GABAR
Gender & Age Based Actor Retrieval

Luca Cogo (830045), **Lidia Lucrezia Tonelli** (813114), **Gianluca Giudice** (830694)

OUR GOAL

Given an image with one or more faces, we aim to detect them and estimate their gender and age.
With these information we will propose an actor that resembles the selected person in the image.

CONTENTS OF THIS PRESENTATION

01.

TASK SPECIFICATION

02.

IMAGE ENHANCEMENT

03.

FACE DETECTION

04.

GENDER CLASSIFICATION & AGE REGRESSION

05.

SIMILARITY FOR ACTOR RETRIEVAL

06.

DEMO & FUTURE WORKS

01.

TASK SPECIFICATION

BOT FINAL WORKFLOW

IMAGE IS ENHANCED

The bot receives an image and enhances it to improve detection and classification

FACES ARE DETECTED

One or more faces in the image are detected with YOLO: the user is asked to choose one. If no faces are detected, the bot will ask for a new input

GENDER AND AGE ARE ESTIMATED

A machine learning model estimates gender and age of the selected person

SIMILAR ACTOR IS PROPOSED

Gender, age and the model features are used to compute the faces similarity and find an actor that resembles the selected person

02.

IMAGE ENHANCEMENT

ENHANCEMENT OPERATIONS

ADAPTIVE GAMMA CORRECTION

To correct the brightness in the image - adapted from [1]

BILATERAL FILTER

To reduce eventual noise but preserving edges (useful for classification)
filter size 7, sigma 50 (as suggested in cv2 docs)

[1] Moroney, Nathan. (2000). Local Color Correction Using NonLinear Masking. Final Program and Proceedings - IS and T/SID Color Imaging Conference. 108-111.

ENHANCEMENT EXAMPLE

INPUT IMAGE

ADAPTIVE GAMMA CORRECTION

BILATERAL FILTER

OUTPUT IMAGE

DETAILS ON ADAPTIVE GAMMA

RGB IMAGE → YCbCr COLOR SPACE → Y channel is the intensity → Y inverted and low-pass filtered (bilateral) → Mask is applied to the original Y → Image is reassembled and converted in RGB

$$y_{new} = 255\left(\frac{y}{255}\right)^{2^{\left(0.5\cdot\frac{128-mask}{128}\right)}}$$

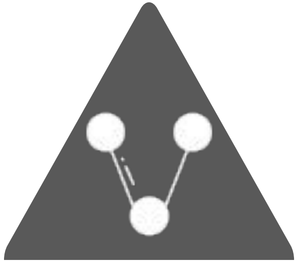
Coefficient 0.5 is added to reduce the weight of the mask

03.

FACE DETECTION

EXPLORED DETECTION METHODS

CASCADE DETECTOR



FINE-TUNED YOLO

Cascade detector

CASCADE DETECTOR

POSITIVE IMAGES: FDDB

Face Detection Data Set and Benchmark (FDDB) is a data set of face regions designed for studying the problem of unconstrained face detection. This data set contains the annotations for 5171 faces in a set of 2845 images taken from the Faces in the Wild data set.

NEGATIVE IMAGES: CALTECH 256

The Caltech 256 is a dataset composed by 30607 images in this dataset spanning 257 object categories. Object categories are extremely diverse, ranging from grasshopper to tuning fork.
The categories that could contain some faces were removed.

CASCADE DETECTOR

Grid search approach for identifying the best choices for the cascade detector:

- Feature type: HAAR; HOG; LBP
- False alarm rate: 0.01; 0.05; 0,1
- Number of stages: 5; 10; 20; 30

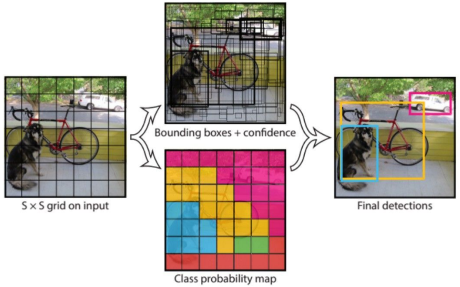
FEATURE TYPE	STAGES	FAR	TRAIN TIME (sec)	PRECISION	RECALL	F-SCORE	AVG PRED TIME
HOG	28	0.01	20559	**0.829**	0.73	**0.776**	0.0598
HAAR	TOO MUCH TIME!						
LBP	20	0.01	**5733**	0.749	**0.773**	0.761	**0.0475**

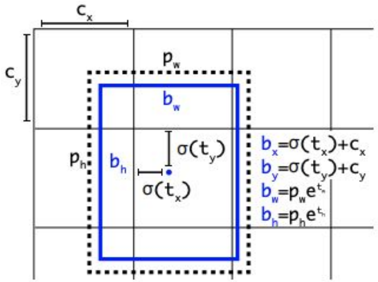
Only the best result for each feature type was included. The object training size was set at [48; 32]

Yolo detector

FINE-TUNED YOLO

- **Yolov3**[1] is SOTA object detection
  - Real time object detection
  - Trained on COCO dataset (a large-scale object detection dataset) that consists in 9000 object categories (only 90 are detected by the model)
  - The “Person” class is a category detected by pretrained Yolov3. However the **“Face” class is not detected**
- Yolov3 **fine-tuning** using Face Detection Data Set and Benchmark (FDDB) (dataset with bounding boxes)
  - A list of 4 points defining the bbox + 1 category
  - The **9 anchor points** are computed using **k-means on the FDDB** dataset (it is important to consider the **aspect-ratio of the faces**)





[1] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

EVALUATION OF FACE DETECTORS

- Predictions and groundtruth are compared by computing the IoU of the bounding boxes.
- A bounding box is a True Positive if IoU ≥ 0.5, otherwise it is a False Positive. If a bounding box from the groundtruth has no matching bounding box from the prediction, that is a False Negative.
- Precision, Recall and F-score are now computed using the total of True Positives, False Positives and False Negatives.

$$\text{Precision} = \frac{tp}{tp + fp} \quad \text{Recall} = \frac{tp}{tp + fn} \quad F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

DETECTORS COMPARISON

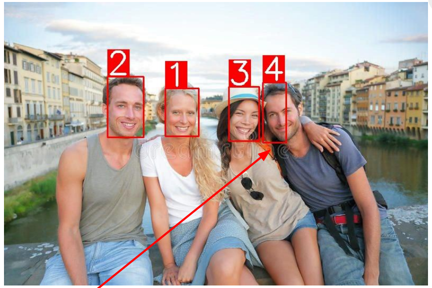
DETECTOR	TRAIN TIME (h)	PRECISION	RECALL	F-SCORE	AVG PRED TIME (on CPU)
LBP CASCADE	**1.6**	0.749	0.773	0.761	**0.0475**
YOLOv3	60	**0.864**	**0.840**	**0.852**	0.357

