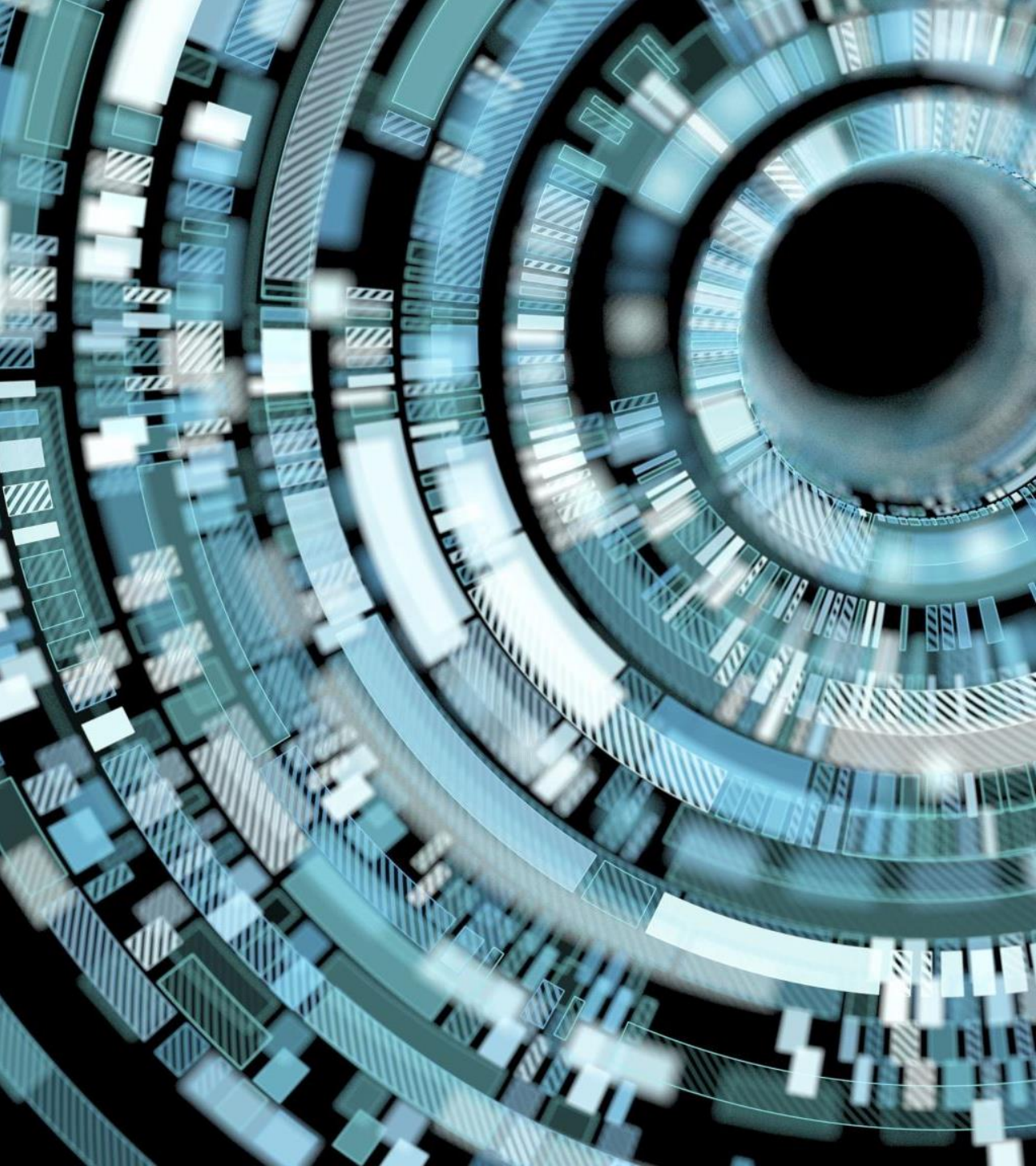


CONTEST VISIONE ARTIFICIALE

GRUPPO 6

**CAVALLARO VIRGINIA, COPPOLA DAVIDE,
NAPOLANO CARMINE, PANDOLFI GIULIANO**

2020/21



Introduzione

La stima dell'età dalle immagini dei volti è oggi una caratteristica rilevante in diverse applicazioni reali, come la segnaletica digitale, la robotica sociale, la business intelligence, il controllo degli accessi e la tutela della privacy.

