Skeleton-free Pose Transfer for Stylized 3D Characters

Zhouyingcheng Liao[1], Jimei Yang[2], Jun Saito[2], Gerard Pons-Moll[3,4], Yang Zhou[2]

[1]Saarland University [2]Adobe Research [3]University of Tübingen

[4]Max Planck Institute for Informatics, Saarland Informatics Campus

Code & Paper

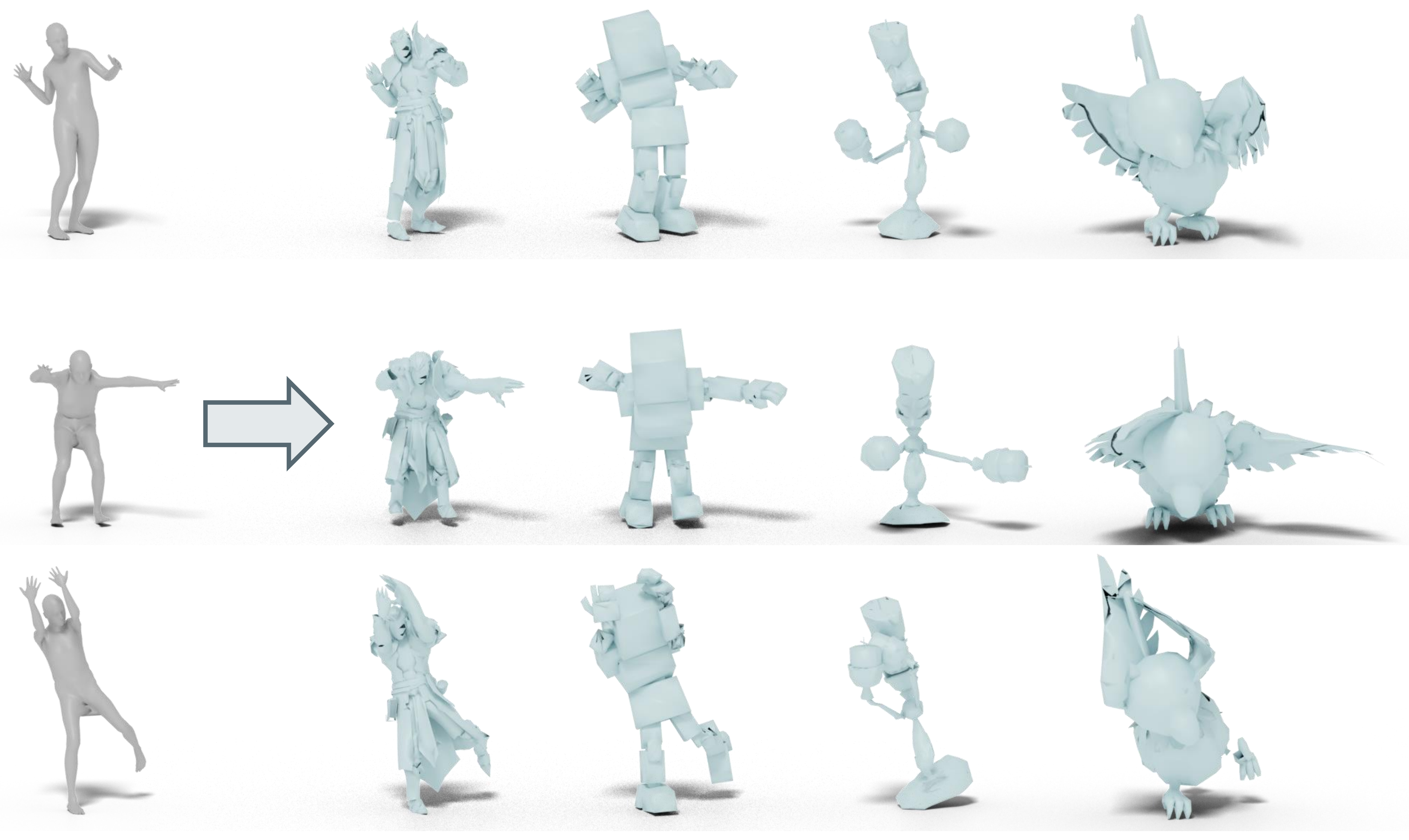
zycliao.com/sfpt

TEL AVIV 2022

EUROPEAN CONFERENCE ON COMPUTER VISION

Goal

Automatic pose transfer between stylized 3D characters without skeletal rigging.



