EUROGRAPHICS 2025 / A. Bousseau and A. Dai
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 44 (2025), Number 2

ASMR: Adaptive Skeleton-Mesh Rigging and Skinning via 2D Generative Prior

Seokhyeon Hong* Soojin Choi* Chaelin Kim Sihun Cha Junyong Noh

Visual Media Lab, KAIST

1. Architecture Details

Tables 1 and 2 show the detailed network architectures of the Skeletal Articulation Prediction module and Skinning Weight Prediction module, respectively. The names correspond to those in Figure 2 of the main paper, except the Skeleton Decoder of Skeletal Articulation Prediction module in Table 1 refers to the combination of the Attention and MLP layers.

2. Additional Experiments

2.1. Evaluations on Out-of-Domain Characters

While our test dataset includes a wide range of stylized characters from Mixamo [Mix], we conducted additional experiments on out-of-domain characters to further evaluate the generalization capabilities of our method. Specifically, we employed characters in the T-pose from the RigNet-v1 dataset [XZK*20], which are unseen during training. These characters were used as source meshes and deformed using source skeletons and motions from our database.

As shown in Figure 1, our method showed robust performance in generating target skeletons that align with the character mesh despite variations in the structures and shapes of the input skeletons. Furthermore, our method produced plausible skinning weights aligned seamlessly with both the source mesh and the predicted target skeletons, deriving plausible deformations for out-of-domain characters, which have shapes and body ratios distinct from Mixamo characters. These results highlight the robustness and generalizability of our approach on unseen characters. For animation results with additional characters, please see the supplementary video.

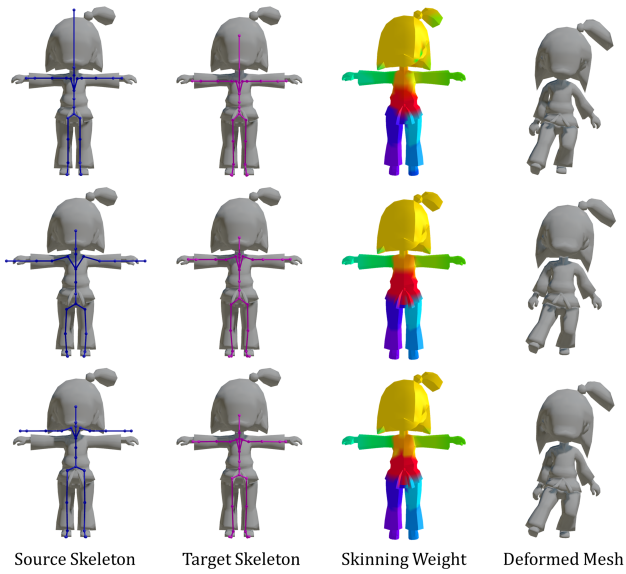

Figure 1: *Qualitative results on out-of-domain characters from the RigNet-v1 dataset. Skeletons are overlaid on source meshes to demonstrate alignments between them.*

2.2. Evaluations on Individual Impact

To rigorously evaluate the impact of individual changes in each independent variable, we conducted two additional experiments: (i) comparison of the performance of the skinning weight prediction module across different baselines, and (ii) analysis on the impacts of individual changes in each element of the skeletal configuration, including body scale, bone lengths, and the number of joints.

Performance of Skinning Weight Prediction To solely compare the performance of the skinning weight prediction components across different baselines, we measured the skinning and deformation metrics using identical skeletons for all methods. Specifically, we used the source skeleton precisely aligned with the source mesh as input to the skinning weight prediction modules of each baseline to generate skinning weights, while bypassing their skeleton prediction modules. For deformation metrics, we used the source poses directly, instead of retargeting poses using SAME [LKP*23]. NBS [LAH*21] was excluded from this experiment because its skinning weight prediction relies solely on the mesh and does not utilize skeleton inputs. To obtain the results of Pinocchio [BP07], we followed the experimental setup of NBS [LAH*21] that uses the auto-skinning tool provided by Blender, which is implemented based on the algorithm of Pinocchio.

