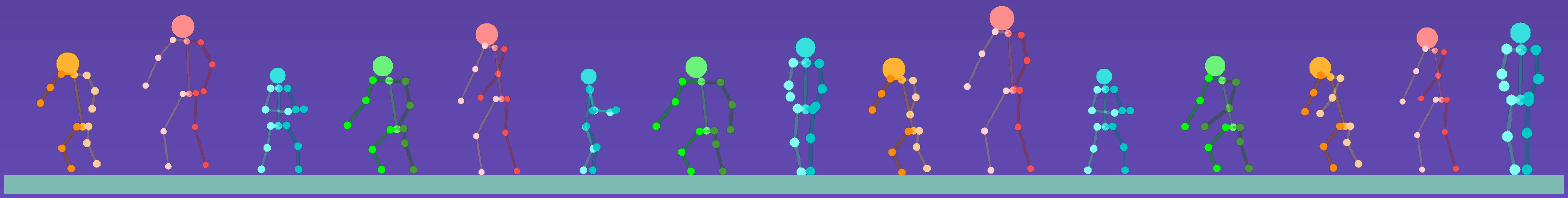
thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

Learning Character-Agnostic Motion for Motion Retargeting in 2D

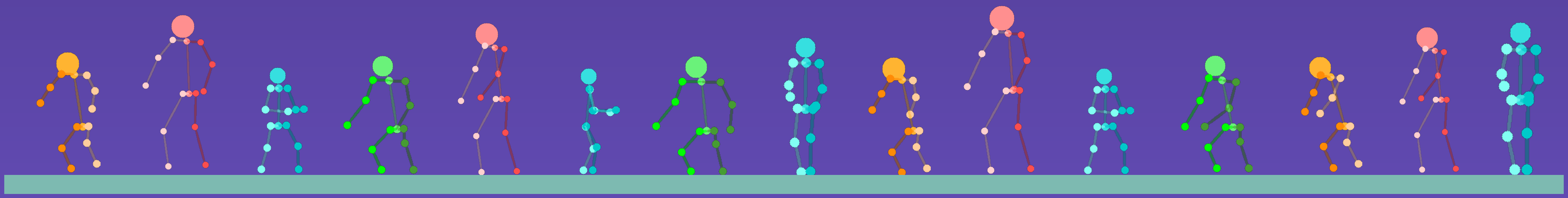
Kfir Aberman, Rundi Wu, Dani Lischinski,
Baoquan Chen, Daniel Cohen-Or



thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

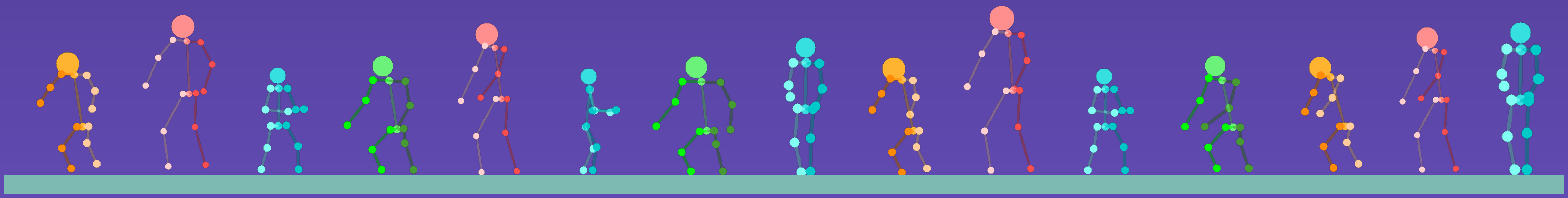
Outline

- Motivation
- Approach
- Results
- Application

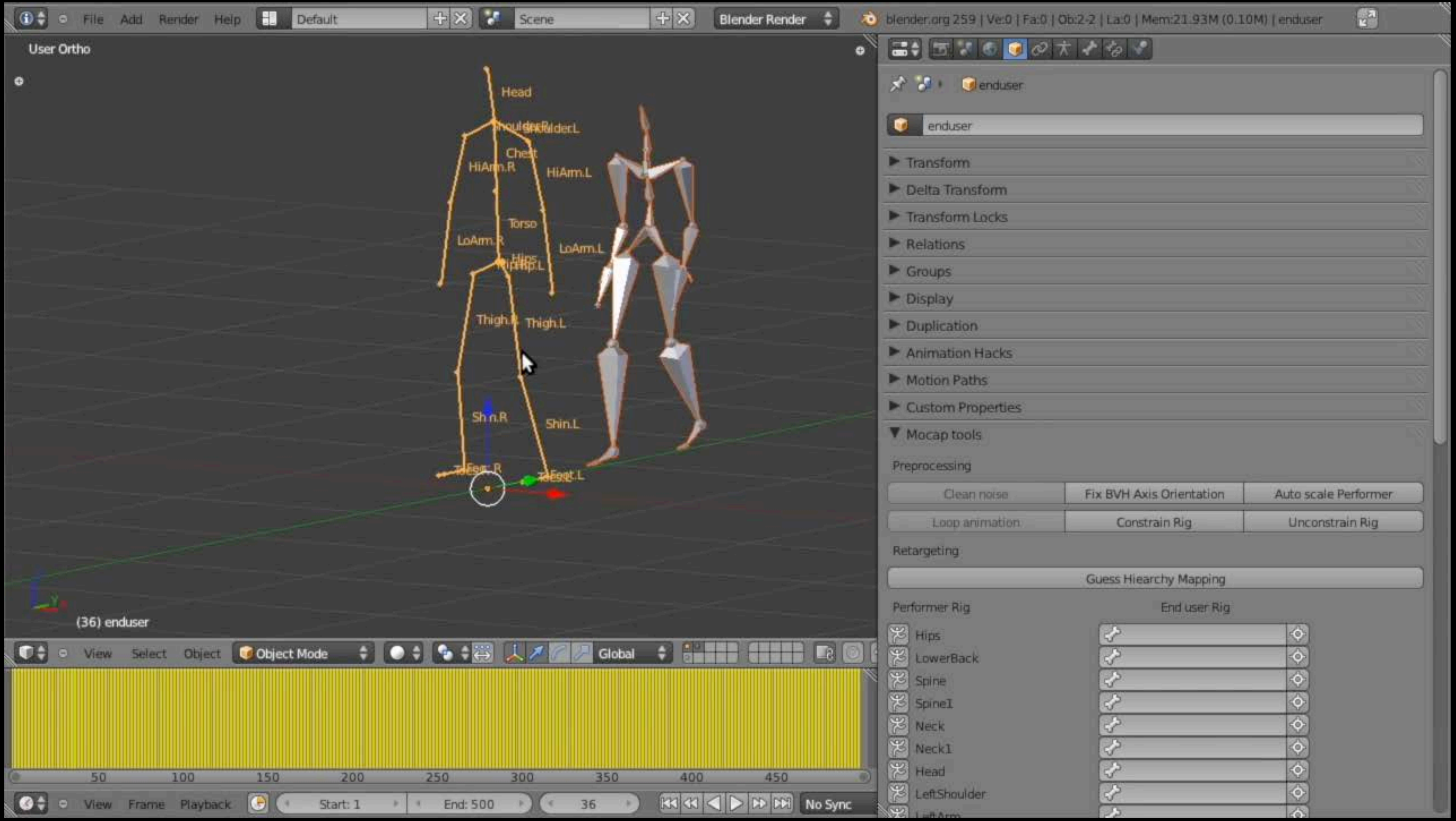


Outline

- Motivation
- Approach
- Results
- Application

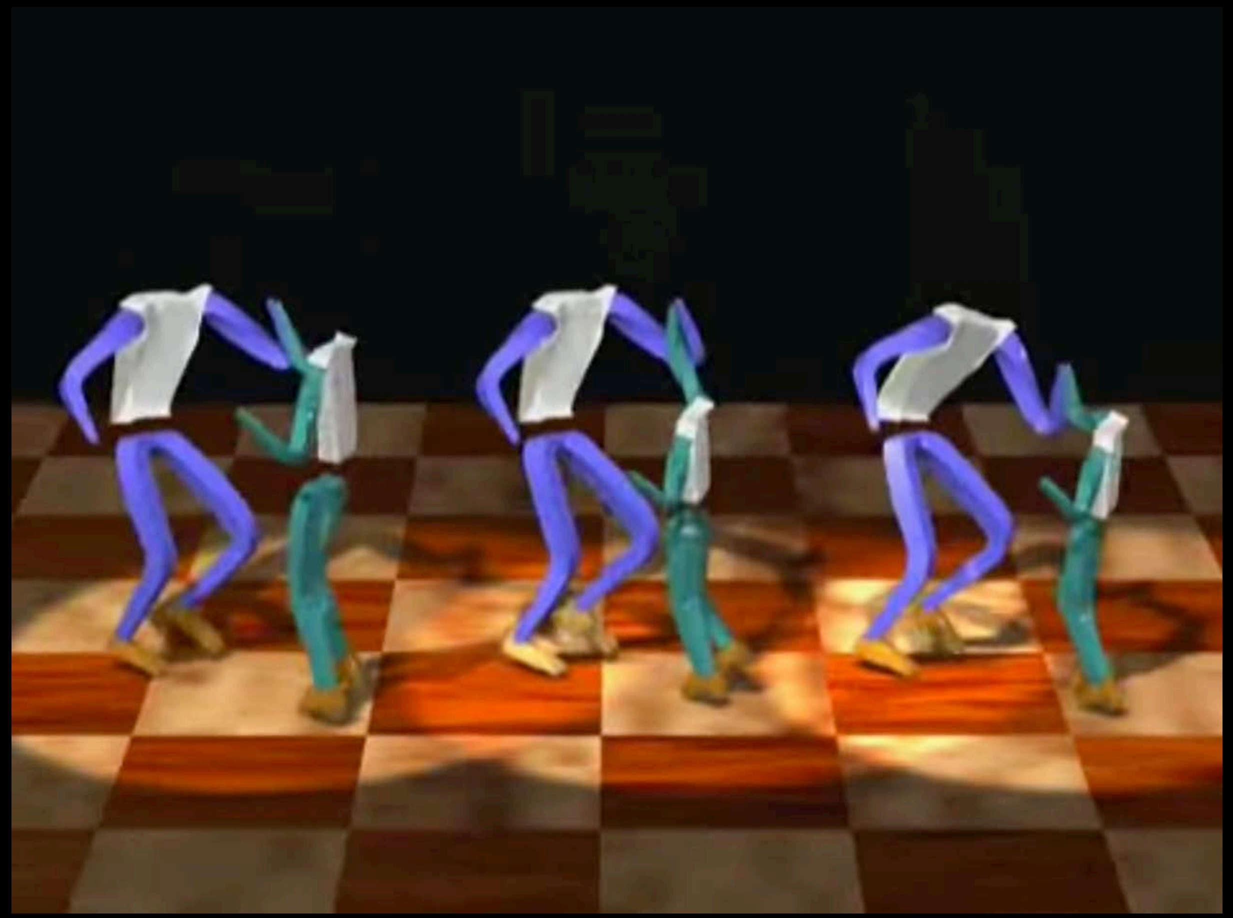


Motion Retargeting in 3D

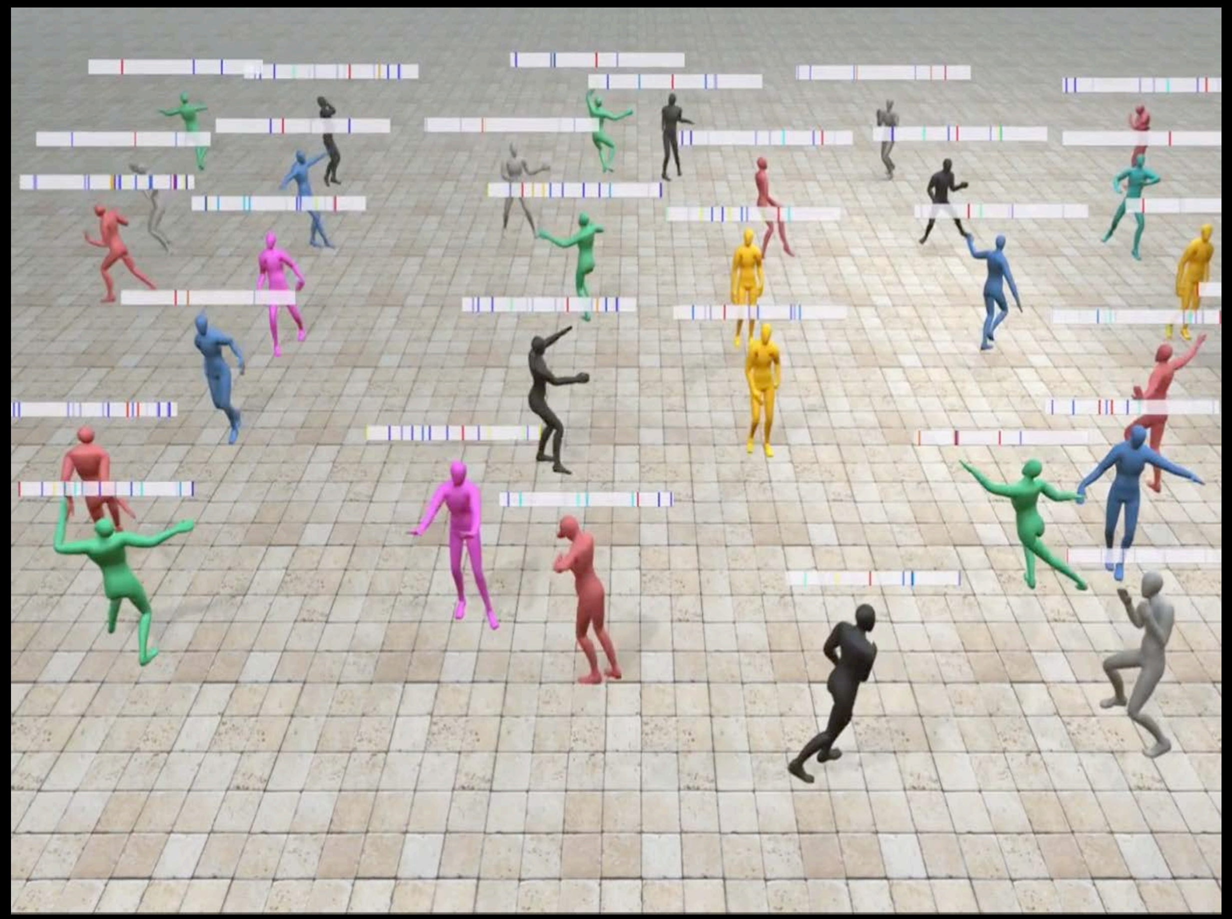


Related Work

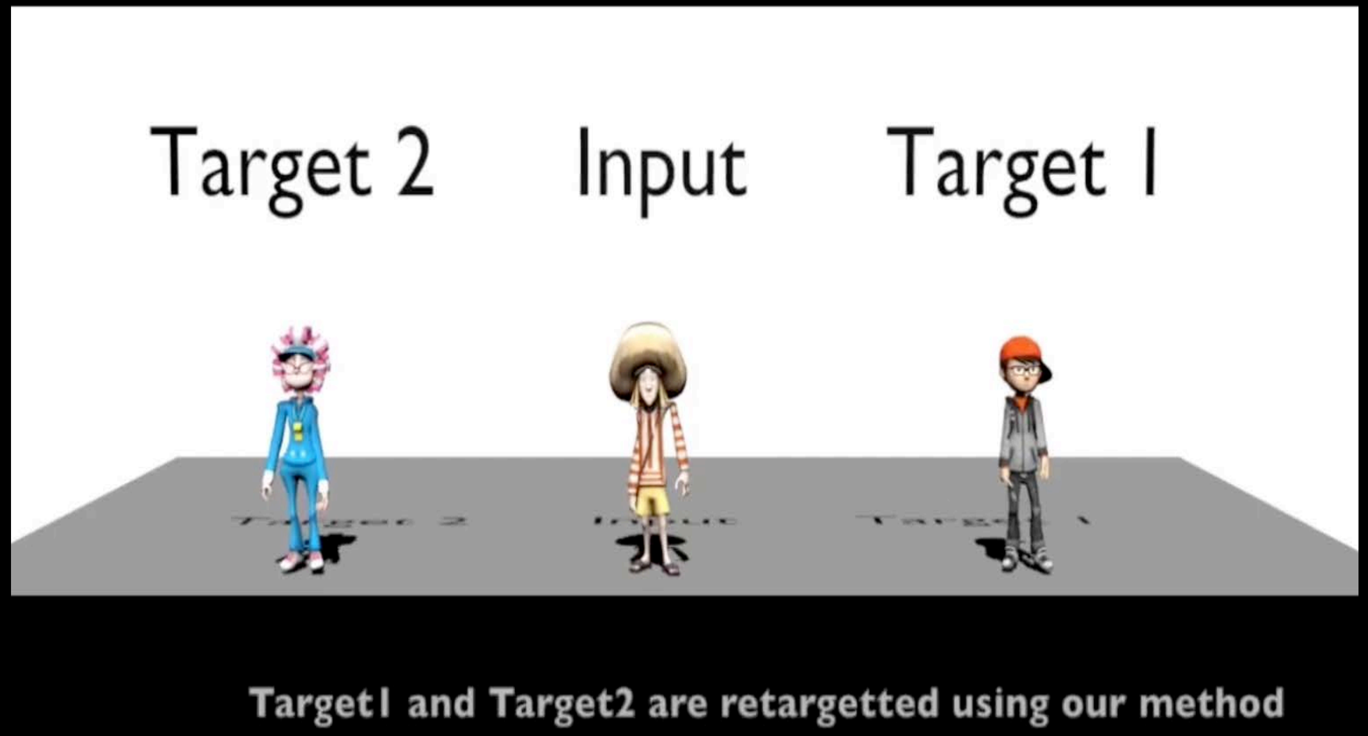
thrive SIGGRAPH2019 LOS ANGELES • 28 JULY – 1 AUGUST



[Gleicher et. al., 1998]

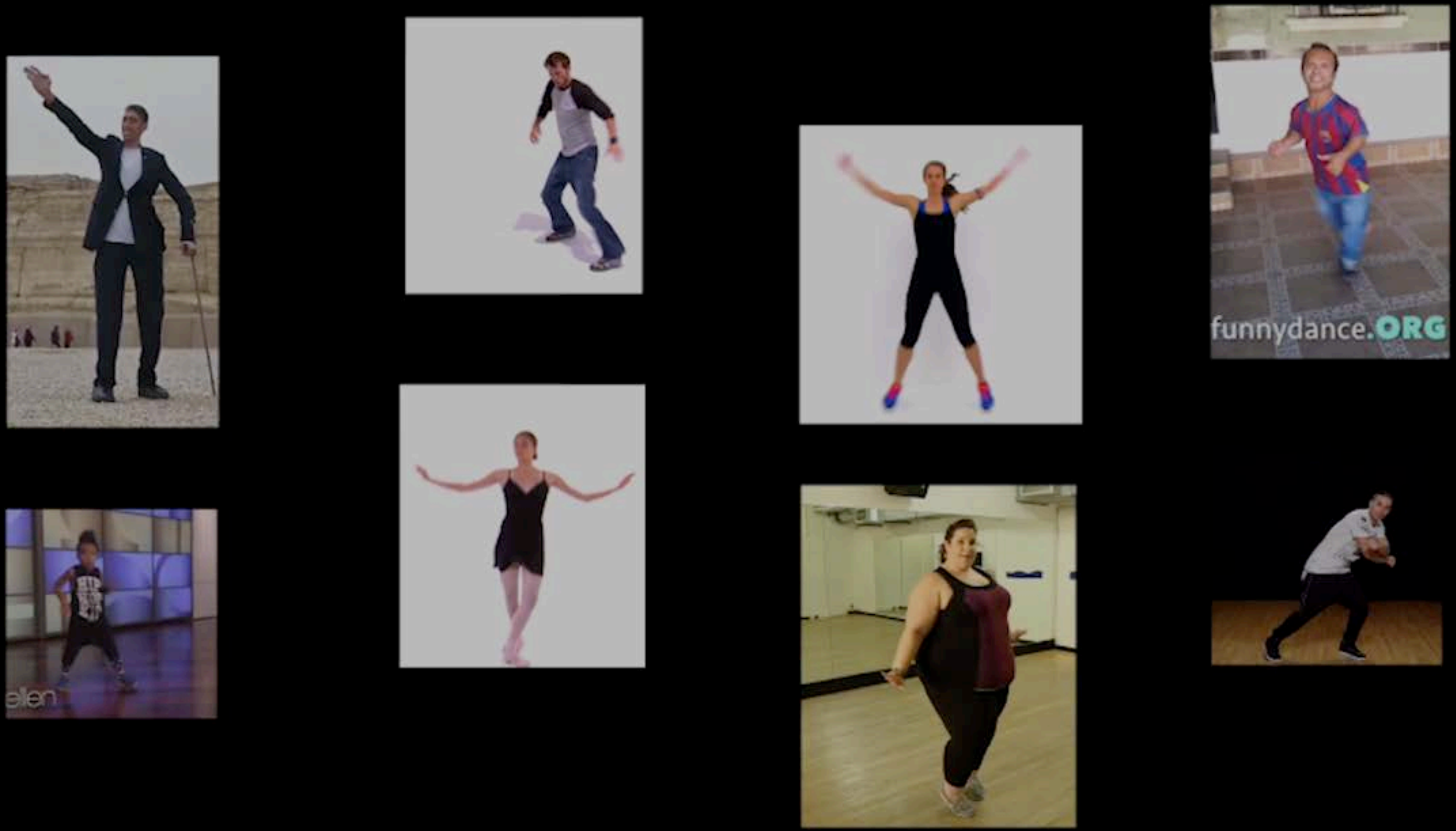


[Aristidou et.al., 2018]

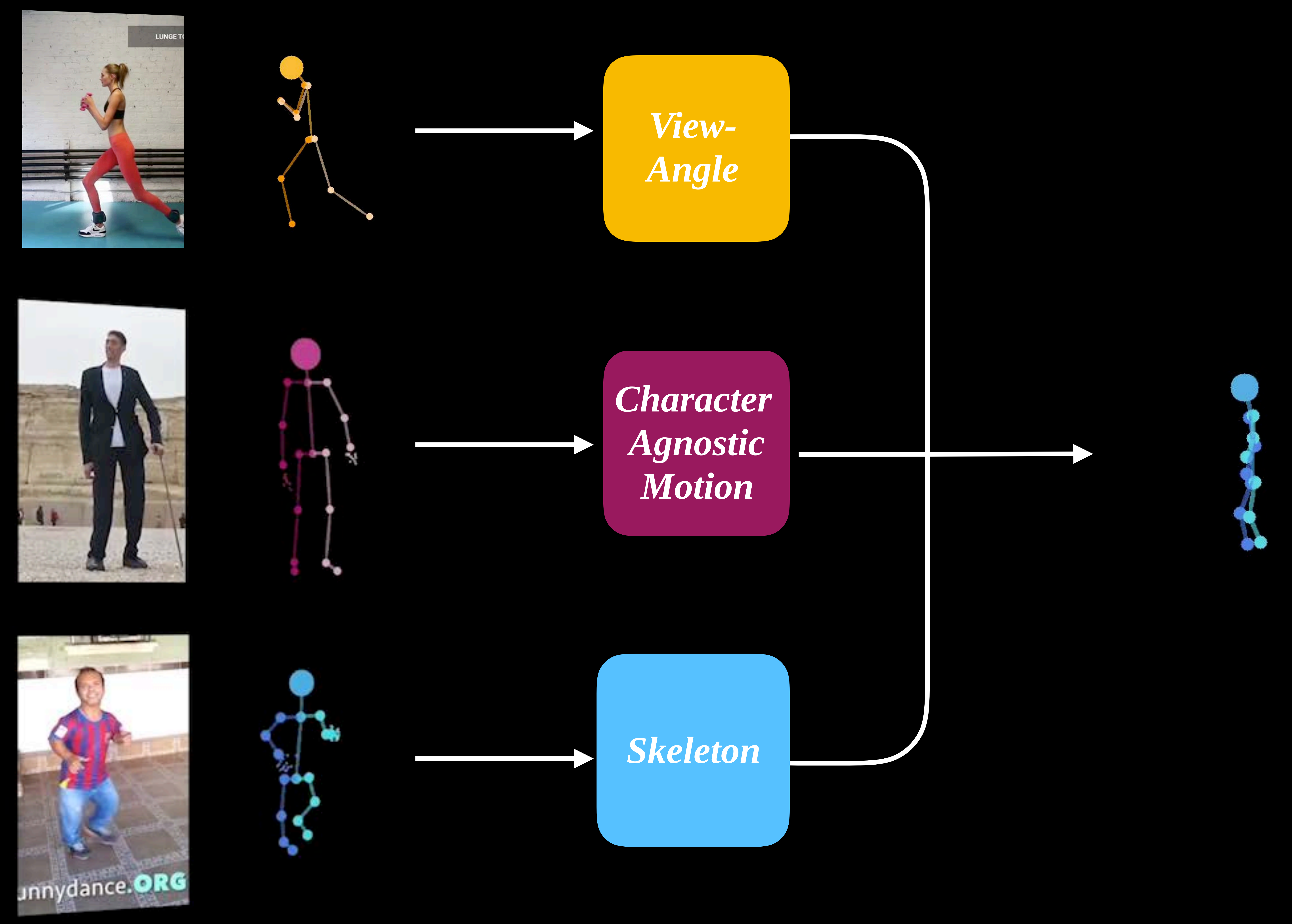


[Villegas et.al., 2018]

