Artificial Intelligence Certification

Summer 2022

Capstone Project:

Document Liveness Detection

Prepared by: Yasmine Maarbani &
Hisham Yassine

Mentor: Mr. Anis Ismael

Table of Contents

I- Problem Definition

a. The Mission

Nowadays, e-services are witnessing a remarkable growth and evolution aiming to provide a state-of-the-art users/customers experience. Knowing that some e-services require a critical user authorization, a new stipulation appears: the need to show identification documents through the device camera of the user to confirm his ID.

The aim of this project is to build an AI model that is capable to detect if the submitted ID was truly in the possession of the user (which we refer to as “Live Document”) or a simple copy (physical or digital) was shown in an act of forgery.

<table>
<tr><th colspan="3">Live Document</th></tr>
<tr><td colspan="3">  </td></tr>
<tr><th colspan="3">Fake Copies</th></tr>
<tr><td>  </td><td>  </td><td>  </td></tr>
<tr><th>Gray Copy</th><th>Colored Copy</th><th>Recaptured Copy</th></tr>
</table>

b. Evaluation Criteria

The model performance will be assessed based on two main criteria:

1. The Live/Fake prediction accuracy
2. The value of the False Acceptance Rate (FAR) and False Rejection Rate (FRR): FAR should be lower than FRR by a multiple of 100

c. Dataset

DLC-2021 (Document Liveness Challenge 2021) dataset will be used for the purpose of this project.

As mentioned in the article titled "Document Liveness Challenge Dataset" [Dmitry V. Polevoy et al], the novelty of this dataset is that it contains shots from video with color laminated mock ID documents, color unlaminated copies, grayscale unlaminated copies, and screen recaptures of the documents. From an ethical perspective, the proposed dataset complies with the GDPR (General Data Protection Regulation) because it contains images of synthetic IDs with generated owner photos and artificial personal information.

Image frames were randomly selected from this dataset and sorted between Live and Fake. Vis-a-vis every 3 live frames, 3 fake copies were retrieved: one frame from each fake category (gray copies, colored copies and recaptured copies) so that a data kept well balanced.

II-Previous Solutions

a. Technique in Use

When it comes to performance, neural networks became the default option in many cases especially when a complicated problem is put in place having a large dataset and sufficient computational resources. Accordingly, previous solutions related to Document Liveness Detection were based on CNN using the same suitable database reaching, according the above-mentioned article [Dmitry V. Polevoy et al] where a ResNet-50 is applied, the following results:

CNN – ResNet-50			
Metrics	Recaptured Copies	Colored Copies	Gray Copies
Accuracy	89.67%	83.61%	Failed
Precision	85.89%	96.01%	
Recall	89.03%	85.56%	

b. Challenges

The use the convolutional neural networks has shown two main disadvantages:

1. The need for extensive computational resources which may not be practical for edge deployment or scenarios where limited resources are available (micro-chips for example...). That's why we were required to present a new technique that does not rely on neural networks or shallow NN in our project.

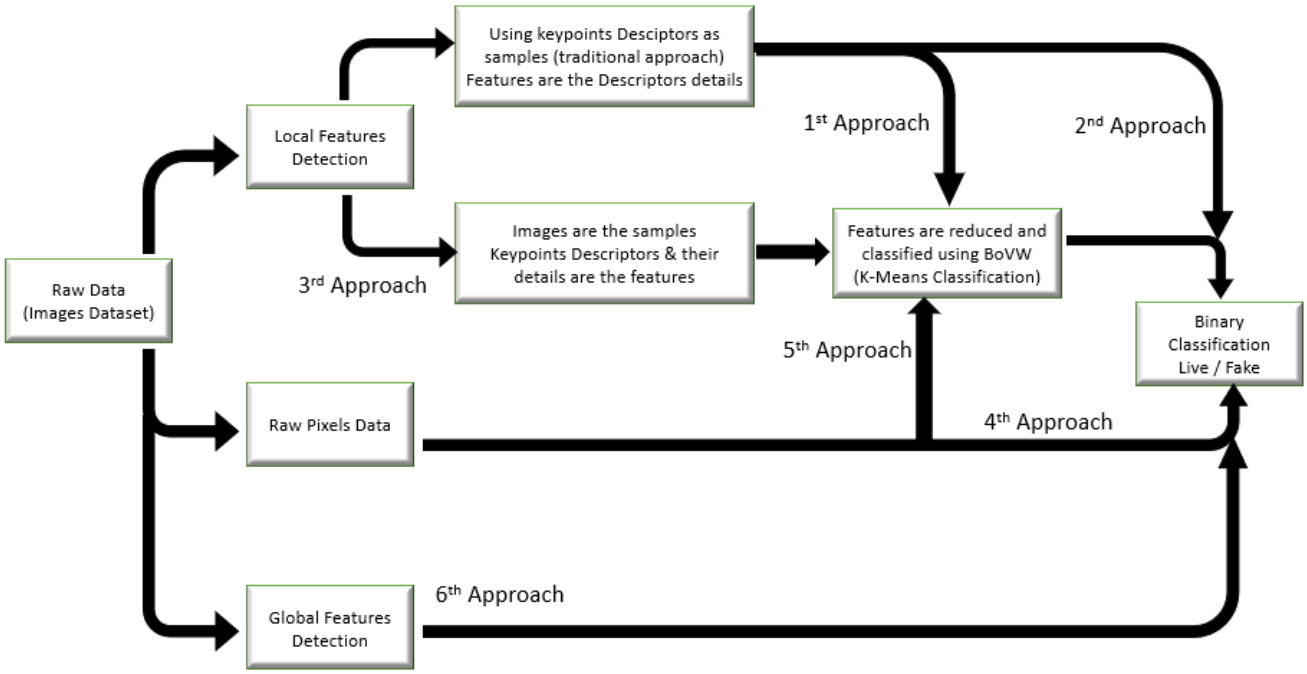
2. Non-sophisticated CNNs were ignoring color features which explains the reason why ResNet-50 and similar or simpler models failed to examine gray samples: models either did not train at all or were overfitted.

