



Polytechnic University of Tirana

Department of Computer Engineering

SIBORA THEODHOR

ELINDA KAJO MECE

Computer Vision

OCR AND BEYOND



THE PRESENTATION IS ORGANISED IN 3 PARTS:



Introduction, previous work on OCR

Idea for the program, OCR
introduction, demonstration and
functionalities



Software Engineering aspect of the program

Designing, development,
implementation, testing



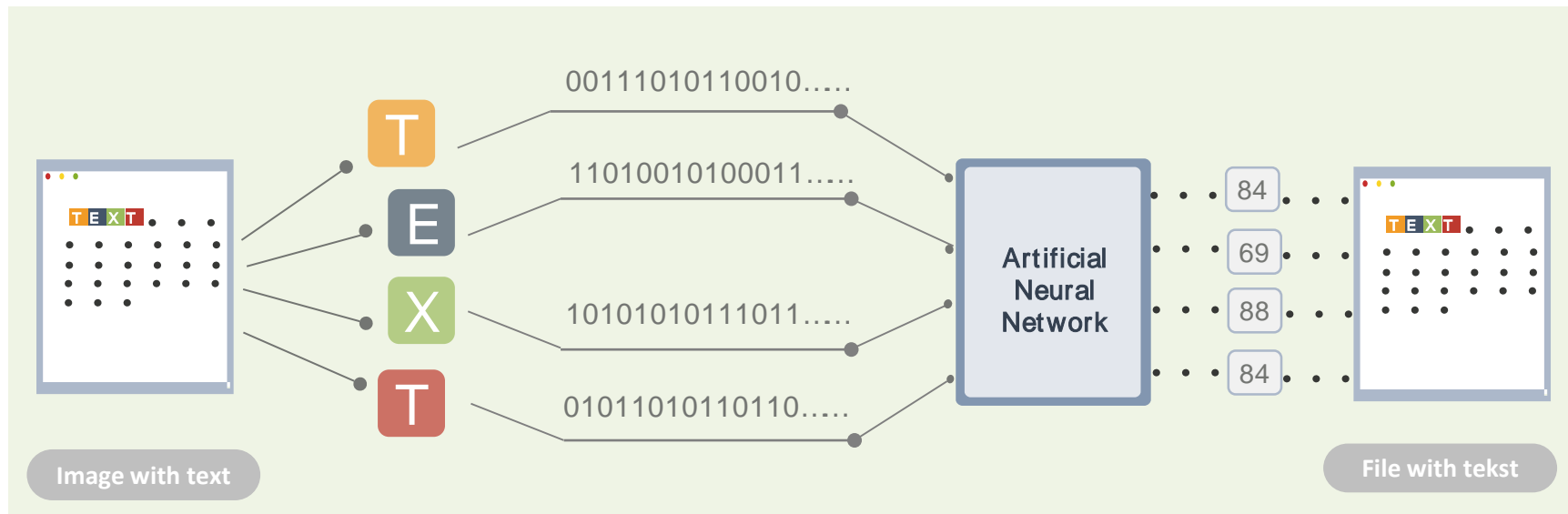
Computer Vision, Machine learning

Introduction to software
engineering approaches in
machine learning

Previous work

SE aspect

Computer vision



Previous work

SE aspect

Computer vision

IDEA & technologies



Developing an app that captures an image with text in albanian language and translates it to english.



Technologies: Android and OCR (Optical Character Recognition)



