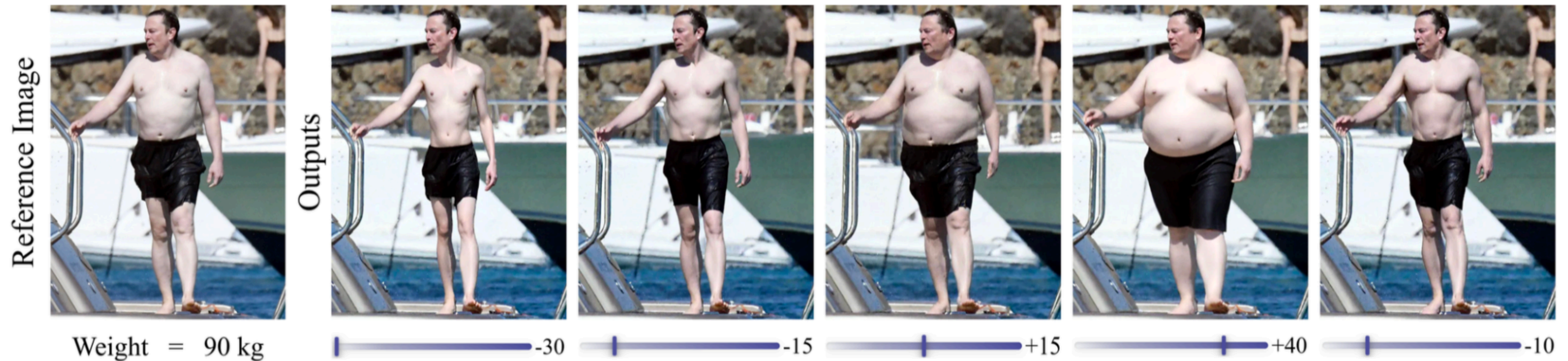


Odo: Depth-Guided Diffusion for Identity-Preserving Body Reshaping



Siddharth Khandelwal, Sridhar Kamath, Arjun Jain
Fast Code AI Consult Pvt. Ltd.

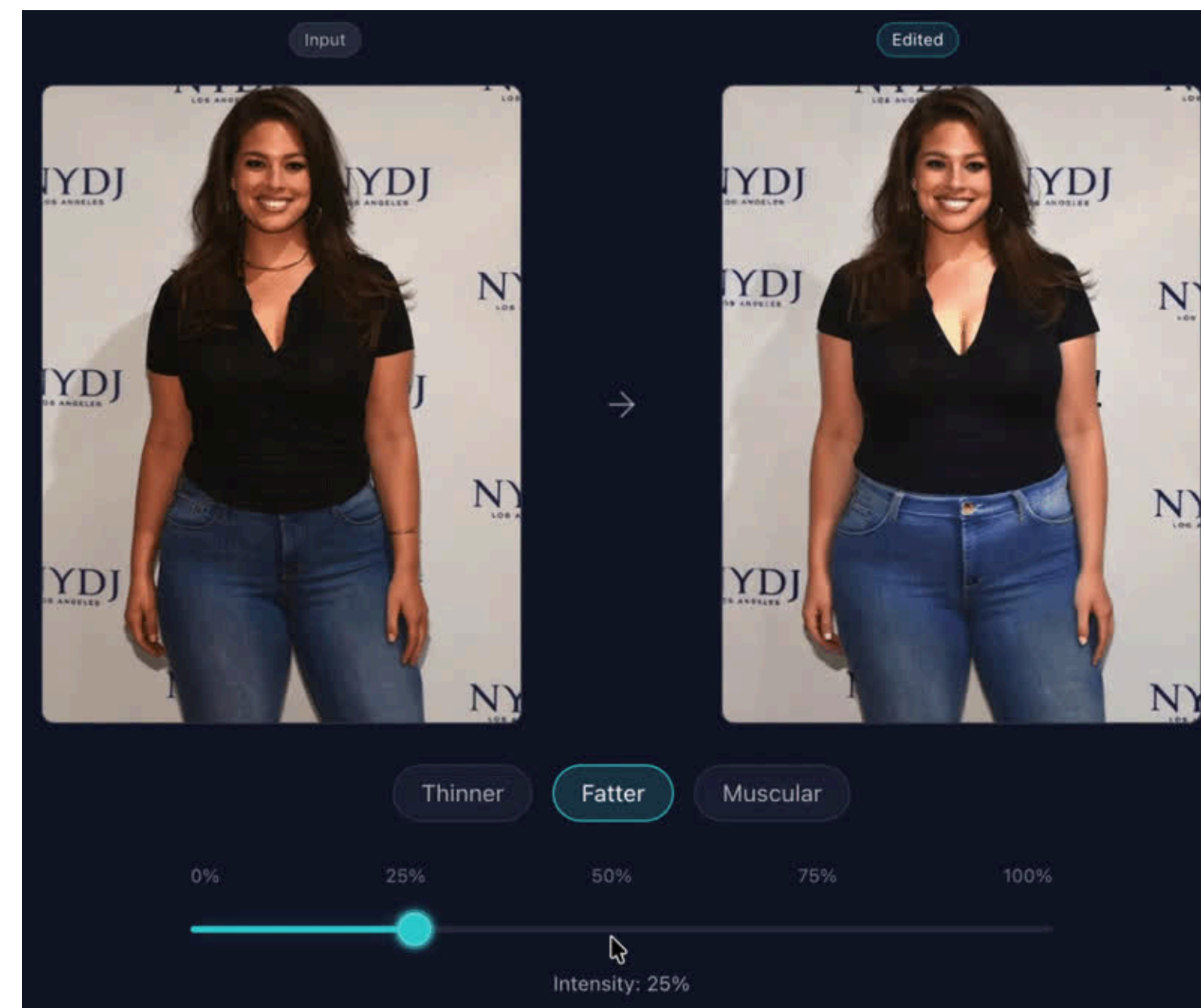


What is human shape editing?

Controllable human body-reshaping (fatter/thinner/more muscular) while preserving identity (face), pose, clothing, and background.

