

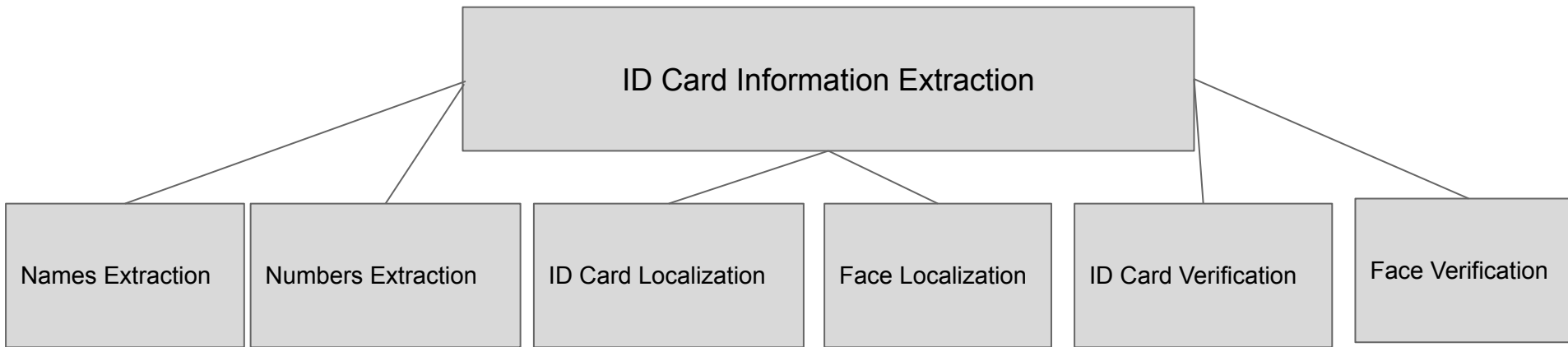
Egyptian ID Card Information Extraction

Ahmed Abdulwahab Zidane
Abdelrahman Ahmed Mansour
Arwa Essam Saad
Nour Mohamed Ahmed
Salma Zakaria Hassan

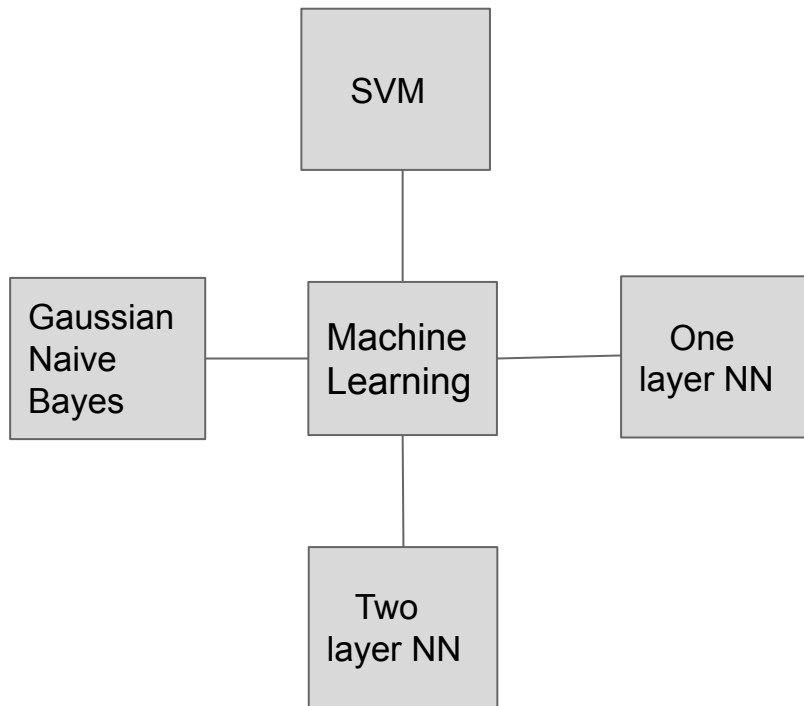
Problems

- No Egyptian ID cards Dataset (only 13)
- Little Resources for Arabic OCR

Project Milestones



OCR - Pretrained Machine Learning Models



Label	Output
احمد عبدالوهاب	ا حكد حيد ا لوها سث
احمد عبدالوهاب	ا حكد حكي ا لوظا نم
احمد عبدالوهاب احمد عبدالوهاب	ا نسد نيش ا دوبا سي ا نسد نيش ا درهاني
دميان	داميطوك