Motivation

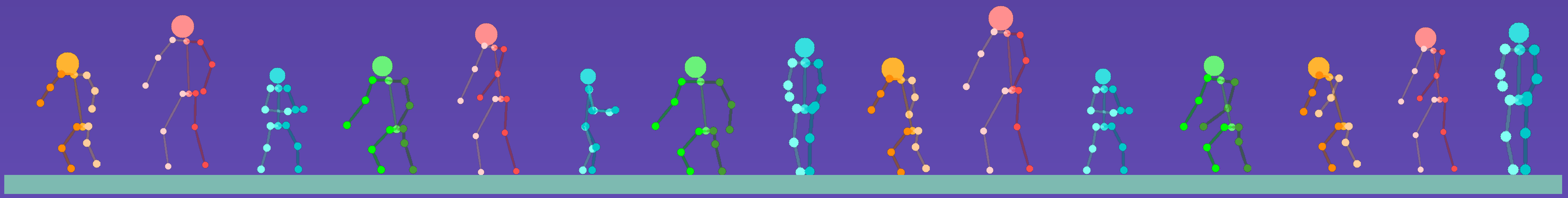


Motion Retargeting in 2D

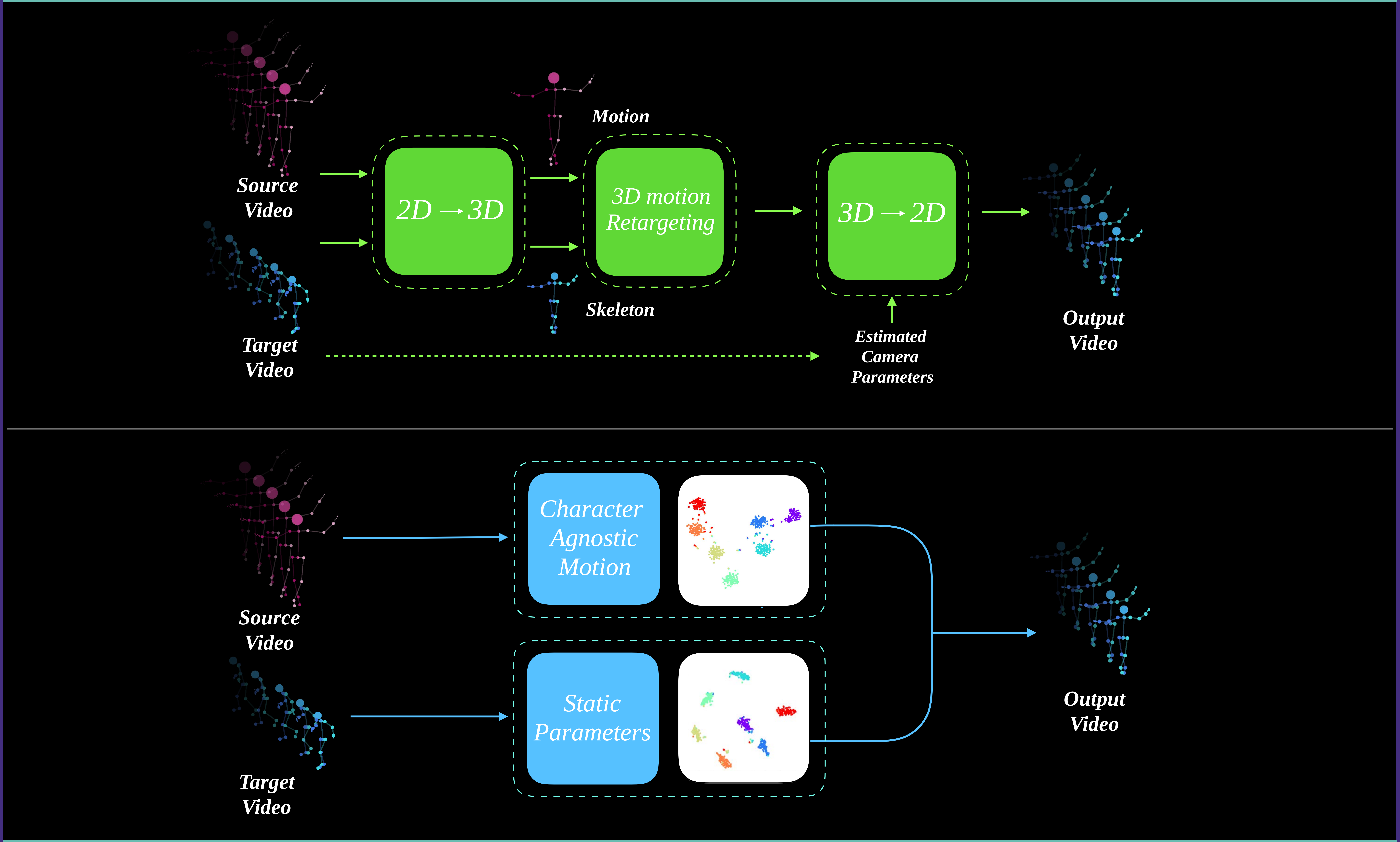


Outline

- Motivation
- **Approach**
- Results
- Application

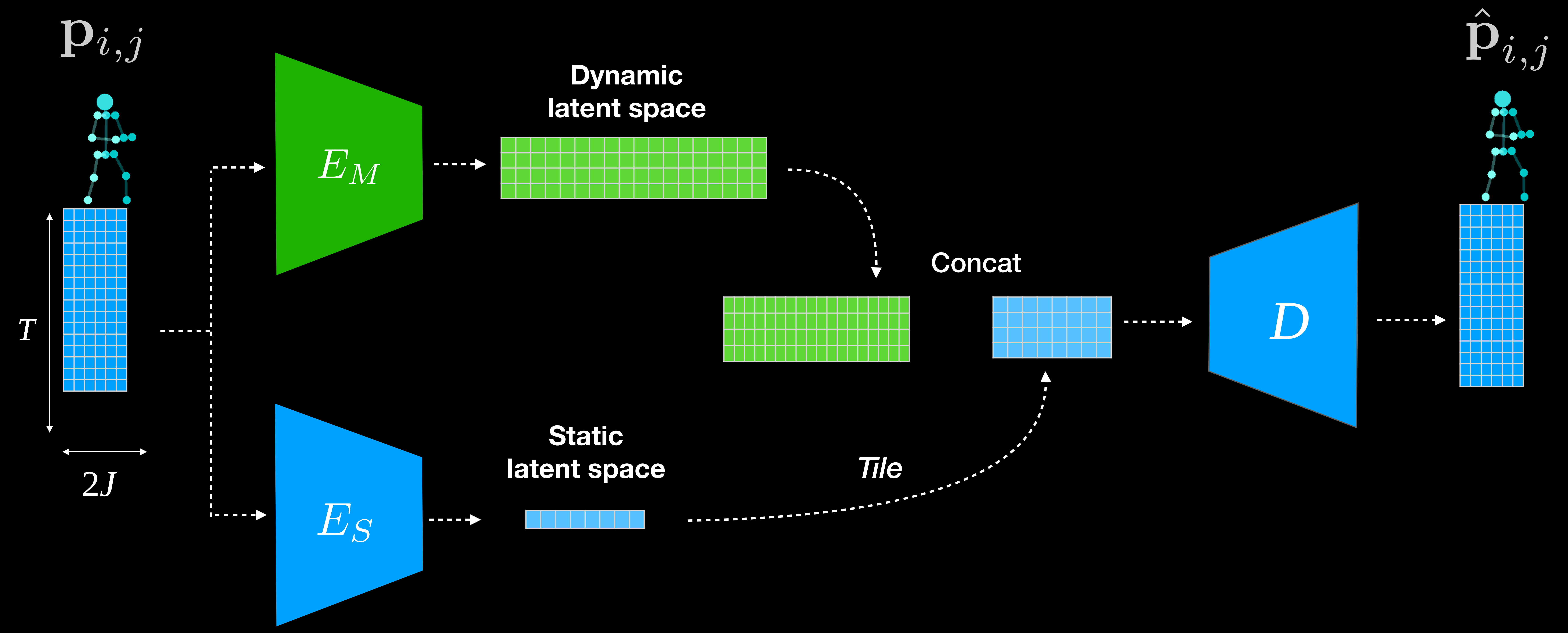


Approach



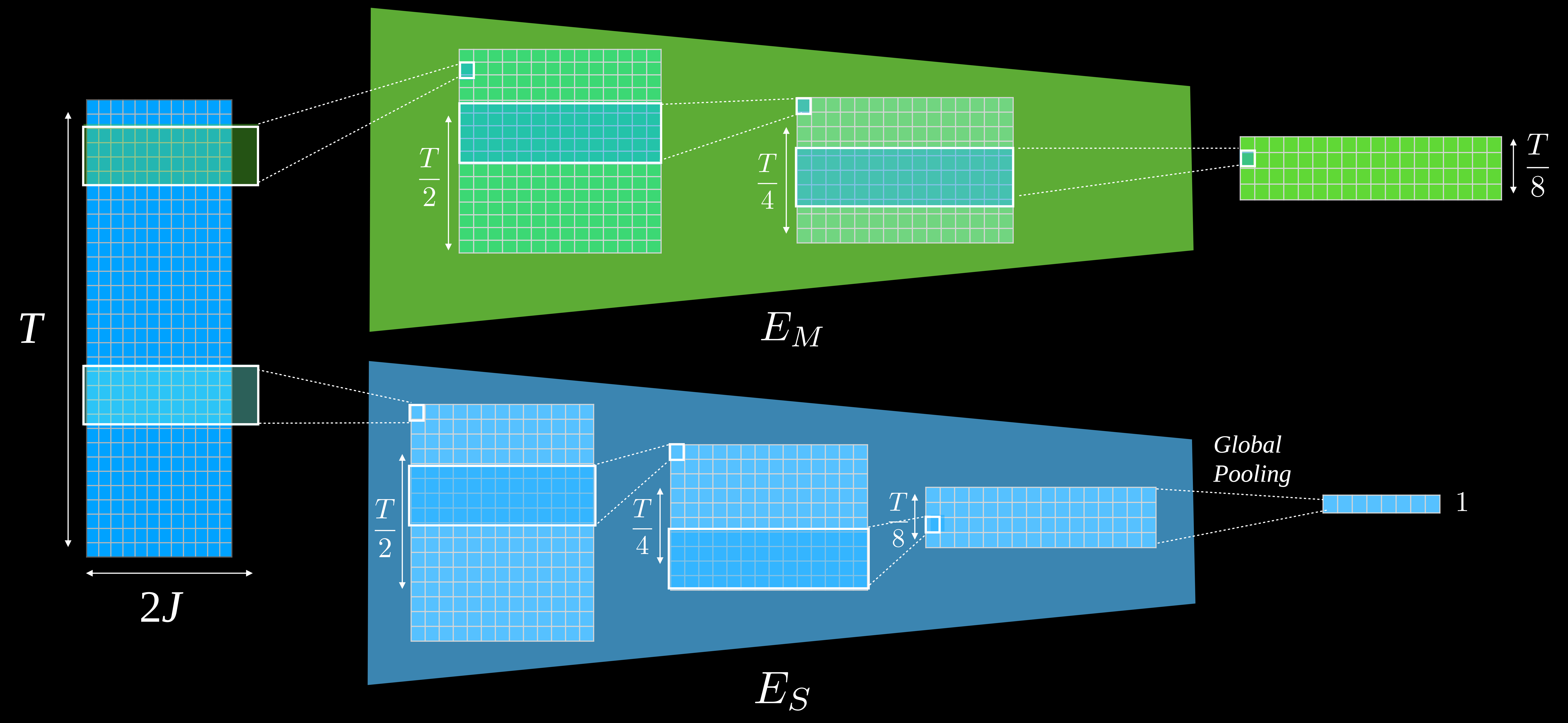
Source Video
Target Video
2D → 3D
Motion
Skeleton
3D motion Retargeting
3D → 2D
Estimated Camera Parameters
Output Video
Source Video
Target Video
Character Agnostic Motion
Static Parameters
Output Video

Architecture

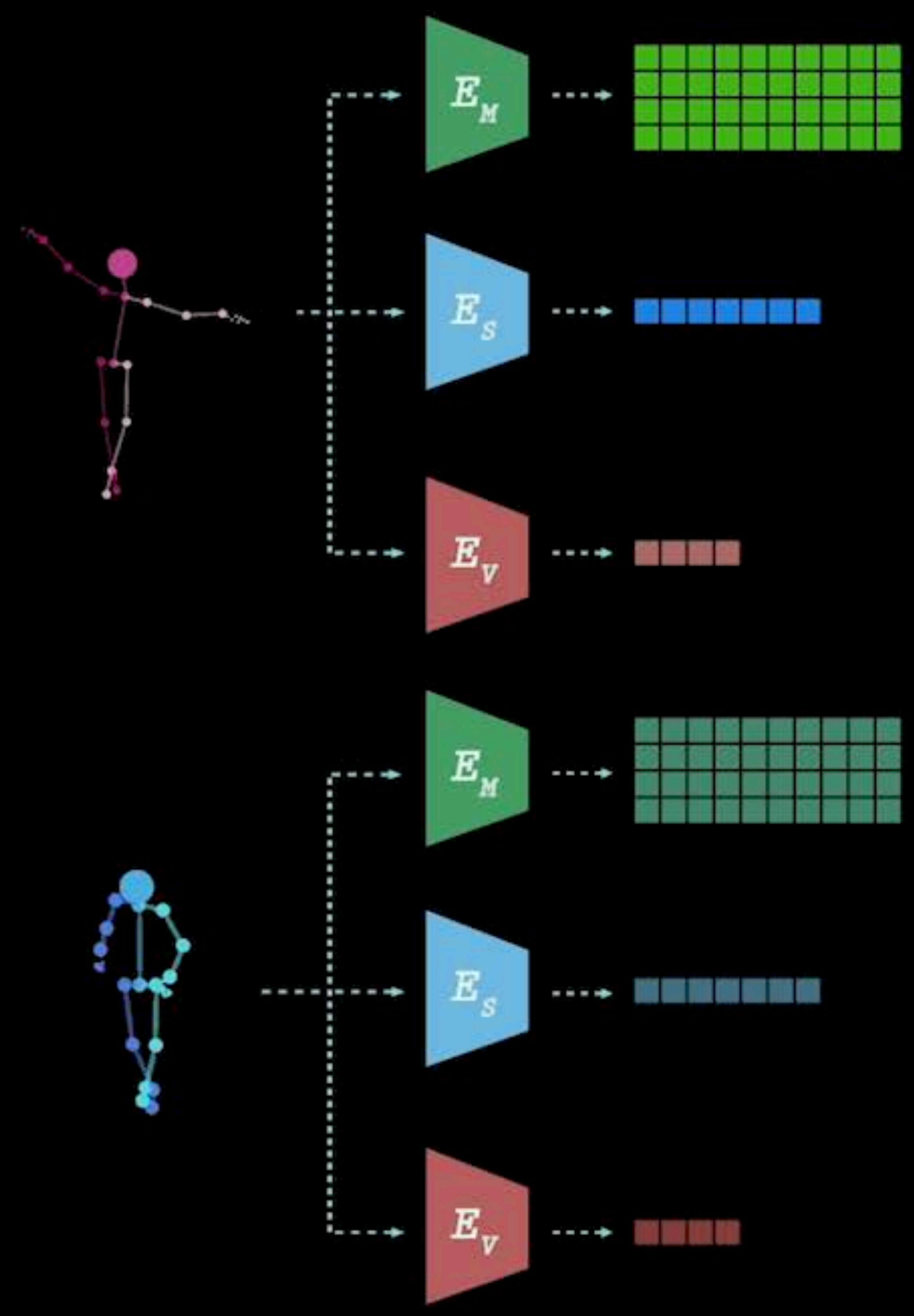


$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{\mathbf{p}_{i,j}\sim\mathcal{P}}\left[\left\| D(E_M(\mathbf{p}_{i,j}), E_S(\mathbf{p}_{i,j})) - \mathbf{p}_{i,j} \right\|^2\right].$$

