#### **INDIAN INSTITUTE OF TECHNOLOGY ROORKEE**



# **Handwritten Character Recognition**



By: TOFIK ALI

17 December, 2020



### Introduction



What is a Character?

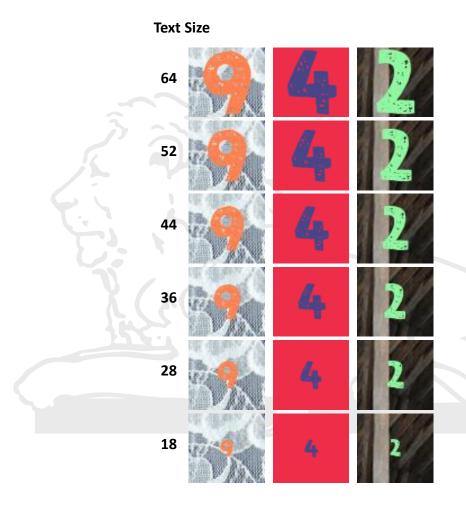
A scripting language has a predefined collection of symbols.

A character is a member of that collection set.

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



Scale



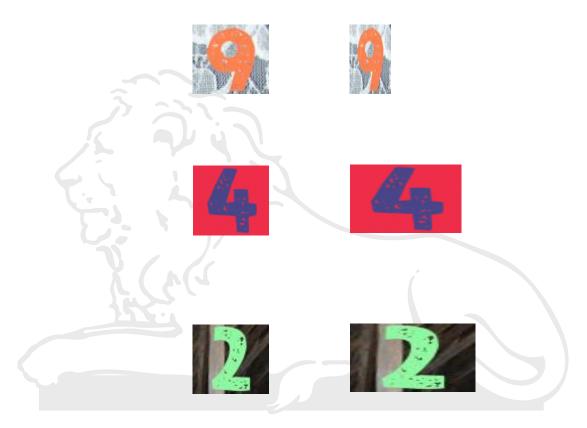


Orientation





Aspect Ratio





• Foreground, Background and Style





• Foreground, Background and Style





#### **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
Bradesco  Brades	DONALD E. KNUTH COMPUTER SCIENCE DEPARTMENT WHISTOPH STANFORD UNIVERSITY STANFORD CA 64305-6645  ENTE 29 Apr 07  ENTO THE BUSHAMPING Chose  S 10.24  LUTHER BURBANK SAVINGS  FROM L(G)12, 175P, Pijel Clai
С	d
Midland Bank  Street  Limited  Openford Kent DA11DG  Pay  S. L. T.S. Utol  Or order  F7-56  H MARRIS  V JOCO000000756.	عدر است الميحارو موست ( مندوق و مر المست عدد 170 مندوق و مر المست عدد 170 مندوق و مر المست عدد 170 مندوق و مر المست عدد مرد منهاه من المرد منهاه من المرد منهاه من المرد منهاه من مرد منهاه من مرد منهاه من مرد منهاه من مرد منهاه من من مرد من
e	f
中國工方集行深圳市分行現金支票 1060337451  ***********************************	Payable at any branch of RB  PAY अर्थिया कृष्णकृत्या भोगवन  RUPEES को भाग भागीया कृष्णकृत्या भागीयान  A-No 0123456789  RESEARCH BANK RB  PUNE-37  #874551# 000012400# 1572244# 80



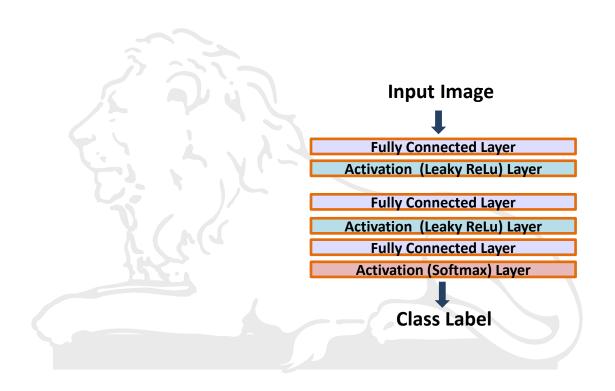
Here we are only covering the Neural Network based methods as they are current state-of-the-art work.

- 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



Simple Feedforward Neural Network





#### **Convolution Neural Network**





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