The other factors are background of the image, different font sizes, colors, aspects of text lines in the images.

Figure 1: Printable Characters with different font size and orientation.

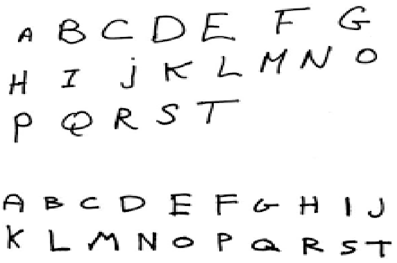


Figure 2: Handwritten Characters with different font size and orientation.

To rectify the above said problems many researchers have employed a deep learning method which is a subset of machine learning for character recognition of OCR. Machine learning is an application of artificial intelligence to learn and improve from experience without explicit programming. Deep learning is a method which performs both learning and classification. Deep learning will be finding wide range of applications in image classification, Text recognition, machine translation and face recognition. The main difference between CNN is that it introduces a deep network architecture, which can capture more discriminative information to infer the optimal solution. The two types of deep neural networks are the convolution neural network (CNN), and the recurrent neural network (RNN). One of the most popular Framework of deep learning is TensorFlow that was released in November 2015 by Google. It is an open source for numerical computation. Deep learning algorithms have been successfully implemented in the object recognition, character recognition field due to its excellent performance improvement. This proposed work will present some of the results using Tensorflow and Tessseract [6] for building Deep Neural Network (DNN) and CNN for solving the problem of OCR. It creates great opportunities to use the applications of deep learning using the Tensorflow and Tesseract powerful libraries to overcome the problems of character recognition.

Literature survey

There are various literatures are available on OCR. In this paper authors have discussed an implementation method for the handwritten character recognition that implements with regularization techniques using deep neural networks techniques dropout and batch normalization and using Tensor Flow framework.

In paper[1]Authors have implemented the algorithm of character recognition for Hand written digits in MNIST data set using batch normalization techniques and have found that increasing of drop out and addition of hidden layer along with ReLu activation function has improved the accuracy by 1%. In this project algorithm is implemented only on Gray scale images.

In the paper [2] authors have discussed about handwritten character recognition of Chinese characters using Convolutional Neural Networks. Authors have also shown that, Deeper the architecture of Convolutional Neural Networks and feature extraction methods, such as Gabor or gradient feature maps are used to enhance the performance of CNN.

In the preprocessing stage authors have normalized the collected database which is suitable to the design architecture of new networks. Authors have concluded that, HCCR-Gabor-GoogLeNet model shows superior results compared to all previous methods in the aspects of accuracy and storage performance.

This paper [3] presents a depth insight into the concept of fast training neural network. To partition the training data a preprocessing stage is used. Data is divided to smaller subsets based on the Euler number feature and symmetry property. In this paper an idea was proposed to reduce network training time. This technique works out well and results in higher recognition rate compared to the conventional neural network approach. Therefore this is suitable for the applications with limited training time constraint. The balance is achieved between the training time and recognition time can be achieved using the proposed method. The above said features are used to partition the data into several groups. The major problem solved in the proposed strategy is that exponential relationship between data and time is resolved and thus it is suitable for many real time applications.

In paper [4] authors have proposed a strategy to identify the text in a scene which is very challenging because of various interference factors. They have proposed a multi scale representation which helps in detecting the text in scene.

This representation consists of set of midlevel primitives, termed as strokelets, which capture the underlying substructures of characters at different granularities.. Here SVM and Random Forest method is used. Authors have worked on different datasets. This strategy can be applied for recognition of various object classes in full images.

This paper [5] is a combination of deep neural networks and recurrent neural networks. The key advantage of CRNN is that it runs directly with coarse level labels. In the training phase, detailed annotations are not required for each individual element. It repudiates fully connected layers used in conventional neural networks; it results in a much more efficient compact model. The above properties make CRNN an excellent approach for image based sequence recognition. In this paper the proposed algorithm performs well in the task of image-based music score recognition, which evidently verifies the generality of it. n. The author proposed the integration of sequence modeling, feature extraction and transcription into a unified framework. The proposed architecture consists of three important layers convolution layers, the recurrent layers, and a transcription layer, from bottom to top.