Dataset

ا	12/21/2019 7:27 AM	File folder
ب	12/21/2019 7:39 AM	File folder
ت	12/21/2019 7:48 AM	File folder
ث	12/21/2019 7:51 AM	File folder
ج	12/21/2019 8:02 AM	File folder
ح	12/21/2019 8:09 AM	File folder
خ	12/21/2019 8:18 AM	File folder
د	12/21/2019 8:23 AM	File folder
ذ	12/21/2019 8:30 AM	File folder
ر	12/21/2019 8:30 AM	File folder
ز	12/21/2019 8:35 AM	File folder
س	12/21/2019 8:42 AM	File folder
ش	12/21/2019 9:18 AM	File folder
ص	12/21/2019 9:22 AM	File folder
ض	12/21/2019 9:27 AM	File folder
ط	12/21/2019 9:31 AM	File folder
ظ	12/21/2019 9:35 AM	File folder
ع	12/21/2019 9:42 AM	File folder
غ	12/21/2019 9:47 AM	File folder
ف	12/21/2019 9:52 AM	File folder
ق	12/21/2019 9:55 AM	File folder
ك	12/21/2019 9:59 AM	File folder
ل	12/21/2019 10:04 AM	File folder
لا	12/21/2019 10:09 AM	File folder
م	12/21/2019 10:11 AM	File folder
ن	12/21/2019 10:15 AM	File folder
ه	12/21/2019 10:17 AM	File folder
و	12/21/2019 10:20 AM	File folder
ي	12/21/2019 10:44 AM	File folder

Results on Large texts

احذروا المهلكات التحريش بين الممن وبين البهام والدواب صح حدن النبي صل الله عليه وسلم نه قال ن الشيطان قد يس ن يسده المصلون في ضير العرب ولكن في التحريش بينهم فكل من حر غثل بين اثنين من بني دم ونقل بينهما ما يذي حدهما فهو نمام من حزب الشيطان من شر الناس كما قال النبي صل الله عليه ولطم لا خبركم بشراكم قالوا بل يا ردظل الله قال شراركم المشاؤون بالنميم المفتدون بين الحب الباغون للبر العنت والعنت المولقا وصح عن رسول الله صل الله عليه وللم نه قال لا يدخل الجن نمام والنمام هو الذي ينقل الحديث بين الناس وبين اثنين بما يذي حدهما و يوحش قلبه جعل صاحبه و صديقه بن يقول له قال حعنك فلان كذا وكذا وفعل كذا وكذا لا ن يكون في ذلك مصلح و فاد كتحذيره من شر يحدث و يترتب وما التحريش بين البهام والدواب والطير وغيرهما فحرام كمنافر الديوك ونطاح الكباش والثيران وتحريش الكلاب بعضها حلل بعضل وما شبه ذلك وقد نه رسول الله صل الله عليه وسلم عن ذلليا فمن فعل ذلك فهو عاصل لله ورسوله ومن ذلك فساد قلب المر عل زوجها والعط عل سطره لما روي ن رسول الله صل الله عليه وسلم قال ملعون من خيب امر عل زهها و عبدا اعل سطره نعوذ بالله من ذلك ونسله تعال ن يجعلنا من الذين يستمعون القول فيتعوبون حسنه ولا يجعلنا من الذين يستمعون فيضلون عنه سوا السبيل نه ولي ذلك والقادر عليه وهو يقول الحق ويهدي السبيل

Image Preprocessing and OCR

Image Preprocessing

- Resize images to the same size. (1280, 819)
- Convert images to gray.
- Convert images to dark black and white.
- Blurring images.
- Cropping images :
 - Around the full name into 1 image.
 - Around the first and last name separately into 2 images.

