One-Shot Learning for Face Recognition

Project TER - Team 6

Lounas, Sosthène, Thanina, Akil

Supervised by Jerôme Pasquet

December 19, 2025

1. Introduction
2. State of the Art
3. Data Exploration
4. Methodology
5. Results Analysis
6. Conclusion

Section 1

Introduction

Context & Problem Statement

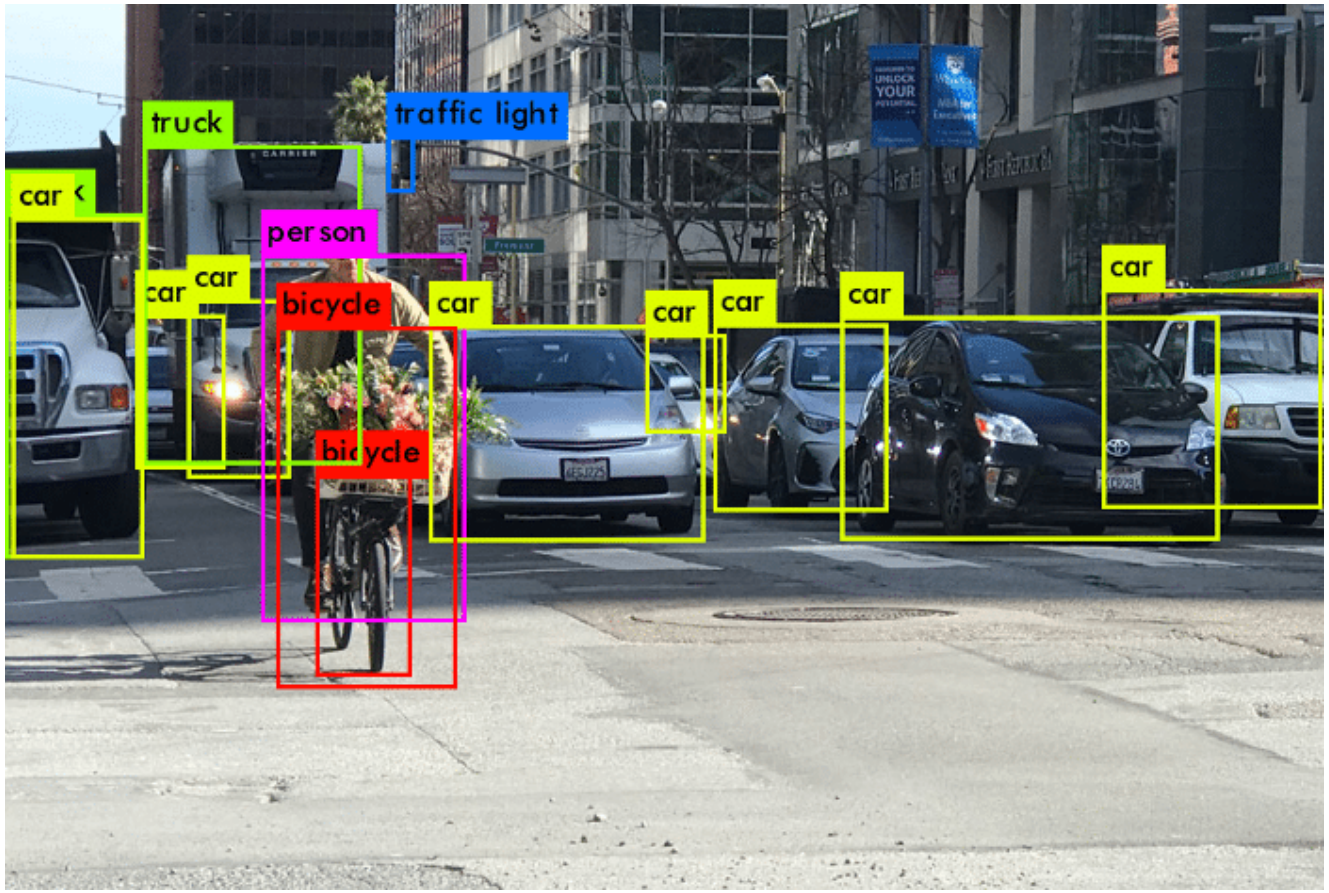
- **The Context** Face recognition is widely used but traditionally requires massive datasets per person.
- **Challenge** In real-world scenarios (police, security), we often have only **one photo** of a suspect/user.
- **The Problematic** How to build a robust identification system capable of recognizing a person from a single reference image (One-Shot)?