The convolution layers automatically perform feature sequence extraction from each input image. Then recurrent layer makes prediction of each frame of the feature sequence, outputted by the convolution layers. The transcription layer translates the per-frame predictions by the recurrent layers into a label sequence. The proposed algorithm is very efficient compared to conventional methods and other DCNN.

The author in this paper [6] has described an efficient implementation of image data into text conversion.They explained various steps required for text extraction from image file and to create a separate text file which consists of information extracted from image file. For image processing, the CV2 Open CV library is used where Python language is used as base and also for text extraction. After preprocessing stage, tessaract is used.To remove special characters from the text file and to convert lowercase characters to uppercase ASCII filtering is used.

In this paper [7] the author clearly differentiated and gives comparison study of optical character recognition services using Tesseract, Google, ABBYY Fine Reader, and Docs OCR. In future, scope is there to make an attempt to develop further OCR services using larger datasets, and to improve the performance parameters such as accuracy and reliability is required. This paper presents new insights on OCR in research area and assists researchers to develop more accurate, novel optical character recognition system using advanced image processing algorithms.

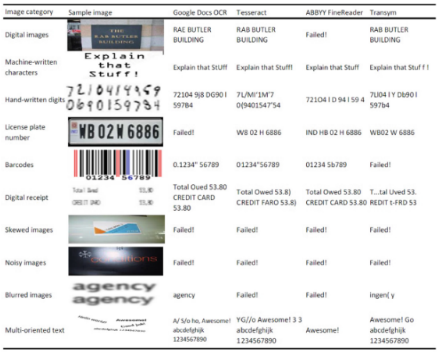


Image category	Sample image	Google Docs OCR	Tesseract	ABBYY FineReader	Transym
Digital images	THE RAB BUTLER BUILDING	RAE BUTLER BUILDING	RAB BUTLER BUILDING	Failed!	RAB BUTLER BUILDING
Machine-written characters	Explain that Stuff!	Explain that StUff	Explain that Stuff!	Explain that Stuff	Explain that Stuf f !
Hand-written digits	7210414959 0690159734	72104 9j8 DG90 I 597B4	7L/MI'1M'7 0(9401547'54	721O4 I D 94 I 59 4	7Ll04 I Y Db90 I 597b4
License plate number	WB 02 W 6886	Failed!	W8 02 H 6886	IND HB 02 H 6886	WB02 W 6886
Barcodes	01234 56789	0.1234" 56789	01234"56789	01234 5b789	Failed!
Digital receipt	[illegible]	Total Oued 53.80 CREDIT CARD 53.80	Total Owed 53.8) CREDIT FARO 53.8)	Total Owed 53.80 CREDIT CARD 53.80	T...tal Uved 53. REDIT t-FRD 53
Skewed images		Failed!	Failed!	Failed!	Failed!
Noisy images	conditions	Failed!	Failed!	Failed!	Failed!
Blurred images	agency agency	agency	Failed!	Failed!	ingen(y
Multi-oriented text	[illegible]	A/ S/o ho, Awesome! abcdefghijk 1234567890	YG//o Awesome! 3 3 abcdefghijk 1234567890	Awesome!	Awesome! Go abcdefghijk 1234567890

Figure 3: The qualitative visualization of four the OCR systems using some sample image dataset.

2. Tesseract

It is a open source software for optical character recognition. It is available in more than 30 languages. It was developed at HP in between 1984 to1994. It was improved in 1995 with good accuracy. In 2005, HP released Tesseract for open source. It provides less rejection. It is highly portable. Tesseract version 3.01 is released now and available for use. HP never used it.Google is developing and maintaining the Tessaract. Tesseract runs on Linux, Windows and Mac OS X; however, due to limited resources only Windows and Ubuntu are rigorously tested by developers. Tesseract only accepts TIFF images of simple one column text as inputs up to version 2. The early versions did not include layout analysis and so inputting multi-columned text, images produced a garbled output. Tesseract is the first OCR engine able to handle white-on-black text.

