

# Digital Signal & Image Management

---

CARLONE A. - 726894

MIORI D. - 813692

ONDEI A. - 826399

# Audio Recognition



GOAL: speaker identification

## Feature Extraction



## Data Collection

- Acquisito un totale di 15 minuti per componente del gruppo in 3 tranches differenti
- I componenti leggevano ogni volta un capitolo di un libro diverso
- Sampling di 2 secondi così da ottiene un totale di 1350 tracce



## Modeling



## Evaluation

# Feature Extraction

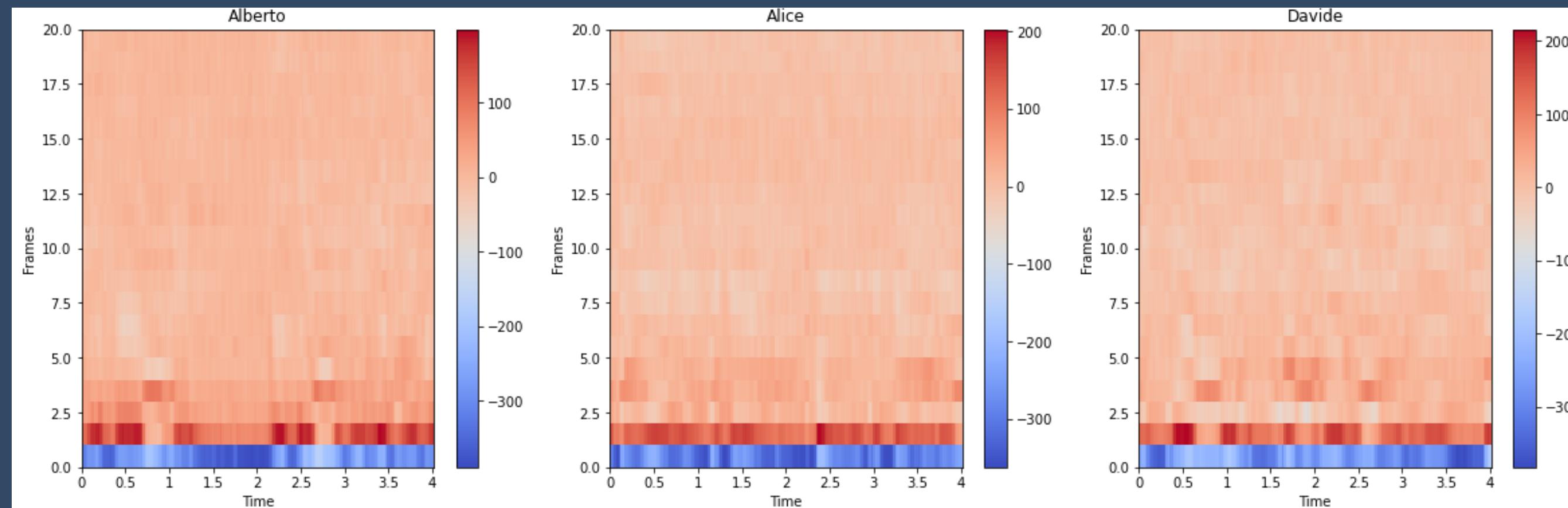


# Evaluation

## Data Collection

## Modeling

- MFCC (Mel-frequency cepstrum coefficients)
- I coefficienti vengono rappresentati come un'immagine per poter essere utilizzati in una CNN



## Feature Extraction



## Evaluation



## Data Collection

## Modeling

- La CNN è composta da 3 layer convoluzionali, 2 di max pooling e 2 dense, allenati su 50 epoch

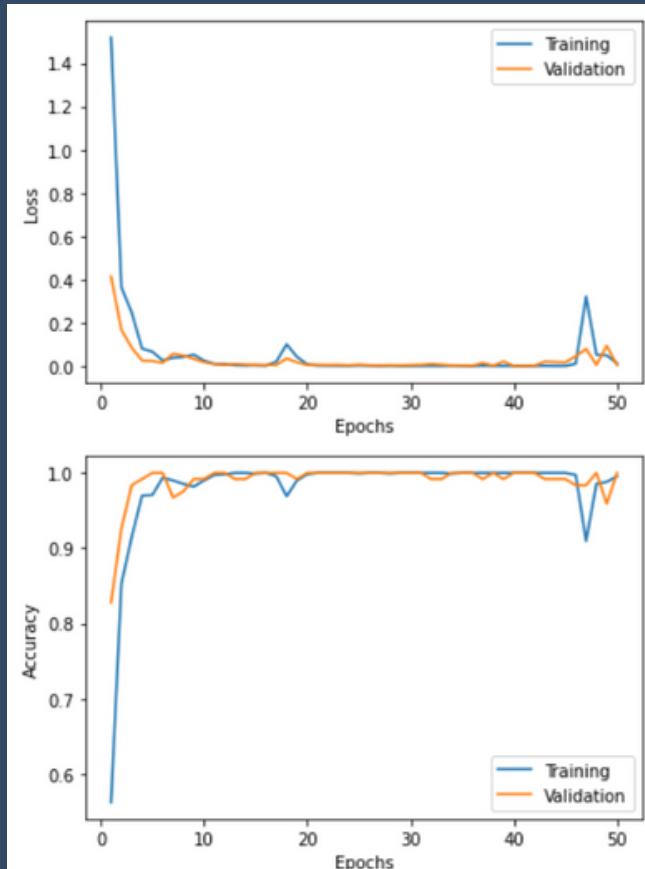


## Feature Extraction

## Evaluation

### Data Collection

### Modeling



	precision	recall	f1-score	support
0	1.00	1.00	1.00	51
1	1.00	1.00	1.00	44
2	1.00	1.00	1.00	40
micro avg	1.00	1.00	1.00	135
macro avg	1.00	1.00	1.00	135
weighted avg	1.00	1.00	1.00	135
samples avg	1.00	1.00	1.00	135

