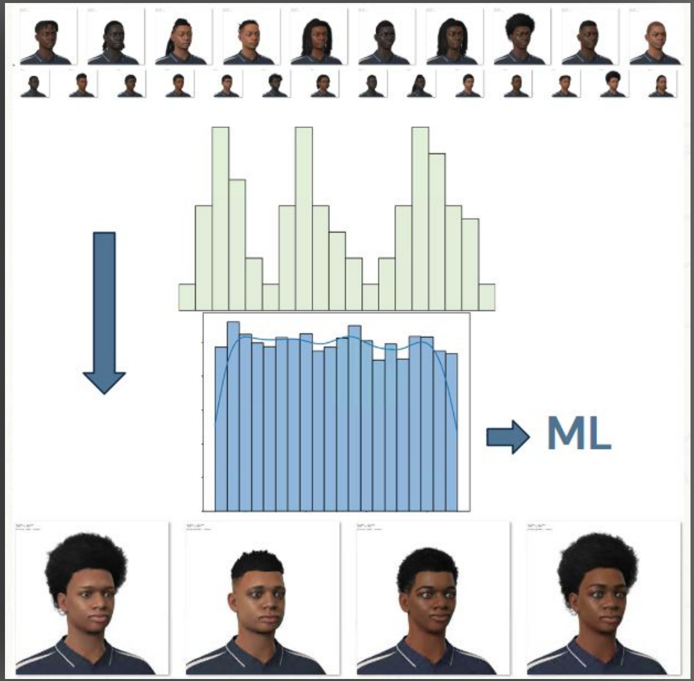
Towards Optimal Training Distribution for Photo-to-Face Parametric Models in Video Games

Igor Borovikov, Karine Levonyan
iborovikov@ea.com, karine@ea.com

Electronic Arts, SEED, Redwood Shores, CA



Agenda

- **Intro**
  - Pseudo-Random Heads
- Motivation
  - Parametric models, FLAME, Photo - to - (Face) Parameters
- **Objectives**
  - Latent and authoring spaces
  - Optimal distributions
  - Pseudo-random heads
- **Conclusion and future work**

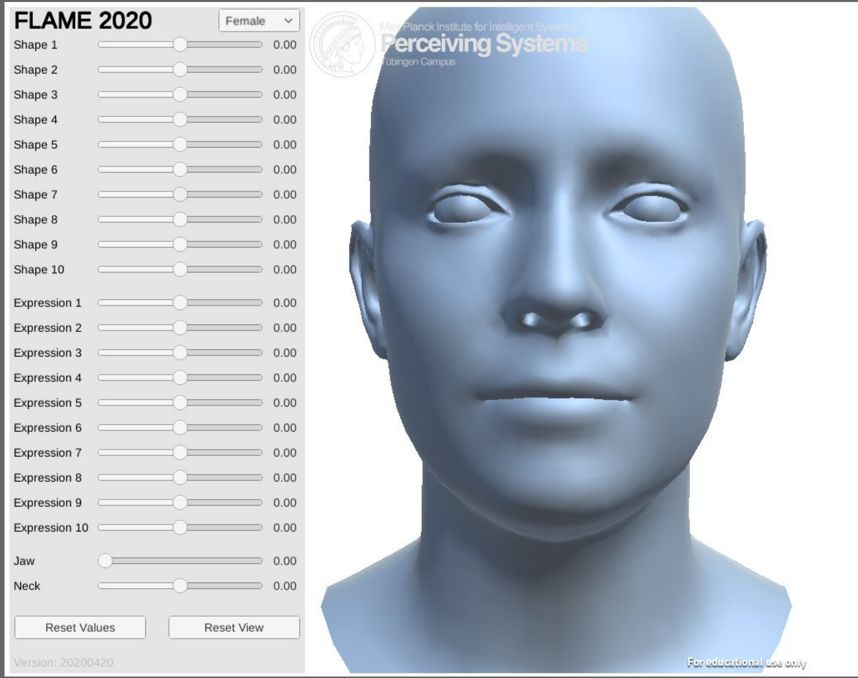
Pseudo-random Heads

- Believable, “natural”
- Variety and Dissimilarity
- “Unlimited” number of heads