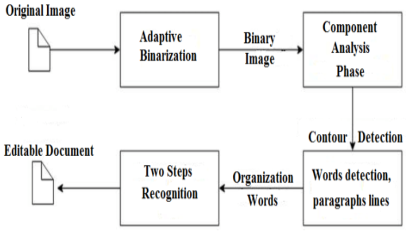


Figure 4: Architecture of Tesseract OCR engine

3. Optical Character Recognition using Tesseract

3.1 Overview

The proposed method for optical character recognition uses a custom designed neural network model using tesseract to classify the user input images as well as open source data sets. In the proposed system, tesseract system is integrated to python with the help of tesseract API library which is compatible with python, c and c++.

Learning process in artificial neurons networks involves three important steps such as acquisition of data, pre-processing step consists of locating, segmenting and normalizing representations and choices of attributes. Choices of attributes step is performed after the pre-processing and extract the attributes that define the data. These attributes serve as network entries of neurons. Before the processing of the data, make the choice of the objects, the definition of the attributes characterizing the objects and the construction of the base learning. At the end of this phase table of two inputs data and attributes of that data can be obtained. Learning process modifies the synaptic weights according to an equation called the learning equation. [8]

3.2 Implementation

Upload the image or image url as well as open source data with allowed extension (png, jpg, jpeg, gif) in to the web page created by the flask web server, then ocr the image ,if the uploaded image is not within allowed extension indicate upload error or if it is within allowed extension continue with the further steps. First step is to preprocess the image; this preprocessed image is feed to the LSTM network and classifier to detect the output.

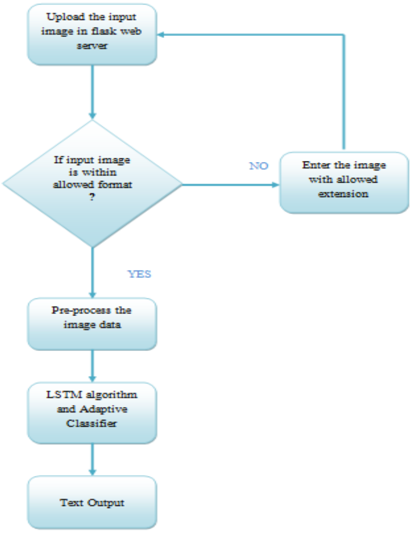

Upload the input image in flask web server
If input image is within allowed format ?
NO
Enter the image with allowed extension
YES
Pre-process the image data
LSTM algorithm and Adaptive Classifier
Text Output

Figure 3.3:Flowchart for image processing and obtaining a text

3.2(a) Pre-Processing

It is used to enhance the features of the image important for future processing, although geometric transformations of images (e.g. rotation, scaling, translation) it suppresses unwilling distortions of image data. Colour image does not help in identifying the importance information like edges of the image so it is importance to convert the RGB image into gray image. In the proposed system binary and Ostu thresholding is used. Thersholding is the simple method to segment the image. Otsu thersholding automatically finds an optimal threshold based on the observed distribution of pixel value. And to remove the noise in the image fastN1 mean denoising is used this helps to improve the image quality. This preprocessed image is feed to the algorithm for extracting the features and to adjust the weights and biases for the prediction of correct output.

3.2(b) Working of LSTM (Long Short Term Memory) algorithm

LSTM network is comprised of different memory blocks called cells. These are used to solve the long term storage problem. Repeating Cell of LSTM has four Neural Network Layers. The Cell state of LSTM allows information to pass through without being changed. The neural network layer of LSTM consists of sigmoid activation function and a point multiplication. The output of sigmoid activation function, varies from 0 to 1.'0' state indicates that not to pass any information and '1' state indicates the complete information to pass through. The below figure shows the block diagram of LSTM network. These networks do not allow any information to be manipulated in the cell state. LSTM cell has three gates Input gate, Forgot gate and output gate.

