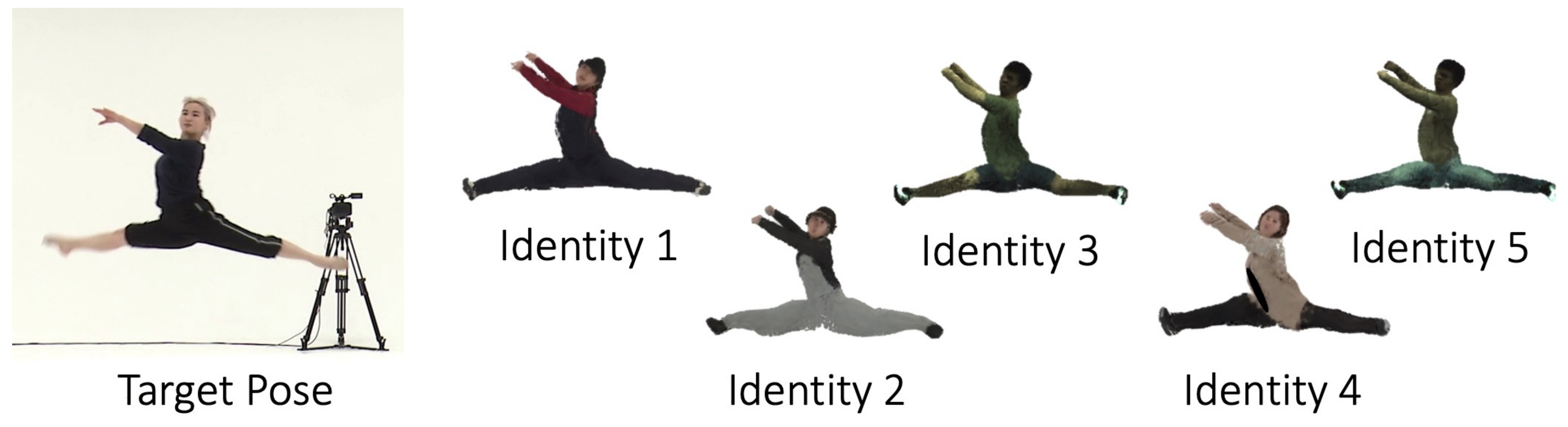
MIGS: Multi-Identity Gaussian Splatting via Tensor Decomposition

Aggelina Chatziagapi, Grigorios G. Chrysos, Dimitris Samaras

EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

Motivation



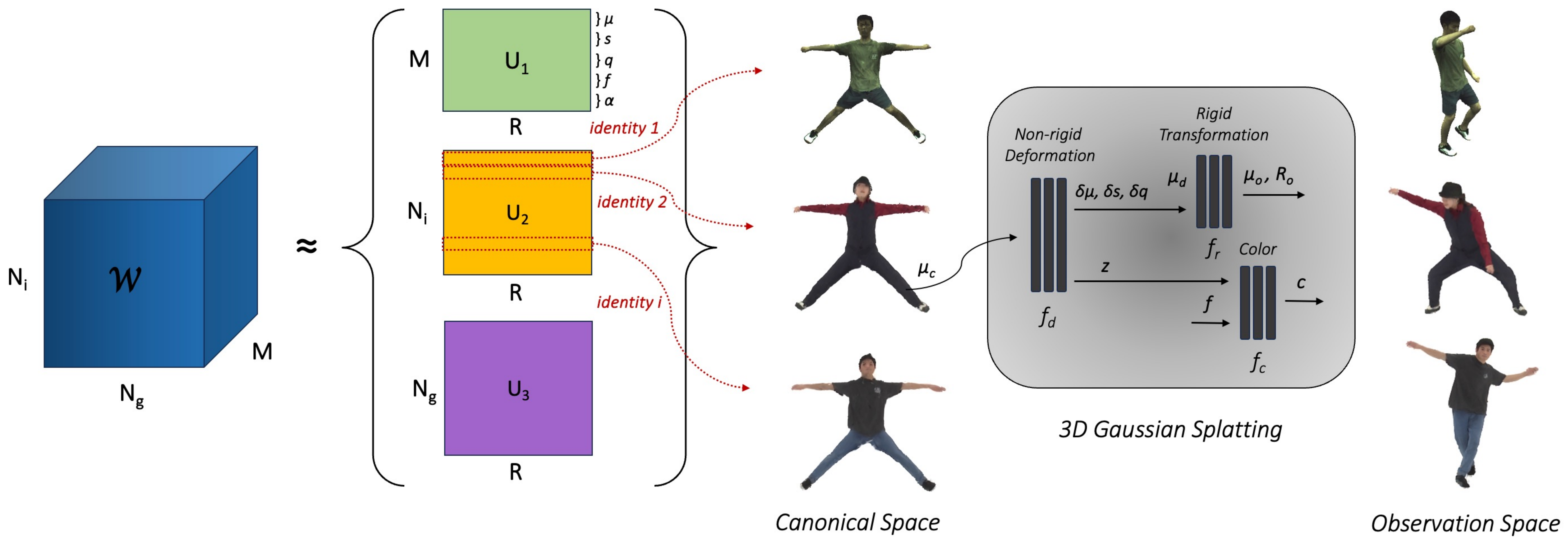
Target Pose
(out of the distribution)