Section 2

State of the Art

1. YOLO

YOLO (You Only Look Once): a single-stage object detection



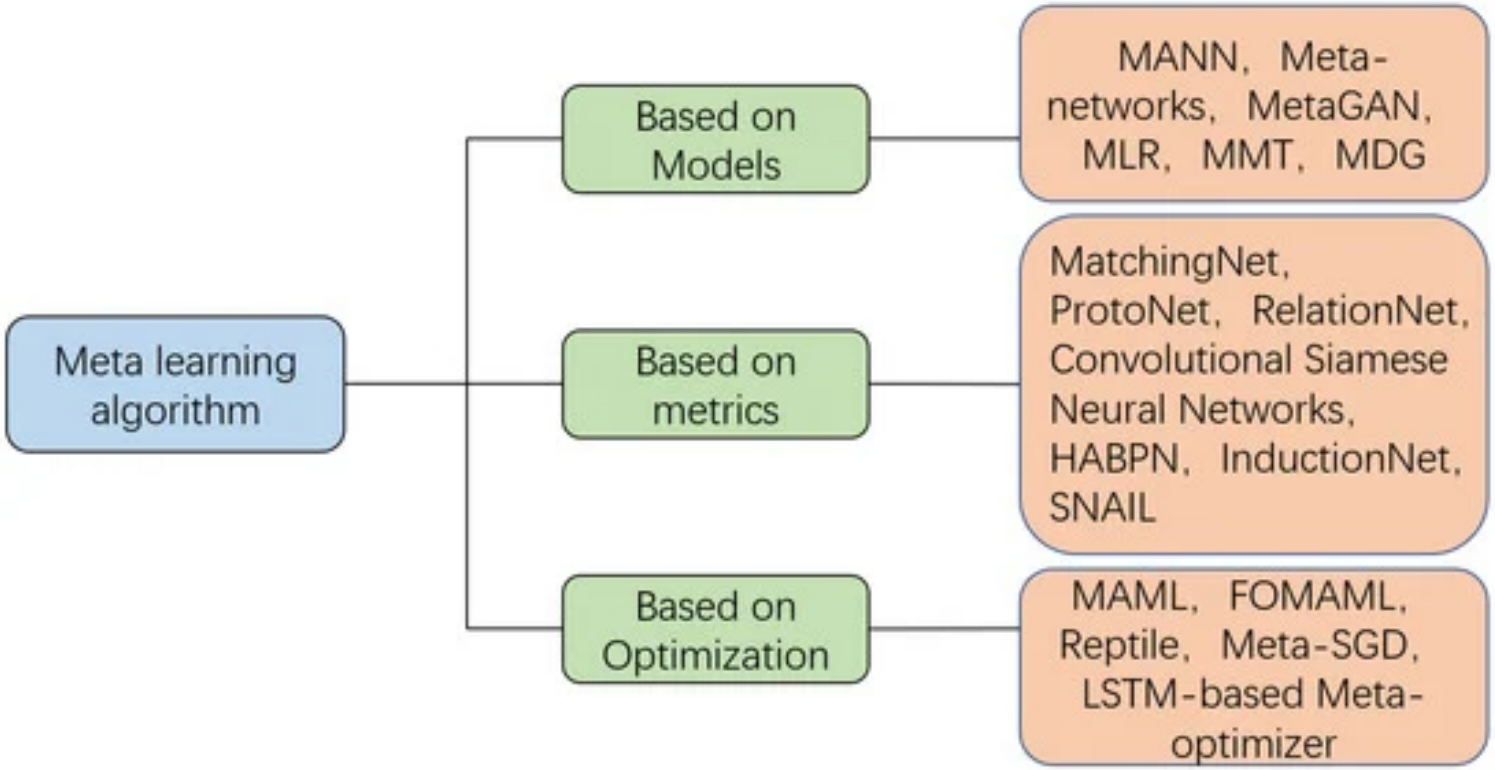
2. SIFT

- "Facial recognition system based on SIFT"
- SIFT algorithm proposed by David G. Lowe in 1999
- Detects and describes keypoints in images
- Robust to scale and rotation changes

3. Meta-Learning Paradigm

- **"Learning to Learn"**
- **Generalization** over Memorization
- **Multitask Strategy** (Episodes)

Different algorithms



ProtoNet

Embedding Space

Metric Learning (Projecting images to vectors)

↓

Prototypes

Class Centroids (Calculating the mean)

↓

Euclidean Distance

Nearest Neighbor Decision

Section 3

Data Exploration

1. YOLO's Dataset

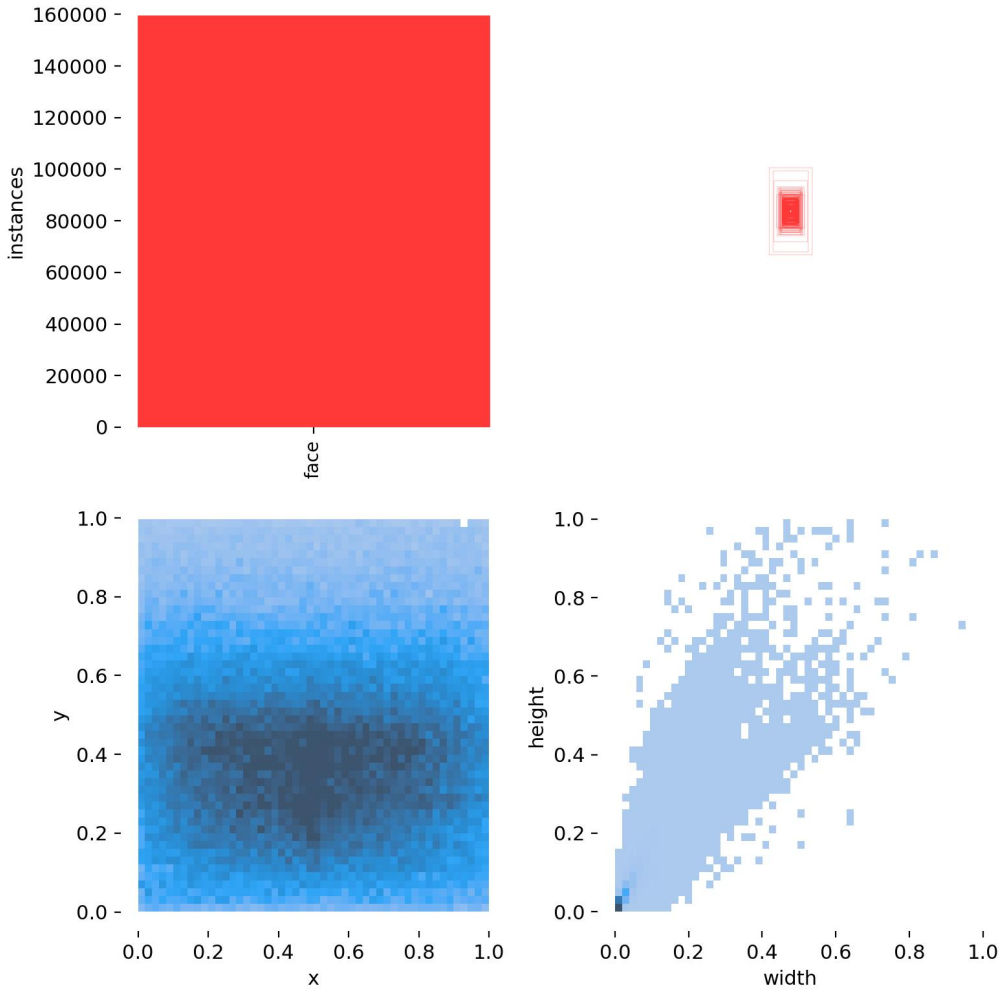
WIDERFACE

- Provides diverse, **real-world** face images for robust fine-tuning
- **Goal**: Increase the performance.

±13000 **Photos**	+3200 **Photos**	+3200 **Photos**
(Train)	(Validation)	(Test)

1.YOLO's Dataset

- **Number of face**: +160,000



2. ProtoNet's Dataset: CelebA

Data Preprocessing & Selection

- **Initial Dataset**: 10,177 unique identities.
- **Filtering Criterion**: Removal of identities with insufficient images to ensure robust support and query sets.
- **Final Selection**: 7,417 identities retained for the project.

5,949 Identities (Training set)

1,468 Identities (Testing set)

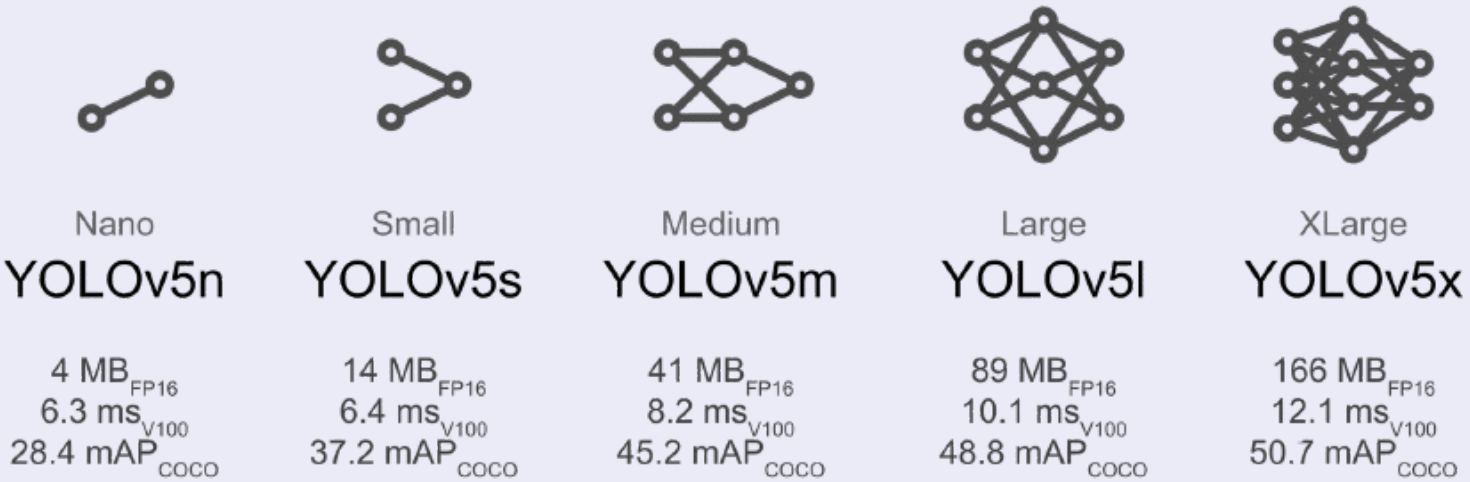
3. Miashs Classmate's Dataset

Section 4

Methodology

1. Face Detection (YOLOv5)

The Type and Version of YOLO

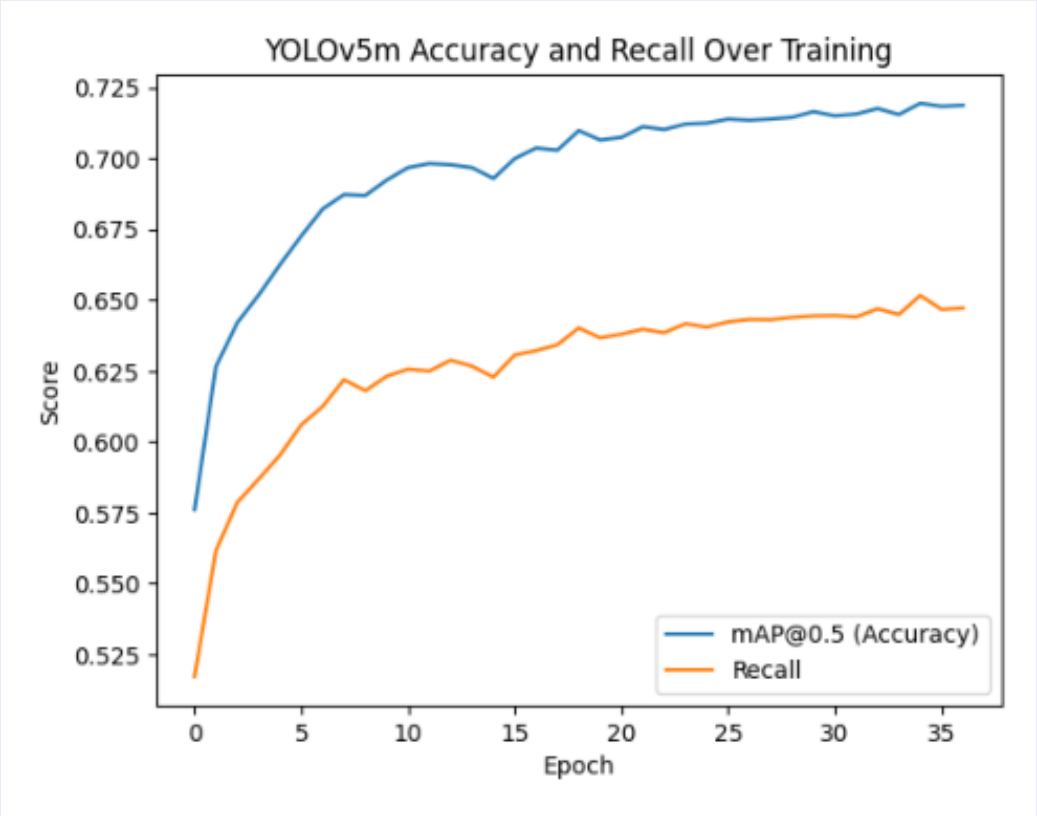


- Selected: YOLOv5m

1. Face Detection (YOLOv5)

Fine-tuning

- Increase performance for face detection



2. Classic Approach (SIFT)

a). Face Detection

Detect faces in images using YOLOv5m fine-tuned Crop and extract each detected face

2. Classic Approach (SIFT)

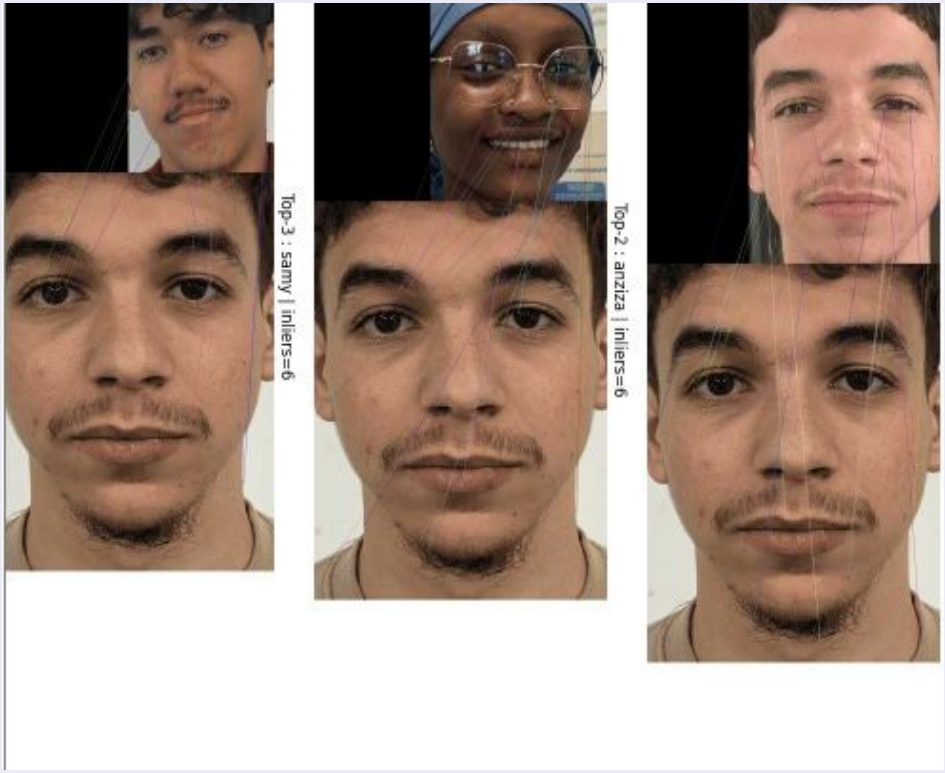
b). Preprocessing and SIFT Feature Extraction

Prepare face images and convert them to grayscale Detect keypoints on the face using SIFT Extract 128 value descriptors to represent facial features

2. Classic Approach (SIFT)

c). Matching

- **Descriptor Matching**: Find pairs of similar keypoints between reference and test images based on their feature vectors.
- **Ratio Test Filter**: Select only reliable matches .
- **RANSAC Algorithm**: Filter wrong matches with RANSAC.



Visualization of feature point connections between two faces.

3. ProtoNet Implementation

a). Backbone Architectures (The Encoders)

Basic CNN

- **Structure**: 4 Conv Blocks.
- **Pros**: Extremely lightweight.
- **Focus**: Fast Inference.

ResNet18

- **Structure**: Residual Blocks (Skip connections).
- **Pros**: Deeper feature extraction.
- **Focus**: Higher Accuracy.

We compared both to find the best speed/accuracy trade-off.

3. ProtoNet Implementation

b). Training Strategy & Experimental Setup

1. Data Split Protocol

- **Disjoint Classes**: The identities used for training are **never** seen during testing.
- **Goal**: Ensure true generalization (Zero-Overlap).

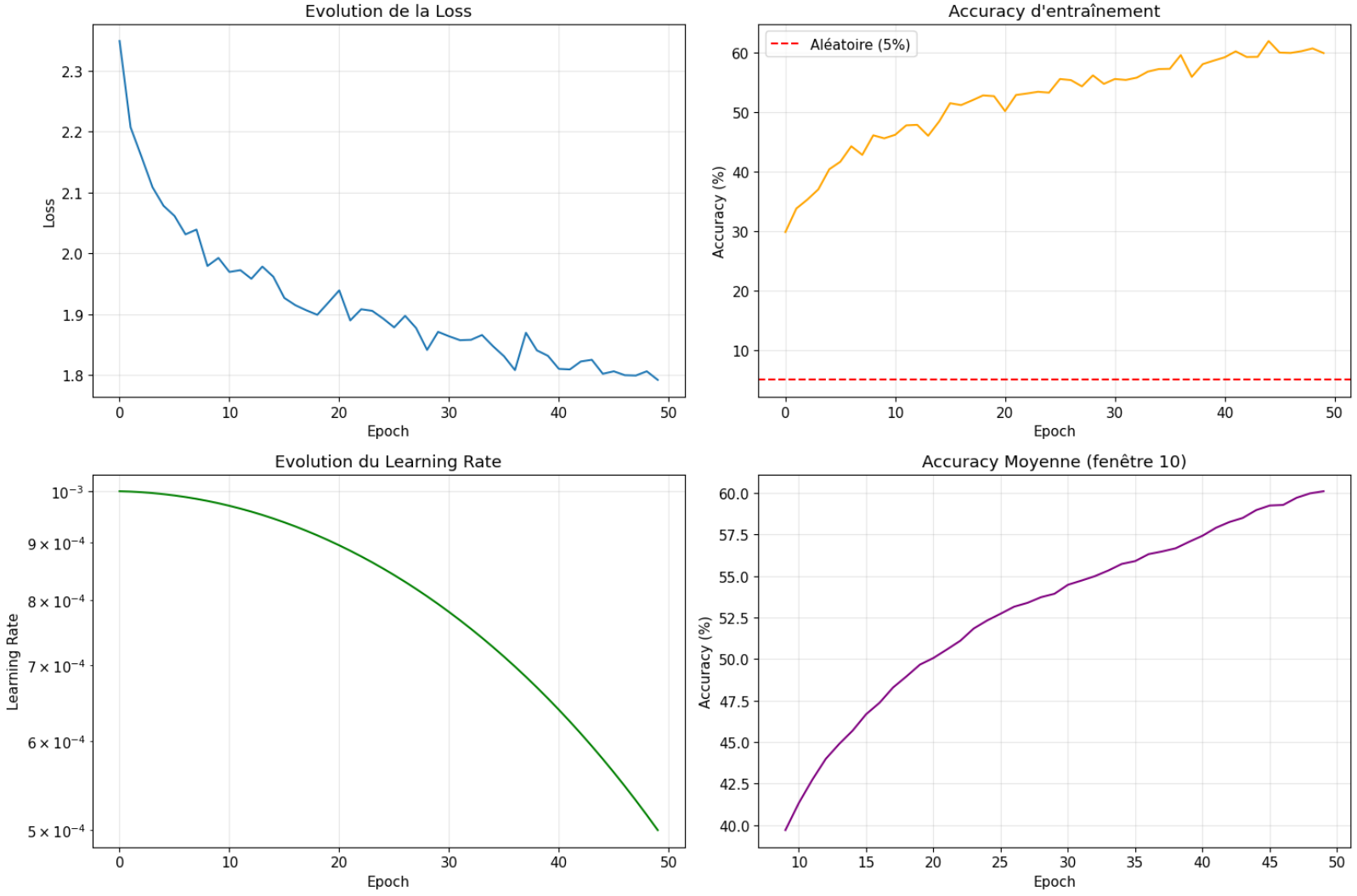
2. Episode Configuration

Parameters for each training iteration:

20	1	10
Way	**Shot**	**Query**
(Classes/Episode)	(Ref Image)	(Test Images)