Approach

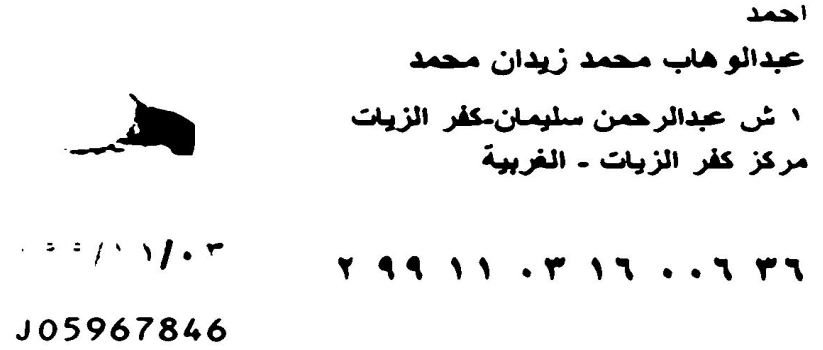
Write Numbers 123456

Write Numbers 123456

كتابة الأرقام ١٢٣٤٥٦٧٨٩٠

كتابة أرقام عربي وانجليزى

فى صفحة ورد واحدة



عبدالوهاب محمد زيدان محمد

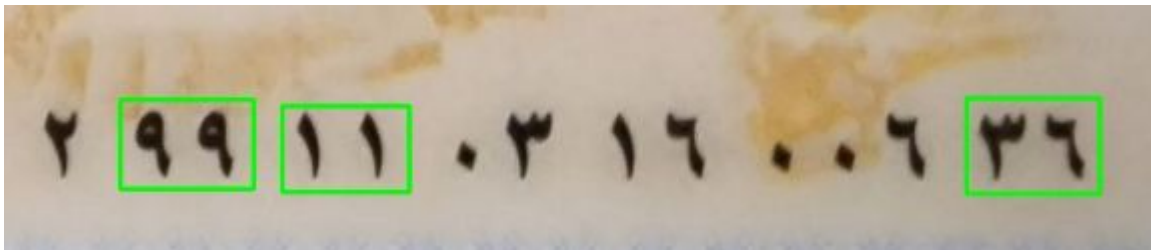
احمد

OCR - Pretrained Models

- Tesseract
 - Arabic numbers
 - Arabic Names
- Arabic OCR
 - Arabic numbers
 - Arabic Names

OCR - Pretrained Models (low accuracy)

- ArabicOCR for arabic number
 - ID numbers were not captured even after preprocessing.



OCR - Pretrained Models (low accuracy)

- Tesseract for arabic characters
 - First names sometimes were not detected.

['رفیق تاوضروس دمیان ' ,
'ساره زکی اسماعیل زکی' ,
'أشرف متولی حسن ' ,
'ب محمد ریدان محمدn' عبدالو هات ' ,
'نور حنفی محمد حامد' ,
'محمد احمد عبدالوارث ' ,
'میان ا . ' ,
'کریم عصام سعد عبدالجواد' ,
'سمیر عبدالقادر عبدالعلیم طویل ' ,
'أروی احمد لطفی هاتم کیلانی' ,
'عماد فاروق مصطفی ' ,
'مصطفی عباس یونس ' ,
'عبدالرحمن احمد منصور بدران']

OCR - Pretrained Models (high accuracy)

- Tesseract for arabic number
 - All ids were captured correctly.

	ids
0	٢٩٩١١.٣١٦.٠٦٣٦
1	٣٠٠.٧١٢٢٦.١٩٣٤
2	٣٠٠.٩.٦١٧.٠٣٢٥
3	٢٩٩.٩٢٨١٤.٣.٨٤
4	٣٠٠.٩.١٢١١.٧.٩
5	٢٩٩.٦.٧٢١.٠.٧٧٦
6	٢٩٩.٨١.٠.١.٣.٤٥
7	٣٠٠.١١٨١٤.١٩.٨
8	٣٠٠.٦٢٧٢١.٣٢٢٧
9	٣٠٠.٢.٨.٢.٠.٩٥٩
10	٢٩٨.٨.٧٨٨.٠.٢٣٩
11	٣٠٠.٥٢.٢١.٠.١٧٨
12	٢٩٩.٨٢٩.١.٣٢٩١