Services: Hp OCR API & Mymemory Translate API

Previous work

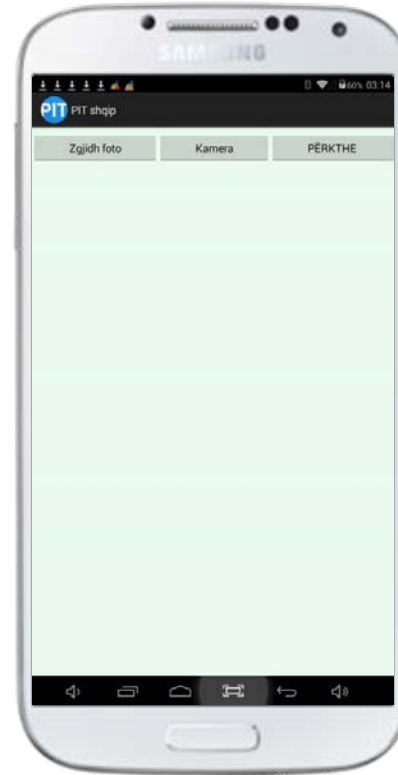
SE aspect

Computer vision

Demonstration

1

Capturing or add
from album.



Previouswork

SE aspect

Computer vision

Demonstration

1

Capturing or add
from album.

2

Image scanning
and Ocr



Previouswork

SE aspect

Computer vision

Demonstration

1

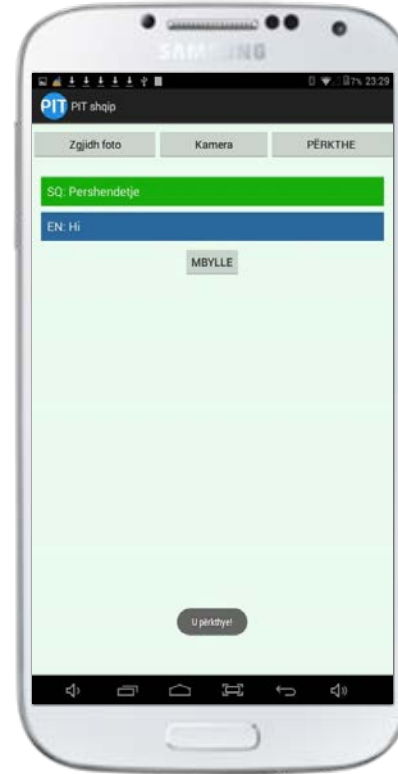
Capturing or add from album.

2

Image scanning and Ocr

3

Conversion from image to char and translation



Previous work

SE aspect

Computer vision



Previous work

SE aspect

Computer vision

App functionality

Flow diagram

Open the application



Select an existing picture from album



OR



Capture a picture from camera

The image is displayed on the screen and it's path is saved in a variable



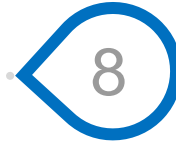
HP OCR API is called with Http protocol (POST request with the picture)

Server returns jobID string which we send in Http GET request



Server returns the text OCR recognised

MyMemory translate API is called (with the text as parameter)



MyMemory translate API returns the translated text



Display translated text

| PLANNING

| DEVELOPING

| VALIDATION

| TESTING

Previous work

SE aspect

Computer vision

Zooming in to OCR: Software engineering in machine learning practices

Developing software versus developing practices and processes

Previous work

SE aspect

Computer vision

Observation from a software engineering perspective

AI solutions have their own processes that are somewhat different from more traditional Software engineering.

