

Identification and Recognition of Handwritten Text from an Image

Aryan Trivedi^a, Rishikesh Pandey^b, Utkarsh Kushwaha^c

^{a,b,c}Dept. of CSE, Galgotias university, G. Noida, U.P.

^arvtrivedi21@gmail.com, ^brishikeshpandey005@gmail.com, ^cut.kush958@gmail.com

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract: The industrial project is about extraction of content from pictures, which is acknowledgment of characters in electronic or mechanical transformation of pictures of composed, manually written or printed text into machine encoded text, regardless of whether from a checked archive, a photograph of a record, a scene- photograph (for instance the content on signs and announcements in a scene photograph) or from caption text superimposed on picture.

Broadly utilized as type of information section from printed paper information – regardless of whether identification reports, solicitations, bank explanations, automated receipts, N business cards, mail, printouts of static-information, or an appropriate documentation m– it is a typical technique for digitizing printed messages with the goal that they can be e electronically altered, put away more minimalistic ally, showed online, and utilized in machine cycles, machine interpretation, (removed) text to- discourse, key information and text mining. OCR is a field of exploration in example acknowledgment, computerized reasoning and PC vision.

The primary goal of this task is to create AI calculation to empower element and information extraction from records with manually written explanations, with an expect to distinguish transcribed words on a picture.

Keywords: Open CV, Pytesseract, OCR, Python, Num Py

Acknowledgement

This project was completed by Aryan Trivedi, Rishikesh Pandey, Utkarsh Kushwaha.

Thanks to all team members who provided insight and expertise that greatly assisted the project.

We thank Mrs Sofia Singh, Assistant Professor, Galgotias University for assistance, and Mr. A. Suresh Kumar, Assistant Professor, Galgotias University for comments that greatly

1. Introduction

The industrial project is about extraction of content from a pictures, which is optical acknowledgment of text in the electronic or mechanical transformation of pictures of composed, manually written or printed text into machine encoded text, regardless of whether from a checked archive, a photograph of a record, a scene- photograph or from caption text superimposed on a picture.

Broadly utilized as a type of information section from printed paper information records – regardless of whether identification reports, solicitations, bank explanations, automated receipts, business cards, mail, printouts of static-information, or any appropriate documentation – it is a typical technique for digitizing printed messages with the goal that they can be electronically altered, looked, put away more minimalistic ally, showed on-line, and utilized in machine cycles, for example, intellectual processing, machine interpretation, (removed) text to- discourse, key information and text mining. OCR is a field

Of exploration in example acknowledgment, computerized reasoning and PC vision.

The primary goal of this task is to create AI calculation to empower element and information extraction from records with

2. Solution Approach

We're using open-source software called Pytesseract using machine learning approach although there are various approaches to implement this project, but at last main aim of this internship is to fulfill project aim and objectives.

To meet venture target we have learned about convolution neural organizations. The plan for picking this field since it empowers the machines to see the world as people do, see it likewise and even utilize the information for a large number of undertakings, for example, Image and Video acknowledgment, Image Analysis and Classification, Media Recreation, Recommendation Systems, Natural Language Processing, and so on The headways in Computer Vision with Deep Learning has been built and culminated with time, basically more than one specific calculation is Convolutional Neural Network. To implement my project I have used tesseract python as suggested in the Webinars as well as Self Learning module.

Used python tools such as OpenCv, Numpy, Python imaging library and Pytesseract

The undertaking configuration is partitioned into parts.

The initial segment intended to take the contribution of the picture whose text is to be separated.

The second piece of the undertaking is the primary piece of the venture which intended to diminish the clamor and actualizes the Pytesseract with convolutional neural organization calculation bit by bit to distinguish the content present in the picture.

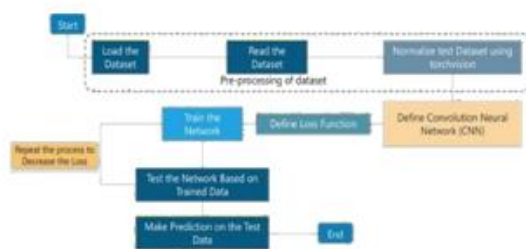
3. Assumptions

The assumptions considered are as follows:

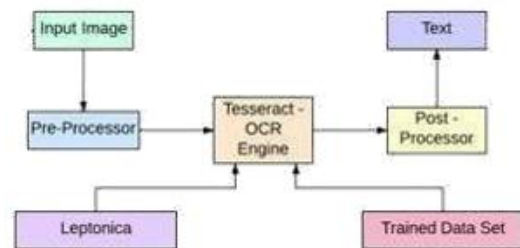
1. The handwritten text across must be in English.
2. The text across the input image must be clearly handwritten in order to achieve good results.
3. All machine dependencies must be installed properly.

