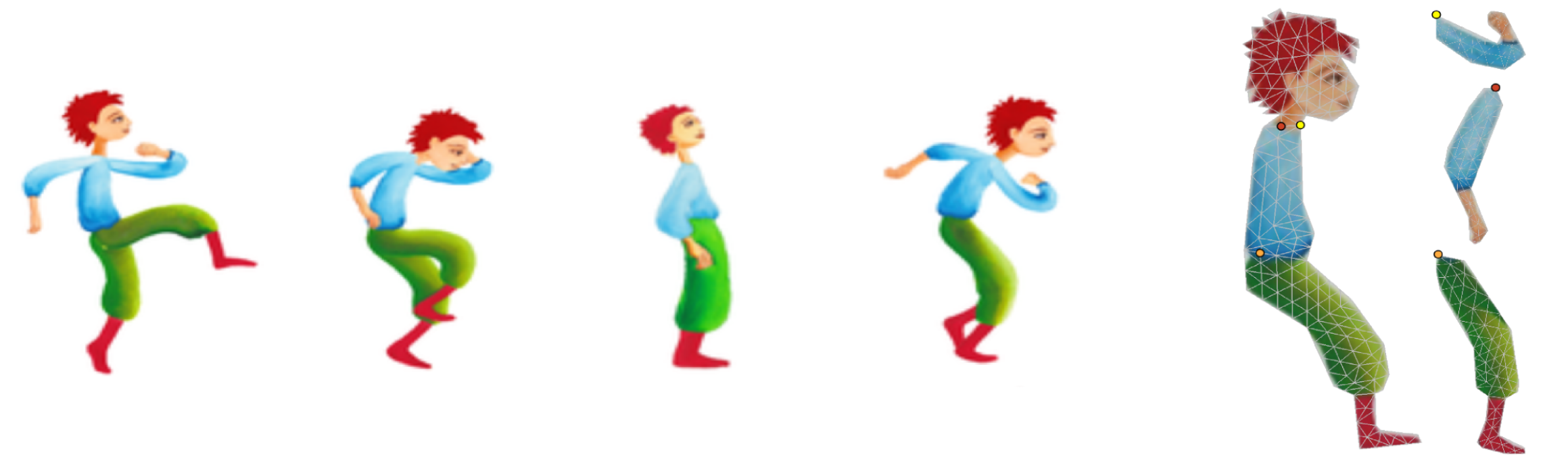
Neural Puppet: Generative Layered Cartoon Characters

