
Celebrity Face Recognition System

AI Models & Applications

Team Members: Camara Ibrahima
Diallo Mamadou Ougailou
Taha Rania

Instructor: Enrico Formenti

Academic Year: 2025-2026

Abstract

Face recognition is a key domain of computer vision, widely applied in areas such as security systems, biometric authentication, and identity verification. The objective of this project, carried out within the *AI Models & Applications* course, is to design and implement a complete face recognition system dedicated to celebrity identification.

The proposed approach relies on a structured multi-stage pipeline that combines person detection, precise face localization, deep feature extraction using pre-trained neural networks, and supervised identity classification. Several deep embedding models and configurations are evaluated and compared in order to analyze their performance, robustness, and computational efficiency.

In addition to quantitative evaluation, a graphical user interface is developed to demonstrate the practical usability of the system, supporting both single-image analysis and large-scale batch processing.

Contents

1	Introduction	2
2	Dataset and Data Preparation	2
2.1	Dataset Description	2
2.1.1	Class Imbalance Analysis	2
2.2	Data Collection and Organization	3
2.3	Preprocessing	3
2.4	Dataset Splitting	4
3	Methodology	4
3.1	Overall Pipeline	4
3.2	Project Architecture	5
3.3	Face Detection	5
3.4	Feature Extraction	6
3.5	Identity Classification	6
3.6	Evaluation Strategy	6
4	Application and User Interface	6
4.1	Application Overview	7
4.2	Processing Modes	7
4.3	Pipeline Visualization	7
4.4	Monitoring and User Experience	7
5	Results and Evaluation	7
5.1	Evaluation Protocol	7
5.2	Evaluation Metrics	8
5.3	Quantitative Results	8
5.3.1	Impact of Reference Database Quality	8
5.3.2	Detection Pipeline Analysis	9
5.3.3	Comparison of Feature Extraction Models	11
5.4	Qualitative Analysis	12
5.5	Discussion of Results	13
6	Discussion	13
6.1	Challenges and Limitations	14
7	Conclusion and Perspectives	14

8 Team Members and Contributions	15
9 References	16

1 Introduction

Face recognition is a major research area in computer vision and artificial intelligence. It aims at identifying or verifying the identity of a person from digital images or video streams. Over the past decade, the emergence of deep learning techniques, particularly convolutional neural networks (CNNs), has led to significant improvements in the accuracy and robustness of face recognition systems [1, 2].

These advances have enabled the deployment of face recognition technologies in a wide range of real-world applications, including biometric authentication, access control, surveillance, and human-computer interaction. Modern systems typically rely on deep feature representations, also known as facial embeddings, which capture discriminative information while remaining robust to variations such as pose, illumination, and facial expressions [4, 3].

The objective of this project, conducted within the *AI Models & Applications* course, is to design and implement a complete face recognition application focused on celebrity identification. The proposed system is based on a structured processing pipeline that integrates multiple stages: person and face detection, deep feature extraction using pre-trained models, and supervised classification for identity prediction.

Beyond the implementation of the recognition pipeline, this project also emphasizes experimental evaluation and practical usability. Several models and approaches are compared in order to analyze their respective strengths and limitations in terms of accuracy and computational efficiency. In addition, a graphical user interface is developed to facilitate interaction with the system, enabling both single-image and batch processing modes.

2 Dataset and Data Preparation

2.1 Dataset Description

The dataset used in this project is the *Pins Face Recognition* dataset, publicly available on the Kaggle platform [8]. This dataset is composed of facial images of well-known celebrities and contains a total of 105 distinct identities. Each class corresponds to a single celebrity and includes a variable number of images, reflecting real-world conditions such as data imbalance and diversity in image quality.

The images present significant variations in pose, illumination, facial expression, background, and resolution. Such variability makes the dataset particularly suitable for evaluating the robustness and generalization capabilities of face recognition systems in realistic scenarios.

It is important to note that the dataset images are not limited to tightly cropped faces. Many images contain complex backgrounds, multiple individuals, or partial occlusions. As a result, a dedicated detection pipeline is required to reliably localize persons and extract facial regions before applying face recognition models.

2.1.1 Class Imbalance Analysis

The dataset exhibits a strong class imbalance, with some identities containing nearly 200 images while others have fewer than 30 samples. This imbalance can bias training and evaluation, and partially explains performance variations across subsets.