Applications

- Virtual Fashion
- Entertainment and digital avatars
- Health visualization (anorexia/bulimia)



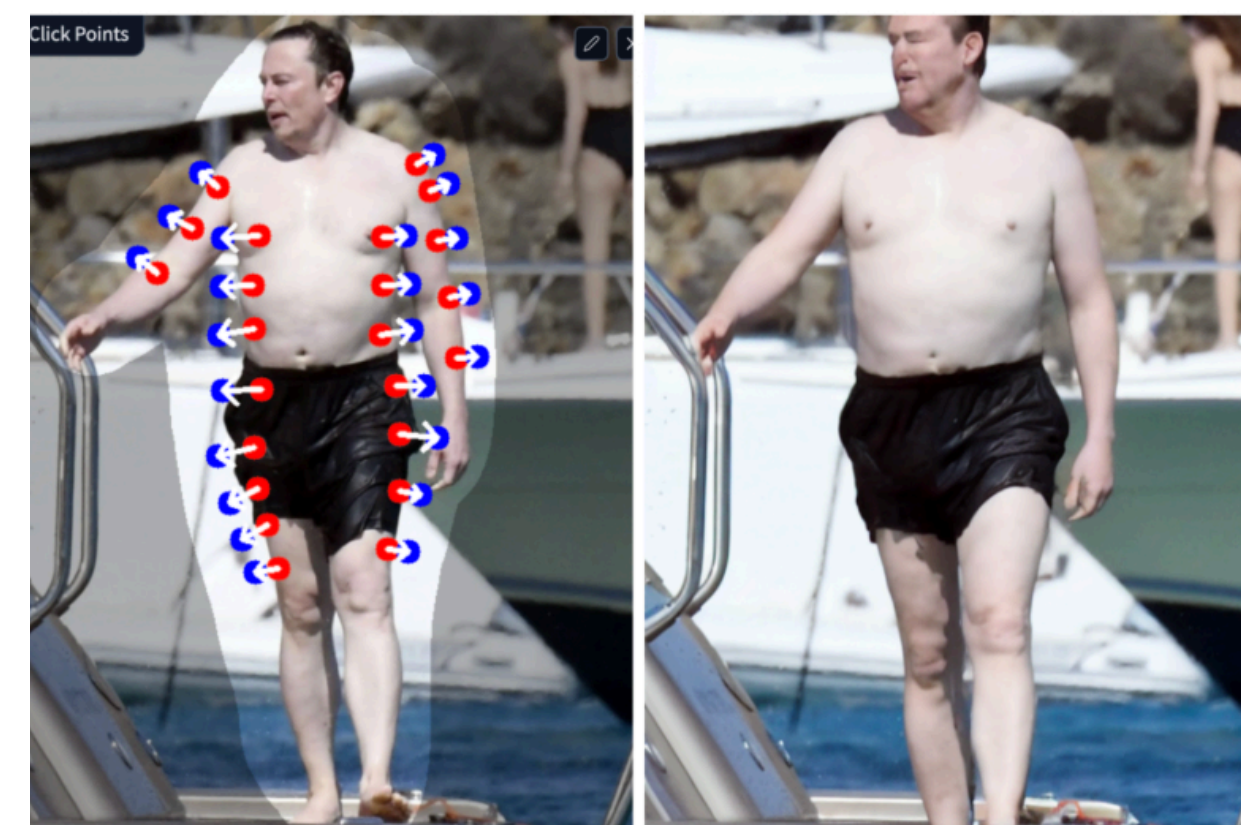


Current Methods

Existing methods for human body reshaping rely on **3D morphable models**, or **2D warping**.

Limitations:

1. Distortions in body and background
2. Manual intervention
3. Struggles with loose clothing
4. Poor generalization to varying shapes





Why is Human Body Reshaping Hard?

Human Body Pose-editing Datasets

- Large-scale
- Variety of poses
- Easy to create



Existing Pose Datasets

Human Body Reshaping Datasets

- Limited number of samples
- Less variations
- Practically impossible to create



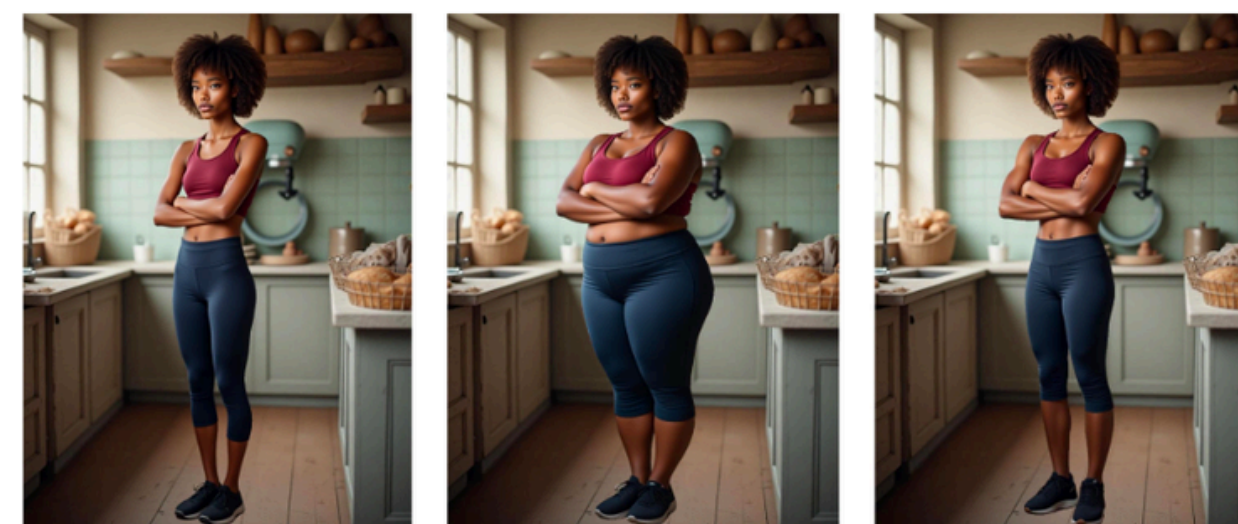
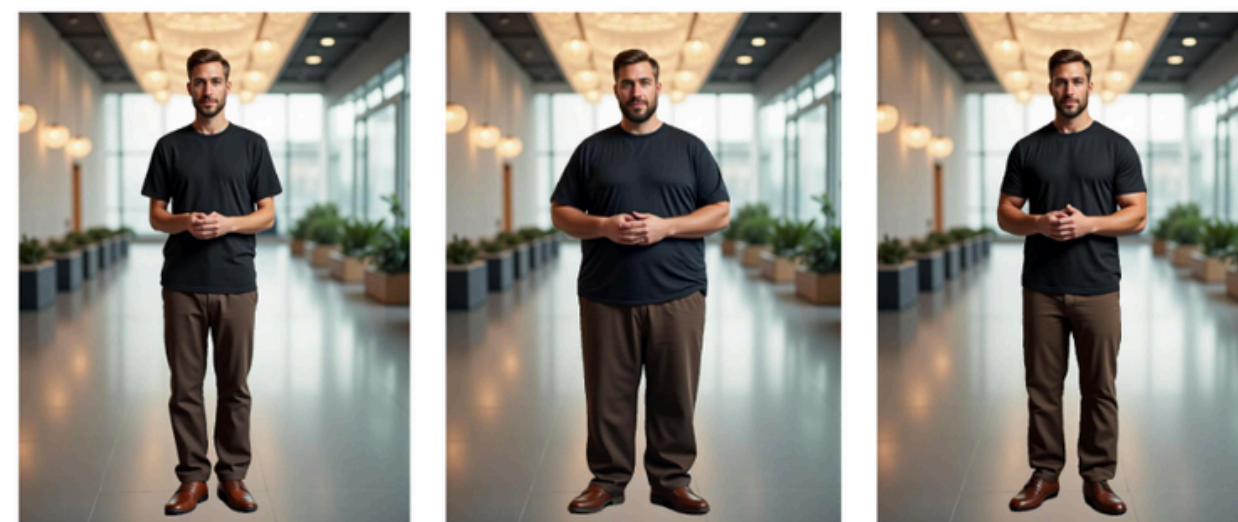
Existing Human Reshaping Dataset



