University of Milano-Bicocca
Master's Degree in Data Science
Digital Signal and Image Management
Academic Year 2022-2023

Facial Image Retrieval on CelebA dataset

Author:
Gianluca SCURI
Niccolò PUCCINELLI

Task: Image Retrieval

- **Content-based image retrieval**: Generation of an ordered list of images related to a query image.
- **Applications**: Surveillance, medical imaging, database research, access control, ...
- Our task: Identification of similarities between the faces of the previous actors and the faces of various celebrities, contained in **CelebA** dataset [3]
- Lack of a proper evaluation metric

[3] Liu, Ziwei and Luo, Ping and Wang, Xiaogang and Tang, Xiaoou (2015) Deep Learning Face Attributes in the Wild

Dataset: CelebA

- **Images**:
  - 202599 face images of various celebrities
  - 10177 unique identities (names not given)
  - Different backgrounds
  - Standard dimensions: 218x178
  - **40 binary facial attribute annotations** per image (e.g. hair color, beard/no beard)
- Binary facial attributes annotations used as **evaluation metric**
  - Images of the 24 actors (RAVDESS) labelled by hand, according to the same binary attributes
  - For each image retrieved: ratio between matching attributes and total number of attributes
  - Sort of **"empirical" accuracy**

Image retrieval: Input processing

Actors

- **Face extraction** as for video processing (Haar Cascade classifier)
- Output shape: 224x224
- Changed background with a **custom** one
  - Avoiding bias due to the retrieval of images with white background
  - Results improvement
- Image values between 0 and 255
- Further processing according to **MobileNetV2** and **ResNet50V2**

Celebrities

- **Face extraction** as for video processing (Haar Cascade classifier)
  - Face celebrities not found deleted from the dataset (5210/202599)
- Output shape: 224x224
- Image values between 0 and 255
- Further processing according to **MobileNetV2** and **ResNet50V2**

Image retrieval: Models

- **Extracting features** by means of 2 different architectures:
  - **MobileNetV2** (3.5M parameters) – 1280 features
  - **ResNet50V2** (25.6M parameters) – 2048 features
- Stop before the final layer: **retrieving an internal feature representation** instead of the final probability distribution
- **KDTree** from scikit-learn in order to build the search tree
  - Euclidean distance
- Query the tree and retrieve top-k similar images based on distance
- Evaluate retrieval through “empirical” accuracy given by facial attributes annotations