FLAME parametric model

Terminology and context:

- We focus on parametric model of a human head (FLAME and an in-house tool)
- FLAME parameters are "authoring parameters"



FLAME web-based editor
https://flame.is.tue.mpg.de/

Motivation: monocular reconstruction face-to-parameters

- **Parametric models are common in video games**

- **Fitting parameters from a single photo (aka "monocular reconstruction")**

  Art pipeline, potential player-facing features

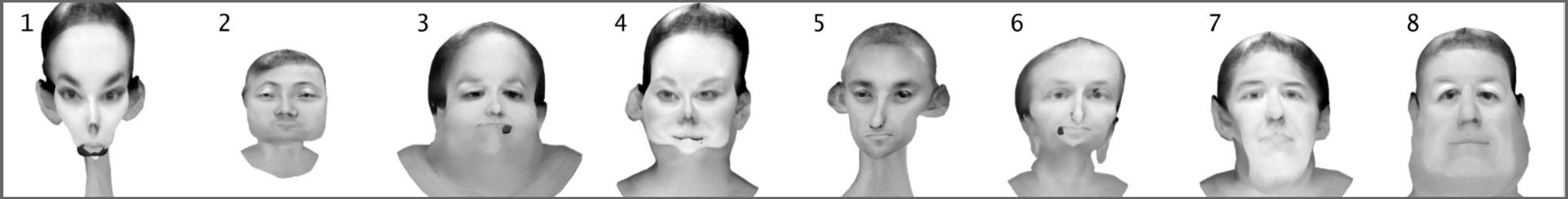
- **Training a large DNN for that requires lots of training data.**

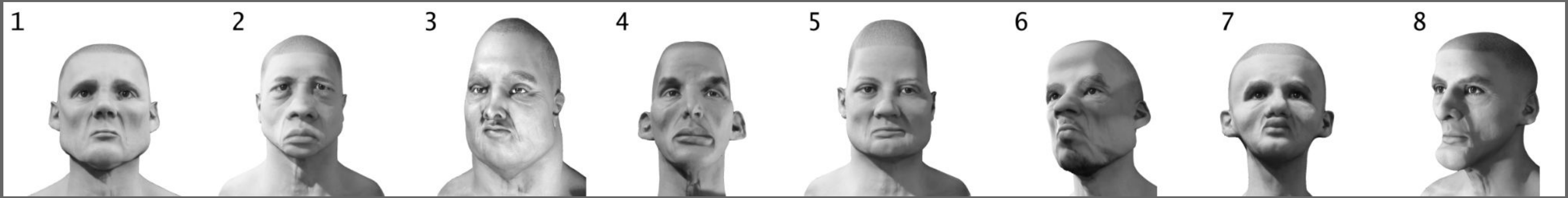
- **Training data is synthetic, generated with authoring tools via automation by varying authoring parameters.**

