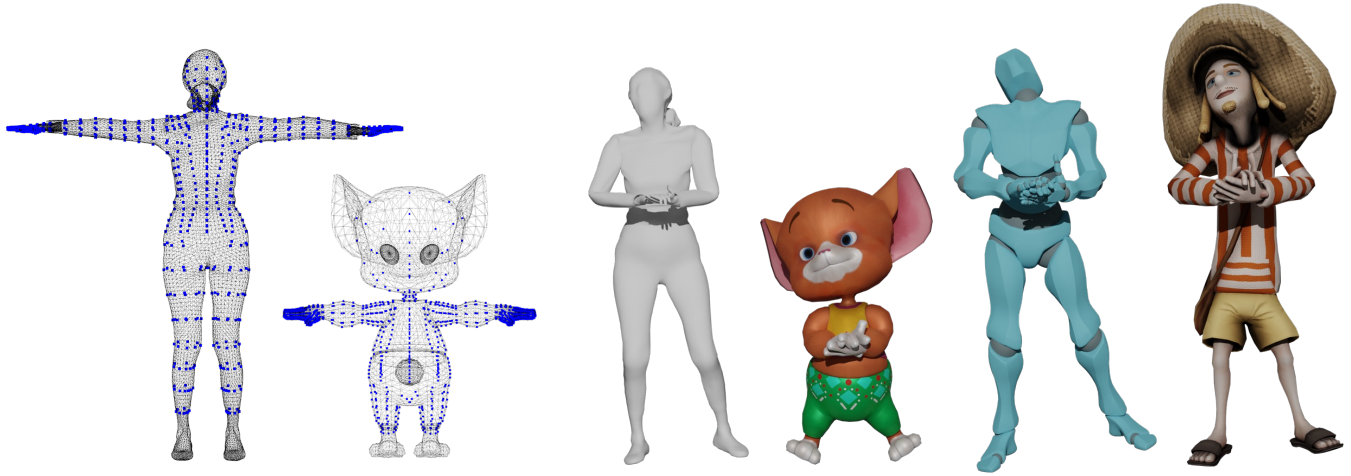
Skinned Motion Retargeting with Dense Geometric Interaction Perception

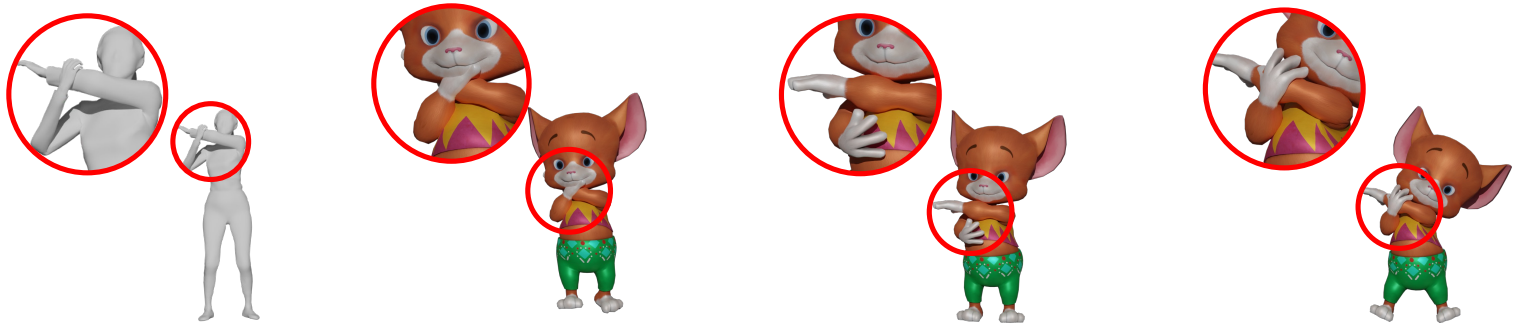
Zijie Ye[1], Jia-Wei Liu[2], Jia Jia[1], Shikun Sun[1], Mike Zheng Shou[2]

[1]Tsinghua University, [2]National University of Singapore



Problem Statement

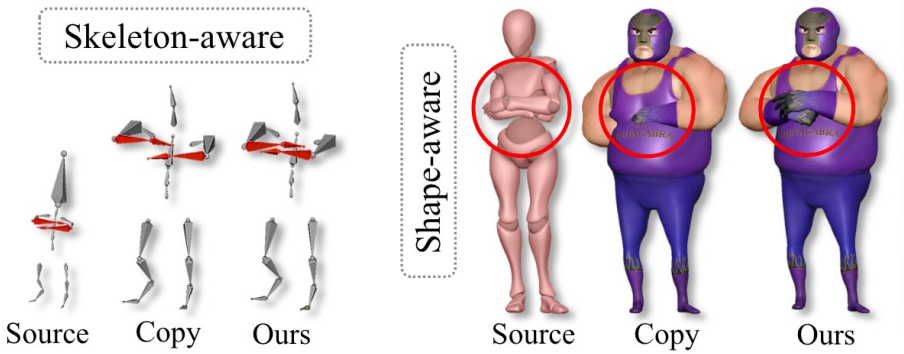
- **Motion Retargeting:** Mapping the motion of source character A to target character B
- **Potential Issue:** Contact Mismatch, Interpenetration
- **Goal:** Maintaining key characteristics by preserving contact & reducing interpenetration