- Le prestazioni sono state monitorate sia in fase di addestramento osservando i plot della loss e dell'accuracy, sia tramite un classification report effettuato sul test set

# Face Recognition



## Pre - processing



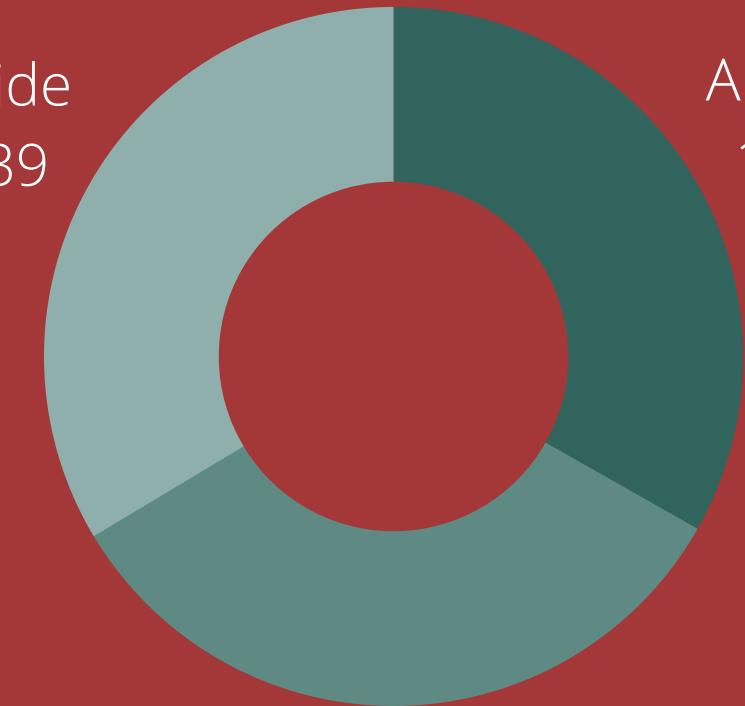
## Evaluation

### Dataset

- Due raccolte di immagini: scatti in università & scatti autonomi
- Si ottengono un totale di 4581 immagini, suddivise equamente nelle 3 classi.
- Le immagini presentano condizioni differente di luce e rappresentazione del volto.

### Modeling

Davide  
1539



Alberto  
1521

Alice  
1521



# Pre - processing

# Evaluation



## Dataset

## Modeling

- Pre - processing in analogia alla rete scelta (VGG16)
- All'interno del data generator vengono inseriti i parametri per la data augmentation, nello specifico si aggiunge il flip orizzontale, shift di altezza e lunghezza, un range per la luminosità, lo zoom e la rotazione

## Pre - processing

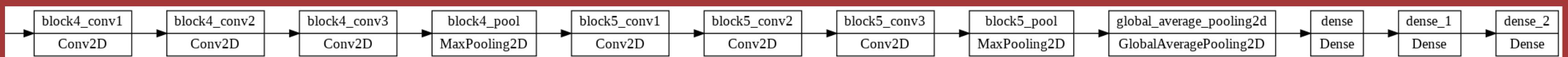
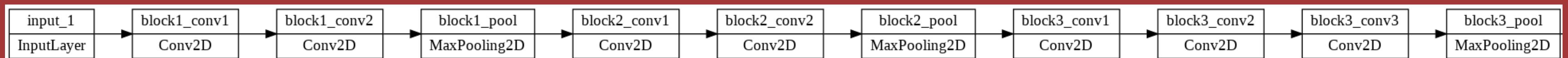
## Evaluation



