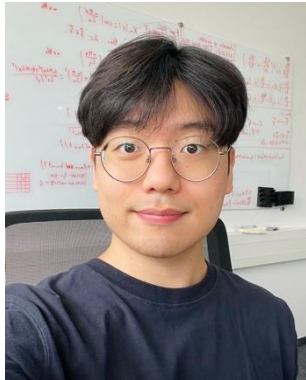


# Dress-Up: Generating Animatable Clothed 3D Humans via Latent Modeling of 3D Gaussian Texture Maps



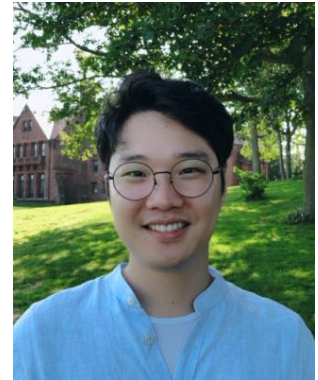
Kim Youwang



Lee Hyoseok



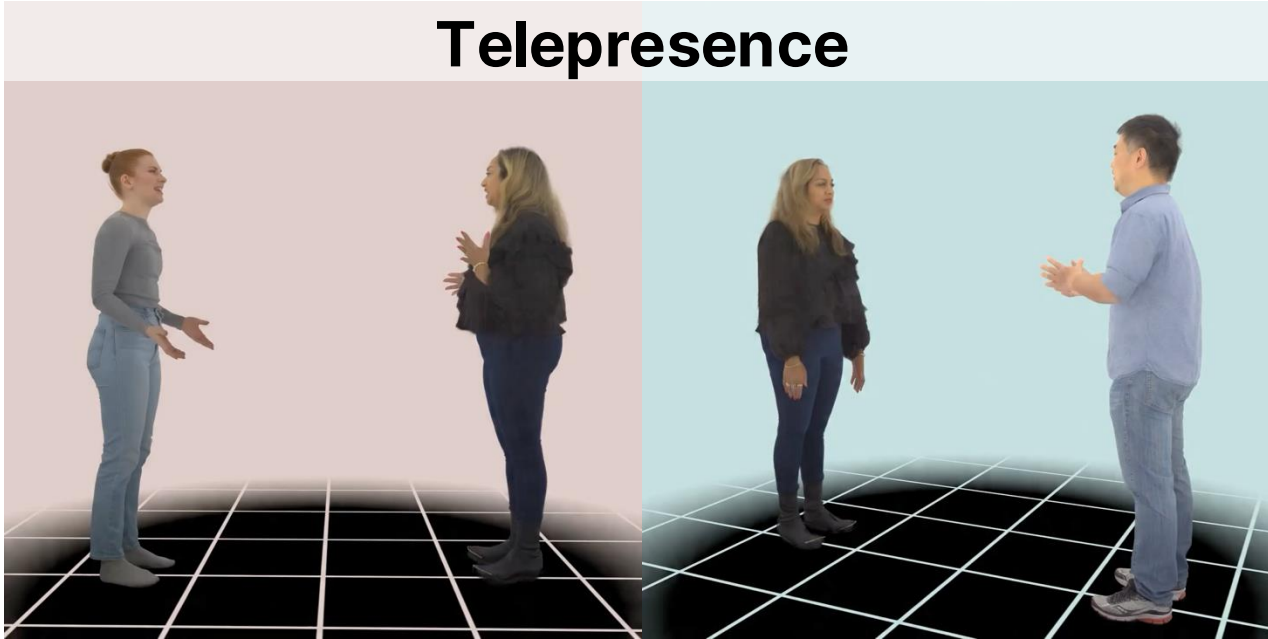
Gerard Pons-Moll



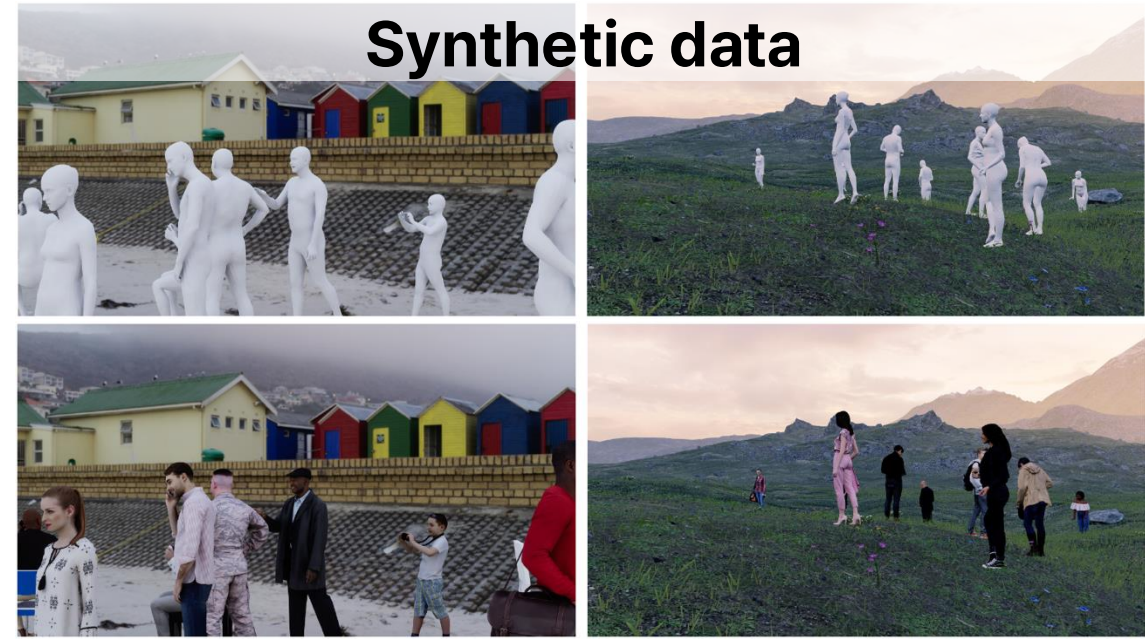
Tae-Hyun Oh