OCR - Pretrained Models (better accuracy)

- Tesseract for arabic characters
 - Better results than ArabicOCR

['احمد عبدالوهاب محمد زيدان محمد',
'دميان ضرروس دميان',
'ساره أشرف متولي حسن',
'ميان محمد حامل',
'نور محمد احمد عبدالوا',
'حذيفه اسماعيل زكي',
'سلمي و',
'أروى عبدالجواد',
'ساره سمير عبدالقادر عبدالعليم طابيل',
'محمل حمل',
'كريم عما',
'عمد عياس يونس',
'عبدالرحمن احمد منصور بدران']

Final results

	A	B
1	ids	Names
2	29911031600636	أحمد عبدالوهاب محمد زبدان
3	30007122601934	دميان ضرورس دميان
4	30009061700325	ساره اشرف متولى حسن
5	29909281403084	مهيار محمد حامل
6	30009012110709	لور محمد أحمد عبدالوا
7	29906072100776	حنيفة اسماعيل زكى
8	29908100103045	سلمى و
9	30001181401908	أروى عبدالجوك
10	30006272103227	ساره سمير عبدالقادر عبدالعليم
11	30002080200959	محمّد احمّل
12	29808078800239	كريم عمار
13	30005202100178	عبد عيسى يونس
14	29908290103291	عبدالرحمن أحمد منصور بدران

احذروا المهلكات التحريش بين الممن وبين البهائم والدواب صح عن النبي صلى الله عليه وسلم انه قال ان الشيطان قد يس ن يعبد المصلون في جرير العرب ولكن في التحريش بينهم فكل من حرش بين اثنين من بني دم ونقل بينهما ما يذي حدهما فهو نمام من حزب الشيطان من شر الناس كما قال النبي صلى الله عليه وسلم لا خبركم بشراكم قالوا بل يا رسول الله قال شراركم المشاؤون بالنميم المفسدون بين الحب الباغون للبر العنت والعنت المشق وصح عن رسول الله صلى الله عليه وسلم انه قال لا يدخل الجن نمام والنمام هو الذي ينقل الحديث بين الناس وبين اثنين بما يذي حدهما ويوحش قلبه على صاحبه وصديقه بن يقول له قال عنك فلان كذا وكذا وفعل كذا وكذا لا يكون في ذلك مصلح وفاد كتحديده من شر يحدث ويترتب وما التحريش بين البهائم والدواب والطير وغيرهما فحرام كمنافر الديوك ونطاح الكباش والثيران وتحريش الكلاب بعضها على بعض وما شبه ذلك وقد نه رسول الله صلى الله عليه وسلم عن ذلك فمن فعل ذلك فهو عاصي لله ورسوله ومن ذلك فساد قلب المؤمن على زوجها والعبد على سيده لما روي عن رسول الله صلى الله عليه وسلم قال ملعون من خيب امره على زوجها وعبد على سيده نعوذ بالله من ذلك ونسله تعالى ن يجعلنا من الذين يستمعون القول فيتبعون حسنه ولا يجعلنا من الذين يستمعون فيضلون عنه سوا السبيل نه ولي ذلك والقادر عليه وهو يقول الحق ويهدي السبيل