### Dataset

### Modeling

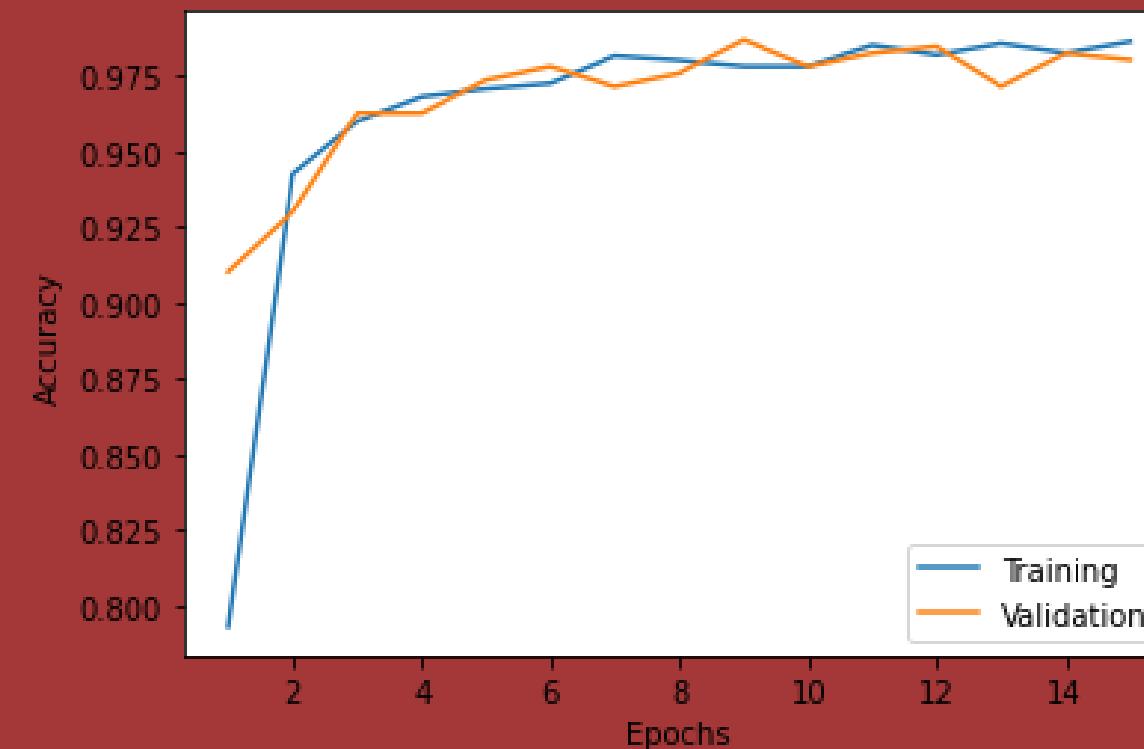
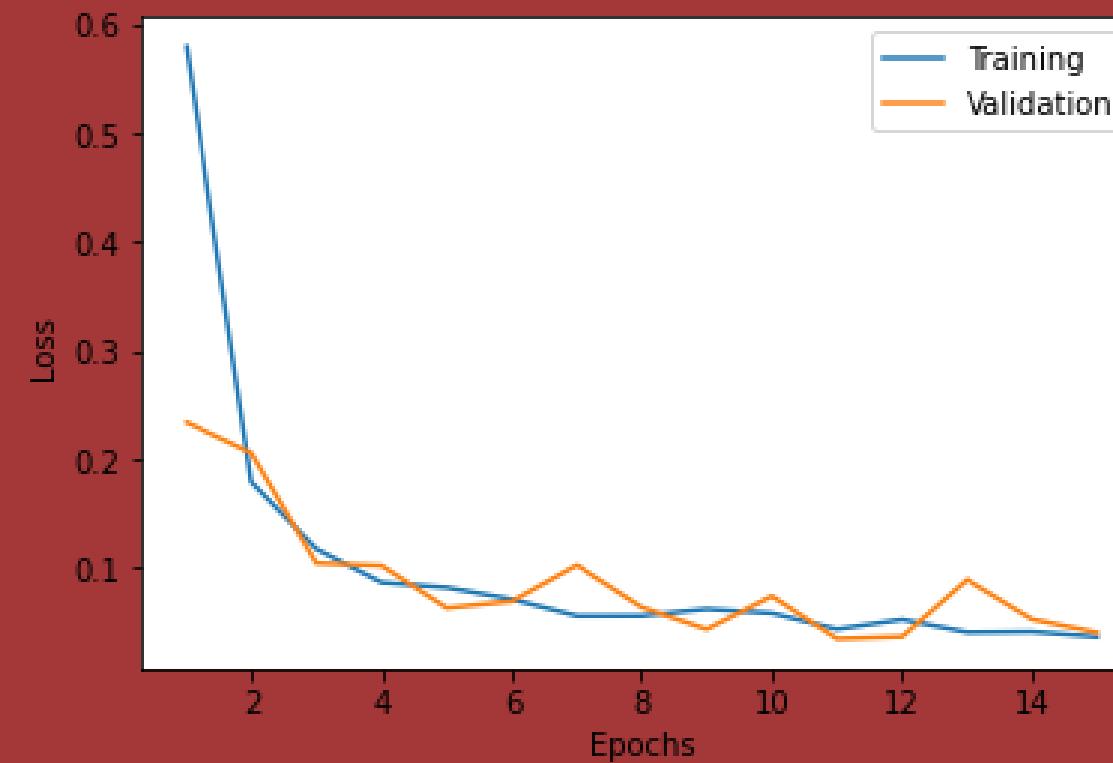
- La rete pre - addestrata scelta è VGG16, la rete viene tagliata prima dei layers fully connected
- Vengono aggiunti due layer densi e uno di output
- Le funzioni di attivazione utilizzate sono ReLu e SoftMax
- Funzione di Loss: categorical crossentropy
- Ottimizzatore: Adam
- Numero Epoche: 15



## Pre - processing



### Dataset



## Evaluation



### Modeling

### Classification Report

	Precision	Recall	F1 - Score
Alberto	0,97	1,00	0,98
Alice	1,00	0,98	0,99
Davide	0,99	0,99	0,99

accuracy 0.99

# Image Retrieval



## Evaluation: Confronto

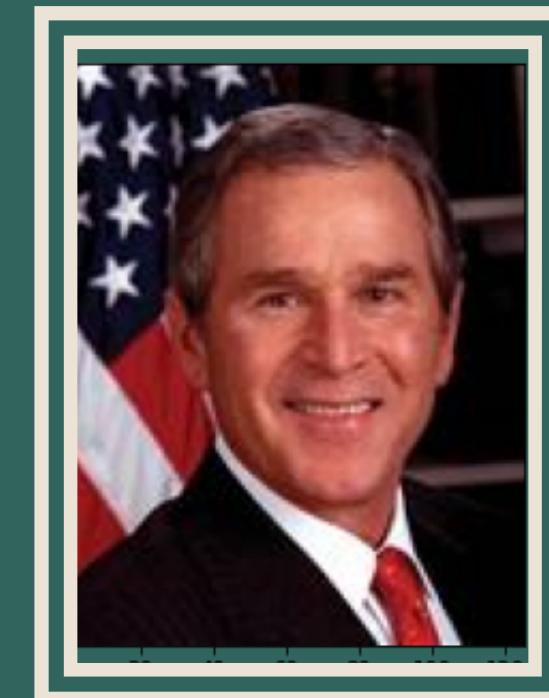
## Retrieval



## Modelli e Test

- MobileNet con Imagenet
  - VGG16 con VGGFace
  - ResNet50 con VGGFace
- 
- Creazione Test Set: estrazione di circa 400 classi per un totale di circa 6000 immagini

