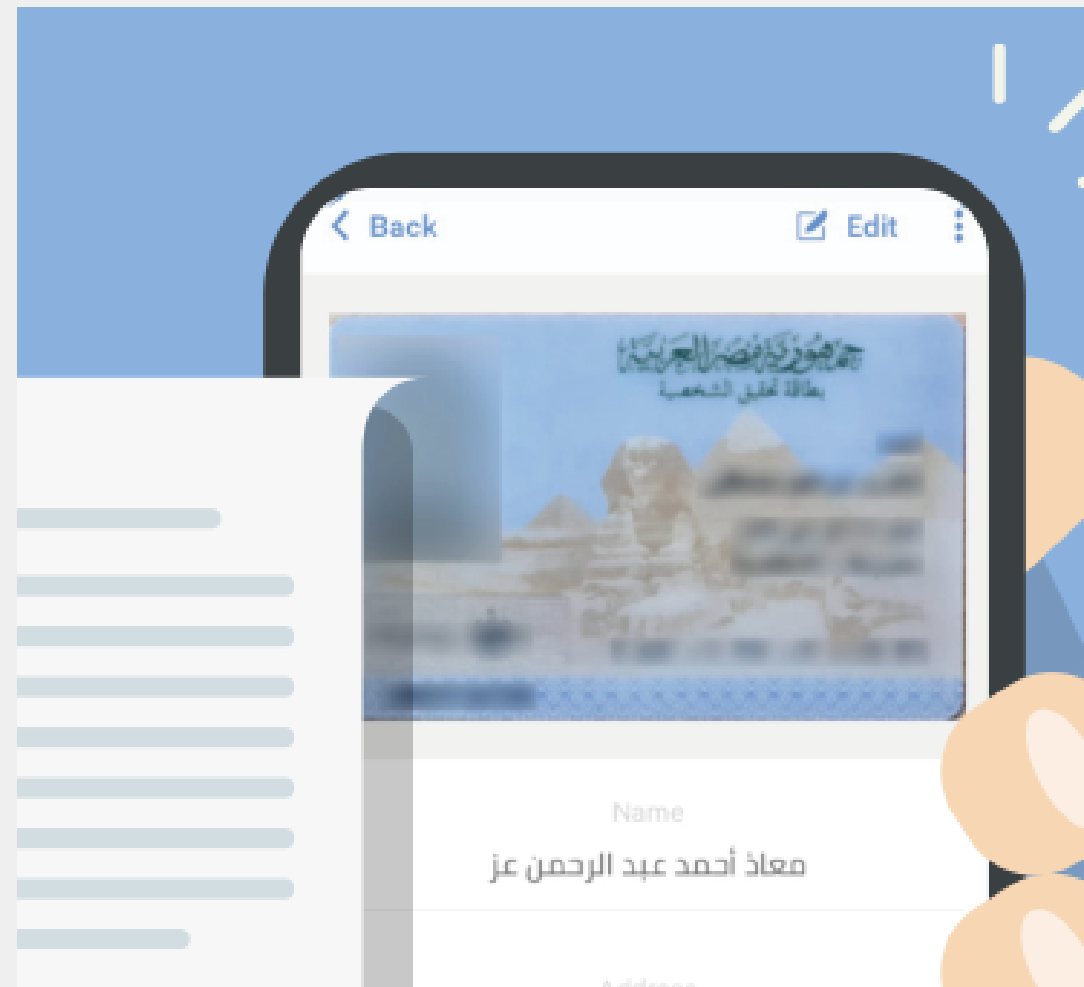


Arabic ID

Information Extraction

Using optical character recognition to scan
Arabic text could not be easier.



