

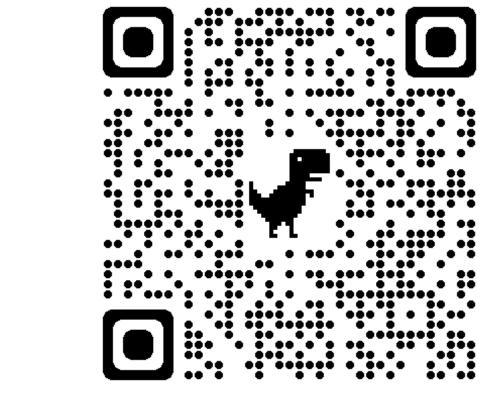
MS-MANO: Enabling Hand Pose Tracking with Biomechanical Constraints

Pengfei Xie^{1,*}, Wenqiang Xu^{2,*}, Tutian Tang², Zhenjun Yu², Cewu Lu²

¹Southeast University

²Shanghai Jiao Tong University

*Equal Contribution





Introduction

Human hand motion relies on the musculoskeletal system, where brain signals to muscles generate torque for joint movement. Most visual hand dynamics frameworks overlook this, modeling hand dynamics as multiody systems, which can result in unnatural, robot-like movements.

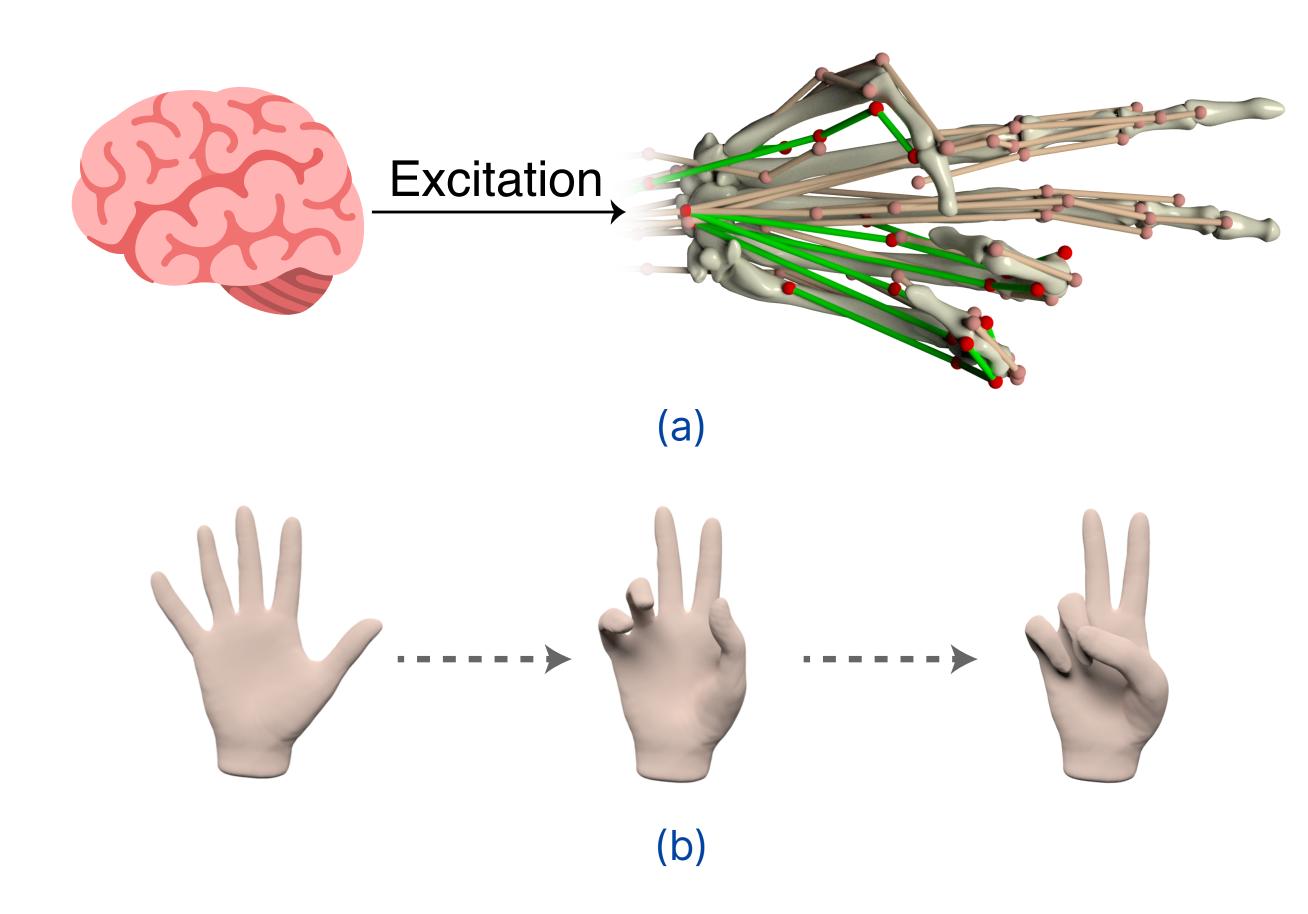


Figure 1. The physiological mechanism of hand dynamics. (a) The excitation signal originating from the brain triggers the contraction or relaxation of muscles. The triggered muscle segments are illustrated in green, while the relaxed ones are in brown. (b) The muscle contraction triggered by excitation manifests as the movement of the hand in appearance.