Contribution

We propose a novel method that

- achieves automatic pose transfer between 3D characters.
- supports characters with diverse shapes, topologies, and mesh connectivities.
- does not require any manual intervention or preprocessing, e.g., rigging, skinning, or correspondence labeling.
- is trained end-to-end in a semi-supervised method and does not require correspondence annotations.

Existing Solutions

Skeleton-based[1,2]:

- Require rigging and skinning.
- Require correspondence between skeletons.
- Skeleton topology has to be the same.

Mesh deformation transfer[3] :

- Not robust to diverse shapes and mesh connectivities.
- Prone to mesh distortion.

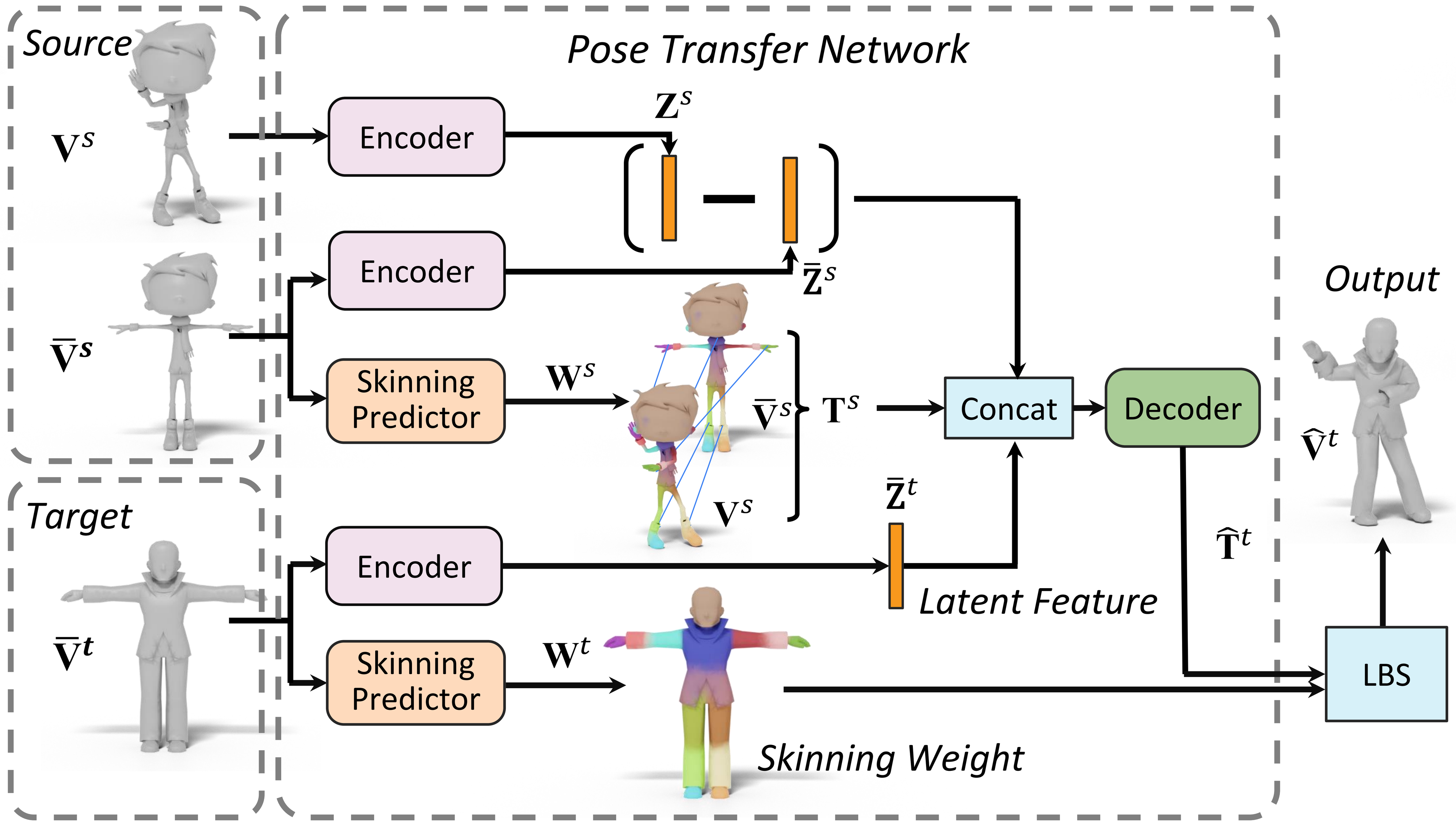
Character Articulation Model

We define K deformation parts for a mesh V. Each part is deformed based on the skinning weight W associated with it. These deformation parts share the same semantics across all characters.

$$\mathbf{V}_i = \sum_{k=1}^{K} w_{i,k} \mathbf{T}_k (\bar{\mathbf{V}}_i - \mathbf{C}_k) \qquad \mathbf{C}_k = \frac{\sum_{i=1}^{N} w_{i,k} \bar{\mathbf{V}}_i}{\sum_{i=1}^{N} w_{i,k}}$$

deformed mesh · skinning weight · rigid transformation · mesh in canonical pose · rotation center for each part

Network



Training and Losses

- Reconstruction loss: Applied when ground truth data is available.

  $L_{rec} = ||\hat{\mathbf{V}}^t - \mathbf{V}^t||_1 \qquad L_{trans} = ||\hat{\mathbf{T}}^t - \mathbf{T}^t||_1$
- Cycle loss: When ground turth is not available, we use the cycle consistency for training. Also, a pseudo ground truth is created by directly copying the source transformation to the target character.