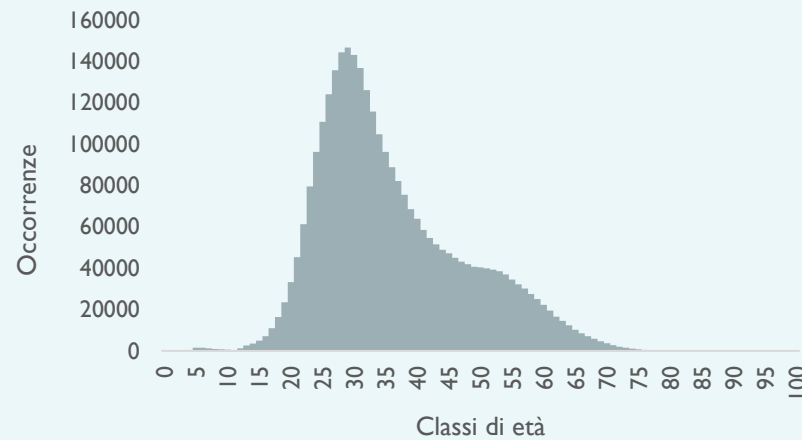
Lo scopo del progetto è stato addestrare una rete sul dataset fornito, tale che stimasse l'età di ogni soggetto presente nelle immagini.



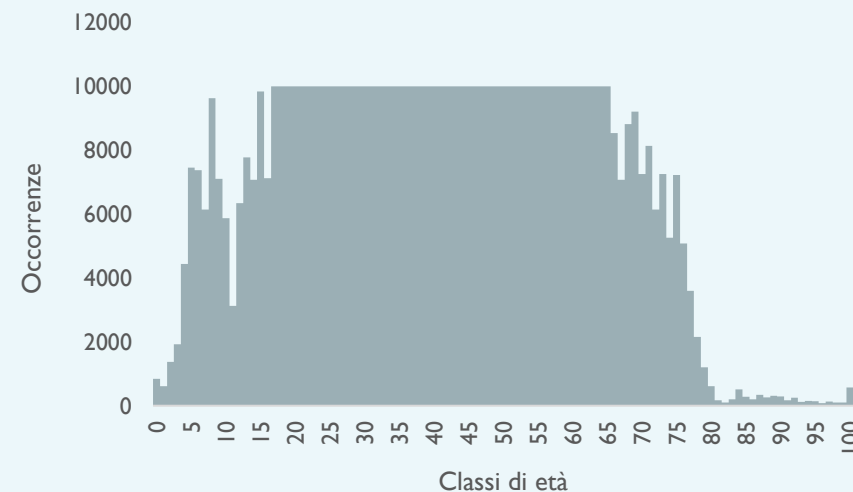
Convolutional Neural Network

- VGGFace
- Rete come REGRESSORE
- Funzione di Costo: MSE
- Metrica: MAE

Procedura di addestramento Dataset



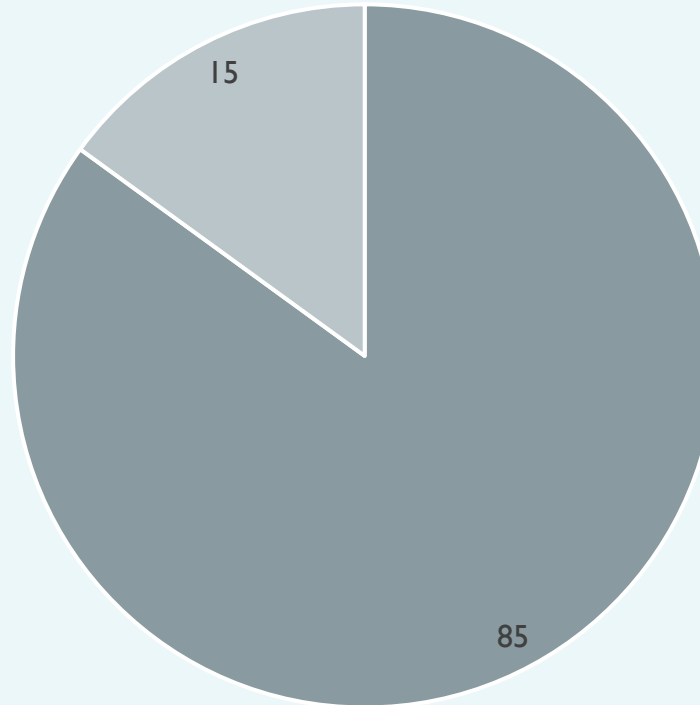
Il dataset originale è composto da più di tre milioni di samples, i quali sono divisi in modo *non uniforme* tra le classi di età.



