Final Project Presentation

Paper Implementation: [CVPR 2023] MoFusion: A Framework for Denoising-Diffusion-based Motion Synthesis

Group 32
110612117 張仲瑜
110550022 賴柏允

Goal

To implement Mofusion, a diffusion-based motion synthesis model that can generate a sequence of motion based on the given text prompt.

Implementation details

- Train the model on Global position
  instead of local joint position and rotation
    - converting the joint positions and rotations to global coordinates with forward kinematics
    - resulting in 22 joints each with their x, y and z coordinates for a total of 66 features.

Implementation details

- Apply time-varying weight to avoid the unstable training when the predicted motion includes lots of noise

  - multiplying all the skeletal losses by a weight
    $\lambda(t) * k = \bar{\alpha}(t)$, where alpha_bar(t)
    is the cumulative product of all alphas up to time step t.

  - alpha = 1 - noise schedule(t)
    As step(t) goes higher, the alpha becomes lower, and so does cumulative product.
    The skeletal hence contributes less to the overall loss.

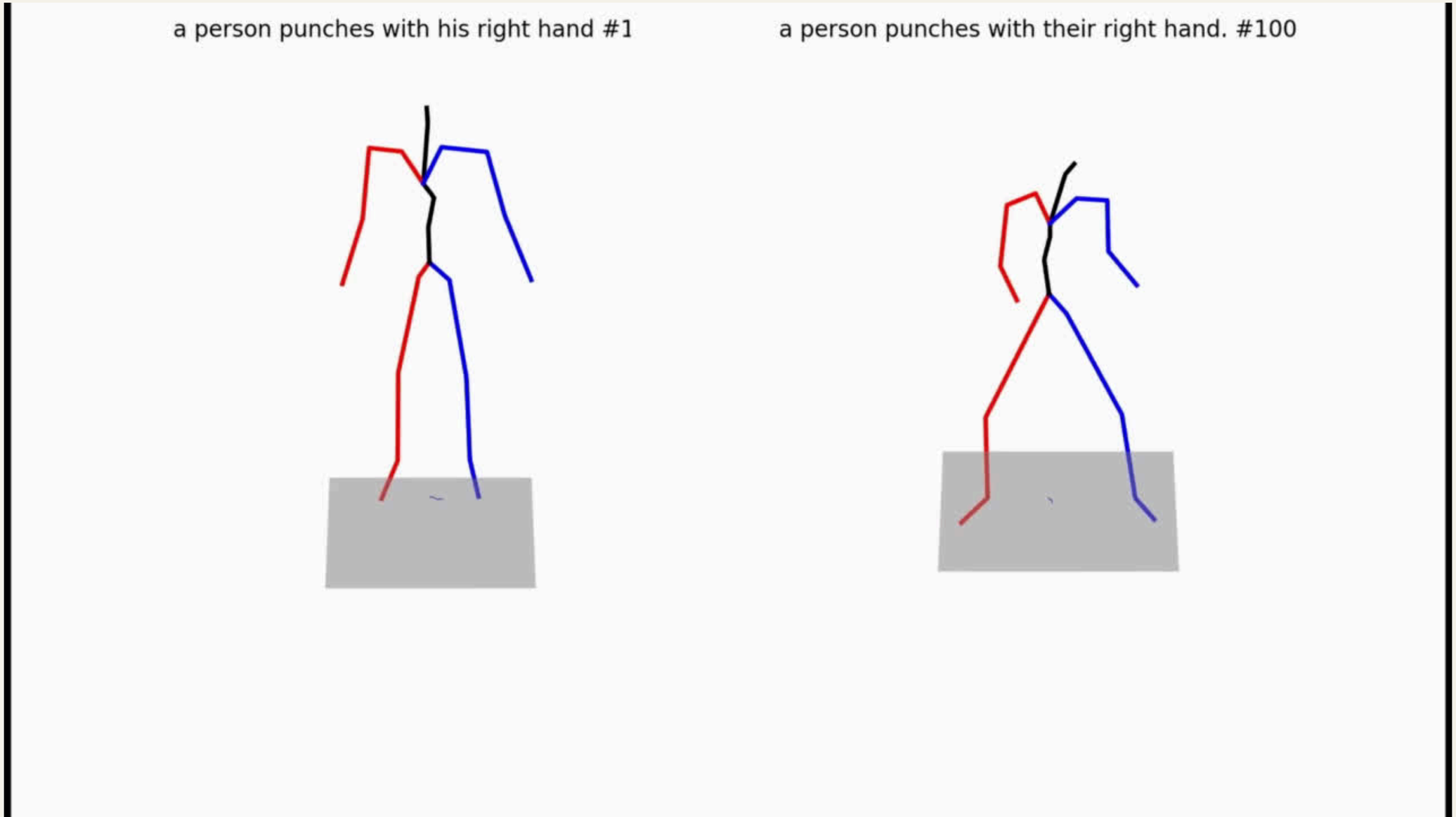
Implementation details

- Implement 3 additional losses proposed in Mofusion
  - Keep the bone lengths steady through the motion by minimizing the temporal variance of bone lengths.
  - Avoid deformation by minimizing asymmetry between paired bones
  - Ground truth supervision on not only the noise, but also on motion.

Discussion

- Original method crashed on the 35th epoch on our machine
- The proposed time-varying weight schedule works well
- The 3 losses work well but some deformation can still appear

Demo video


a person punches with his right hand #1
a person punches with their right hand. #100


THANKS