  Data = {(Image, parameters), …}

Completely random training data

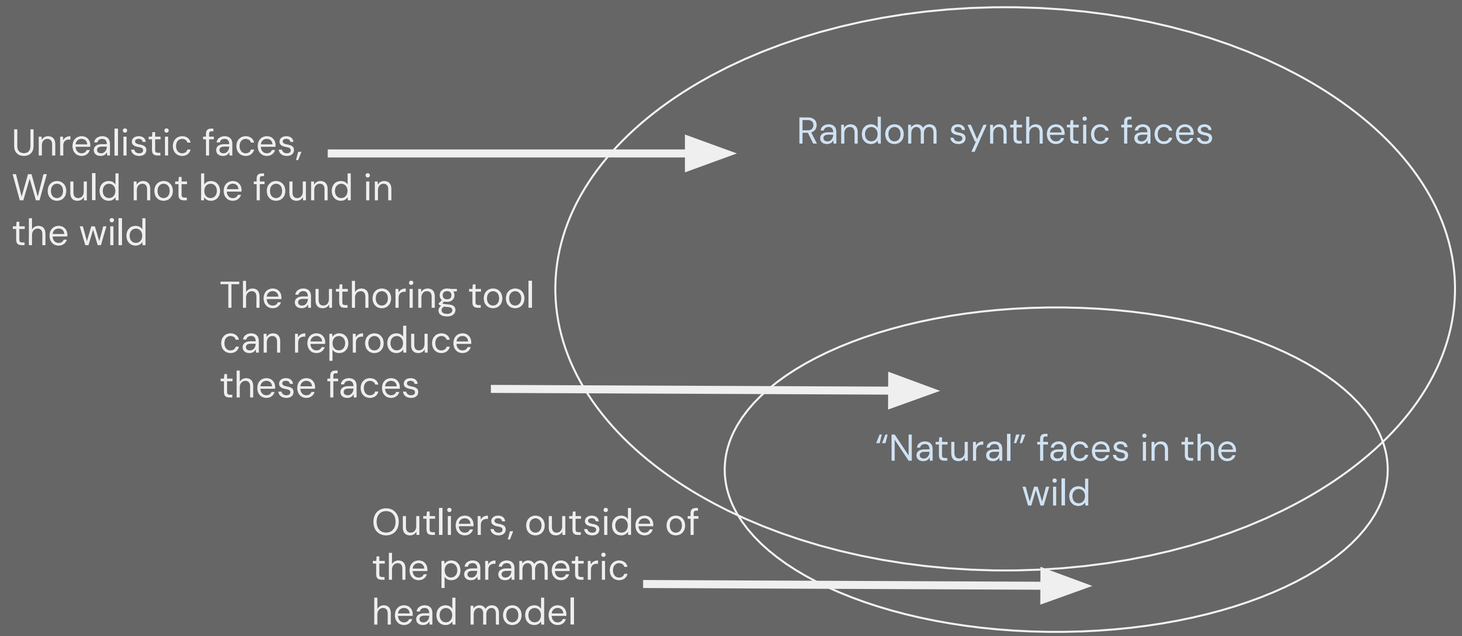
Completely random heads. Top: FLAME, bottom: in-house
We lose correlations between features in completely random data