# Photorealistic, animatable, clothed 3D humans

## Telepresence



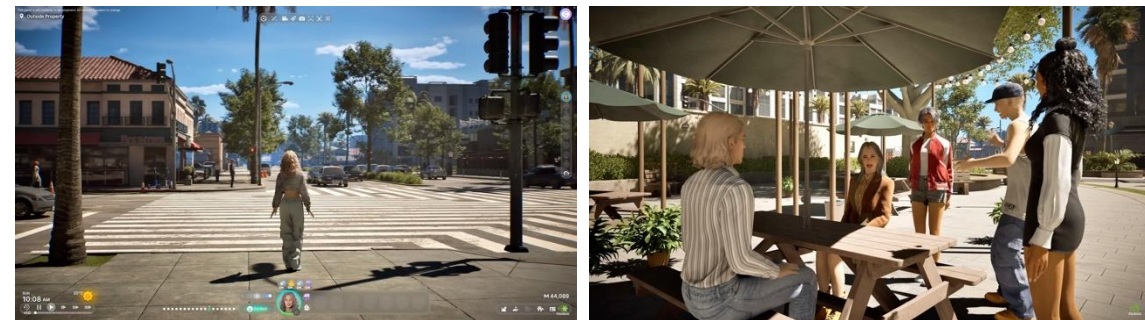
## Synthetic data



## 3D virtual try-on



## Gaming & content creation



Agrawal et al., "Seamless Interaction: Dyadic Audiovisual Motion Modeling and Large-Scale Dataset," arXiv preprint 2506.22554.

Patel et al., "AGORA: Avatars in Geography Optimized for Regression Analysis," CVPR 2021.

