The VoxCeleb Speaker Recognition Challenge (VoxSRC) Workshop 2019

Joon Son Chung and Andrew Zisserman
Introduction

• Creation of the VoxCeleb dataset

• Overview of the speaker recognition challenge
Datasets: VoxCeleb2

*a large-scale audio-visual dataset of human speech*

150,000+ YouTube videos

of

7000+ different celebrity speakers

1 million+ utterances

2000+ hours of video

Clips from the same identity
YouTube videos are a great source

- Multi-speaker environments
- Varying audio quality and background channel noise
- Freely available
Aim: Automatically obtain audio segments of speakers from videos uploaded to YouTube

To do this we need to solve the following:

- **When is a person speaking?**  
  - Done using Active Speaker Verification (ASV)

- **Which speaker is the celebrity that we want?**  
  - Done using Face Verification
Fully Automated Pipeline

Audio feature extraction → Face detection and Tracking → Active Speaker Verification

When is a person speaking?

Felicity Jones

Face Verification
Who is the speaker?

match

VOXCELEB

match

Felicity Jones
Fully Automated Pipeline

1. Candidate List

Felicity Jones

Face Verification

match

VOXCELEB

Active Speaker Verification

match

Audio feature extraction

Face detection and Tracking

Download videos Face detection
Face tracking Audio feature extraction
Active speaker verification Face verification VoxCeleb database
Elon Musk
Celebrities are the ideal choice – many ‘interview’ videos. Many identities, ranging from actors and sportspeople to entrepreneurs.

- 7000+ identities, from actors and sportspeople to entrepreneurs.

**1. Candidate List**

<table>
<thead>
<tr>
<th>A.J. Buckley</th>
<th>Alex Trebek</th>
<th>Andy Samberg</th>
<th>Ben Whishaw</th>
<th>Bruno Ganz</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.R. Rahman</td>
<td>Alexa Davalos</td>
<td>Anwar Barnard</td>
<td>Beth Grant</td>
<td>Bruno Mars</td>
</tr>
<tr>
<td>Aamir Khan</td>
<td>Alexander Siddig</td>
<td>Angela Lee</td>
<td>Bethany Mota</td>
<td>Hurt Reynolds</td>
</tr>
<tr>
<td>Aaron Tveit</td>
<td>Alexandra Daddario</td>
<td>Angela Kinsey</td>
<td>Betty White</td>
<td>CCH Pounder</td>
</tr>
<tr>
<td>Aaron Yoo</td>
<td>Alexandra Roach</td>
<td>Anne Hathaway</td>
<td>Bill Nighy</td>
<td>Caitriona Balfe</td>
</tr>
<tr>
<td>Abbie Cornish</td>
<td>Alex Johnson</td>
<td>Ansel Elgort</td>
<td>Bill Pullman</td>
<td>Callan McAuliffe</td>
</tr>
<tr>
<td>Abigail Breslin</td>
<td>Alfre Woodard</td>
<td>Anthony Anderson</td>
<td>Billie Joe Armstrong</td>
<td>Callie Thorne</td>
</tr>
<tr>
<td>Abigail Spencer</td>
<td>Alice Eve</td>
<td>Anthony Mackie</td>
<td>Binging Li</td>
<td>Cameron Boyce</td>
</tr>
<tr>
<td>Adam Beach</td>
<td>Alija Bachleda</td>
<td>Anthony Rapp</td>
<td>Blair Underwood</td>
<td>Candace Cameron Bure</td>
</tr>
<tr>
<td>Adam Brody</td>
<td>Alison Azero</td>
<td>Anton Yelchin</td>
<td>Blake Michael</td>
<td>Candide Accola</td>
</tr>
<tr>
<td>Adam Copeland</td>
<td>Allison Pill</td>
<td>Antonio Cupo</td>
<td>Blake Shelton</td>
<td>Candide Patton</td>
</tr>
<tr>
<td>Adam Driver</td>
<td>Allison Williams</td>
<td>Arden Cho</td>
<td>Bob Barker</td>
<td>Caroline Rhea</td>
</tr>
<tr>
<td>Adrienne Curry</td>
<td>Amanda Seyfried</td>
<td>Armand Hammer</td>
<td>Bobby Cannavale</td>
<td>Carolyn Hennesy</td>
</tr>
<tr>
<td>Adrienne Palicki</td>
<td>Amazay Nolacco</td>
<td>Asa Butterfield</td>
<td>Bonnie Wright</td>
<td>Carrie Ann Inaba</td>
</tr>
<tr>
<td>Agness Deyn</td>
<td>Amber Riley</td>
<td>Ashley Greene</td>
<td>BooBoo Stewart</td>
<td>Carrie Fisher</td>
</tr>
<tr>
<td>Aidan Turner</td>
<td>America Ferrera</td>
<td>Ashley Jensen</td>
<td>Brad Paisley</td>
<td>Carrie Underwood</td>
</tr>
<tr>
<td>Ajay Devgn</td>
<td>Anil Nigam</td>
<td>Audra McDonald</td>
<td>Bradley Steven Perry</td>
<td>Casey Affleck</td>
</tr>
<tr>
<td>Akshay Kumar</td>
<td>Amanda Peet</td>
<td>Audrina Patridge</td>
<td>Brenda Blethyn</td>
<td>Casey Wilson</td>
</tr>
<tr>
<td>Alain Delon</td>
<td>Ana Gasteyer</td>
<td>Avan Jogia</td>
<td>Brendan Gleeson</td>
<td>Cassandra Peterson</td>
</tr>
<tr>
<td>Alan Alda</td>
<td>Ana Kasparian</td>
<td>B.J. Novak</td>
<td>Bret Davern</td>
<td>Caterina Murino</td>
</tr>
<tr>
<td>Alan Cumming</td>
<td>Ana Luzardo</td>
<td>Barbara Eden</td>
<td>Brian Blessed</td>
<td>Caterina Scorsone</td>
</tr>
<tr>
<td>Alan Rickman</td>
<td>Andrew Braugher</td>
<td>Barbara Hershey</td>
<td>Brian Cox/Brian Dennehy</td>
<td>Catherine Hardwicke</td>
</tr>
<tr>
<td>Alan Tudyk</td>
<td>Andrea Riseborough</td>
<td>Bear Grylls</td>
<td>Bridget Moynahan</td>
<td>Cedric the Entertainer</td>
</tr>
<tr>
<td>Alba Rohrwacher</td>
<td>Andrew Dice Clay</td>
<td>Bellamy Young</td>
<td>Bridget Regan</td>
<td>Celia Imrie</td>
</tr>
<tr>
<td>Alids Hodge</td>
<td>Andrew Garfield</td>
<td>Ben Feldman</td>
<td>Britt Martin</td>
<td>Choice Crawford</td>
</tr>
<tr>
<td>Alex Borstein</td>
<td>Andrew Lee Potts</td>
<td>Ben McKenzie</td>
<td>Brooke Burke-Charet</td>
<td>Chudzik Roseman</td>
</tr>
<tr>
<td>Alex Kingston</td>
<td>Andrew Rannells</td>
<td>Ben McKenzie</td>
<td>Bruno Boleitner</td>
<td>Charice</td>
</tr>
<tr>
<td>Alex Pettyfer</td>
<td>Andrew Scott</td>
<td>Ben Stiller</td>
<td>Bruno Mans</td>
<td>Charles Dance</td>
</tr>
</tbody>
</table>