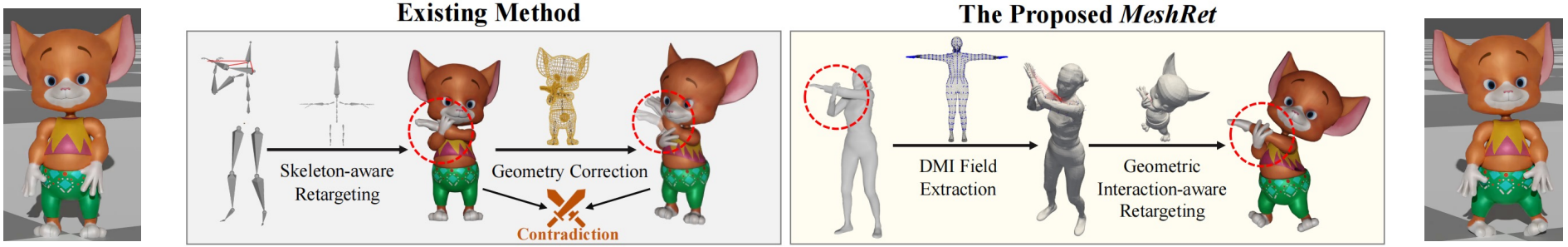
Source | Interpenetration | Contact Mismatch | Desirable Result

Motivation

- Previous Methods
  - Skeletal motion retargeting
    - NKN (CVPR 2018), Villegas et al.
    - PMnet (BMVC 2019), Lim et al.
    - SAN (TOG 2020), Aberman et al.
  - Geometry-aware motion retargeting
    - Contact-aware (CVPR 2021), Villegas et a
    - R2ET (CVPR 2023), Zhang et al.
    - SMTNet (CVPR 2024), Zhang et al.



- Our Goal
  - Instead of correcting skeletal retargeting results, we model dense interactions between body geometries directly.



Challenges & Solutions

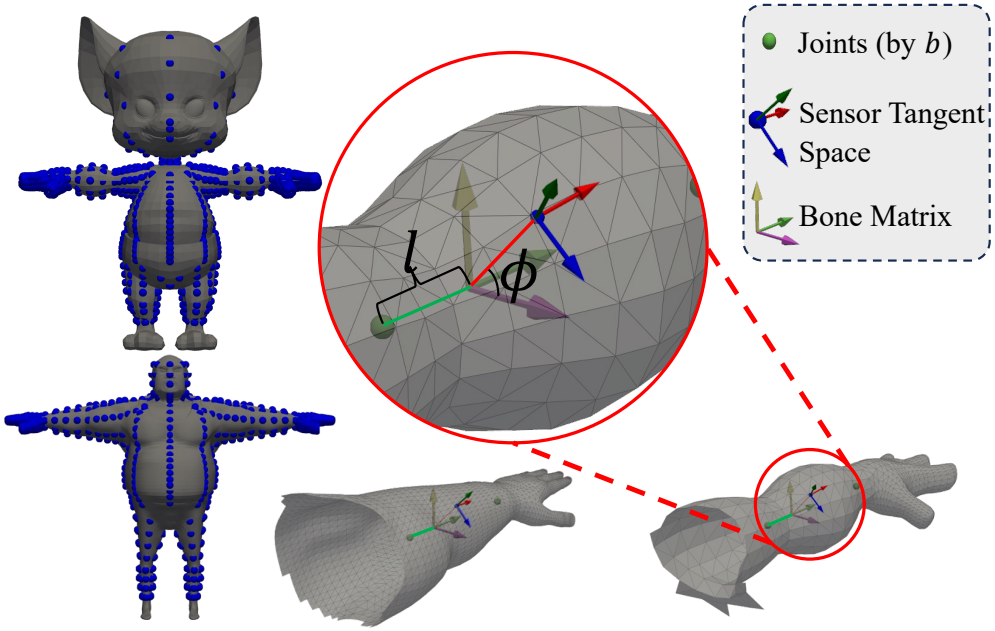
- Challenge: Lack of dense mesh correspondence
- Solution: Semantically consistent sensors (SCS)
  - Inspired by the medial axis inverse transform
  - Automatically derive dense mesh correspondence from sparse skeleton correspondence
  - Each virtual sensor is described by a semantic coordinate (b, l, ϕ)

