

Optical Character Recognition

Assignment

— Vedha Krishna Yarasuri

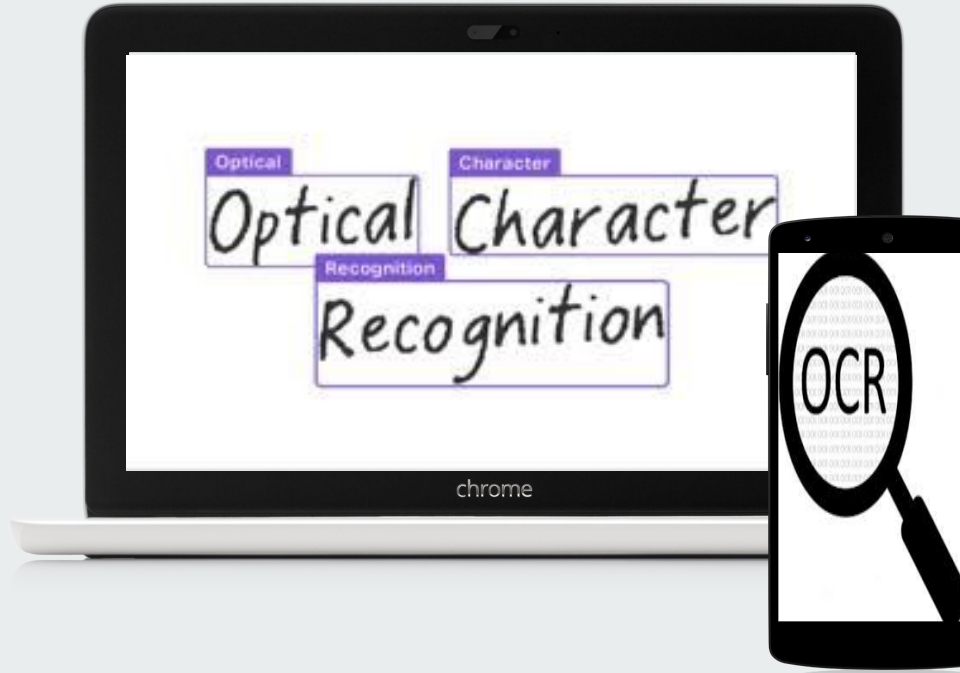


Table Of Contents

Introduction to OCR

Various Approaches

My Approach

Dataset

Data Augmentation Techniques

Feature Extraction Results

Sequence Models

Hyperparameters

Training, Saving Evaluating

Web App Interface

Challenges Faced

Conclusion



Introduction To OCR

— The Problem Statement

Optical Character Recognition

Optical Character Recognition (OCR) systems mainly focus on extracting text from scanned documents.[1] Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo)

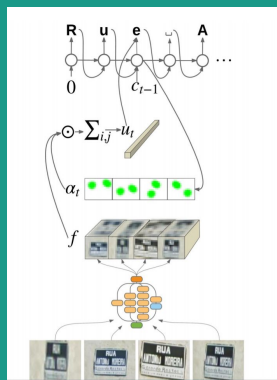
A close-up photograph of a person's hand holding a white handheld scanner over a printed document. The scanner's sensor is positioned over a line of text. The document text is partially visible and includes the heading 'Optical character recognition' and a paragraph explaining the process of converting images of typed text into a searchable digital format. The lighting is bright, and the focus is on the scanner and the text it is reading.

LEROY-MERLIN POLSKA SP. Z O.O.	
ul. Targowa 72 03-734 Warszawa	
SKLEP LEROYMERLIN GTWICE	
A.I.J.Nowska Jeziorskańska 3	
Nr rej	DO-00003773 Tel: (032)30-08-80
NIP 113-00-89-950	
nr wyd.: 127273	
PARAGON FISKALNY	
43826671 SZYNIA DO MONTAZU SZAPEK 200 CM	19,80 19,80 A
43826713 ZAWIESZKA DO SZAPEK TWORZYWO B	7,80 7,80 A
43826713 ZAWIESZKA DO SZAPEK TWORZYWO B	7,80 7,80 A
40479355 TYP D WKRETY UNIWERSALNE	28,06 28,06 1,01 A
Sprzed opod. PTU A 36,41	
Kwota A 23,00%	5,81
Podatek PTU	5,81
SUMA PLN 36,41	
000594001 Serwisant	16:15
39CKE-YUSVL-VOMCH-KJ6BF-T23DA	
ZEP BAE 09235164	
Platność KARTA BANKU 36,41	
Nr transakcji	022-005-000006 4114
Karta DOM	2046026857936
DOM - Twoje punkty naliczymy jutro	
ZAMONIENIA NA TELEFON	32/300 88 39
ZAPRASZAMY PONOWNIE	
www.leroymerlin.pl	