## Evaluation: Metriche

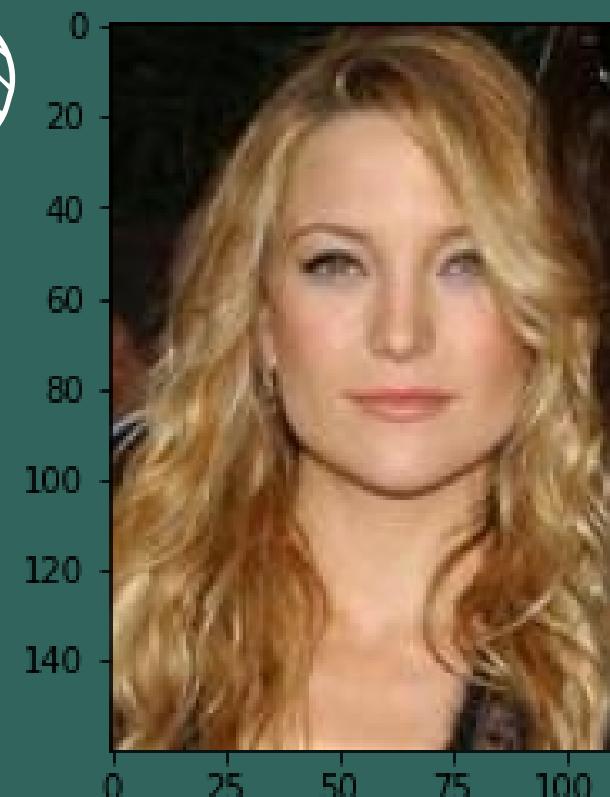


# Evaluation: Confronto

Retrieval



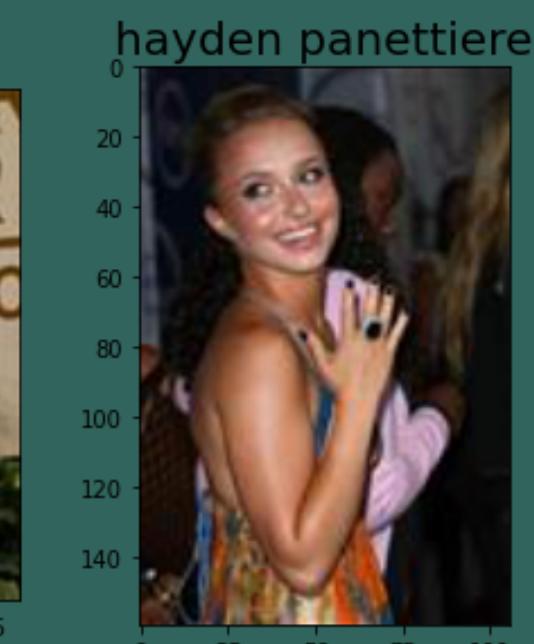
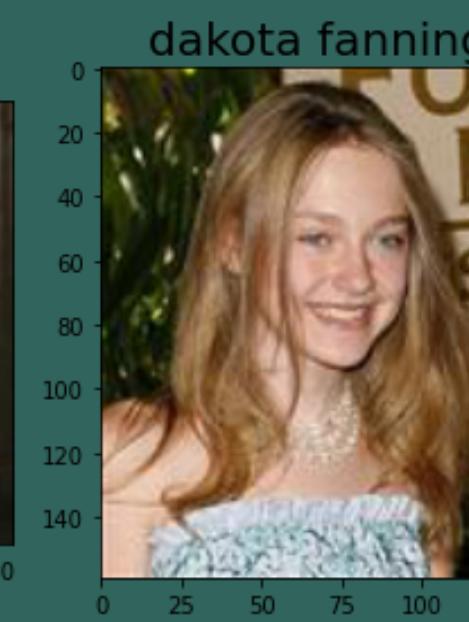
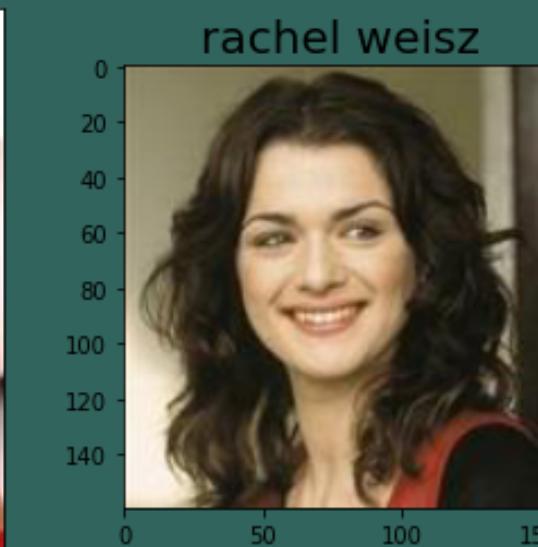
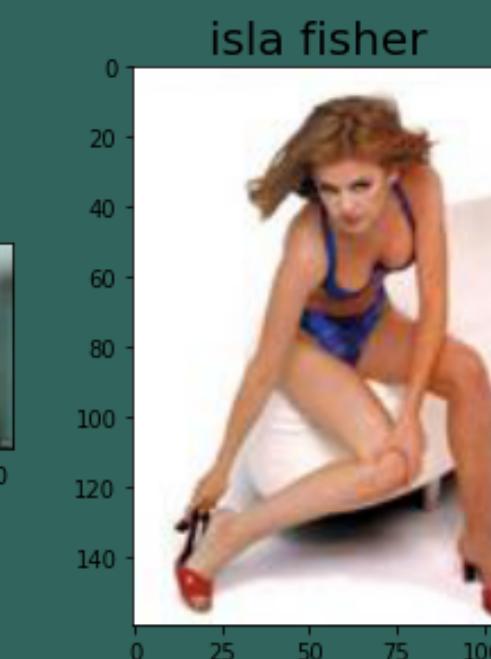
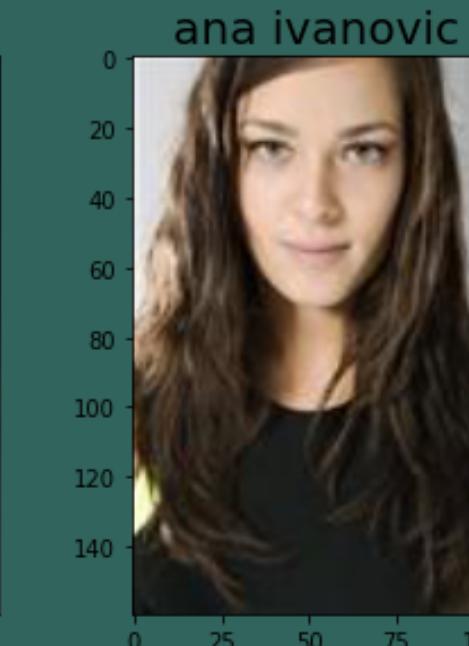
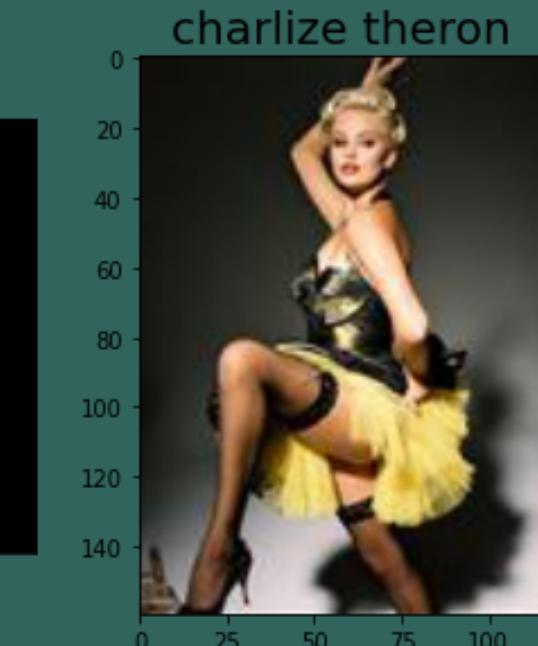
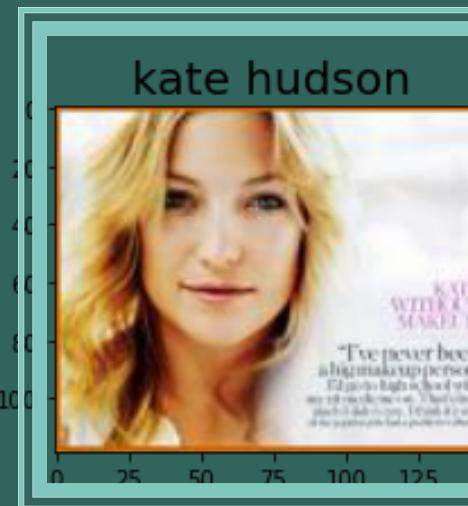
Modelli e Test



