Handwritten Character Recognition

By: TOFIK ALI

17 December, 2020

Introduction

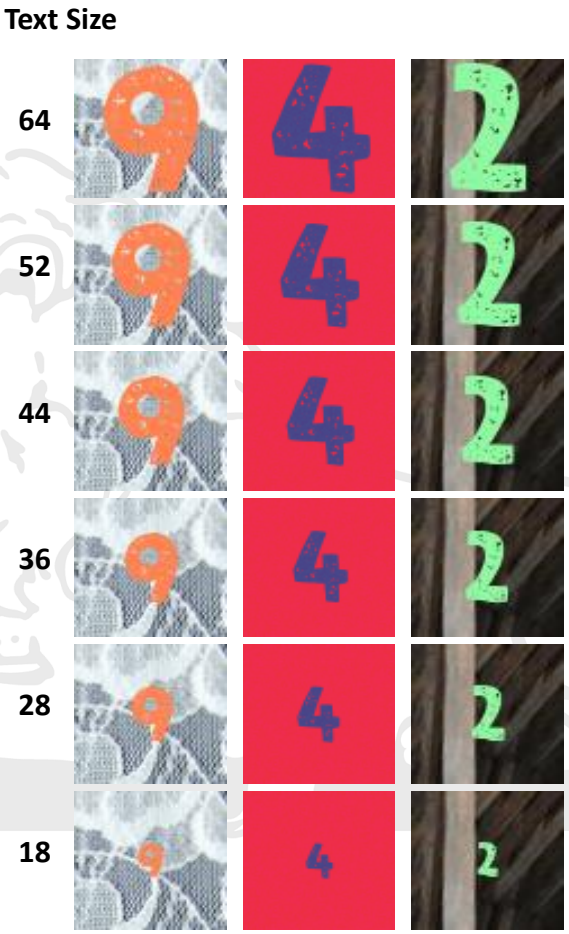
- What is a Character?

A scripting language has a predefined collection of symbols.
A character is a member of that collection set.

It is formed by some predefined different arrangement of line strokes.

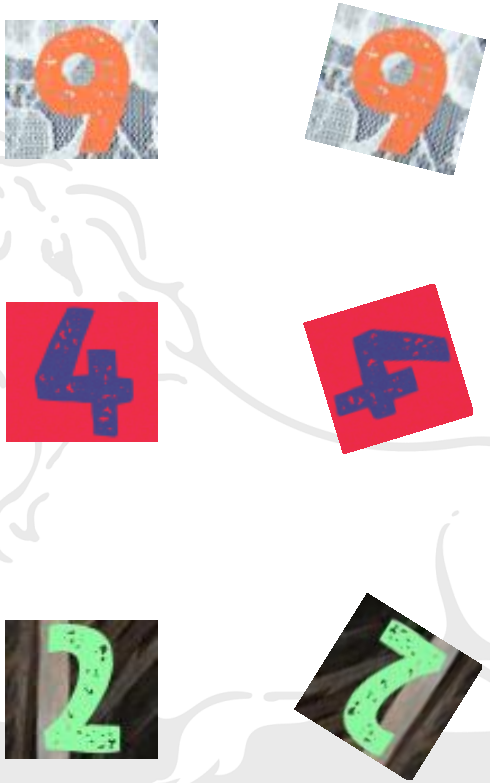
Issues

- Scale



Issues

- Orientation

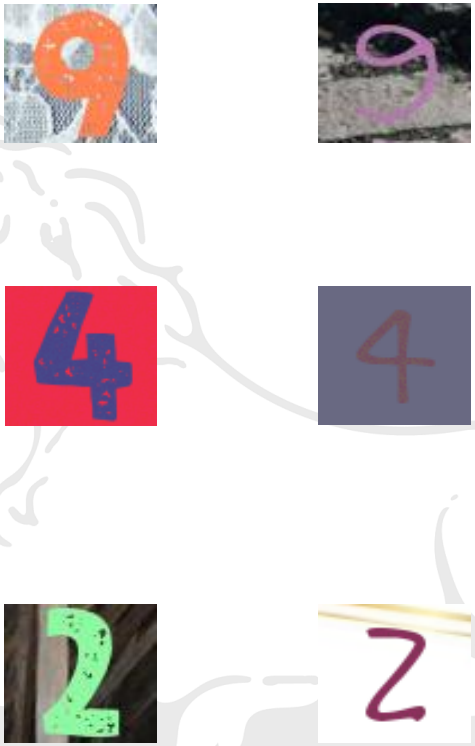


Issues

- Aspect Ratio

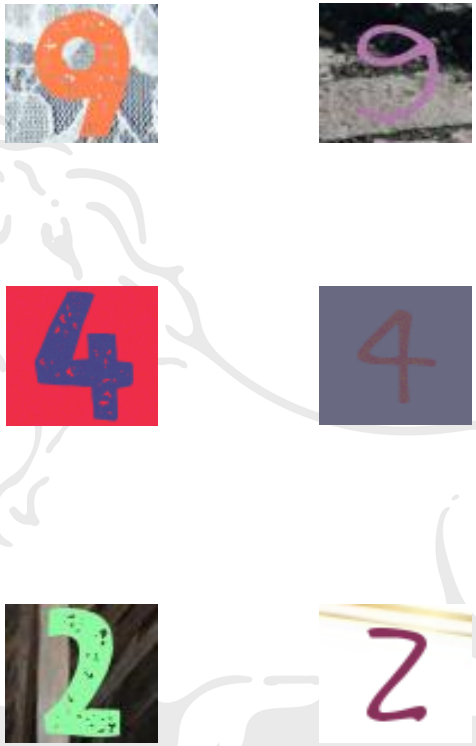
Issues

- Foreground, Background and Style



Issues

- Foreground, Background and Style



Issues

Challenges in Handwritten word Recognition

I. Huge variability and ambiguity of strokes from person to person.

II. Handwriting style of an individual person also varies time to time and is inconsistent.

III. Poor quality of the source document/image (physical source) due to degradation over time.

IV. Text in printed documents sit in a straight line whereas humans need not write a line of text in a straight line on white paper.