ENSEMBLE APPROACH

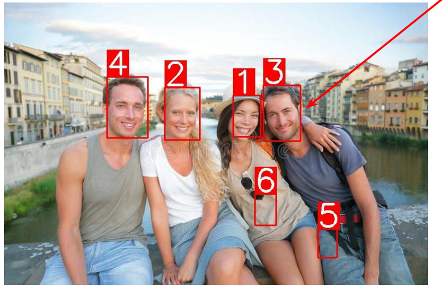
- The best method for face detection is the combination of both cascade and yolo
- **Idea**: given the high recall of cascade, it is possible to filter the bounding boxes using yolo
  - Starting from the yolo boxes, compute the IOU w.r.t. every boxes founded by cascade
  - If the IOU is greater than 0.5, keep the larger bounding box

INPUT IMAGE

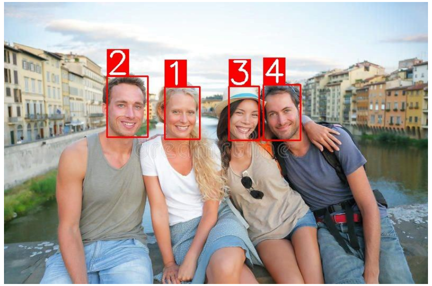
YOLO DETECTOR



CASCADE DETECTOR



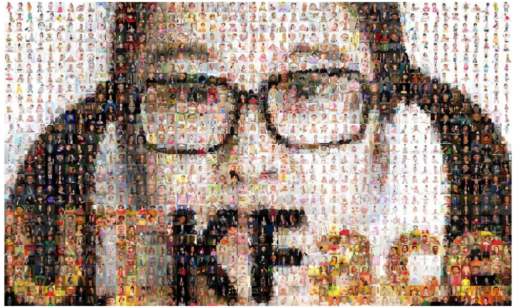
ENSEMBLE APPROACH



04.

GENDER CLASSIFICATION & AGE REGRESSION

DATASET FOR CLASSIFICATION AND REGRESSION



UTKFace Dataset

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity.

EXPLORED METHODS

GENDER CLASSIFICATION

- HANDCRAFTED FEATURES + SVM
- DEEP NN FROM SCRATCH
- FINE-TUNED VGG-FACE

AGE REGRESSION

- HANDCRAFTED FEATURES + DECISION TREE
- DEEP NN FROM SCRATCH
- FINE-TUNED VGG-FACE

Handcrafted features

HAND-CRAFTED FEATURES

SIFT + BOW REPRESENTATION

HOG

COLOR HISTOGRAM ON IMAGE DIVIDED IN 4 PARTS

LBP

SVM FOR GENDER CLASSIFICATION

+

DECISION TREE FOR AGE REGRESSION

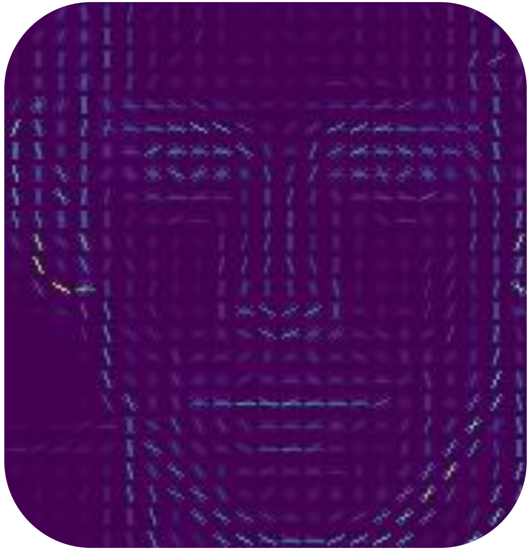
SIFT + VISUAL BOW REPRESENTATION

- **SIFT descriptors extraction** from every image in the training dataset:
    - Convert image to **grayscale**
    - Extract 25 SIFT keypoints and descriptors (vectors of 128 numbers)
    - Build **dictionary** of descriptors with **k-means** (k = 150)

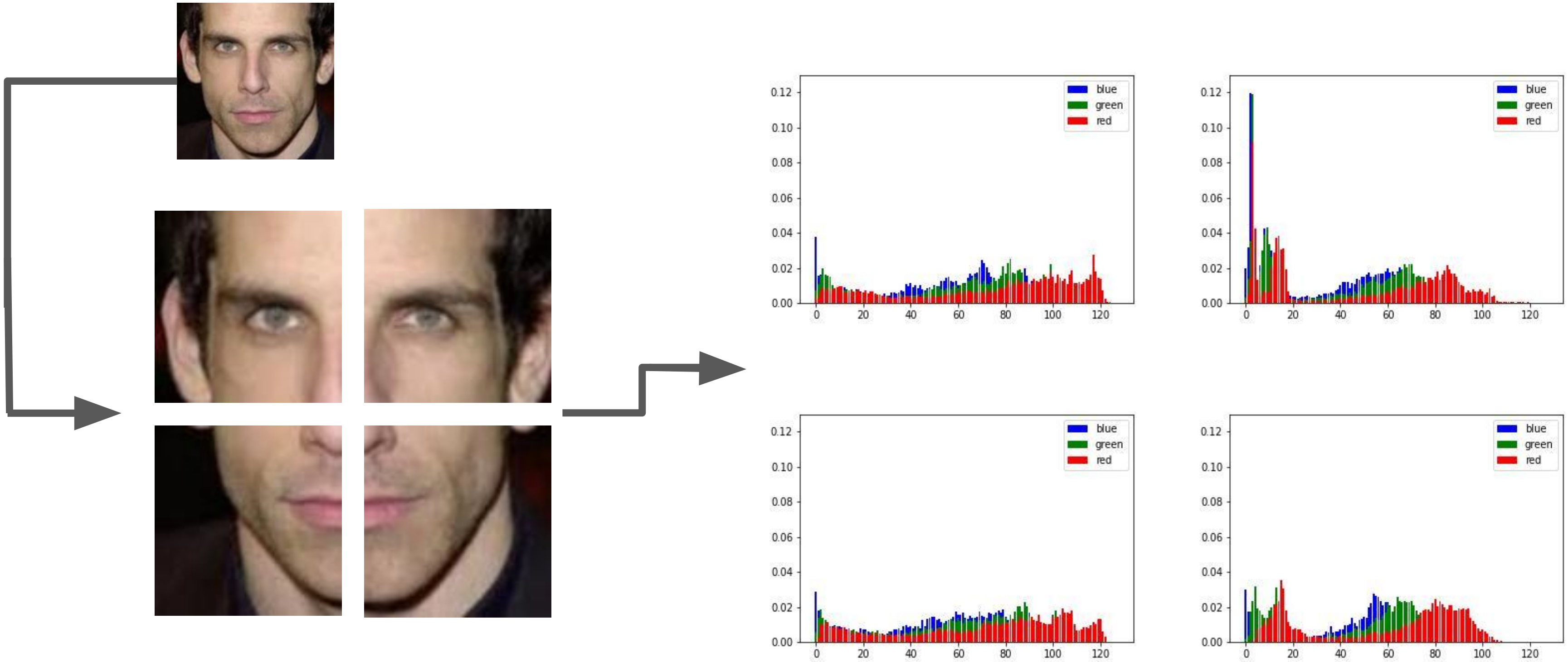
- For each training image:
    - Extraction of **SIFT descriptors** from gray image
    - For each descriptor predict the **dictionary word**
    - Compute **Histogram** → BOW representation of image