Data Annotation

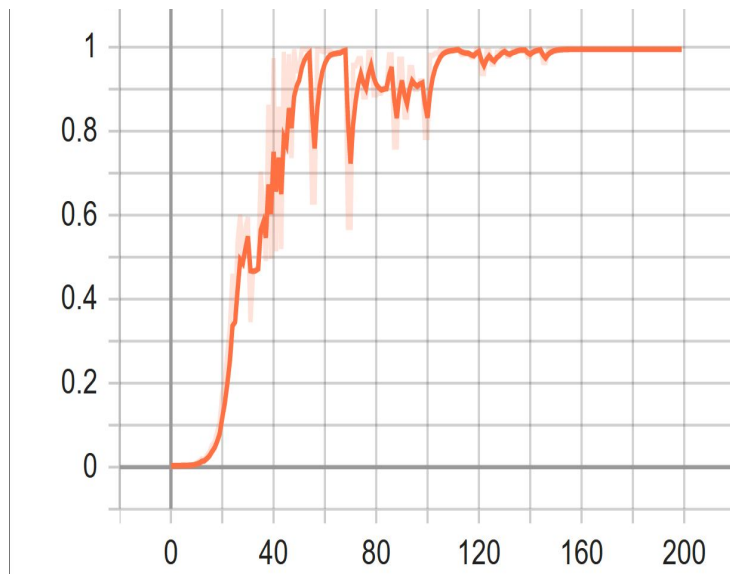
Data annotation using make sense ai



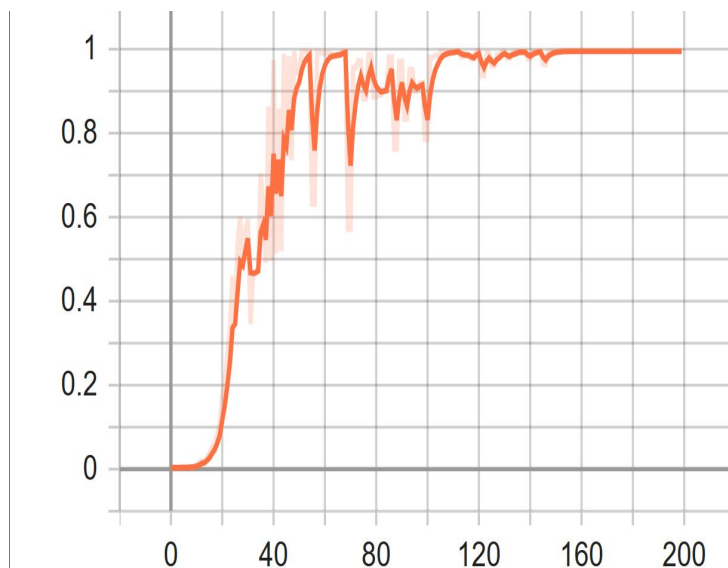
Face & ID Detection

Face Detection & ID Detection using YOLOv5

Face Detection Results

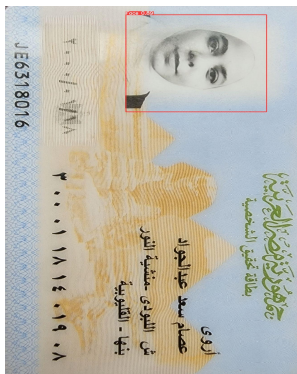
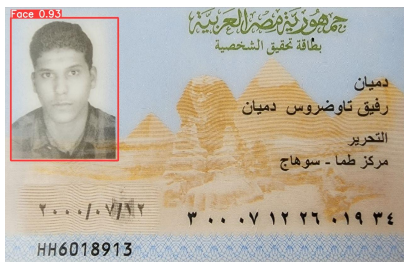


ID Detection Results

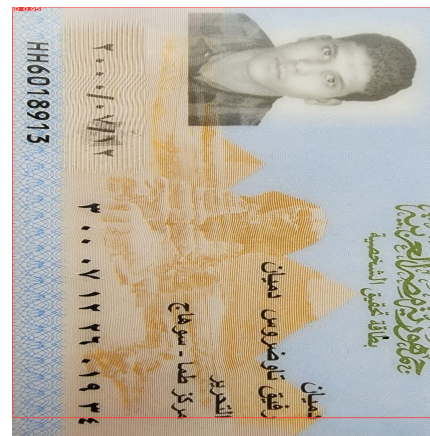
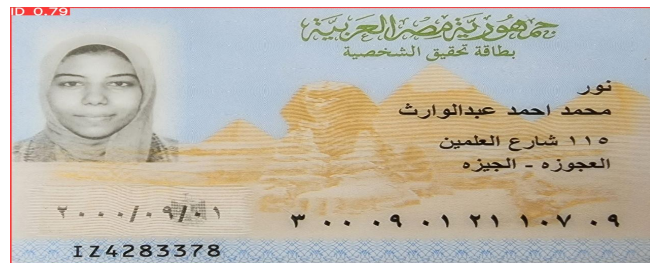


