ELIS CONSULTING & LABS

# Automated Identity Verification


## Abstract

Nowadays, the line separating physical world from digital one is getting thinner. In this context, ensuring reliability and security in the digital world, dominated by anonymity, is a priority for companies that offer online services. ELIS Consulting & Labs has recently developed an Artificial Intelligence-based system that automates the identification of users registering to an online service. The system was designed into two steps: Integrity Check and Face Detection and Matching. For each of them, we report the deep learning-based methods that have been adopted.




Erasmus+
Key competences for an European model of Industry 4.0
**i4EU Project**

Learn more about us here:
https://www.i4eu-pro.eu

Co-funded by the
Erasmus+ Programme
of the European Union

# Introduction

Nowadays, identity verification plays an important role. In fact, in a lot of private and public sectors the identity validation is required for access control, international border or physical security. The boom in Internet usage occurred in the past two decades, leading to more frequent usage of daily online transactions for payments, data exchange, e-commerce, money transfers etc. All these online operations require an authentication strategy, in which the risks of fraud and identity theft became minimal. Consequently, a new need emerged: seeking trust and security in the online services, where anonymity is predominant. Hence, identifying interaction between people is crucial for both the public and private sectors. In response to these phenomena, some governments allowed their citizens to prove their identity in the Internet world. In this scenario, manifesting electronic identity became crucial to access all the information registered in the state databases or to access to private online services. To clarify, electronic identification is referred to a digital way demonstrate the identity of a person or organizations, with the aim of executing online transactions.

Through this case study, we describe the system developed by ELIS Consulting & Labs to solve the problem of digital identity verification for a big italian corporate. Specifically, the goal was to automate the authentication of new users registering to an online platform.

# Challenges

The process of verifying digital identity, is not a onetime process. The specific phases that take place in different temporal instants. These steps are the following:

- *Loading Identity document* from the customer who wants to use an online service.
- *Validation* by the authority or company that offers the online service. It includes document authenticity and validity determination.
- *Verification* of the matching between the identity declared in the document and the real person.
- *Authentication* by verifying manually the information provided by the user with that ones present in the uploaded document.
- *Vetting* of the user's profile.

The steps required to develop this system can be summarized as follows: (i) integrity check of the document, i.e., the uploaded images contain a document id that is completely visible and valid, and (ii) a face matching step in which the user has to demonstrate that he or she owns the document.

To solve this problem, several challenges must be overcome. First, you need to train a computer vision model to recognize documents. Next, it is necessary to extract the text from the image to process the data contained within it. Finally, the model must be able to recognize faces and compare them with each other.

> *To solve this problem, several challenges must be overcome. First, you need to design a computer vision model to recognize documents. Next, you should find a way to extract the text from the image to process the data contained within it. Finally, the model must be able to recognize faces and compare them with each other.*

# Solution

First of all, identifying the problem in an appropriate dimension is fundamental for any machine learning application as it impacts over all the subsequent phases from the initial choice to the implementation and testing of the model. For each task, documents chosen to be accepted for the online user identification are: identity card, driving license and passport. The tasks executed for

automated the identity document verification, can be treated as binary classification problems, one of the most common machine learning tasks on vector data. The model differentiates between two classes: valid/invalid and accepted/rejected.

## Integrity check

For the integrity task, the model selection activity takes into consideration the two sub-problems of integrity check, namely legibility and completeness. Choosing an architecture to perform this task is actually an activity of selecting multiple architectures to be implemented to solve the whole task. The three models identified for the entire integrity checking process is illustrated in Figure 1. Firstly, to estimate the image quality ELIS implemented a quality estimator model. This is selected to address the legibility sub-task. Secondly, the input image is segmented by cutting out the background through a U-Net network [1]. After that, a NetVLAD [2] is executed to extract meaningful features, including a final Fully Connected Network that act as a classifier to determine whether the uploaded image is complete or not. The used dataset is MIDV-500, downloaded from a public domain and therefore available for its application. MIDV-500 is a dataset which shows the video frames of several kinds of identity documents in different conditions. This dataset was chosen over the other available ones mainly for the quantity and the completeness of images, which made it particularly suitable for developing a recognition architecture for identity images. The Dataset Analysis is performed to identify dataset information before using it. It contains 50 different identity document types, including 17 types of identity cards, 13 types of driving licenses, 14 types of passports and 6 other identity documents of various countries. For each document, it is shoot a video using 2 mobile devices: Apple iPhone 5 and Samsung Galaxy S3. For each document 10 videos are obtained, in 5 different conditions. The shooting situations are the follows: Table, Hand, Keyboard, Clutter and. Each video is shot for at least 3 seconds and the first 3 seconds of each video is split into 10 frames per second. Therefore, having shot 500 videos (50 documents) x (5 conditions) × (2 devices), 15000 images are obtained in the dataset, with a resolution of 1080 × 1920 pixels for each image. Furthermore, for each document, the standard image of the aforementioned document is also

shown, for the sake of clarity.