- ✓ **Unified multi-identity neural representation** of human avatars
- ✓ **Robust animation** under novel poses, out of the training distribution
- ✓ **Significant decrease** in the total number of learnable parameters

Single-Identity Models (Standard) | Multi-Identity Model (Ours)

Standard approaches learn multiple single-identity representations. We learn a ***single*** representation for ***multiple*** identities from monocular videos.

Method



Tensor $\mathcal{W}$ | CP Decomposition | Multi-Identity Training

Multi-Representation with 3D Gaussian Splatting

For each identity *i*, we learn a set of 3D Gaussians $\{G^{(i,g)}\}_{g=1}^{N_g}$. Each Gaussian is associated with a position $\boldsymbol{\mu}^{(i,g)} \in \mathbb{R}^3$, a scaling vector $\boldsymbol{s}^{(i,g)} \in \mathbb{R}^3$, a quartenion $\boldsymbol{q}^{(i,g)} \in \mathbb{R}^4$, a feature vector $\boldsymbol{f}^{(i,g)} \in \mathbb{R}^{32}$, and an opacity $\alpha^{(i,g)} \in \mathbb{R}$. Given N_i identities, we construct a tensor:

$$\boxed{\boldsymbol{\mathcal{W}} \in \mathbb{R}^{N_i \times N_g \times M}, \text{where } \boldsymbol{w}_{i,g,:} = \left[\boldsymbol{\mu}^{(i,g)}; \boldsymbol{s}^{(i,g)}; \boldsymbol{q}^{(i,g)}; \boldsymbol{f}^{(i,g)}; \alpha^{(i,g)}\right]}$$

We assume a low-rank structure and learn a **CP Tensor Decomposition** [1] with rank R:

$$\boxed{\boldsymbol{W}_{(2)} \approx \boldsymbol{U}_3 \left(\boldsymbol{U}_2 \odot \boldsymbol{U}_1\right)^T}, \quad \text{where} \quad \boldsymbol{W}_{(2)} \in \mathbb{R}^{N_g \times (N_i M)}$$

Learnable parameters: $\boldsymbol{U}_1 \in \mathbb{R}^{M \times R}$, $\boldsymbol{U}_2 \in \mathbb{R}^{N_i \times R}$, $\boldsymbol{U}_3 \in \mathbb{R}^{N_g \times R}$

i.e., ***($M + N_i + N_g$)R* parameters** instead of *$M N_i N_g$*.