To address this, we

- Introduce MS-MANO, a musculoskeletal extension of the MANO hand model, with learning support and shape adaptability.
- Demonstrate MS-MANO's effectiveness in hand pose tracking using the BioPR framework, benchmarking on DexYCB and Oaklnk datasets.

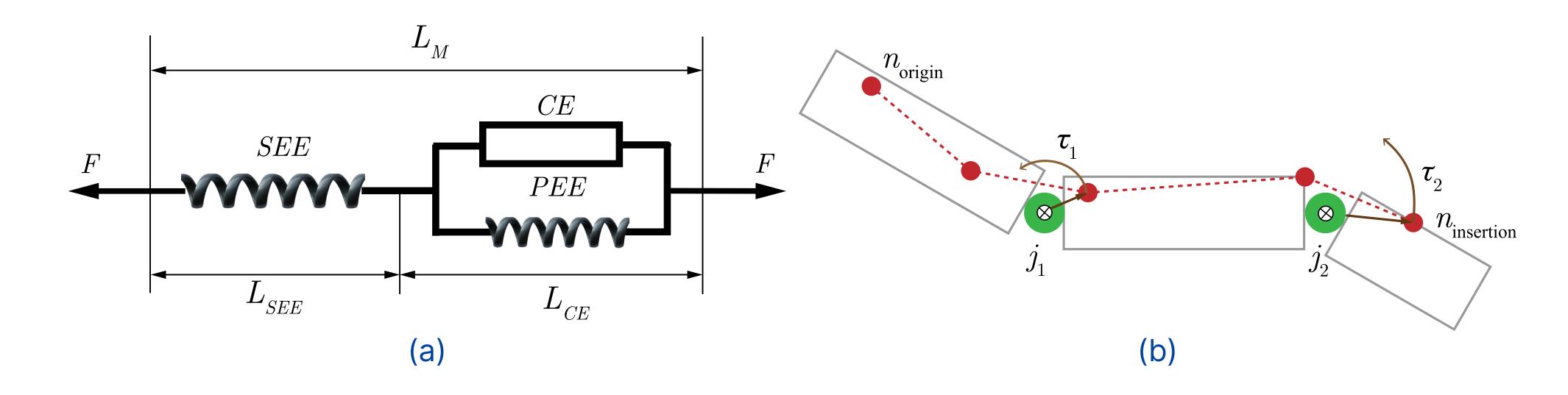


Figure 2. The hill-type muscle. (a) Each muscle segment is composed of the contractile element CE, the parallel elastic element PEE, and the serial elastic element SEE. (b) Each muscle segment originates from a certain point n_{origin} and ends at $n_{\text{insertion}}$. A joint j connects two bones. Once triggered, the muscle segment can apply torque τ on the joint.

Biomechanical Pose Refiner

BioPR refines hand pose estimates by incorporating muscle dynamics using **IDNet** (Inverse Dynamics) and a simulation-in-the-loop framework. Initial poses are corrected via a reference pose p^{ref} generated by forward kinematics with muscle excitation signals a from IDNet. A refinement network then produces a refined pose $p^{\text{refined}} = \mathcal{M}(p^{\text{pred}}_{t+1}, p^{\text{ref}}_{t+1})$.