Evaluation: Metriche



10 volti più simili con MobileNet

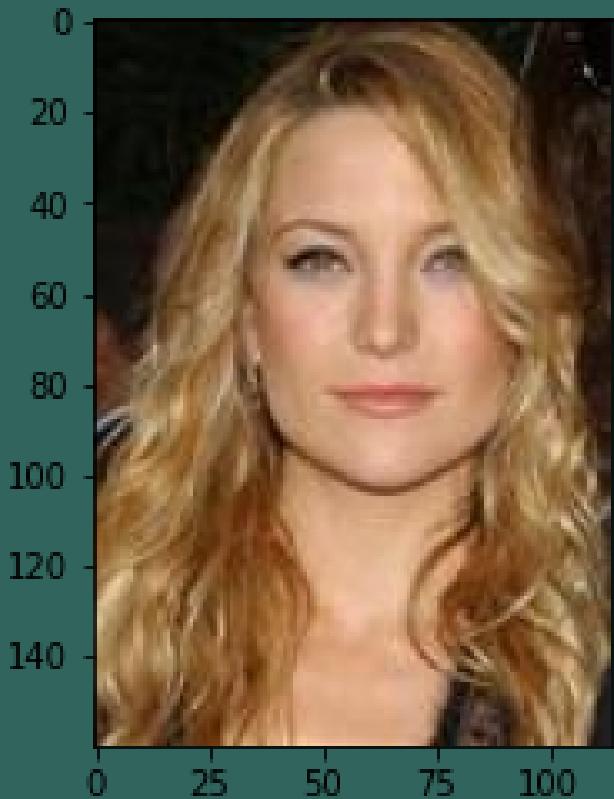


# Evaluation: Confronto

Retrieval



Modelli e Test



## Evaluation: Metriche

10 volti più simili con VGGFace-vgg16

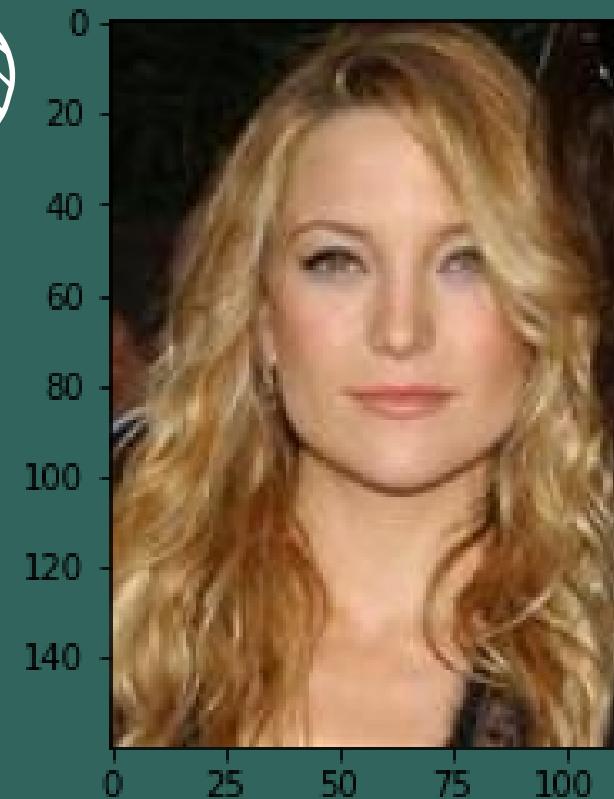