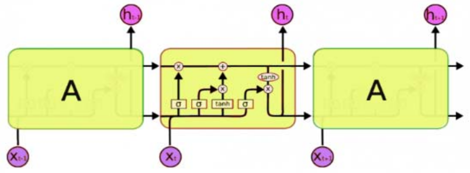


Figure 4.2 (b): Long Short Term Memory network

i. Forgot Gate

As the name indicates the forget gate decides whether the information should be kept or thrown away. It also indicates the proportion of data to be kept. Most widely logistic sigmoid function is used as activation function in LSTM networks. The output of previous hidden state and information from the current input is passed through the activation function. The output of the activation function lies between 0 and 1.If the value is closer to 1 indicates to keep and the value close to zero indicates to forget the information.

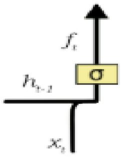


Figure (i): Forgot gate

i. Input Gate

The input gate is responsible for the addition of information to the cell state. This addition of information is as shown in figure.

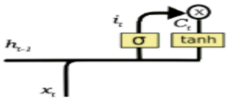


Figure (ii): Input Gate

- In the Input gate we pass previous output state and present state input into a sigmoid function.
- The sigmoid function squish the values between 0 and 1.The output of sigmoid function is passed through the tanh function. The tanh function outputs a vector containing all possible values that which sqinches outputs from -1 to +1.
- The sigmoid gate output to the tanh function output are multiplied and then adding the useful information to the cell state via addition operation.

Once this three-step process is done with, ensure that only that information is added to the cell state is important and is not redundant.

ii. Output gate

Output gate is responsible for selecting useful information from the current cell state and showing it out as an output. The functioning of an output gate can again be broken down to three steps:

- In the first step,The output of previous hidden state and current state input is passed into the sigmoid function.
- The new output is multiplied with the tanh function and sigmoid output. The output is new cell state.
- The new cell state and the new hidden state output is passed to the next step.

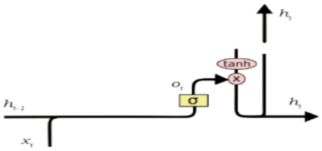


Figure (iii) : Output gate

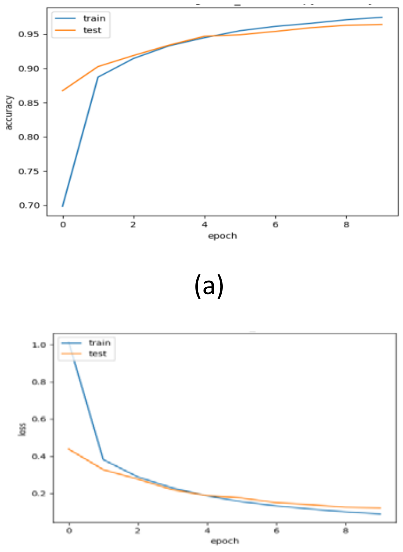
3.2(c) Adaptive Classifier

Adaptive classifier used to correctly classify the input image. Adaptive iterative process allows learning from the input training data. The features used in classification are the components of the polygonal approximation of the outline of a shape. Since the adaptive classifier learns during the first run, it can only make significantly less contribution near the top of the page if deployed during the first run. Therefore, a second pass is run over the page, in which words that were not recognized well enough in the first run are recognized again.

After the image is processed tesseract is used as the final step for OCR to get optimum output. Classifier correctly classifies the image to recognize the character. Tesseract supports many languages the purpose is to extract English text from the images which must be kept in the tess data folder. Tesseract will give the better accuracy for gray scale mode images compared to colored image.

4.Comparision

The proposed cross entropy method is compared with other entropy method and with error mean squared. Proposed method uses the binary cross entropy during validation all the input image with numbers are grouped as class one with different features corresponding to different numbers in this way the binary cross entropy is used in the work. The blue line indicates the training the dataset and orange line indicates the testing dataset. Below figure shows the categorical cross entropy accuracy and loss graph.Graph(a) shows during first epoch accuracy was less as epoch increases accuracy reaches maximum upto 95% Graph(b) shows the loss during testing and training here loss percentage decreases with increase with epoch. Using categorical cross entropy maximum accuracy achieved is 95% for 8epoch as shown in figure (5.1)



(a)

(b)

Figure 4.1: (a) CNN model with categorical cross entropy accuracy.