Various Approaches

1 .The First approach is by using Attention Mechanism

Wojna Z et al has proposed a model with 84.2% accuracy for recognition of textual data from images using French street dataset. They have trained a model which takes various versions of an image and plots attention masks on the image where the text is available and predicts text in the image.



2 .The Second method for extracting text from an image is by using Connectionist temporal classification (CTC).

It can be used for tasks like on-line handwriting recognition or recognizing phonemes in speech audio. CTC refers to the outputs and scoring, and is independent of the underlying neural network structure.

3 .The Third method is by using image processing techniques



My Approach

The Attention Based Approach

Model Architecture

Feature Extraction

The features of the given image are extracted using pre-Trained Inception Resnet V2 model. The features obtained are then stored using numpy.

Sequence Model

The features are given to an encoder (conv net) and then passed through attention model (neural net) to compute attention weights (context). These attention weights are sent to a decoder to decode the text embedded in the image.



The Dataset

IIIT Hyderabad's Dataset

The IIIT 5K dataset contains 5000 cropped word images from Scene Texts and born-digital images.

Few Instances of the images are shown in the following slide.

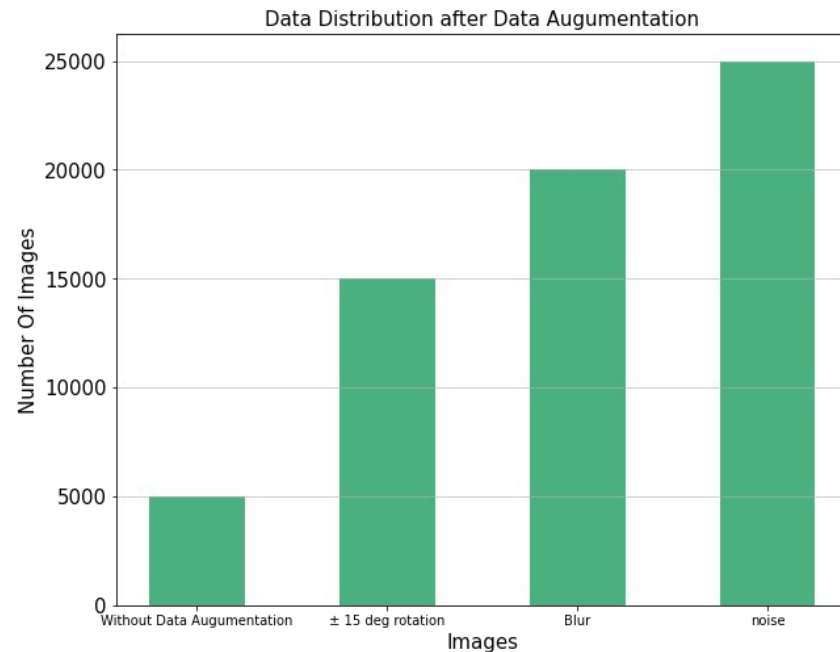
The Dataset



Data Augmentation

Data Augmentation Techniques :

- 1 . Rotating images ± 15 deg
- 2 . Image Blur using PIL
- 3 . Uniform Noise to the Images