ChangeLing18K

First Large-scale dataset for Human Body Reshaping

- **18573 total pairs across 1523 subjects**
- Triplets of Thin - Fat - Muscular
- Consistent Face Identity, Clothing, Pose and Background



Thin

Fat

Muscular



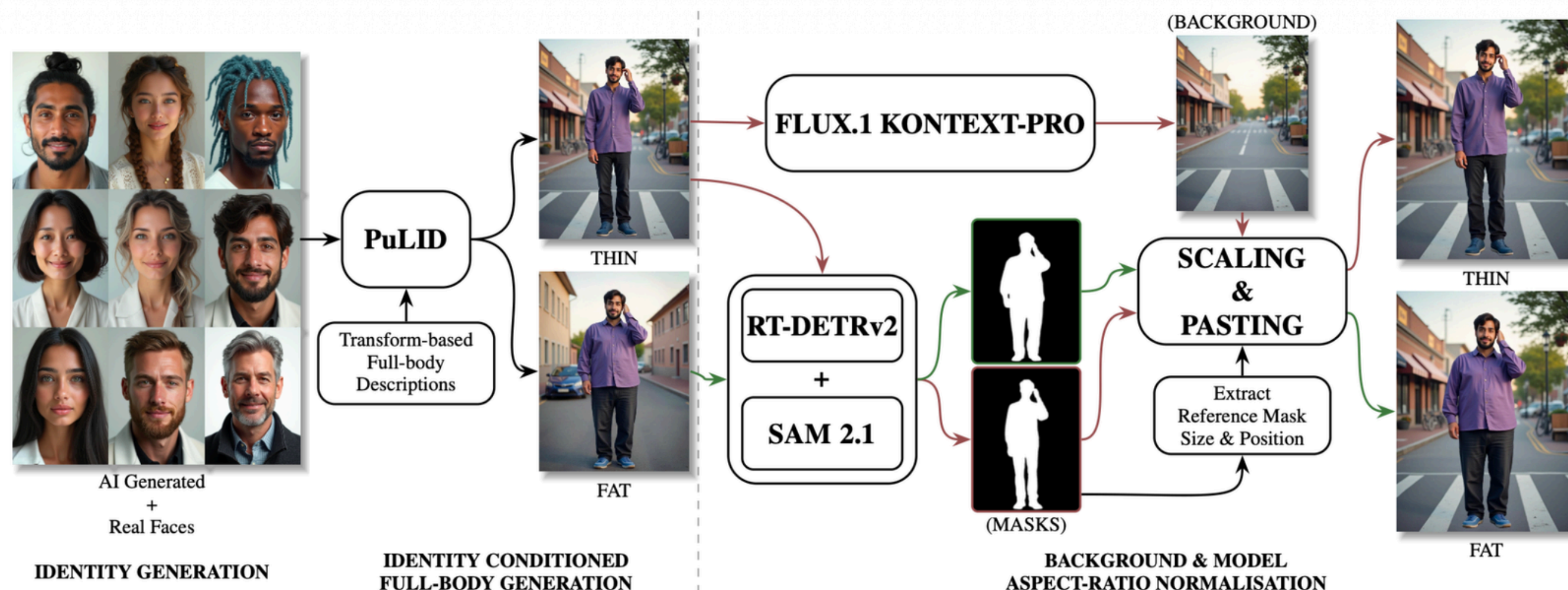
Dataset Generation Pipeline

4-stage, model-driven pipeline that guarantees identity fidelity, pose and attire consistency, and high-resolution realism across three body types: thin, fat, and muscular