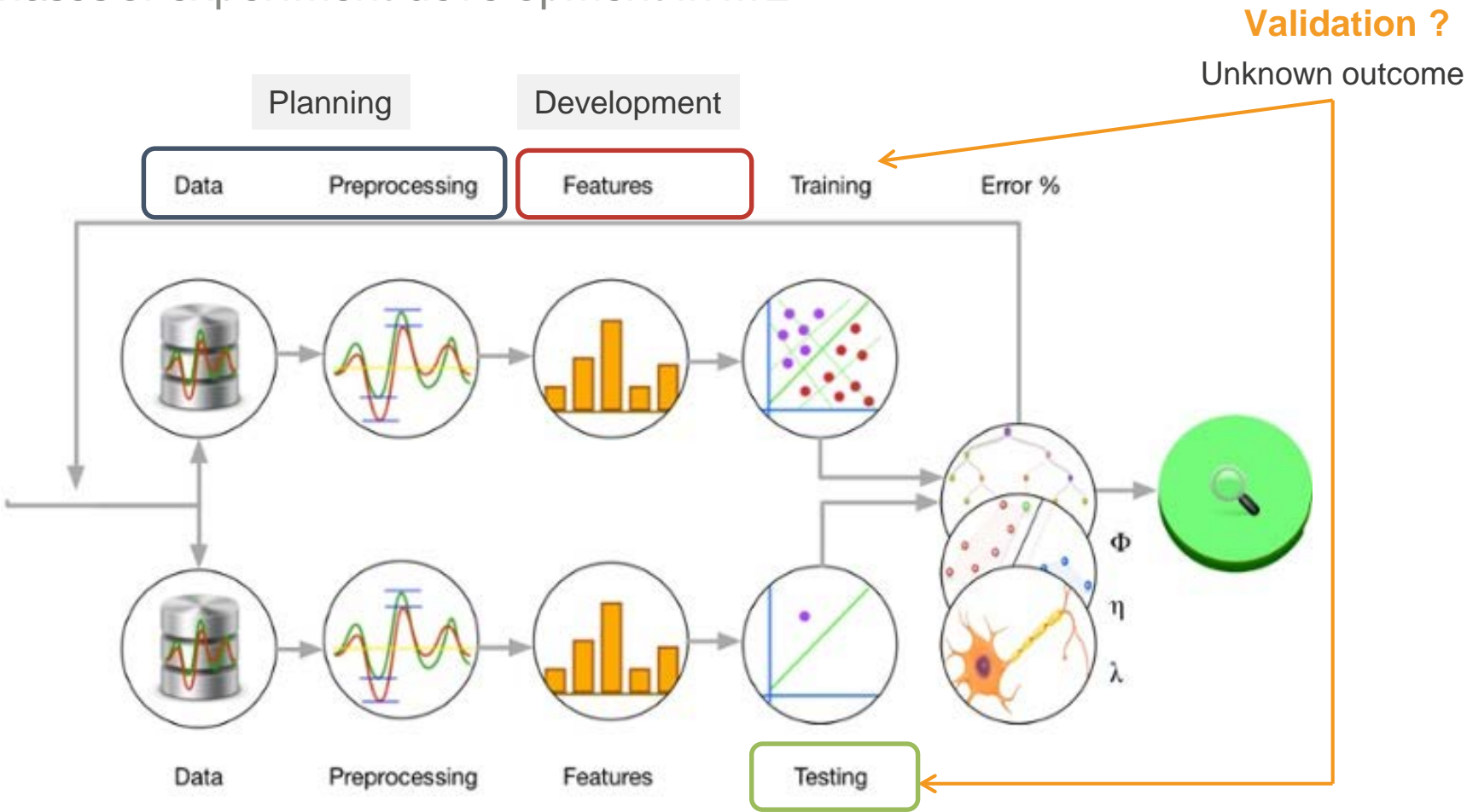
Two tutorials on building character recognition through the algorithms and capabilities in Azure Machine Learning and the other one through Tensor Flow.