Architecture

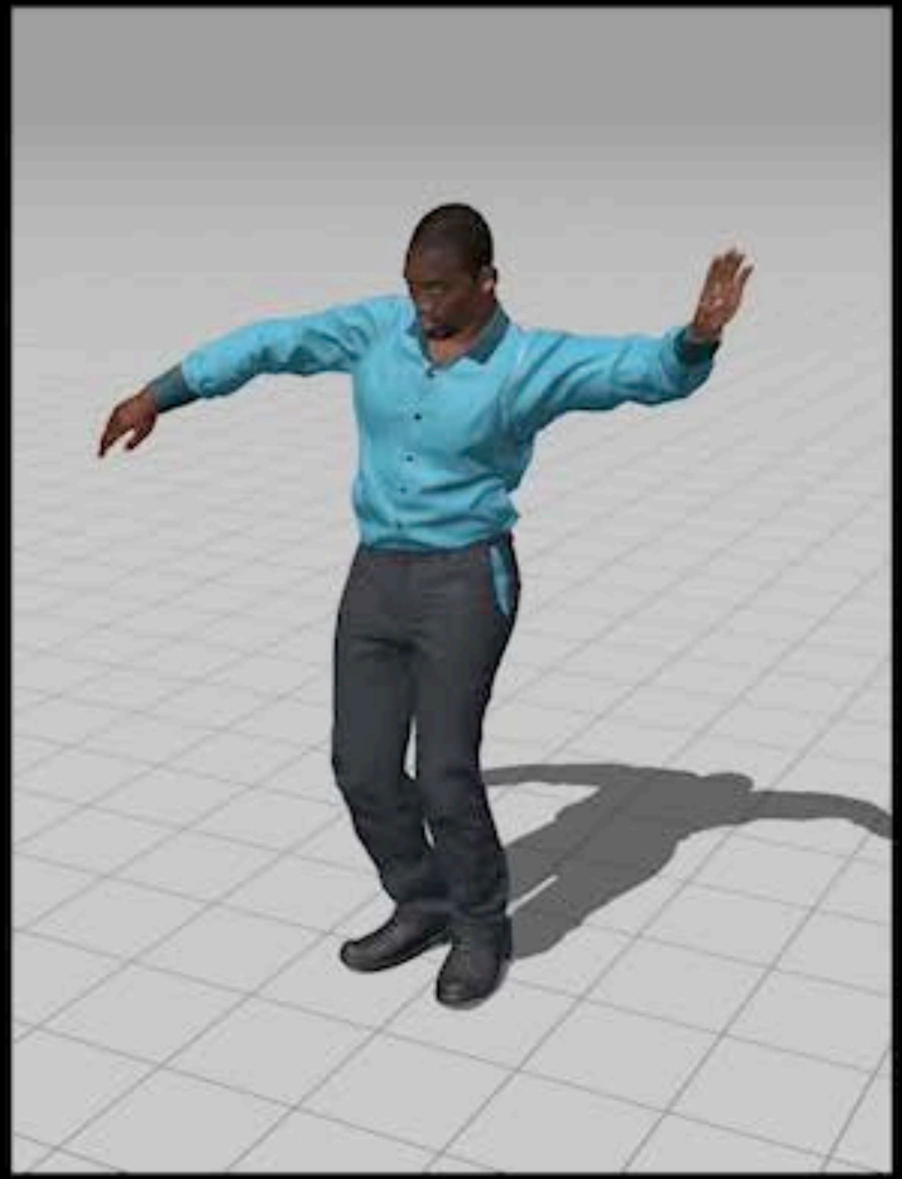


Decompose and Re-compose

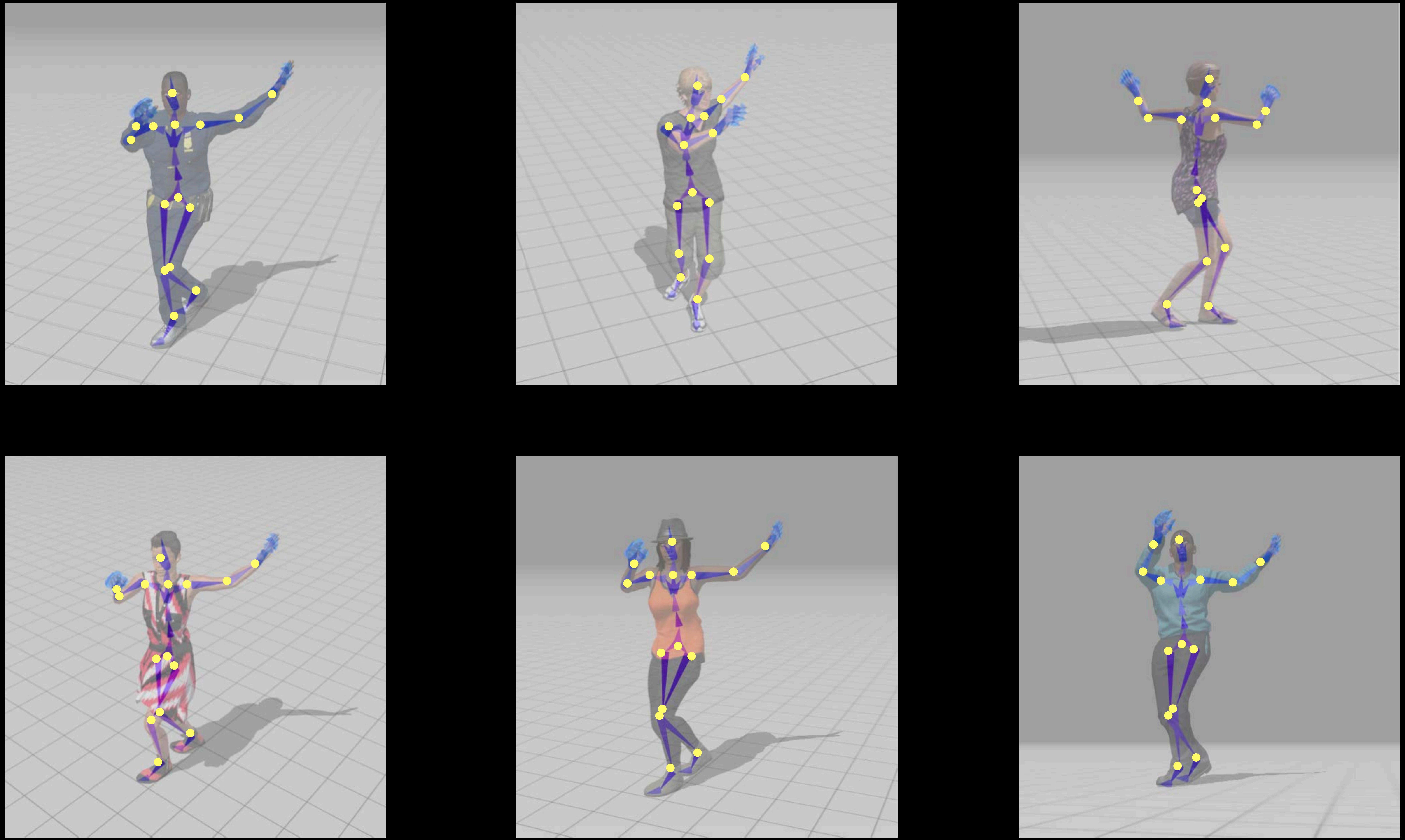


$$\begin{aligned}\mathcal{L}_{\text{cross}} = \; & \mathbb{E}_{\mathbf{p}_{i,j},\mathbf{p}_{k,l}\sim\mathcal{P}\times\mathcal{P}}\left[\|D(E_M(\mathbf{p}_{i,j}), E_S(\mathbf{p}_{k,l})) - \mathbf{p}_{i,l}\|^2\right] \\ + \; & \mathbb{E}_{\mathbf{p}_{i,j},\mathbf{p}_{k,l}\sim\mathcal{P}\times\mathcal{P}}\left[\|D(E_M(\mathbf{p}_{k,l}), E_S(\mathbf{p}_{i,j})) - \mathbf{p}_{k,j}\|^2\right]\end{aligned}$$