As shown in Table 3, Pinocchio [BP07] achieved the best performance across all metrics, with our method producing comparable results. While ours did not achieve the best quantitative scores, the

Table 1: *Network architectures of the Skeletal Articulation Prediction module.*

Name	Layers	Channels	Attention Heads
Mesh Encoder	Linear - BatchNorm - ReLU - Dropout	$35 \rightarrow 256$	-
	Linear - BatchNorm - ReLU - Dropout	$256 \rightarrow 256$	-
	Linear - BatchNorm - ReLU - Dropout	$256 \rightarrow 32$	-
	Pooling & Concatenation	$32 \rightarrow 64$	-
	Linear	$64 \rightarrow 32$	-
Skeleton Encoder	GAT - BatchNorm - ReLU - Dropout	$6 \rightarrow 16$	16
	GAT - BatchNorm - ReLU - Dropout	$256 \rightarrow 16$	16
	GAT - BatchNorm - ReLU - Dropout	$256 \rightarrow 16$	16
	GAT - BatchNorm - Dropout	$256 \rightarrow 32$	1
Skeleton Decoder	CrossAttention (QKV) - Dropout - Residual Connection	$32 \rightarrow 2$	16
	Linear - ReLU	$32 \rightarrow 32$	-
	Linear - ReLU	$32 \rightarrow 32$	-
	Linear - BatchNorm	$32 \rightarrow 3$	-

Table 2: *Network architectures of the Skinning Weight Prediction module.*

Name	Layers	Channels	Attention Heads
Skeleton Encoder	GAT - BatchNorm - ReLU - Dropout	$6 \rightarrow 16$	16
	GAT - BatchNorm - ReLU - Dropout	$256 \rightarrow 16$	16
	GAT - BatchNorm - ReLU - Dropout	$256 \rightarrow 16$	16
	GAT - BatchNorm - Dropout	$256 \rightarrow 32$	1
Skinnning Weight Predictor	CrossAttention (QK)	$N_V \times 32$ and $N_J \times 32 \rightarrow N_V \times N_J$	1

Table 3: *Quantitative results on skinning and deformation using identical source skeletons that precisely align with the source meshes. The best result for each column is in bold.*

	Skinning L1↓	CD↓	ADE↓	MDE↓	ELS↑
Pinocchio	**0.0188**	**2.83**	**2.56**	**13.74**	**0.96**
RigNet	0.0178	2.95	4.71	95.60	0.10
Ours	0.0469	4.72	4.72	21.52	0.89

discrepancies were minor with visually imperceptible variations as shown in Figure 2. Furthermore, our method still has an advantage in generalizability of rigging and skinning for various skeletal configurations, in that our method produced results consistent to Table 2 of the main paper, whereas Pinocchio produced results with significant deviation. RigNet [XZK*20] achieved comparable results in CD and ADE metrics to other methods, but its results exhibited noticeable artifacts, such as stretched vertices, which were reflected in significantly higher MDE and lower ELS values than those from other methods.

Changes of Individual Skeletal Configuration Starting with an initial source skeleton containing 25 joints, precisely aligned in size and proportion with the source mesh, we modified three key elements of the skeletal configuration to generate new source skeletons, as follows:

- Body scale: A uniform scaling factor of 0.5 was applied to all bones to adjust the overall body size.

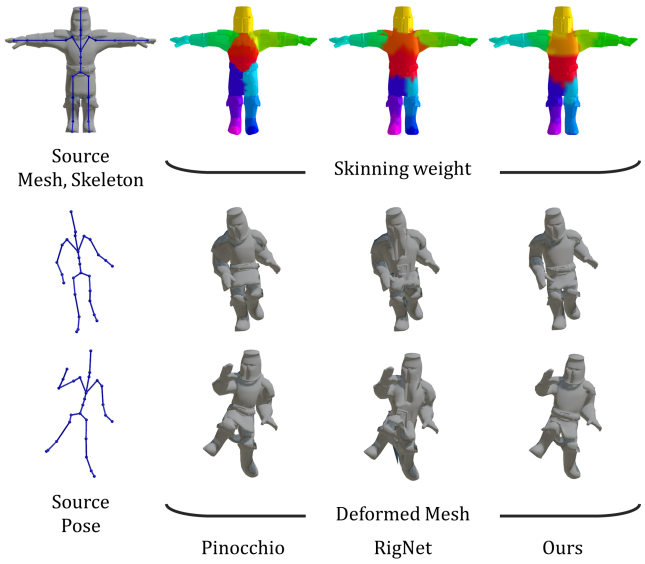

Figure 2: *Qualitative comparison with baselines on predicted skinning weights and mesh deformation. The first column shows the source mesh and skeleton given to the skinning weight prediction module of each baseline, with source poses to deform the mesh. In the second to fourth columns, the first row demonstrates the skinning weights predicted by each baseline, while the second and third rows show the resulting deformed meshes.*