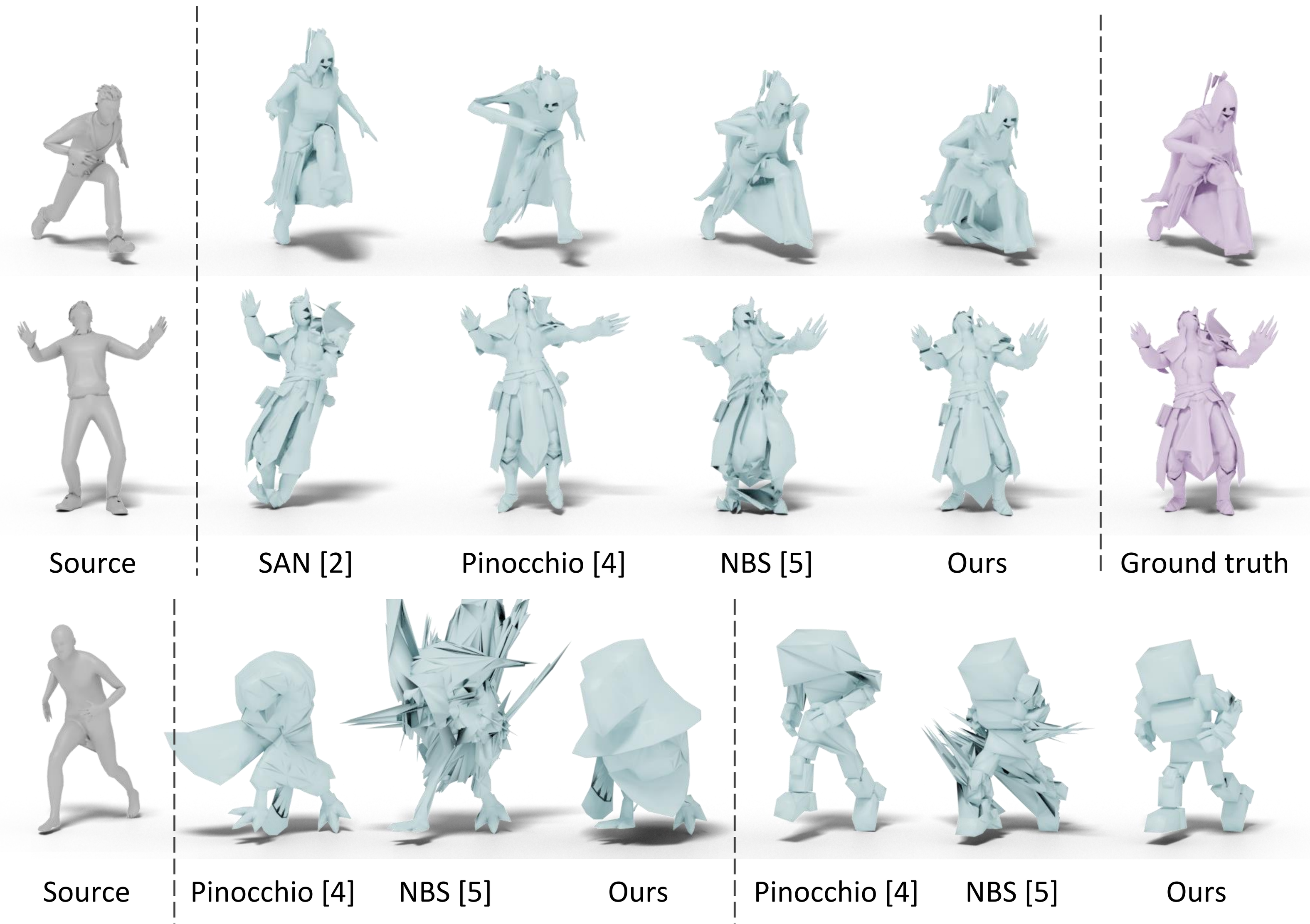
  $L_{cyc} = ||\hat{\mathbf{V}}^s - \mathbf{V}^s||_1 + w_{\text{pseudo}}||\hat{\mathbf{V}}^t - \tilde{\mathbf{V}}^t||_1$
- Skinning weight loss: If two vertices belong to the same body part on the ground truth skinning, their predicted part should also be the same.

  $L_{skin} = \gamma_{i,j} \sum_{k=1}^{K} (w_{i,k} log(w_{i,k}) - w_{i,k} log(w_{j,k}))$
- Edge length regularization: the edge lengths before and after deformation should be similar. It prevents the character intrinsic shape.

  $L_{edge} = \sum_{\{i,j\} \in \mathcal{E}} |\, ||\hat{\mathbf{V}}^t_i - \hat{\mathbf{V}}^t_j||_2 - ||\mathbf{V}^t_i - \mathbf{V}^t_j||_2 \,|$