Omid Poursaeed, Vladimir G. Kim, Eli Shechtman, Jun Saito, Serge Belongie

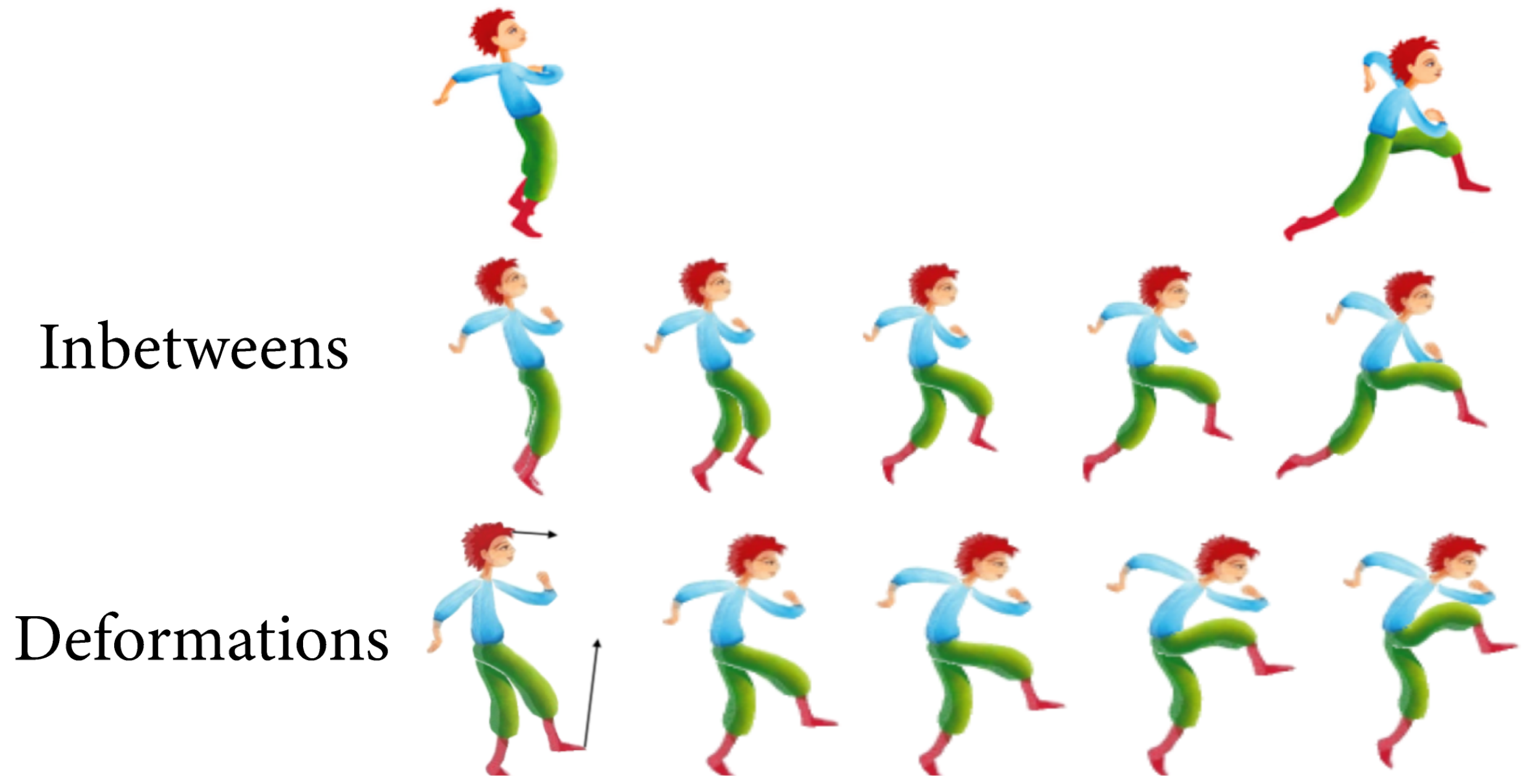
Motivation: Animating 2D characters requires extensive manual effort

Goal: Learning to animate characters with minimum expert supervision

- The user provides a set of unlabeled images and a layered template mesh



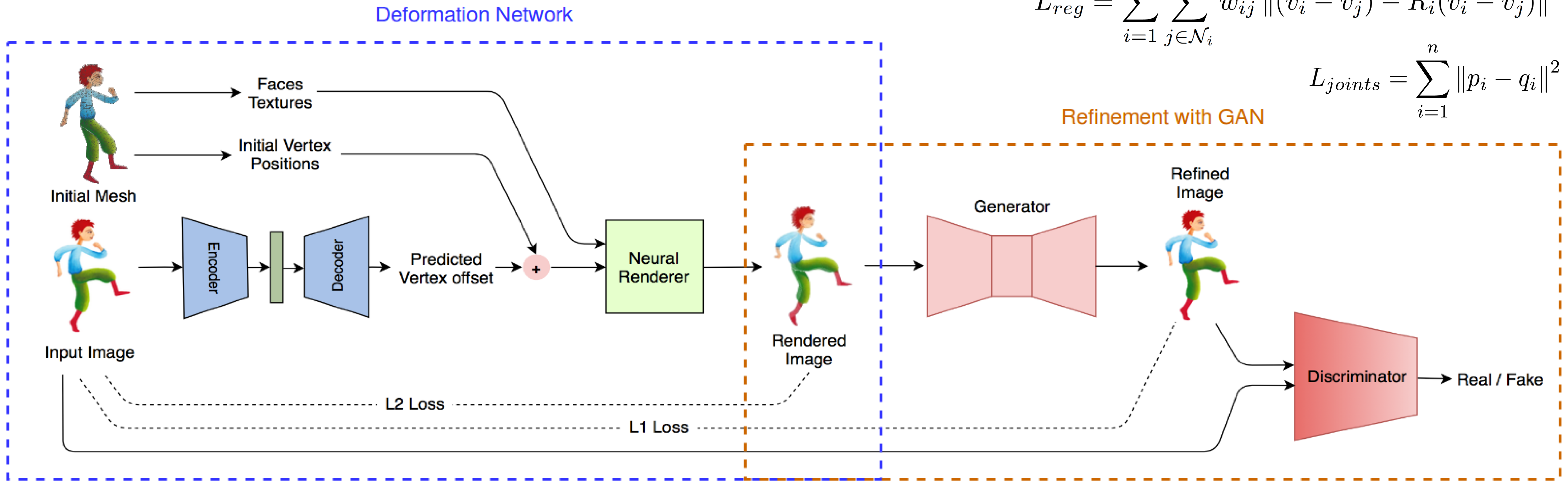
- The model generates in-between frames and constrained deformations