(b) CNN model with categorical cross entropy loss.

Binary cross entropy gives the constant output. Binary cross entropy measures how far away from the true value (which is either 0 or 1) the prediction is for each of the classes and then averages these class-wise errors to obtain the final loss. This cross enrtopy can achieve accuracy rate up to 98% as shown in figure (5.2).

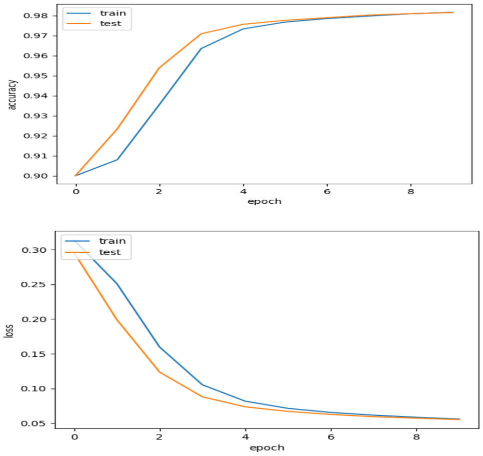
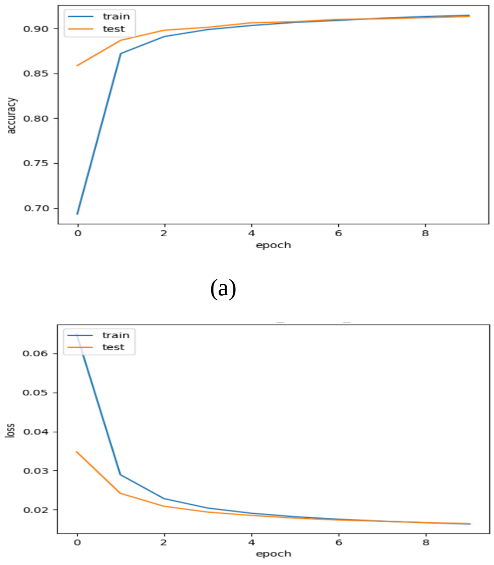


Figure 4.2: (a) CNN model with binary cross entropy accuracy. (b) CNN model with binary cross entropy loss.



(a)

(b)

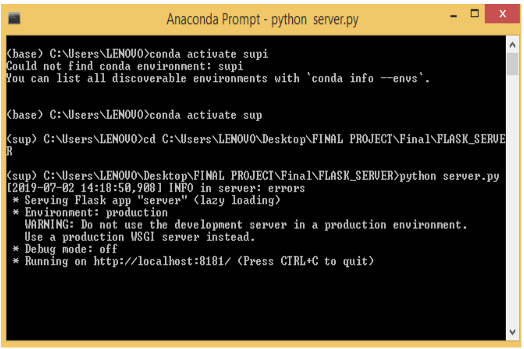
Figure 4.3: (a) NN model with mean squared error accuracy. (b) NN model with mean squared error loss.

The mean squared error (MSE) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors—that is, the average squared difference between the estimated values and what is estimated figure (4.3) shows the neural network model loss and accuracy. Comparing with different method proposed work shows the better accuracy with 98% for training and testing validation .

Results and Discussion

5.1 Results for OCR using Tesseract

The proposed work shows the resuts for RNN algorithm using Tesseract which are explained below.Here the web server page is created to display the output image instead of displaying in the prompt, it will store the process and deliver web page to clients.



```
Anaconda Prompt - python  server.py

(base) C:\Users\LENOVO>conda activate supi
Could not find conda environment: supi
You can list all discoverable environments with `conda info --envs`.


(base) C:\Users\LENOVO>conda activate sup

(sup) C:\Users\LENOVO>cd C:\Users\LENOVO\Desktop\FINAL PROJECT\Final\FLASK_SERVE
R

(sup) C:\Users\LENOVO\Desktop\FINAL PROJECT\Final\FLASK_SERVER>python server.py
[2019-07-02 14:18:50,908] INFO in server: errors
 * Serving Flask app "server" (lazy loading)
 * Environment: production
   WARNING: Do not use the development server in a production environment.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://localhost:8181/ (Press CTRL+C to quit)
```

Figure 5.1(a): Running server program in anaconda prompt.

