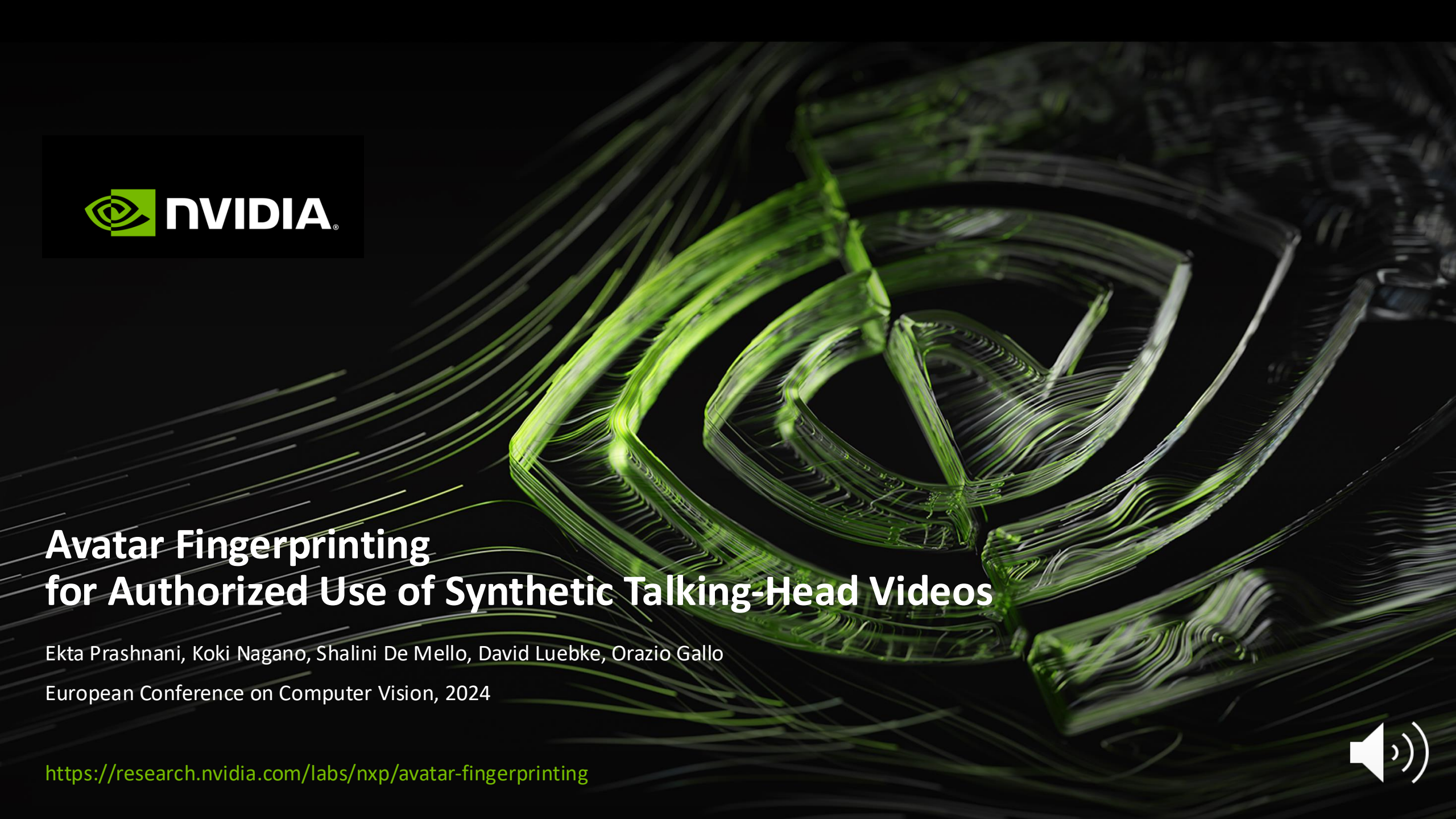


Avatar Fingerprinting for Authorized Use of Synthetic Talking-Head Videos

Ekta Prashnani, Koki Nagano, Shalini De Mello, David Luebke, Orazio Gallo

European Conference on Computer Vision, 2024

https://research.nvidia.com/labs/nxp/avatar-fingerprinting

Context: Talking-Head Video Generation

Self-Reenactment: Driving Identity = Target Identity

Driving Video

Talking-Head Generator*

Target Image

Synthesized Video (or Target Video)