HOG

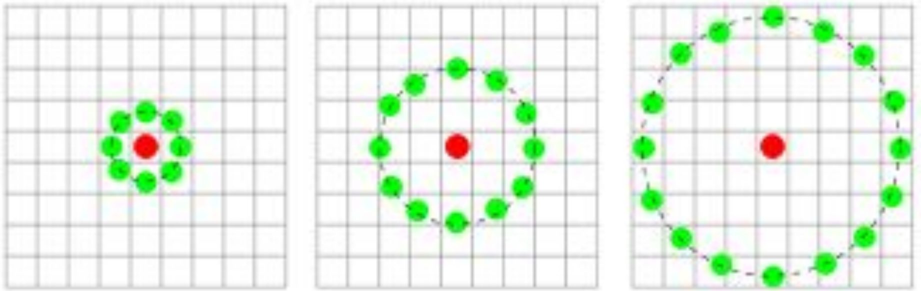
- **HOG descriptors extraction** from every image in the training dataset:
    - HOG features describe gradient and orientation of edges



COLOR HISTOGRAM



LBP



- **LBP descriptors extraction** from every image in the training dataset:
    - LBP describes local spatial patterns and gray scale contrast

METHODS

- Gender classification: **SVM**
  - Metrics: **accuracy**, F-score
- Age regression: **decision tree**
  - Metrics: Mean Absolute Error, **top-k accuracy** (defined by us)
  - Top-k accuracy considers expected ages correct if they are distant less or equal k from the true ages
- Search for best parameters combinations, possible values:
  - # SIFT points: 10 (>10 has been tried…)
  - # bins of color histogram: 32, 64, 128
  - # LBP points and LBP radius: (8,1), (16,2), (24,3)

TOO MUCH TIME FOR >10 SIFT POINTS!

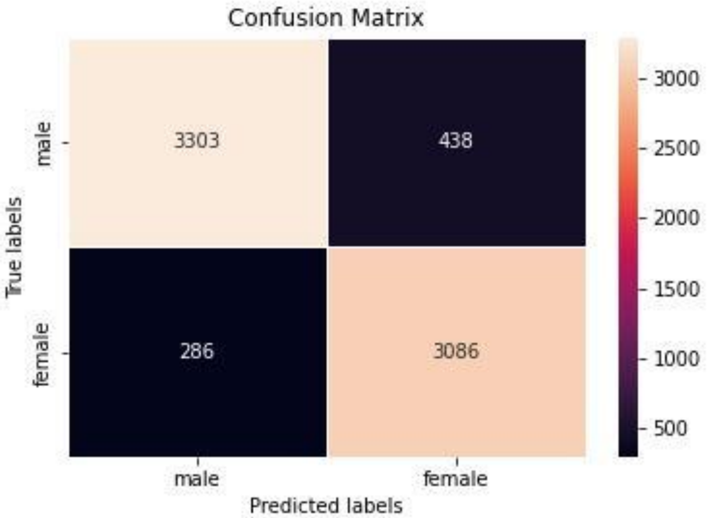
PARAMETERS SEARCH

- Grid-search to find the **best combination of features parameters**: number of SIFT keypoints, color histogram bins, LBP radius, LBP points number

SIFT	HIST	HOG	LBP	Gender accuracy	Gender F-score	Age MAE	Age Top-5	Age Top-10	Training time
#points=10	#bins=32	Y	p=24, r=3	88,21 %	87,42 %	14,22	31,71 %	50,28 %	6h04m
#points=10	N	N	N	59,86 %	55,27 %	21,18	21,68 %	35,56 %	4h26m
N	#bins=128	Y	p=16, r=2	88,06 %	87,25 %	14,18	**32,98 %**	**51,41 %**	1h42m
N	#bins=128	N	N	72,34 %	69,15 %	19,63	24,27 %	38,82 %	**31m**
N	N	N	p=24, r=3	66,71 %	63,13 %	16,71	26,86 %	43,05 %	42m
N	N	Y	N	**88,31 %**	**87,50 %**	15,15	31,12 %	48,64 %	1h17m

BEST PARAMETERS COMBINATION

SVM Gender Classifier:



Accuracy	89,82 %
F-Score	89,5 %

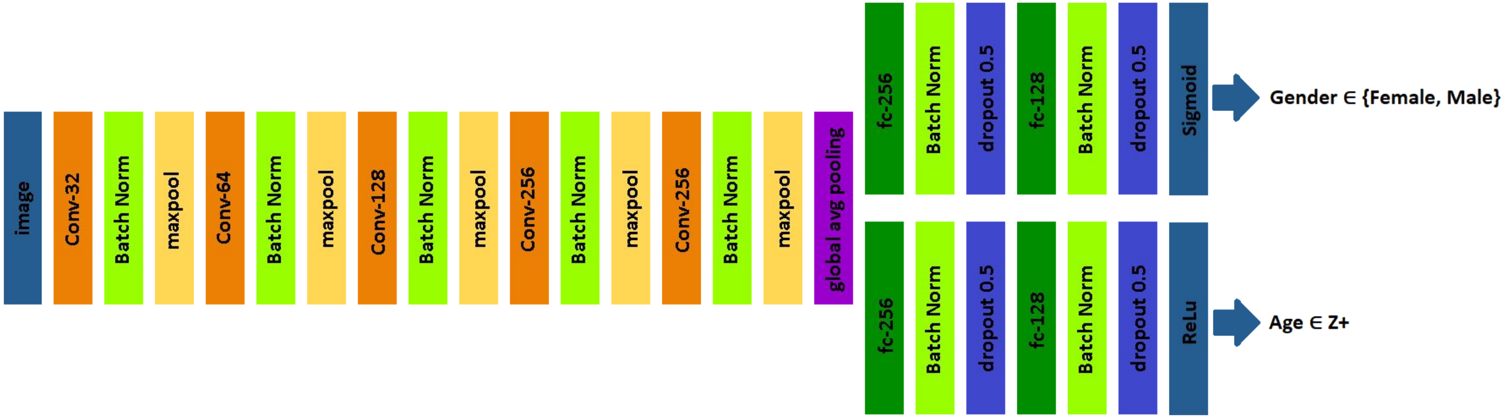
Decision Tree Age Regressor:

MAE	14,95
Top-5 accuracy	31,12 %
Top-10 accuracy	48,9 %
Top-15 accuracy	62,17 %
Top-20 accuracy	71,71 %


LET'S SPRINKLE
SOME DEEP LEARNING
memegenerator

NN from scratch

DEEP NN FROM SCRATCH



Multi-Task learning problem:

- Speed-up in training phase
- The nature of MTL force to capture general features for faces, thus leading to some sort of regularization

Custom loss function:

- Train jointly the *gender* and *age* heads using a custom loss function *(y = gender; z = age)*

$$L = \lambda_1(y\log(\hat{y}) + (1-y)\log(1-\hat{y})) + \lambda_2(z-\hat{z})^2$$

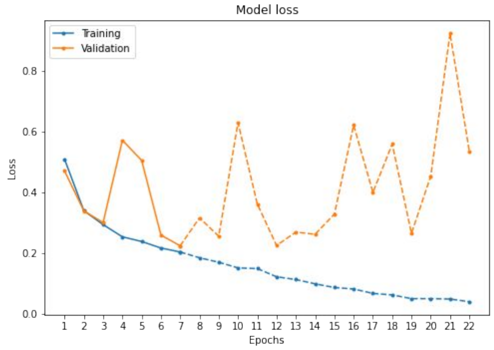
$\lambda_1 = 3; \quad \lambda_2 = 1$ *i.e.*, gender 75%; age 25%

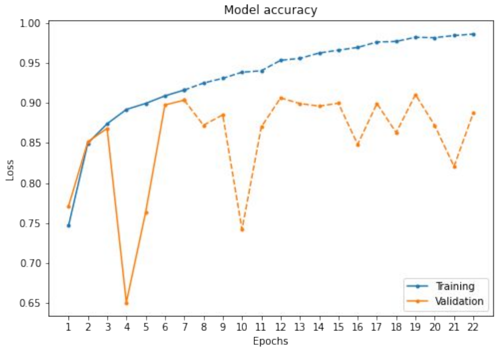
DEEP NN FROM SCRATCH

Bayesian Optimization was used for hyperparameters optimization.

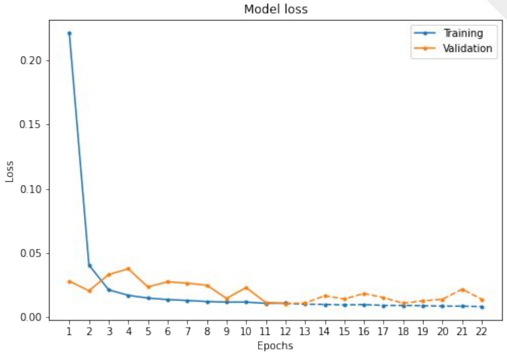
	POSSIBLE VALUES	BEST VALUE
Dropout rate (same for each layer)	0.1; 0.2; 0.3; 0.4; 0.5	0.5
Learning rate for Adam	1e-1; 1e-2; 1e-3; 1e-4	1e-4

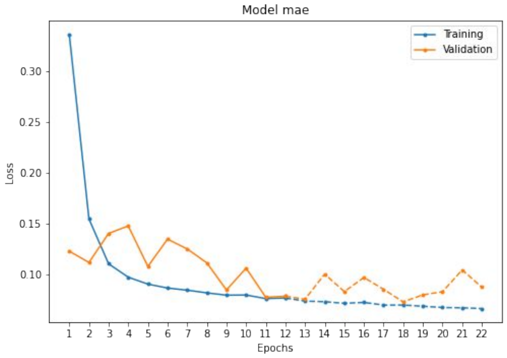
Gender Head





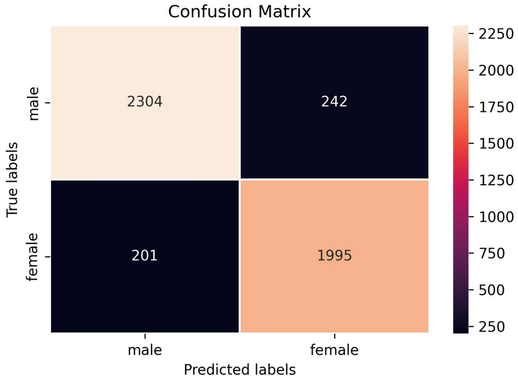
Age Head





DEEP NN FROM SCRATCH

Gender Classifier



Accuracy	90.6 %
F-Score	90.0 %

Age Regressor

MAE	8.91
Top-5 accuracy	36.2 %
Top-10 accuracy	64.45%
Top-15 accuracy	82.73%
Top-20 accuracy	92.18%

Fine-tuning VGG

FINE-TUNED VGG-Face

VGG-Face is a CNN for face recognition that was trained using 2.6 million faces
This network is the starting point for the fine tuning using the UTKFace dataset