Si è provveduto ad ottenere un dataset di dimensioni minori che fosse più bilanciato, per il quale ogni classe avesse un massimo di 10.000 samples.

Procedura di addestramento Dataset

PARTIZIONAMENTO DATASET



Il dataset finale è
costituito da 673.202
samples.

■ TRAIN ■ VALIDATION



Procedura di addestramento Data Augmentation

Lista Filtri

1	Flip orizzontale
2	Random Gaussian Noise
3	Random Brightness
4	Hue
5	Erosion
6	Shear
7	Salt and Pepper

Una prima forma di data augmentation è stata eseguita al fine di aumentare il numero di samples nelle classi con meno di 10.000 samples



Procedura di addestramento Data Augmentation

Lista Filtri

1	Flip orizzontale
2	Random Gaussian Noise
3	Random Brightness
4	Hue
5	Erosion
6	Shear
7	Salt and Pepper

- **Extreme:** Ogni sample è salvato altre 9 volte, applicando 9 filtri differenti (1, 2, 1+2, 1+3, 1+2+3, 2+4, 5, 6, 7).
- **High:** Ogni sample è salvato altre 4 volte, applicando 4 filtri differenti (1, 2, 1+2, 1+3).
- **Medium:** Ogni sample è salvato altre 2 volte, applicando 2 filtri differenti (1,2).
- **Basic:** Ogni sample è raddoppiato: viene salvata una copia sulla quale è applicato un filtro “base”(1).
- **Sporadica:** Per solo la metà dei samples viene creata una copia alla quale è applicata il filtro “base” (1).



Procedura di addestramento Data Augmentation

L'applicazione dell'augmentation è stata differenziata a seconda del numero di samples presenti in ogni classe

N. Samples	Classi	Tipo di Augmetation
Da 1 a 1000	0-4, 8-11, 75-100	Extreme
Da 1001 a 2000	5-7, 12, 73, 74	High
Da 2001 a 3000	13, 71, 72	Medium
Da 3001 a 5000	14, 15, 69, 70	Basic
Da 5001 a 6000	68	Sporadica

Procedura di addestramento Face Pre-processing

- Ritaglio attorno al volto principale
- Ridimensionamento 124x124px
- Sharpen

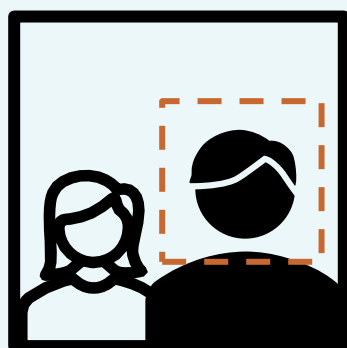


Immagine
originale



Ritaglio volto
principale



Ridimensionamento
124x124 px



Sharpen

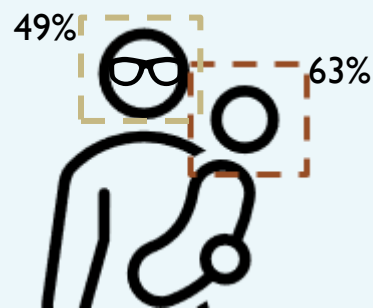
Procedura di addestramento Face Detection

Come Face-Detector è stata utilizzata una rete neurale profonda di OpenCV:

la **res10_300x300_ssd_iter_140000_fp16.caffemodel**, basata su Caffe, un deep learning framework, sviluppato da Berkeley AI Research (BAIR).