4. Project Diagrams

Logic Flow of Convolution Neural Networks



Logic Flow of Pytesseract OCR Engine

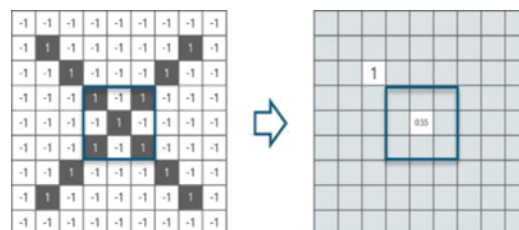


5. Algorithms

The algorithm used to implement the project is the Pytesseract ocr engine which uses convolutional neural network algorithm which is used by the tesseract optical character recognition engine in python. There are four layered concepts we should understand in convolutional neural networks:

1. Convolution
2. Rectified Linear Unit
3. Pooling Layers
4. Full Connectedness (Fully Connected Layer)

Convolution of an Image



Convolution has the pleasant property of being translational invariant. Naturally, this implies that every convolution channel speaks to a component of premium (e.g pixels in letters) and the Convolutional Neural Network calculation realizes which highlights contain the subsequent reference (for example letter set).

We have 4 stages for convolution:

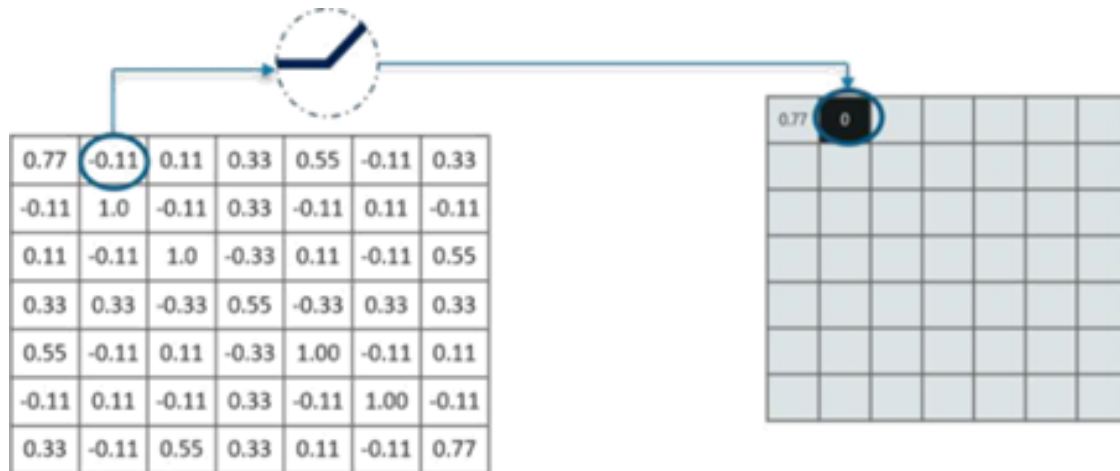
- Line up the component and the picture
- Multiply each picture pixel by comparing highlight pixel
- Add the qualities and discover the total
- Divide the total by the absolute number of pixels in the element

The yield signal strength isn't reliant on where the highlights are found, yet essentially whether the highlights are available. Thus, a letters in order could be sitting in various positions and the convolutional neural organization calculation would at present have the option to remember it.

Rectified Linear Unit

Change work possibly initiates a hub if the info is over a specific amount, while the information is under zero, the yield is zero, however when the info transcends a specific limit, it has a straight relationship with the needy variable.

The principle point is to eliminate all the negative qualities from the convolution. All the positive qualities continue as before however all the negative qualities get changed to zero as demonstrated as follows:



Contributions from the convolution layer can be smoothed to decrease the affectability of the channels to commotion and varieties. This smoothing cycle is called sub testing and can be accomplished by taking midpoints or taking the most extreme over an example of the sign.