# Evaluation: Confronto

Retrieval



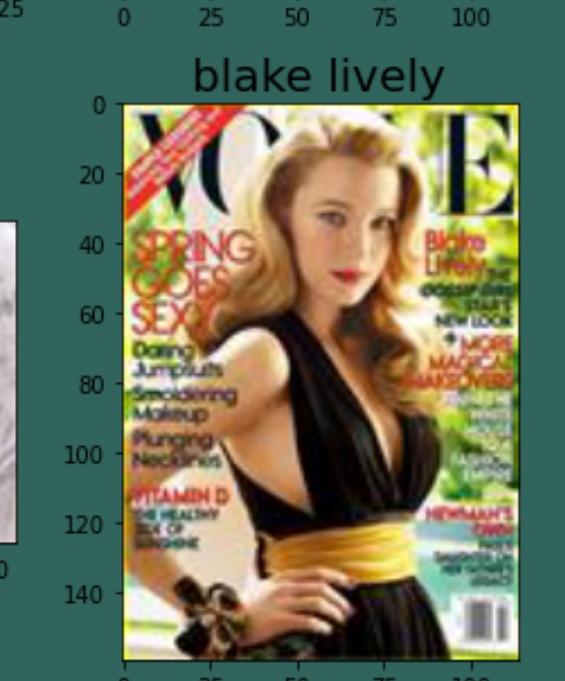
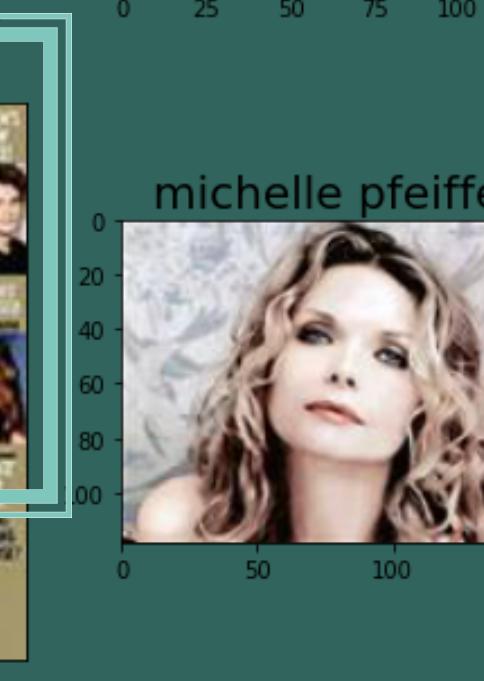
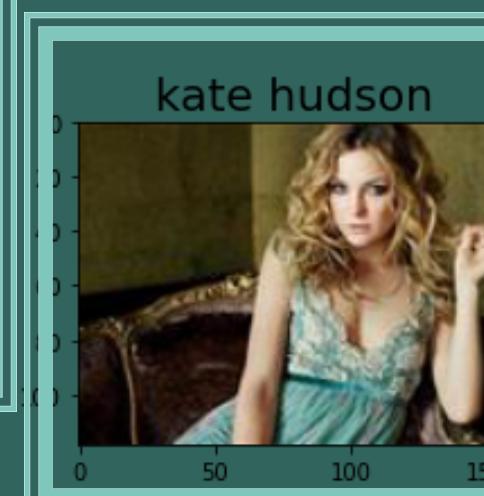
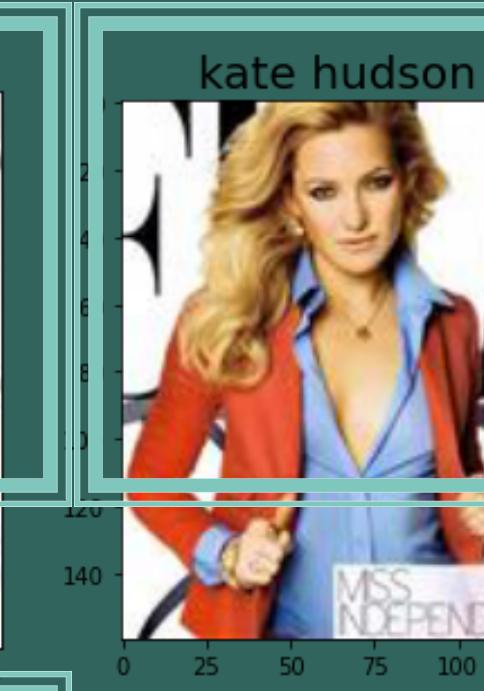
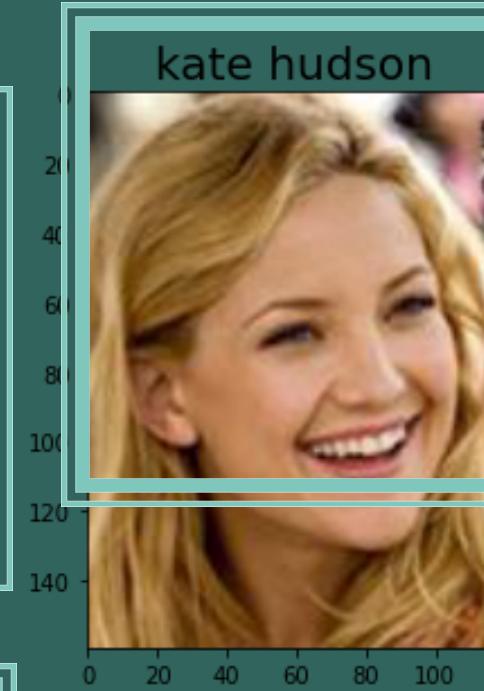
Modelli e Test



