DiverseFake:
A Comprehensive Dataset of DeepFake Generation Techniques

Mu-Ruei Tseng, Xixiang Luo, Ann McNamara, and Xin (Shane) Li*
Emails: mtseng@tamu.edu, xixiangluo999@gmail.com, ann@tamu.edu, xinli@tamu.edu

TEXAS A&M UNIVERSITY

Introduction

•Existing deepfake datasets are outdated and only cover a narrow range of generation methods, leading to saturation in detectors trained on them, which consequently do not perform well on recent deepfake media.
•We present a pipeline to create a diverse, state-of-the-art dataset.

Contributions:

- Developed a multi-faceted deepfake dataset with latest deep fake generation models.
- Fine-tuned latest detectors using this dataset to improve cross-dataset accuracy.

Datasets & Models

•**Public Datasets we integrated:**
- A. FaceForensics++ *(2019)*
- B. Celeb-DF (v2) *(2020)*
- C. FakeAVCeleb *(2021)*

•**Models adopted for new data generation**
- **Identity Swapping** Techniques:
  - G1: SimSwap *(ACMMM 2020)*
  - G2: MobileFaceSwap *(AAAI 2022)*
- **Content Manipulation** Techniques:
  - G3: DaGAN *(CVPR 2022)*
  - G4: MCNet *(ICCV 2023)*
  - G5: SadTalker *(CVPR 2023)*

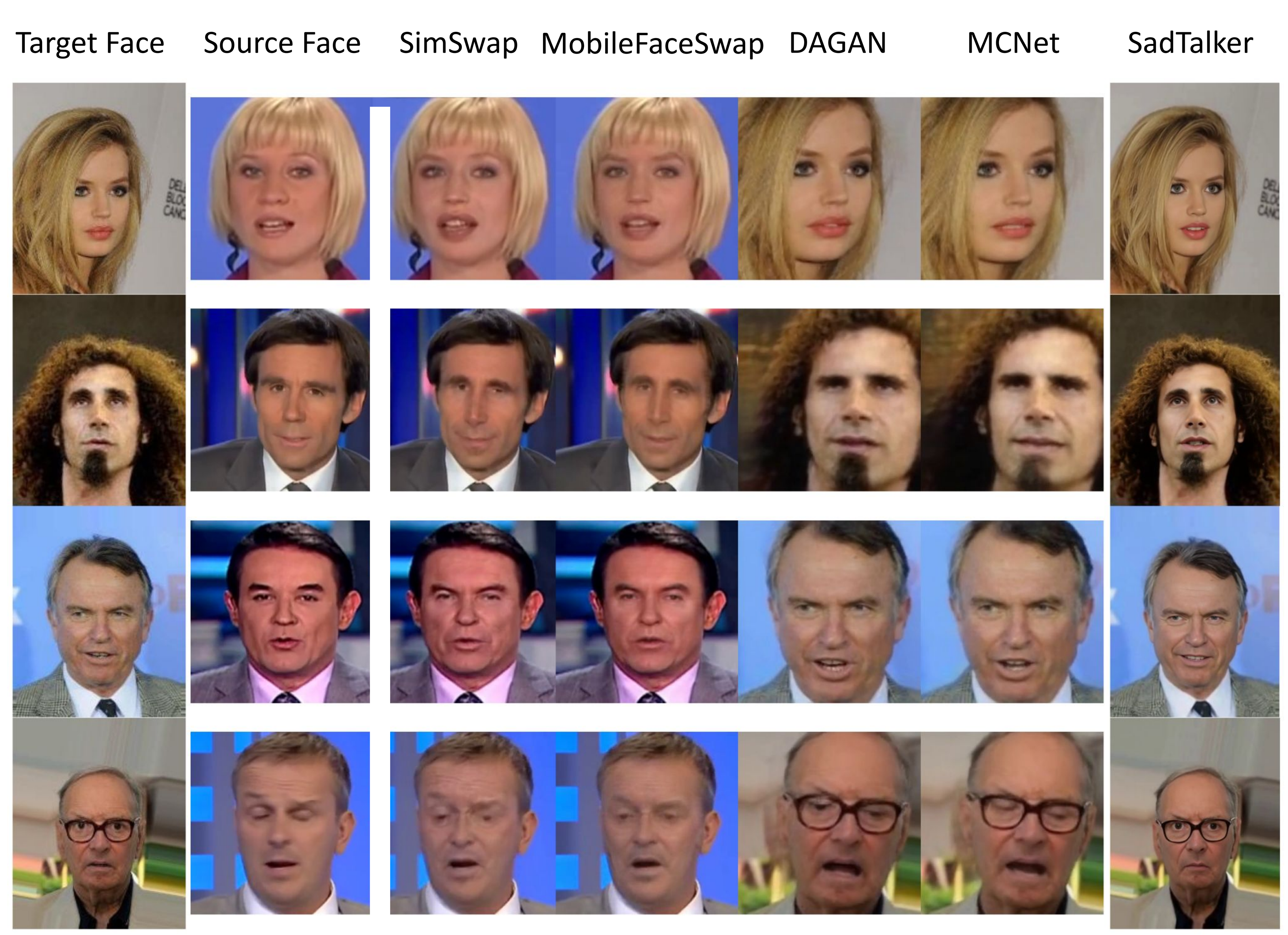
- **Top Detectors from DeepfakeBench**
  - D1: Xception *(ICCV 2019)*
  - D2: SPSL *(CVPR 2021)*
  - D3: RECCE *(CVPR 2022)*
  - D4: UCF (*ICCV 2023*)

