Supplementary Material

Network architecture

Figs. 6 to 9 show the network architecture of pose refiner, skeleton motion, residual non-rigid motion, and NeRF networks.

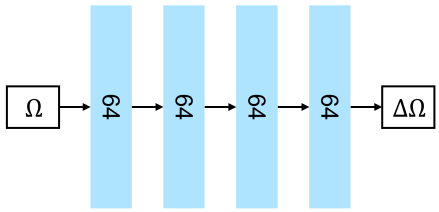

Figure 6. Pose refiner network. Given a body pose $\theta = (J, \Omega)$ in an image, this network takes in joint angles Ω and outputs joint angle relatives $\Delta\Omega$, which is used to obtain an updated body pose θ^o.

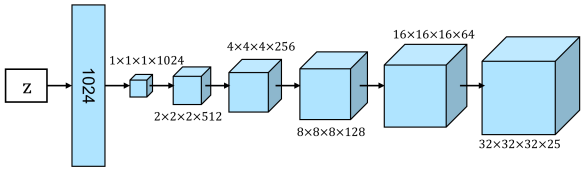

Figure 7. Skeleton motion network. This network generates a weight volume W^c to explicitly represent the skeleton motion field. This network is composed of a fully-connected layer, a tensor reshaping operator, and five 3D transposed convolutions in sequential. It takes in a random constant latent variable $\mathbf{z}$ of a size 256 and outputs a volume of size $32 \times 32 \times 32 \times 25$. The generated weight volume W^c is then used to derive the blend weights w_i^o.

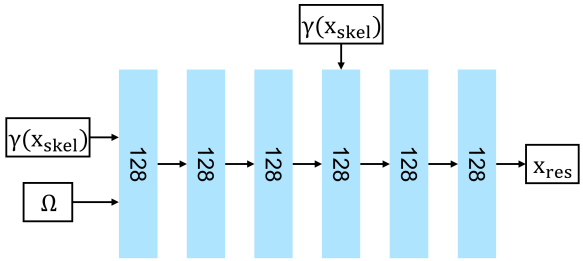

Figure 8. Residual non-rigid motion network. This network is conditioned on the body pose θ^o and the skeleton motion field M_{skel}. Specifically, it takes in the updated joint angles Ω^o ($\Omega^o = \Delta_\Omega \otimes \Omega$), and the positional encoding of the points in skeleton motion field $\gamma(\mathbf{x}_{\text{skel}})$. At the fourth layer of the network, we use a skip connection for $\gamma(\mathbf{x}_{\text{skel}})$. This network produces a residual motion field as an offset $\mathbf{x}_{\text{res}}$ to $\mathbf{x}_{\text{skel}}$. The addition of $\mathbf{x}_{\text{skel}}$ and $\mathbf{x}_{\text{res}}$ completes the full motion field.

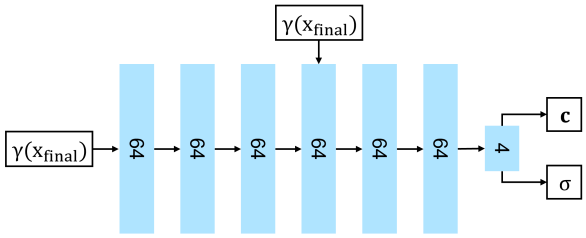

Figure 9. NeRF network for obtaining $\mathbf{c}$ and σ. This network takes in the positional encoding of the points in full motion field $\gamma(\mathbf{x}_{\textbf{final}})$, where $\mathbf{x}_{\textbf{final}} = \mathbf{x}_{\text{skel}} + \mathbf{x}_{\text{res}}$. At the fourth layer of the network, we use a skip connection for $\gamma(\mathbf{x}_{\text{skel}})$. This network outputs color $\mathbf{c}$ and density σ for volume rendering.

Enlarged rendering images

Figs. 10 and 11 are enlarged versions of Figs. 4 and 5, which showcase a visual quality comparison between our method and HumanNeRF for rendering a human from 4 different viewpoints in the same time frame, and from the same viewpoint at 4 different time frames, respectively.