1. Identity Creation

2. Identity-conditioned Full-body Generation

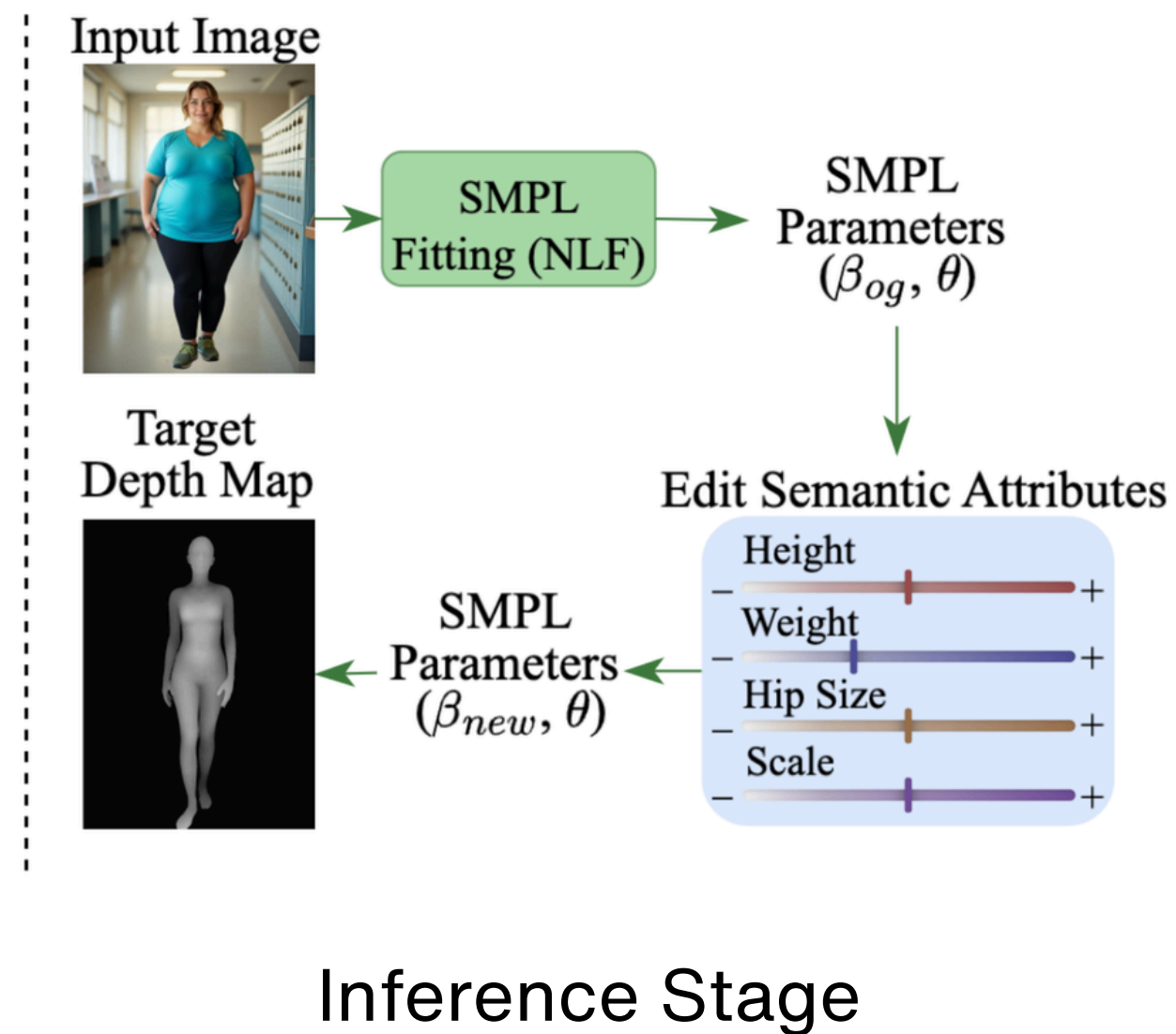
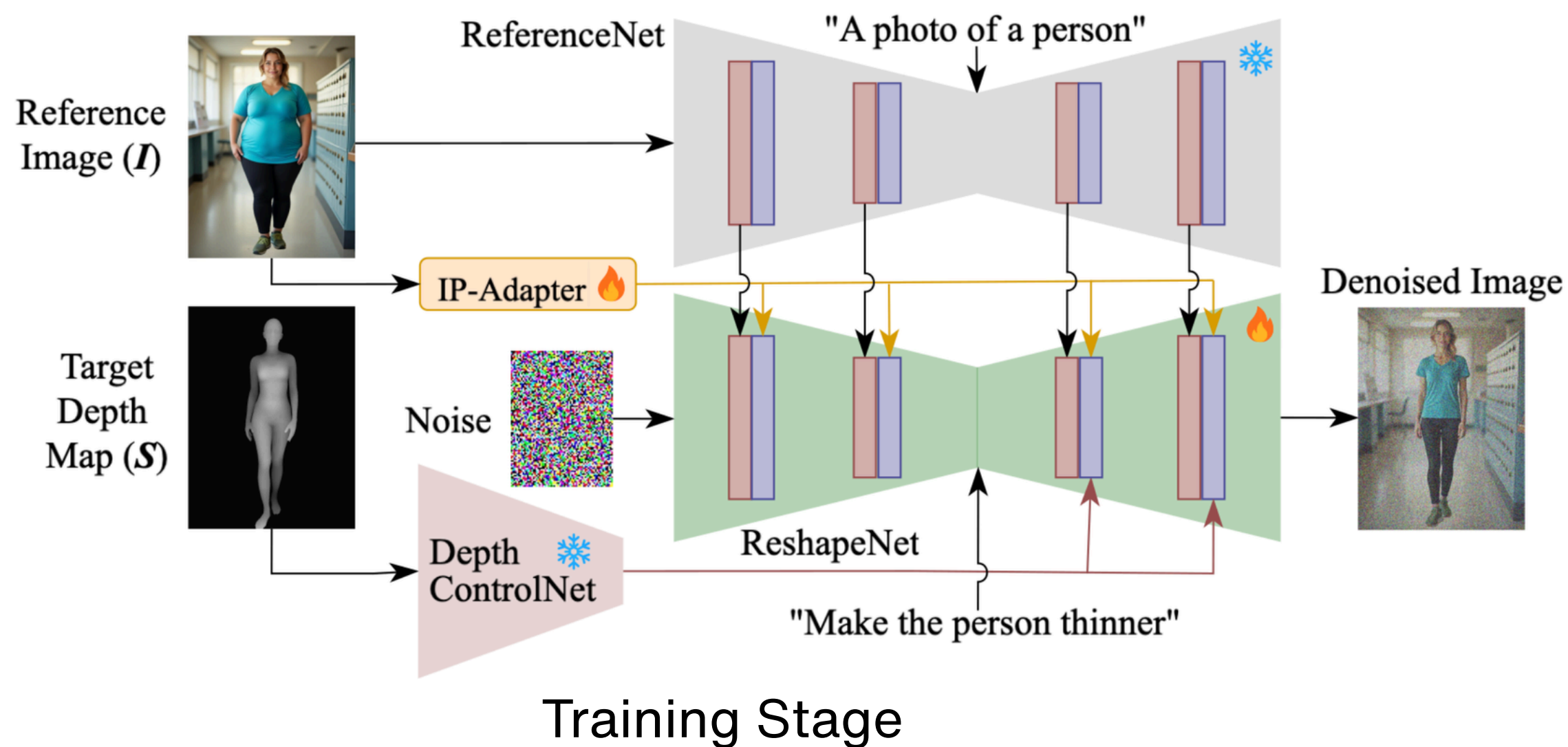
3. Background and Aspect-ratio Normalization 4. Manual Curation





