

# Image Text Extractor Documentation

## Overview

The Image Text Extractor is a web application built using Streamlit, designed to extract text from uploaded images using Tesseract OCR.

## Features

- **Upload Image:** Users can upload JPEG or PNG images via the web interface.
- **Text Extraction:** Once an image is uploaded, users can click "Extract Text" to initiate OCR processing.
- **Display Text:** Extracted text is displayed below the uploaded image.

## Installation and Setup

### Prerequisites

- Python (version 3.6 or higher)
- pip package manager

### Installation Steps

#### 1. Clone the repository:

```
git clone <repository_url>
cd image-text-extractor
```

#### 2. Install dependencies:

```
pip install -r requirements.txt
```

#### 3. Specify Tesseract OCR executable path:

- Ensure Tesseract OCR is installed. If not, download and install from [Tesseract-OCR](#).
- Set the Tesseract path in `app.py`:

```
pytesseract.pytesseract.tesseract_cmd =
r'<path_to_tesseract_executable>'
```

#### 4. Run the Streamlit application:

```
streamlit run app.py
```

5. **Access the application:** Open your web browser and go to `http://localhost:8501` to use the Image Text Extractor.

## Usage

1. **Upload an image:**
  - Use the sidebar to upload a JPEG or PNG image file.
2. **Extract Text:**
  - Click the "Extract Text" button to process the uploaded image with OCR.
3. **View Extracted Text:**
  - The extracted text will be displayed below the uploaded image.

## Notes

- Ensure the uploaded images are clear and well-lit for better OCR accuracy.
- Supported image formats: JPEG (.jpg, .jpeg) and PNG (.png).
- Adjust OCR settings or preprocess images for specific requirements using `app.py`.

## Application User Interface:

