VIPM Project: Are you good or bad?

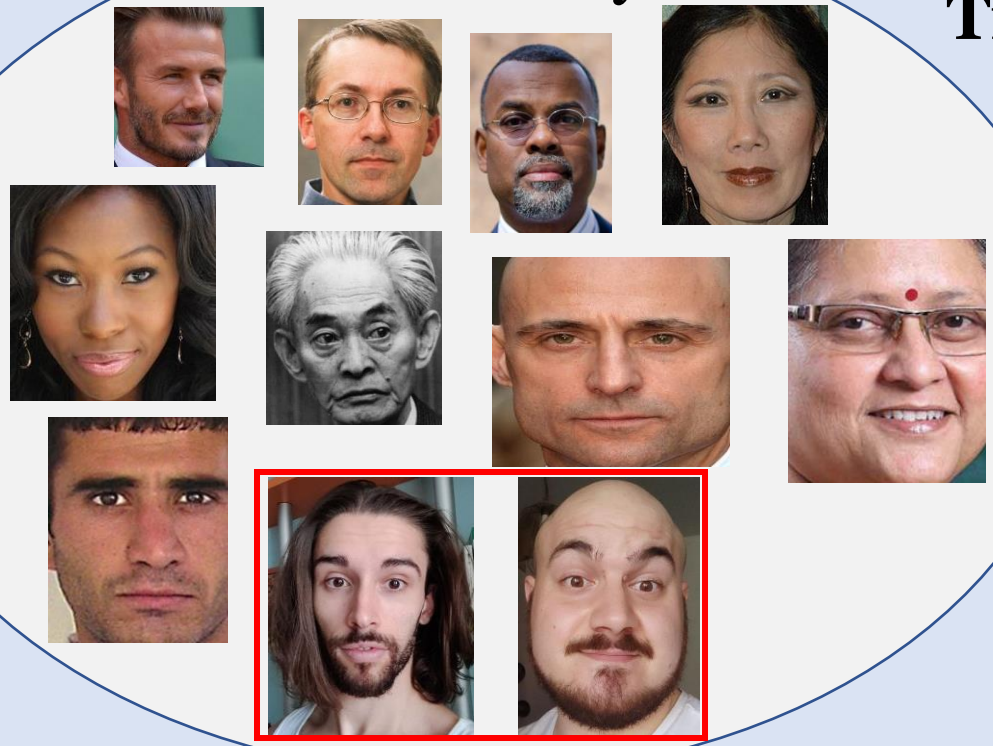
Alessandro Gherardi – mat.817084
Simone Giuseppe Locatelli – mat.816781

Introduzione

- Classificazione: buono/cattivo
- Content Based Image Retrieval
- Iterative Search Refinement

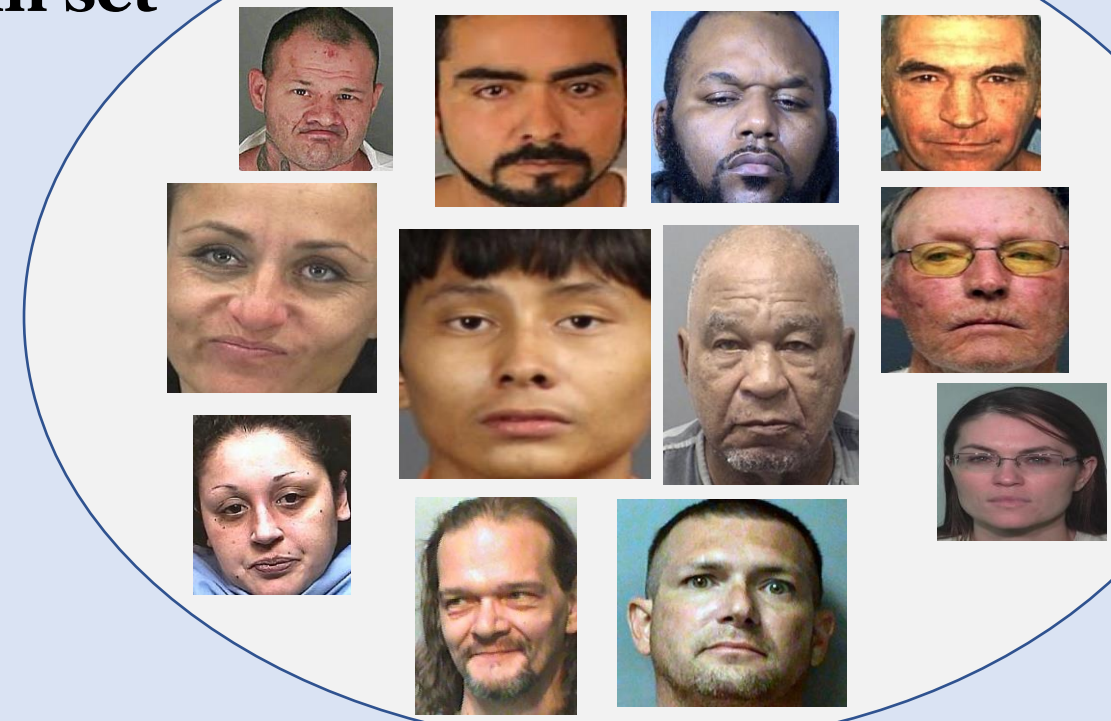
Dataset

5670 Savory



Train set

5610 Unsavory



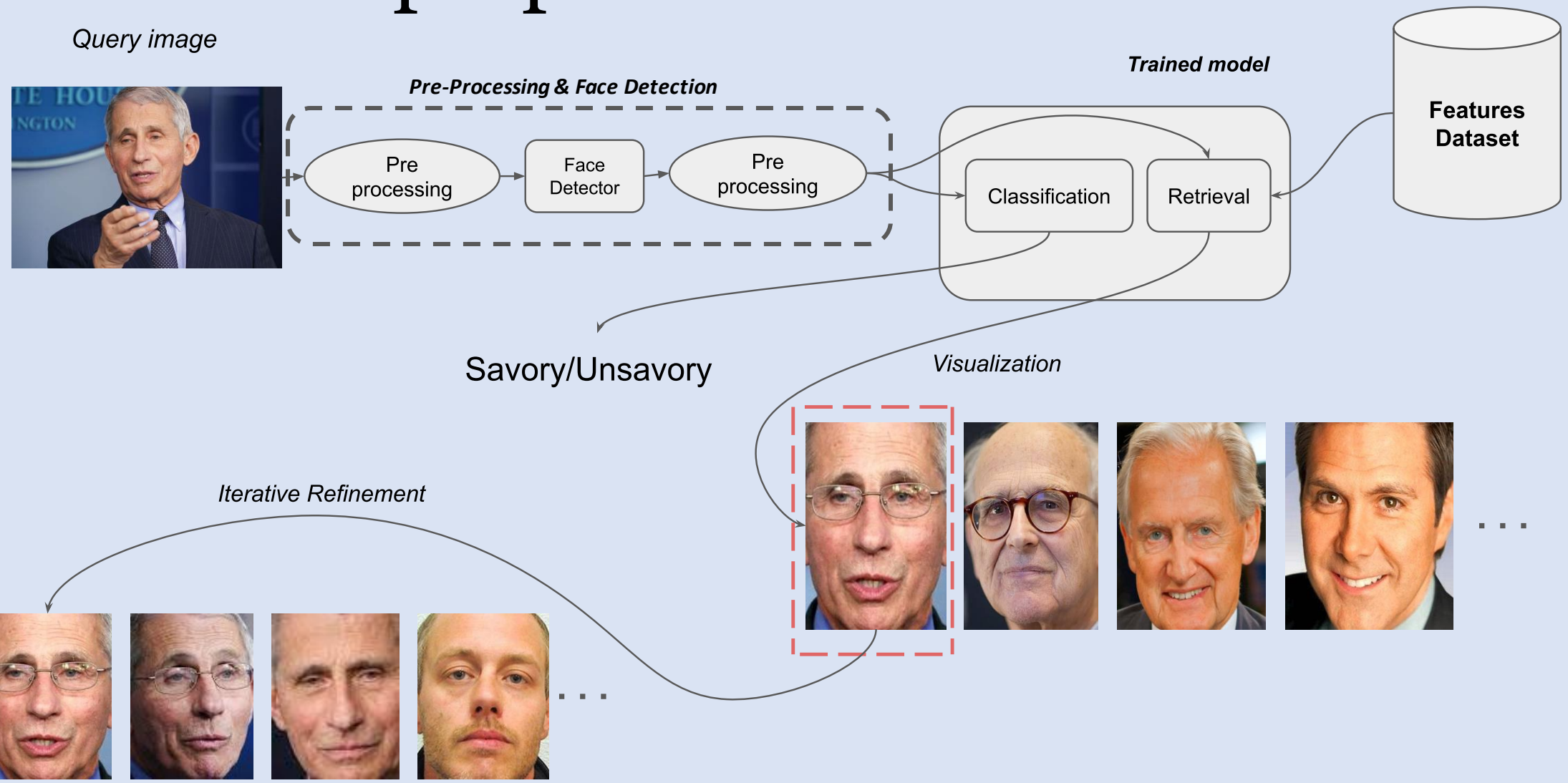
Validation set:

- 300 Savory
- 300 Unsavory

Test set:

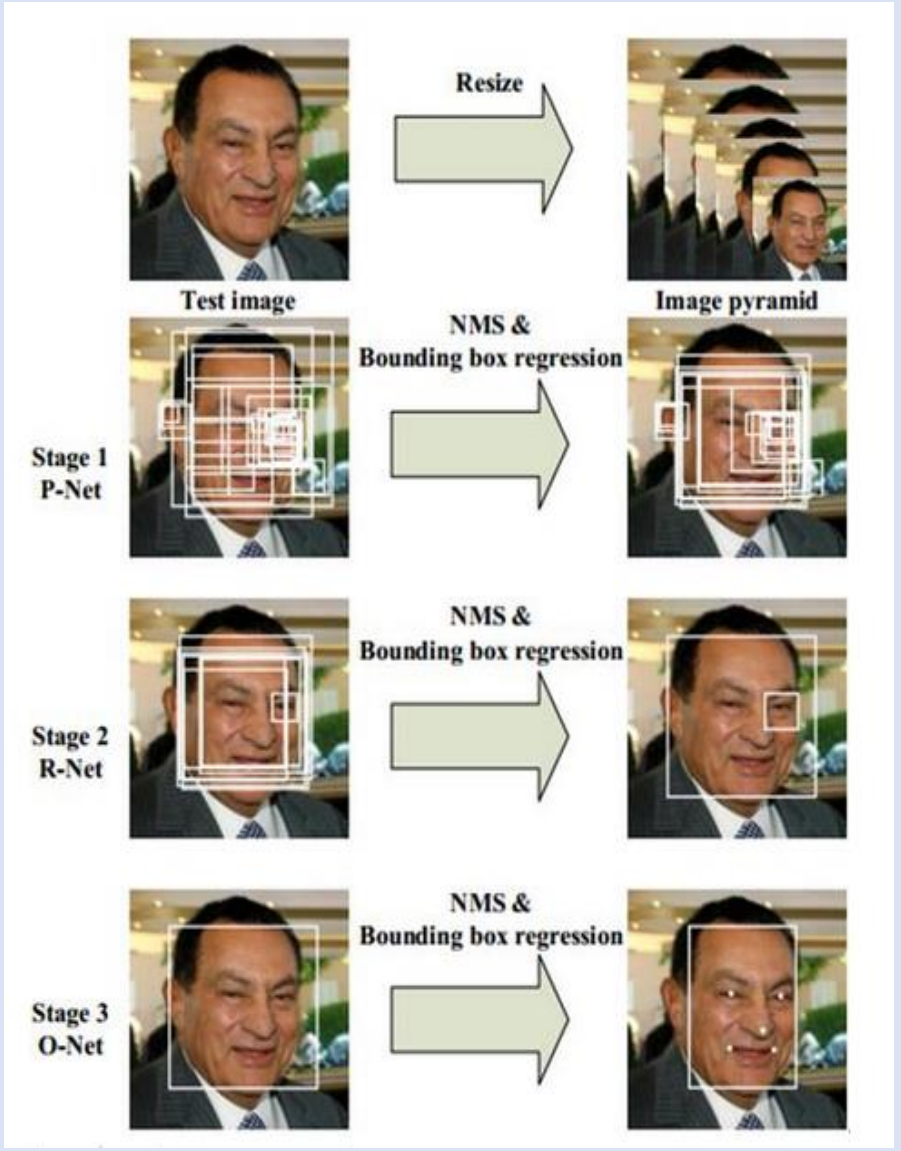
- 300 Savory
- 300 Unsavory

Soluzione proposta

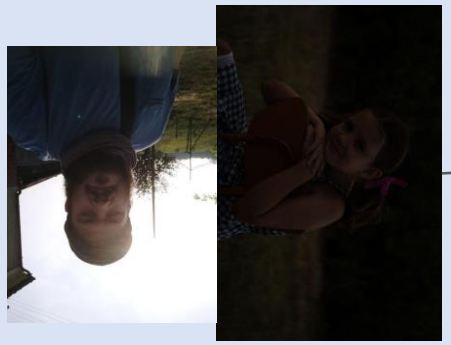


Face Detection

- MTCNN (Multi-Task Cascaded Convolutional Neural Networks)
- Costruzione dell'immagine piramidale come input dell'algoritmo
- Algoritmo suddiviso in 3 fasi:
  - P-Net -> FCN per estrarre bounding-box candidate
  - R-Net -> CNN per raffinare risultati ottenuti
  - O-Net -> CNN per produrre bounding-box finale e posizione punti chiave (occhi, naso, ...)



Credit: Zhang et al. ***Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks***



Face Detect

No Face

Preprocess all image

Gamma → Sharp → Face Detect → Rotation and detection

No Face → OUT

Face

Face

Crop Based on MTCNN bbox

Face Detect

No Face

Rotation and detection

No Face

Face

Face

Crop with bord

Preprocessing and Face Detect

Value:
(# pixels <128) / (# pixels > 128)

Soglie di stop:
Value < 5
Value > 0.2

Value → Gamma → Value → Sharp

Value OK

Value not OK

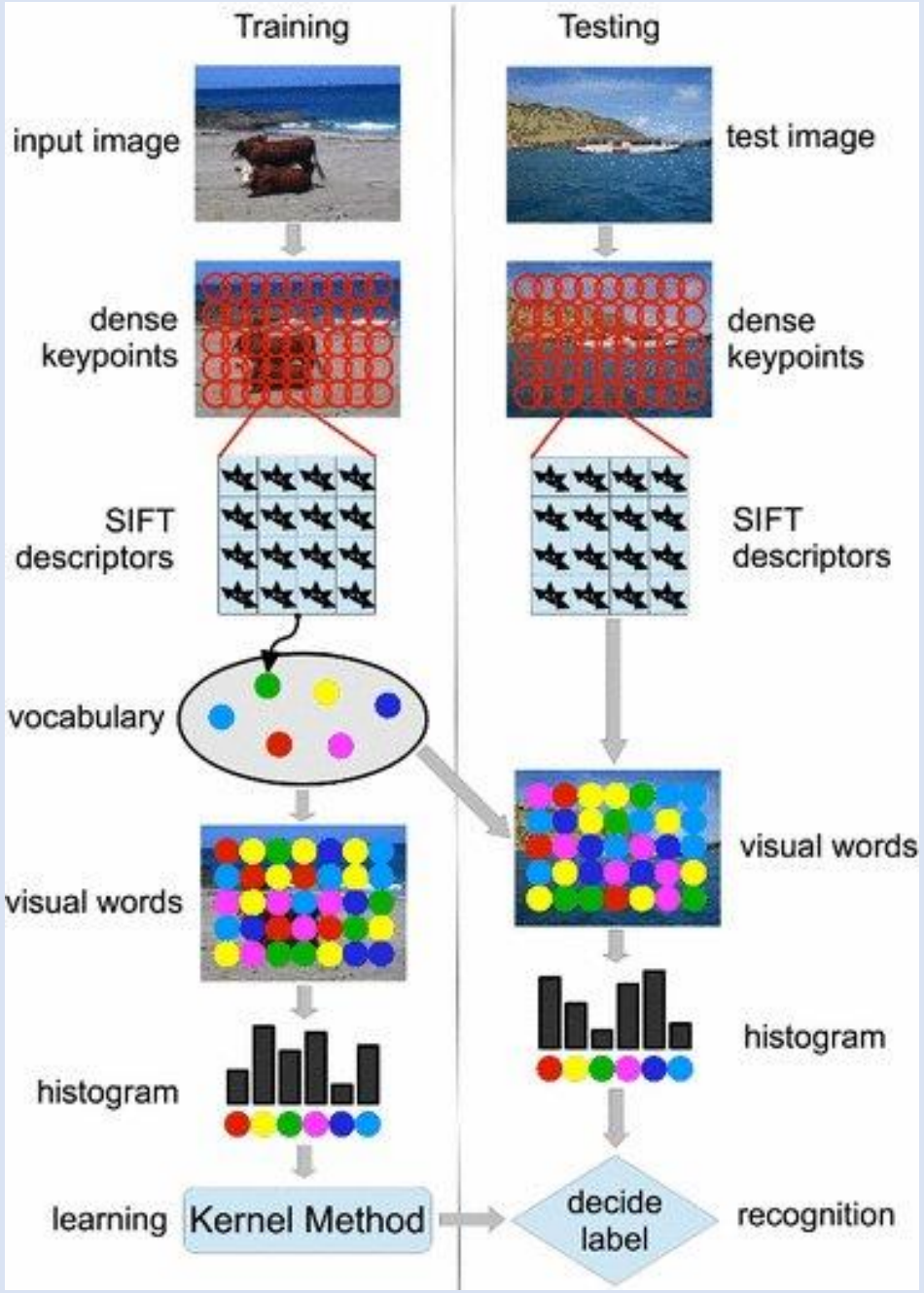
Models

- Addestramento di modelli sul dataset di training
- Modelli basati su features hand-crafted:
  - SVM -> BOVW sfruttando features SIFT
  - Random Forest -> Istogramma Colore RGB
  - Random Forest -> Combinazione SIFT + Colore

- Modelli Neurali
  - CNN -> Features apprese autonomamente, robuste al rumore e gerarchiche

Models – BOVW

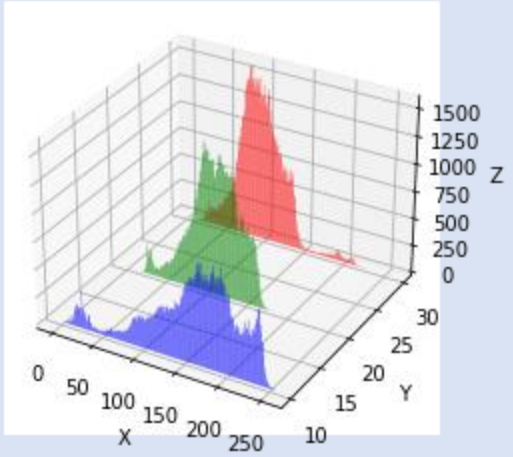
- Costruzione modello BOVW utilizzando descrittori SIFT.
- Estrazione descrittori SIFT (128-dim) da una griglia regolare.
- Creazione del vocabolario tramite K-Means -> 150 parole visuali (i centroidi individuati).
- Creazione istogramma normalizzato delle parole visuali.

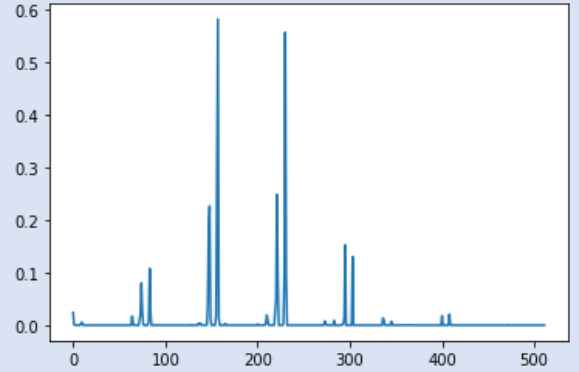


Credit: Patch Autocorrelation Features: a translation and rotation invariant approach for image classification

Feature – Color Histogram

- Istogramma RGB normalizzato con 8 bin per canale.
- Descrittore di dimensione 512 ad immagine.

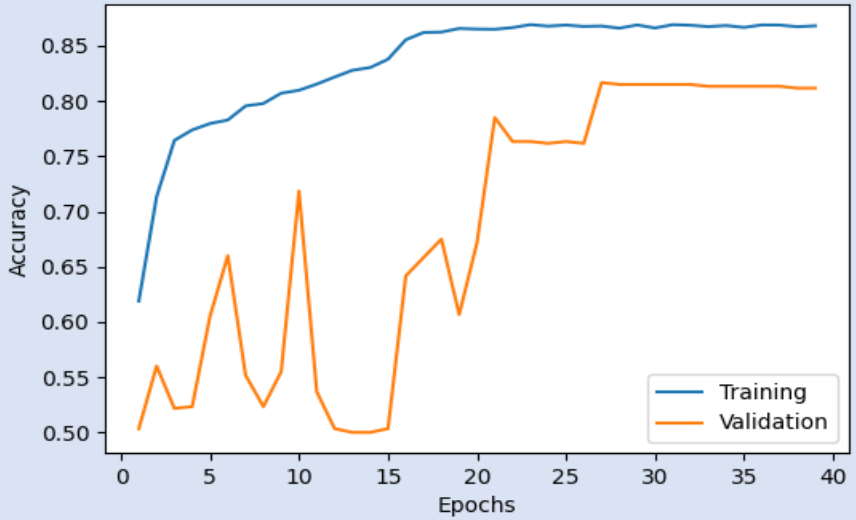


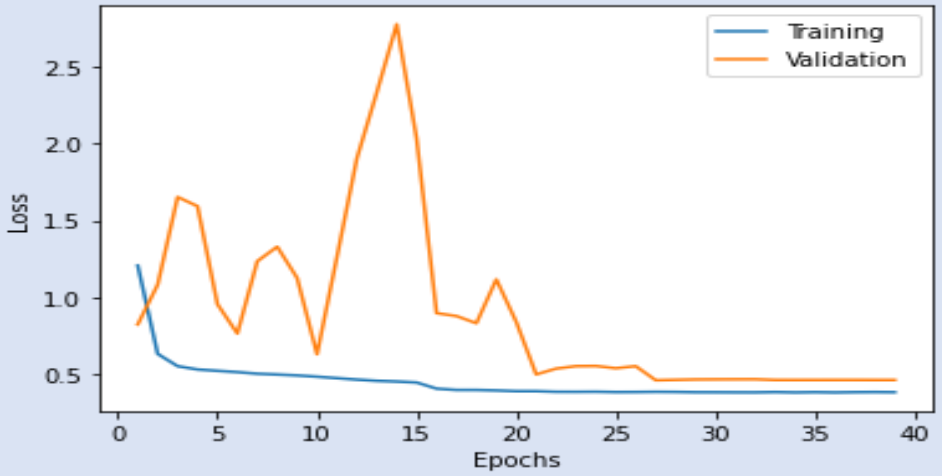


Models - CNN

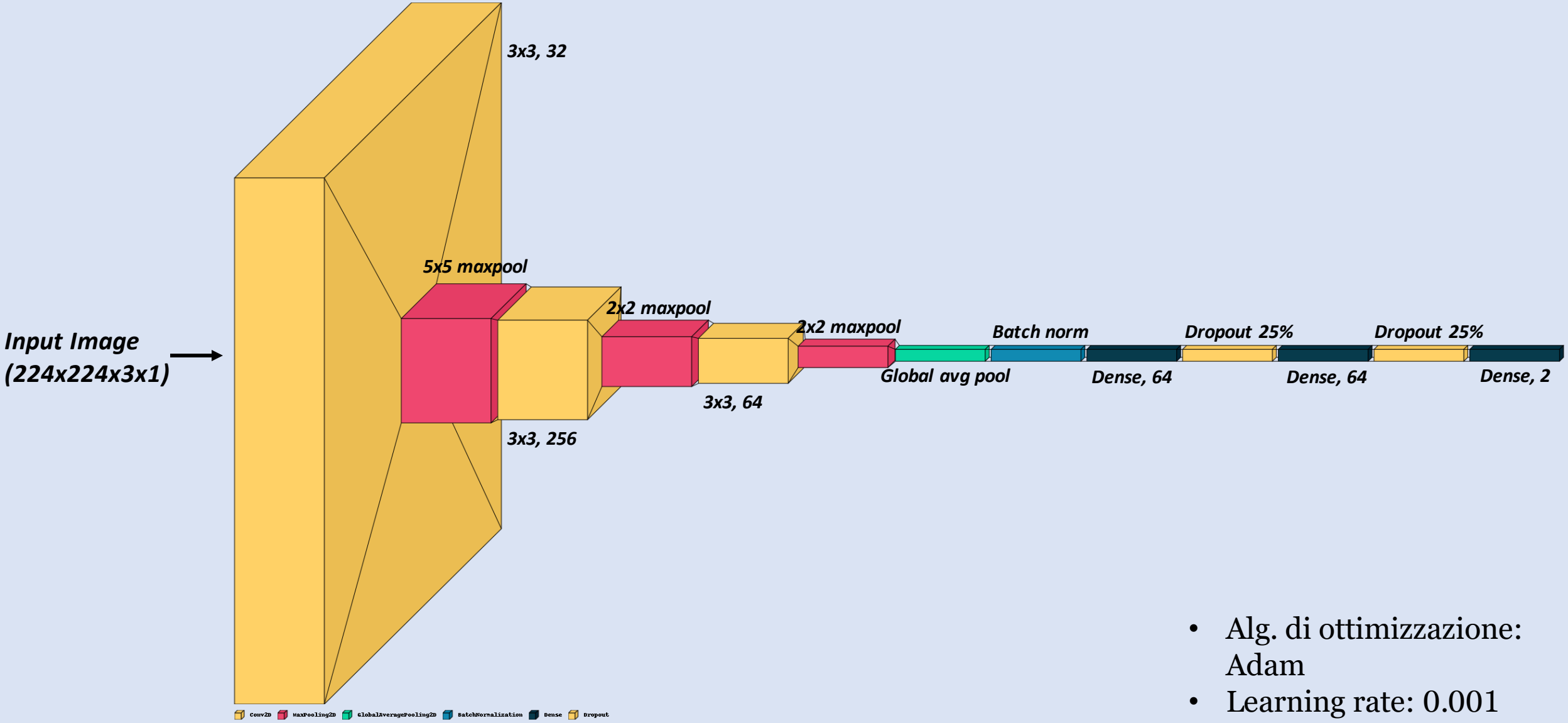
- Struttura CNN e Iperparametri individuati tramite ottimizzazione Bayesiana.
- Ottimizzazione effettuata su un sottoinsieme (25%) di train e validation set.

Phase	Accuracy	Loss
Train	86.8%	0.38
Validation	81.1%	0.46
Test	88.6%	0.34





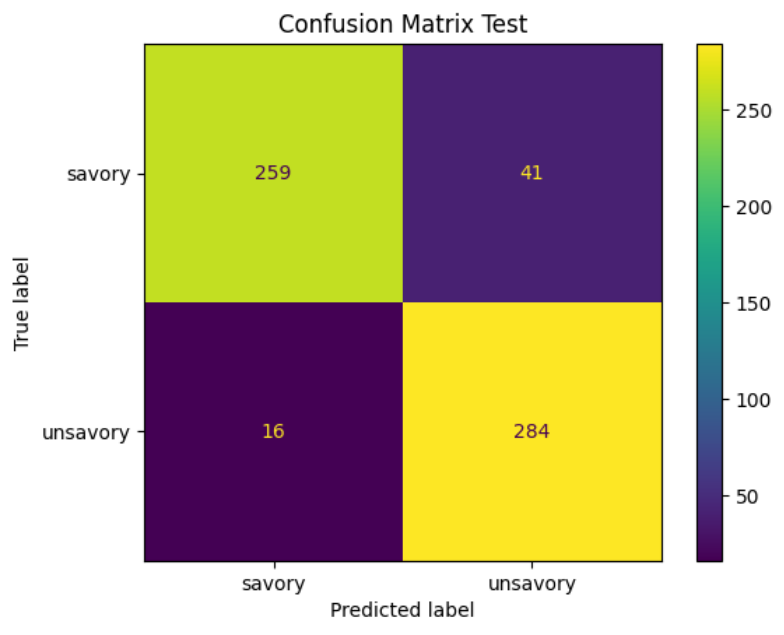
Struttura CNN



- Alg. di ottimizzazione: Adam
- Learning rate: 0.001

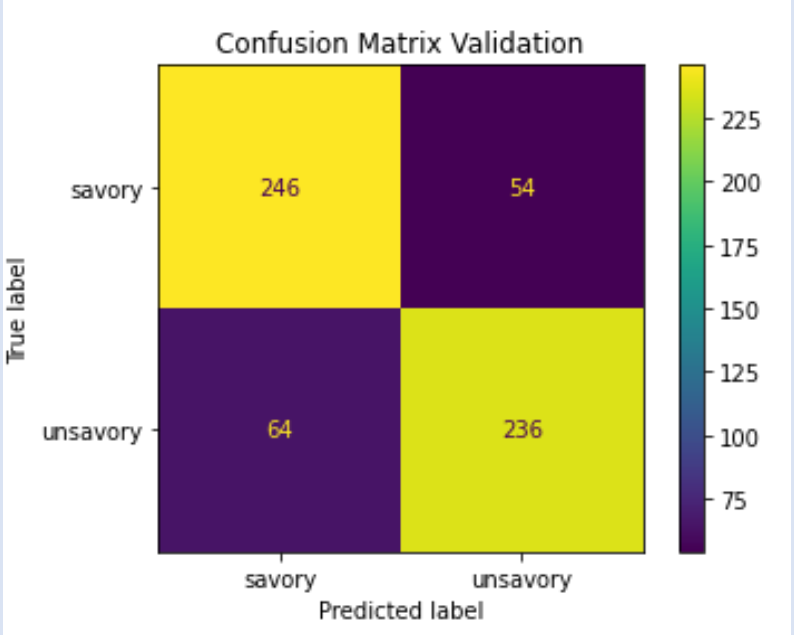
Classification Performance

SVM: BOVW



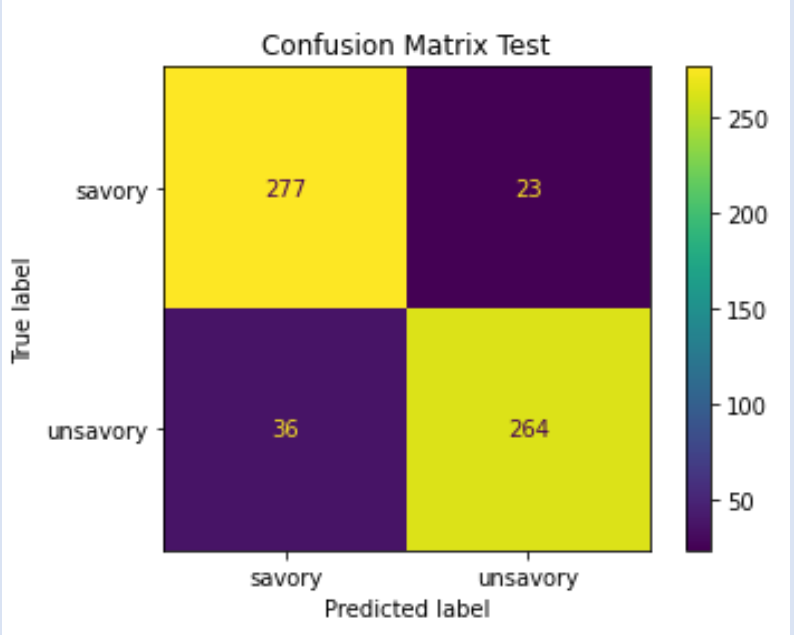
	Precision	Recall	F1-score
Savory	94%	86%	90%
Unsavory	87%	95%	91%

Random Forest: Color



	Precision	Recall	F1-score
Savory	82%	95%	88%
Unsavory	94%	80%	86%

Random Forest: BOVW+Color



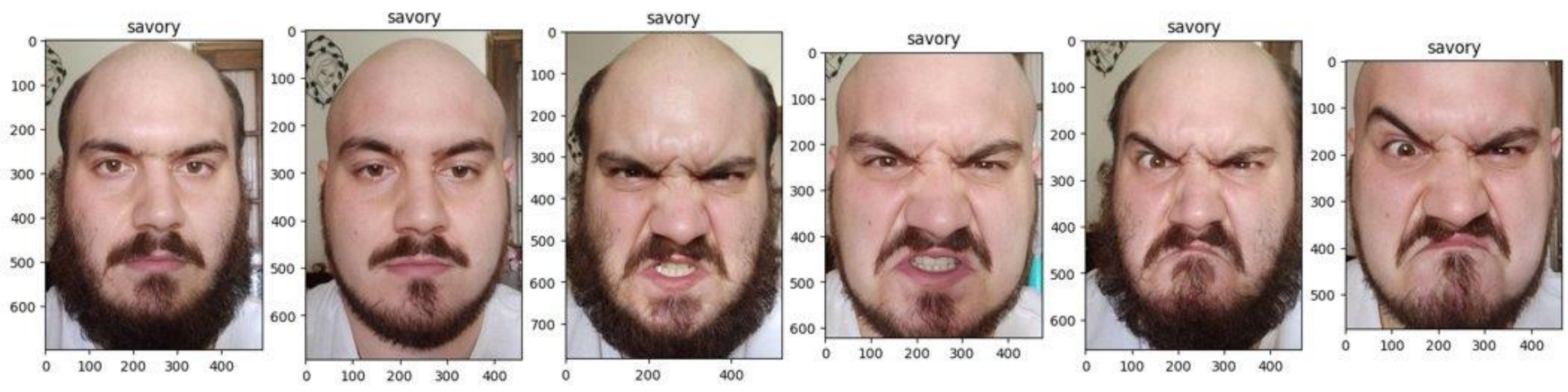
	Precision	Recall	F1-score
Savory	88%	92%	90%
Unsavory	92%	88%	90%

BOVW Robustezza (1)

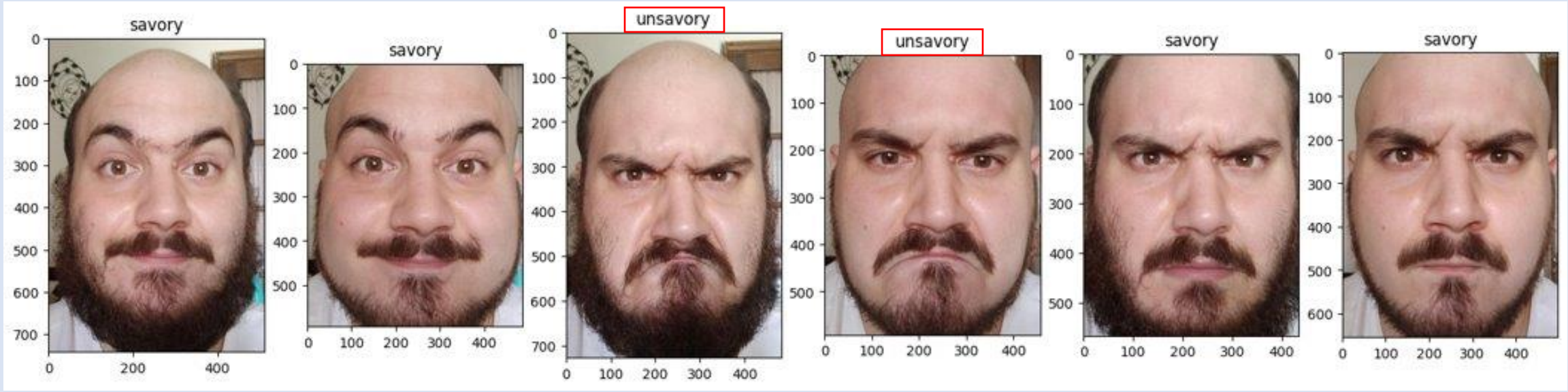
Passate alla classificazione, dopo preprocessing, 16 foto:

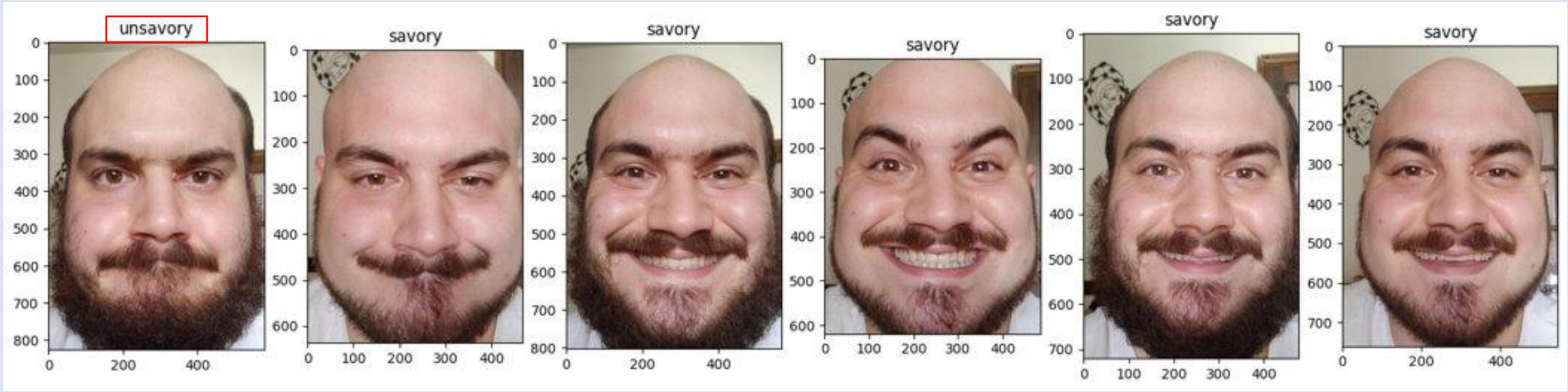
- 8 con barba lunga e "capelli"
- 8 con barba corta e rasato
- Foto accoppiate per espressione facciale

Test effettuato con modello trainato senza le nostre facce nel database



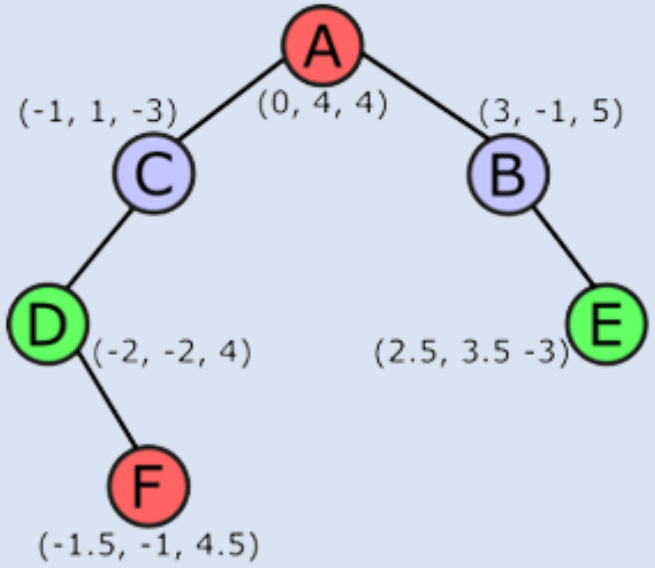
BOVW Robustezza (2)

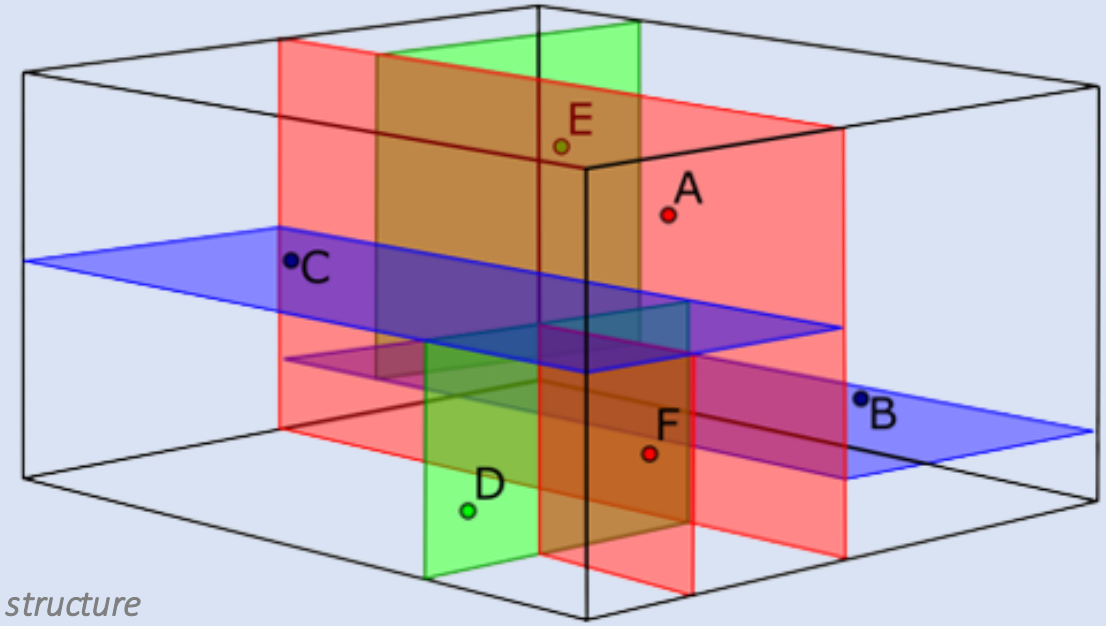




CBIR

- Retrieval tramite struttura dati ***KDTree***.
- Permette di organizzare punti (le features estratte dalle immagini) in uno spazio K-D partizionato.
- Garantisce una ricerca in tempo O(n).
- Per ogni modello utilizzato viene costruito il rispettivo albero con le relative features estratte dal train set.
    - Per la CNN, le features sono state estratte dall'ultimo layer di maxpooling (5184-D).