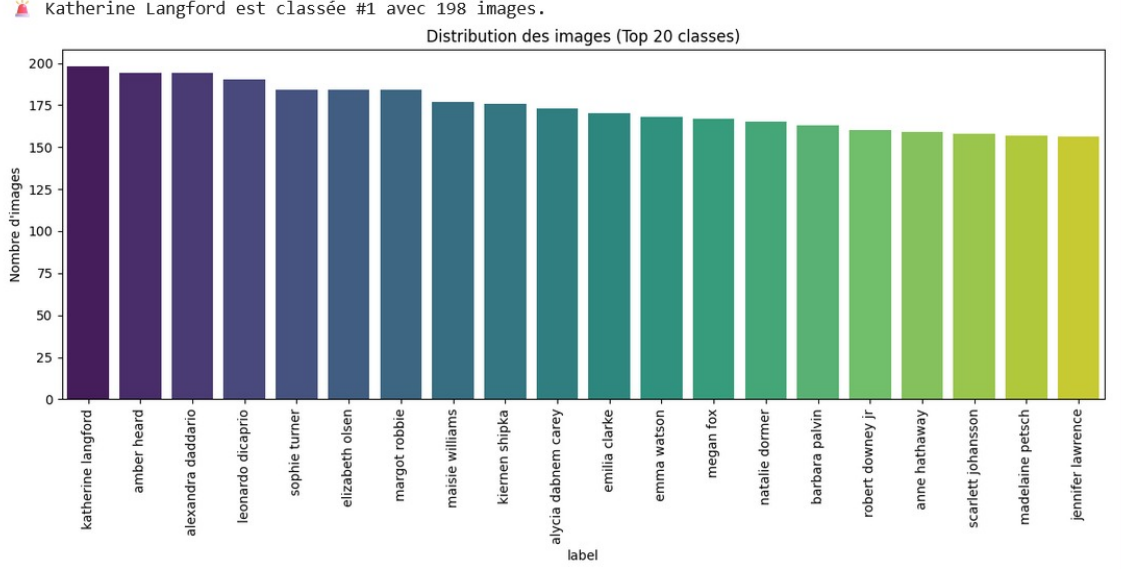


Figure 1: Distribution of images per identity (Top 20 classes), highlighting dataset imbalance.

Top 5 classes	#images	Bottom 5 classes	#images
katherine_langford	198	morgan_freeman	28
amber_heard	194	jeff_bezos	48
alexandra_daddario	194	lionel_messi	48
leonardo_dicaprio	190	mark_zuckerberg	50
sophie_turner	184	wayne_johnson	56

Table 1: Most and least represented identities, confirming strong class imbalance.

2.2 Data Collection and Organization

The dataset was retrieved directly from Kaggle and organized into a structured directory hierarchy, where each subfolder corresponds to a specific identity label. This organization allows for straightforward loading of the data using standard deep learning frameworks and facilitates the management of class labels during training and evaluation.

To ensure reproducibility and proper project management, raw datasets were excluded from the version control system using appropriate configuration files, while all processing scripts and trained models were tracked.

During the preprocessing stage, intermediate working directories are used to store the outputs of each detection step. In particular, person detection results and cropped face images are saved separately in order to facilitate debugging, quality analysis, and reproducibility of the pipeline.

2.3 Preprocessing

The preprocessing stage plays a critical role in the overall performance of the face recognition system. Instead of directly resizing raw images, a multi-step detection and filtering strategy is applied in order to ensure high-quality facial inputs.

First, a person detection model based on YOLO is used to localize human subjects in each image. This step aims to maximize recall by detecting all potential persons, even in complex scenes or cluttered backgrounds.

Second, a dedicated face detection model (RetinaFace) is applied to the person bounding boxes in order to precisely localize facial regions. This stage acts as a quality filter by discarding

low-confidence detections, extremely small faces, and ambiguous regions.

Only faces satisfying minimum size and confidence constraints are retained for subsequent processing. The selected facial images are then resized to a fixed resolution of 224×224 pixels and normalized according to the requirements of the pre-trained deep learning models used for feature extraction.

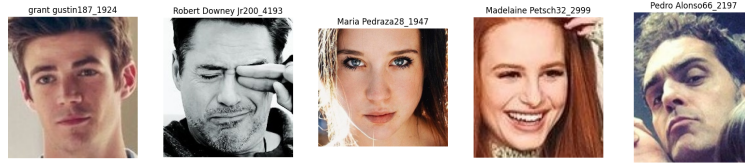


Figure 2: Examples of face detections obtained after the preprocessing stage, illustrating accurate localization and cropping of facial regions in complex scenes.

2.4 Dataset Splitting

The dataset was split into training, validation, and test sets following a 70% / 20% / 10% ratio, respectively. A fixed random seed was used during the splitting process in order to guarantee reproducibility of the experimental results. This separation ensures that model evaluation is performed on previously unseen data, providing a reliable estimate of generalization performance.

3 Methodology

This section describes the methodology adopted to design and implement the face recognition application. The proposed approach follows a modular and structured pipeline, allowing each component of the system to be developed, evaluated, and improved independently. The methodology was designed to address the main challenges of face recognition, including robust face localization, discriminative feature extraction, and reliable identity classification.

This splitting strategy ensures that identity distribution is preserved across subsets while preventing data leakage between training and evaluation phases.