After Data Augmentation



Column 1 : Normal Image

Column 2 : + 15 Deg Rotated Image



Column 3 : - 15 Deg Rotated Image



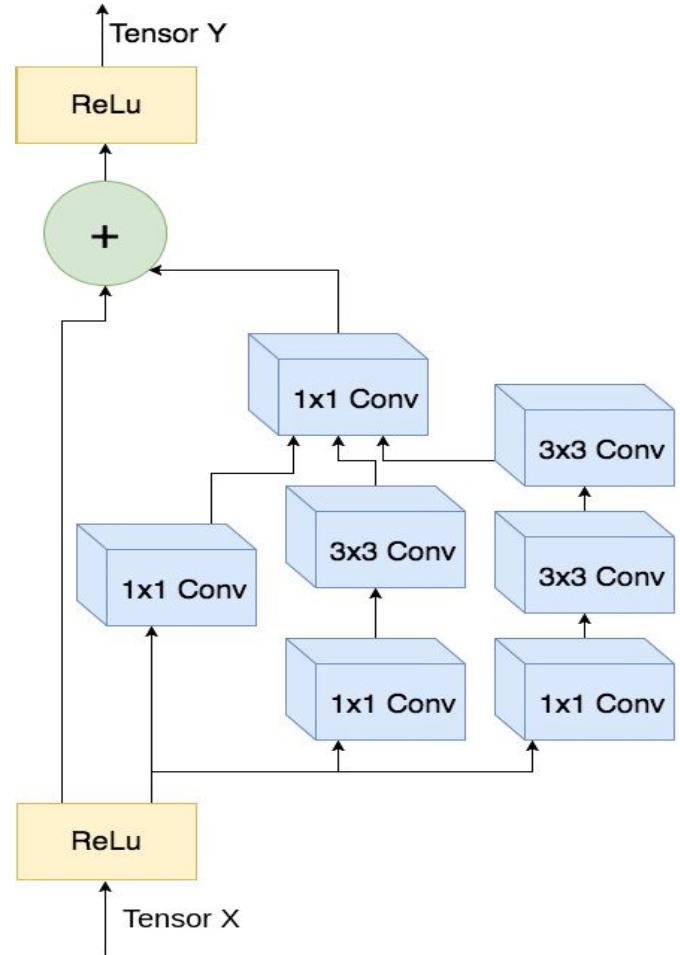
Column 4 : Image Blur

Column 5 : Image Noise

Feature Extractor

– Inception Resnet V2

- The inputs of the dataset are given to Inception Resnet model for extracting features at the layer “block8 1 conv”.
- The Layer “mixed 7a” and “block8 1 conv” nearly gives similar results.

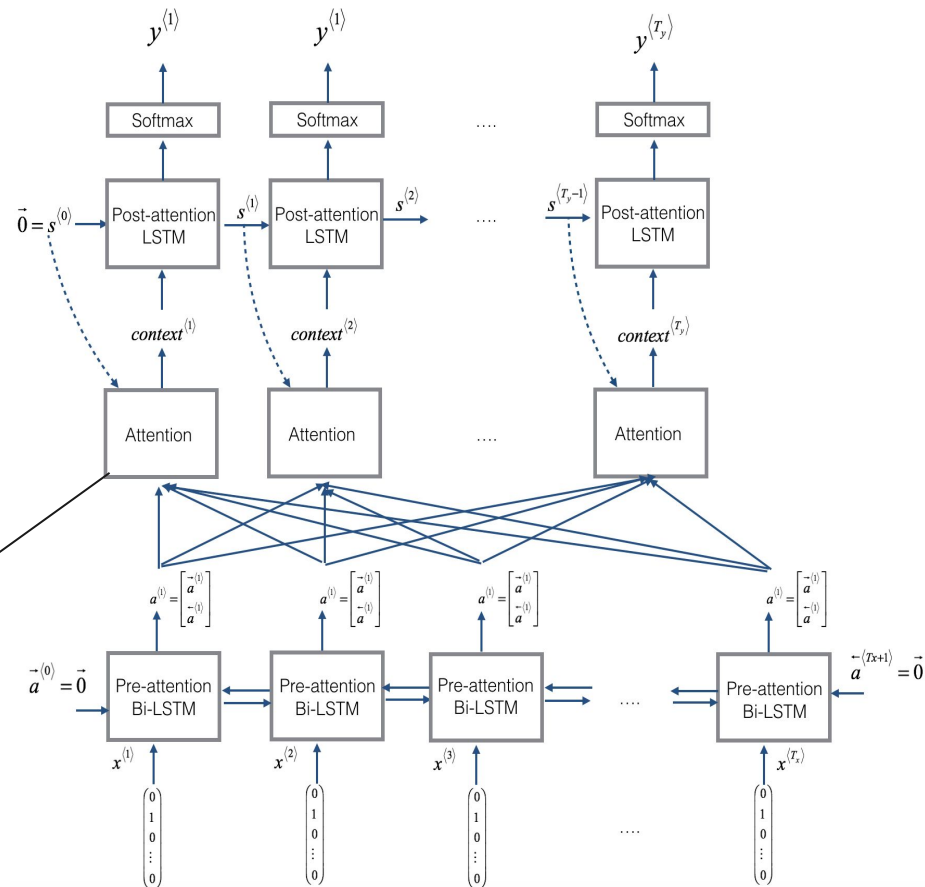
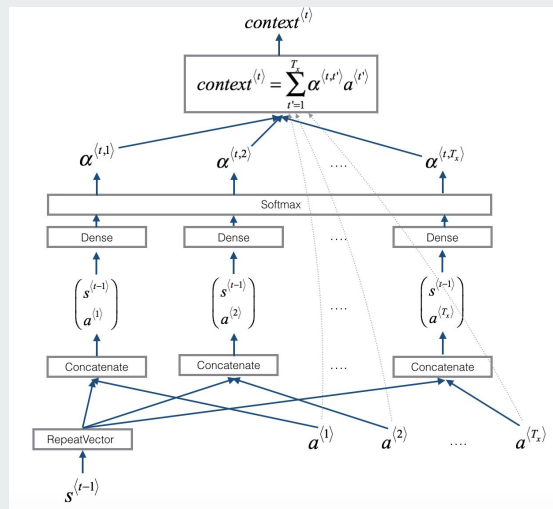