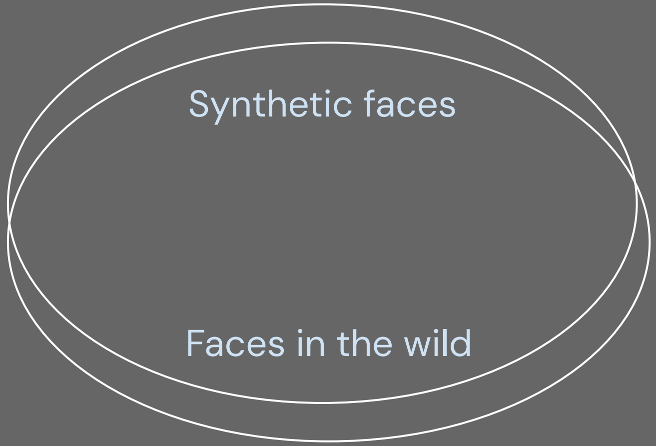
Domain gap

Synthetic data diversity and domain coverage



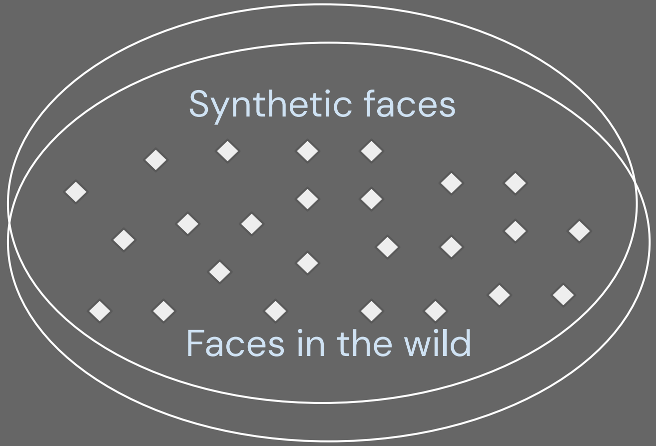
Training data optimization objectives

- **Diversity of the generated heads**
- **Maximum coverage of possible heads in the wild**
- **Parameters correlations:**
  - Avoid spurious correlations
  - Capture “natural” correlations

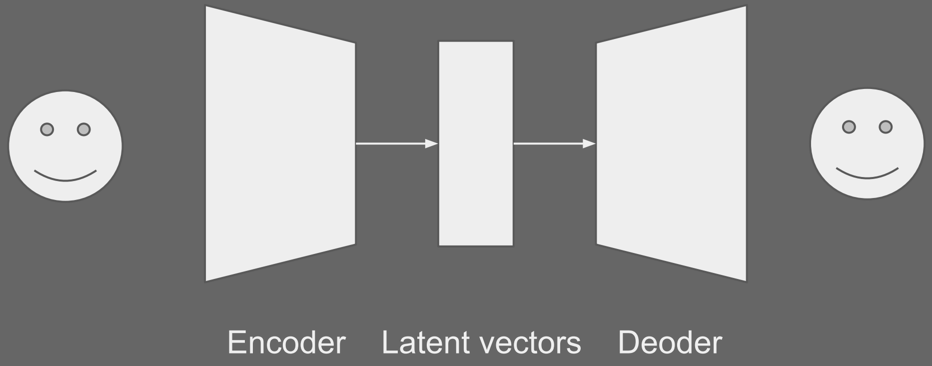


Training data optimization objectives

- **Diversity of the generated heads**
- **Maximum coverage of possible heads in the wild**
- **Parameters correlations:**
  - Avoid spurious correlations
  - Capture “natural” correlations
- **Optimal sampling of the domain:**
  - Prevents biases
  - Reduces the amount of required data



FaceNet and its latent Space



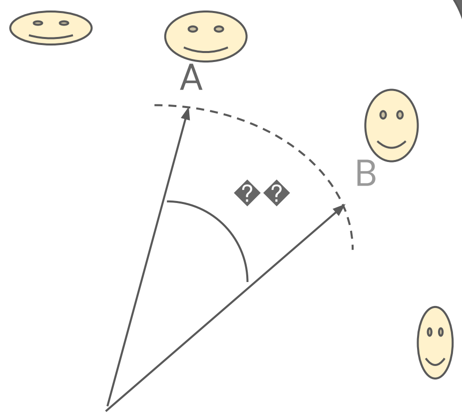
Latent space offers a compressed representation of facial features.