Esempio KDTree data structure

Retrieval Performance

- Sono state scelte 20 immagini casuali, 10 per classe.
- Come metrica per il retrieval è stata utilizzata la Precisione (P) media, a diverse soglie: 1, 5 e 10.

$$P = \frac{|\{\text{No. relevant images}\} \cap \{\text{No. retrieved images}\}|}{|\{\text{No. retrieved images}\}|}$$

Model	P@1	P@5	P@10
BOVW	50%	45%	44%
Color Histogram	60%	44%	43%
BOVW+Color	55%	47%	43%
CNN	**75%**	**68%**	**65%**

Esempio Retrieval

NB: Il bot utilizza due tree separati per le due classi, ritornando i 5 piu simili savory e i 5 più simili unsavory

* You are: savory, with 99.0 confidence! *

/1 - unsavory - Rank: 1 - dist: 0.061 /2 - savory - Rank: 2 - dist: 0.066 /3 - unsavory - Rank: 3 - dist: 0.066 /4 - unsavory - Rank: 4 - dist: 0.066 /5 - savory - Rank: 5 - dist: 0.067

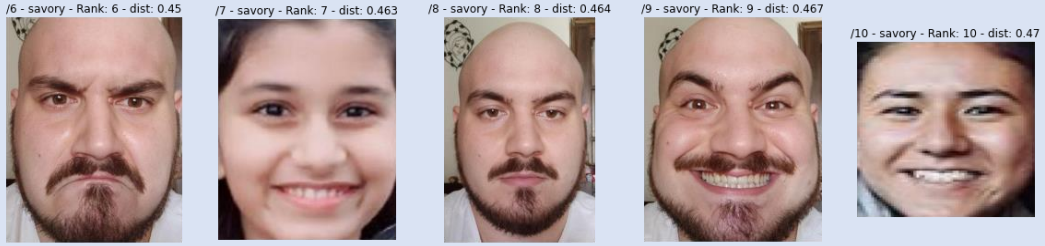
/6 - savory - Rank: 6 - dist: 0.067 /7 - unsavory - Rank: 7 - dist: 0.067 /8 - savory - Rank: 8 - dist: 0.067 /9 - savory - Rank: 9 - dist: 0.067 /10 - unsavory - Rank: 10 - dist: 0.067

BOVW

* You are: savory, with 69.0 confidence! *

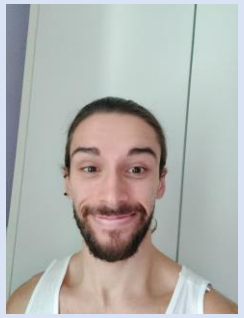
/1 - savory - Rank: 1 - dist: 0.392 /2 - savory - Rank: 2 - dist: 0.431 /3 - savory - Rank: 3 - dist: 0.434 /4 - savory - Rank: 4 - dist: 0.439 /5 - savory - Rank: 5 - dist: 0.443

/6 - savory - Rank: 6 - dist: 0.45 /7 - savory - Rank: 7 - dist: 0.463 /8 - savory - Rank: 8 - dist: 0.464 /9 - savory - Rank: 9 - dist: 0.467 /10 - savory - Rank: 10 - dist: 0.47



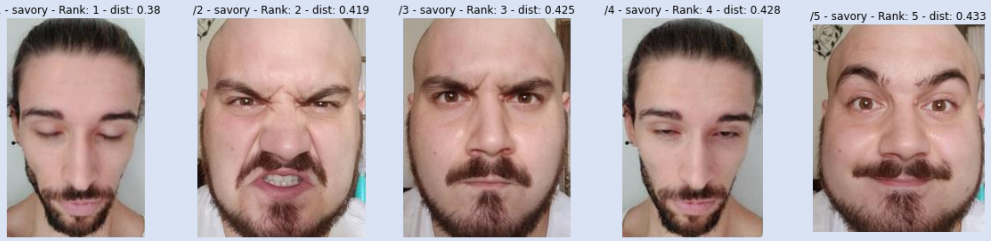
BOVW+Color

Query-by-example: savory

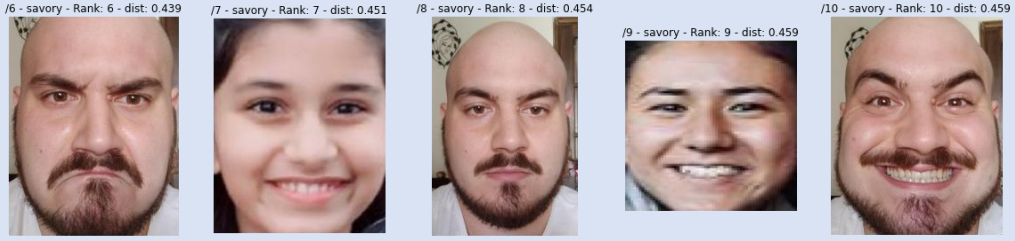


* You are: savory, with 80.0 confidence! *

/1 - savory - Rank: 1 - dist: 0.38 /2 - savory - Rank: 2 - dist: 0.419 /3 - savory - Rank: 3 - dist: 0.425 /4 - savory - Rank: 4 - dist: 0.428 /5 - savory - Rank: 5 - dist: 0.433



/6 - savory - Rank: 6 - dist: 0.439 /7 - savory - Rank: 7 - dist: 0.451 /8 - savory - Rank: 8 - dist: 0.454 /9 - savory - Rank: 9 - dist: 0.459 /10 - savory - Rank: 10 - dist: 0.459