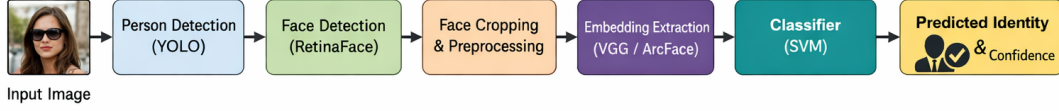
The proposed methodology strictly follows the project specifications by adopting a multi-stage processing pipeline. The system sequentially combines person detection, precise face localization, deep facial embedding extraction, and supervised identity classification, allowing each stage to be evaluated independently.

3.1 Overall Pipeline

The face recognition system is based on a sequential pipeline composed of three main stages: face detection, feature extraction, and identity prediction. Given an input image, the system first detects the presence of persons and localizes facial regions. These detected faces are then processed by a deep neural network to extract compact and discriminative feature representations, commonly referred to as facial embeddings. Finally, a supervised classification model is applied to these embeddings in order to predict the corresponding identity.

Figure 3 illustrates the overall architecture of the proposed face recognition system, highlighting the sequential processing stages from input image to identity prediction.

This modular design makes it possible to analyze the contribution of each component to the overall system performance and to identify potential bottlenecks or sources of error within the pipeline.



Overview of the proposed face recognition pipeline.

Figure 3: Overview of the proposed face recognition pipeline

3.2 Project Architecture

The project follows a modular and well-structured architecture, designed to separate data preparation, core processing, visualization, and reporting. This organization facilitates code readability, maintenance, and reproducibility.

The **pre** directory contains all scripts related to data preparation, preprocessing, and post-processing tasks, including face extraction, statistics computation, and result visualization. These scripts are used to prepare the data and analyze intermediate results without interfering with the core application logic.

The **src** directory includes the main implementation of the face recognition system. It contains the core pipeline responsible for orchestrating detection, feature extraction, and classification, as well as the graphical user interface used for interaction and visualization.

All graphical resources, such as figures, diagrams, and generated plots, are stored in the **img** directory. This includes the pipeline overview diagram and the experimental result visualizations used in the report.

Finally, the **rep** directory contains the report sources written in \LaTeX , ensuring full reproducibility of the final document and compliance with submission requirements.

This clear separation between preprocessing, core implementation, visualization, and reporting ensures reproducibility of the experiments and facilitates systematic evaluation of intermediate results.

3.3 Face Detection

Face detection is a critical component of the proposed recognition system, as imprecise localization directly impacts the quality of extracted facial features. To address this challenge, a two-stage detection strategy is adopted.

First, a person detection model based on YOLO is used to identify all human subjects present in an image [6]. This stage is designed to maximize recall by detecting a wide range of potential candidates, even in complex scenes with cluttered backgrounds.

Second, RetinaFace is applied to the detected person regions in order to precisely localize facial areas. Compared to generic detectors, RetinaFace provides more accurate and stable face bounding boxes, while also acting as a quality filter by discarding low-confidence detections and extremely small faces.

This two-step strategy allows the system to combine robustness and precision, ensuring that only reliable facial regions are forwarded to the feature extraction stage.

3.4 Feature Extraction

Once facial regions are detected and cropped, deep feature extraction is performed using pre-trained convolutional neural networks. These networks transform facial images into fixed-length embedding vectors that capture high-level and discriminative facial information.

Multiple deep models were considered in order to analyze their impact on recognition performance. Pre-trained architectures were selected due to their strong generalization capabilities and their ability to leverage large-scale training on external datasets. Using pre-trained models also reduces computational cost and training time while maintaining high recognition accuracy.

Several pre-trained face embedding models were evaluated in this project, including VGG-Face, ArcFace, and FaceNet. These models were selected in order to analyze the trade-offs between discriminative power, embedding dimensionality, and computational cost within the same recognition pipeline.

3.5 Identity Classification

The extracted facial embeddings are used as input to a supervised classification model responsible for identity prediction. This approach decouples feature learning from classification, allowing the classifier to operate on compact and meaningful representations.

A classical machine learning classifier is employed to learn decision boundaries between different identities based on the embedding space. This choice provides interpretability, efficient training, and good performance, particularly when combined with high-quality deep facial embeddings.

Operating in the embedding space allows the classifier to exploit compact and discriminative representations, resulting in efficient training and stable decision boundaries between identities.

3.6 Evaluation Strategy

To objectively assess the effectiveness of the proposed methodology, the system is evaluated on a held-out test set containing unseen images. Quantitative evaluation is performed using standard classification metrics, enabling a fair comparison between different model configurations. This evaluation strategy ensures that the reported results accurately reflect the generalization capabilities of the face recognition system.

4 Application and User Interface

To ensure practical usability of the proposed face recognition system, a graphical user interface was developed on top of the underlying recognition pipeline. The main objective of this interface is to allow users to interact with the system in an intuitive manner, without requiring prior knowledge of the implemented models or algorithms.

The application was designed with a focus on responsiveness, transparency of the processing steps, and real-time feedback. It provides clear visual indicators for each stage of the pipeline and supports both qualitative inspection and quantitative evaluation of the system.

