

Documentation for Fake Invoice Detection

Using Tesseract OCR and Computer Vision

Maha Qaiser

22i-2348

Description

A simple web app to detect fake invoices using Tesseract OCR, SHA256 hashing, and layout template matching with OpenCV.



Upload Original Invoice

Choose file No file chosen

Add Original Invoice

Fake Invoice Detector

Choose file No file chosen

Analyze Invoice

How to Run

#0. Make sure Tesseract OCR is installed and set the correct path in main.py:

```
pytesseract.pytesseract.tesseract_cmd =  
r"C:\Users\mahaq\AppData\Local\Programs\Tesseract-OCR\tesseract.exe"
```

Then on the terminal:

1. Install dependencies

```
pip install flask pillow pytesseract opencv-python
```

2. Make sure Tesseract-OCR is installed and added to PATH

3. Run the Flask app

```
python app.py
```

Flow of Web App's Fake Invoice Detection

Step 1: Upload Image

The user clicks the 'Choose File' button and uploads an image

Step 2: Click to Analyze

The user clicks the 'Analyze Invoice' button and the results are displayed

How it Works

1. Analyze Invoice

- Extracts text from uploaded invoice using Tesseract OCR
- Computes SHA256 hash of the text
- Checks if this hash matches any known original invoice's hash and further uses template matching and field presence to classify the invoice

2. Add Original Invoice

- User uploads an original invoice image
- The SHA256 hash of its OCR text is stored in a file 'original_hashes.txt'

Example Results

Test case 1

Original Invoice Input:

Invoice no: 40394912

Date of issue: 03/23/2020

Seller:

Mccullough-Bradley
457 Vargas Brook Suite 967
East James, MA 63095

Tax Id: 905-94-2945
IBAN: GB76GUQ007697012433171

Client:

Cross-Estrada
7297 Smith Roads
Lake Sherri, TN 88469

Tax Id: 915-81-8637

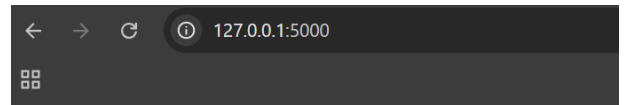
ITEMS

No.	Description	Qty	UM	Net price	Net worth	VAT [%]	Gross worth
1.	Microsoft Xbox One X 1TB 4k Ultra HD Console black controller works great 1787	5,00	each	249,00	1 245,00	10%	1 369,50

SUMMARY

	VAT [%]	Net worth	VAT	Gross worth
	10%	1 245,00	124,50	1 369,50
Total		\$ 1 245,00	\$ 124,50	\$ 1 369,50

Output:



Upload Original Invoice

Choose file No file chosen

Add Original Invoice

Fake Invoice Detector

Choose file No file chosen

Analyze Invoice

- Invoice No: 51109338
- Date of Issue: 04/13/2013
- Seller Tax ID: 945-82-2137
- Client Tax ID: 942-80-0517
- IBAN: GB75MCRL06841367619257

Original Invoice

Test case 2

Fake Invoice Input:

INVOICE

Date Issued:

01 January 2030

Invoice No:

01234

Issued to:

Sacha Dubois

123 Anywhere St., Any City,

ST 12345

NO	DESCRIPTION	QTY	PRICE	SUBTOTAL
1	Item 1	1	\$ 100	\$ 200
2	Item 2	2	\$ 50	\$ 100
GRAND TOTAL				\$ 300

Note:

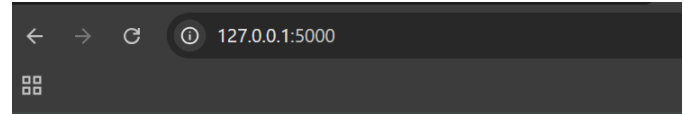
Bank Name: Rimberio

Account No: 0123 4567 8901

Claudia

Finance Manager

Output:



Upload Original Invoice

Choose file

No file chosen

Add Original Invoice

Fake Invoice Detector

Choose file

No file chosen

Analyze Invoice

Fake Invoice