Cao et al., "GS-VTON: Controllable 3D Virtual Try-on with Gaussian Splatting," arXiv preprint 2410.05259

NVIDIA ACE | inZOI - Create Simulated Cities with Co-Playable Characters

# Conventional avatar creation pipeline

Stage 1) Real-world performance captures

Stage 2) Optimize animatable 3D representations

Stage 3) Animate avatars at test-time

## Performance capture



Usually takes  
several hours :(

Optimizing  
3D representations  
(3DGS, NeRF, etc)

## Test-time animation





# Our goal: Scalable avatar creation pipeline

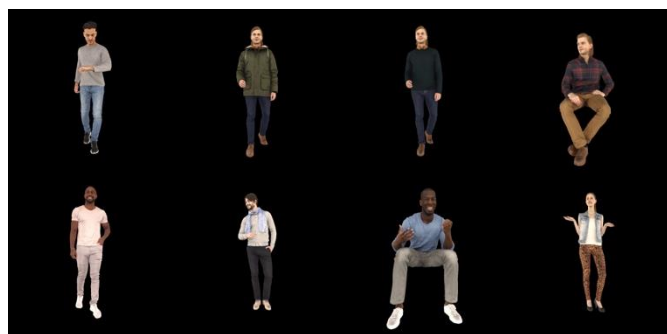
We build a feed-forward, unconditional generative model for creating **Photorealistic, Animatable, Clothed** 3D human avatars



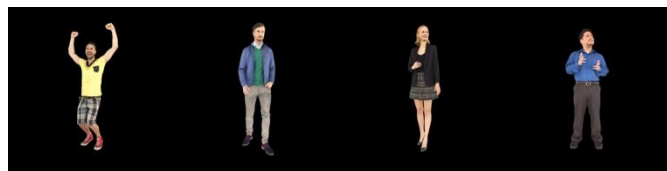
# Dress-Up: Generative modeling of 3D humans

## Learn latent space from real human scans

- Build a compact latent space of **diverse human textures and geometries**
- Encoder-free latent space modeling, i.e., auto-decoding, for efficient training



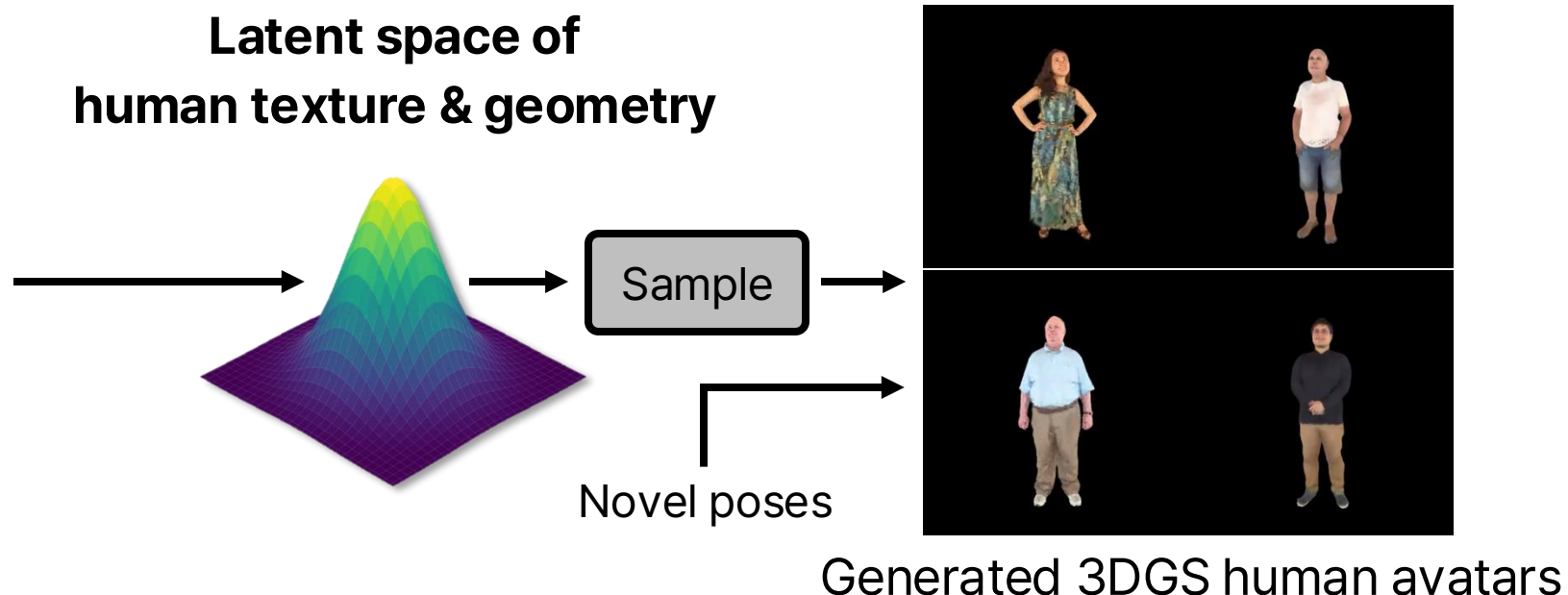
...