3. ProtoNet Implementation

Meta Train



3. ProtoNet Implementation

Meta Test

	N-WAY	K-SHOT	Q-QUERY	ACCURACY (%)
5-WAY 1-SHOT	5	1	10	90 %
5-WAY 5-SHOT	5	5	10	86 %
10-WAY 1-SHOT	10	1	10	63 %
10-WAY 5-SHOT	10	5	10	87 %

Section 5

Results Analysis

1. Baseline Results: SIFT Approach

Accuracy: 0.33

(33% Success Rate)

Interpretation:

- **Limited Recognition**: The system recognizes only a few faces correctly.
- **Cause 1: Dataset Size**: The available data is too small to build robust feature references using classic methods.
- **Cause 2: Low Visual Detail**: The images have few visual details.

2. Embedding Analysis:

PCA

Quantitative Insights

- Dimension Reduction: 128D $\rightarrow$ 2D
- Variance Explained: 59.2%

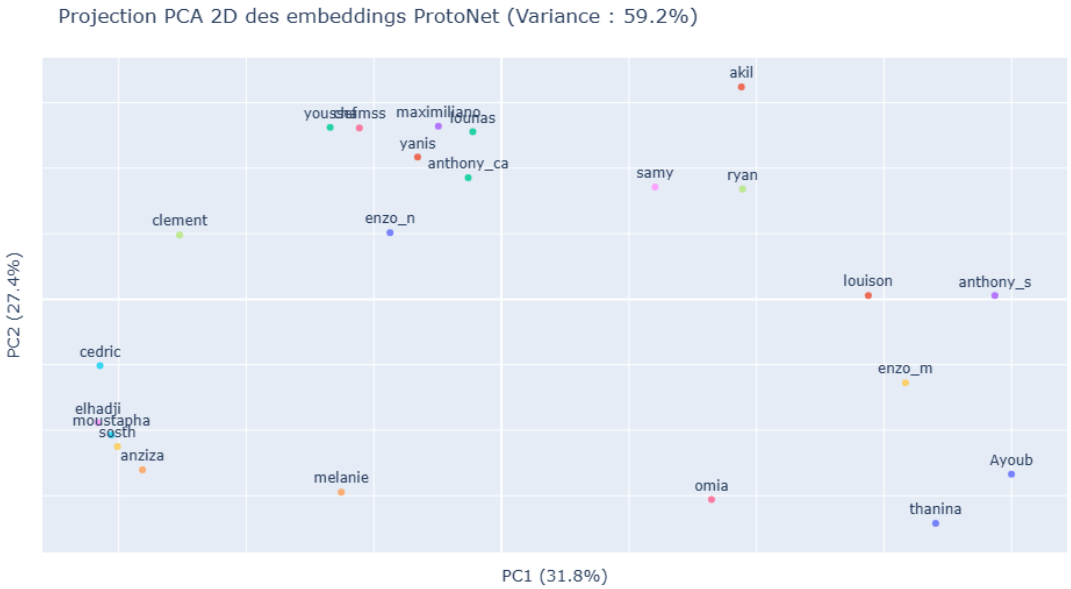


Figure 1: 2D Projection of Embeddings

2. Embedding Analysis:

t-SNE

Analysis

- Non-linear Mapping
- Compact Clusters

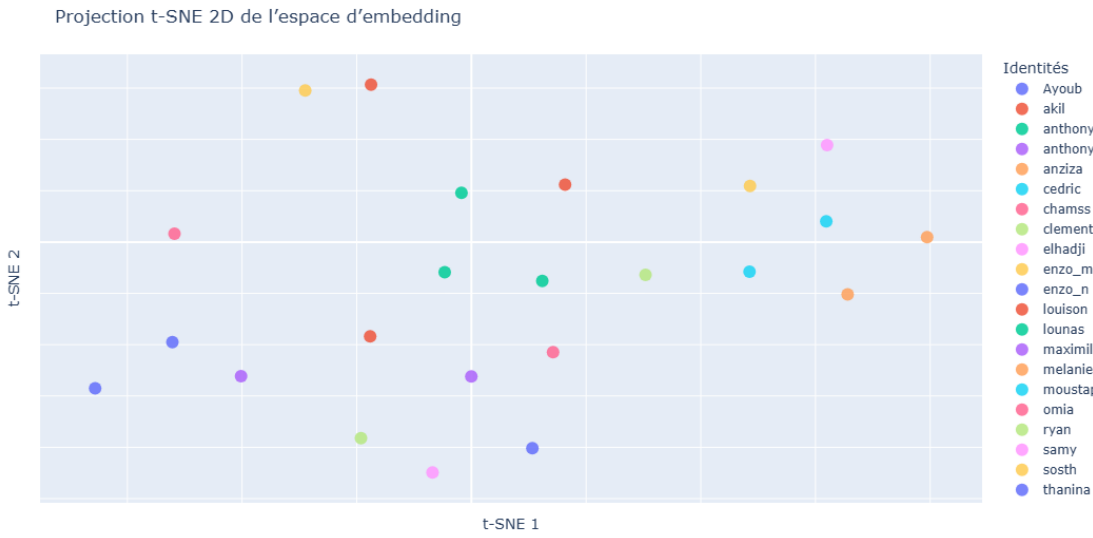


Figure 2: 2D Projection of Embeddings

3. Real-world Evaluation:

Figure 3: Initial detection: 23 faces identified by the model.

3. Real-world Evaluation: Accuracy Results

Performance on 18 Validated Identities

Metric	Calculation	Result
Top-1 Accuracy	3 / 18	16.67 %
Top-5 Accuracy	**14 / 18**	**77.78 %**

- **Interpretation**: The Top-5 score (77.78%) validates the quality of the ProtoNet clusters.
- **Finding**: In most cases, the correct identity is found among the 5 closest prototypes.

3. Real-world Evaluation: ROC Curve

Biometric Metrics

- **AUC**: 0.8547
- **EER**: 25.00 %
- **Optimal Threshold (τ)**: 0.9511

Quick Analysis

- **AUC 0.85**: High ability to distinguish between identities.
- **EER 25%**: Acceptable for non-critical use, but shows real-world limits (lighting, pose).

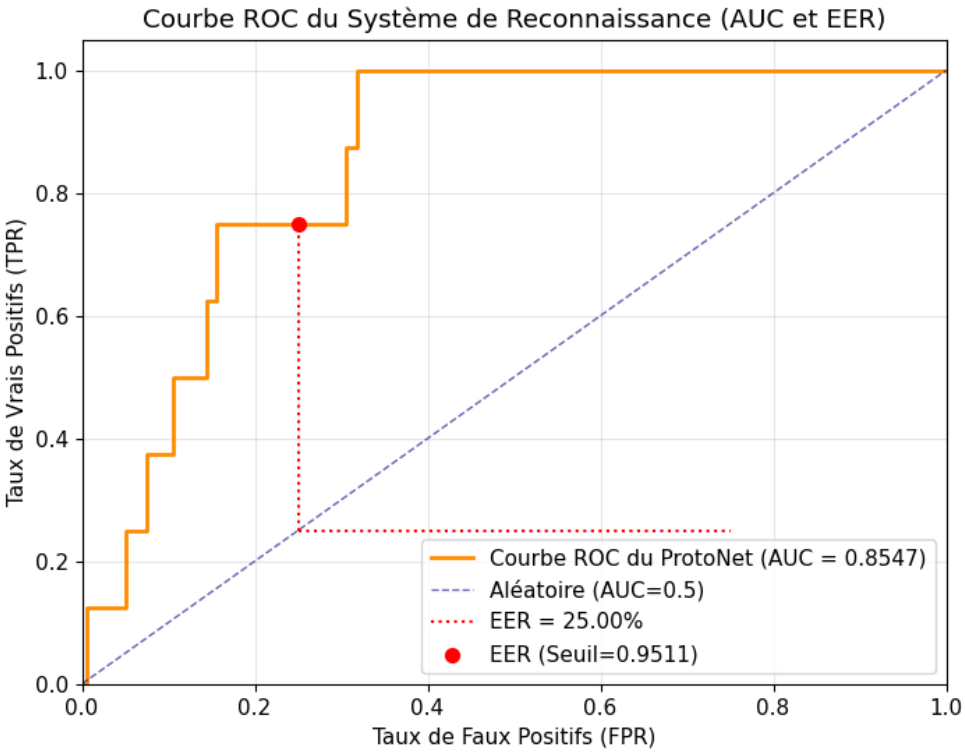


Figure 4: ROC Curve (168 distance pairs)

Section 6

Conclusion

Conclusion

- The goal was to evaluate if a ProtoNet model combined with YOLO could be successufuly
- the results are encouraging
- However, achieving accurate Top-1 identification remains challenging