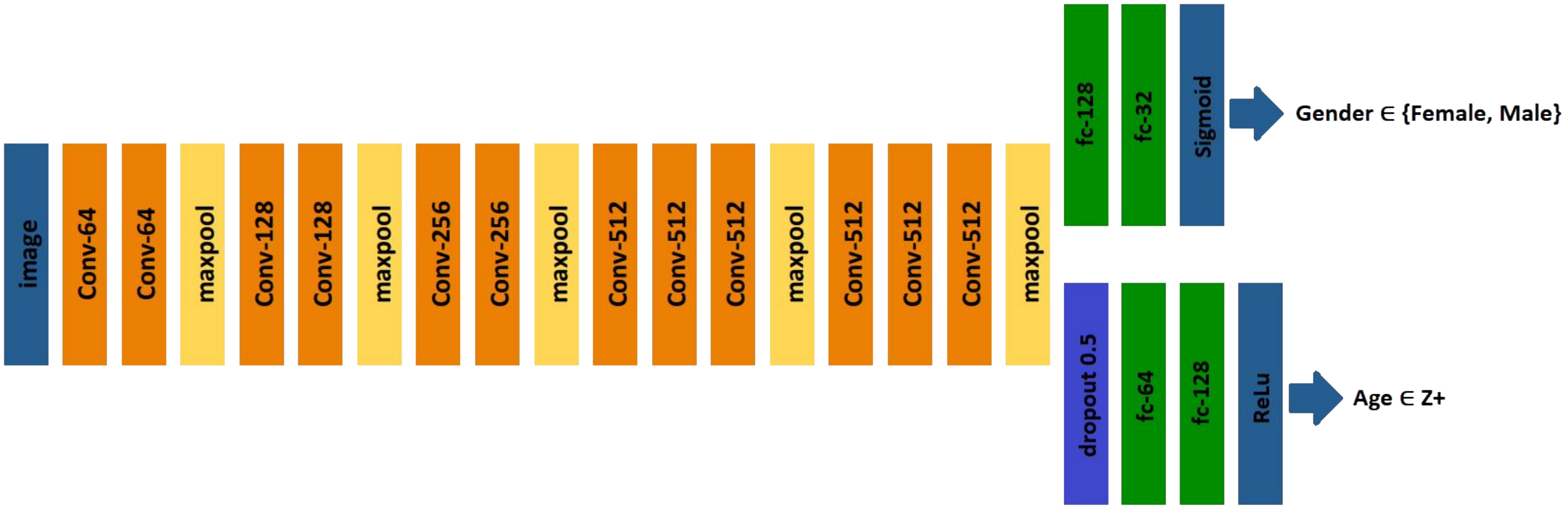
FINE-TUNED VGG-Face

The **bayesian optimization** algorithm was used In order to choose the best hyperparameters for the last layers of the network.

	POSSIBLE VALUES	GENDER CLASSIFICATOR BEST VALUES	AGE REGRESSOR BEST VALUES
Dropout after convolutional blocks	None; 0.2; 0.5	None	0.5
Number of dense layers	1; 2	2	2
1° dense layer size	64; 128	128	64
2° dense layer size	32; 64; 128	32	128
Dropout after dense layers	None; 0.2; 0.5	None	None
Learning rate	1e-2; 1e-3; 1e-4	1e-2	1e-4

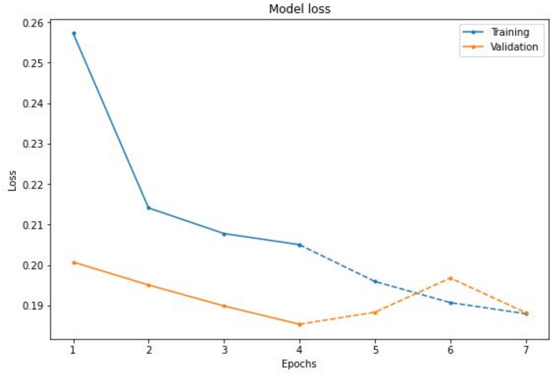
Both the models were trained using Adam as optimization function. For the Gender classifier, binary-crossentropy was used as loss function, while mean squared error was used for the Age regressor.