III- New Proposed Solution

a. Suggested Approaches

Since CNN is discouraged to be used to accomplish our mission, traditional Computer Vision techniques were examined to solve the problem. Three main approaches were presented in this regard in an attempt to reach the best results in terms of accuracy, FAR (False Acceptance Rate) and FRR (False Rejection Rate).

The following diagram shows the different approaches that has been tackled in our project:



It’s important to mention that we have applied numerous optimizations/solutions in order to reach the best metrics of each approach including such as:

1- Reducing features dimensionality using PCA and BoF (Bag of Features)
2- Background removal especially while extracting local features so that the local extractor ORB (Oriented FAST and Rotated BRIEF) focuses on the ID document itself regardless of the background
3- The use of different classification algorithms (SVM, Logistic Regression, XGBoost, RandomForest, etc.)
4- Hyperparameters Tuning using GridSearch
5- Data Balancing and cross validation using KFold technique…

b. Chosen Technique

Since FAR and FRR are directly related to the accuracy, we focused mainly on the accuracy score reached by each approach/technique. Although our efforts were concentrated at the beginning on the Local Features Extraction, experiments show that using raw pixels data (considering the pixels values as direct features for the model) registered a better performance and finally building our model based on Global Features guided us to even better prediction results.
Accordingly, our final model has been structured using the global features provided by OpenCV library. This technique examines the whole picture as one complete sample in opposition to the local features that detect numerous detailed key points and their corresponding descriptors inside each image.

Using Global Features, we succeeded to designate the following images characteristics:

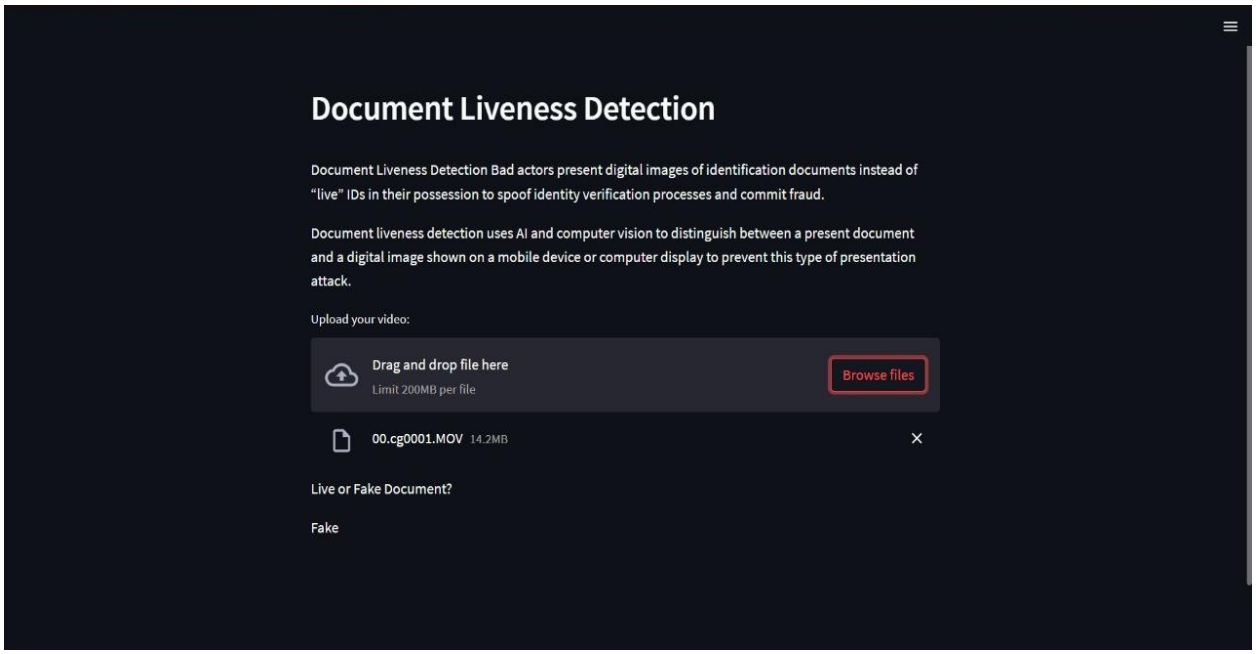
1- The Shape: HuMoments technique was applied (as part of OpenCV cv2)
2- The Colors: Images colorspace were converted to HSV and then colors histograms were calculated, normalized and flattened
3- The Texture: Haralick technique was applied (as part of Mahotas library)