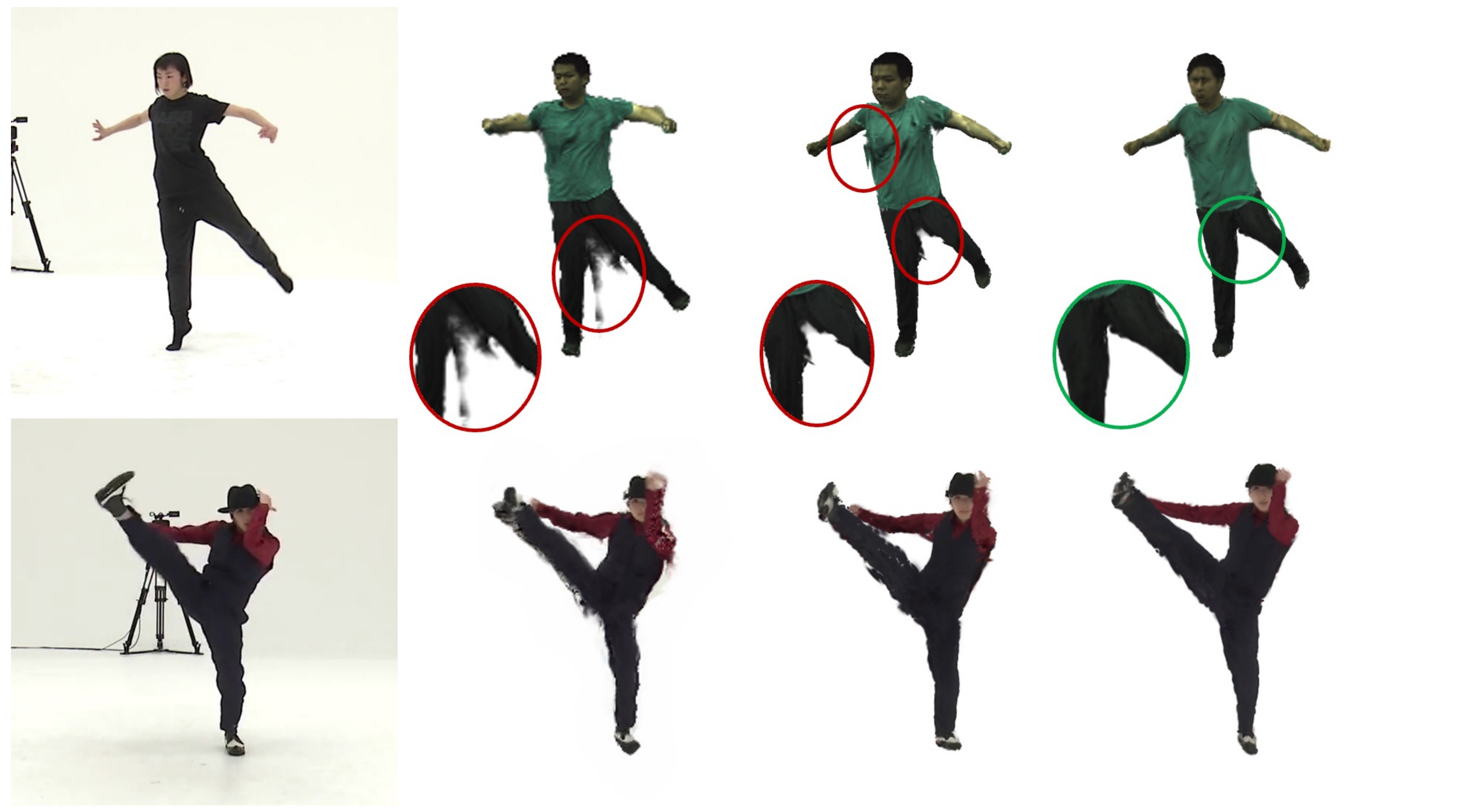
For N_i = 30, N_g = 5 x 10^4, M = 43, R = 100, MIGS learns 5 x 10^6 instead of 6.5 x 10^7 parameters, leading to a decrease by at least one order of magnitude.

References

[1] Kolda, T.G. and Bader, B.W., Tensor decompositions and applications, SIAM Review, 2009

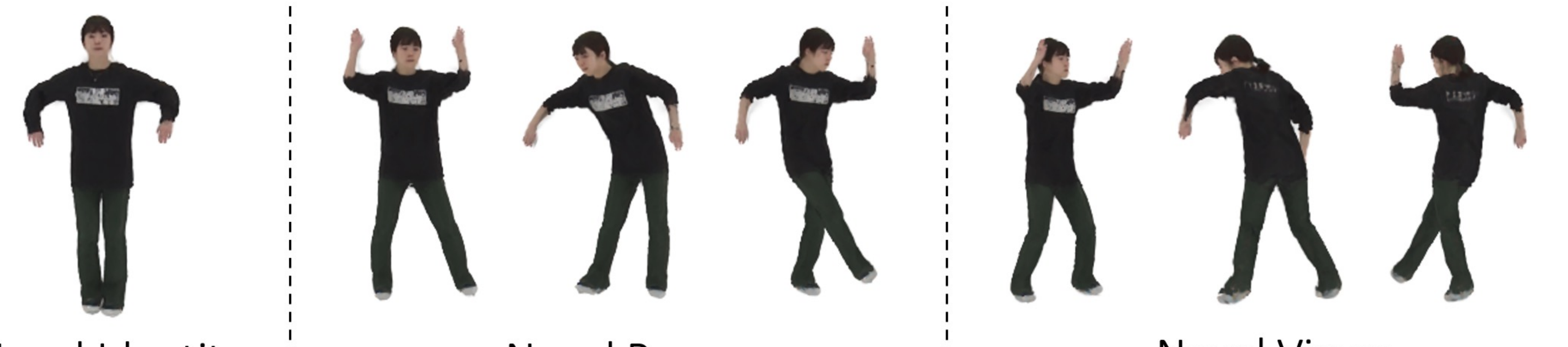
[2] Qian, Z. et al., 3DGS-Avatar: Animatable Avatars via Deformable 3D Gaussian Splatting, CVPR, 2024

Animation under Novel Poses



Target Pose | HumanNeRF | 3DGS-Avatar | MIGS (Ours)

Novel Identity



Novel Identity | Novel Poses | Novel Views

Ablation Study on Rank *R and N_i*