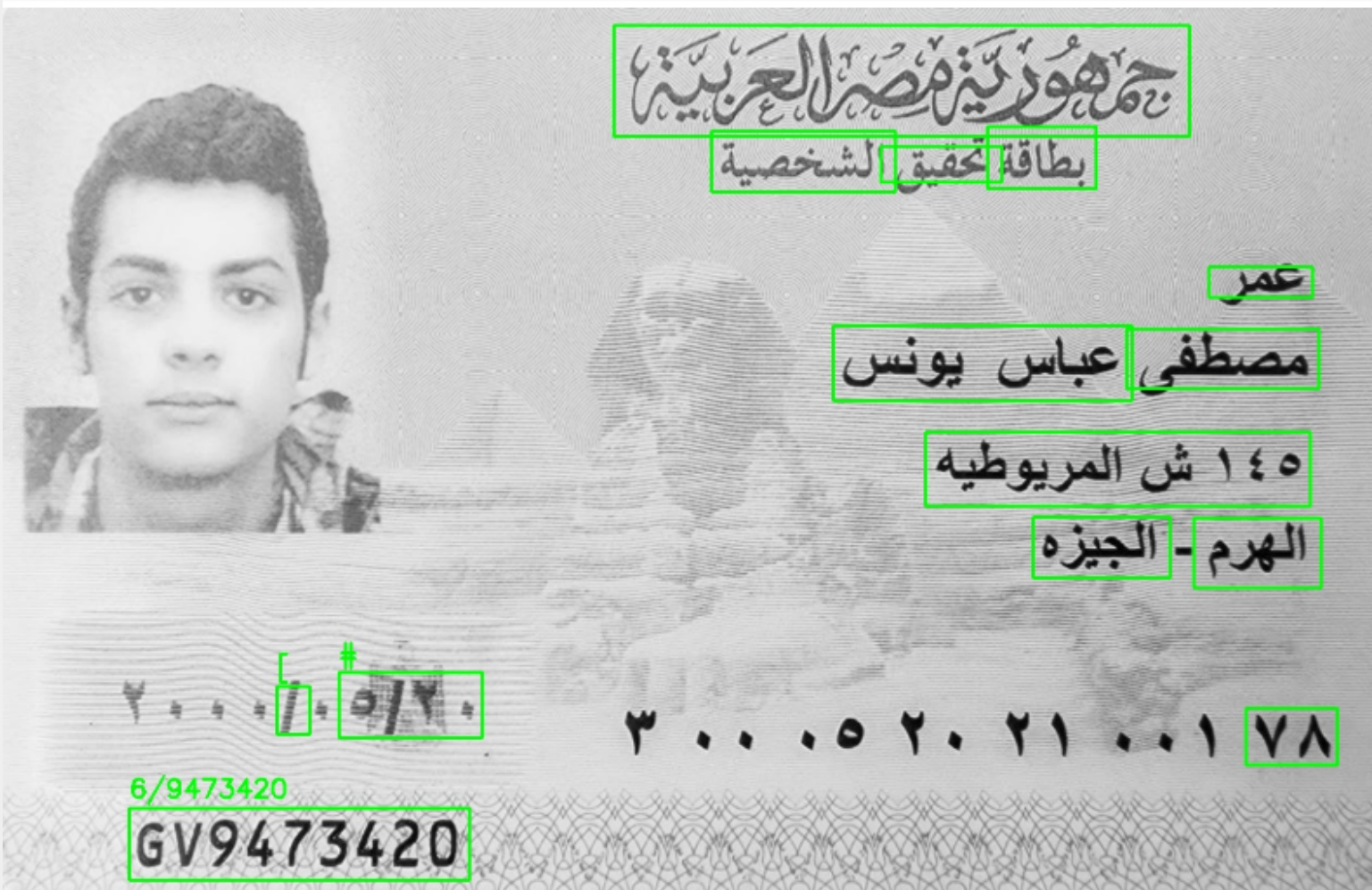
Arwa
Abdelrahman
Nour
Salama
Zidane

NTI

1. Extract id numbers



we used Arabic ocr and its result were bad



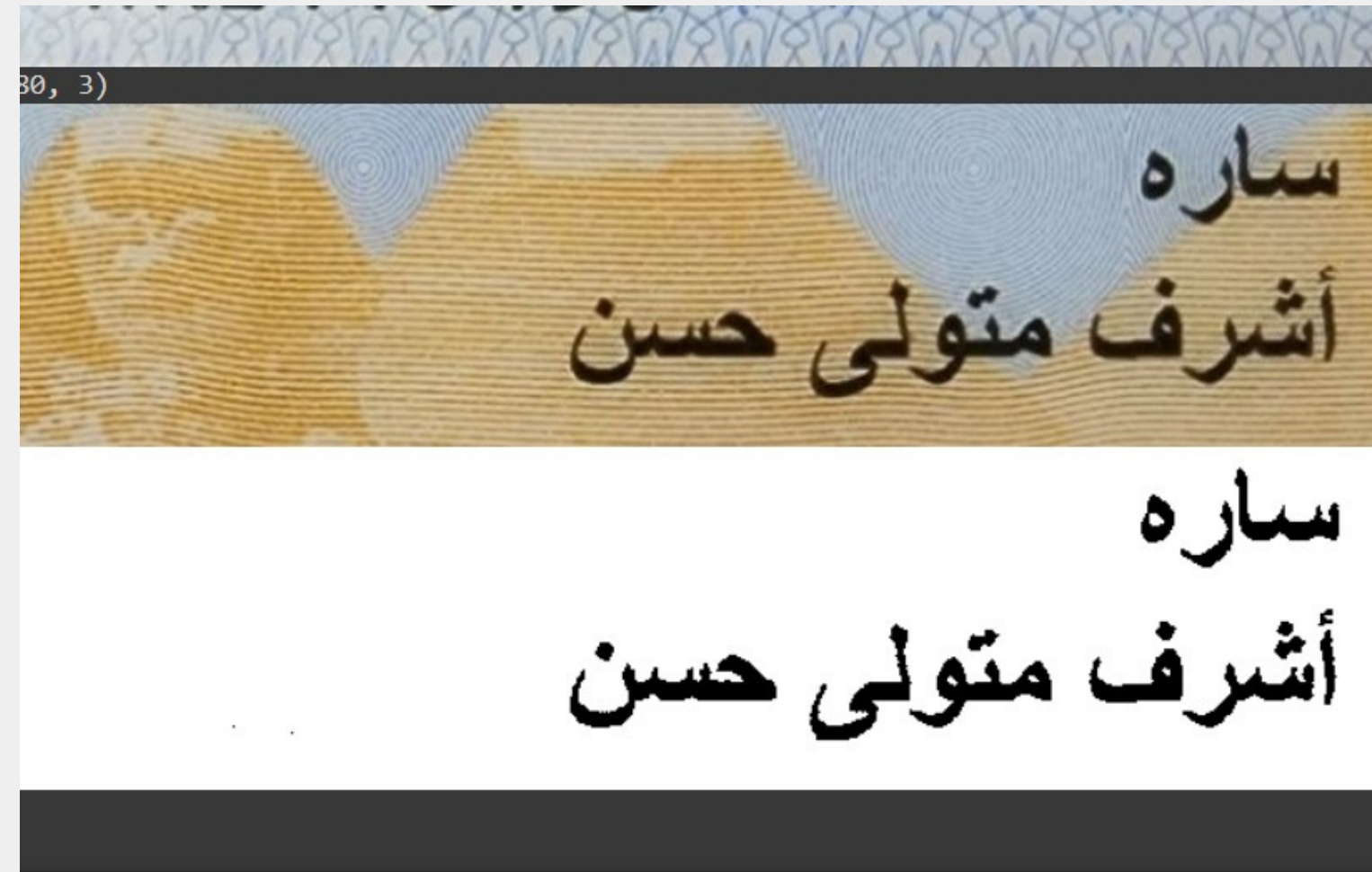
| | | |
|--------|---------|-----------------|
| [INFO] | 0.6050: | الشخصية |
| [INFO] | 0.7422: | بطاقة |
| [INFO] | 0.7002: | عباس يونس |
| [INFO] | 0.6749: | ١٤٥ ش المريوطيه |
| [INFO] | 0.8461: | الجيزه |
| [INFO] | 0.6549: | الهرم |
| [INFO] | 0.3469: | [|
| [INFO] | 0.2798: | ٢٠٠٩/٥/٢٠ |
| [INFO] | 0.9999: | ٧٨ |
| [INFO] | 0.7130: | 6/9473420 |
| [INFO] | 0.6050: | تحقيق |
| [INFO] | 0.8414: | عمل |
| [INFO] | 0.5874: | مصطفى |