Algorithm 1: Derive Semantically Consistent Sensors from Semantic Coordinate

Input: Mesh $\mathbf{O}$, joint locations $\mathbf{J} \in \mathbb{R}^{N\times 3}$, bone index $b \in \{0, 1, \cdots, N\}$, origin parameter $l \in [0, 1)$, direction parameter $\phi \in [0, 2\pi)$

Output: Sensor feature $\mathbf{s} \in \mathbb{R}^{4\times 3}$

$i_{\text{parent}} \leftarrow \text{bone\_parent\_joint}(b)$, $i_{\text{child}} \leftarrow \text{bone\_child\_joint}(b)$;
$\mathbf{x}_{\text{parent}} \leftarrow \mathbf{J}[i_{\text{parent}}]$, $\mathbf{x}_{\text{child}} \leftarrow \mathbf{J}[i_{\text{child}}]$;
$\mathbf{o} \leftarrow (1-l)\mathbf{x}_{\text{parent}} + l\mathbf{x}_{\text{child}}$; /* Ray origin */
$\mathbf{d}_{\text{forward}} \leftarrow \text{forward\_direction}(\mathbf{O})$; /* Face forward direction */
$\mathbf{d}_{\text{bone}} \leftarrow \text{normalize}(\mathbf{x}_{\text{child}} - \mathbf{x}_{\text{parent}})$; /* Bone unit direction vector */
$\mathbf{d}_{\text{other}} \leftarrow \mathbf{d}_{\text{forward}} \times \mathbf{d}_{\text{bone}}$;
$\mathbf{n} \leftarrow \cos(\phi)\mathbf{d}_{\text{forward}} + \sin(\phi)\mathbf{d}_{\text{other}}$; /* Ray direction */
$\mathbf{B} \leftarrow \text{bone\_mesh}(\mathbf{O}, \text{b})$; /* Bone associated mesh */
$\mathbf{r} \leftarrow \text{ray}(\mathbf{o}, \mathbf{n})$;
$\mathbf{p} \leftarrow \text{ray\_mesh\_intersection}(\mathbf{B}, \mathbf{r})$;
if $\mathbf{p} \neq \emptyset$ **then**
  $\mathbf{t} \leftarrow \text{tangent\_matrix}(\mathbf{x}_{\text{p}}, \mathbf{B})$;
  $\mathbf{s} \leftarrow \text{concat}(\mathbf{p}, \mathbf{t})$;
else
  $\mathbf{s} \leftarrow \mathbf{0}$;
end