Inbetweens
Deformations

Approach:

- Template-based mesh deformation based on input pose
- Refining the rendered image with a GAN

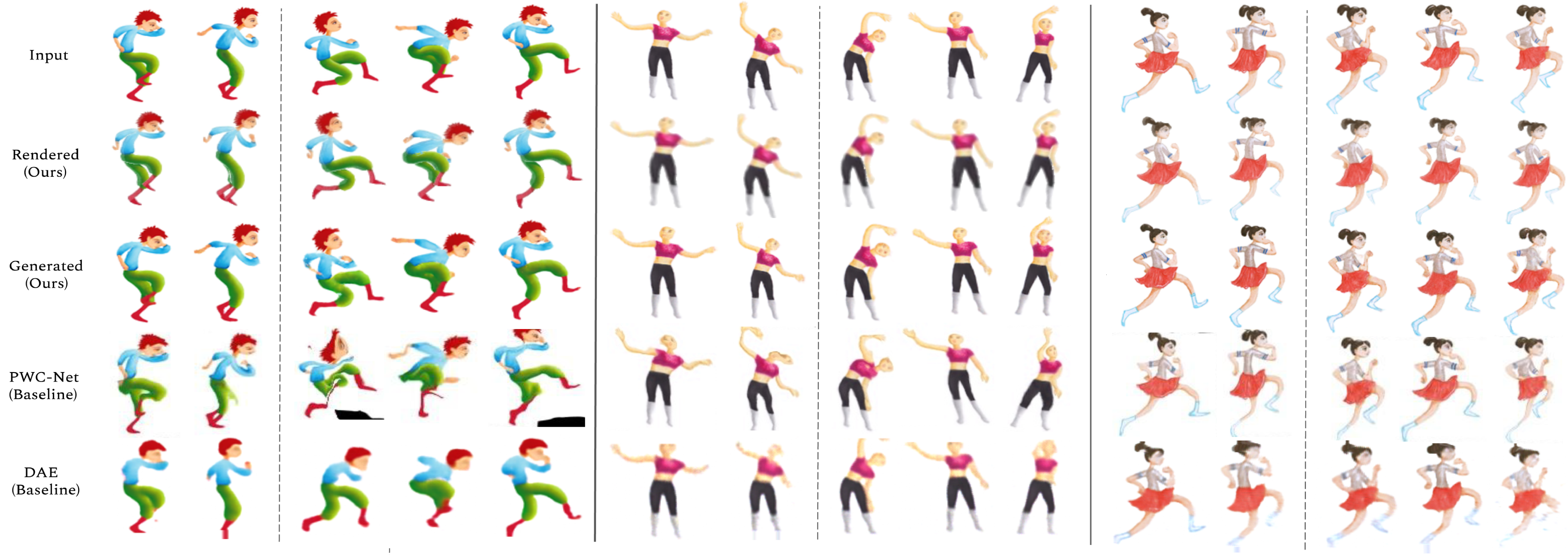
Architecture:

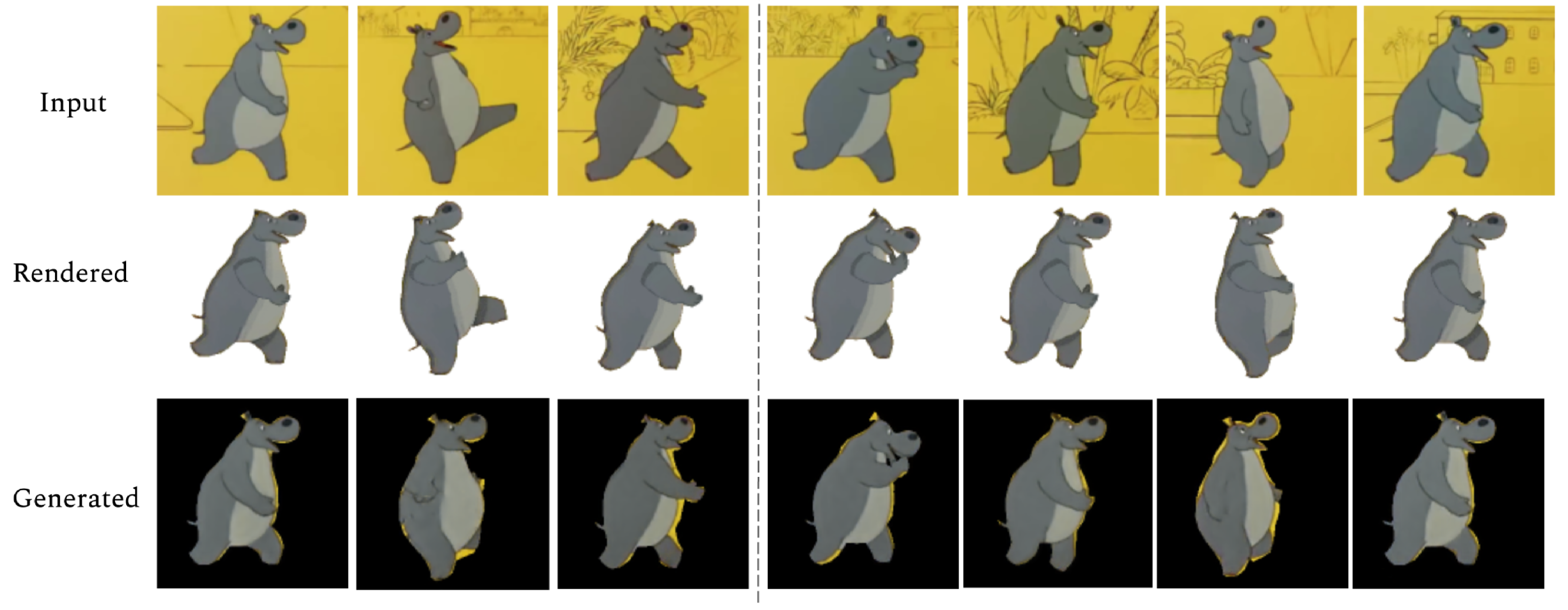

Deformation Network
Initial Mesh
Faces Textures
Initial Vertex Positions
Input Image
Encoder
Decoder
Predicted Vertex offset
+
Neural Renderer
Rendered Image
L2 Loss
L1 Loss
Refinement with GAN
Generator
Refined Image
Discriminator
Real / Fake

$$L_{reg} = \sum_{i=1}^{|V|} \sum_{j \in \mathcal{N}_i} w_{ij} \left\| (\hat{v}_i - \hat{v}_j) - R_i(v_i - v_j) \right\|^2$$

$$L_{joints} = \sum_{i=1}^{n} \|p_i - q_i\|^2$$

Results:

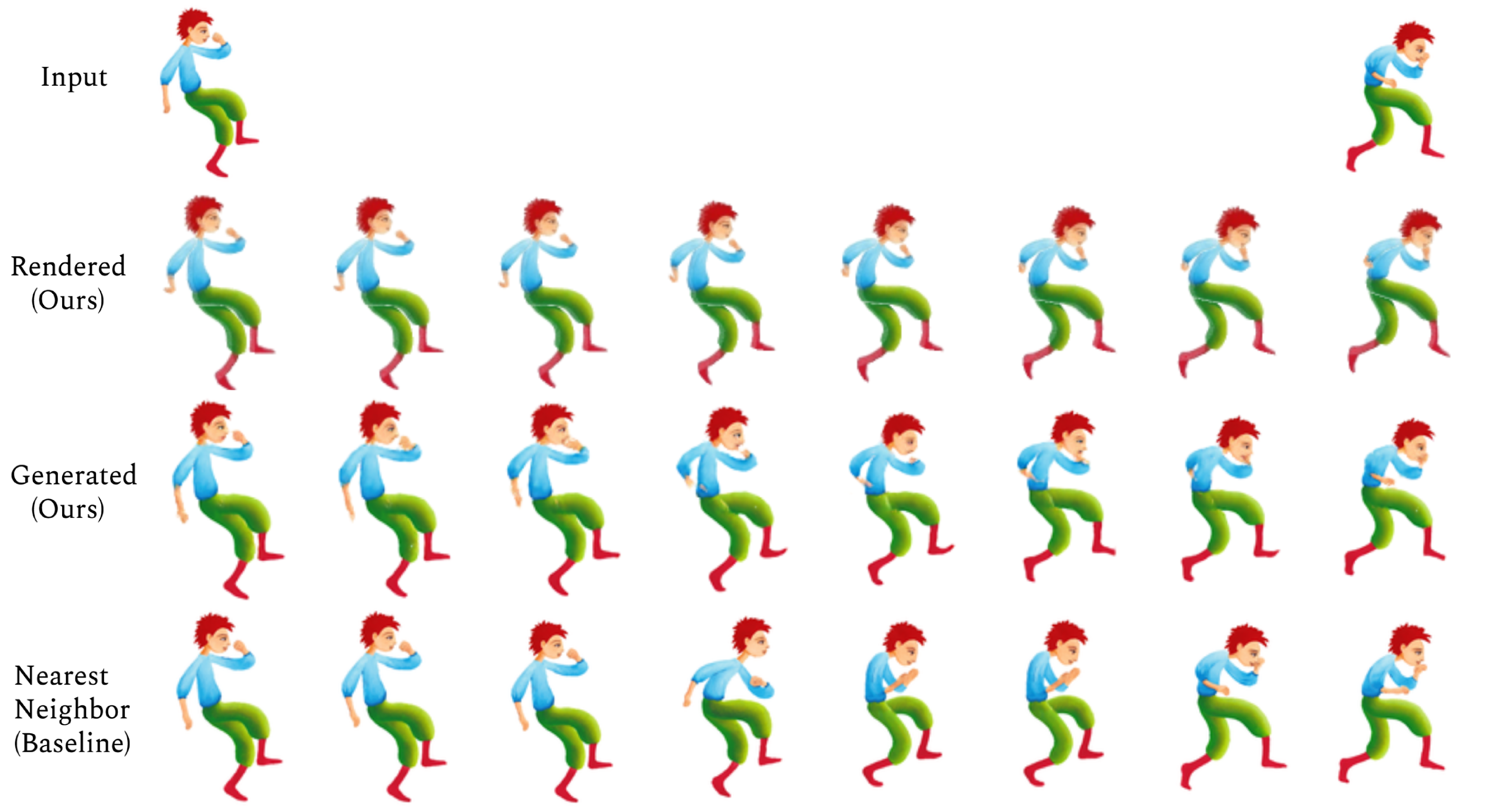

Input
Rendered (Ours)
Generated (Ours)
PWC-Net (Baseline)
DAE (Baseline)