—

2. Extract names

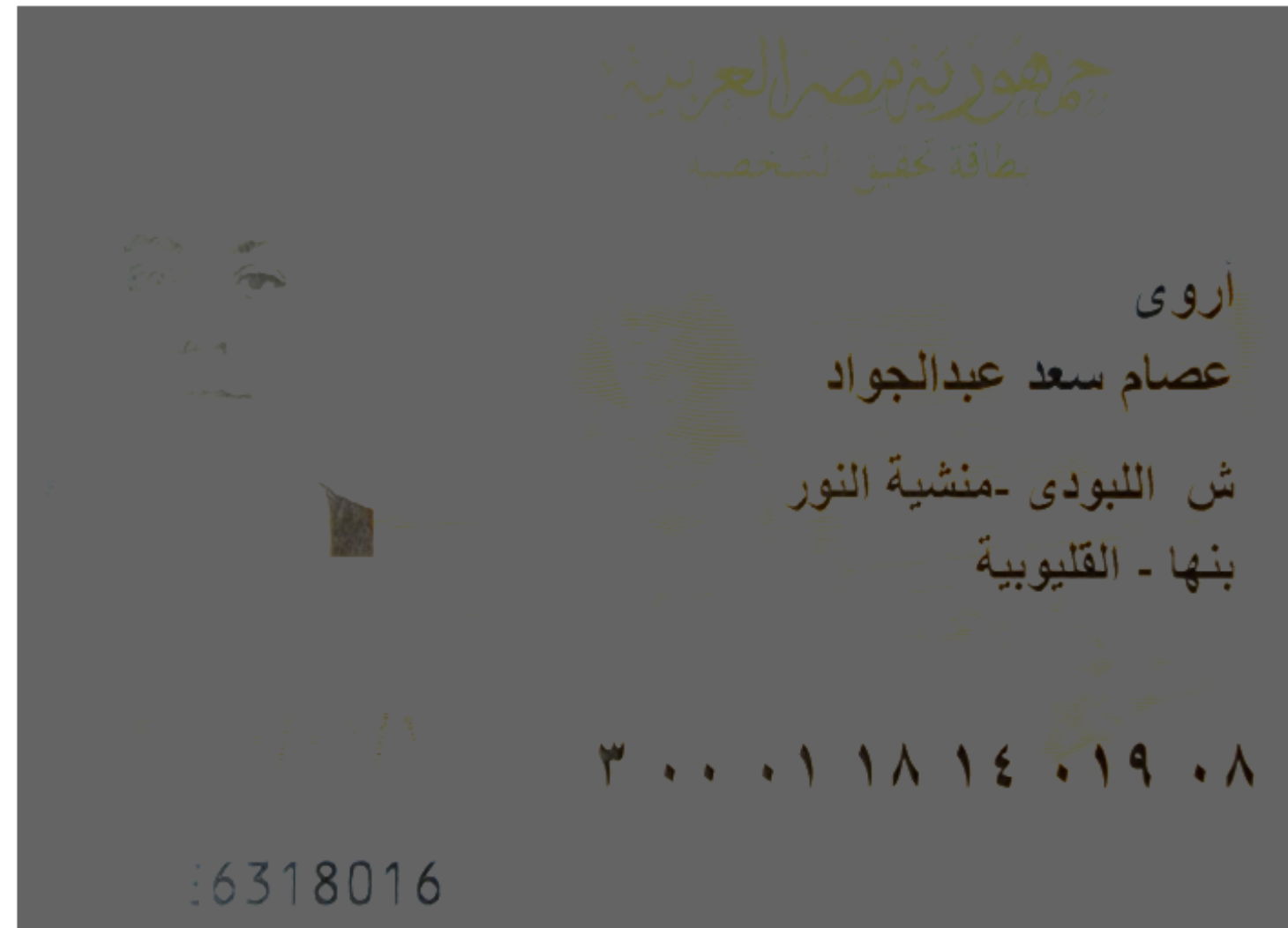


We made some preprocessing to the images to improve the results

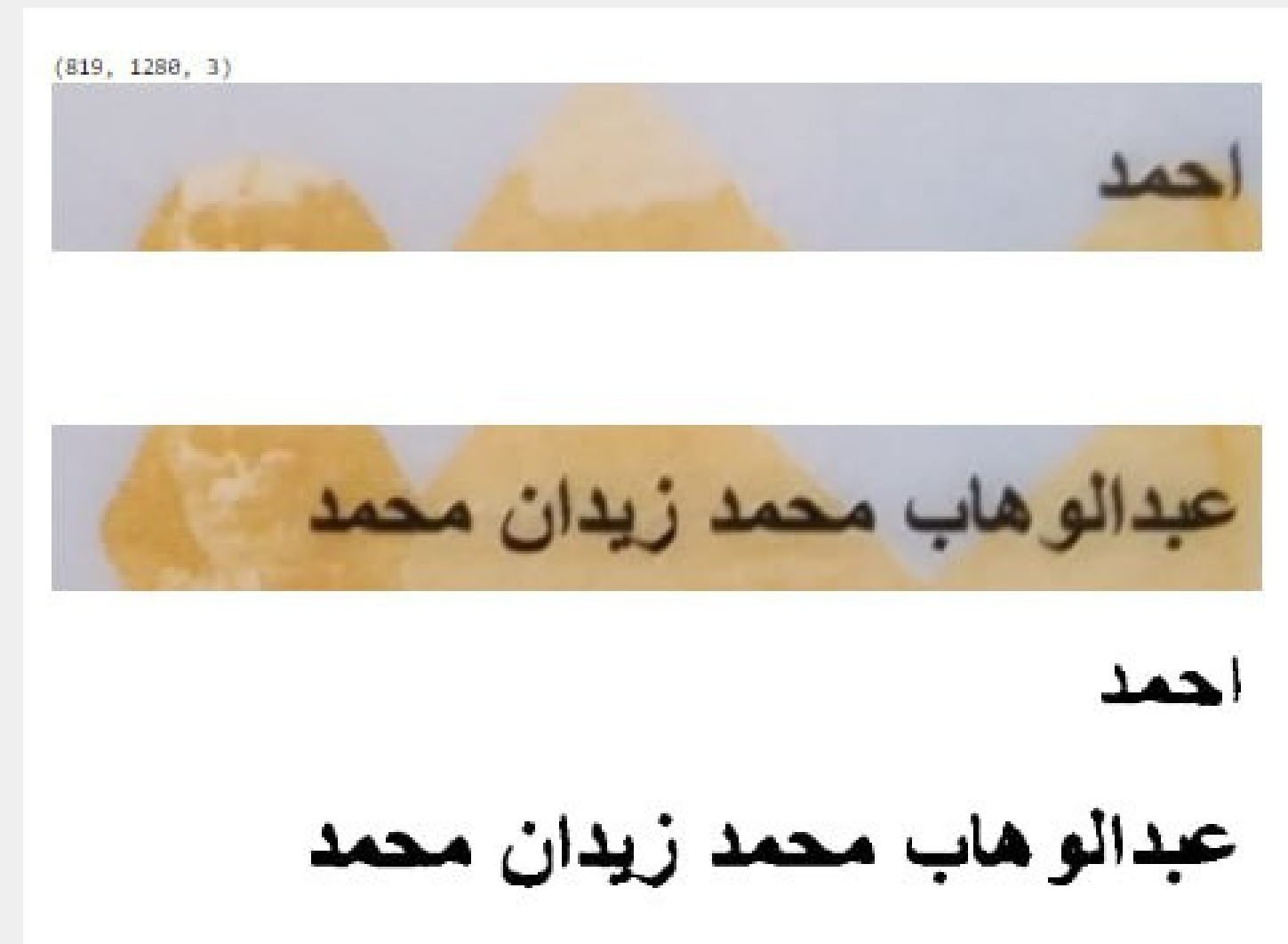


We made some preprocessing to the images to improve the results using BGR2GRAY

```
[ ] img3 = threshold_eng_num(imgg)  
cv2_imshow(img3)
```