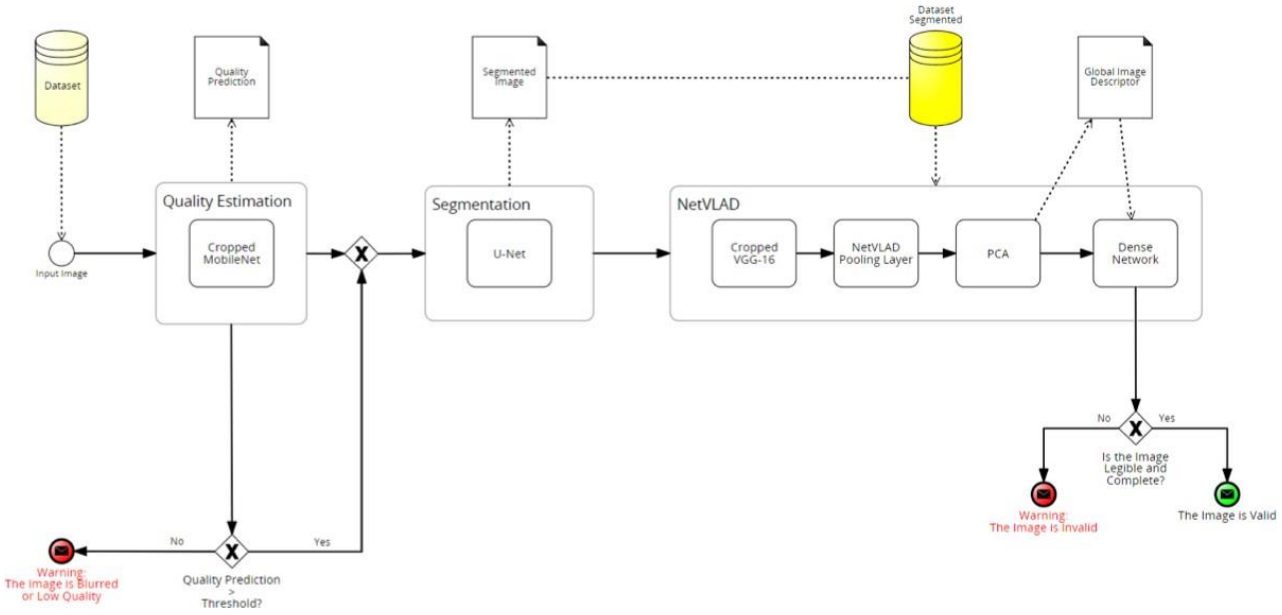


**Figure 1 Integrity check pipeline**

To execute the integrity check task on identity documents, a cascade approach is used which consists, at the operational level, of three phases, as shown in the pipeline in Figure 1.

1. **Quality Estimation**: the first step for accepting or discarding the uploaded image is a usability verification of the image based on its quality. A prediction model for the technical quality estimation is implemented, based on the work done by Google and presented in the paper [3]. This architecture returns a quality score in the range (1, 10), which is used as the first pass/fail filter to pass the image in the subsequent steps. If this score is lower than a certain established threshold, it means that the image is not legible and it is returned a warning message, requesting a new shot;
2. **Segmentation**: once the image legibility is ensured, its completeness must also be guaranteed. Before moving in that direction, the image is given as input to an image segmentation model that uses a U-Net to generate an identity document mask. These masks are applied to the image, to obtain new segmented image, in which the background has been removed and only the part of the image with the identity document is present. This step is used to allow the next model to better extract the image features related to the document part;
3. **NetVLAD**: segmented images are fed to a VGG-16 CNN for the extraction of local image features. In this architecture, the last pooling layer is replaced with a NetVLAD layer (Vector of Locally Aggregated Descriptors), which allows to group the extracted local descriptors into a

global image representation (Global Image Descriptors GID). GIDs are classified using a Fully Connected Network (consisting in a 3 fully connected layer) which returns as an output whether the image is valid or not. If not, it is shown a warning message, requesting a new shot from the user.

