

Supplementary Material

Visualization of attentions for 3D human pose and shape estimation. To further verify whether the NSSM can recalibrate the attention map generated by classic self-attention module [1] to improve the 3D human pose and shape estimation, we conduct the following qualitative experiments. Specifically, we visualize the results of ADAM-Net-only L-ADAM and ADAM-Net-only MHA (both without including the S-ADAM attention module) under the setting of 3D human pose and shape estimation. Both methods use one attention head for ease of visualizing the attention map. The visualization in Figure 1 confirms that by integrating NSSM into the L-ADAM attention module, our ADAM-Net can indeed focus

attention on a more appropriate range of action sequence, thereby improving 3D human pose and shape estimation. On the contrary, the classic self-attention module [1] generates an unreliable attention map, focusing attention on less correlated frames and ignoring the continuity of human motion in the action sequence, which reduces its accuracy.

REFERENCES

- [1] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

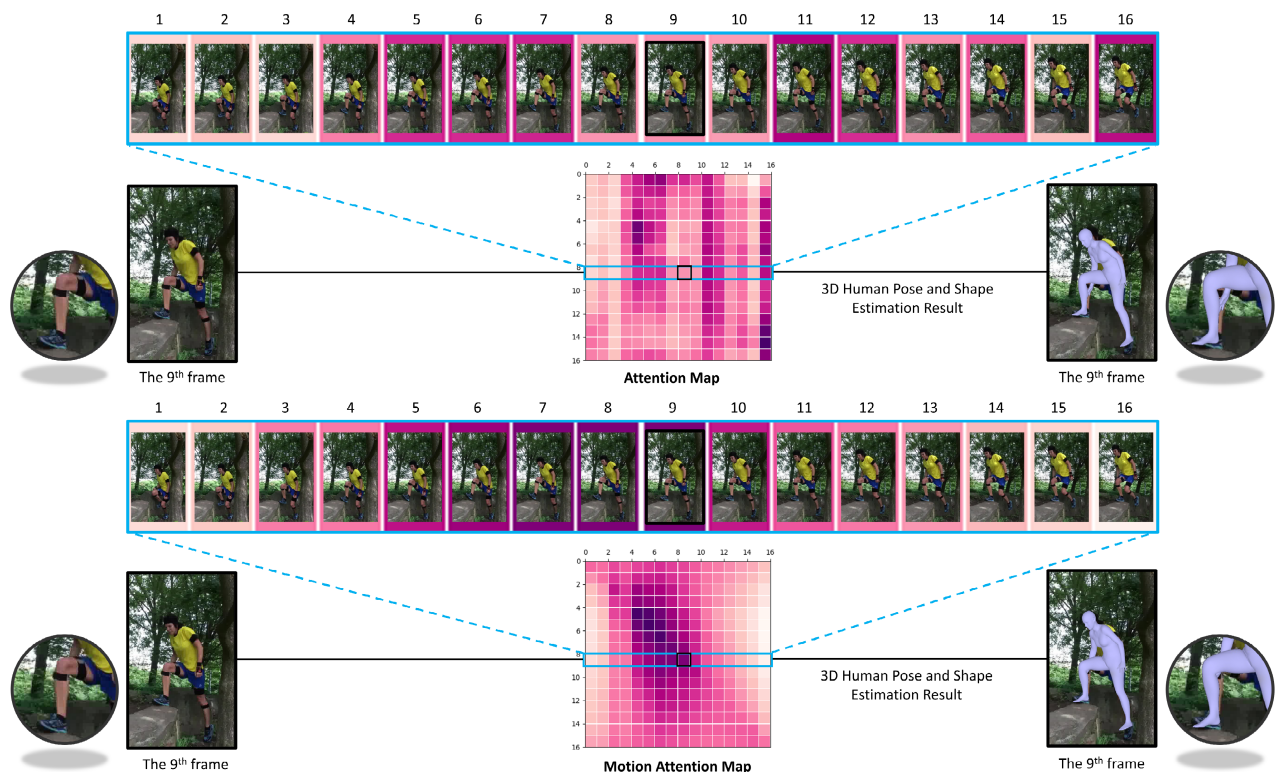


Fig. 1. Visual comparison of 3D human pose and shape estimation between MHA [1] and L-ADAM. Where the attention map and motion attention map are generated from the MHA and L-ADAM, respectively. In the attention and motion attention maps, the darker the purple, the higher the attention value. The results demonstrate that by integrating NSSM into the L-ADAM attention module, the motion attention map generated by our ADAM-Net can indeed focus attention on a more appropriate range of action sequence, thereby improving 3D human pose and shape estimation.