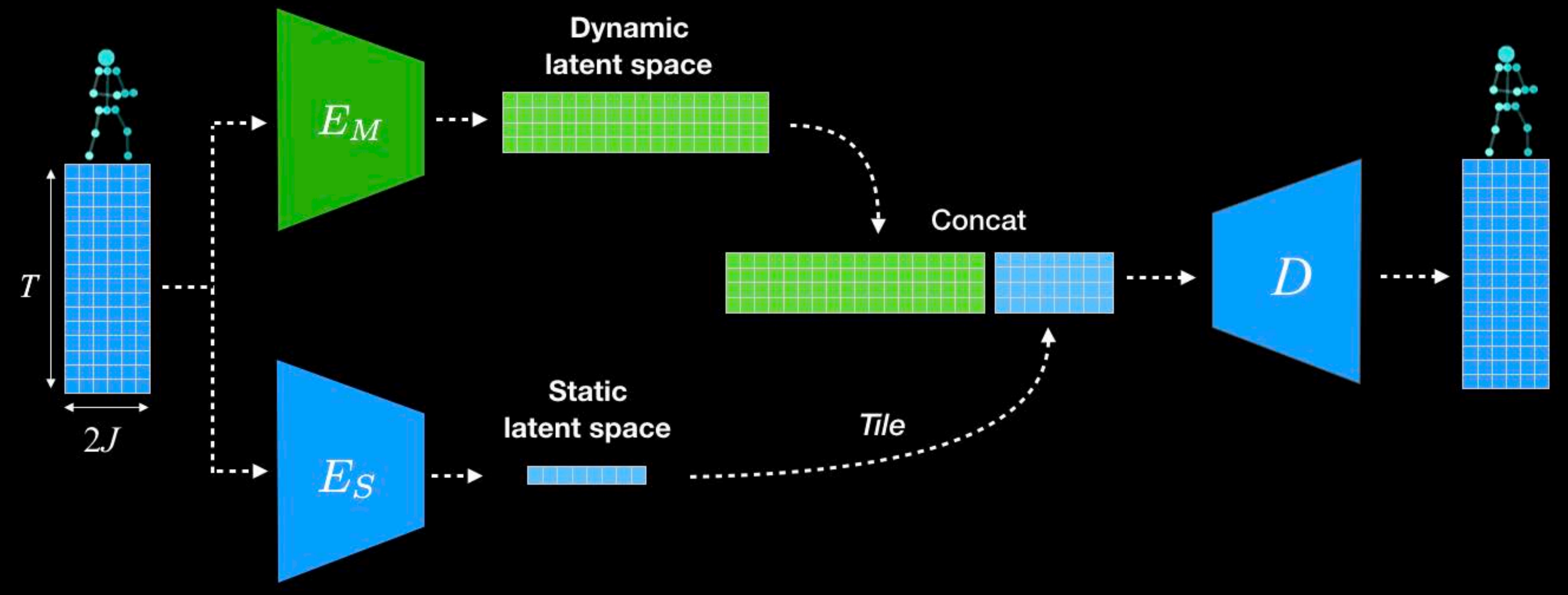
Synthetic Data



Synthetic Data

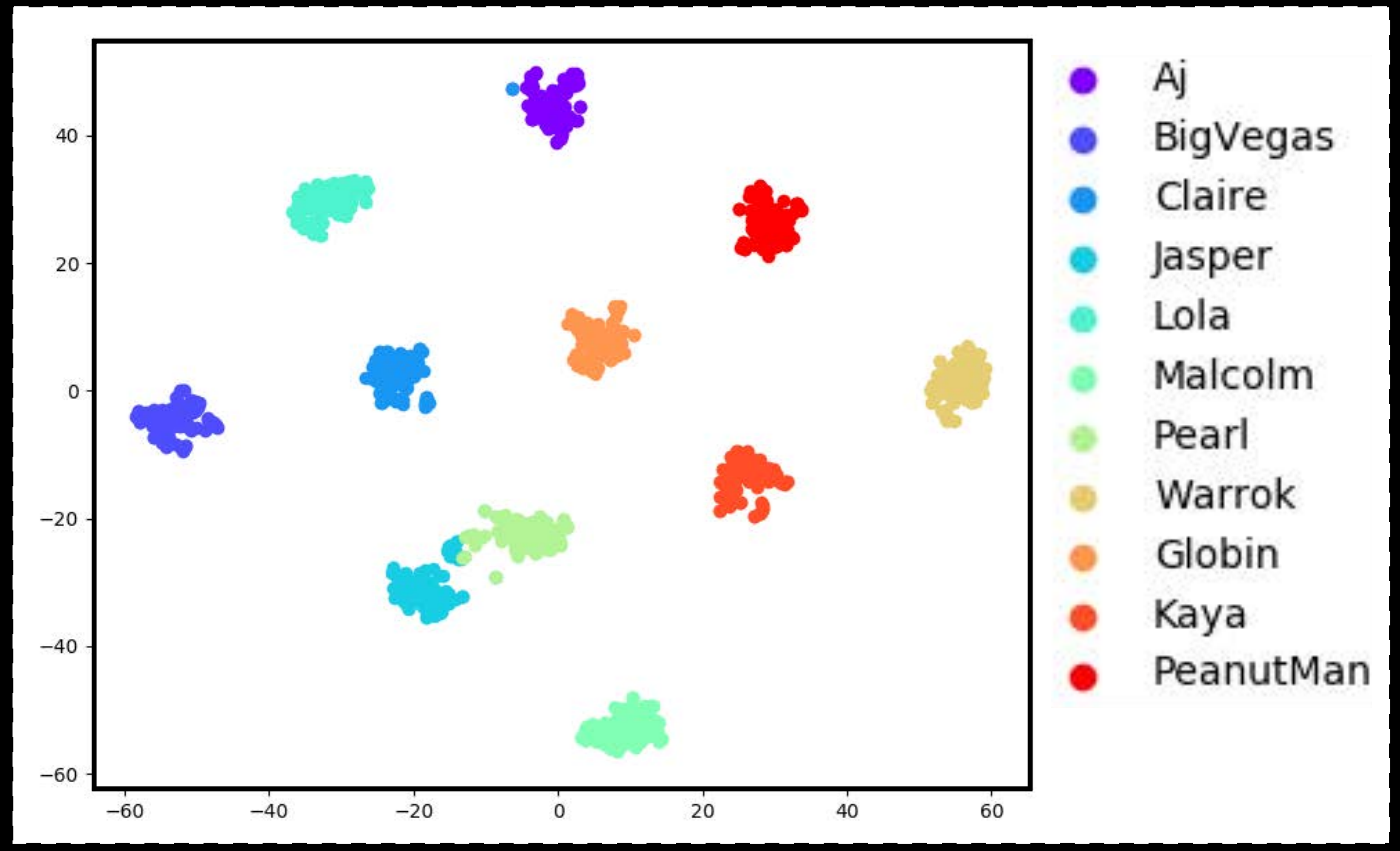


Learning Clusters Implicitly

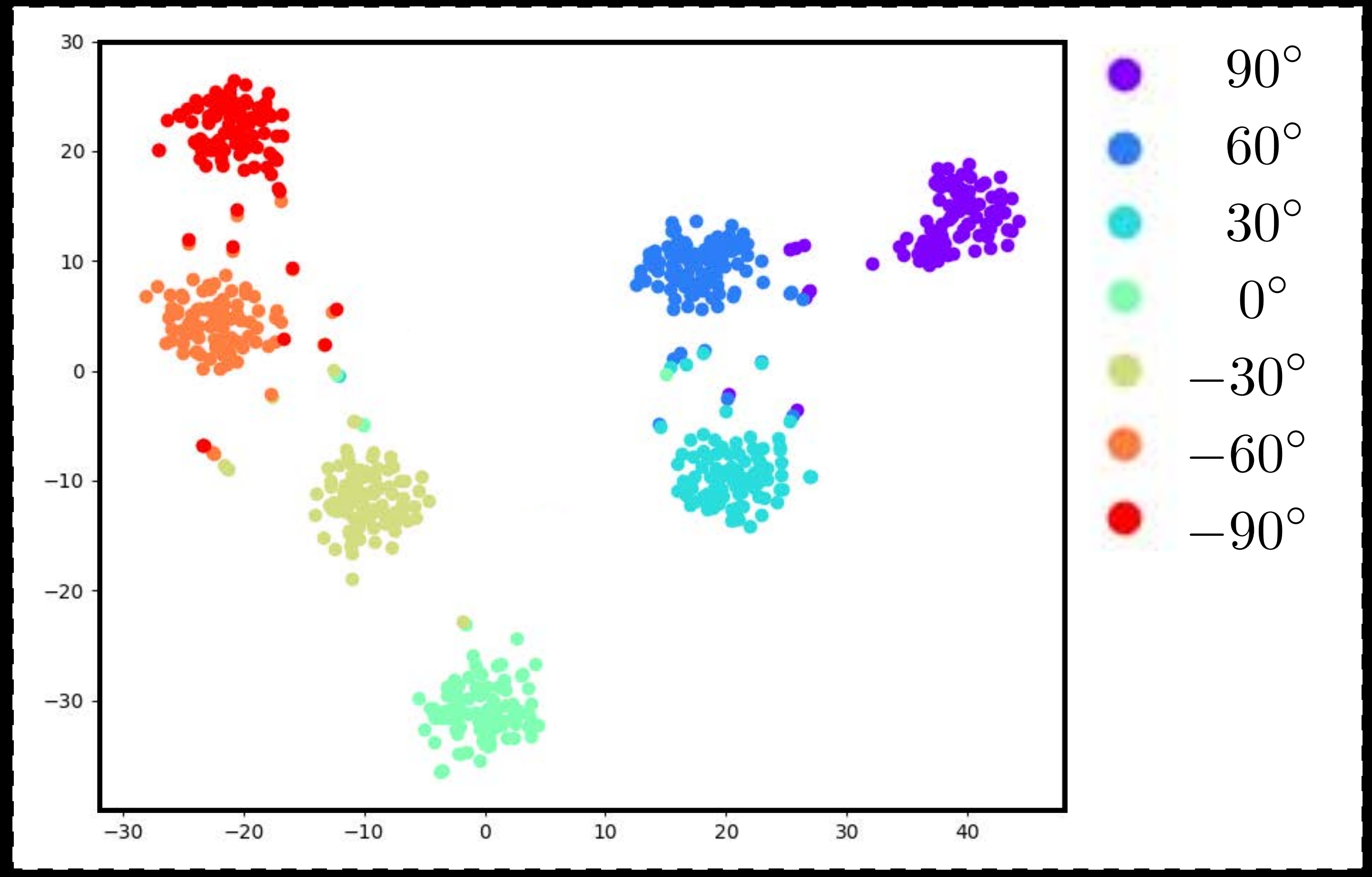


$$\mathcal{L}_{rec} + \lambda \mathcal{L}_{cross}$$

Skeleton Latent Space

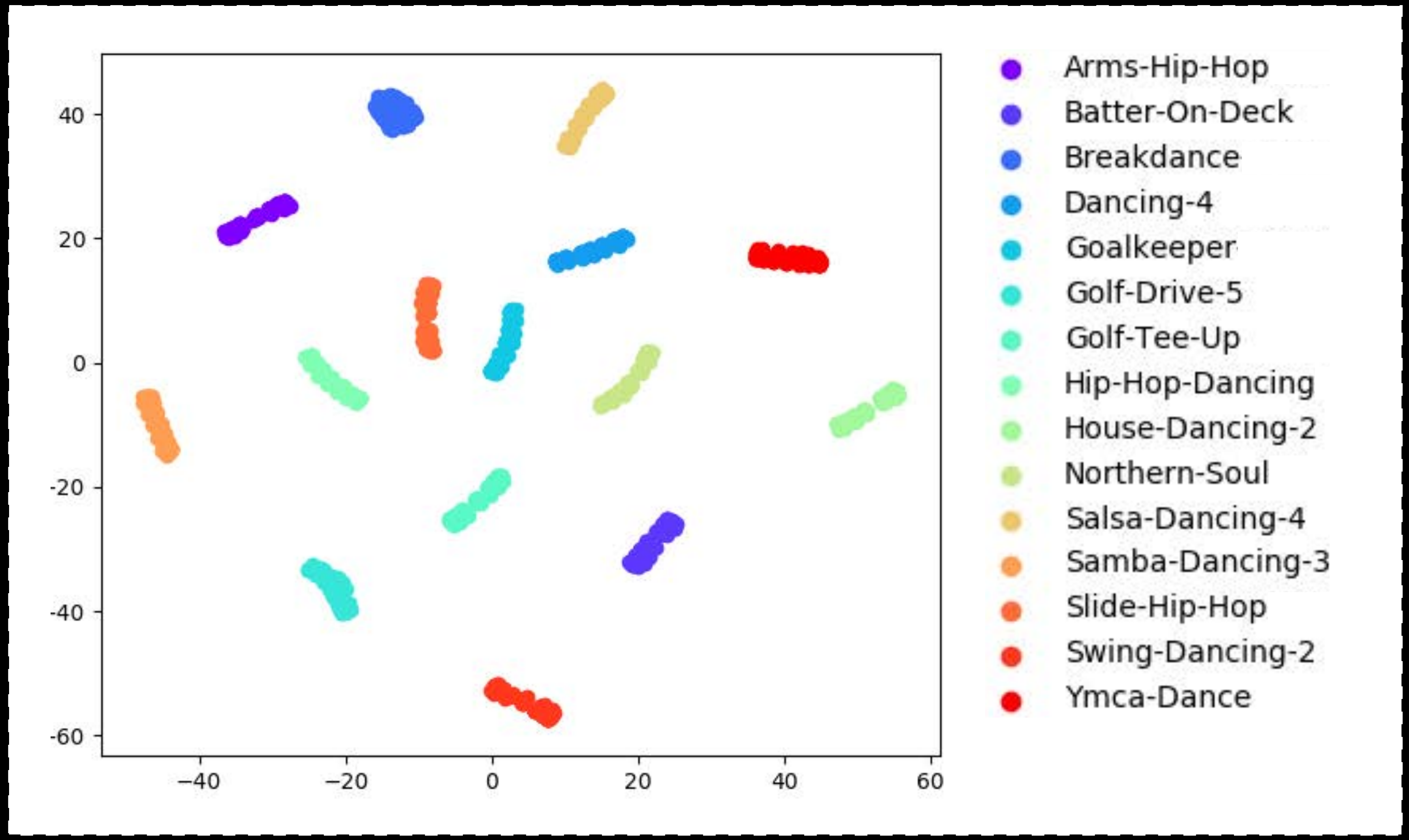


View-Angle Latent Space

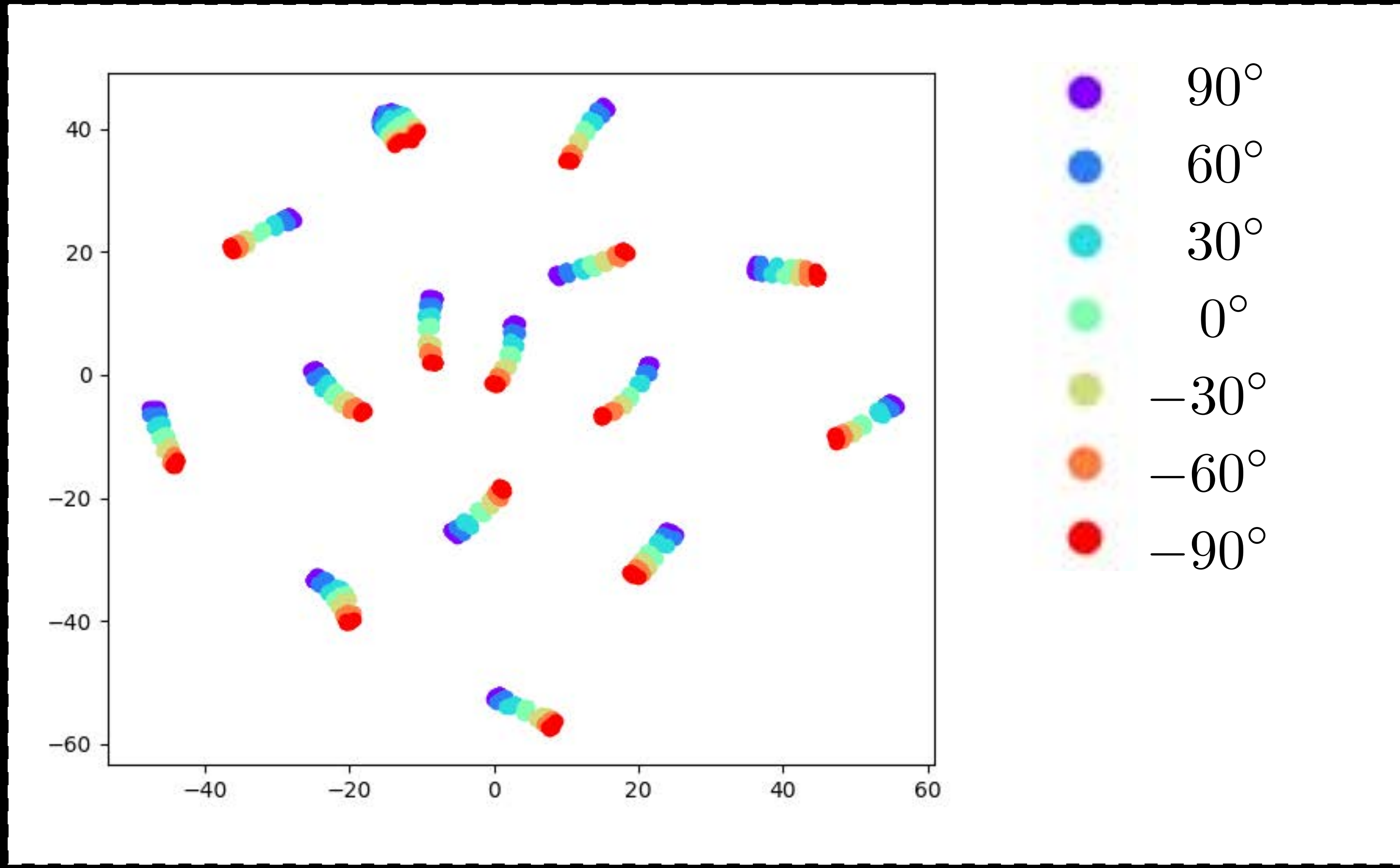


Implicitl Clusters Learning

Motion Latent Space



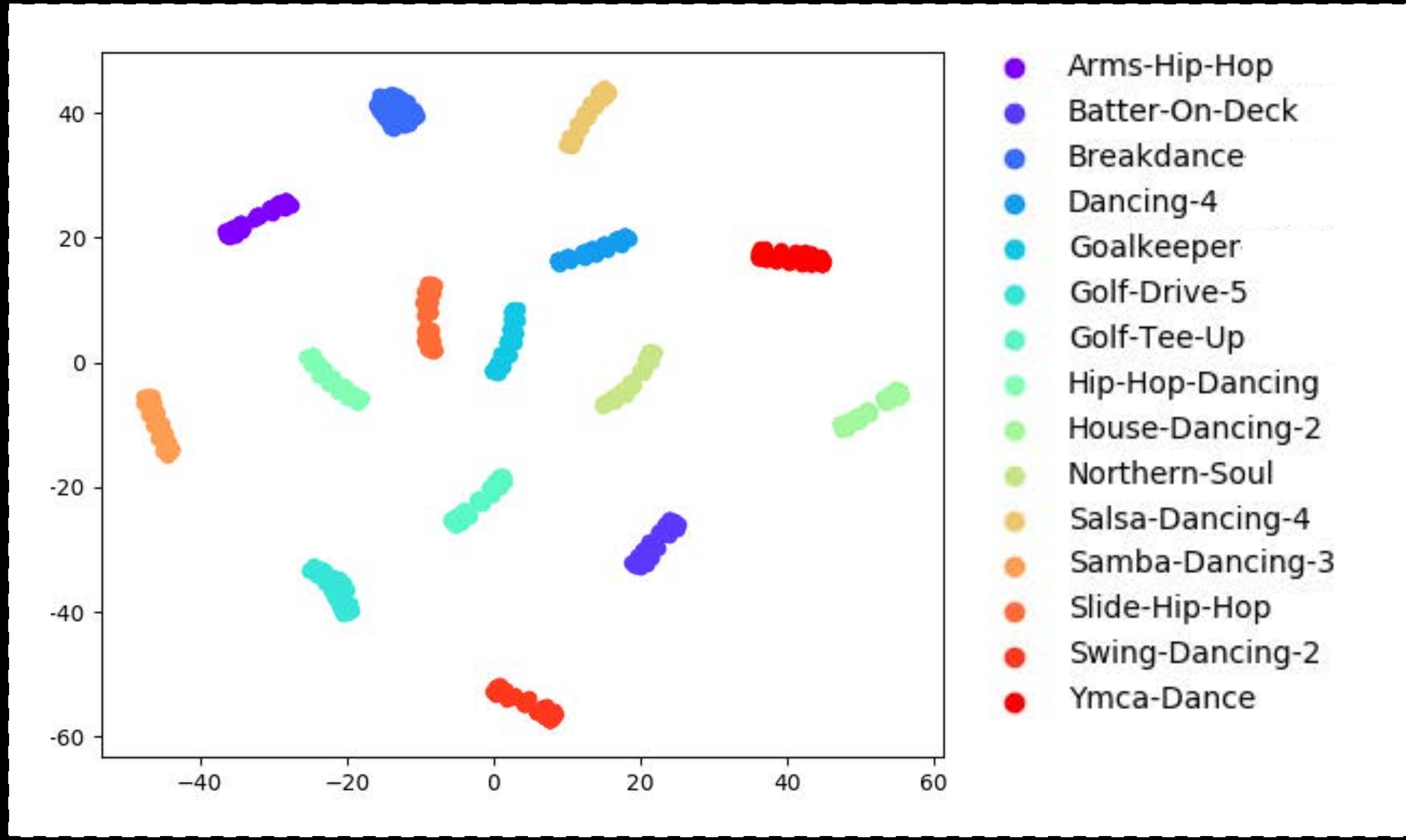
Motion Latent Space - View Angle labels



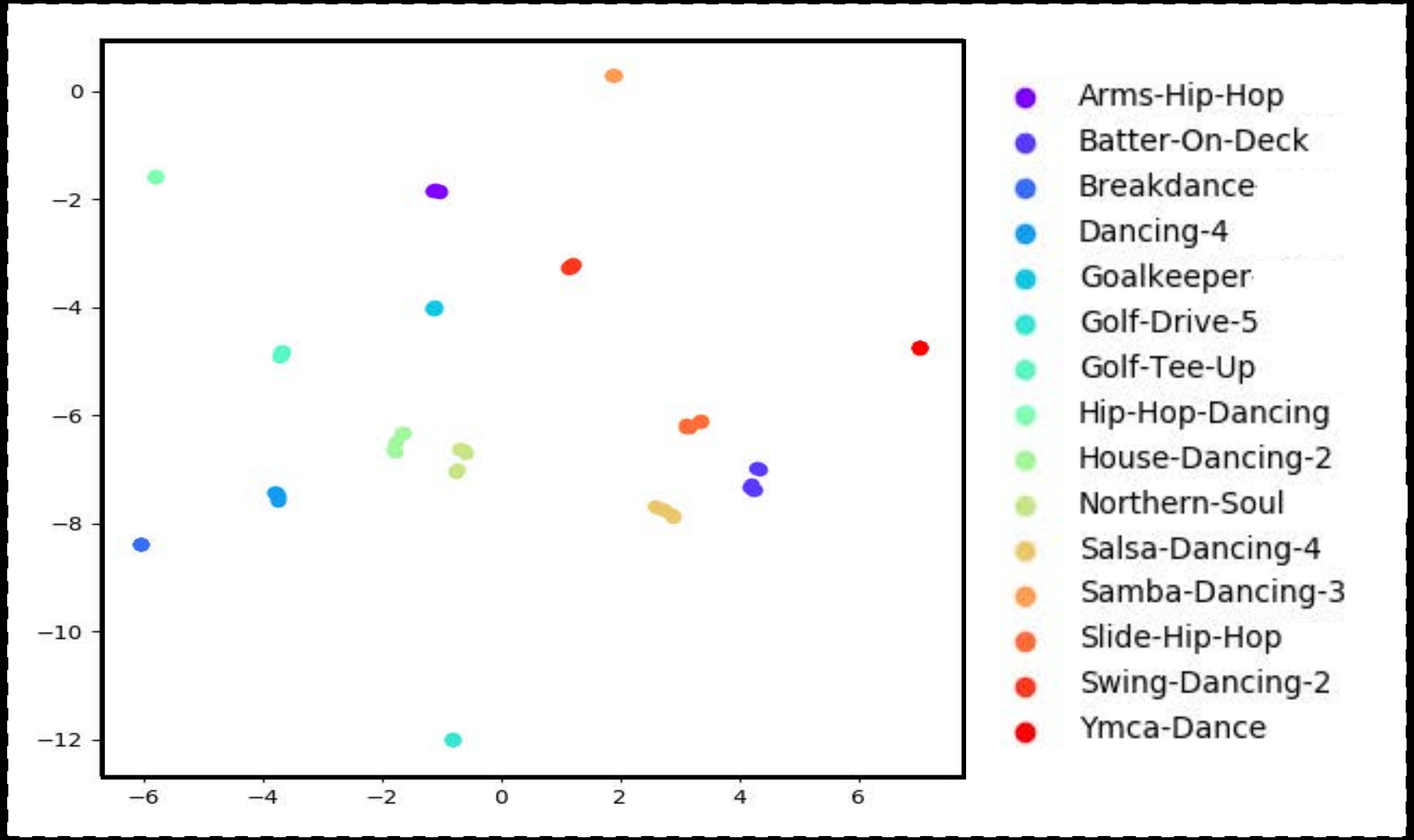
$$\mathcal{L}_{trip\_M} = \mathbb{E}_{p_{i,j}, p_{i,l}, p_{k,l} \sim \mathcal{P}} [\|E_M(p_{i,l}) - E_M(p_{i,j})\| - \|E_M(p_{i,l}) - E_M(p_{k,l})\| + \alpha]_+,$$