4.1 Application Overview

The interface integrates the complete face recognition workflow, from image selection to final identity prediction. All processing stages, including person detection, face extraction, feature embedding computation, and classification, are executed automatically once an input is provided.

The modular architecture of the application allows each component of the pipeline to operate independently. This design choice improves robustness, facilitates debugging, and enables future extensions or model replacements with minimal changes to the overall system.

4.2 Processing Modes

The application supports two complementary processing modes.

The first mode focuses on single-image processing. In this mode, the user selects an image through the interface and receives immediate visual feedback, including detected regions, predicted identity, confidence level, and processing time. This mode is particularly useful for demonstrations and qualitative analysis of individual predictions.

The second mode enables batch processing of large image collections. Images are processed sequentially in an asynchronous manner, ensuring that the user interface remains responsive throughout execution. This mode allows efficient large-scale evaluation of the system and supports the generation of aggregated statistics, such as processing time, throughput, and confidence distribution across the dataset.

4.3 Pipeline Visualization

To improve interpretability and user understanding, the interface provides a visual representation of the face recognition pipeline. Each processing stage is associated with a dedicated status indicator that reflects its current state, such as waiting, running, successful completion, or error.

Detected persons and extracted faces are highlighted directly on the input images using bounding boxes, while predicted identities are displayed alongside confidence indicators. This visualization facilitates qualitative assessment of the system’s behavior and helps identify potential sources of error.

4.4 Monitoring and User Experience

In addition to prediction results, the application provides real-time monitoring information during execution. This includes progress indicators, elapsed time, and summary statistics related to the recognition outcomes. Informative messages and explicit error notifications are displayed when necessary, contributing to a smooth and transparent user experience.

Overall, the graphical interface transforms the face recognition pipeline into a complete and user-oriented application, bridging the gap between algorithmic development and practical deployment.

5 Results and Evaluation

This section presents the experimental results obtained with the proposed face recognition system and analyzes its performance under different configurations. The evaluation focuses on the accuracy of identity prediction, as well as on the practical behavior of the system when applied to real-world image collections.

5.1 Evaluation Protocol

To ensure a fair and objective evaluation, experiments were conducted on a held-out test set composed of images that were not used during training or validation. The evaluation protocol

follows a closed-set identification scenario, where each test image is assumed to belong to one of the known identities present in the dataset.

All models were evaluated under identical conditions, using the same data splits and preprocessing pipeline. This protocol allows a direct comparison between different feature extraction models and classification strategies.

The evaluation therefore reflects the behavior of the complete recognition pipeline, including detection, preprocessing, feature extraction, and classification stages.

5.2 Evaluation Metrics

System performance was primarily assessed using classification accuracy, defined as the ratio of correctly identified images over the total number of test samples. In addition to accuracy, qualitative indicators such as confidence scores and error cases were analyzed in order to better understand the behavior of the system.

For batch processing experiments, aggregated statistics were computed, including the number of correctly identified images, unknown predictions, and processing failures. These metrics provide a global view of system reliability and robustness.

Accuracy is a relevant metric in this closed-set identification scenario, where each test image is assumed to belong to one of the known identities. However, qualitative analysis is also necessary to interpret misclassifications caused by challenging conditions such as occlusions, low resolution, or visually similar identities.

5.3 Quantitative Results

The quantitative evaluation focuses on the impact of data quality and model configuration on face recognition performance. Rather than reporting isolated metrics, the analysis emphasizes comparative trends across pipeline stages and experimental settings, providing insight into the robustness and reliability of the proposed system.

5.3.1 Impact of Reference Database Quality

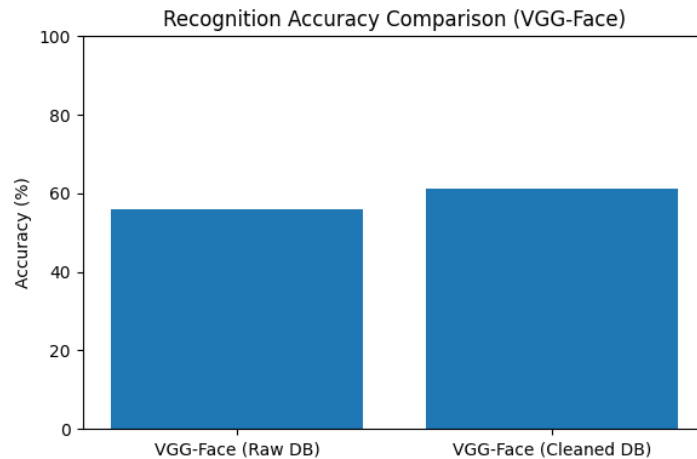


Figure 4: Recognition accuracy comparison before and after reference database cleaning using the VGG-Face model

Figure 4 shows the impact of reference database quality on recognition performance. Cleaning the celebrity reference images leads to a clear improvement in accuracy, increasing from 56% to 61.1%.