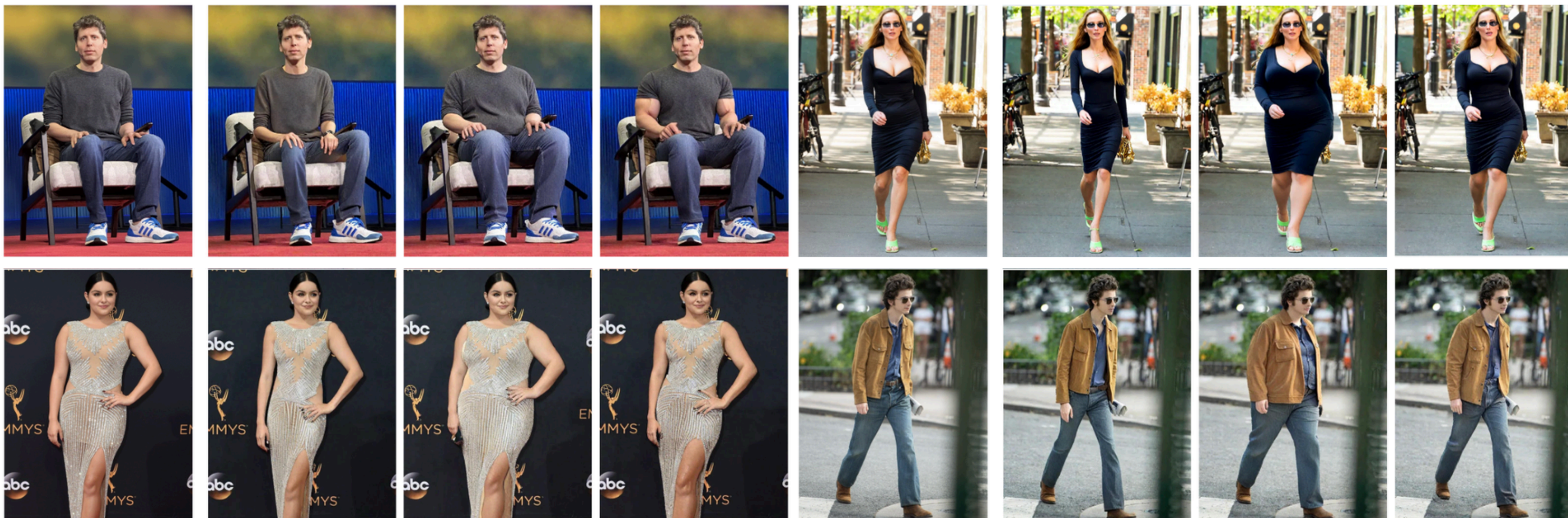
Odo

We introduce the first diffusion based model Odo trained on our dataset for controllable human body reshaping.