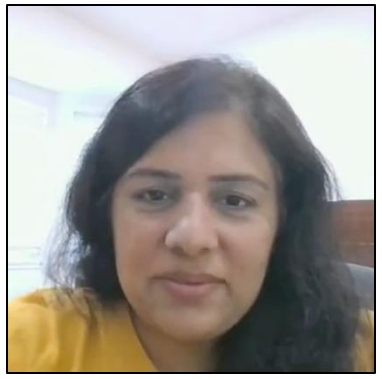
*face-vid2vid: https://nvlabs.github.io/face-vid2vid/

Context: Talking-Head Video Generation

Cross-Reenactment: Driving Identity ≠ Target Identity

Identity: ID1

Talking-Head Generator



Target Identities:
ID2, ID3

Target Videos:
ID1 → ID2, ID1 → ID3

Novel Task: Avatar Fingerprinting

Identity: ID1

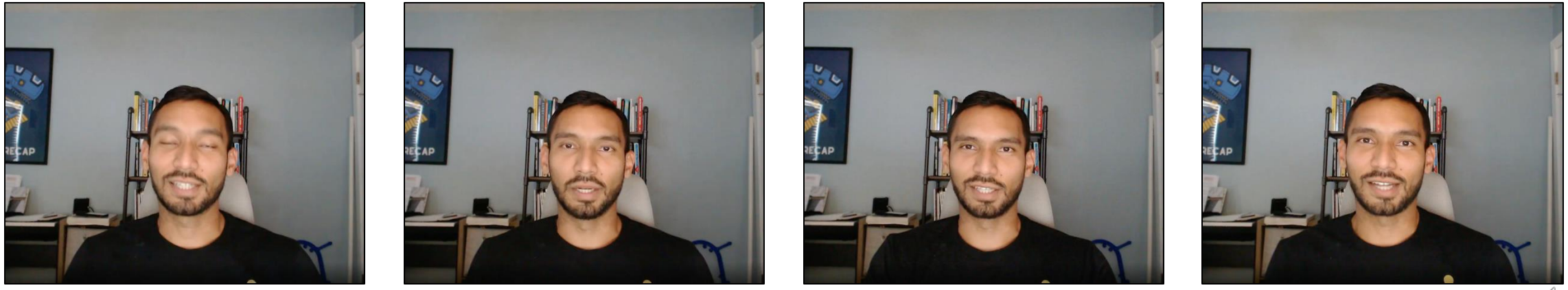
Identity: ID4

Original Videos

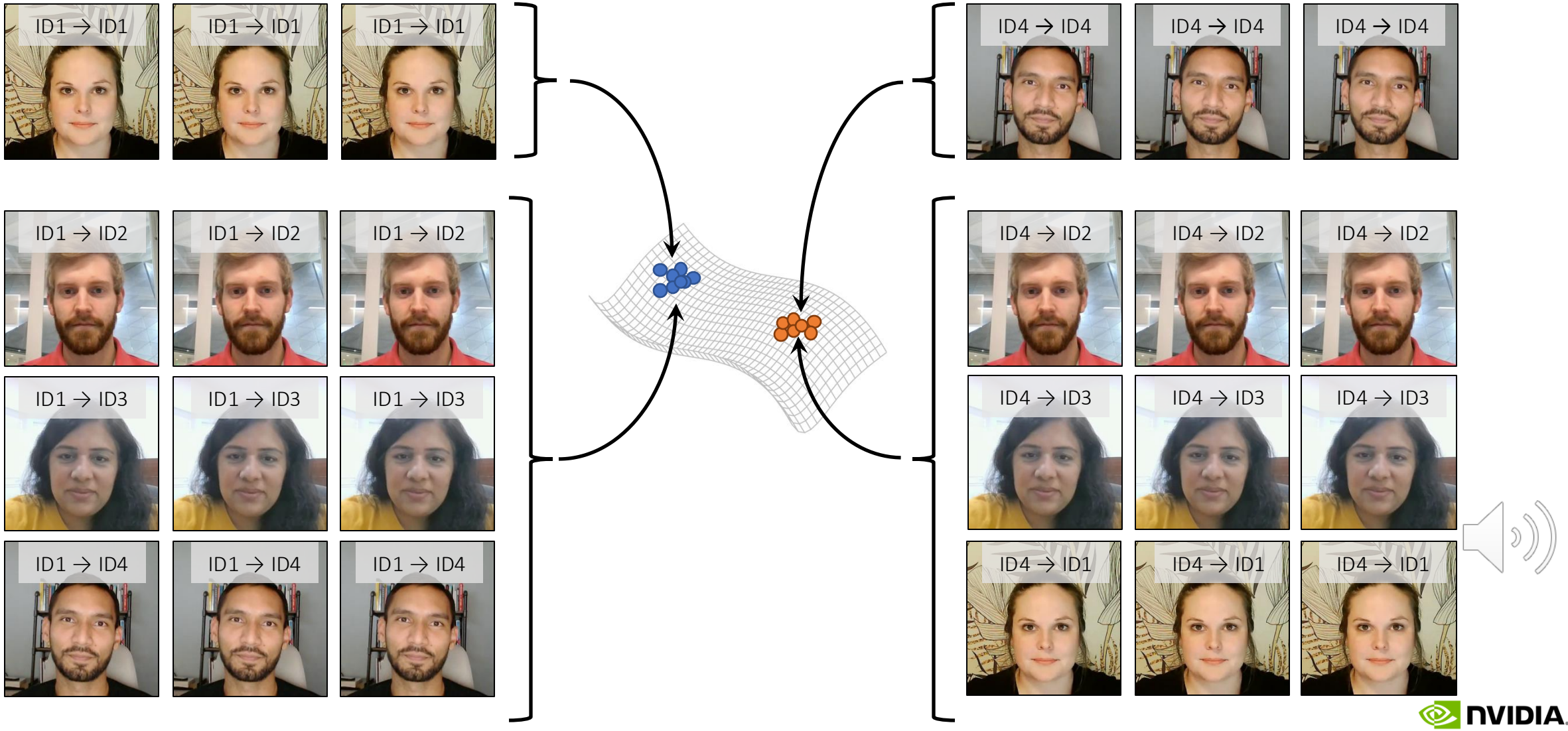
Novel Task: Avatar Fingerprinting

Identity: ID1

Identity: ID4



Novel Task: Avatar Fingerprinting