Previous work

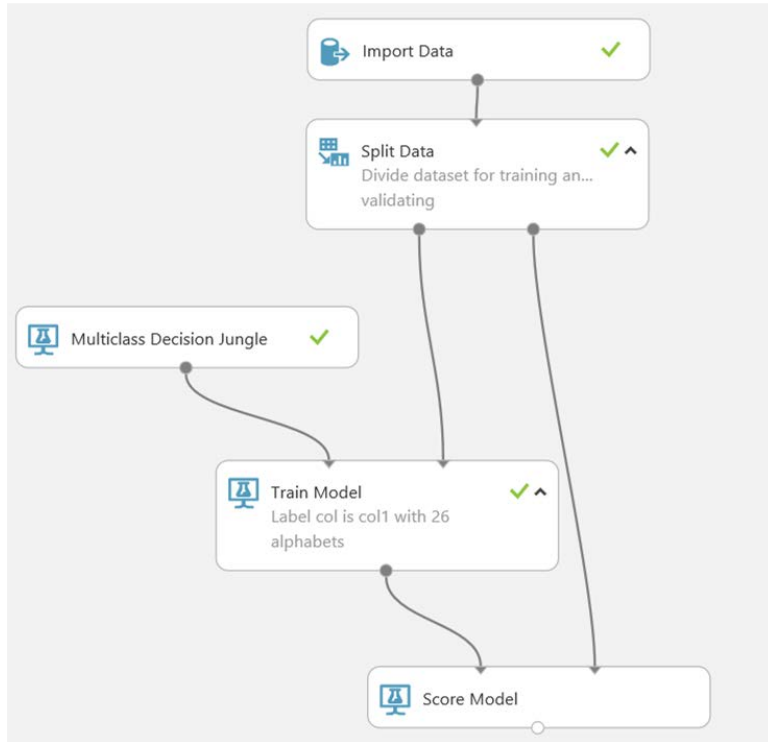
SE aspect

Computer vision

Phases of experiment development in ML



Character recognition using Multiclass Decision Jungle - Experiment



Set of 20,000 unique letter images generated by randomly distorting pixel and preprocessed.

First the data is split 50% to train the model and the other 50 % for testing.

Multiclass Decision Jungle is the algorithm that does the training and the classification.

After training, the model is used to score the other 50 % of the data and see how well the model predicted and classified new data .

Deep learning approach

Building digit recognition in tensor flow

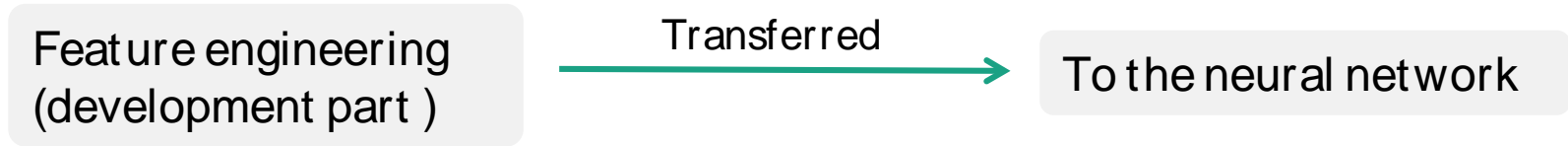


Previous work

SE aspect

Computer vision

Another alternative to doing ML is using deep learning and neural networks.



🕒 Time allocation now shifts to modeling the neural network.



Data wrangling

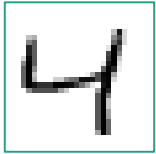
5 0 4 1

MNIST - database of handwritten digits used for training and testing

MNIST Data is split for training (55,000) , testing(10,000) and validation (5,000).

28 px

28 px



"x"

4 (label)

"y"



28

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .6 & .8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .5 & 1 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .9 & 1 & .1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .3 & 1 & .1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Previous work

SE aspect

Computer vision

Data wrangling

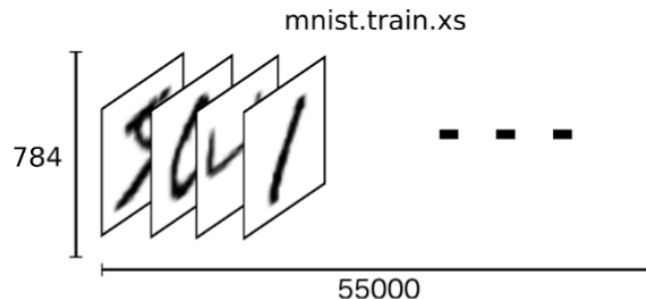
Each image is 28 pixels by 28 pixels which can be interpreted as a big array of numbers.

We can flatten this array into a vector of $28 \times 28 = 784$ numbers



MNIST images are just a bunch of points in a 784-dimensional vector space

The result in our case is a tensor with a shape of $[55000, 784]$.