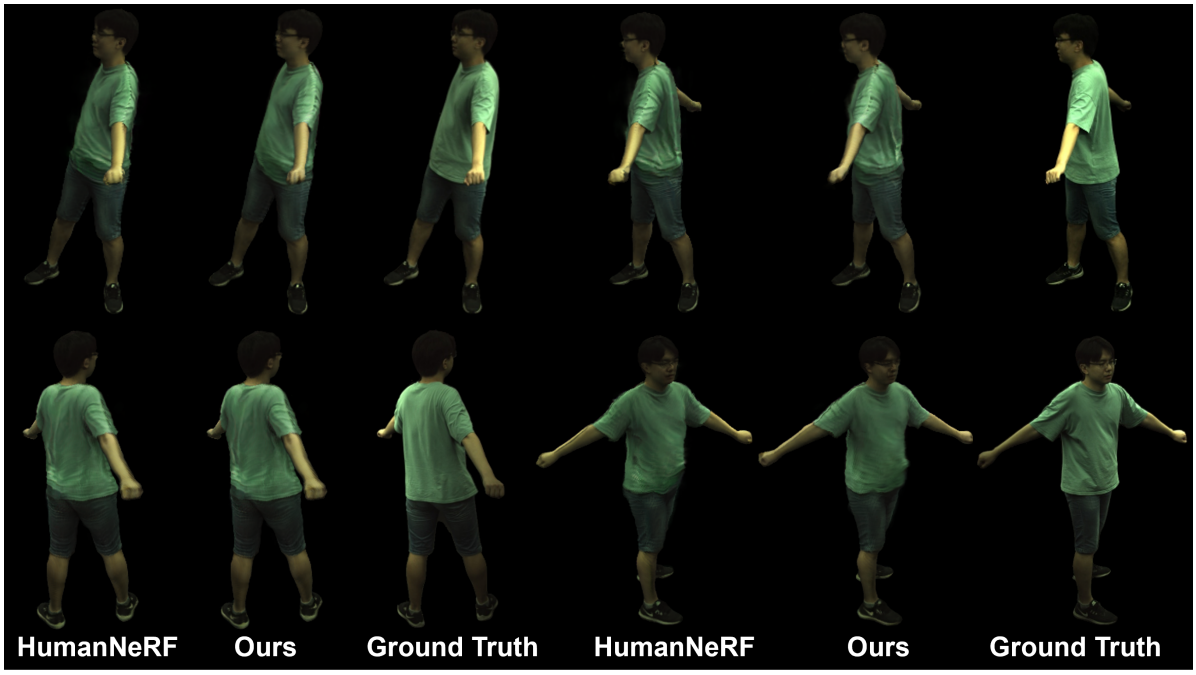

Figure 10. Enlarged Fig. 4 for detailed inference comparison: rendering a human from 4 different viewpoints in the same time frame.

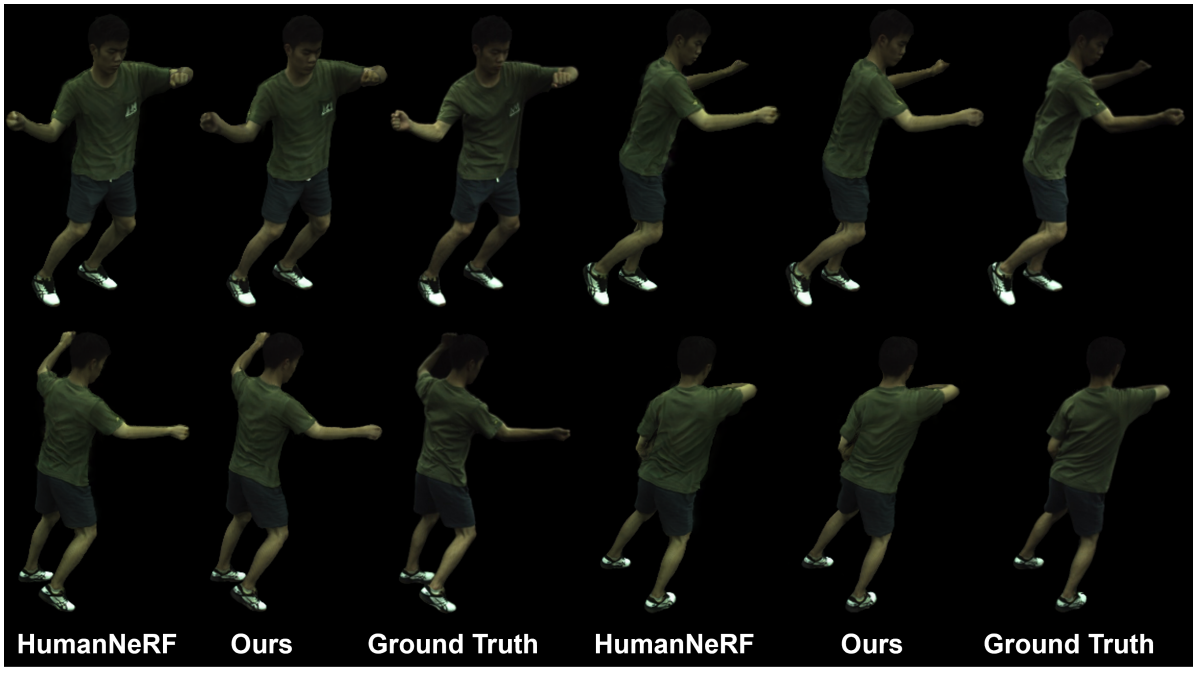

Figure 11. Enlarged Fig. 5 for detailed inference comparison: rendering a human from the same viewpoint at 4 different time frames.