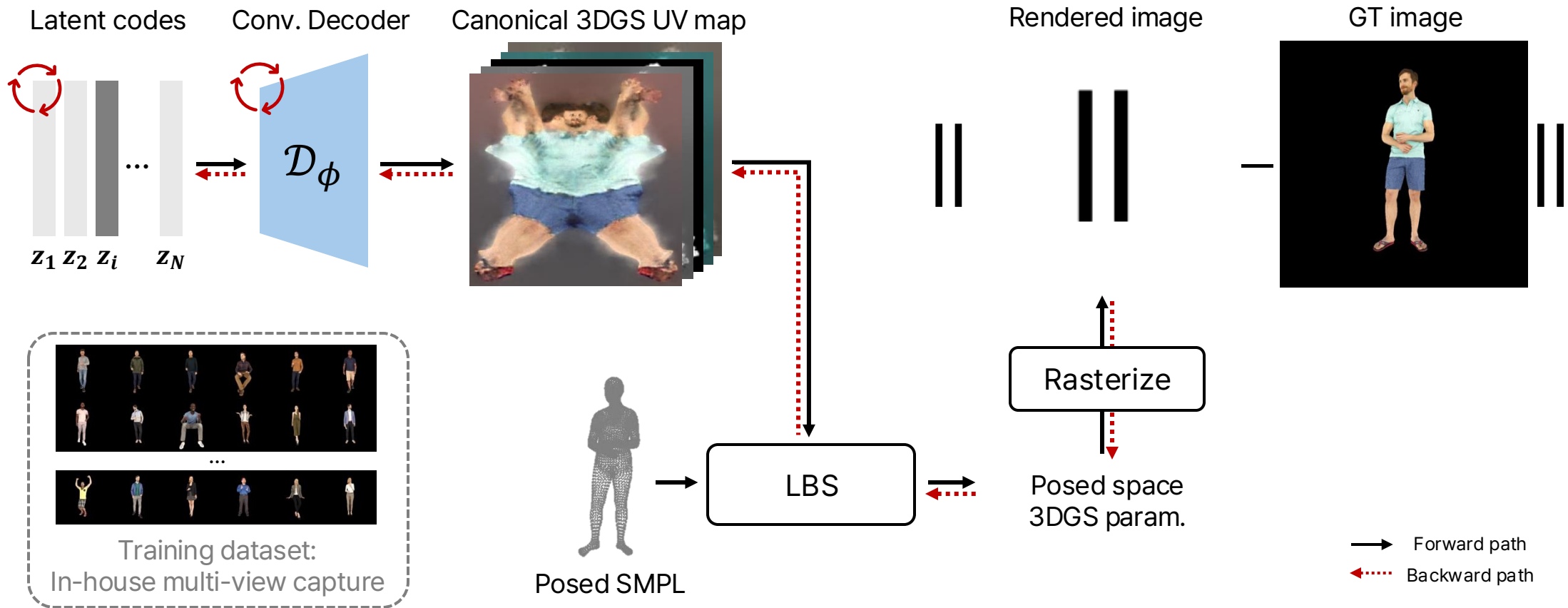
Multi-view real-human scans

## Learn to sample from the latent space

- Train a **latent diffusion model (LDM)** using the built latent codes as a dataset
- LDM maps randomly sampled noise into the valid latent codes, later decoded into 3DGS avatars