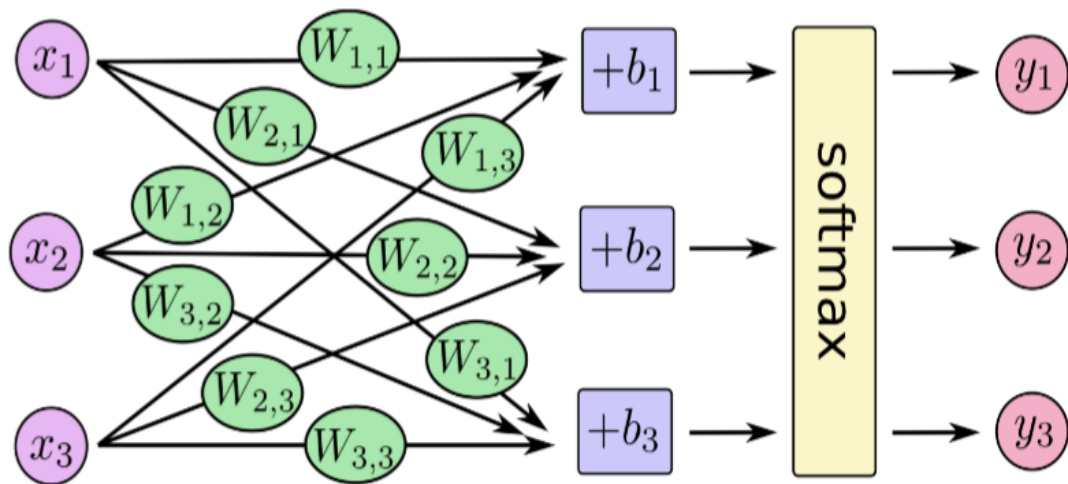


Previous work

SE aspect

Computer vision

Engineering the neural network



Softmax regression – model used to assign probabilities to an object being one of several different things, it gives a list of values between 0 and 1 that add up to 1.

It has two steps: first we add up the evidence of our input being in certain classes, and then we convert that evidence into probabilities.

To tally up the evidence that a given image is in a particular class, we do a weighted sum of the pixel intensities :

The weight is negative if that pixel having a high intensity is evidence against the image being in that class, and positive if it is evidence in favor.

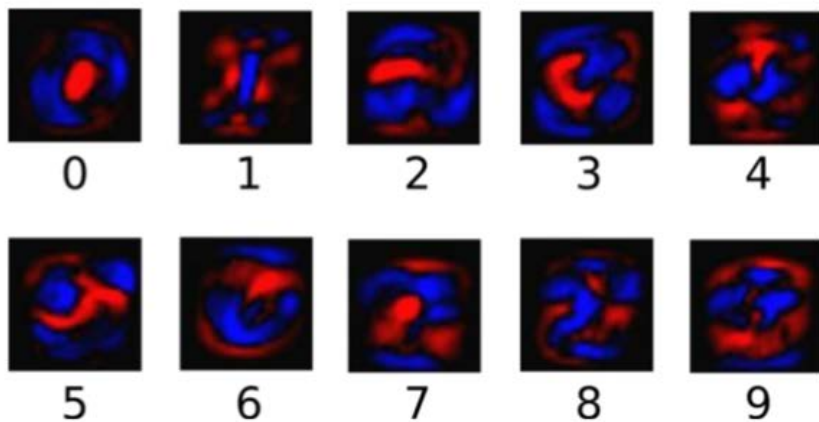


Diagram shows the weights one model learned for each of these classes.

- Red negative weights
- Blue positive weights

Previous work

SE aspect

Computer vision

“In machine learning, having more data is almost always as important as having better algorithms and the amount of effort that goes into data engineering is more than the effort that goes on algorithm picking.”

References

- [Machine Learning is Fun! Deep Learning and Convolutional Neural Networks : Part 1,2,3,4](#)
- [Distributed Machine Learning Pipelines in Apache Spark](#)
- [Train, Test, Evaluate for Multiclass Classification: Letter Recognition Dataset](#)
- [TFLearn: Deep learning library featuring a higher-level API for TensorFlow](#)
- [The MNIST Data](#)

Thank you !