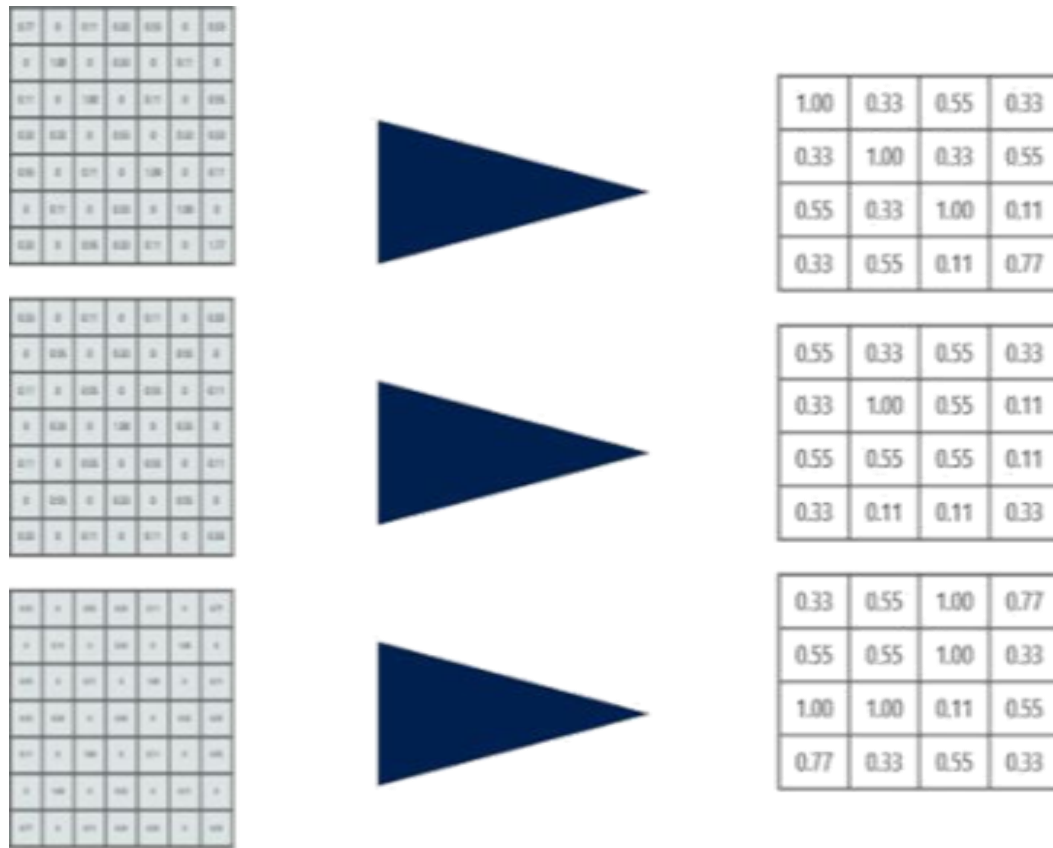
Pooling Layer

In this layer the therapist the picture stack into a more modest size. Pooling is done in the wake of going through the enactment layer. We do this by executing the accompanying 4 stages:

- Pick a window size (generally 2 or 3)
- Pick a step (generally 2)
- Walk your window across your sifted pictures
- From every window, take the greatest worth

We took window size to be 2 and we got 4 qualities to browse. From those 4 qualities, the greatest incentive there is 1 so we pick 1. Likewise, note that we began with a 7x7 lattice yet now a similar framework in the wake of pooling boiled down to 4x4.

Yet, we need to get the window across the whole picture. The strategy is actually as same as above and we need to rehash that for the whole picture. Do take note of that this is for one channel. We need to do it for 2 different channels too. This is done and we show up at the accompanying outcome:



Well the easy part of this process is over. Next up, we need to stack up all these layers!

Stacking Up the Layers

So to get the time period in one picture we're here with a 4x4 framework from

a 7x7 framework subsequent to going the contribution through 3 layers – Convolution, Rectified Linear Unit and Pooling.