Note on evaluation scope. The reported accuracy values depend on the evaluation protocol. The 56% \rightarrow 61.1% improvement corresponds to the full closed-set evaluation using the curated reference database. In addition, on a reduced subset of 10 classes (1,341 images) used for visualization purposes, VGG-Face reaches 73.66% accuracy, which is not directly comparable to the full evaluation because the number of classes and class imbalance differ [2, 4].

This improvement confirms that embedding-based recognition systems are highly sensitive to the quality of reference images. Poorly aligned or low-quality reference faces can distort similarity measurements, even when using strong deep embedding models.

5.3.2 Detection Pipeline Analysis

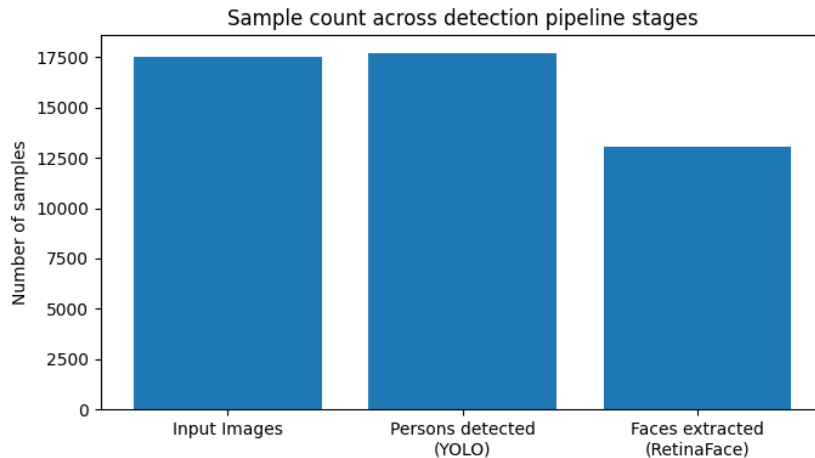


Figure 5: Number of samples retained at each stage of the detection pipeline

Figure 5 illustrates the evolution of the number of samples throughout the detection pipeline. This progressive reduction highlights the impact of successive filtering stages, which aim to balance recall preservation and face quality, a critical trade-off in practical face recognition systems.

Pipeline Stage	Method	Samples / Images	Remarks
Input dataset	Raw images	17,513	High variability and noise
Person detection	YOLO	17,707	Background individuals detected
Face detection	RetinaFace (thr = 0.5)	13,043	Quality-based filtering
Reference database	Celebrity_db	105	One high-quality image per identity
Face recognition	VGG-Face + SVM	61.1% accuracy	Cleaned reference data

Table 2: Summary of detection and recognition results across the pipeline

Table 2 provides a global overview of the face recognition pipeline, highlighting the progressive filtering of samples and the impact of data quality on recognition performance.

The summarized results emphasize the cumulative effect of design choices across the pipeline. While early detection stages aim to preserve recall, later filtering steps prioritize face quality, which directly impacts recognition accuracy. This confirms that system performance is not determined by a single component, but rather by the interaction between detection, data preparation, and feature representation.

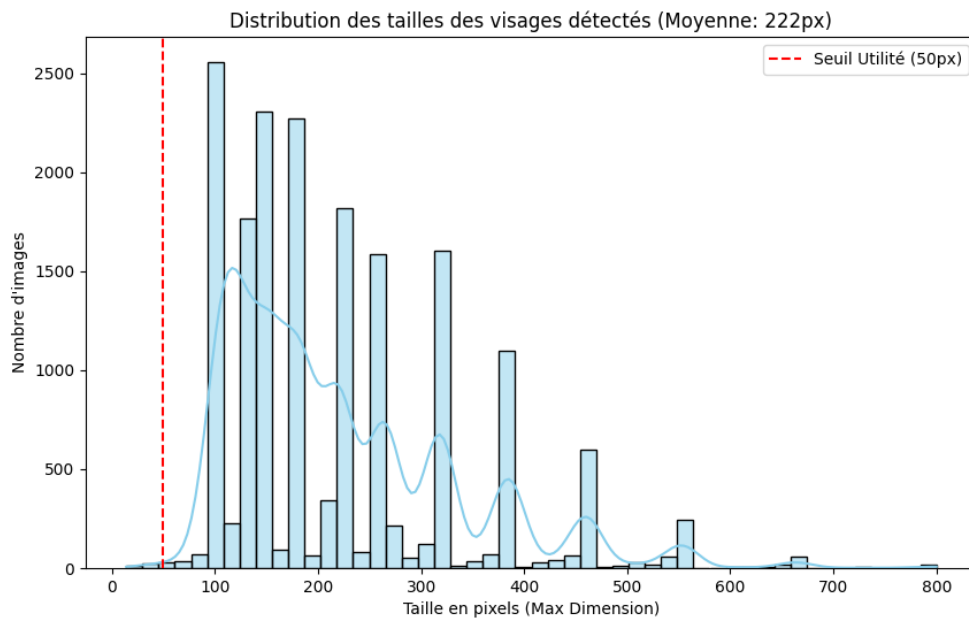


Figure 6: Distribution of detected face sizes (maximum dimension). The red line indicates the 50px usability threshold.