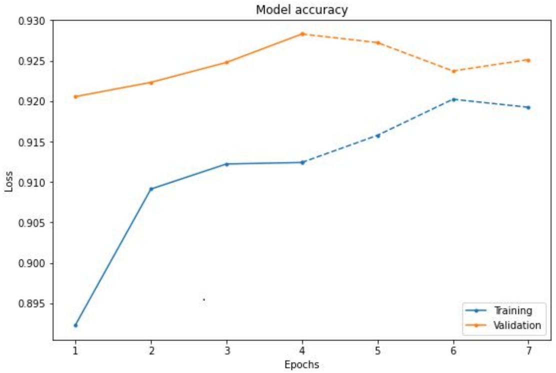
FINE-TUNED VGG-Face



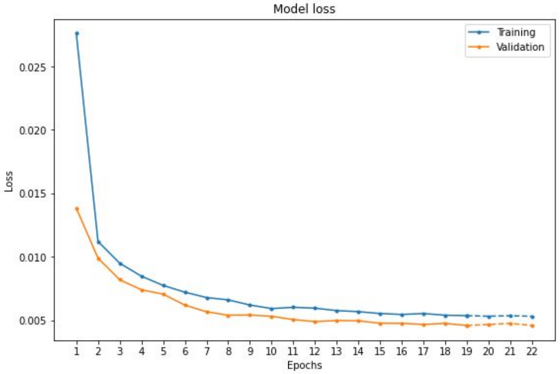
FINE-TUNED VGG-Face

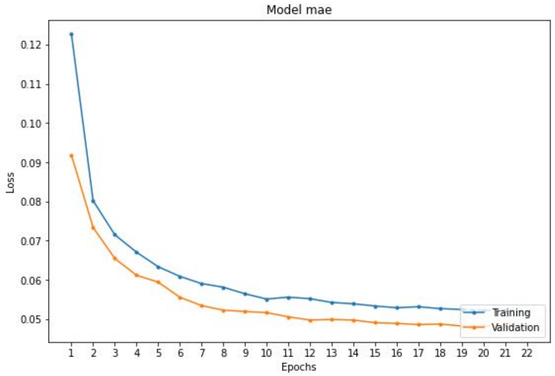
Gender Classifier





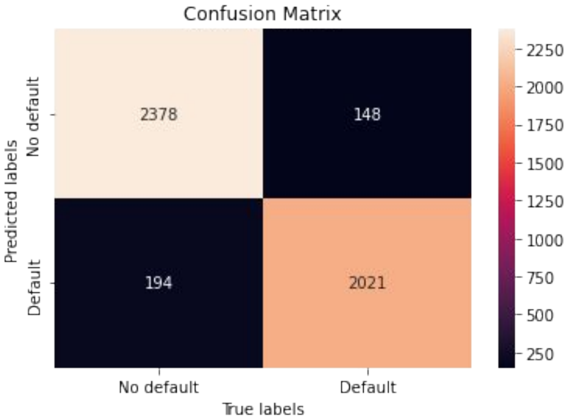
Age Regressor





FINE-TUNED VGG-Face

Gender Classifier



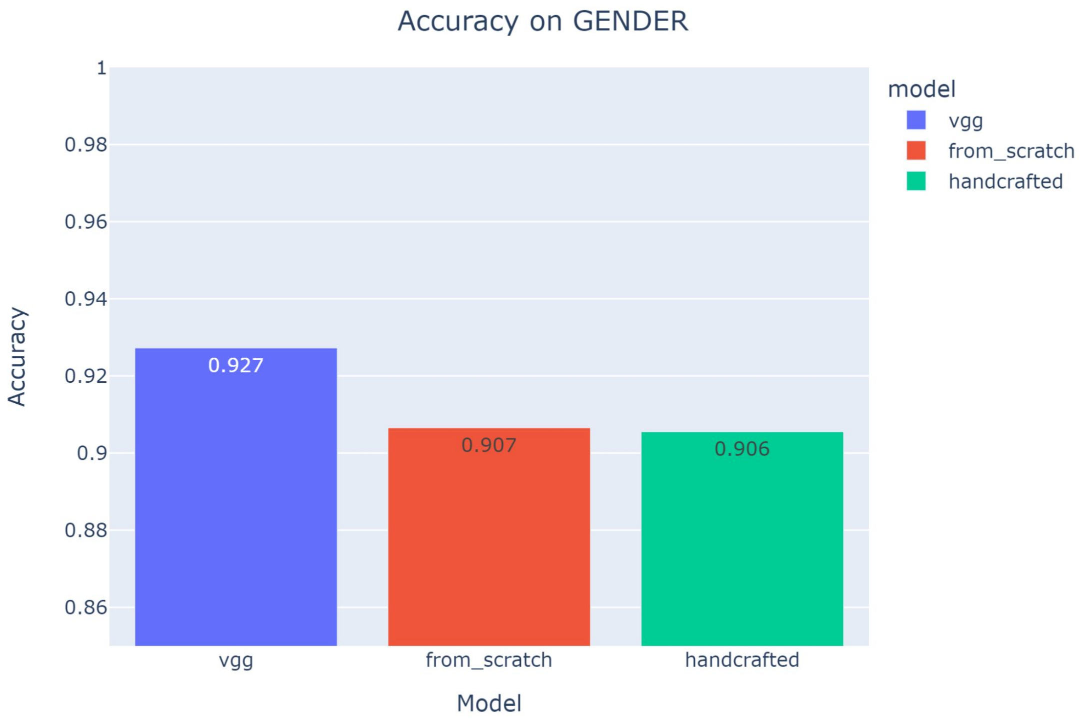
Accuracy	93%
F-Score	93%

Age Regressor

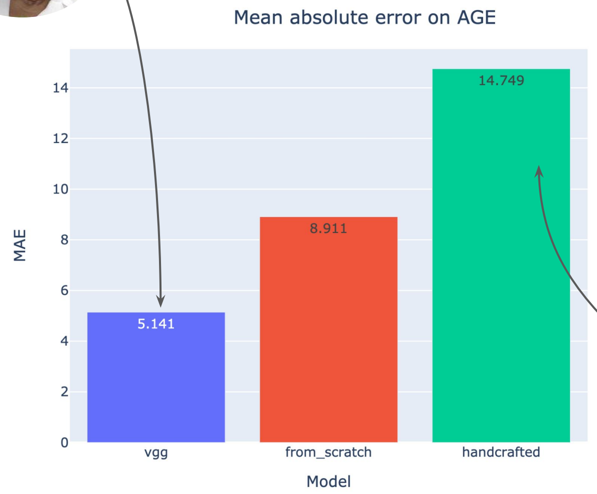
MAE	5.452
Top-5 accuracy	59.21%
Top-10 accuracy	84%
Top-15 accuracy	94%
Top-20 accuracy	97.5%

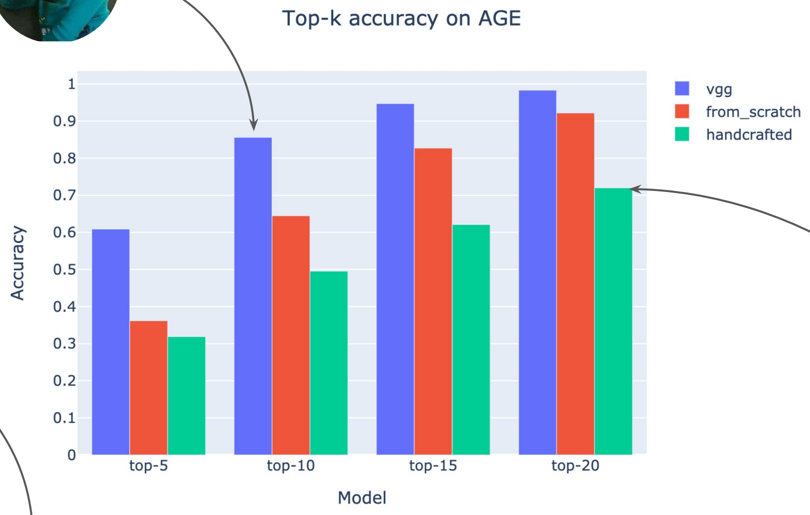
Models comparison

MODELS COMPARISON: GENDER

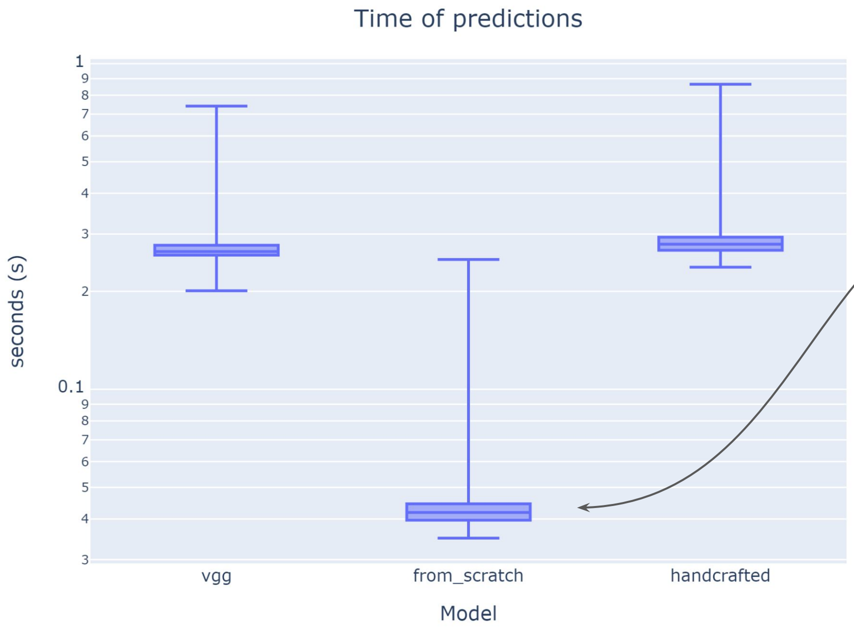


MODELS COMPARISON: AGE





MODELS COMPARISON: TIME



MODELS COMPARISON – SUMMARY

MODEL	Gender accuracy	Age MAE	Age Top-5 accuracy	Age Top-10 Accuracy	Avg prediction time
VGG-Face	**92.7 %**	**5.14**	**60.9 %**	**85.6 %**	0.2751
From scratch	90.7 %	8.91	36.2 %	64.5 %	**0.0429**
handcrafted	90.6 %	14.74	31.9 %	49.5 %	0.2886

MODELS COMPARISON

- Our chosen model is **fine-tuned VGG-Face**
- It has the best MAE and top-5 accuracy on age regression
- It has an acceptable time of prediction

05.