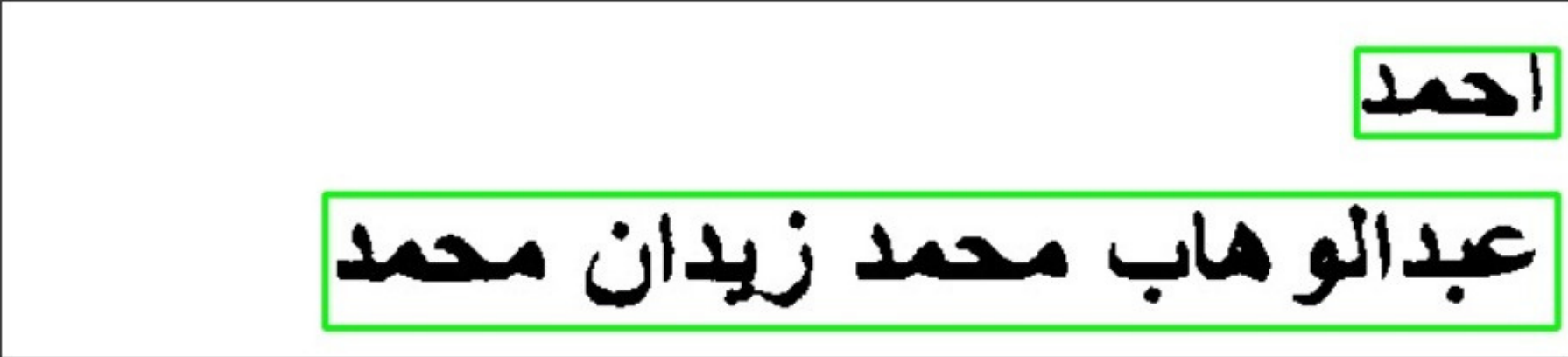


We tried to improve the results by dividing the name into two parts



We used Arabic ocr

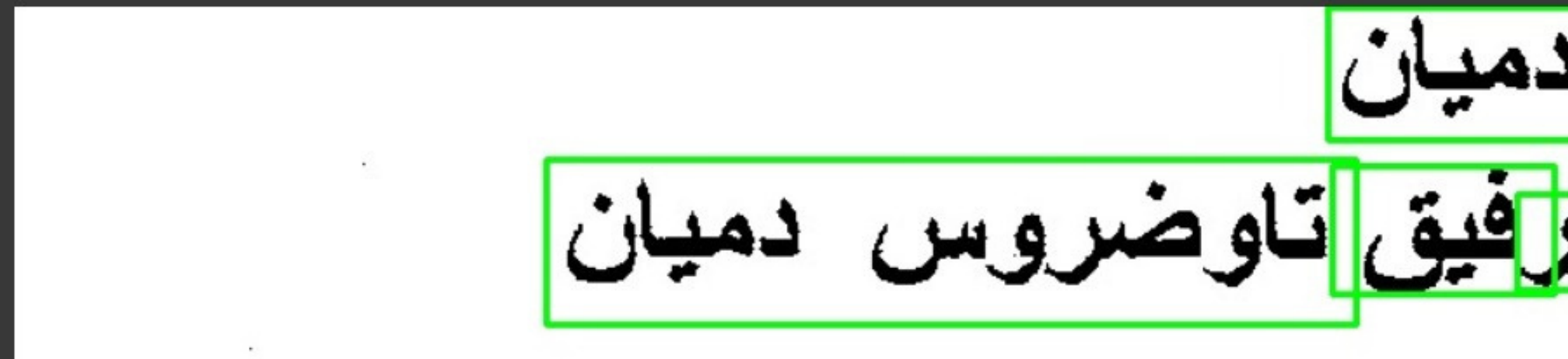
```
[305] results  
[[[[[626, 22], [719, 22], [719, 62], [626, 62]], 'احمد',  
[[[149, 89], [719, 89], [719, 151], [149, 151]],  
'عبدالوهاب محمد زيدان محمد',  
0.7947361137563755]]  
  
img = cv2.imread(out_image)  
cv2_imshow(img)
```



Results were bad

```
[309] results
```

```
[310] img = cv2.imread(out_image)
      cv2_imshow(img)
```