The classification was performed using a Random Forest classifier as it achieved the highest accuracy score and best confusion matrix. It’s important to note that a regular Logistic Regression classifier has also shown exceptional results.

c. Model Deployment

- A basic web page was designed and linked to our model using Flask technique and actual deployment was achieved using streamlit.
- The above-mentioned web page allows the user to upload either an image or a video showing the claimed identification document.
- The image will be then processed by the model that is responsible to detect whether the presented ID is Live or Fake. The result will be shown on the same web page.
- In case a video was uploaded, the system will be responsible to extract one image frame before being processed by the model similar to what was described just above

Example shown as follows:



Deployment Public Link:

https://yasmine-maarbani-document-liveness-detection-app-aq9mxz.streamlit.app/

IV- Results

a. Our Solution Outcomes

- Accuracy Score: 90%
- Confusion Matrix:

243 (TN)	**0** (FP)
50 (FN)	**166** (TP)

- Classification Report:

	precision	recall	f1-score	support
0	0.84	1.00	0.91	243
1	0.99	0.78	0.88	216
accuracy			0.90	459
macro avg	0.92	0.89	0.89	459
weighted avg	0.91	0.90	0.89	459

- FAR and FRR

- FAR = 1 – TNR = 1- Specificity = 8.23%
- FRR = 1 – TPR = 1- Recall = 1.85%

b. Comparison with the Previous Solution

Our new proposed solution can be compared to the previously presented one based on three main criteria: (a) the resulting metrics, (b) the computational resources and (c) the reliability:

		CNN	Traditional Global Features Extraction
a	Metrics	Accuracy: 86.64% (mean value for colored and recaptured copies)	Accuracy: 94%
b	Computational Resources	CNN contained 7 conv2d layers	No NN
c	Reliability	Failed on Fake Gray Copies	Trained on all types of documents including the gray copies that represented 25% of the dataset

c. Implication on Real-World Scenarios

Achieving a model that have the above stated characteristics will allow a great evolution in the e-services world. In fact, being able to detect the liveness of the uploaded documents is helping the emerging of services that cannot be limited to a username/password authentication, rather we will be capable to include banking, governmental, airport e-services, etc. where the actual official ID documents are a necessity.

Also, the ability of having such a robust model without the need for a neural-network structure will allow to implement these e-services on edge devices or in environment of limited resources like micro-chips for example.

In brief, a better customer/user experience and more profitable business.

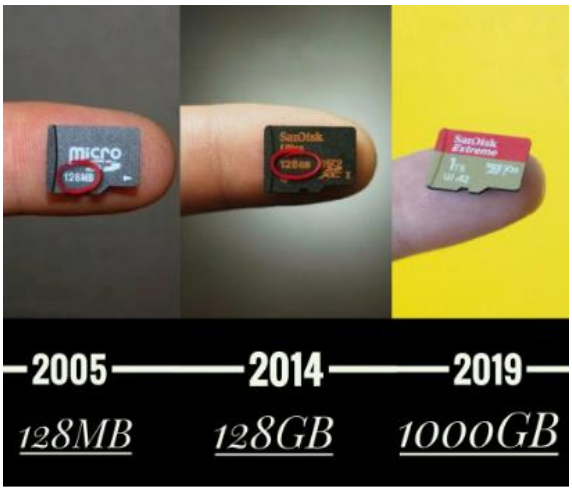
V- Potential Advancement

a. Improving Our Solution

While global features extraction has proven a better performance for our specific mission comparing to local features extraction, new researches are conducted to bring together both techniques i.e. combining both global and local features to understand better images: local features will then focus on specific internal characteristics (borders, corners, …) whereas global ones will focus on general patterns and traits of the whole image.
The module would be robust enough to predict any kind of presented ID and still with no need for a resources-demanding neural network.

b. Applying New Techniques

Thinking for future should be aligned with tech growth taking into consideration that not only AI techniques are developing, but also the corresponding resource are witnessing contentious and quick development. The following image is just an example of the enormous scale of progress we are making (removable memory devices for instance):



What we are trying to say that we do encourage the use of newly state-of-the-art technique trying to optimize/minimize their demand for computational resources in order to benefit from their exceptional performance such as the Vision image Transformers (ViT) that are making their fast step into the domain nowadays.