FaceNet Latent Space

- **Similar latent vectors (cosine distance) represent similar faces**

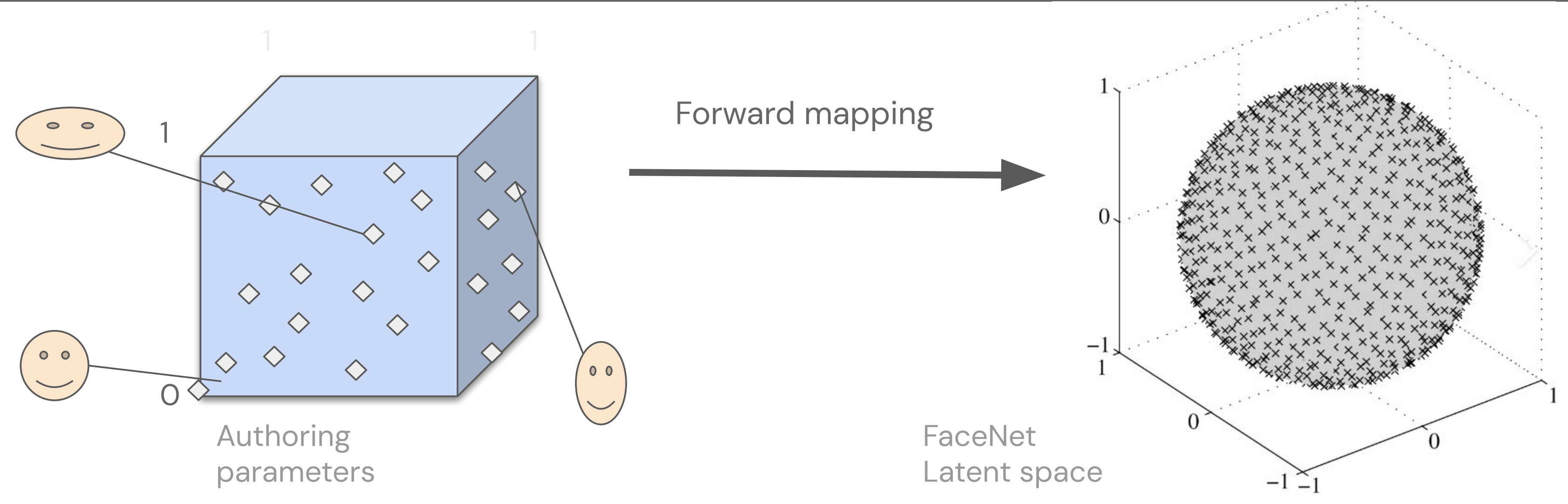
- **Cosine distance < 0.51 ⇒ "same face"**
  ~10k "substantially different" human faces