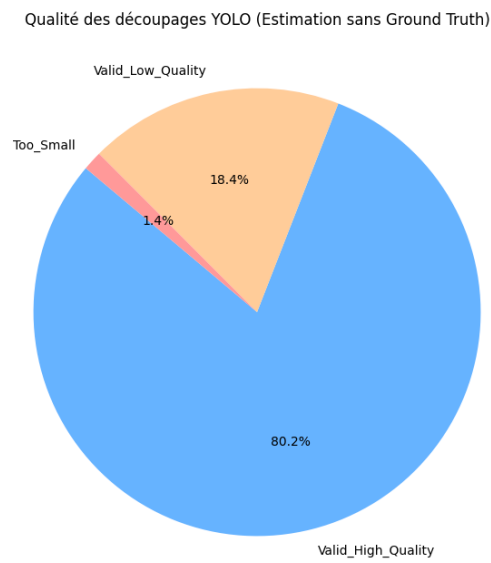


Figure 7: Qualitative distribution of detected regions after YOLO-based person detection

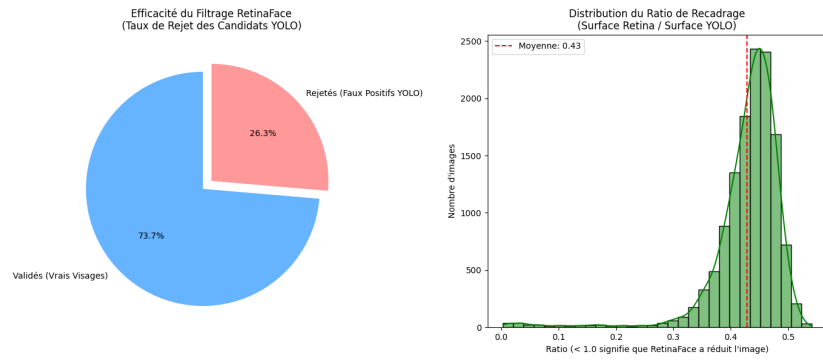


Figure 8: Average surface ratio between YOLO detections and RetinaFace face crops

While quantitative metrics provide an overall measure of system performance, qualitative analysis is essential to better understand error patterns and failure cases.

5.3.3 Comparison of Feature Extraction Models

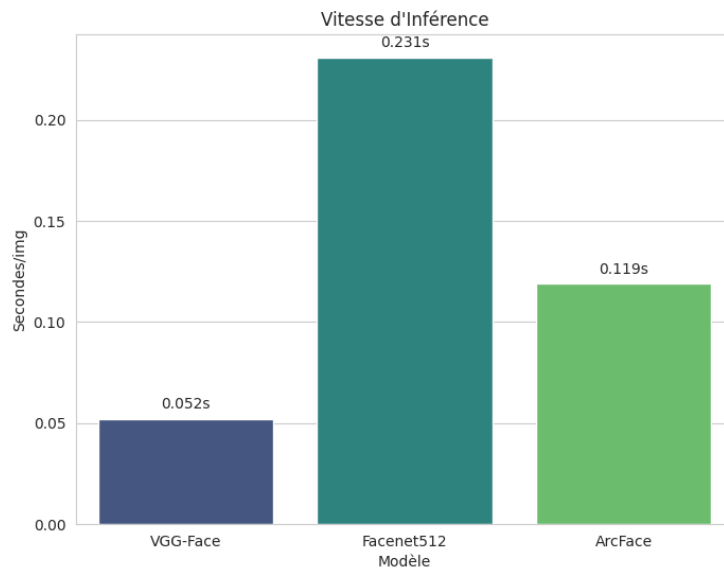


Figure 9: Inference time comparison between different face embedding models

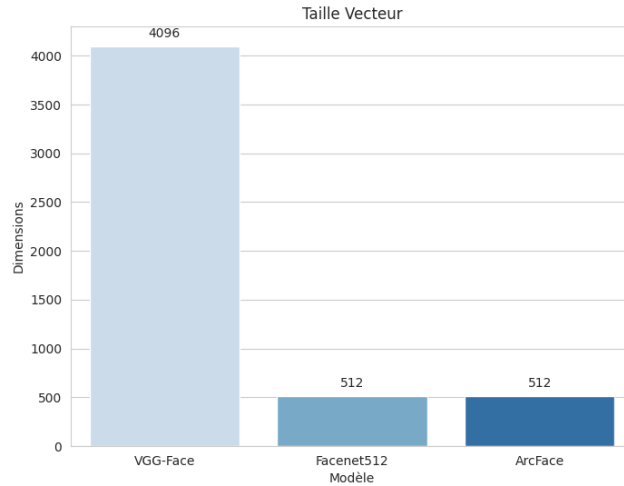


Figure 10: Embedding dimensionality comparison across evaluated models

These results illustrate the trade-off between computational efficiency and embedding compactness when selecting a face representation model.