Methodology

•Building new database:
- The original content to be changed is referred to as the "source"; while the intended identity (to appear in the new video) is termed the "target".
- Source and target videos were collected from public datasets such as FaceForensics++, Voxceleb2, and CelebA. We also included multiple videos in the wild.
- For effective training, used the VGG-Face model to select source and target face pairs, ensuring they are distinct enough to be distinguishable, yet similar enough to maintain realism.

- Deep fake techniques were categorized into two types, with the latest representative generators selected from each category:
  - **Face Swapping**: Modify a source video by replacing the source face with a target face.
  - **Content Manipulation**: Edits or creates new content directly by using driving images, audio, or videos.

Generated Result Examples



For more examples, scan the QR code to visit the project website.

Quantitative Evaluation

•SOTA detectors that nearly reached saturated performance on existing datasets now demonstrate lower accuracy on our dataset.

Table 1. Detectors were trained on the training set of Dataset A and tested on the test sets of Datasets A, B, C, and our dataset. A noticeable drop in performance (Video AUC) was observed. Similar trends were seen when detectors were trained on Datasets B and C.

Detectors / Datasets	A	B	C	Ours
Xception	99.61%	81.65%	94.88%	**76.61%**
SPSL	98.55%	79.92%	87.95%	**68.63%**
RECCE	99.34%	82.21%	90.55%	**72.91%**
UCF	99.80%	83.79%	91.31%	**68.17%**

•Detectors fine-tuned with our dataset yields most significant improvement across all test sets.

Table 2. Detectors initially pre-trained on Dataset A, fine-tuned with the training sets of B, C, and our dataset; and then evaluated on a mixed test set comprising samples from B, C, and ours with the same percentage. Detectors fine-tuned with our dataset yields largest performance gains.

Detectors	Pretrained Detector (on A)	Fine tuning with B	Fine tuning with C	Fine tuning with Ours
Xception	81.75%	80.26%	83.23%	**84.01%**
SPSL	78.75%	81.54%	83.07%	**89.07%**
RECCE	76.29%	77.41%	79.31%	**83.92%**
UCF	79.25%	76.42%	80.03%	**81.24%**

Ongoing Work

•Generate and integrate more challenging datasets to enhance the training of advanced deep fake detectors.
•Incorporate eye-tracking and attention-tracking to efficiently produce produce higher-quality annotations for more effective training.

Acknowledgement:
This project was partly supported by GCRI Seed Funds and NIH R15HD108765.