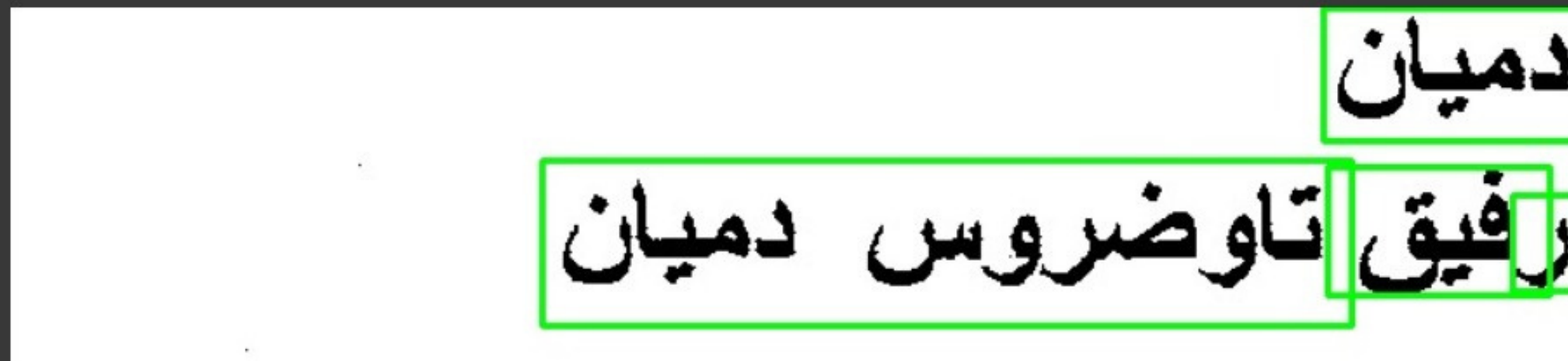


```
[INFO] 0.8498: دقیق تاوضروس دمیان
```

```
[INFO] 0.7528: دمیان
```

```
[309] results
```

```
[310] img = cv2.imread(out_image)  
      cv2_imshow(img)
```



```
[INFO] OCR'ing input image...
```

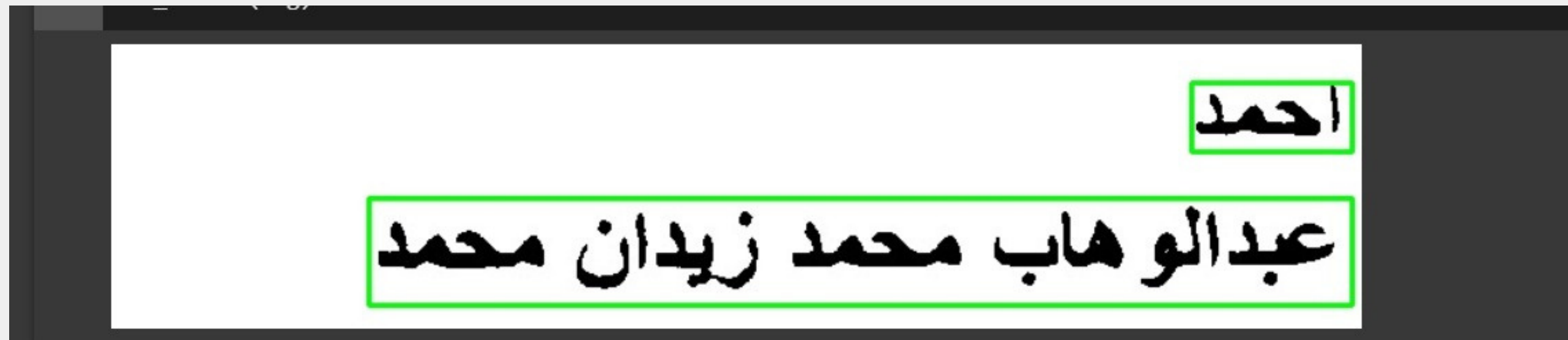
```
[INFO] 0.4785: دَمیان
```

```
[INFO] 0.3683: تاوَضروس
```

```
[INFO] 0.8257: دقیق
```



tesseract ocr



```
[312] print(names)
```

```
['احمد عبدالوهاب محمد زيدان محمد', 'دميان رفيق تاوضروس دميان']
```