## Evaluation: Metriche



10 volti più simili con VGGFace-resnet



## Evaluation: Confronto

Modelli e Test

- Precision = n° immagini rilevanti selezionate / n° immagini selezionate
- Recall = n° immagini rilevanti selezionate / n° immagini rilevanti totale
- F1 Score = media armonica tra Precision e Recall

## Retrieval

Evaluation: Metriche

Risultati

	MobileNet	VGG16	ResNet50
Recall	0,09	0,21	0,37
Precision	0,13	0,29	0,52
F1 Score	0,11	0,25	0,44

## Evaluation: Confronto

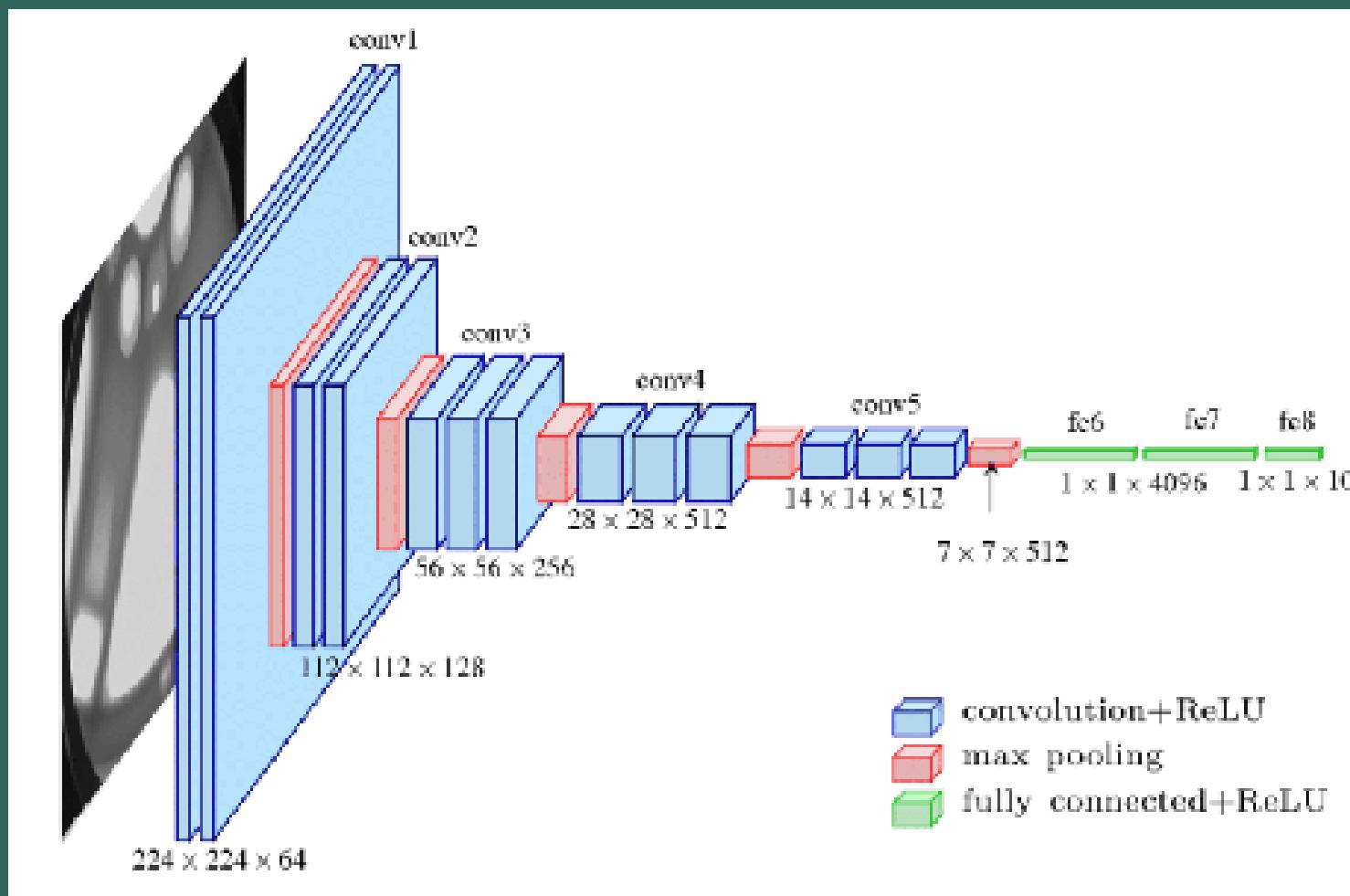
## Retrieval



Modelli e Test



Evaluation: Metriche



Il modello migliore risulta essere ResNet50 da VGGFace