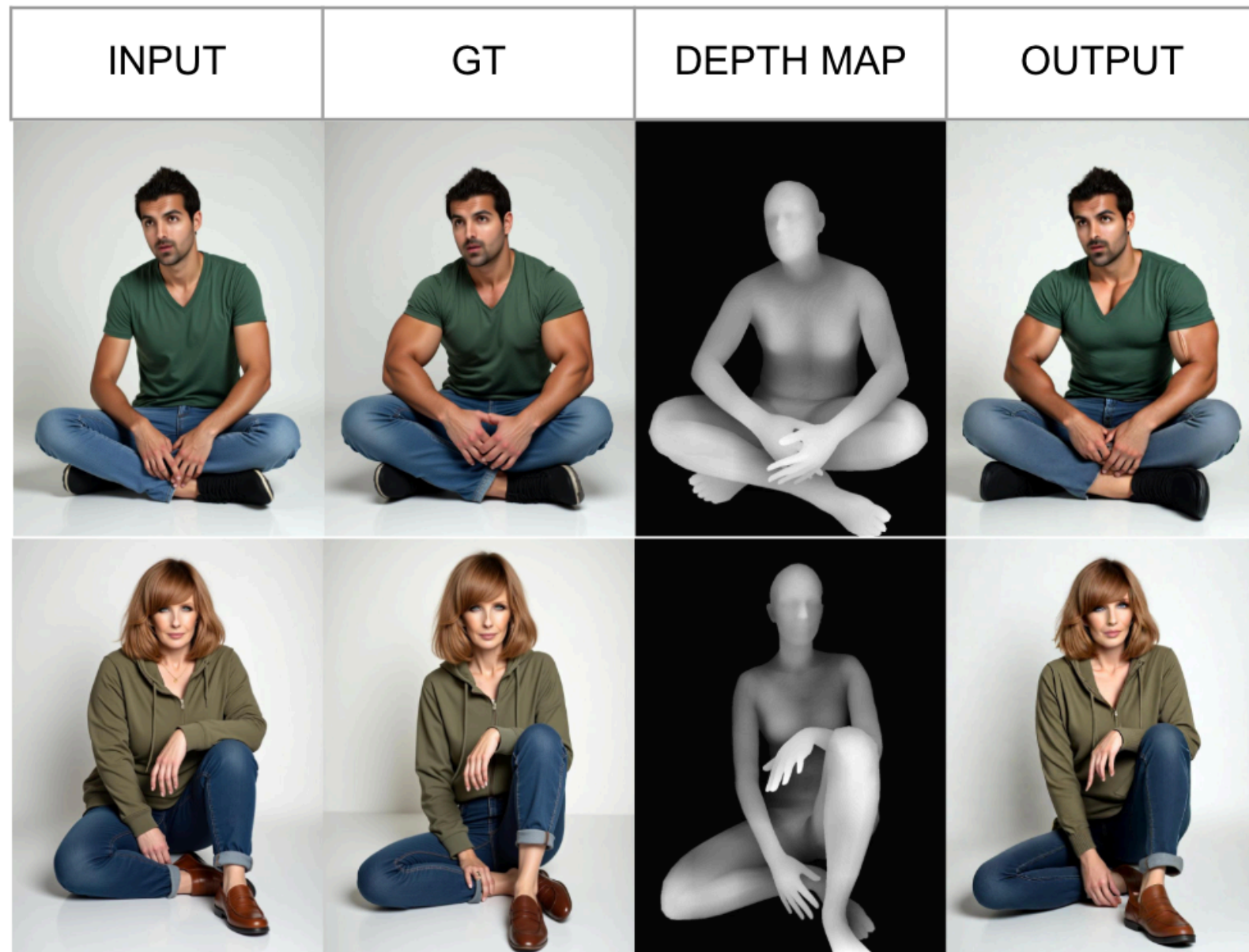


Qualitative Results on In-the-wild Images

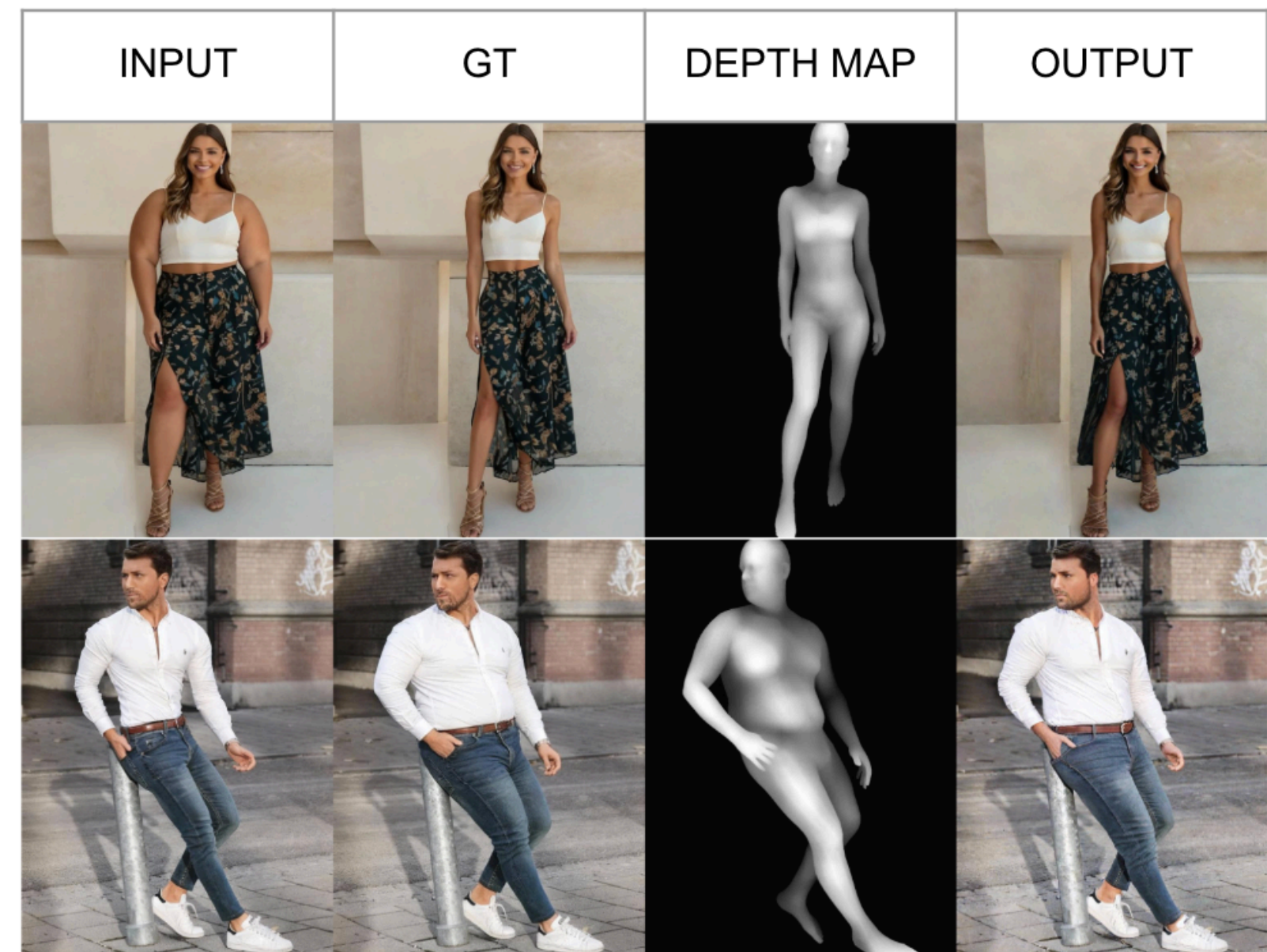




Qualitative Results on our Benchmark



Results

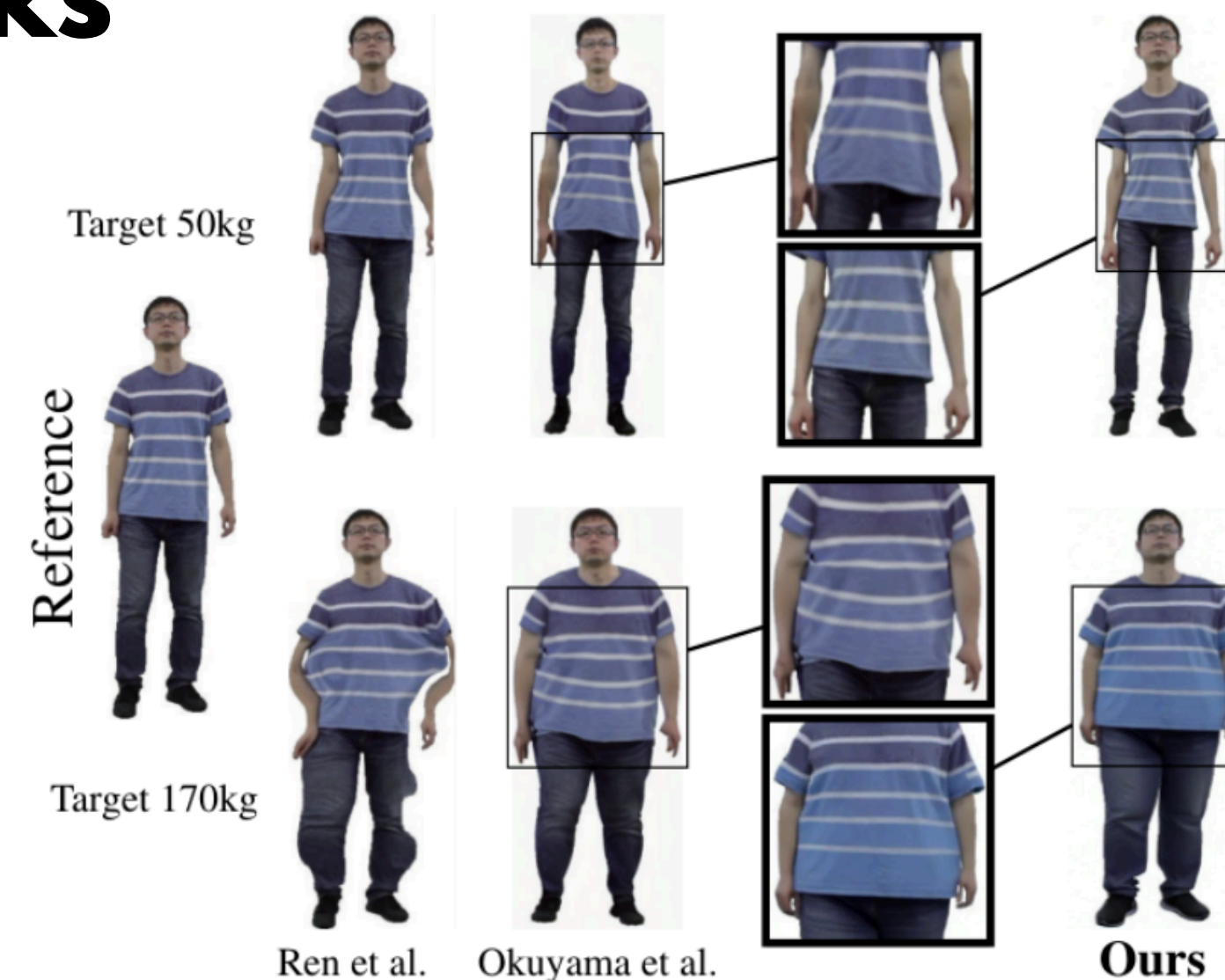


Results on Benchmark
generated using Seedream 4.0



Quantitative & Qualitative Comparison with other works

| <u>Method</u> | <u>SSIM</u> ↑ | <u>PSNR</u> ↑ | <u>LPIPS</u> ↓ | <u>PVE-T-SC</u> ↓ |