3. Face detection



4. Dataset Annotation



Data annotation using make sense ai



Detection by using YOLO5



جمهورية مصر العربية
بطاقة تحقيق الشخصية



ميّار
حنفي محمد حامد

١ ش محمد دموع - مشروع عثمان - بهتيم
شبرا الخيمة ثان - القليوبية

١٩٩٩/٠٩/٢٨

٢٩٩٠٩٢٨١٤٠٣٠٨٤

GZ4025798

5. Id card detection and localization



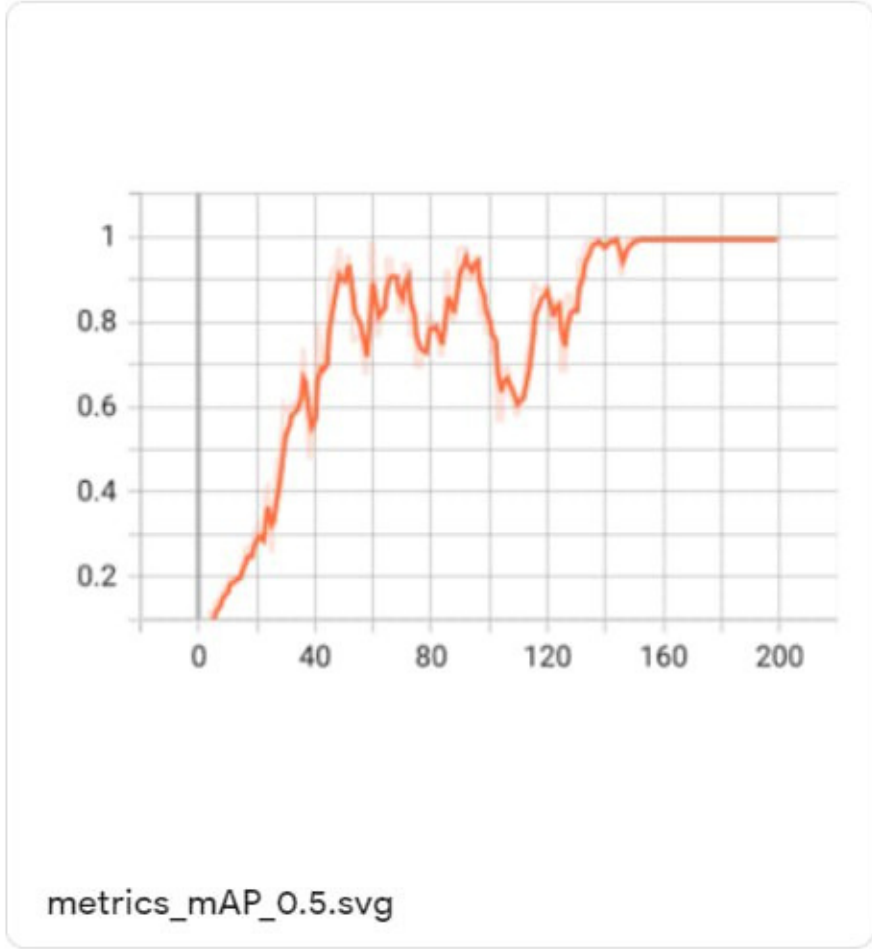
Detection by using YOLO5



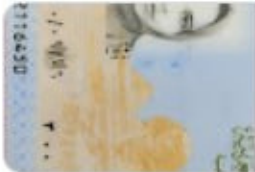


ID Detection Results

ID Detection results



photo_58997568513718187...
Image



20230109_105828.jpg
Image



20230109_105916.jpg
Image



6. Face verification



We used mse



—
We used face net vgg



Cosine similarity: 0.45216697454452515
They are not same person!



Cosine similarity: -1.1920928955078125e-07
They are same person





Final Output



[REDACTED]

| | A | B |
|----|----------------|-------------------------------------|
| 1 | IDs | Names |
| 2 | 29911031600636 | عبدالوهاب محمد ز محمد زيدان محمد |
| 3 | 30007122601934 | حنفي محمد حامد |
| 4 | 30009061700325 | مصطفى عباس يونس |
| 5 | 29909281403084 | |
| 6 | 30009012110709 | نور محمد احمد عبدالوارث |
| 7 | 29906072100776 | ب زكى اسماعيل زكى |
| 8 | 29908290103291 | أروى عصام سعد عبد أ لجو اد |
| 9 | 29908100103045 | روس دميان رفيق تاوضروس |
| 10 | 30001181401908 | عبدالرحمن احمد منصور بدران |
| 11 | 30006272103227 | احمد لطفى هاشم كيلانى |
| 12 | 30002080200959 | كريم عماد فاروق مصطفى |
| 13 | 29808078800239 | سمير عبدالقادر عبدالعليم طليل |
| 14 | 30005202100178 | أشرف متولى حسن |



Thank You
—