# Latent space modeling of 3D clothed humans

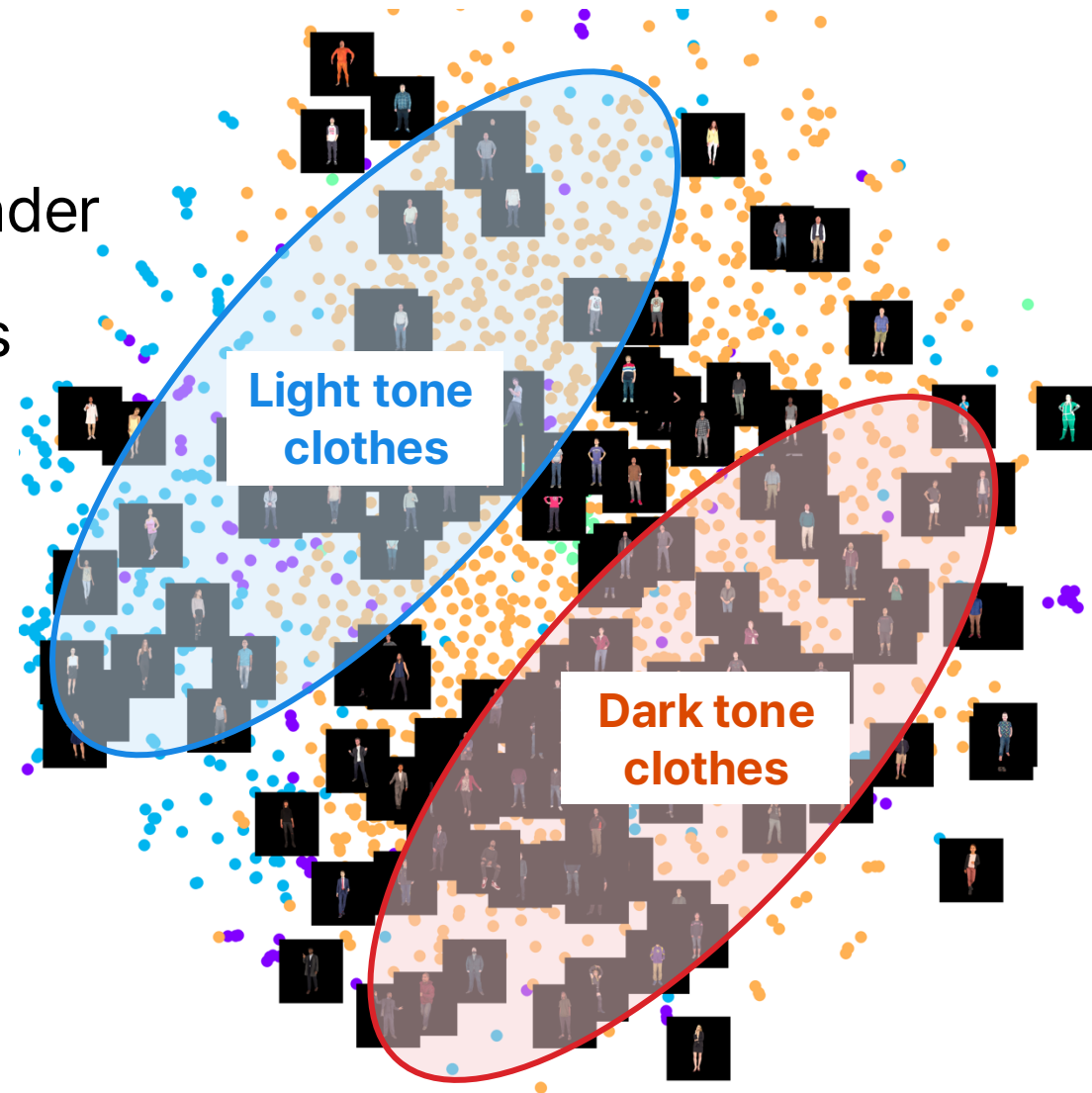
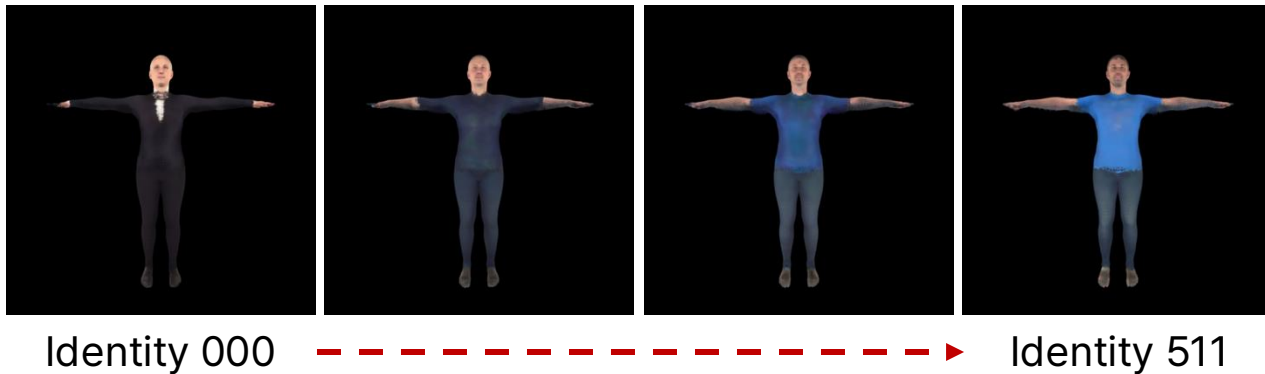