**Names are taken from the existing VGGFace2* dataset**

Fully Automated Pipeline

1. Data Extraction
- MFCC extraction
- Face Detection
- Landmark Detection

2. Face Verification
- Match

Active Speaker Verification
- Match

Face Verification

MATCH

Felicity Jones

VOXCELEB
2. Face Tracking and Landmark Detection
Fully Automated Pipeline

Audio feature extraction

Face detection and Tracking

SyncNet

Active Speaker Verification

Face Verification

match

match

3. Active Speaker Detection

Felicity Jones

Face Verification

VOXCELEB
3. Active Speaker Detection - SyncNet

VoxSRC

3. Active Speaker Detection
4. Face Verification

4. Face Verification
High thresholds – no manual intervention

![Graph showing precision and recall for active speaker verification and face verification.](image)
The VoxCeleb1 Dataset

22,496 YouTube videos of 1,251 different celebrity speakers

The VoxCeleb2 Dataset

150,480 YouTube videos

of

7000+ different celebrity speakers

The VoxCeleb Speaker Recognition Challenge
VoxSRC 2019

• The goal of this challenge is to probe how well current methods can recognize speakers from speech obtained 'in the wild'.
• A new dataset has been collected for evaluation.
• VoxCeleb1 test sets are used for validation.
VoxSRC 2019 test set

- 500+ speakers, 19K utterances, 208K pairs
- Collected using a similar pipeline to VoxCeleb
- From YouTube videos of celebrities that do not appear in the VoxCeleb datasets
- ~90% of the impostor pairs are from same gender
- All utterances are at least 4 seconds in length
VoxSRC 2019 test set

- Manual verification of all speech segments
- In addition, annotators pay particular attention to examples whose speaker embeddings are far from cluster centres
VoxSRC 2019 - Tracks

• **Fixed:** Participants can train only on the VoxCeleb2 dev dataset for which we have already released speaker verification labels.

• **Open:** Participants can use the VoxCeleb datasets and any other data (including that which is not publicly released) except the challenge's test data.