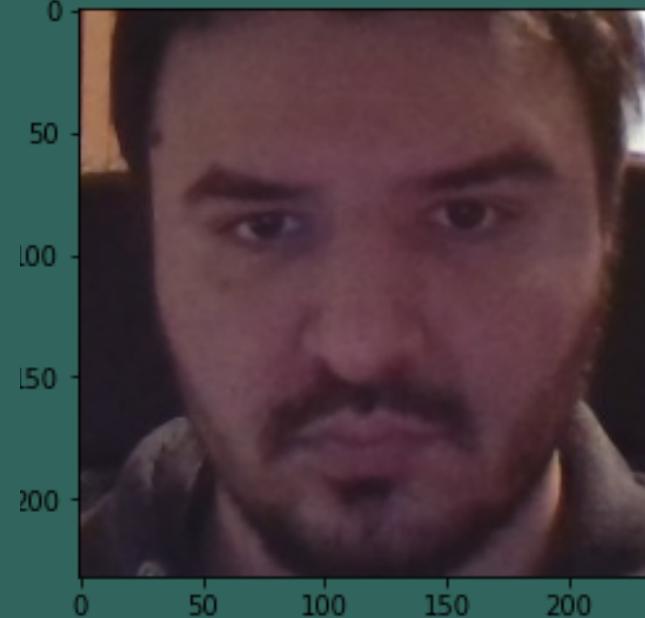
Il Retrieval finale è stato effettuato su un pool di circa 24000 immagini (15 per classe)

# Evaluation: Confronto

# Retrieval



Modelli e Test



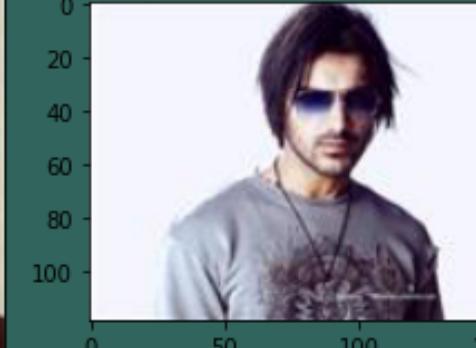
Evaluation: Metriche

10 volti più simili ad Alberto con VGGFace-ResNet50

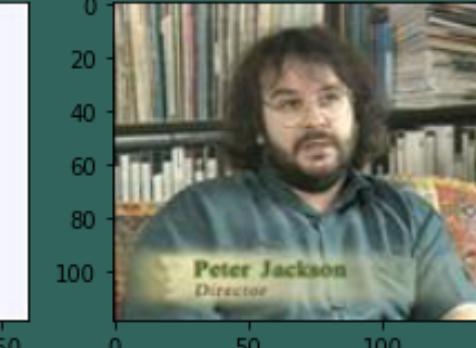
john belushi



john abraham



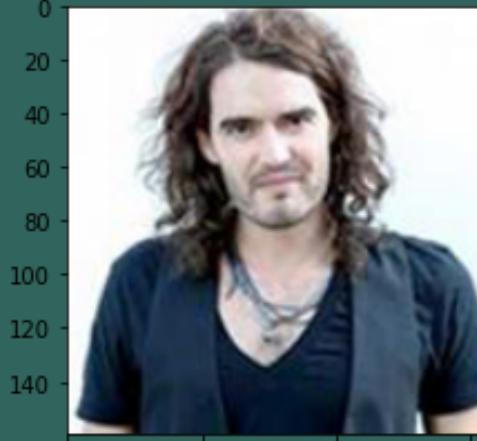
peter jackson



arielle kebbel



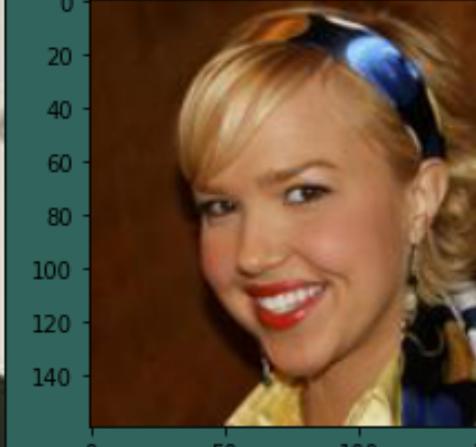
russell brand



jeremy piven



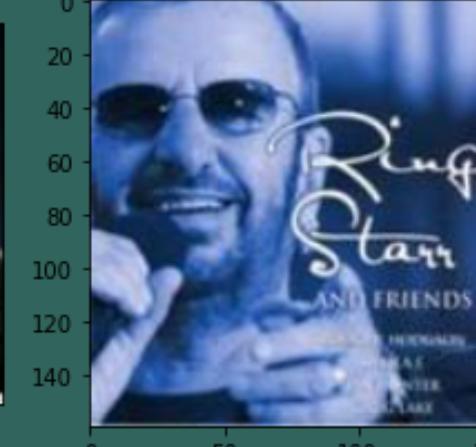
arielle kebbel



steve jobs



ringo starr



ringo starr



**Thank you**