Training objective

$$\mathcal{L}_{\text{AD}} = \sum_{i=1}^N \sum_{v=1}^V \underbrace{\|\mathbf{I}_{i,v} - \mathcal{R}(\boldsymbol{\theta}, \mathcal{D}_\phi(\mathbf{z}_i))\|_1}_{\text{Recon. loss}} + \underbrace{\|\mathbf{z}_i\|_2^2}_{\text{Regularization}}$$

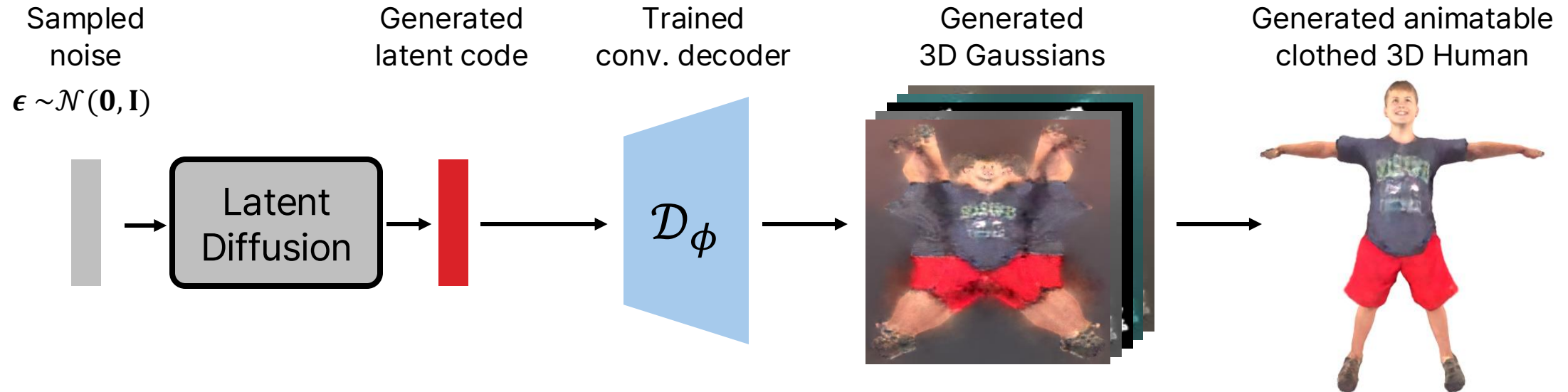
# Learned latent space of 3D clothed humans

- We trained a latent space with **2K identities**
- The learned latent space **inherently finds the semantic clusters**, e.g., cloth colors, gender
- **Interpolation** across the learned latent codes are possible → Already can create new IDs