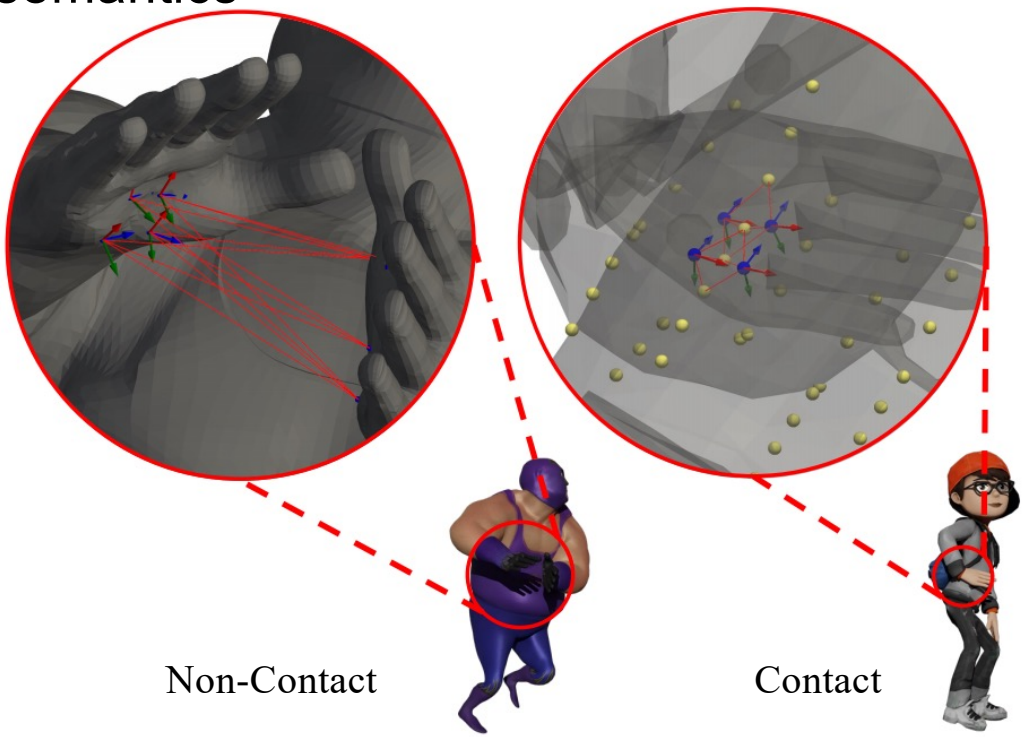


Challenges & Solutions

- Challenge: Model dense mesh interaction between body parts
- Solution: Dense mesh interaction (DMI) field
  - Relative sensor positions in sensor tangent space
  - Semantic coordinates instead of spatial coordinates
  - Captures contact & non-contact semantics

$$\mathbf{d}^{t,i,j} = \mathbf{t}_i^{-1}(\mathbf{p}_j^t - \mathbf{p}_i^t),$$

$$\overline{\mathbf{D}}^t = \{(\mathbf{d}^{t,i,j}, b_i, b_j, l_i, l_j, \phi_i, \phi_j)\}_{i=1:S}^{j=1:S}$$



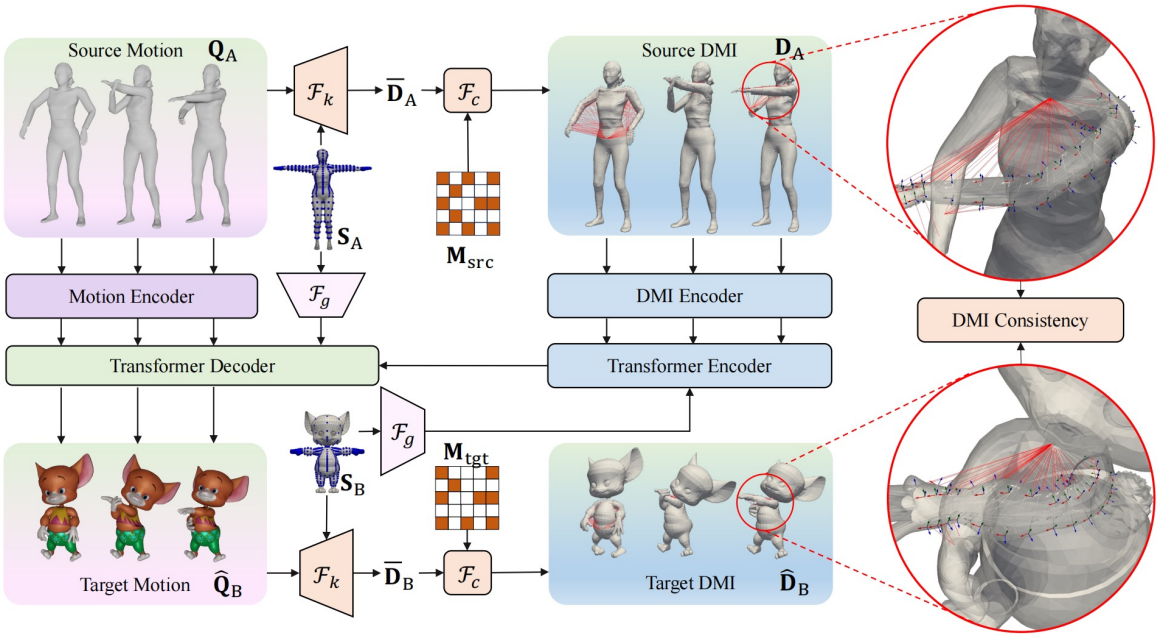
Pipeline

- DMI Consistency Loss

$$\mathcal{L}_{\text{dmi}} = -\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{L} c(k,l)\frac{\mathbf{d}_{\text{A}}^{t,k,l}\cdot\hat{\mathbf{d}}_{\text{B}}^{t,k,l}}{||\mathbf{d}_{\text{A}}^{t,k,l}||_2\cdot||\hat{\mathbf{d}}_{\text{B}}^{t,k,l}||_2}$$

- End-effector loss

$$\mathcal{L}_{\text{ef}} = \frac{1}{T|\mathcal{X}|}\sum_{t=1}^{T}\sum_{i\in\mathcal{X}}||R(\mathbf{Q}_{\text{A}}^{t},i) - R(\hat{\mathbf{Q}}_{\text{B}}^{t},i)||$$



- Key designs
  - SCS for dense mesh correspondence
  - DMI field to model complex geometric interactions
  - Sparsify $\overline{\mathbf{D}} \in O(S^2)$ by sensor pair selection, $\mathbf{D} = \mathcal{F}_c(\overline{\mathbf{D}}) \in O(S)$
  - Align both skeletal semantics and geometric semantics in one single stage

Comparison