A user interface was designed which options to select image file and its equivalent text had output was generated. The image files as well as the OCR output were simultaneously displayed as shown in figure 5.1(b).

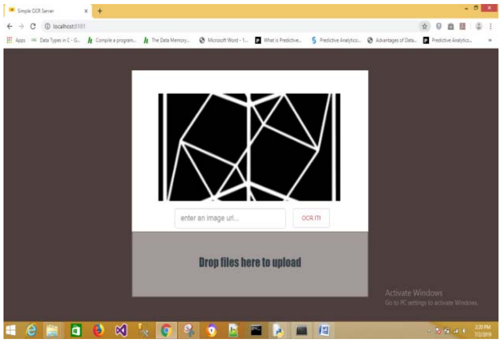
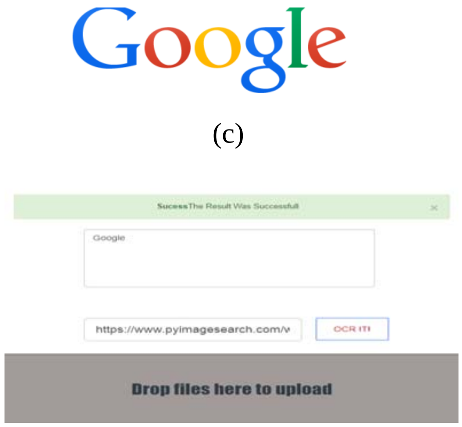


Figure 5.1(b): Web server page of Tesseract OCR engine.

Fig.5.1 (c) and (d) shows the extracted text output from sample image file. Here image URL is given as input. The top task bar shows that text image is successfully predicted which contains the text Google in the given URL for the image.



(c)

Sucess The Result Was Successfull

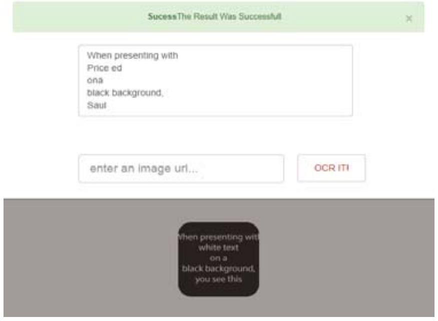
Google

https://www.pyimagesearch.com/w

(d)

Figure 5.1(c): Text input for the given URL (d)Predicting the output for the given url.

The bottom task bar shows the input file dropped which is to be detected once the key OCR IT ! is pressed it will start the process displayed in figure 5.1(e) and it also shows that image obtained is nearly equal the droped input file.

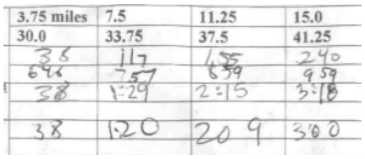

SucessThe Result Was Successfull

```
When presenting with
Price ed
ona
black background,
Saul
```

When presenting with
white text
on a
black background,
you see this

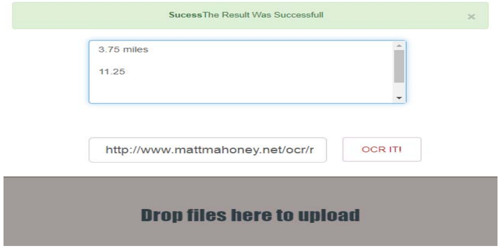
Figure 5.1 (e): Predicting the output for the uploaded image file.

To detect the number URL containing the number is feed has the input . The figure 5.1(g) shows that image obtained is nearly equal the droped input file.