# Latent diffusion for generating 3DGS avatars

- Train a **latent diffusion model (LDM)** using the learned per-ID latent codes
- LDM learns the complex **data manifold of the clothed canonical 3DGS avatars**
- LDM can sample novel codes, and these codes will be decoded into **3DGS parameters in SMPL UV coordinates**





# Randomly generated canonical 3DGS avatars



# Animating the generated avatars at test-time



# Current limitations & future directions

- **Lack of high-quality multi-view human scan datasets**

→ Can benefit from synthetic image datasets, paired with 3D skinned meshes<sup>[1]</sup>

- **Do not model the dynamic properties of the clothes**

→ Can extend with Physics-integrated 3D Gaussians<sup>[2]</sup>

- **Lack of appearance controllability (e.g., text control)**

→ Can add text or other modalities as a condition for LDM, to add controllability (we have some early results!)



A male wearing  
a black suit,  
white shirt



A black man  
wearing  
a gray suit



A kid wearing  
a blue t-shirt  
and jeans




An old man  
wearing a  
sportswear

[1] Zhuang et al., "IDOL: Instant Photorealistic 3D Human Creation from a Single Image," CVPR 2025

[2] Xie et al., "PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics," CVPR 2024

# Future direction: Trinity for virtual 3D humans

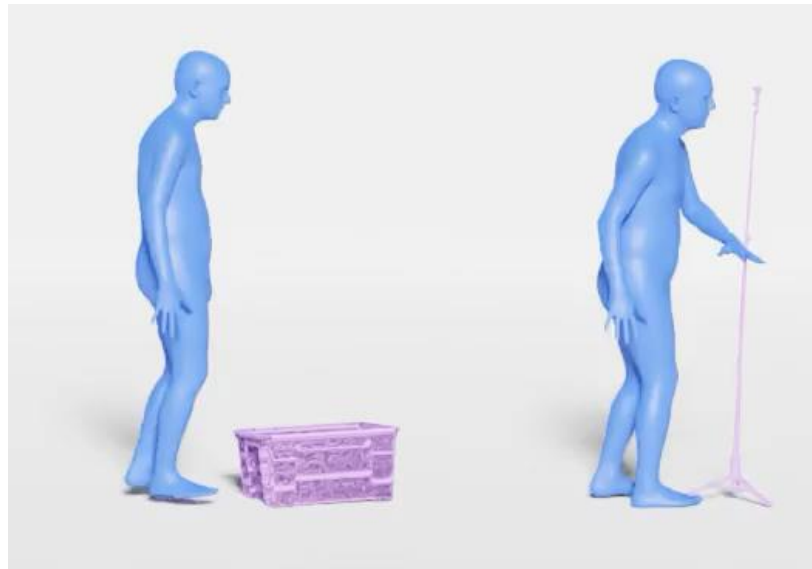
- A virtual 3D human, which satisfies the following:
  - 1) Exhibit **photorealistic** appearance  **Dress-Up**
  - 2) **Interact seamlessly** with the 3D objects and scenes
  - 3) Make lifelike **emotional expressions**
- Could pave the way toward the **next-generation of human-centric content creation.**

} **More ways to go!**

**Photorealism**



**Physical interaction**



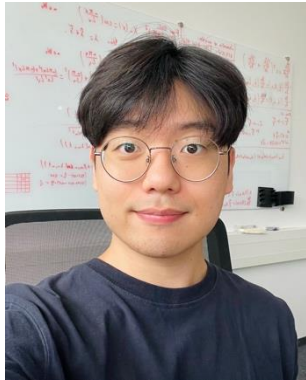
**Emotional expression**







# Dress-Up: Generating Animatable Clothed 3D Humans via Latent Modeling of 3D Gaussian Texture Maps



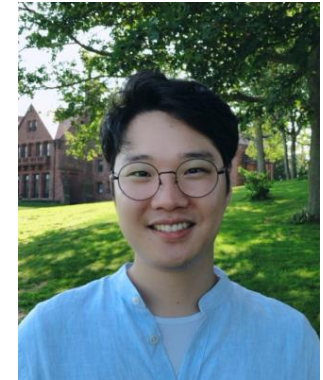
Kim Youwang



Lee Hyoseok



Gerard Pons-Moll



Tae-Hyun Oh