SIMILARITY FOR ACTOR RETRIEVAL

RETRIEVAL WORKFLOW

GENDER FILTERING

The actors dataset is filtered to find the suitable actors w.r.t. the estimated gender and age

FEATURE EXTRACTION

VGG-Face features are extracted from the selected face

SIMILARITY COMPUTATION

The extracted features are compared with all other actor’s features and ages (by a LUT) and a measure of distance is computed using a metric

SIMILAR ACTOR IS PROPOSED

The actor with the highest similarity measure is proposed

ACTORS DATASET

- Names and images **scraping** from https://today.yougov.com/ratings/entertainment/popularity/all-time-actors-actresses/all
- Gender and age scraping from **Wikidata**
- Dataset of **614 actors** obtained

Jamie Lee Curtis
Gender: female
Age: 63

Keanu Reeves
Gender: male
Age: 57

Emma Roberts
Gender: female
Age: 30

VGG-Face FEATURE EXTRACTION

SIMILARITY

- The similarity between two faces is computed considering both the **extracted features** of the CNN and the predicted **age** + **gender**

Given a face with the features *A*, age *z* and gender *y*:

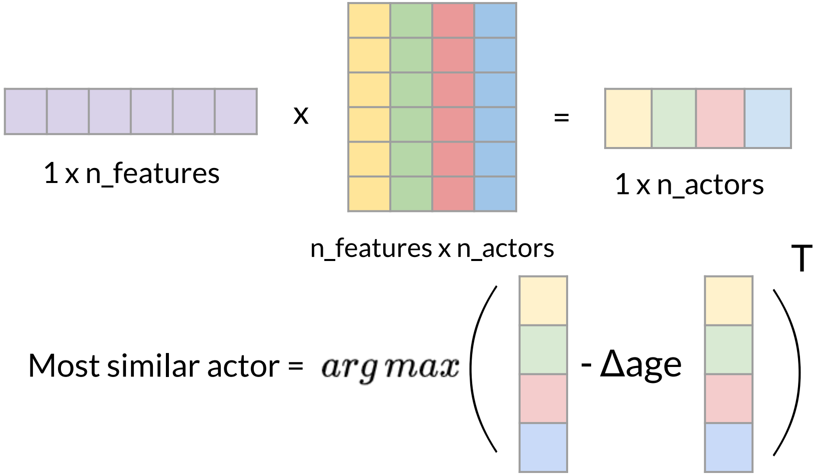
1. First we filter the actors by gender
2. For each actor in the dataset with the corresponding features *B* (retrieved using a LUT) we use a custom metric as a distance between *A* and *B*

$$S(A,B) = \lambda_1 \frac{A \cdot B}{\|A\|\|B\|} - \lambda_2 \frac{|age_A - age_B|}{max\_age}$$

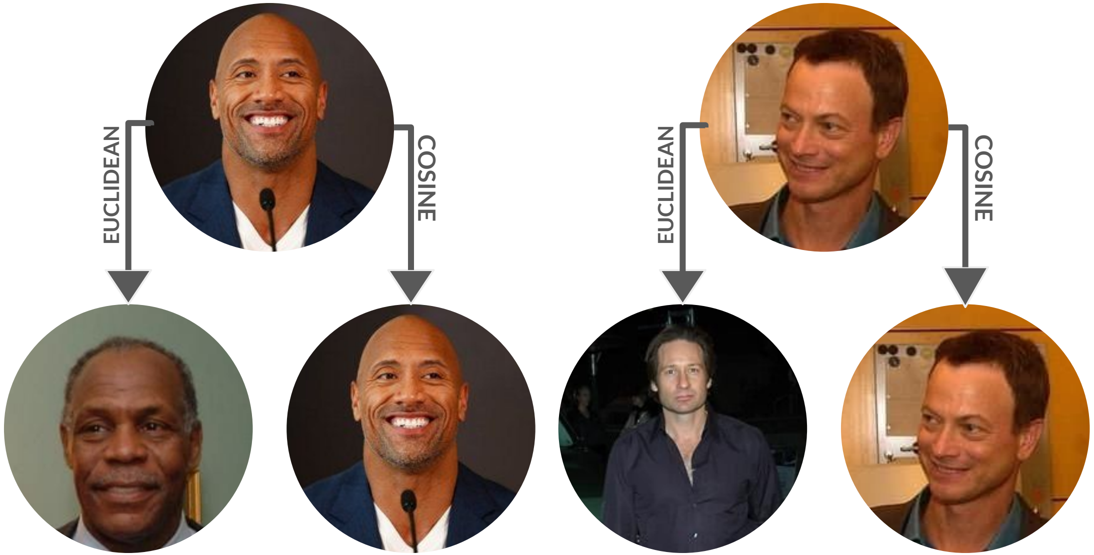
$$\lambda_1 = 7, \quad \lambda_2 = 1$$

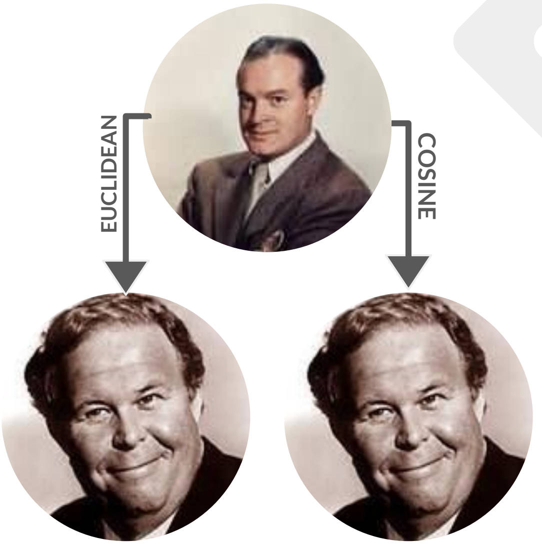
Idea: Compute the **cosine similarity** by matrix multiplication.

- Normalize the precomputed vectors, stack the features into a matrix, and save the transposed matrix
- During inference time, simply normalize the input vector and compute the matrix multiplication
- Significant speed-up:
  0.10347s -> 0.00055s (185x faster)



SIMILARITY – WHY NOT EUCLIDEAN DISTANCE?