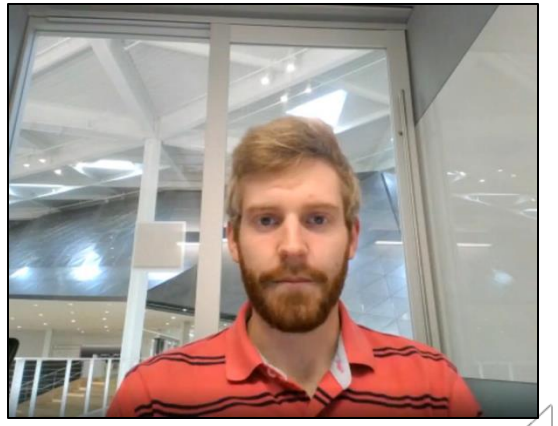

NVIDIA Facial Reenactment (NVFAIR) dataset

2-Stage Data Capture over Video Calls

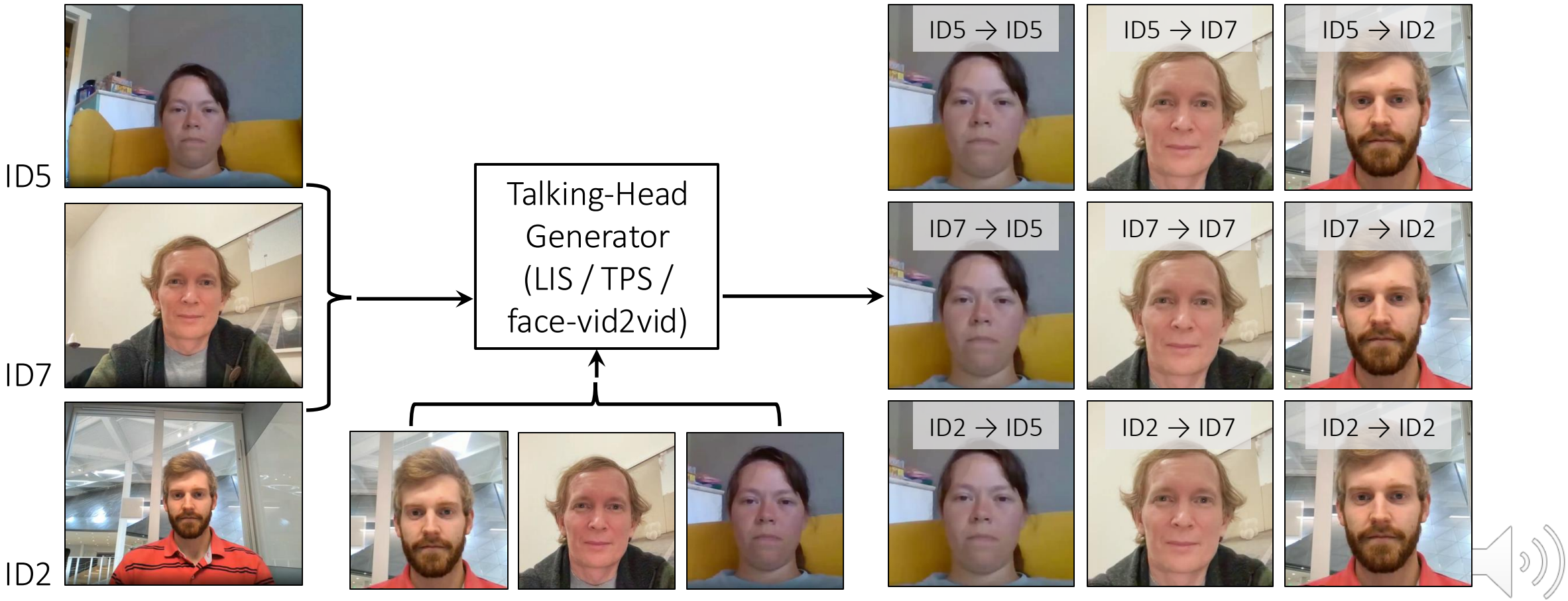
Stage I: Free-Form Monologues

prompt: “Is there a household chore you don’t like doing?"