Face Detection & ID Detection using YOLOv5

Face Detection Outputs



ID Detection Outputs



Face & ID Verification

Face Verification Using MSE

```
: h, w = img_1.shape                # getting the height and width of the image
scale_percent = 60                 # percent of original size
width = int(img1.shape[1] * scale_percent / 100) # width normalization
height = int(img1.shape[0] * scale_percent / 100) # width normalization
dim = (width, height)              # diminsions as a set of width and length

# resize image
resized_1 = cv2.resize(img_1, dim, interpolation = cv2.INTER_AREA)
```

Trying Image Verification

```
] new_images = []

for i in range(len(list_images)):
    read = cv2.imread(list_images[i])           # read images
    resized = cv2.resize(read, dim, interpolation = cv2.INTER_AREA) # resize images
    gray = cv2.cvtColor(resized, cv2.COLOR_BGR2GRAY) # Converting images to gray
    new_images.append(gray)

print(new_images)
```

```
] for i in range(len(new_images)):
    print(new_images[i].shape)
```

```
] len(list_images)
```



```

: def mse(resized1, resized2):
    ...
    Input: Two images
    Operation: Apply mean square error on the two images
    output: error between the images

    ...

    diff = cv2.subtract(resized1, resized2)
    err = np.sum(diff**2)
    mse = err/(float(h*w))

    return mse

```

```

: errors = [] # List of the errors
import numpy as np

for i in range(len(new_images)):
    # print(new_images[i].shape)
    errors.append(mse(resized1, new_images[i]))
    m = min(errors)
print(errors.index(m)) # get the image location with lowest error

```

ID Verification - Same Approach

Face & ID Verification using VGG Face

```
def preprocess_image(image_path):  
    img = load_img(image_path, target_size=(224, 224))  
    img = img_to_array(img)  
    img = np.expand_dims(img, axis=0)  
    img = preprocess_input(img)  
    return img  
  
def find_cosine_similarity(source_representation, test_representation):  
    a = np.matmul(np.transpose(source_representation), test_representation)  
    b = np.sum(np.multiply(source_representation, source_representation))  
    c = np.sum(np.multiply(test_representation, test_representation))  
    return 1 - (a / (np.sqrt(b) * np.sqrt(c)))  
  
vgg_face_descriptor = Model(inputs=model.layers[0].input, outputs=model.layers[-2].output)
```

```
epsilon = 0.4  
  
def verify_face(img1, img2):  
    img1_representation = vgg_face_descriptor.predict(preprocess_image(img1))[0, :]  
    img2_representation = vgg_face_descriptor.predict(preprocess_image(img2))[0, :]  
    cosine_similarity = find_cosine_similarity(img1_representation, img2_representation)  
  
    f = plt.figure()  
    f.add_subplot(1, 2, 1)  
    plt.imshow(image.load_img(img1))  
    plt.xticks([]); plt.yticks([])  
    f.add_subplot(1, 2, 2)  
    plt.imshow(image.load_img(img2))  
    plt.xticks([]); plt.yticks([])  
    plt.show(block=True)  
  
    print("Cosine similarity: ", cosine_similarity)  
  
    if (cosine_similarity < epsilon):  
        print("They are same person")  
    else:  
        print("They are not same person!")
```

Face Verification using VGG Face

Face Verification Results:



Cosine similarity: 0.4464693069458008
They are not same person!



Cosine similarity: 0.0
They are same person

Face Verification using VGG Face

ID Verification Results:



Cosine similarity: 0.7341
They are not same person!



Cosine similarity: 0.0
They are same person

References

- [Arabic-OCR](#)
- [An Efficient, Font Independent Word and Character Segmentation Algorithm for Printed Arabic Text.](#)
- [A Robust Line Segmentation Algorithm for Arabic Printed Text with Diacritics.](#)
- [yolov5](#)
- [ArabicOcr](#)
- [Arabic Character Segmentation Using Projection Based Approach with Profile's Amplitude Filter.](#)
- [keras-vggface](#)
-

References

- [tesseract-ocr](#)