|-------------------------|---------------|----------------|----------------|----------------------|
| Ren et al. | 0.679 | 17.4567 | 0.2363 | 13.6337 |
| FLUX.1 Kontext [dev] | 0.6788 | 16.5195 | 0.2826 | 19.1911 |
| Ours | 0.7716 | 19.0714 | 0.2035 | 7.4902 |



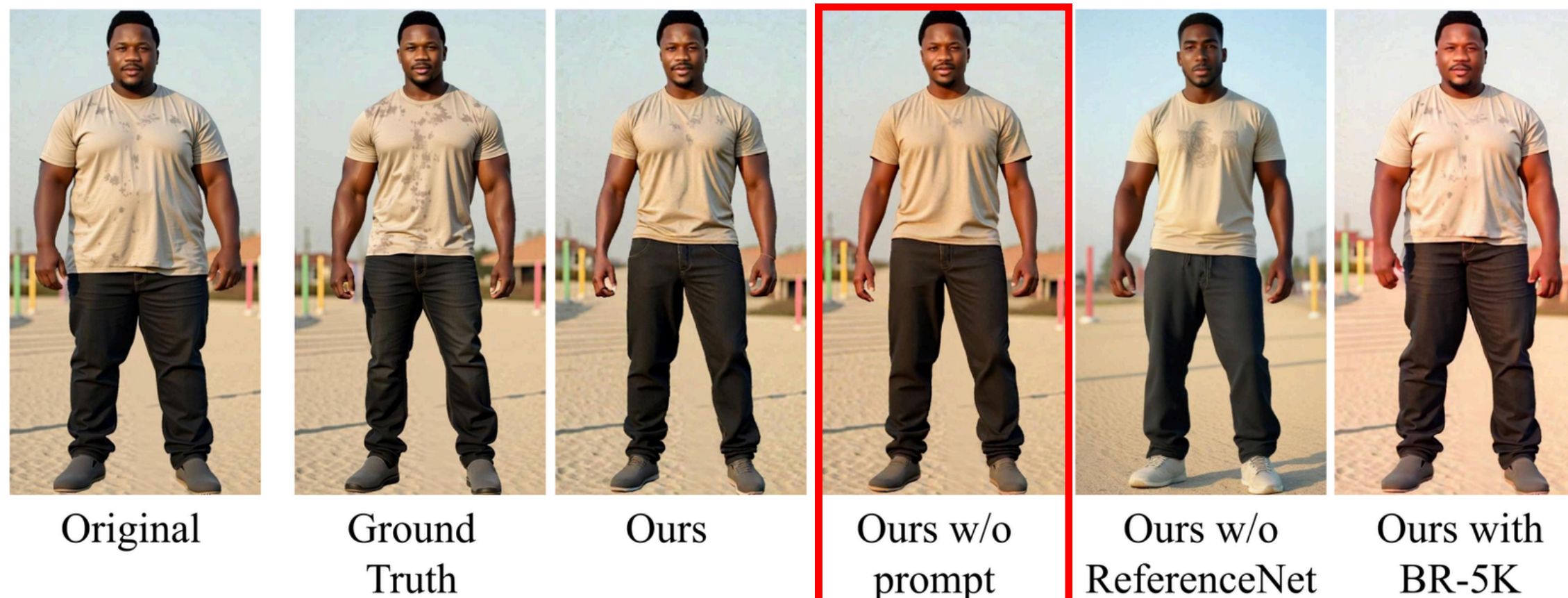
- Anatomical Consistency
- Texture Fidelity
- Zero Distortion

Jianqiang Ren et al., "Structure-aware flow generation for human body reshaping," CVPR 2022.