We further diminish the picture from 4x4 to 2x2 to accomplish this we need to play out the 3 tasks in emphasis after the main pass. So after the second pass we show up at a 2x2 framework as demonstrated as follows:



The last layers in the organization are completely associated, implying that neurons of going before layers are associated with each neuron in resulting layers.

This impersonates significant level thinking where all potential pathways from the contribution to yield are thought of.

Likewise, completely associated layer is the last layer where the order really occurs. Here we take our separated and yelled pictures and put them into one single rundown as demonstrated as follows:

1	0.55
0.55	1.00

1	0.55
0.55	0.55

0.55	1.00
1.00	0.55

So next, when we feed in, 'X' and 'O' there will be some component in the vector that will be high.

Consider the picture beneath, as should be obvious for 'X' there are various components that are high and correspondingly, for 'O' we have various components that are high:

All things considered, what did we comprehend from the above picture is the point at which the first, fourth, fifth, tenth and eleventh qualities are high; we can arrange the picture as 'x'. The idea is comparative for different letter sets too – when certain qualities are orchestrated the manner in which they will be, they can be planned to a genuine letter or a number which we require.

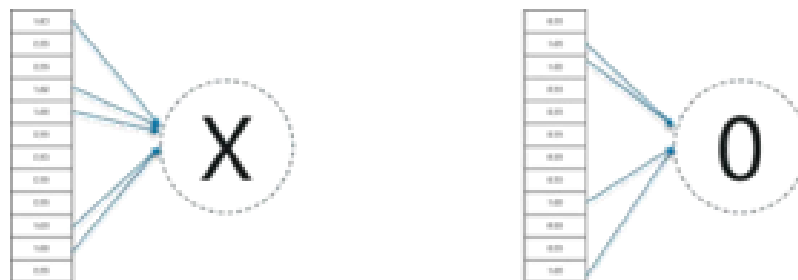
Prediction of Image Using Convolutional Neural Networks – Fully Connected Layer

Now, we're finished preparing the organization and we can start to foresee and check the working of the classifier. We should look at a basic model:

0.9
0.65
0.45
0.87
0.96
0.73
0.23
0.63
0.44
0.89
0.94
0.53

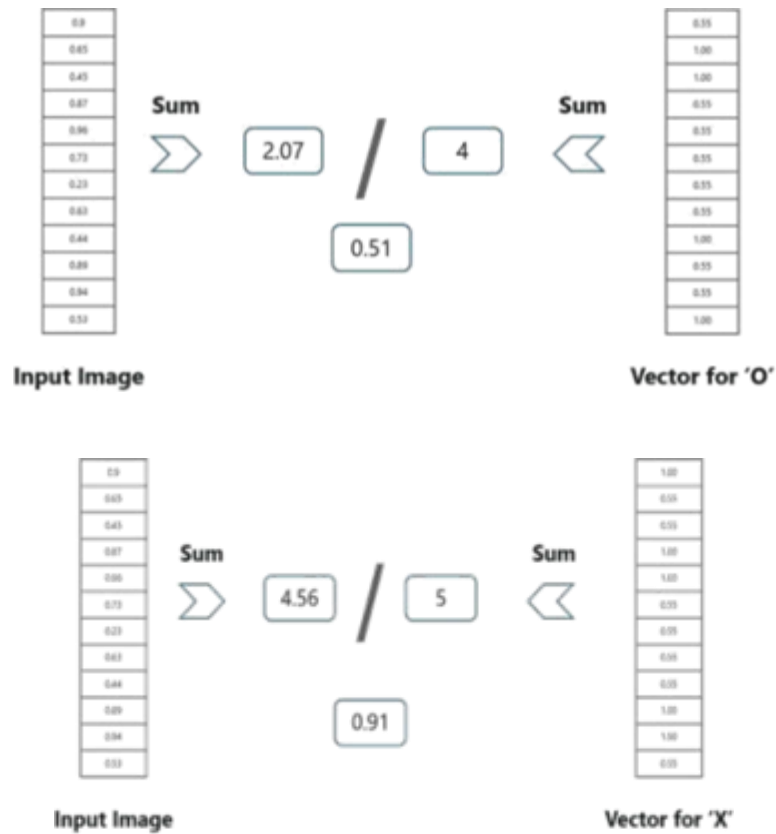
We have a 12 component vector got subsequent to passing the contribution of an arbitrary letter through all the layers of our organization.