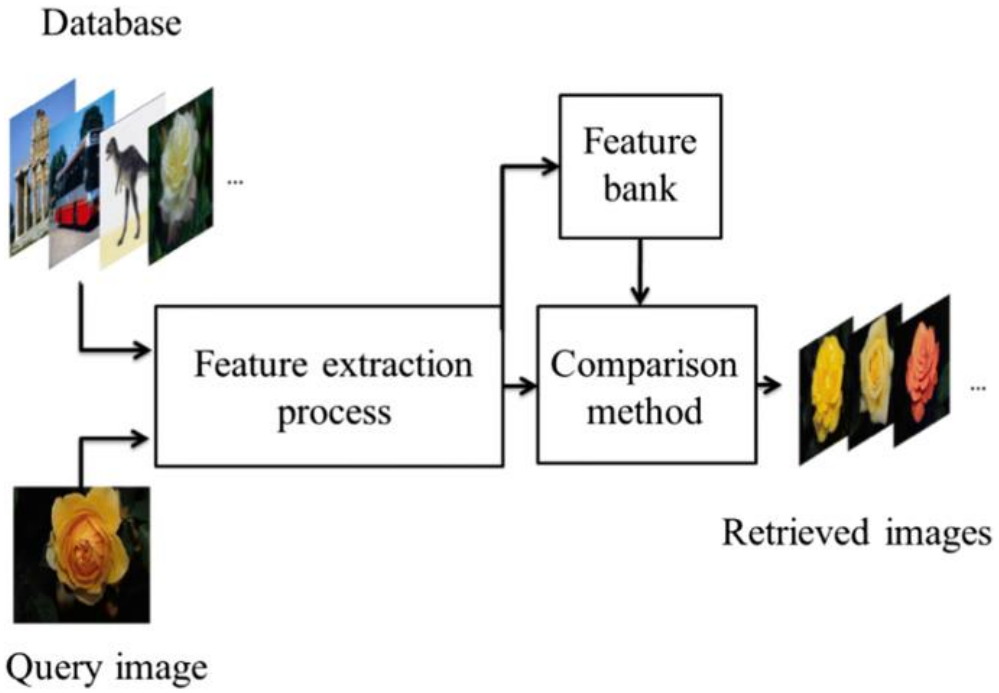
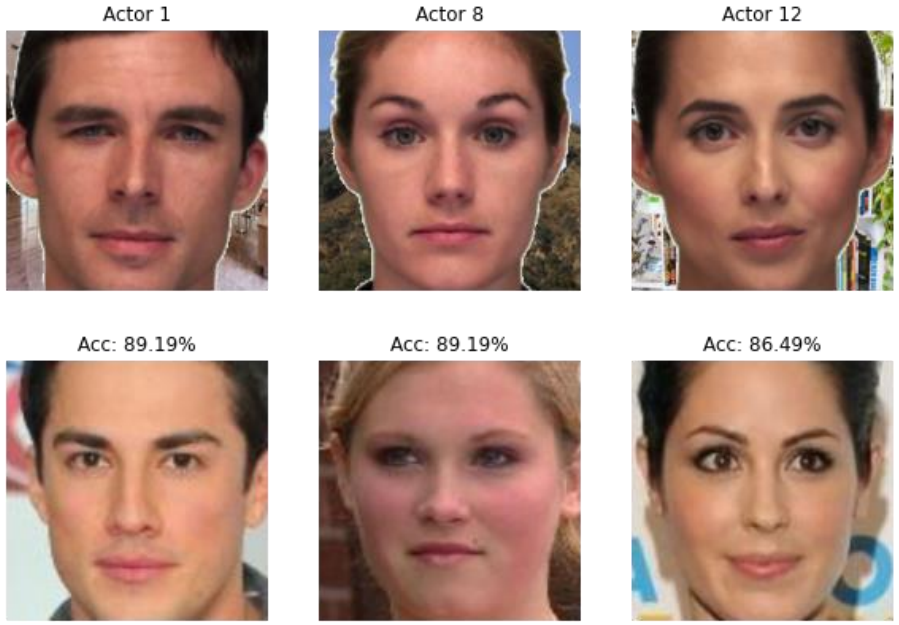


Image retrieval: Results

MobileNetV2

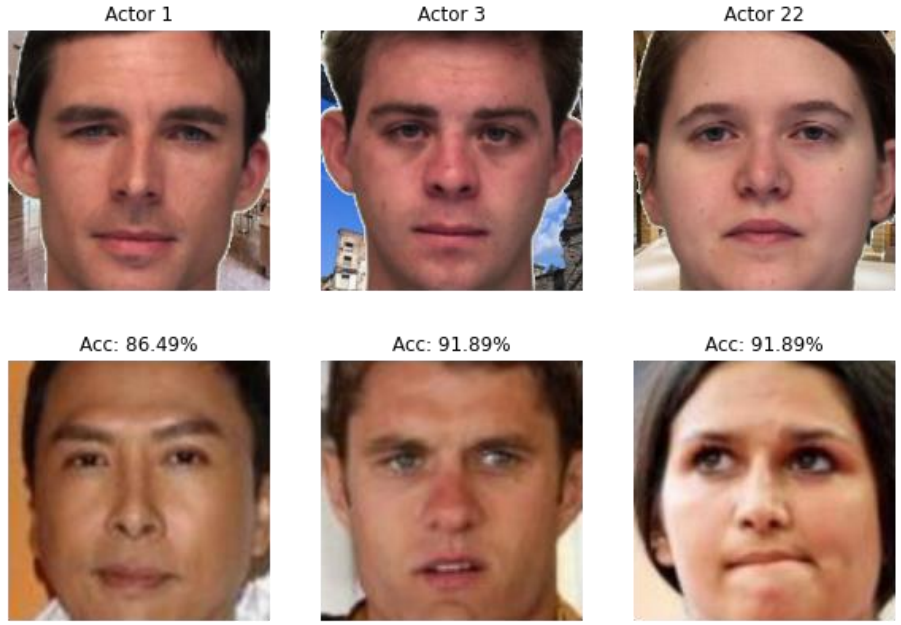
- Mean top-2 “accuracy” for 24 actors (top-2 images): **81.08%**



Top-3 actor results for MobileNetV2

ResNet50V2

- Mean top-2 “accuracy” for 24 actors (top-2 images): **80.86%**



Top-3 actor results for ResNet50V2

Image retrieval: Demo