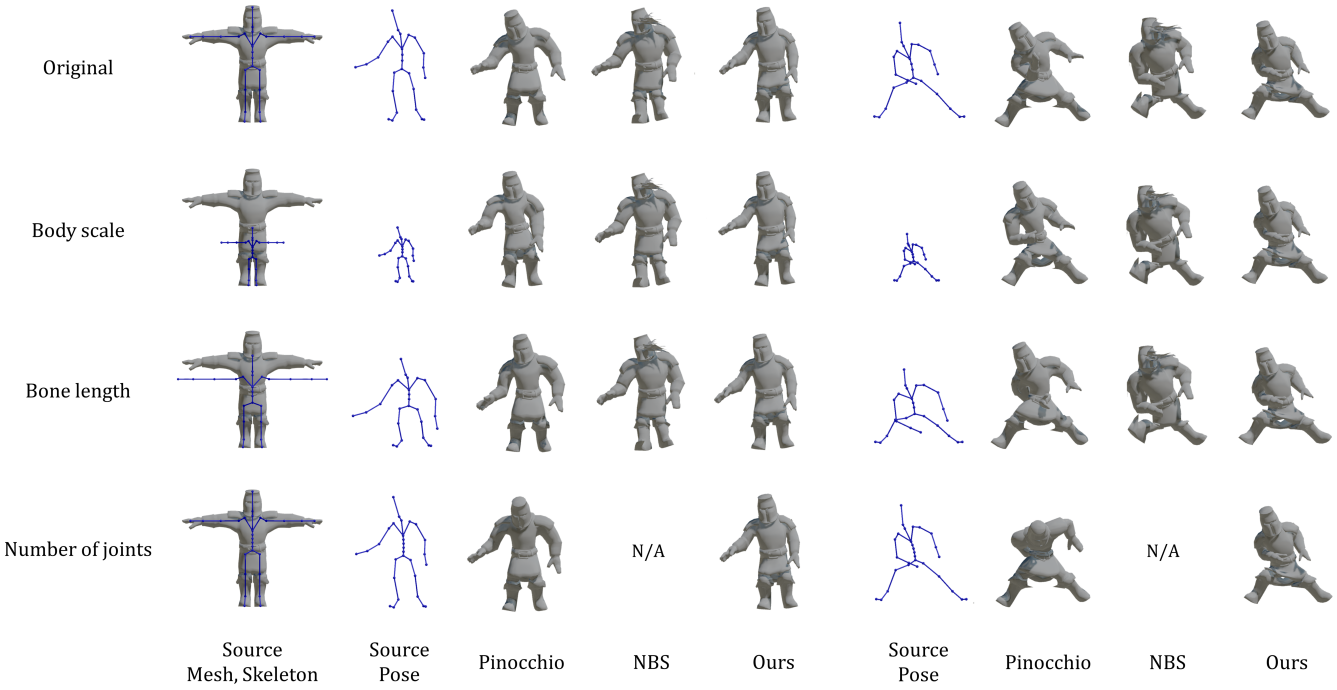

Figure 3: *Qualitative comparison of deformation results under varying skeletal configurations. Beginning with a source skeleton that precisely aligns with the source mesh, each subsequent row includes the source skeleton generated by modifying one factor: body scale, bone length, or the number of joints, respectively, while maintaining the other elements fixed.*

- Bone length: A non-uniform scaling was applied using scaling factors of 0.8 and 1.2 along the vertical and lateral axes of each joint, respectively.
- Number of joints: The number of joints was randomly adjusted, resulting in two additional joints to the initial skeleton.

Figure 3 shows the generated source skeletons with their corresponding poses, along with the deformed meshes driven by each method based on the source mesh. Our approach consistently produced plausible deformations that follow the source poses, regardless of changes in the skeletal configuration. In contrast, Pinocchio [BP07] failed to preserve the volume of the source mesh, resulting in excessive expansion around shoulders and contraction around the spine. NBS [LAH*21] resulted in distorted meshes due to improper skinning weights applied to certain body parts. Because NBS relies on a pre-defined set of joints, the results for the last source skeleton, which contains additional joints, were excluded.

References

[BP07] BARAN I., POPOVIĆ J.: Automatic rigging and animation of 3d characters. *ACM Transactions on graphics (TOG) 26*, 3 (2007), 72–es. 1, 3

[LAH*21] LI P., ABERMAN K., HANOCKA R., LIU L., SORKINE-HORNUNG O., CHEN B.: Learning skeletal articulations with neural blend shapes. *ACM Transactions on Graphics (TOG) 40*, 4 (2021), 1–15. 1, 3

[LKP*23] LEE S., KANG T., PARK J., LEE J., WON J.: Same: Skeleton-agnostic motion embedding for character animation. In *SIGGRAPH Asia 2023 Conference Papers* (2023), pp. 1–11. 1

[Mix] Mixamo. https://www.mixamo.com/. Accessed: 2024-08-21. 1

[XZK*20] XU Z., ZHOU Y., KALOGERAKIS E., LANDRETH C., SINGH K.: Rignet: Neural rigging for articulated characters. *arXiv preprint arXiv:2005.00559* (2020). 1, 2