Yuta Okuyama et al., "Diff-Body: Diffusion-based pose and shape editing of human images," WACV 2024.



Ablation Studies



| <u>Method</u> | <u>SSIM</u> ↑ | <u>PSNR</u> ↑ | <u>LPIPS</u> ↓ | <u>PVE-T-SC</u> ↓ |
|-----------------------|---------------|---------------|----------------|-------------------|
| Ours | 0.7716 | 19.0714 | 0.2035 | 7.4902 |
| Ours w/o prompts | 0.7281 | 15.1745 | 0.2566 | 9.8032 |
| Ours w/o ReferenceNet | 0.6035 | 13.9982 | 0.4625 | 9.4157 |
| Ours with BR-5K data | 0.6939 | 14.1581 | 0.3432 | 18.6143 |



Ablation Studies



Original

Ground Truth

Ours

Ours w/o prompt

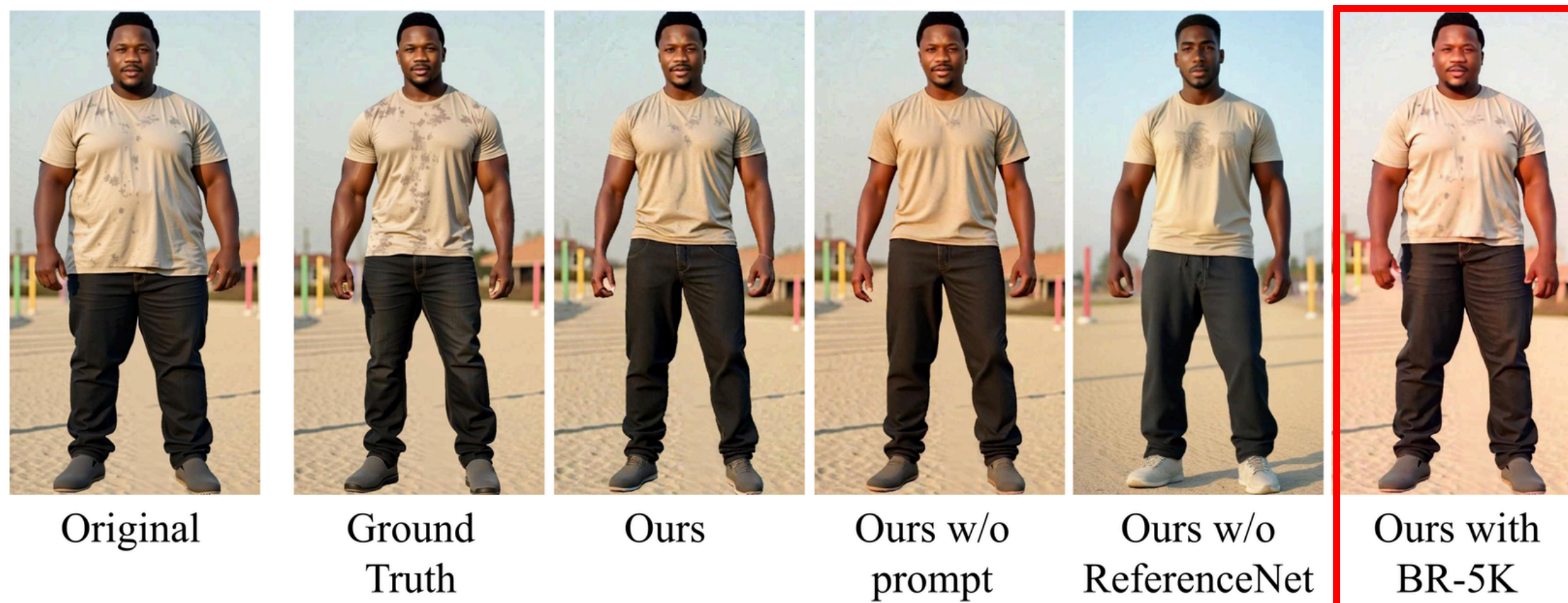
Ours w/o ReferenceNet

Ours with BR-5K

| <u>Method</u> | <u>SSIM</u> ↑ | <u>PSNR</u> ↑ | <u>LPIPS</u> ↓ | <u>PVE-T-SC</u> ↓ |
|-----------------------|---------------|---------------|----------------|-------------------|
| Ours | 0.7716 | 19.0714 | 0.2035 | 7.4902 |
| Ours w/o prompts | 0.7281 | 15.1745 | 0.2566 | 9.8032 |
| Ours w/o ReferenceNet | 0.6035 | 13.9982 | 0.4625 | 9.4157 |
| Ours with BR-5K data | 0.6939 | 14.1581 | 0.3432 | 18.6143 |



Ablation Studies



| <u>Method</u> | <u>SSIM</u> ↑ | <u>PSNR</u> ↑ | <u>LPIPS</u> ↓ | <u>PVE-T-SC</u> ↓ |
|-----------------------|---------------|----------------|----------------|-------------------|
| Ours | 0.7716 | 19.0714 | 0.2035 | 7.4902 |
| Ours w/o prompts | 0.7281 | 15.1745 | 0.2566 | 9.8032 |
| Ours w/o ReferenceNet | 0.6035 | 13.9982 | 0.4625 | 9.4157 |
| Ours with BR-5K data | 0.6939 | 14.1581 | 0.3432 | 18.6143 |



Conclusion

Key Contributions of our work:

1. **ChangeLing18K Dataset** - 1st large-scale, high-resolution dataset of 18,573 body-shape pairs featuring consistent identity, clothing, and backgrounds.
2. **Odo** - End-to-end diffusion-based model for photorealistic, identity-preserving human body reshaping.
 - a. **SOTA performance** - achieves least geometric reconstruction error (7.5mm PVE-T-SC) compared to prior warping-based baselines.

We are publicly releasing our dataset to break the paired-data bottleneck and accelerate future research in robust human shape editing.



Thank You



Scan to learn more or visit
<https://research.fastcode.ai/odo>