## Face Detection and Matching

The final task to be performed is Face Detection and Matching, which is used to make sure that the person who claims the document ownership is the real possessor. To execute the Face Detection and Matching task on identity documents, a cascade approach is used which consists, at the operational level, of three phases, as shown in the pipeline in Figure 2:

1. **Face and Landmark Detection**: this first phase aims to detect the two faces in the uploaded image that correspond to the user's selfie and the identification photo in the identity document. To do this, it is used an MTCNN [4] which is able to detect faces and to place 5 landmarks identifying key points in the faces.
2. **Face Embedding**: the second step is the image transformation into a features vector that represent it. Before doing this task, a face alignment activity is needed to align the faces in the event that the user or the photo in the document is not aligned with the standard reference system. This activity is fundamental as faces with different incli- nations correspond to different vectors and therefore there could be errors in the final classification. Subsequently, an image flip activity is performed and a mirror image is generated. The two images are added together. In this way, the corresponding feature vector will be a more robust vector with a more decisive direction. Finally, the vector is normalized to a vector of length equal to 1.
3. **Classification**: this last step aims to determine whether the two images correspond to the same person or if they are two different people. To do this, the cosine between the two vectors is calculated using the cosine similarity method. Once a threshold is established, it is possible to classify the two photos as belonging to the same person, if the cosine similarity between the two vectors is above a certain threshold, or as belonging to different people.

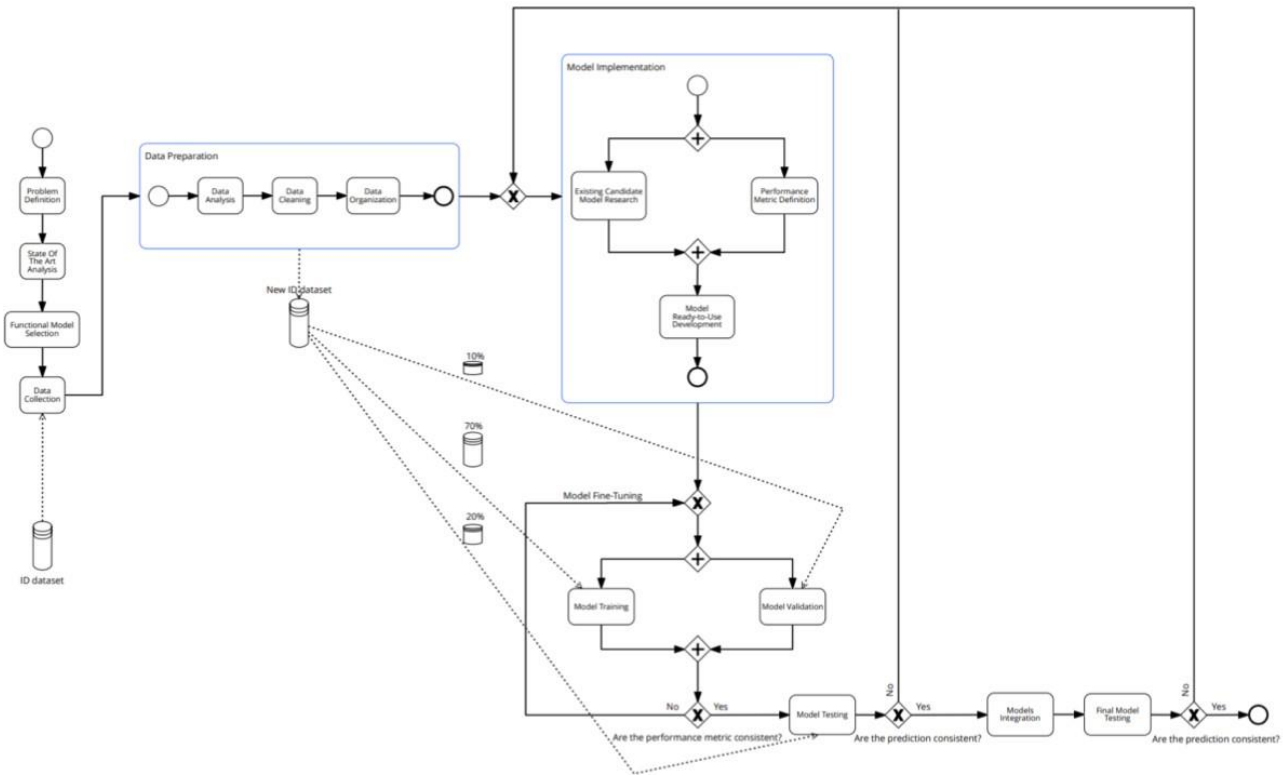


**Figure 2 Face Detection and Matching Pipeline**

# References


1. OlafRonneberger,PhilippFischer,andThomasBrox.U-net:Convolutionalnetworksfor biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.
2. Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5297–5307, 2016.
3. Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. IEEE Transac- tions on Image Processing, 27(8):3998–4011, 2018.
4. Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Let- ters, 23(10):1499–1503, 2016.


This showcase has been collected in the framework of the Erasmus+ project Key competences for an European model of Industry 4.0 (pr. n° 2019-1-FR01-KA202-062965), funded by European Commission.
For more information: www.i4eu-pro.euLegal notice: This publication / communication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.