

ct](<https://pypi.org/project/pytesseract/>)

Character Recognition (OCR) library that will extract text from images. This library requires the [Python Image Library](<https://pypi.org/project/Pillow/>) to be installed on the system as well as [Google Tesseract OCR](<https://github.com/tesseract-ocr/tesseract>), which powers the character recognition.pip. [pdf2image](<https://github.com/Belval/pdf2image>) converts PDFs to images that OCR can work on.

Linux (Linux)

```
sudo apt-get install python3-pil`
```

```
sudo apt-get install pillow`
```

Linux (Linux)

```
sudo apt-get install tesseract-ocr`
```

```
sudo apt-get install tesseract`
```