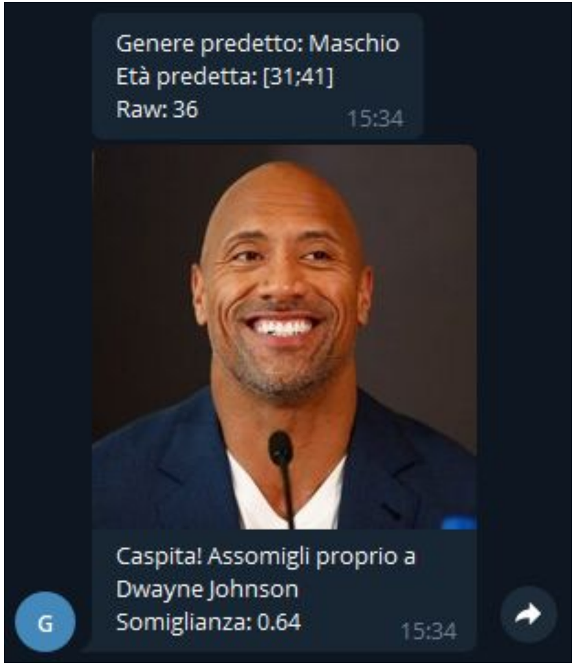




Both distances get some actors wrong!

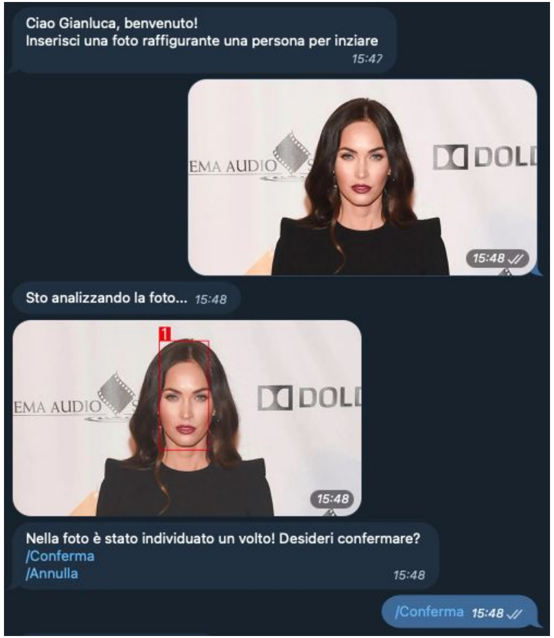
- In some experiments cosine similarity seems to make less errors than Euclidean
- Probably in the underlying feature space similar vectors have similar angles and not similar magnitude

SIMILARITY – SOME RESULTS



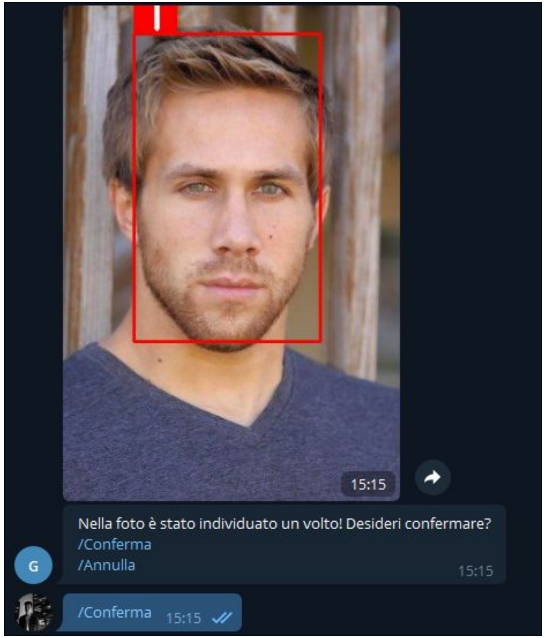
Actors in different images result in the same actor

SIMILARITY – SOME RESULTS



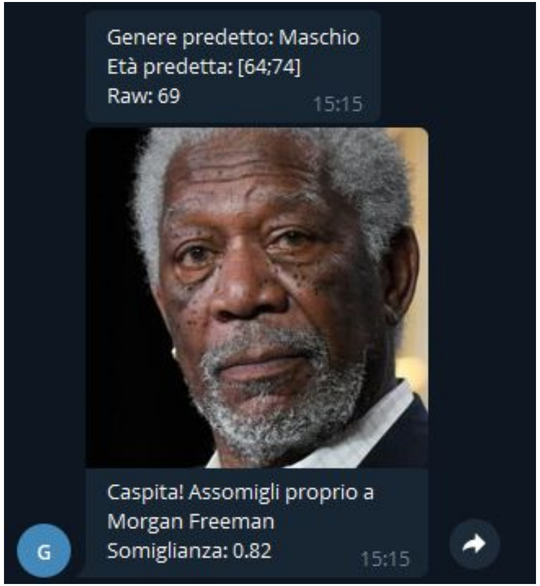
Actors in different images result in the same actor

SIMILARITY – SOME RESULTS



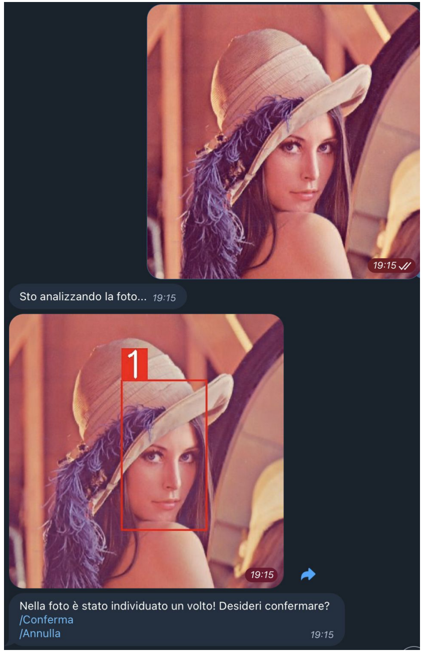
People similar to an actor result in that particular actor

SIMILARITY – SOME RESULTS



People similar to an actor result in that particular actor

SIMILARITY – SOME RESULTS



People similar to an actor result in that particular actor

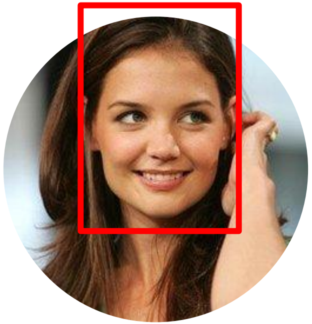
06.

DEMO & FUTURE WORKS

OBSERVATIONS AND FUTURE WORKS

- Similarity features on actors should be computed by using face detection
- Similarity depends too much on face orientation → faces should be rotated using key points
- Similarity depends too much on facial expressions
- Some choices for similarity were made without a proper validation metric. A possible metric for the task would be the elicitation of the average opinion from the users

STARRING



KATIE HOLMES

as Lidia Lucrezia Tonelli

MATT DAMON

as Luca Cogo

ZACH BRAFF

as Gianluca Giudice

Thank you for your attention