Stage II: Scripted Monologues



Synthetic Talking-Head Video Generation



Original Videos of Driving Identities

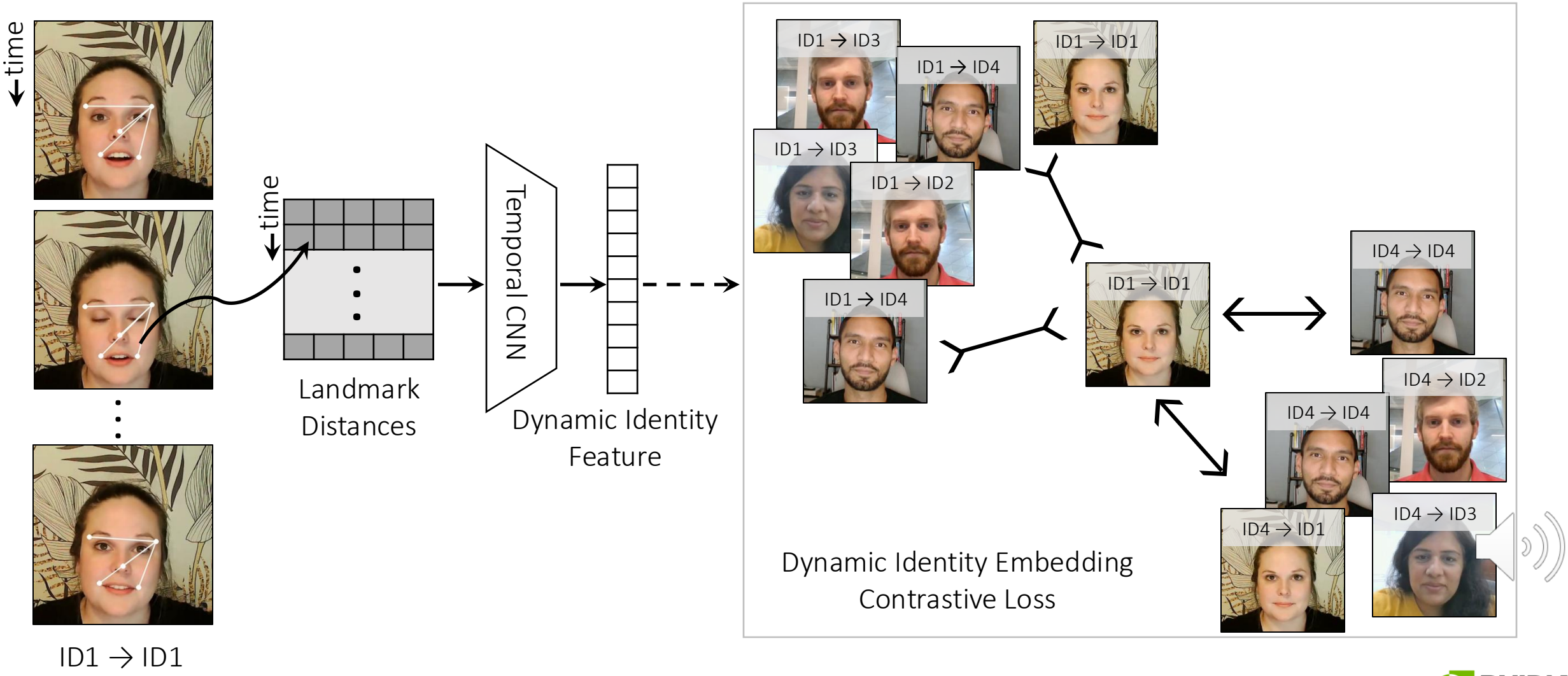
Target Identities: ID5, ID7, ID2

Synthetic Talking-Head Videos

[1]face-vid2vid: https://nvlabs.github.io/face-vid2vid/ [2] https://github.com/wyhsirius/LIA [3] https://github.com/yoyo-nb/Thin-Plate-Spline-Motion-Model