- 512-dimensional unit sphere



$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

Latent and authoring spaces

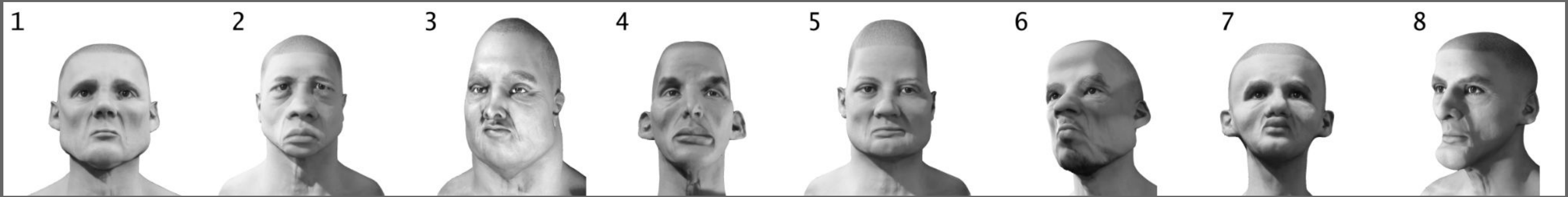


Random authoring parameters ⇒ image ⇒ Latent vector

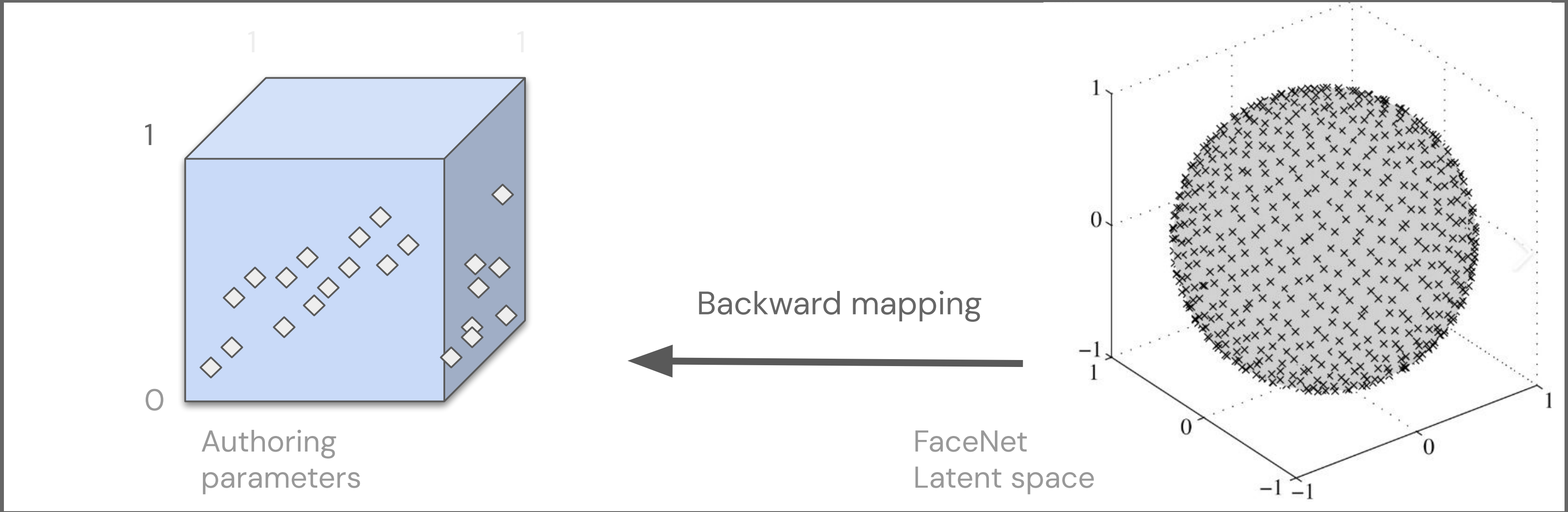
Authoring parameters ⇒ **neural net F** ⇒ Latent vector

Completely random training data

Completely random heads. Top: FLAME, bottom: in-house
We lose correlations between features in completely random data