Triplet Loss

Motion Latent Space Without Triplet loss

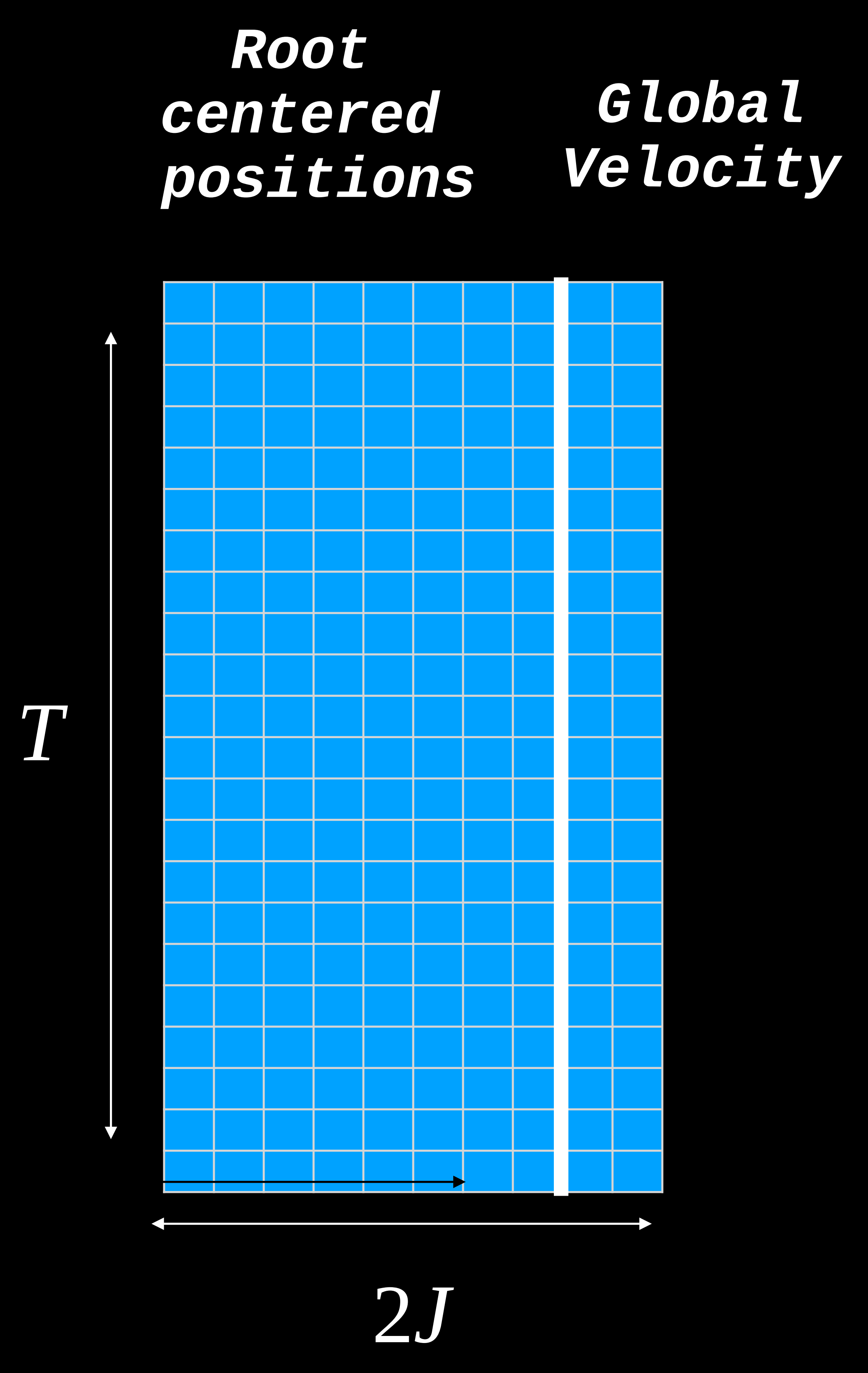


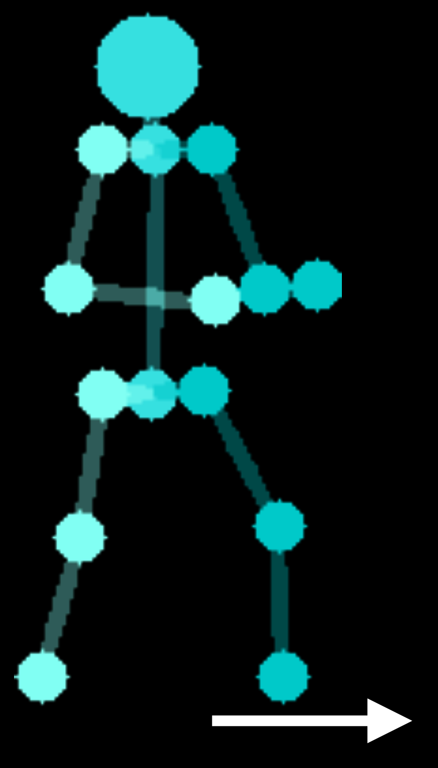
Motion Latent Space With Triplet loss



$$\mathcal{L}_{\text{trip\_M}} = \mathbb{E}_{\mathbf{p}_{i,j},\mathbf{p}_{i,l},\mathbf{p}_{k,l}\sim\mathcal{P}}[\|E_M(\mathbf{p}_{i,l}) - E_M(\mathbf{p}_{i,j})\| - \|E_M(\mathbf{p}_{i,l}) - E_M(\mathbf{p}_{k,l})\| + \alpha]_+,$$

Foot Velocity Loss

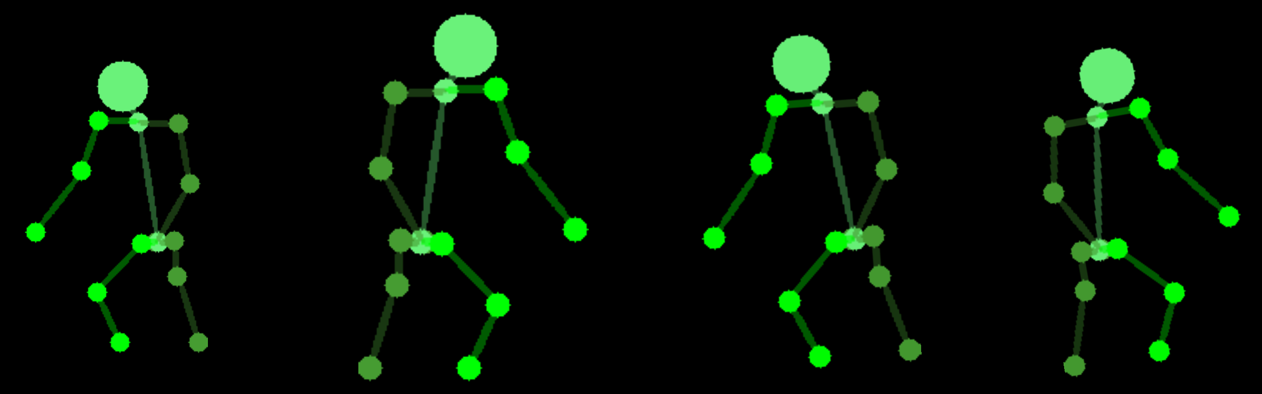




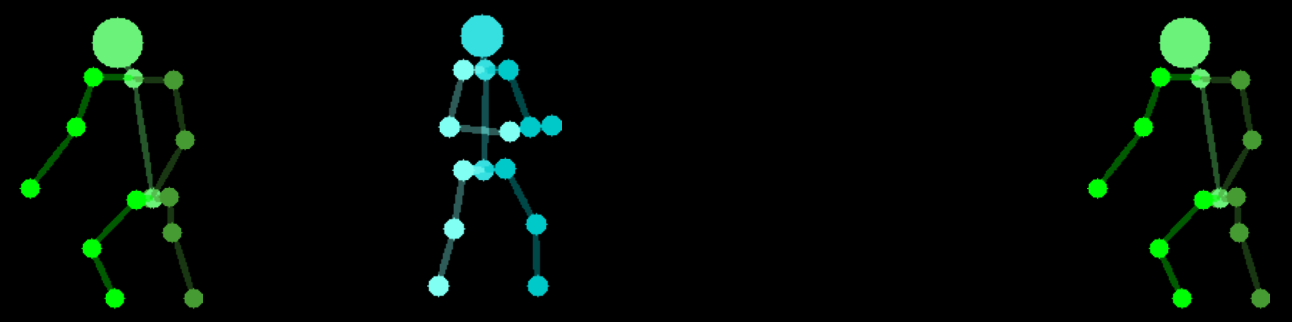
$$\mathcal{L}_{\text{foot}} = \mathbb{E}_{\mathbf{p}_{i,j}\sim\mathcal{P}} \sum_{n\in\mathcal{J}_{\text{end}}} \|V_{\text{global}}(\hat{\mathbf{p}}_{ij}) + V_{\text{joint}_n}(\hat{\mathbf{p}}_{ij}) - V_{\text{orig}_n}(\mathbf{p}_{ij})\|^2,$$

Supporting Videos in the wild

- *Augmentation (Temporal trimming, flips, rotation, scale)*



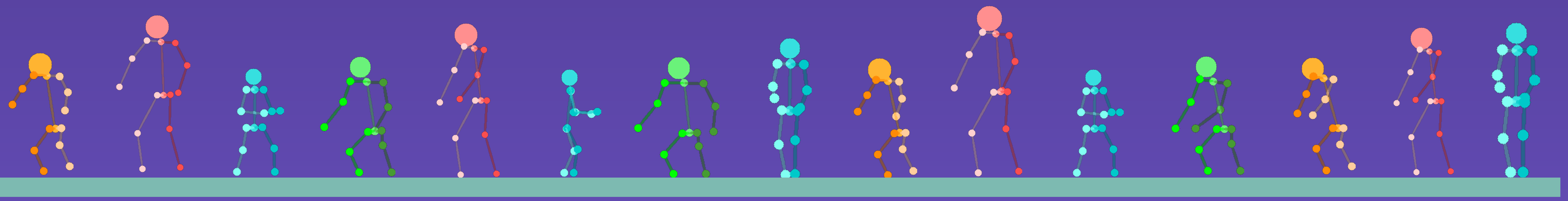
- *Adding noise to the training data*



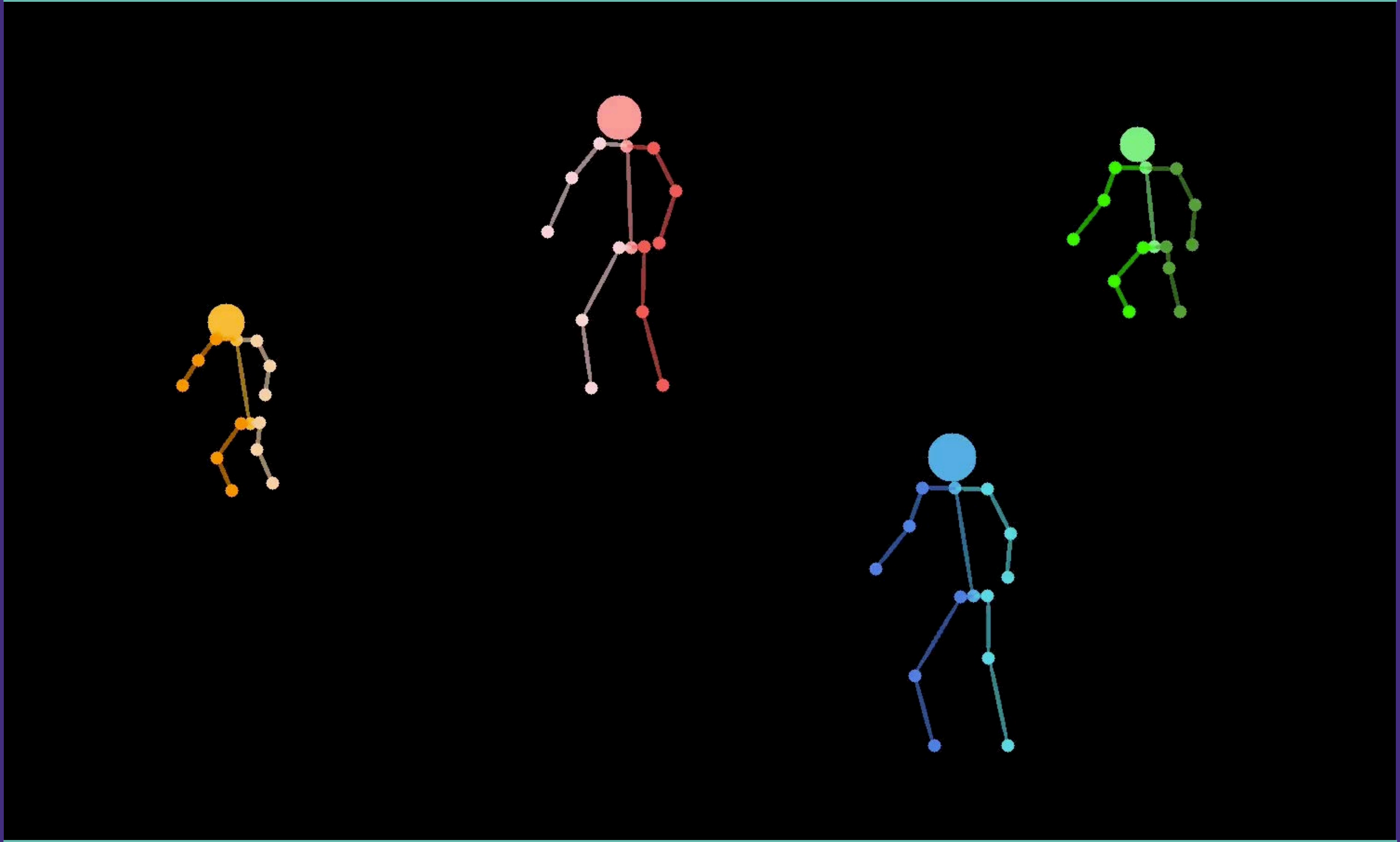
- *Reconstruct real videos using (only) the reconstruction loss.*

Outline

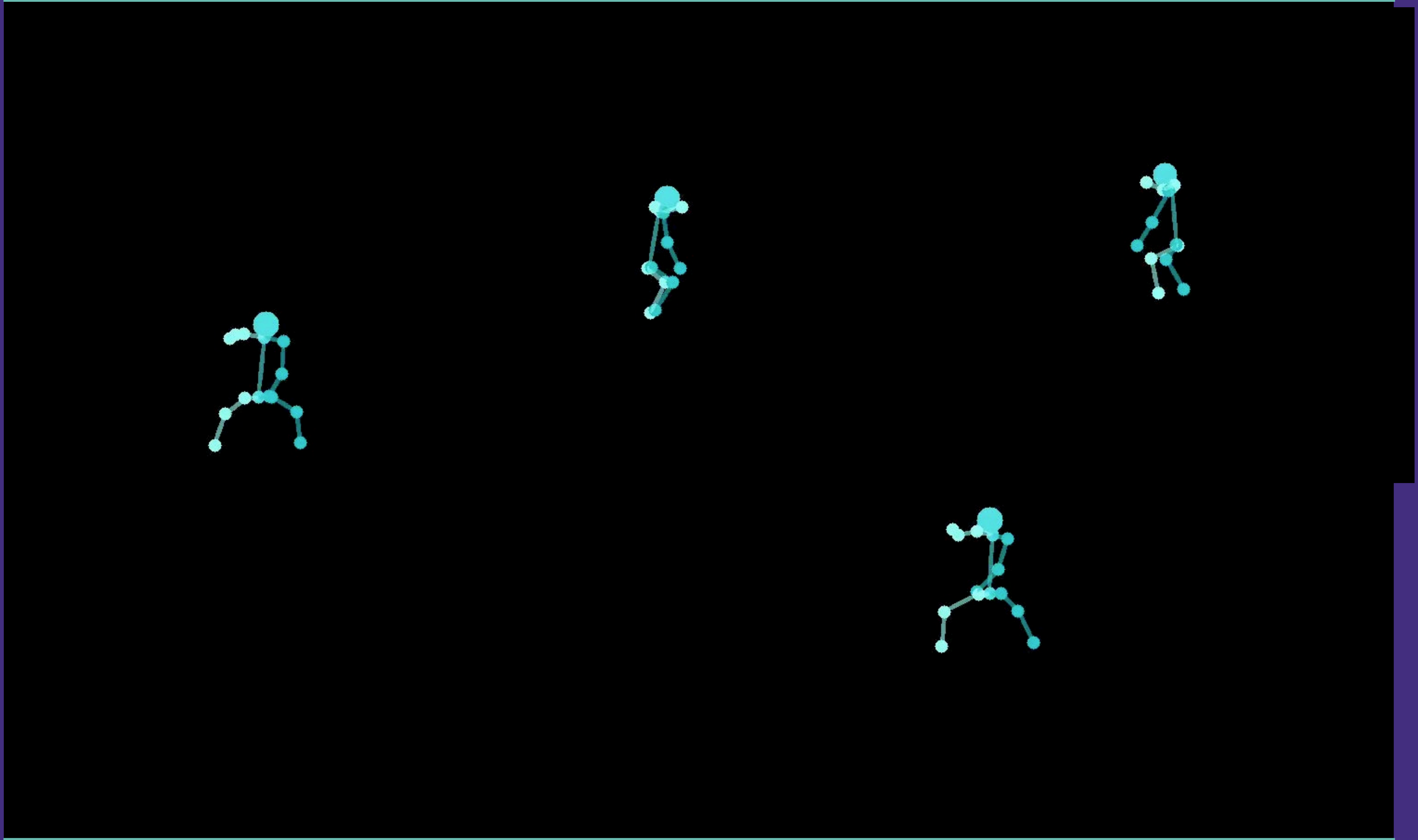
- Motivation
- Approach
- **Results**
- Application