Features Extracted From “block8_1_conv”



Features Extracted from the image labelled “YOU”

Attention Based Sequence Models





Attention Based Sequence Model

Model Architecture

Encoder


A Shallow Neural Network with 256 units, outputs the activations of the encoder.

Attention

A Deep neural network with 1 hidden layer sandwiched between input and output layers. Input and hidden layer contains 512 units.

Decoder

Contains an embedding layer, GRU layer and followed by 2 deep layers. GRU and 1st deep layer contains 512 units. The 2nd deep layer contains 41 (Vocab size) units.

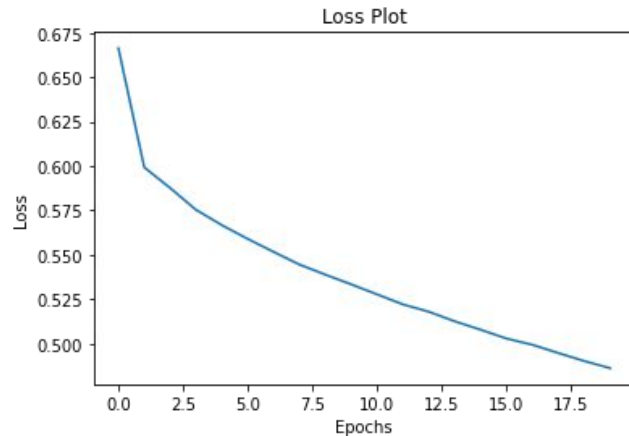
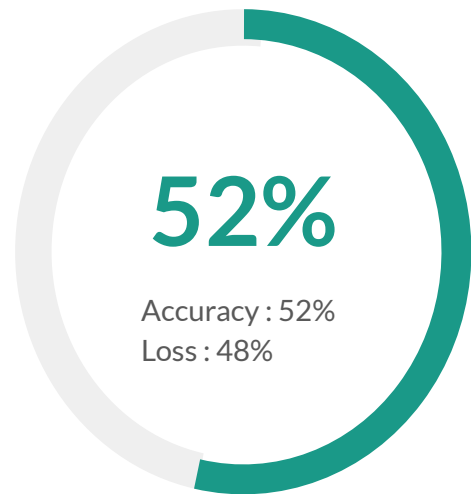


Hyperparameters

Hyperparameter Name	Value
BATCH_SIZE	64
BUFFER_SIZE	1000
embedding_dim	256
units	512
vocab_size	41
num_steps	609
features_shape	2080
attention_features_shape	64



Training, Checkpoints, Saving The Weights





Web App Design

Web Application has been designed using :

- 1 . Node.js
- 2 . Bootstrap

Hey, There 🙋

You need to upload a picture of sample form.

Upload a textual image and get its textual readings in it.

Choose file

Browse

Upload

Working


When the user uploads then the image will be sent to the server using REST api and gets downloaded using multer.

Later The steps remains same.

Hey, There 🤖

You need to upload a picture of sample form.

Upload a textual image and get its textual readings in it.





Challenges Faced and Solutions

My Challenges and Solutions

1. Lack of Experience and Practice
-- Time Management and Hard working
2. Designing Sequence Models
-- TF Image Caption Example
3. Tensorflow.js -- Learning
4. Shell Scripting -- Learning

Thank You