We make forecasts dependent on the yield information by contrasting they got qualities and rundown of 'x' and 'o'.



We just added the qualities we which discovered as high (first, fourth, fifth, tenth and eleventh) from the vector table of X and we persuaded the aggregate to be 5. We did precisely the same thing with the info picture and gotan estimation of 4.56.

At the point when we partition the worth we have a likelihood match to be 0.91! How about we do likewise with the vector table of 'o' presently:



Being 0.51 is less than 0.91, isn't it?

So we can conclude that the resulting input image is an 'x'. And this is how prediction work is done.

6. Outcome

The calculation can identify and section manually written content from a picture. The model effectively ready to identify greatest words in a given line of sentence or words, which makes it about 90% precise while execution and testing.

For instance the info picture having the manually

Written content is given as following: We have the output as 0.51 with this table.

The model processes the image removes the noise from the image and the Pytesseract performs the convolution neural networks and predicts the text to be extracted.

Extracted Text: Jim Morrison

Jim Morrison

```

File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
# Recognize text with tesseract for python
result = pytesseract.image_to_string(Image.open(img_path))

# Remove template file

return result

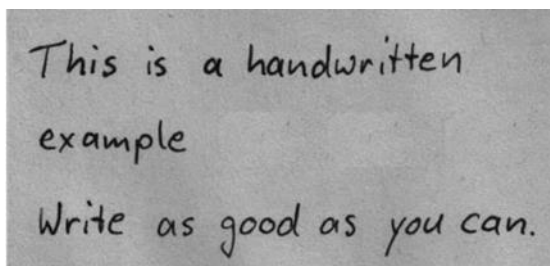
print ("____The Recognition Process Of Text From The Input Image Started Successfully____")
print ("The Recognized Text From The Input Image IS:")
print (get_string(filename))

print ("____ Process Finished ____")
____The Recognition Process Of Text From The Input Image Started Successfully____
The Recognized Text From The Input Image IS:
Jim Morrison
____ Process Finished ____
    
```

As we can see the model is quite accurate and successfully able to extract the handwritten text.

The model predicts and extracts the text from the image as follows:

Whereas another image having the handwritten text is given as following:



The model processes the image removes the noise from the image and the Pytesseract performs the convolution neural networks and predicts the text to be extracted.

Extracted Text : This is the handwritten example

Write as good as you can

As we can see the model is quite accurate and successfully able to extract the handwritten text.

The model predicts and extracts the text from the image as follows:

```

File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
cv2.imwrite(img_path, img)

# Recognize text with tesseract for python
result = pytesseract.image_to_string(Image.open(img_path))

# Remove template file

return result

print ("____The Recognition Process Of Text From The Input Image Started Successfully____")
print ("The Recognized Text From The Input Image IS:")
print (get_string(filename))

print ("____ Process Finished ____")
____The Recognition Process Of Text From The Input Image Started Successfully____
The Recognized Text From The Input Image IS:
This is a handwritten
example
Write as good as you can.
____ Process Finished ____
    
```

7. Exceptions Considered

The exceptions considered in this are as follows:

1. The text across the input image must be of the same color not multicolor handwritten text.
2. The image doesn't have too aggressive multicolor backgrounds across the text of the image.
3. The image doesn't have any kind's objects in the background across the text of the image.

8. Enhancement Scope

The enhancement scope of this project are follows:

1. The accuracy of the model can increased with predefined models and powerful machine learning GPU processors can be used to attain a good percentage of accuracy.
2. In future we can use this algorithm with more than one particular language.

References

1. <https://software.intel.com/content/www/us/en/develop/training/course-artificial-intelligence.html>
<https://software.intel.com/content/www/us/en/develop/training/course-machine-learning.html>
https://www.python-course.eu/machine_learning.php
<https://numpy.org/doc/>
<https://pypi.org/project/pytesseract/>
<https://www.geeksforgeeks.org/python-using-pil-imagegraband-pytesseract/>
2. <https://medium.com/analytics-vidhya/convolutional-neuralnetworks-cnn-explained-step-by-step-69137a54e5e7>
3. 69137a54e5e7
4. <https://drive.google.com/drive/folders/1OfofbBPm5XDAZLVaAfFm0ASW937P03Hs?usp=sharing>