Input
Rendered
Generated

Average L2 distance to ground truth

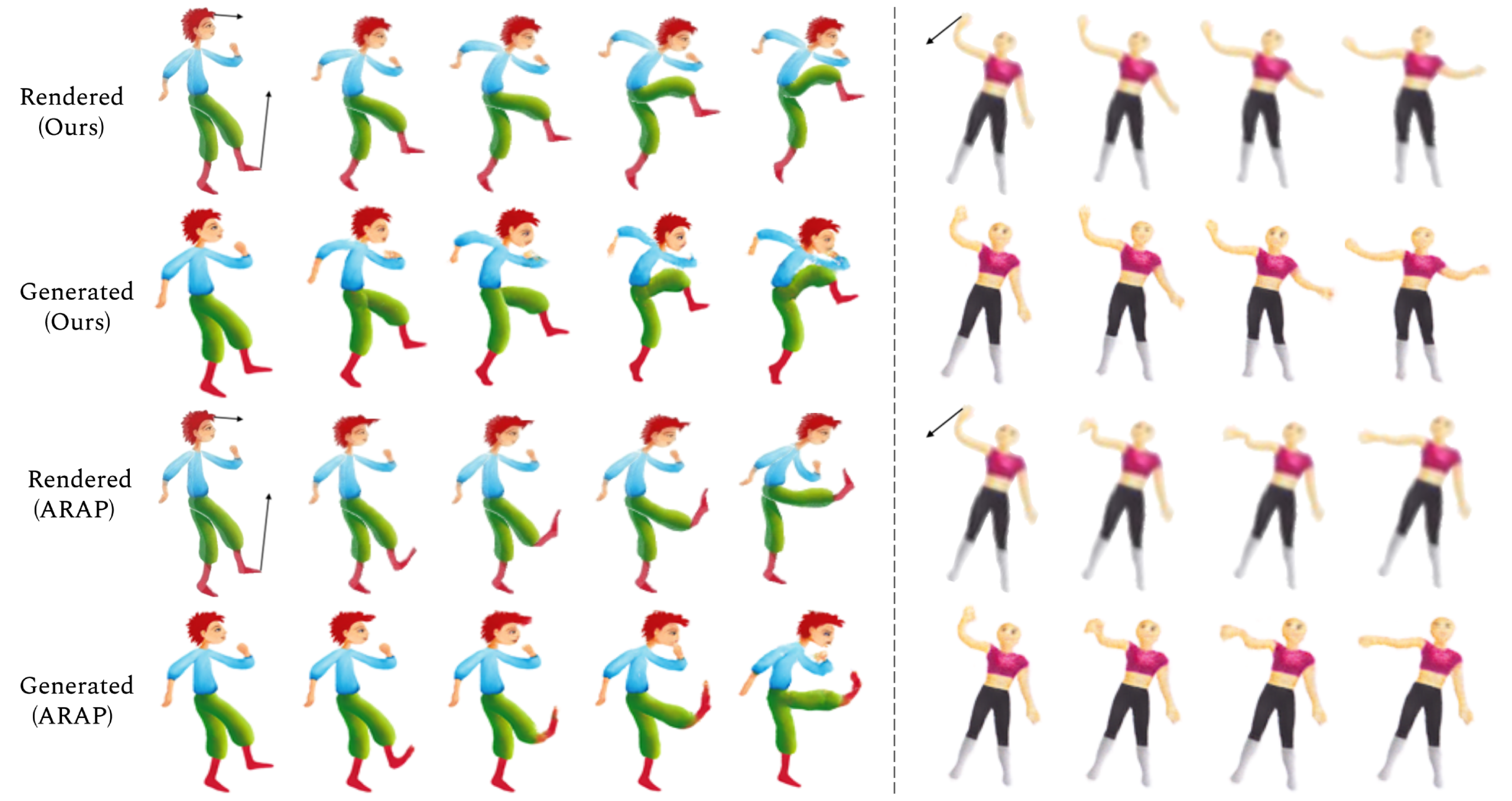
	Char1	Char2	Char3	Char4	Avg
Rendered	819.8	732.7	764.1	738.9	776.1
Generated	**710.0**	**670.5**	**691.7**	**659.2**	**695.3**
PWC-Net	1030.4	1016.1	918.3	734.6	937.1
DAE	1038.3	1007.2	974.8	795.1	981.6

Interpolation:


Input
Rendered (Ours)
Generated (Ours)
Nearest Neighbor (Baseline)

User-contrained Deformation:

$L_{user} = \sum_i \left\| p_i(\mathbf{v}) - p_i^{\text{trg}} \right\|^2$ $\quad L_{deform} = L_{user} + \alpha_1 \cdot L_{reg} + \alpha_2 \cdot L_{joints}$ $\quad z \longleftarrow z - \eta \nabla_z L_{deform}$


Rendered (Ours)
Generated (Ours)
Rendered (ARAP)
Generated (ARAP)

Correspondence Estimation (PCK):

	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$
Ours	**67.18**	**46.39**	**24.17**
UCN	67.07	43.84	21.50
PWC-Net	62.92	40.74	18.47