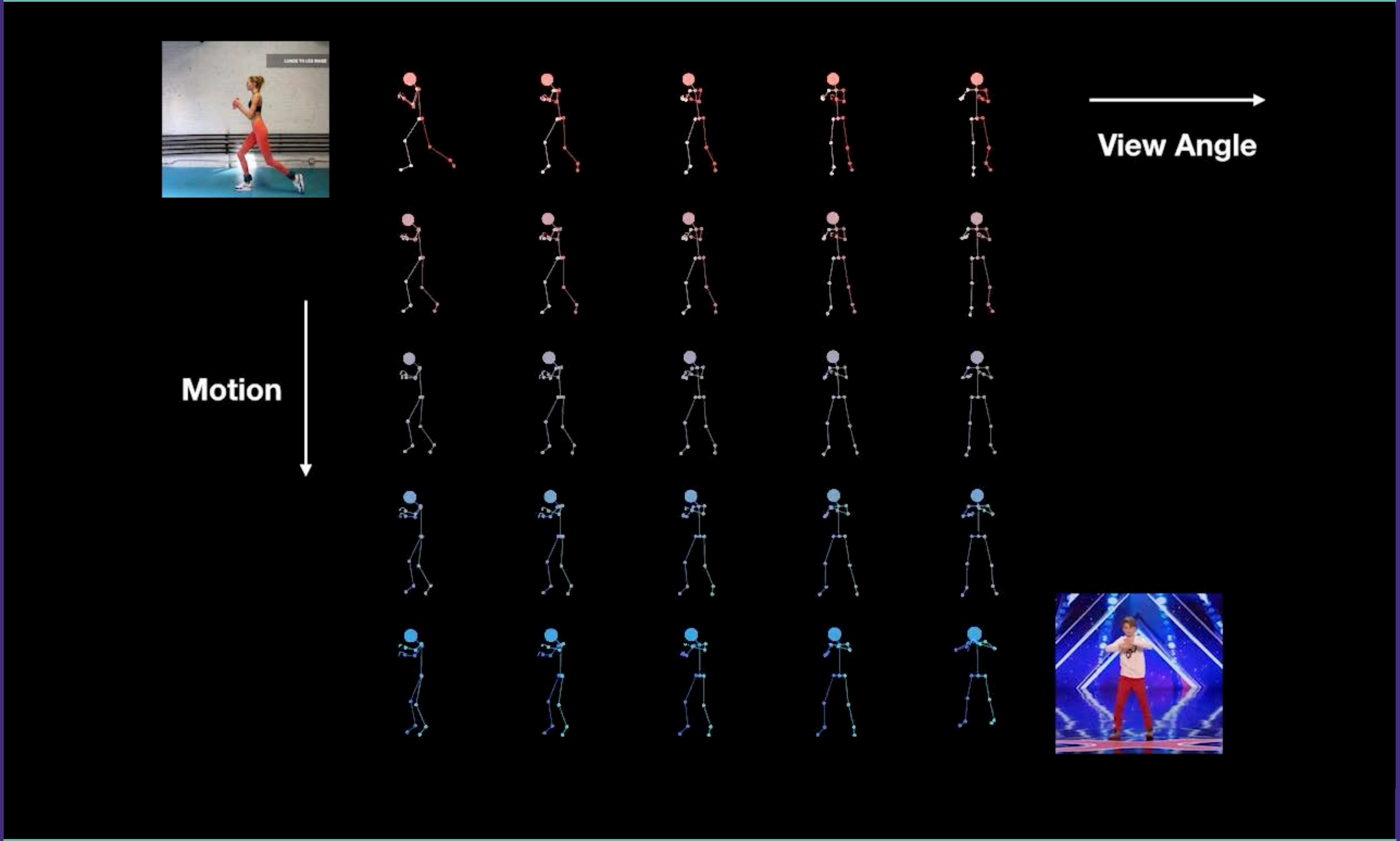
Results-skeleton



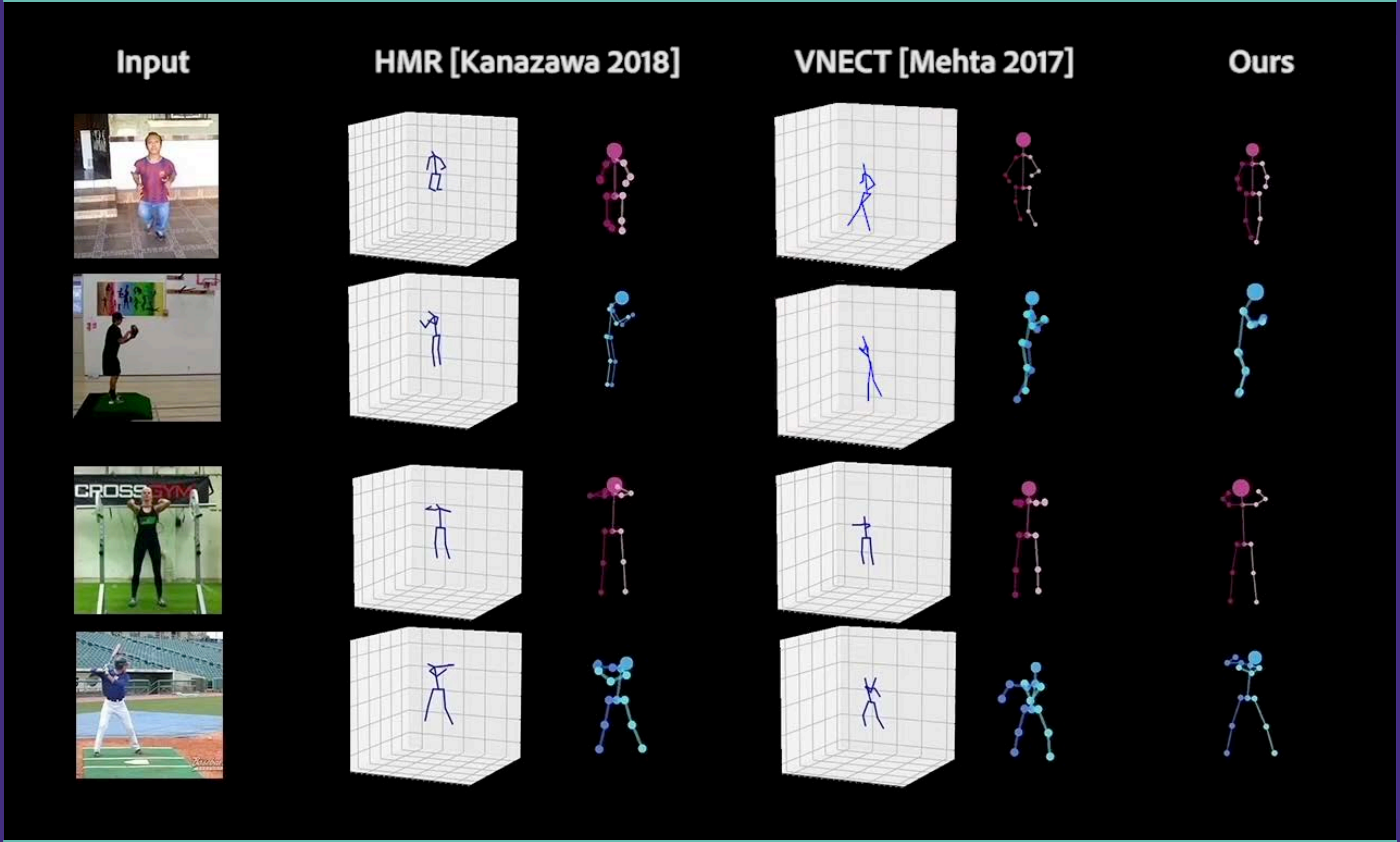
Results – view



Interpolation

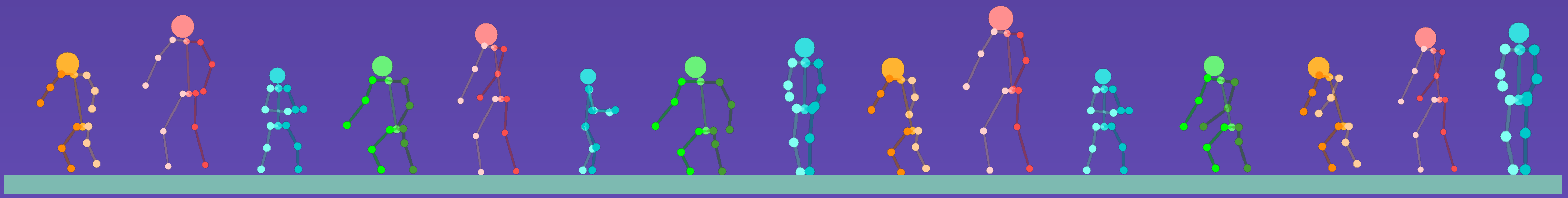


Comparison

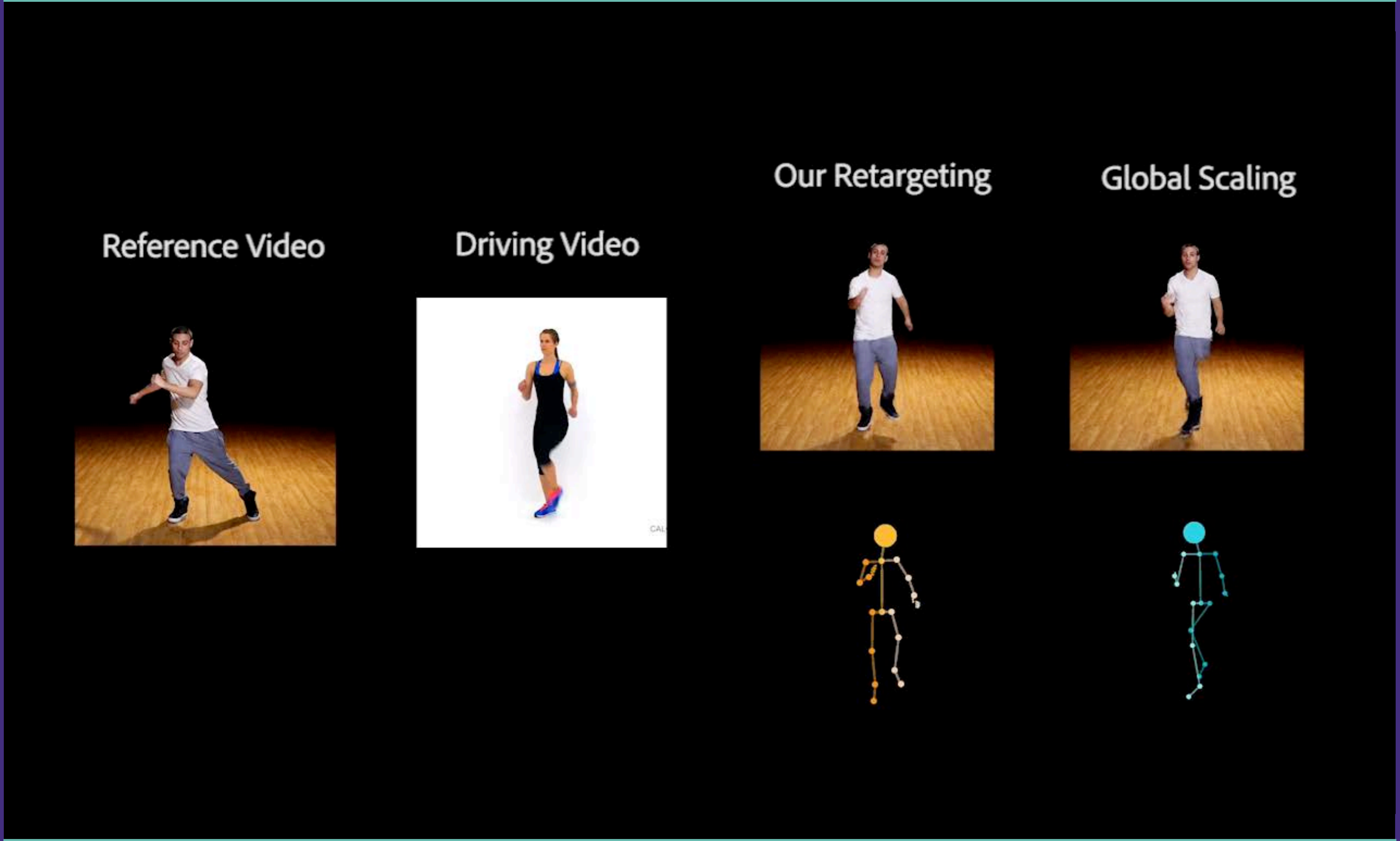


Outline

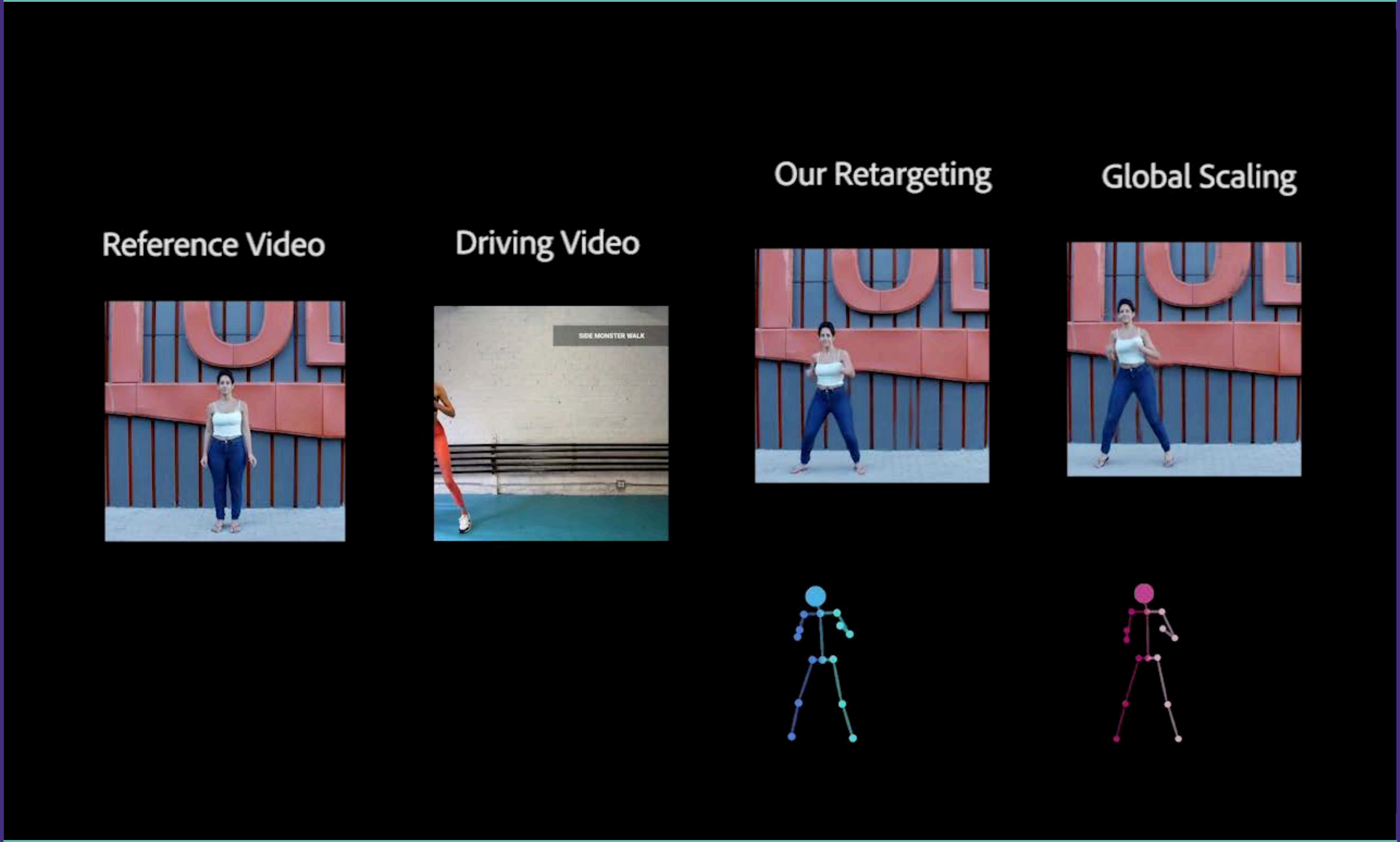
- Motivation
- Approach
- Results
- Application



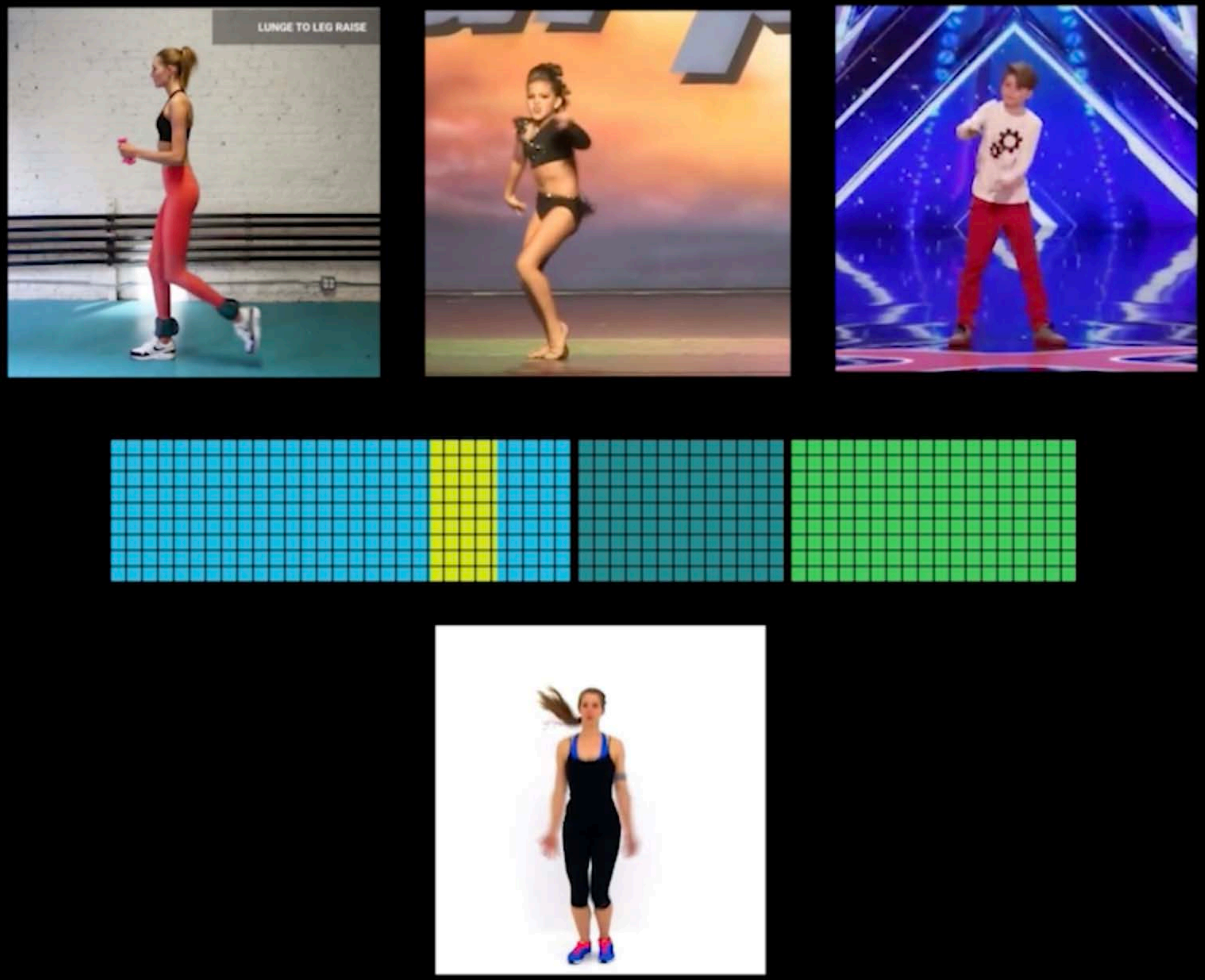
Applications-performance cloning



Applications-performance cloning



Applications – Motion Retrieval