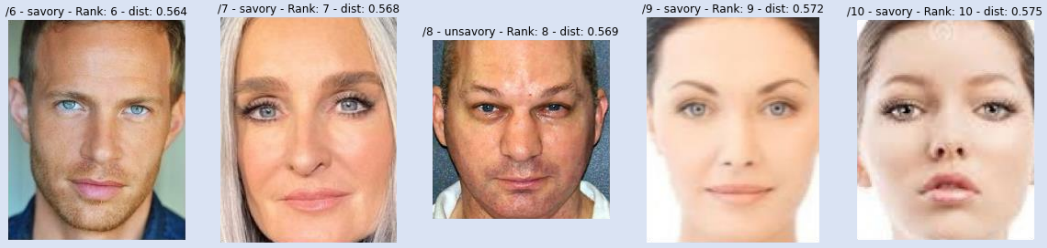


Color

* You are: savory, with 97.00000286102295 confidence! *

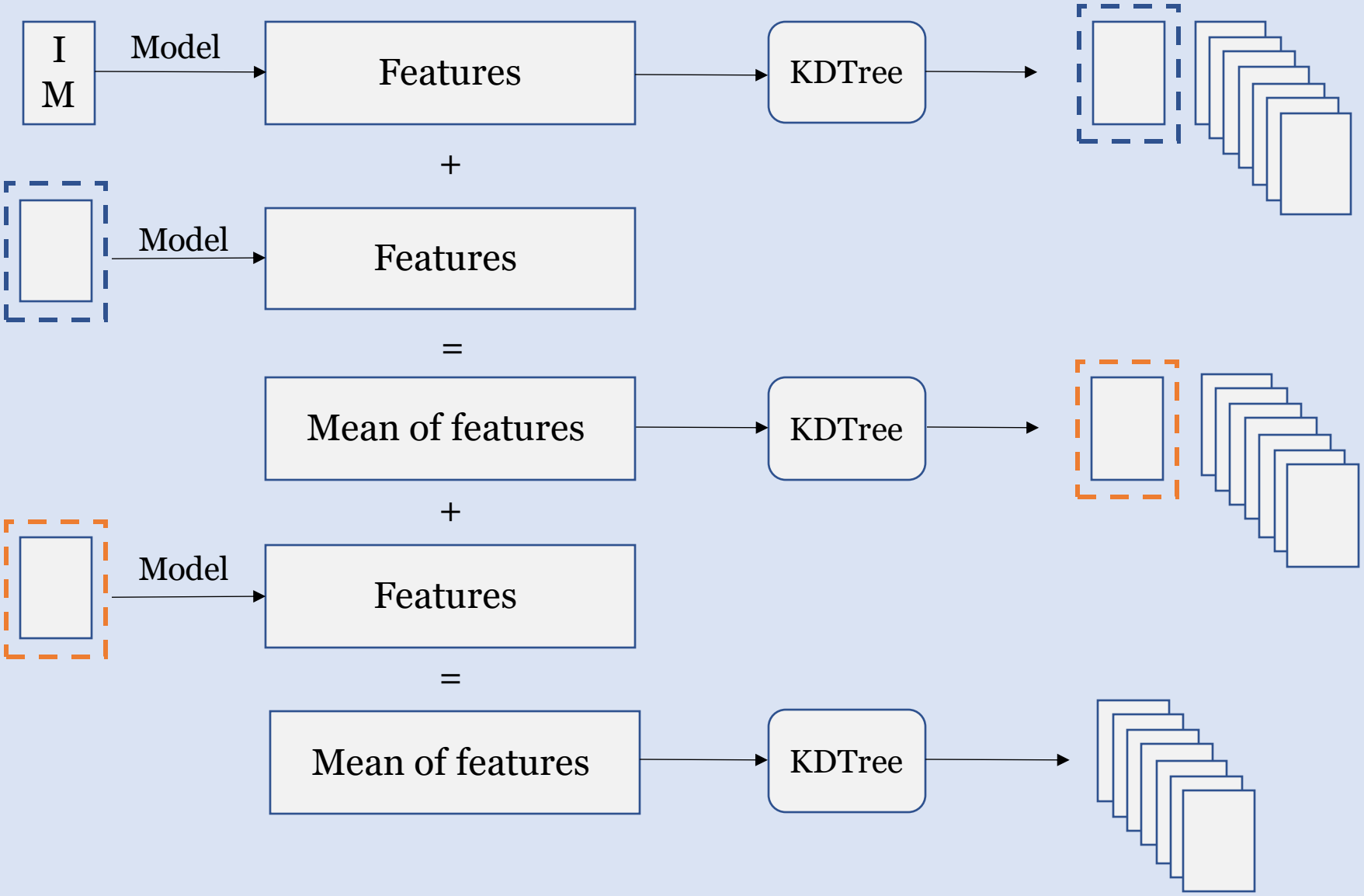
/1 - savory - Rank: 1 - dist: 0.501 /2 - savory - Rank: 2 - dist: 0.518 /3 - savory - Rank: 3 - dist: 0.532 /4 - savory - Rank: 4 - dist: 0.557 /5 - savory - Rank: 5 - dist: 0.562

/6 - savory - Rank: 6 - dist: 0.564 /7 - savory - Rank: 7 - dist: 0.568 /8 - unsavory - Rank: 8 - dist: 0.569 /9 - savory - Rank: 9 - dist: 0.572 /10 - savory - Rank: 10 - dist: 0.575



CNN

Iterative Relevance Feedback



Possibili sviluppi da test scartati

Siamese Network: modello scartato, ottima classificazione, troppo oneroso computazionalmente, scarso retrieval

Utilizzo di **hand-crafted features** estratte dalle sezioni del viso per il retrieval

Fine-tuning di CNN come DenseNet e ResNet

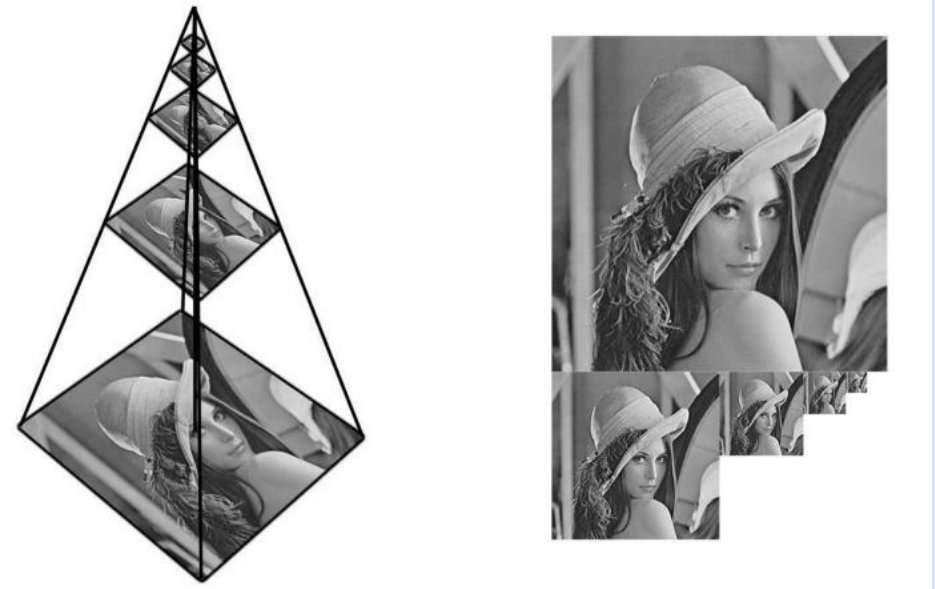
Conclusioni

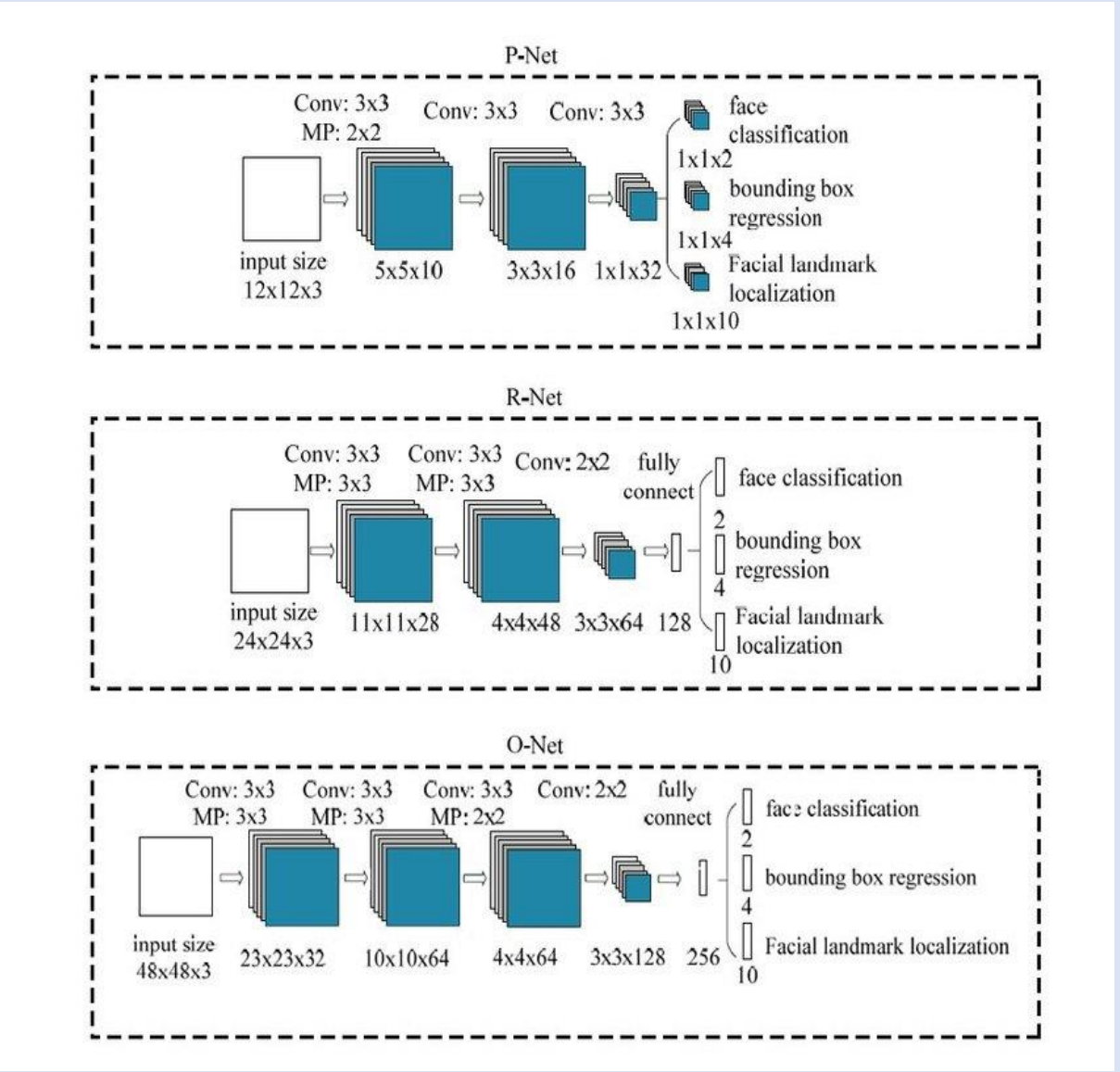
Il task di classificazione binaria risulta, anche grazie alla qualità del dataset, facilmente risolvibile con vari modelli

Il task di retrieval risulta più complesso poiché è sono necessarie feature più discriminanti per discernere tra un volto e un altro

In classificazione, quindi, il descrittore migliore è risultato essere BOVW, mentre in fase di retrieval la CNN riesce ad estrapolare una rappresentazione più efficacie.