5.4 Qualitative Analysis

Beyond numerical results, qualitative analysis was performed by inspecting prediction outputs and visual annotations produced by the application. Correctly classified examples demonstrate the system’s ability to generalize across different facial appearances, while misclassified cases often reveal limitations related to occlusions, low resolution, or ambiguous facial features.

The visualization of detected faces, bounding boxes, and confidence indicators proved particularly useful for diagnosing errors and understanding the decision process of the recognition pipeline.

Most misclassified cases are associated with challenging detections, such as extreme poses, partial occlusions, or low-resolution faces, which confirms the strong dependency between detection quality and recognition accuracy. These quantitative and qualitative observations provide a comprehensive view of the system behavior and motivate the following discussion.

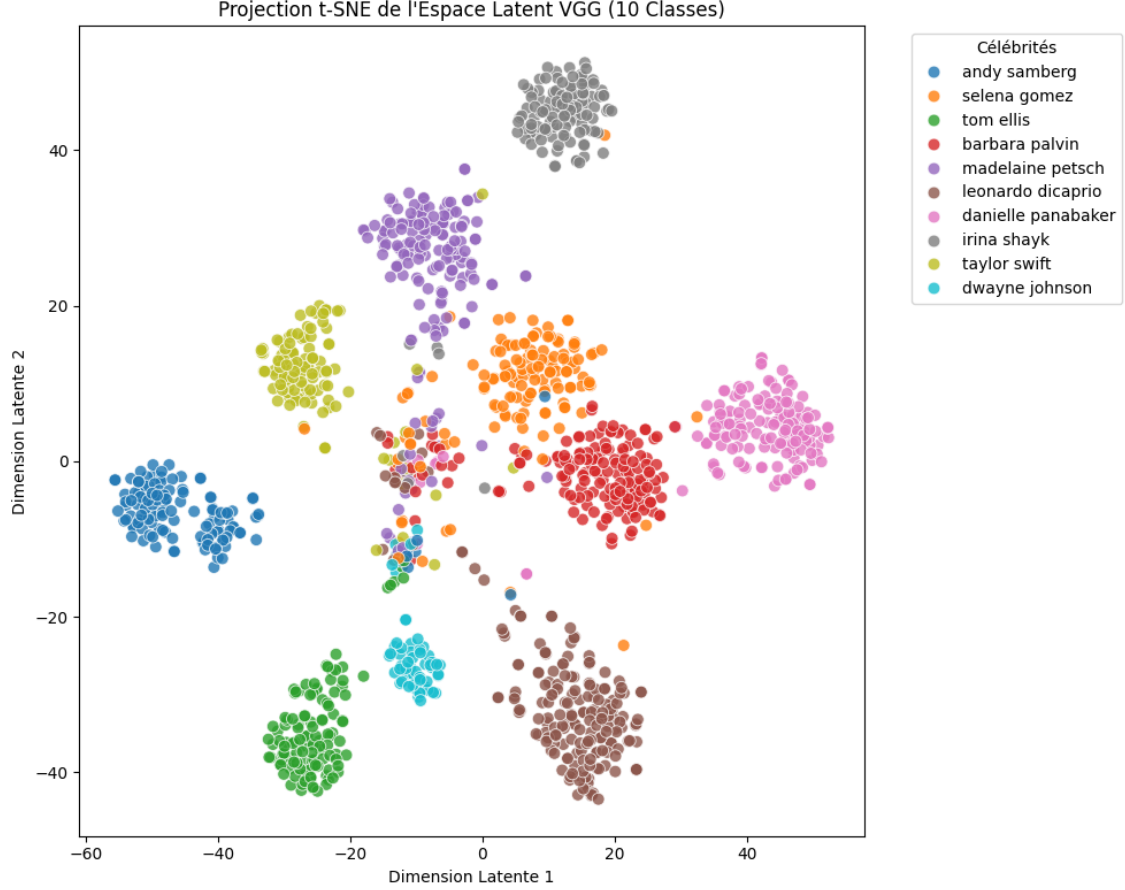


Figure 11: t-SNE projection of VGG-Face embeddings on 10 classes, illustrating cluster separability in the latent space.

5.5 Discussion of Results

Overall, the results confirm the effectiveness of the proposed methodology for celebrity face recognition. While higher-performance models yield better accuracy, they also introduce additional computational cost. This trade-off highlights the importance of balancing accuracy and efficiency depending on the intended application scenario.

The batch processing experiments further demonstrate the practical applicability of the system, enabling large-scale evaluation and consistent performance monitoring.

