

OCR

July 16, 2024

```
[6]: import easyocr
import joblib
import pytesseract
from PIL import Image
import os
import matplotlib.pyplot as plt
```

0.1 Recognizing the text from images - using easyOCR

```
[7]: def recognise_text_easyocr(img_path,min_confidance=0.5):

    reader = easyocr.Reader(['en','hi'])
    text = ''

    for (bbx,txt,prob) in reader.readtext(img_path):
        if prob > min_confidance:
            text += ' '+txt

    text = text.strip() #Strip the leading/trailing spaces

    file_name = os.path.splitext(os.path.basename(img_path))[0] #Extract the
    ↪file name from image path
    output_path = '/kaggle/working/'+ file_name + '_easyocr_+'.txt'
    with open(output_path,'w') as file:
        file.write(text)

    return print(f'Easyocr_output: {text}')
```

0.2 Recognizing the text from images - pytesseract

```
[8]: def recognise_text_pytesseract(img_path):

    image = Image.open(img_path)

    text = pytesseract.image_to_string(image,lang='eng+hin')

    text = text.strip()
```

```

    file_name = os.path.splitext(os.path.basename(img_path))[0] #Extract the
↪file name from image path
    output_path = '/kaggle/working/' + file_name + '_pytesseract_'.txt'
    with open(output_path,'w') as file:
        file.write(text)

    return print(f'pytesseract_output: {text}')

```

[]:

0.3 Comparing the read texts from images of Easyocr and pytesseract

```

[9]: for file in os.listdir('/kaggle/input/images'):

    path = '/kaggle/input/images' + '/' + file

    print(file)
    recognise_text_easyocr(path)
    recognise_text_pytestesseract(path)
    print('\n')

```

text6.png
Easyocr_output:
pytesseract_output: wa at aA

text0.jpg
Easyocr_output: MAKE TEXT STAND OUT FROM
pytesseract_output: "" MAKE TEXT
STAND OUT FROM
_. BACKGROUNDS 7? "Se

text4.jpg
Easyocr_output: ISAYIT TO REMIND YOU THAT
pytesseract_output: Mere ee A eld
TN A Re
De UR ELL
SELES
Dae

all

text8.jpg
Easyocr_output: ()
MatraShabd com

pytesseract_output: Ha Haz Aer Weg

faa (tated)

ala oer aret a16e

MatraShabd.com

text2.jpg

Easyocr_output: NEW MTNDSET NEW RESULTS

pytesseract_output:

text3.jpg

Easyocr_output: roads in your neighborhood. Work will commence on June 3rd, as

shown in the project schedule This work will require road closures at certain times throughout the project. We will be starting work work zone limits and park your car in your driveway at night, please this may be a short-term inconvenience and we hope to work together with you in a professional manner for the long-term improvement of your neighborhood.

pytesseract_output: SHADOWING PRACTICE 08

The city has contracted with Acme Construction to improve certain roads in your neighborhood. Work will commence on June 3rd, as shown in the project schedule. This work will require road closures at certain times throughout the project. We will be starting work

at 7:00 AM and will finish around 5:00 PM. If you are within the active work zone limits and park your car in your driveway at night, please be sure to move it prior to the 7:00 AM start time. We understand that this may be a short-term inconvenience and we hope to work together with you in a professional manner for the long-term improvement of your neighborhood.

text9.png

Easyocr_output:

pytesseract_output: ai Fae Re ak afteri & Aad A AST
CAAA SAL SAAT HTT 1 TS SEK GAR SATE

012 FI @ sik RER Se USAR & wa a sala

Hal afew!

text7.png

Easyocr_output:

pytesseract_output:

text5.jpg

Easyocr_output: Allusion Alliteration Allegory Analogy EVERYTHING you need to know about LANGUAGE FEATURES English Language FEATURES Anaphora Antithesis Emotive language Assonance MY ASSIGNMENT

pytesseract_output: Allusion

Alliteration

"_ "= EVERYTHING

you need to know about

FEATURES

Anaphora Emotive language

Antithesis

Assonance

Q® MY Assicnmenrrr HELP

text1.jpg

Easyocr_output: scared? good. we dont grow when we stay inside our comfort zone.

pytesseract_output: scared? good.

we don't grow

when we stay

inside our

comfort zone.