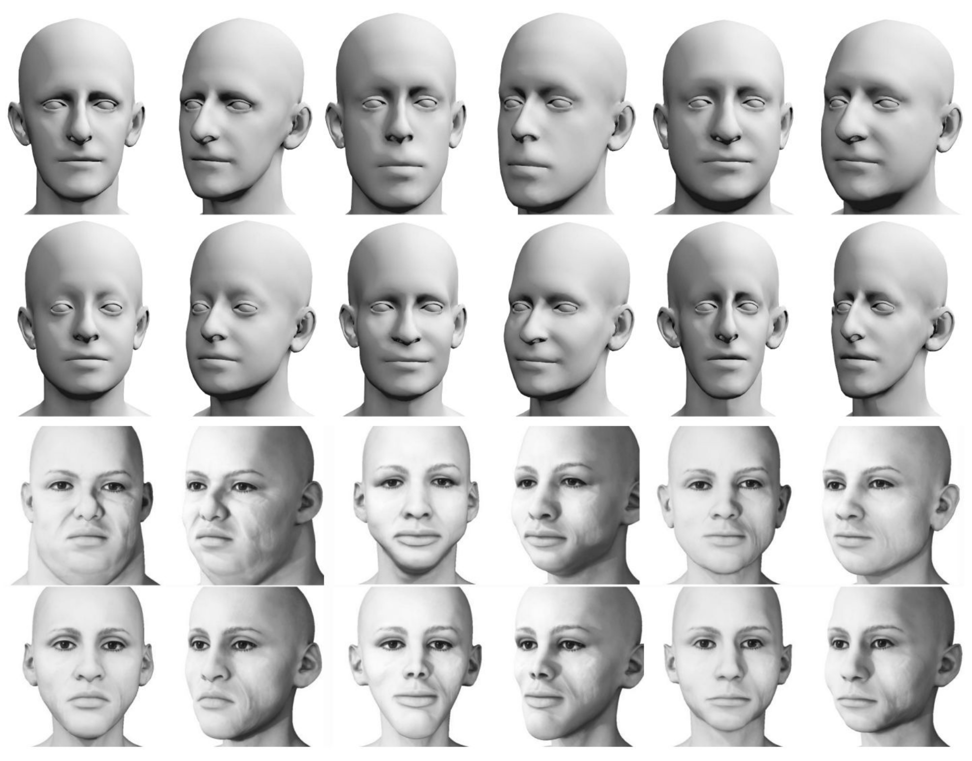


Latent and authoring spaces

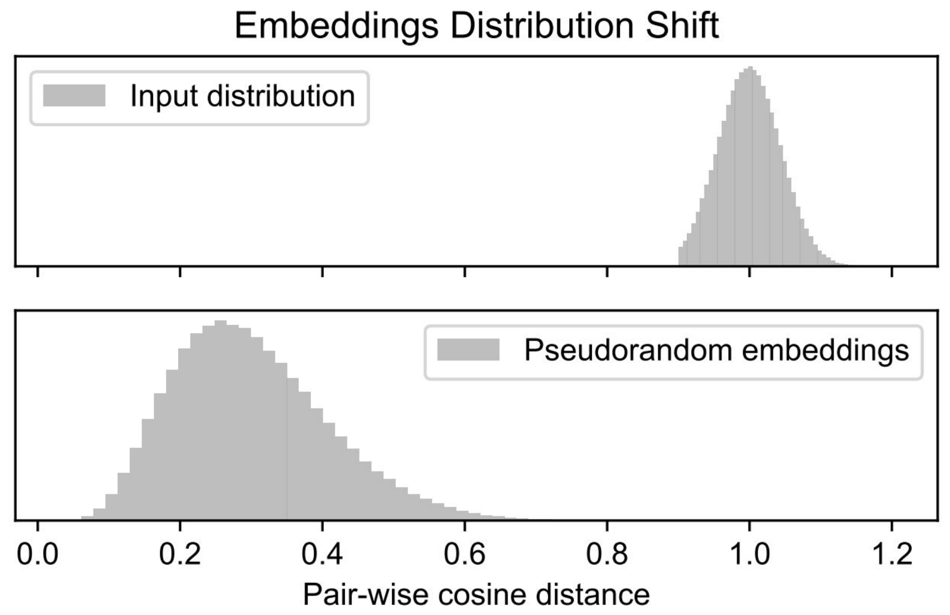


Authoring parameters ⇐ **neural net B** ⇐ Latent vector

Sampling with the backward mapping



Distribution shift: $B \neq F^{-1}$



Distribution shift of the pair-wise distances in embedding space. Top: samples s in L. Bottom: rendered $B(s)$ samples with reduced separation.

Research item – fix $B \neq F^{-1}$

Q: Can we learn "rotation correction" $\rho(s)$?

Define $\rho(s) = s - B \circ A \circ F(s)$ for the "roundtrip" of a sample from L to A and then back to L. Then, instead of $B(s)$, use $B(s - \rho(s))$ when sampling. The goal is to reduce the distribution shift on the plot.

Future work, work in progress

- **From shape to full head including color palette elements, facial expression, hairstyle, facial hair, etc**
  - Expectation: the distribution shift will reduce
  - Better understanding of the latent space structure and the influence of concrete features on embeddings
  - Ranking authoring parameters by influence on visual variety to reduce number of sliders (e.g., mobile applications)
  - Completeness of the authoring space: how large are gaps in the latent space that we can't populate with the authoring tool?

- **Other face-related latent spaces in generative domain?**

Conclusion and future work

Thank you for attending!
Q&A