3.75 miles	7.5	11.25	15.0
30.0	33.75	37.5	41.25
35	117	155	240
646	757	659	959
38	1:29	2:15	3:18
38	1:20	209	3:00

(f)


SucessThe Result Was Successfull

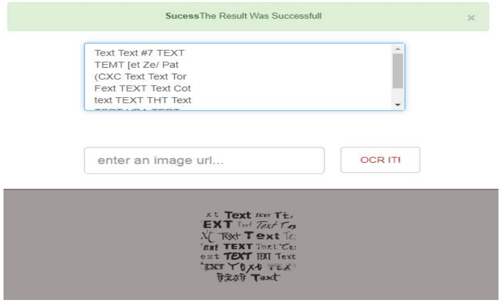
```
3.75 miles

11.25
```

http://www.mattmahoney.net/ocr/r

(g)

Figure 5.1(f):Data contained in the url.
(g) Predicting the output for the given url which contains number.

To detect the result for image with different font size and orientation following example is taken as the input which shows some text are detected correctly and for image with orientaion failed to get the satisfied output.


SucessThe Result Was Successfull

```
Text Text #7 TEXT
TEMT [et Ze/ Pat
(CXC Text Text Tor
Fext TEXT Text Cot
text TEXT THT Text
```

Figure 5.1(h): Predicting the output for the given data with different font size and oriented image.

6.Conclusion and Future Scope

6.1 Conclusion

Image processing is used to extract crucial information from the image. Image processing algorithm is applied first and those images are feed to different methodologies to get the accurate results. One of the methods used is CNN; processed image is feed to convolution layers to get the optimized output and another method is tesseract Tessaract is a good commercial engines in terms of improving accuracy.

The character is recognized effective and reliable manner using LSTM algorithms. Comparative analysis of image on different entropy showed that one can achieve the output near to perfection and the output will vary from one epoch to another. Hence the output of each stage is observed and subjected to OCR and the best output is considered for the result.

Future Work

In future an OCR system can be developed for multi lingual script in country like India where more languages are in use officially. There is future scope for an OCR system which efficiently recognizes manuscripts and cursive scripts. To reduce errors in OCR neuro-Fuzzy networks can be used to train the software.

References

[1] Khaled S. Younis, Abdullah . Alkhateeb A New Implementation of Deep Neural Networks for Optical Character Recognition and Face Recognition. The University of Jordan Computer Engineering Department, 1949.

[2] Haifeng Zhao, Yong and Jinxia Zhang "Character Recognition via a Compact Convolutional Neural Network", School of Software Engineerring, Jiniliing Institute of Technology, Hongjing Avenue, China.

[3] Huei-Yung and Chin-Yu Hsu ,"Optical Character Recognition with Fast Training Neural Network", Department of Electrical Engineering Advanced Institute of Manufacturing with High-Tech Innovation.

[4] Cong Yao, Xiang Bai_ Baoguang and Shi Wenyu," Strokelets: A Learned Multi-Scale Representation for Scene Text Recognition. Department of Electronics and Information Engineering, Huazhong University of Science and Technology, IEEE TRANSACTIONS ON IMAGE PROCESSING. volume 25, 6,JUNE , 2016

[5] Baoguang Shi, Xiang Bai, Senior Member, IEEE, and Cong Yao , "Short Papers:An End-to-End Trainable Neural Network for Image-Based Sequence Recognition", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,volume,39,11,NOVEMBER ",2017

[6] Rahul R. Palekar, Sushant U. Parab and Dhrumil P. Parikh, Member, IEEE, Prof. Vijaya N. Kamble,"Real Time License Plate Detection Using OpenCV and Tesseract", International Conference on Communication and Signal Processing vol-96, no-6,april-2017.

[7] Ahmad P. Tafti, Ahmadreza Baghaie2, Mehdi Assefi, Hamid R. Arabnia,Zeyun Yu, and Peggy Peissig, OCR as a Service: An Experimental Evaluation of Google Docs OCR, Tesseract, ABBYY,FineReader, and Transym

[8] Chirag Patel - Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA) Charotar University of Science and Technology(CHARUSAT), Atul Patel, PhD. Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA) Charotar University of Science and Technology(CHARUSAT), Dharmendra Patel -Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA) Charotar University of Science and Technology(CHARUSAT),"Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study", *International Journal of Computer Applications (0975 – 8887) Volume 55– No.10, October 2012*