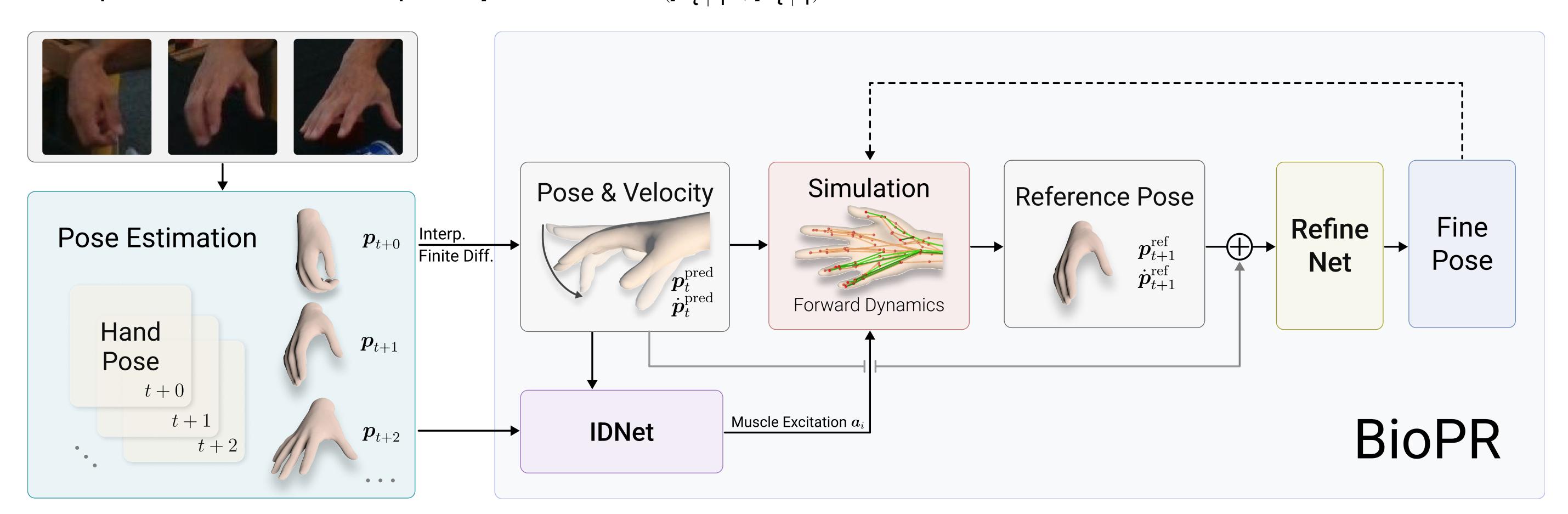


Figure 3. The simulation-in-the-loop pipeline of BioPR. Given a sequence of RGB images and the corresponding predictions of an existing hand pose estimator, BioPR first interpolates and differentiates the poses to get the joint velocities. Then, the IDNet is used to infer the muscle excitation signals. The joint poses, velocities, excitation signals, and the poses of the previous frame (denoted by dotted lines) are sent into the simulator, which will generate the next reference pose by forward kinematics. The Refine Net will do the final refinement based on the pose, velocity, and reference pose. On the next frame, the refined pose can be fed back to the simulator.

MusculoSkeletal-MANO

MS-MANO integrates hill-type-based (fig. 2) musculoskeletal dynamics with MANO, mapping Open-Sim's bone-centric muscles into a joint-centric representation.

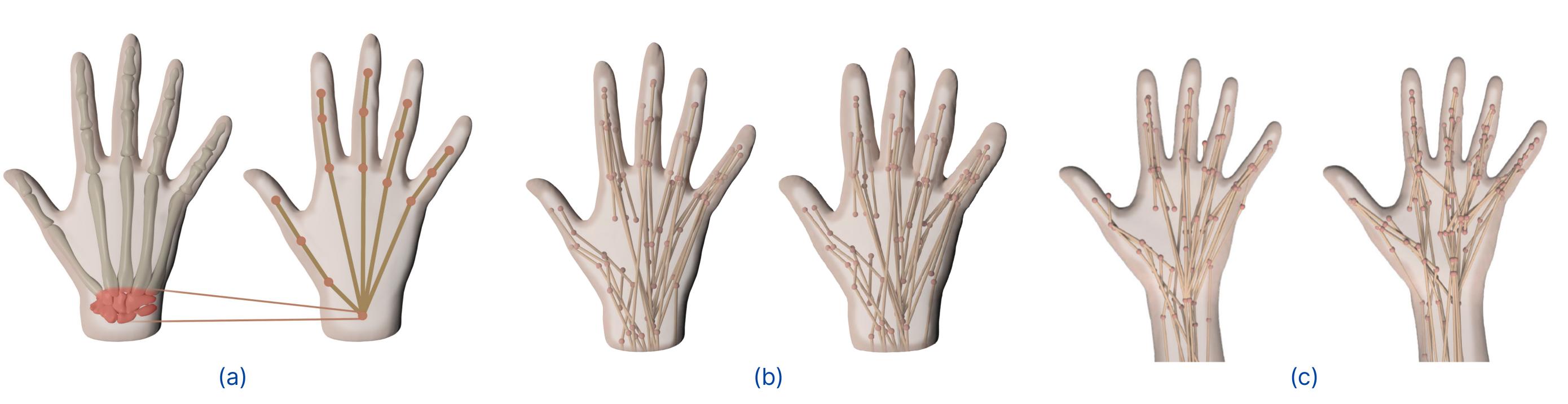


Figure 4. Joint-centric muscle adaptation. (a) A set of smaller bones in the MyoHand model is mapped into a single joint in the MANO model. (b) The bone-centric muscle segments can adapt to different shapes. (c) (Left) The raw skeleton after the automatic mapping will result in issues like intersection. (Right) The manually revised skeleton can perfectly fit with the MANO model.