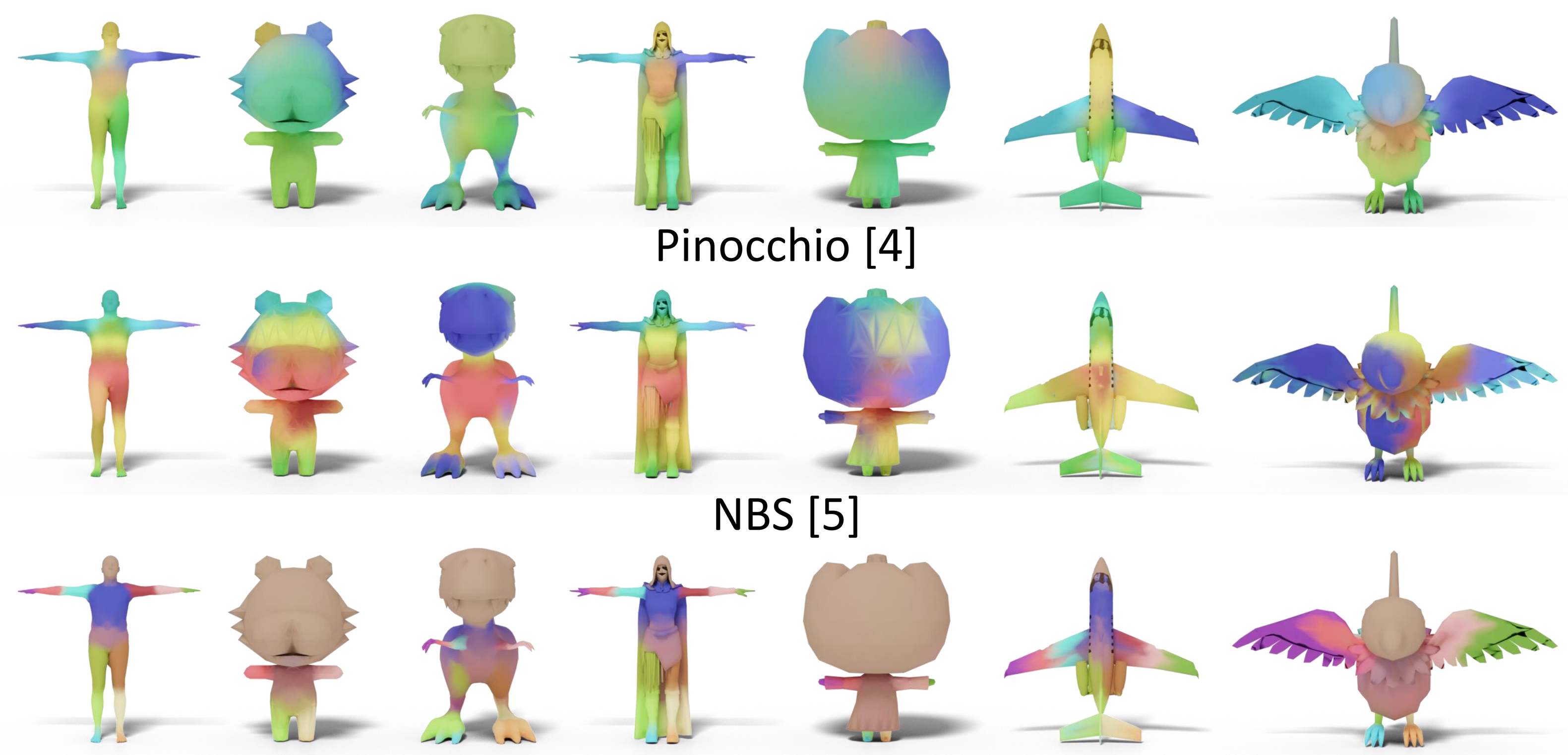
Pose Transfer Results

Compared to other methods, our results match the source pose the best and preserve the character shape.



Skinning Weight Visualization

Our skinning paints characters consistently on semantic regions while the comparison methods fail on some body parts.



References

[1] Villegas, R., et al. "Neural kinematic networks for unsupervised motion retargetting." *CVPR*, 2018.
[2] Aberman, K., et al. "Skeleton-aware networks for deep motion retargeting." *TOG*, 2020.
[3] Zhou, K., et al. "Unsupervised shape and pose disentanglement for 3d meshes." *ECCV*, 2020.
[4] Baran, I., et al. "Automatic rigging and animation of 3d characters." *TOG*, 2007.
[5] Li, P., et al. "Learning skeletal articulations with neural blend shapes." *TOG*, 2021.