I bounding box considerati sono unicamente quelli con confidence maggiore o uguale al 50%, ove non vi fossero bounding box con questa confidence, l'immagine non è stata inserita nel dataset.

Nonostante ciò, durante una revisione del dataset, ci si è accorti che, in alcuni casi isolati, la detection non sia stata accurata (scegliendo, ad esempio, parte del volto o un volto secondario), causando un errore nella stima dell'età.





Procedura di addestramento

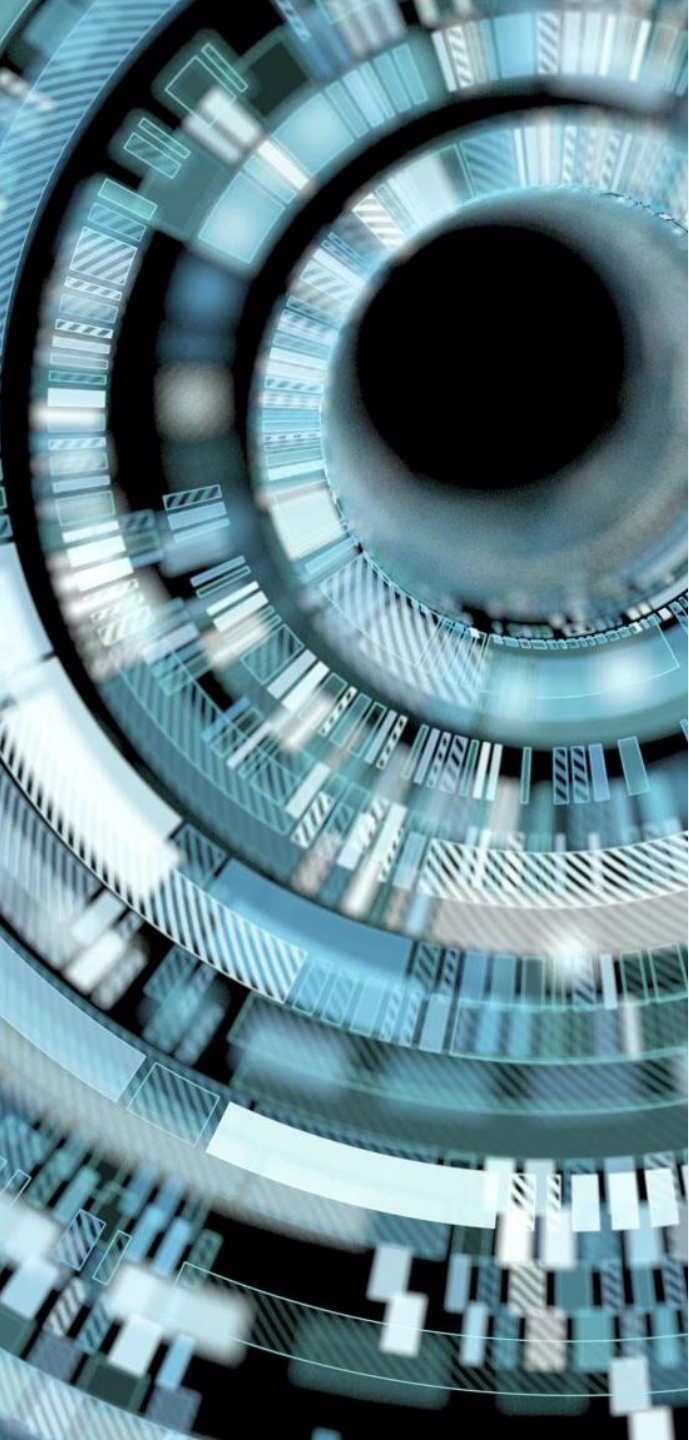
Training from Scratch o Fine-Tuning

Sono stati aggiunti quattro layers per il **fine-tuning**:

- *Flatten*
- *Dense + ReLu*
- *Dense + ReLu*
- *Dense + ReLu*

Il tipo di Pooling utilizzato per il modello è stato *average pooling*

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 124, 124, 3)]	0
conv1_1 (Conv2D)	(None, 124, 124, 64)	1792
conv1_2 (Conv2D)	(None, 124, 124, 64)	36928
pool1 (MaxPooling2D)	(None, 62, 62, 64)	0
conv2_1 (Conv2D)	(None, 62, 62, 128)	73856
conv2_2 (Conv2D)	(None, 62, 62, 128)	147584
pool2 (MaxPooling2D)	(None, 31, 31, 128)	0
conv3_1 (Conv2D)	(None, 31, 31, 256)	295168
conv3_2 (Conv2D)	(None, 31, 31, 256)	590080
conv3_3 (Conv2D)	(None, 31, 31, 256)	590080
pool3 (MaxPooling2D)	(None, 15, 15, 256)	0
conv4_1 (Conv2D)	(None, 15, 15, 512)	1180160
conv4_2 (Conv2D)	(None, 15, 15, 512)	2359808
conv4_3 (Conv2D)	(None, 15, 15, 512)	2359808
pool4 (MaxPooling2D)	(None, 7, 7, 512)	0
conv5_1 (Conv2D)	(None, 7, 7, 512)	2359808
conv5_2 (Conv2D)	(None, 7, 7, 512)	2359808
conv5_3 (Conv2D)	(None, 7, 7, 512)	2359808
pool5 (MaxPooling2D)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
fc6 (Dense)	(None, 1024)	4719616
fc7 (Dense)	(None, 512)	524800
fc8 (Dense)	(None, 1)	513
Total params: 19,959,617		
Trainable params: 5,244,929		
Non-trainable params: 14,714,688		



Procedura di addestramento

Procedura di training

Epoche di addestramento: 15 (con Early Stopping)

Optimizer: rmsprop.

Rmsprop ha un momentum standard di 0.9.

I valori sono stati scelti
manualmente a seguito di varie
sessioni di addestramento.

LR_schedule:

- initial_learning_rate: 0.001
- decay_steps: 20
- decay_rate: 0.96
- staircase: True

Early Stopping:

- monitor: val_loss
- mode: min
- patience: 3
- min_delta: 0.01

Model checkpoint:

- monitor: val_loss
- mode: min
- save_best_only: True

ReduceLROnPlateau:

- monitor: val_loss
- mode: min
- patience: 2
- min_delta: 0.02
- factor: 0.5



Risultati Sperimentali

Model	Function	Optimizer	Loss	Val_Loss	Metrics	Val_metrics
VGGFace	Regressor	Adam	MSE: 10.35	MSE: 39.85	MAE: 2.41	MAE: 4.07
VGGFace	Regressor	Rmsprop	MAE: 2.36	MAE: 3.95	MSE: 14.07	MSE: 38.95
VGGFace	Regressor	Rmsprop	MSE: 10.80	MSE: 40.03	MAE: 2.80	MAE: 4.15
VGGFace + data augmentation	Regressor	Rmsprop	MSE: 13.51	MSE: 40.24	MAE: 2.59	MAE: 4.09
VGGFace	Classifier	SGD	CCE: 4.34	CCE: 4.93	AUC: 0.72	AUC: 0.65

Inoltre, è stato costruito un test set non etichettato di pochi samples per controllare l'età stimata sul singolo soggetto.

La rete scelta è quella che si è comportata meglio nella stima dell'età di bambini e anziani.

An abstract graphic on the left side of the slide, featuring concentric circles and a complex pattern of blue and white lines, resembling a stylized eye or a data visualization.

Conclusioni

Il primo addestramento effettuato sul dataset originale era influenzato dalla poca rappresentatività delle classi più giovani e anziane, ottenendo come output esclusivamente età di cui si avevano più samples (>100.000).

Con l'ausilio della tecnica del data augmentation siamo riusciti ad ottenere da un numero ristretto di dati, una varietà di samples appartenenti alla stessa classe. Grazie ai filtri applicati, ogni sample ha dato origine a nuove immagini che hanno arricchito la rappresentatività della classe.

A seguito del bilanciamento del dataset, come ci si aspettava, sono fortemente migliorate le performances del modello.

La scelta di operare fine-tuning su VGGFace ha influenzato considerevolmente la capacità di stima delle età del nostro regressore.