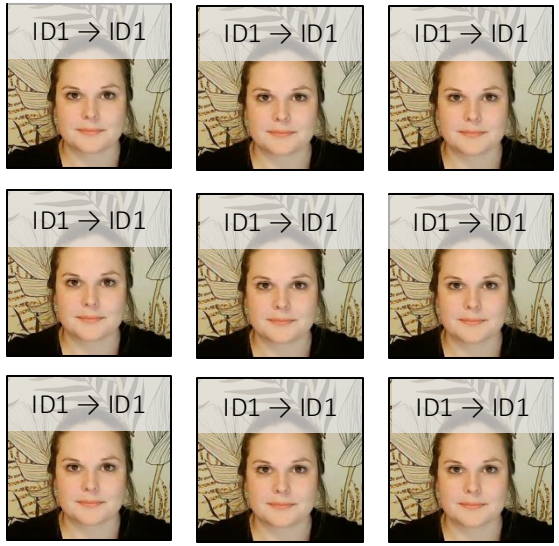
NVIDIA

Training an Avatar Fingerprinting Model

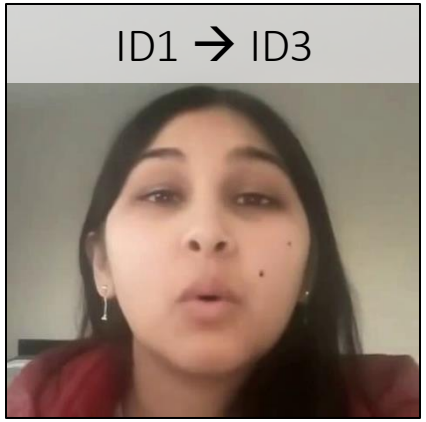


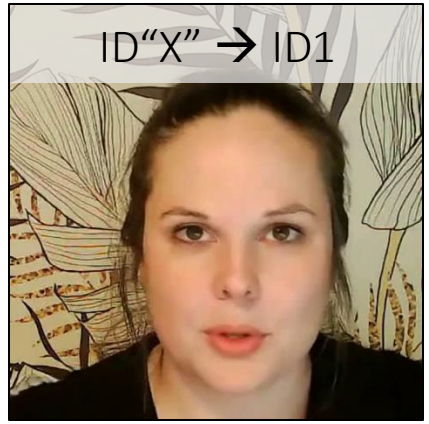
Visual Results with Reference = ID1

reference videos of ID1



Avatar Fingerprinting applied to three types of synthetic videos:

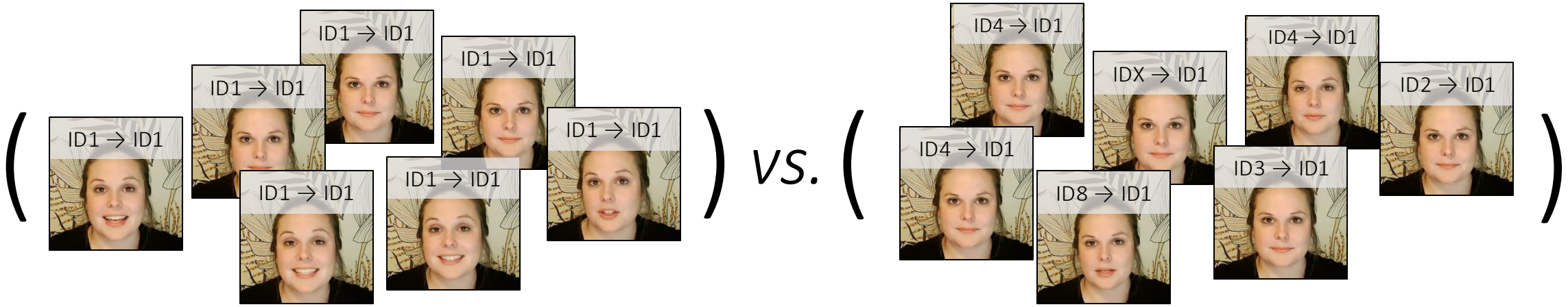




$d = 0.590$ $d = 0.961$ $d = 1.787$

videos driven by ID1 are closer to the reference set

Results: Self vs. Cross-Reenactments



test set: 35 identities disjoint from training set

AUC on test set:

- ✓ face-vid2vid[1] generator: 0.87
- ✓ LIA[2] generator: 0.84
- ✓ TPS[3] generator: 0.82

} robustness to new generators not seen during training

[1]face-vid2vid: https://nvlabs.github.io/face-vid2vid/
[2] https://github.com/wyhsirius/LIA
[3] https://github.com/yoyo-nb/Thin-Plate-Spline-Motion-Model

NVIDIA

Contributions

- ✓New task: Avatar Fingerprinting
  - ✓to enable authorized use of synthetic talking-head videos
- ✓NVFAIR: A novel large-scale face-reenactment dataset
  - ✓largest collection of face-reenactments (650,000+)
  - ✓three talking-head generators
  - ✓both self- and cross-reenactments per subject
  - ✓scripted and free-form monologues
  - ✓videoconference-based recording
- ✓The first avatar fingerprinting algorithm

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Distribution Statement A – Approved for Public Release, Distribution Unlimited

NVIDIA