V. Cursive handwriting makes separation and recognition of characters challenging.

VI. Text in handwriting can have variable rotation to the right which is in contrast to printed text where all the text sits up straight.

VII. Collection of a good labelled dataset for training is time consuming and costly.

Examples

a

b

c

d

e

f

Solutions

Here we are only covering the Neural Network based methods as they are current state-of-the-art work.

I. Simple Feedforward Neural Network

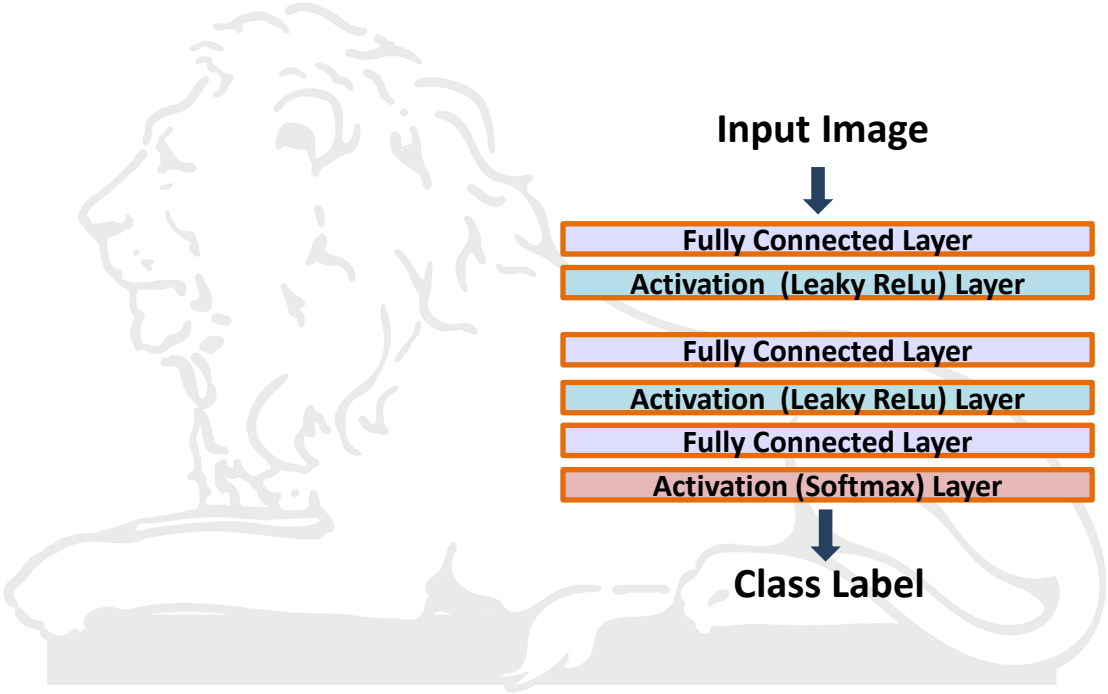
II. Convolution Neural Network

III. Convolution + Recurrent Neural Network for word/text-line recognition

The respective code is available at my GitHub
https://github.com/tali7c/ML-classifications-with-python

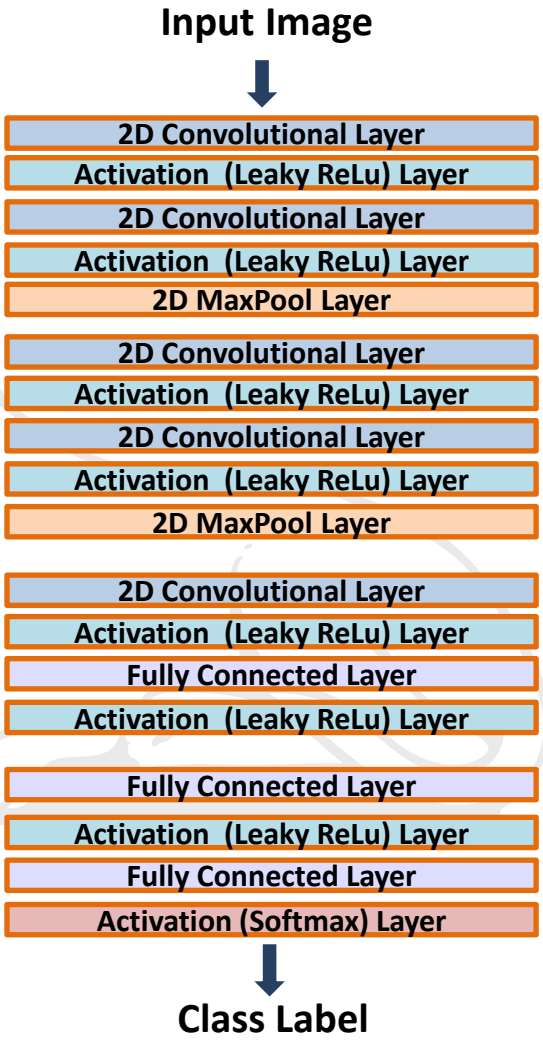
Solutions

Simple Feedforward Neural Network



Solutions

Convolution Neural Network



Solutions

Convolution + Recurrent Neural Network for word/text-line recognition

The dataset can be downloaded from
https://drive.google.com/drive/folders/1ZiGxk6ZN5IBjtNAI4JkJK9YFehS8KNmT?usp=sharing

Thank you