Appendice - MTCNN



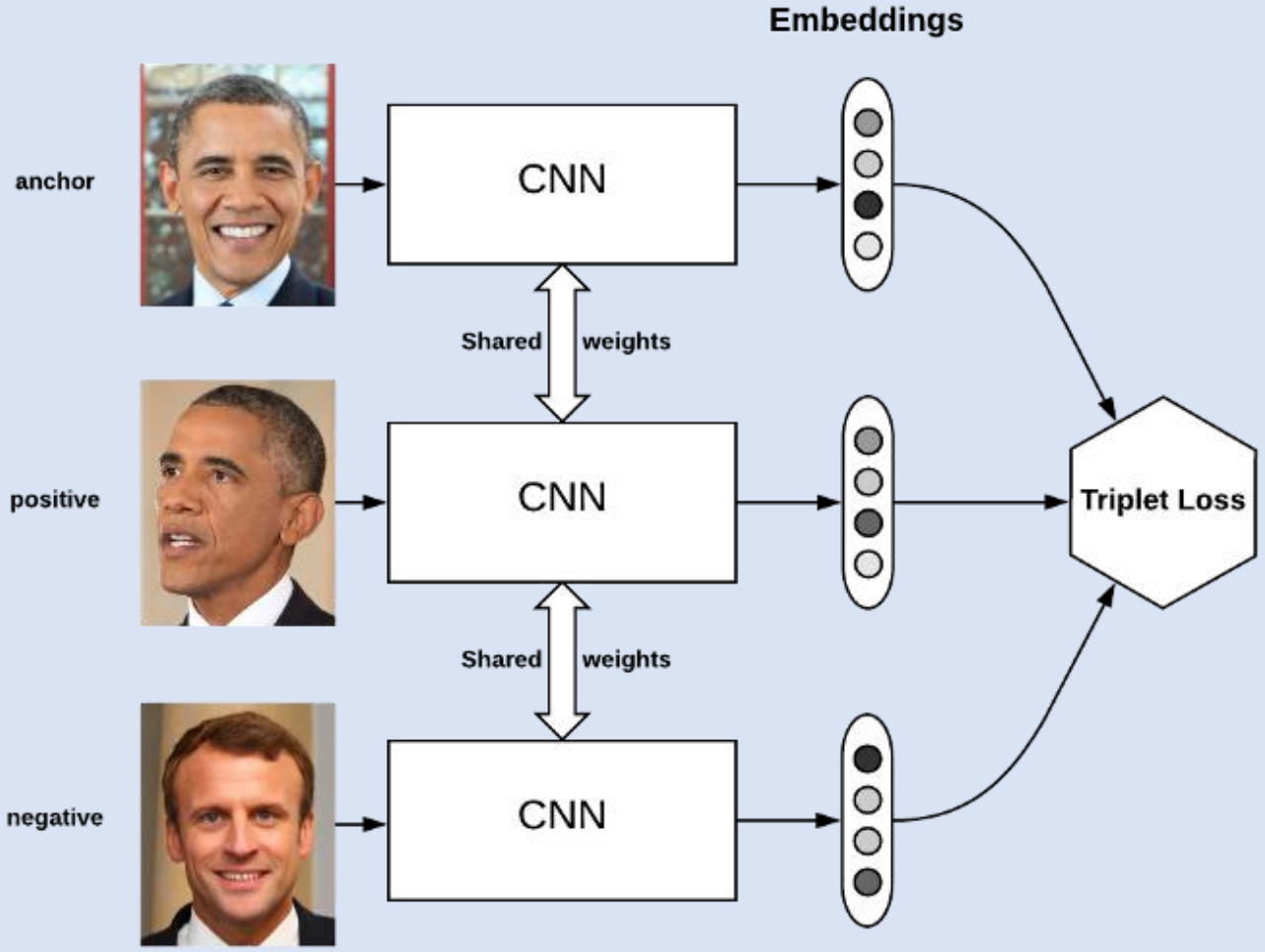


Appendice - Bayesian optimization

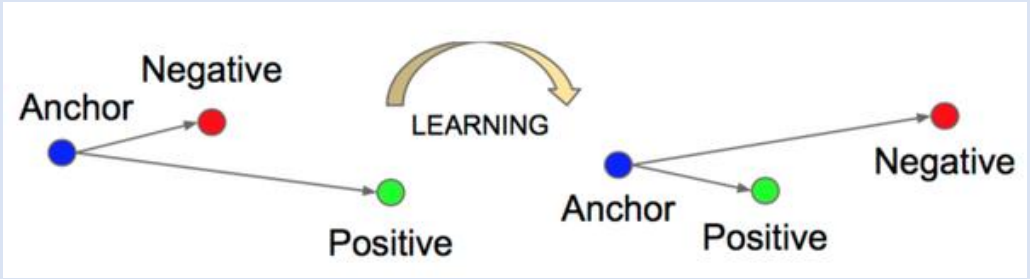
- Approcci manuali o Grid Search per l'individuazione degli iperparametri migliori non efficienti.
- Sfrutta il teorema di Bayes, per limitare lo spazio di ricerca degli iperparametri di un modello.
  - P(A|B) = P(B|A)*P(A) / P(B)
- L'ottimizzazione sfrutta la conoscenza ottenuta durante i vari tentativi, per guidare la ricerca verso gli iperparametri migliori.

# Filtri I conv.	Dim. Filtro I conv.	Dim. I Max pooling	# layer Conv	# Filtri i-esima conv.	# Dense layer	Dim. I-esimo Dense layer	Apply Dropout	Optimizer	Learning rate
[32, 64]	[(3x3), (5x5)]	[(3x3), (5x5)]	[1, 2]	[64, 128, 256]	[1, 2]	[64, 128, 256]	[True, False]	[adam, rmsprop, adamax]	[0.001, 0.01]

Appendice - Tentativi – Siamese CNN

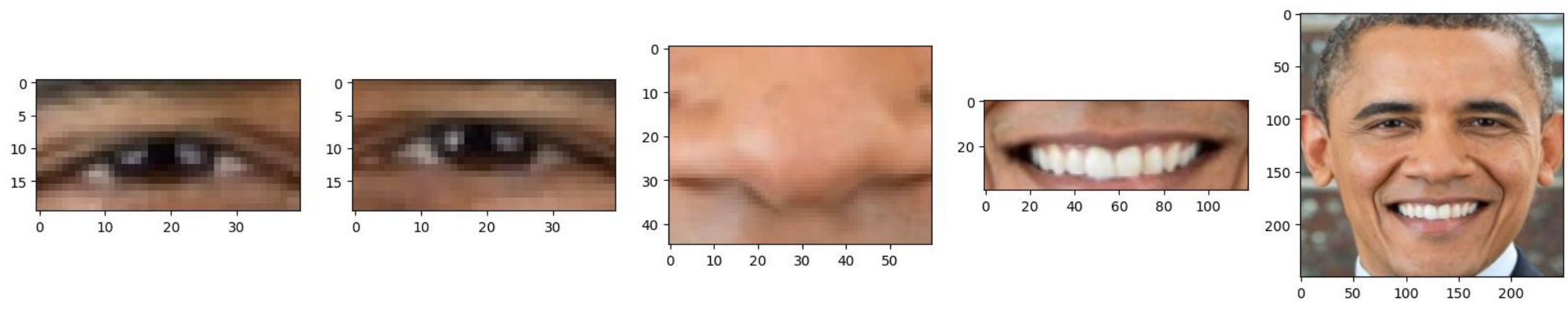


Triplet loss on two positive faces (Obama) and one negative face (Macron). (Source: Olivier Moindrot)



Before (left) and after (right) minimizing triplet loss function (Source: FaceNet)

Appendice - Tentativi – Hand-crafted features



Features:

- Colore medio occhi
- Distanza occhio-occhio, occhi-naso, naso-bocca, lunghezza bocca
- Per tutti gli elementi, dopo trasholding:
  - Asse maggiore e minore
  - Area
  - Perimetro
  - Eccentricità