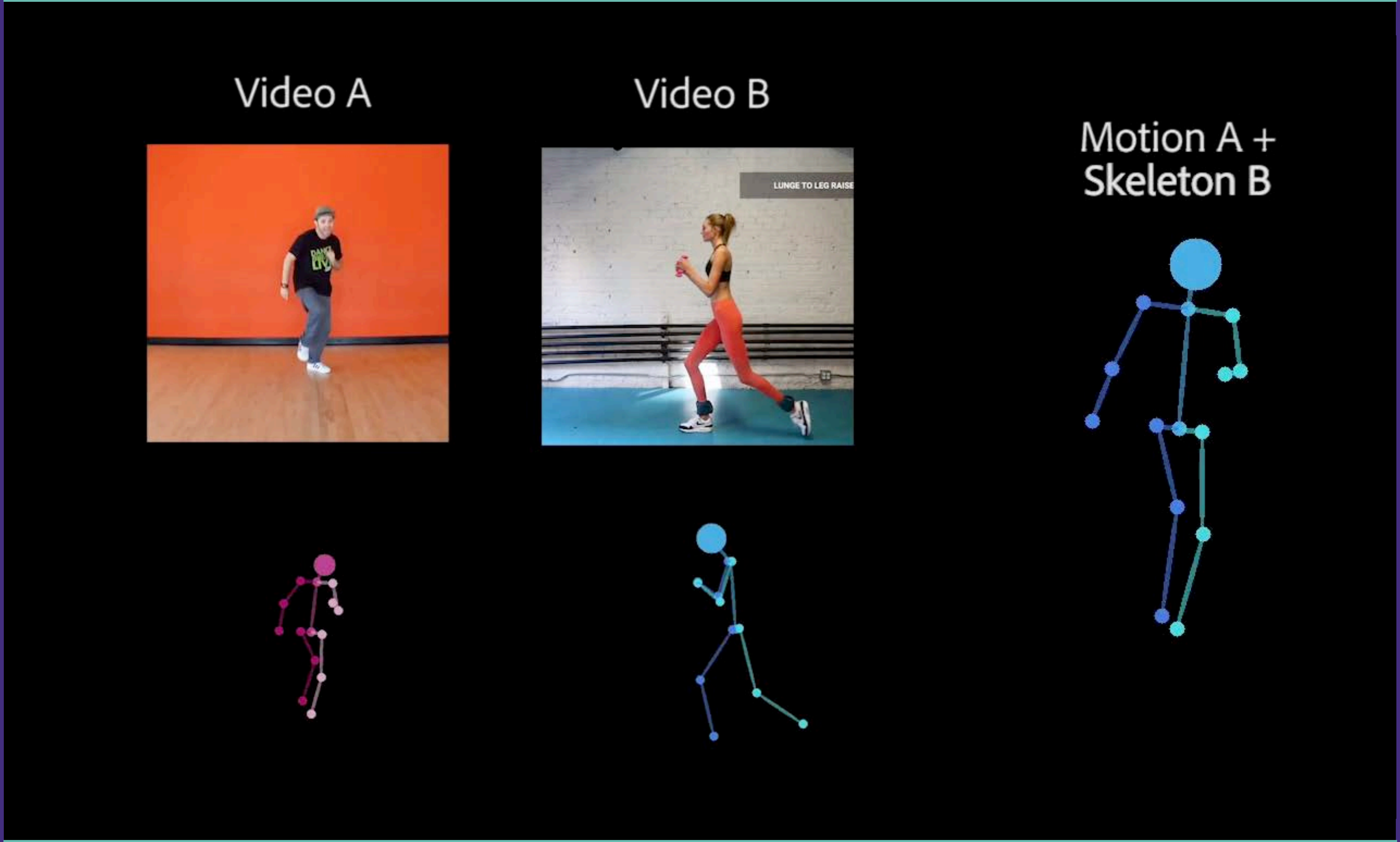
Applications – Motion Retrieval

thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

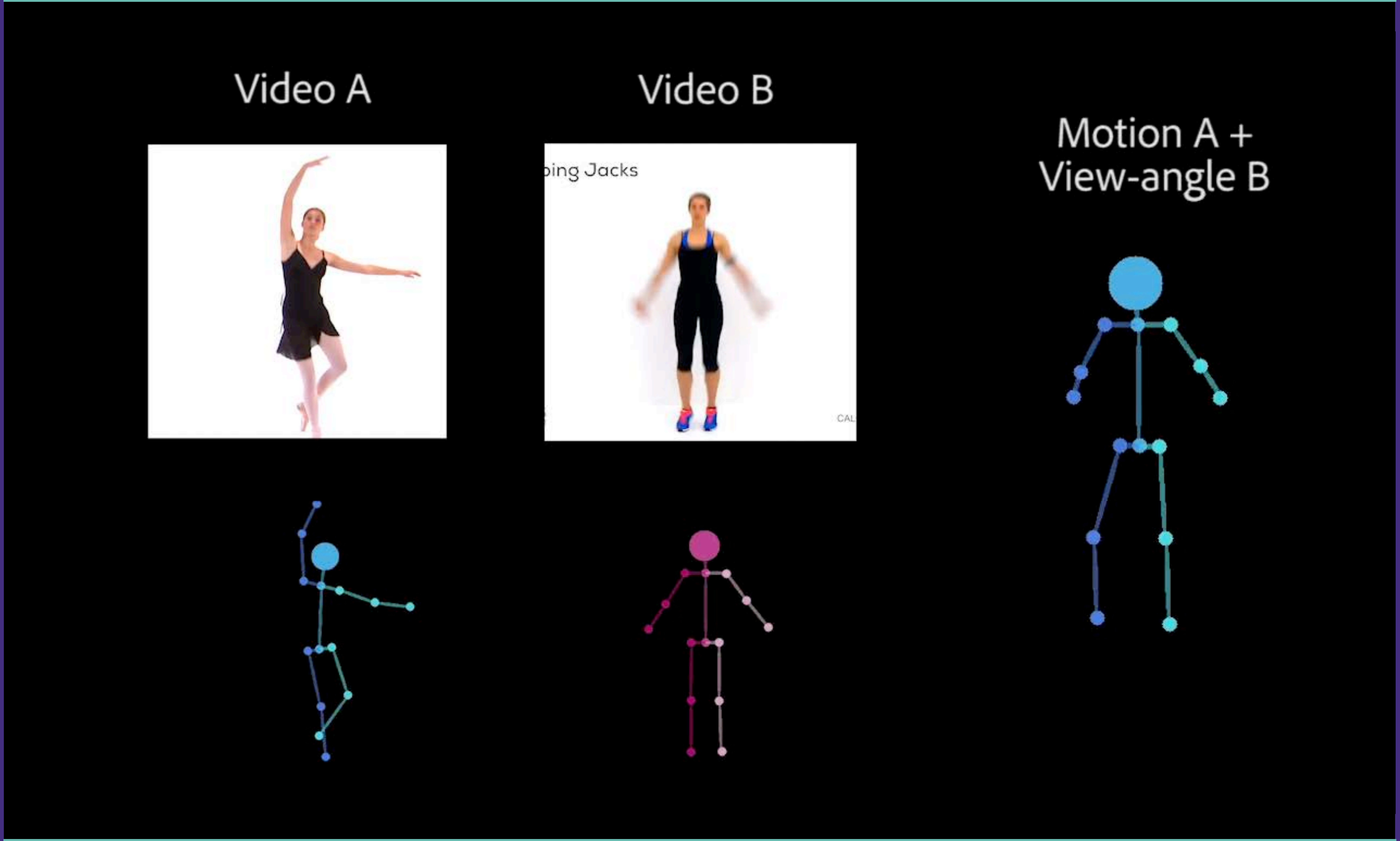
Query Video

Top 4 Results

Failure cases



Failure cases

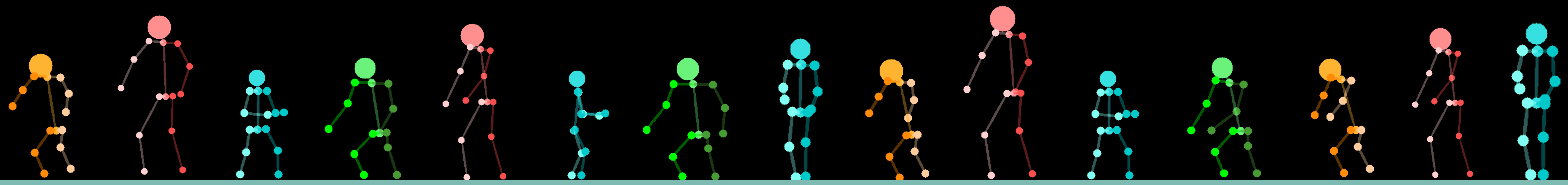


Conclusions

Take home message:

Deep networks can constitute a better solution for specific sub-tasks, which do not strictly require a full 3D reconstruction.

Synthetic data can really help with deep neural network training.



Questions?