6 Discussion

The experimental results obtained in this project provide valuable insights into the behavior and limitations of face recognition systems based on deep facial embeddings. Beyond raw accuracy scores, the analysis highlights the strong dependency of recognition performance on data quality and pipeline design.

One of the most significant observations concerns the impact of face detection and preprocessing stages. The two-stage detection strategy combining YOLO and RetinaFace proved effective in balancing recall and precision. While YOLO favors recall by detecting a wide range of potential candidates, RetinaFace acts as a quality filter by discarding ambiguous or low-quality detections. This design choice directly influences the quality of extracted embeddings and, consequently, the final recognition accuracy.

The comparison between different embedding models further emphasizes the importance of feature representation. High-capacity models demonstrate stronger discriminative power but introduce increased computational cost. In contrast, more compact models reduce memory requirements but may impact inference time or separability in the embedding space. These observations confirm that the choice of an embedding model should be guided by the intended deployment scenario rather than accuracy alone.

Error analysis reveals that most misclassifications occur under challenging conditions, such as low-resolution images, partial occlusions, or visually similar identities. These limitations are largely inherent to the dataset and highlight the critical role of high-quality reference images in embedding-based recognition systems.

From a practical perspective, the results demonstrate that system performance emerges from the interaction between multiple pipeline components rather than from a single model. Careful preprocessing, reliable detection, and appropriate model selection are therefore essential for deploying face recognition systems in unconstrained environments.

These findings highlight the importance of adopting a system-level perspective when designing face recognition pipelines, where detection, representation, and classification choices jointly determine overall performance.

6.1 Challenges and Limitations

Despite the effectiveness of the proposed approach, several limitations must be considered when interpreting the reported results.

A first limitation is related to data quality. The original dataset contains a significant amount of noise, including blurred images, occlusions, and background individuals. Although the use of strict face detection thresholds improves the quality of extracted faces, it also leads to a reduction in the number of usable samples, which may impact recall.

Another limitation concerns the dependency of recognition performance on the reference database. Experimental results clearly show that the accuracy of embedding-based recognition systems is highly sensitive to the quality of reference images. Poorly aligned or low-quality reference faces can significantly degrade performance, even when using strong deep models.

In addition, the current system operates in a closed-set recognition scenario, where all test identities are assumed to be known in advance. This assumption limits applicability in real-world settings, where unknown identities may frequently appear.

Finally, computational cost represents a practical constraint. Deep feature extraction, particularly with high-capacity models, introduces non-negligible processing time, which may limit scalability or real-time deployment without further optimization.

7 Conclusion and Perspectives

In this project, a complete face recognition application was designed and implemented within the framework of the *AI Models & Applications* course. The proposed system integrates all key components of a modern face recognition pipeline, including face detection, deep feature extraction, supervised classification, and a user-friendly graphical interface.

The experimental evaluation demonstrates that the chosen methodology is effective for celebrity identification and highlights the advantages of deep embedding-based approaches. The application successfully supports both single-image analysis and large-scale batch processing, enabling comprehensive evaluation and practical usability.

Several perspectives for future work can be considered. These include extending the system to open-set recognition scenarios, fine-tuning deep models on domain-specific data, and exploring alternative classification strategies. Additionally, deploying the application as a web-based or real-time system could further enhance its applicability.

Overall, this project demonstrates that effective face recognition in unconstrained environments relies as much on data quality and pipeline design as on the choice of deep models. The results underline the importance of careful preprocessing, robust detection, and rigorous evaluation when deploying face recognition systems in unconstrained real-world scenarios.

8 Team Members and Contributions

This project was carried out collaboratively by the members of Team 3. Each member was responsible for a specific part of the system in order to ensure an efficient and structured development process.

- **Diallo Mamadou Ougailou** was responsible for the design and implementation of the detection pipeline, including person detection and face localization. This contribution also involved qualitative evaluation of detection results and the analysis of associated limitations.
- **Camara Ibrahima** focused on the face recognition component of the system. This included feature extraction using deep learning models, training and evaluation of the classification module, and implementation of the inference pipeline for identity prediction.
- **Taha Rania** was in charge of system evaluation and result analysis. This role included the computation of evaluation metrics, visualization of results, and the rédaction of the final report, as well as verification of compliance with the project requirements.

9 References

References

- [1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
- [2] O. M. Parkhi, A. Vedaldi, and A. Zisserman, *Deep Face Recognition*, British Machine Vision Conference (BMVC), 2015.
- [3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, *ArcFace: Additive Angular Margin Loss for Deep Face Recognition*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [4] F. Schroff, D. Kalenichenko, and J. Philbin, *FaceNet: A Unified Embedding for Face Recognition and Clustering*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [5] J. Deng et al., *RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild*, arXiv preprint arXiv:1905.00641, 2019.
- [6] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*, arXiv preprint arXiv:1804.02767, 2018.
- [7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [8] Kaggle, *Pins Face Recognition Dataset*, <https://www.kaggle.com/datasets/havingfun/pins-face-recognition>, accessed 2025.