Attention Mechanism Exploits Temporal Contexts: Real-time 3D Human Pose Reconstruction

Ruixu Liu, Ju Shen, He Wang
Chen Chen, Sen-ching Cheung, Vijayan Asari

Code is available at: (https://github.com/lrxjason/Attention3DHumanPose)
Attentional Mechanism

 Temporal Attention Scores: \( \{ w_0^{(0)}, w_1^{(0)}, \ldots, w_{26}^{(0)} \} \)

Top Layer
Target frame \( P_t \)

Kernel size: \( k_0 = 3 \)

Layer 0
\( \theta_1^{(1)} \)
Input

\( \{ P_0, P_1, \ldots, P_{n-1} \} \)

Layer 1
\( \theta_2^{(2)} \)

\( \{ w_0^{(1)}, w_1^{(1)}, \ldots, w_8^{(1)} \} \)

Layer 2
\( \theta_3^{(3)} \)

\( \{ w_0^{(2)}, w_1^{(2)}, w_2^{(2)} \} \)

Layer 3

Kernel size: \( k_1 = 3 \)

Kernel size: \( k_2 = 3 \)

Kernel size: \( k_3 = 1 \)

Bottom Layer

Kernel Attention Module
GAP
Softmax

TCN Unit
Linear Projection Unit
Element-Wise Multiplication
Element-Wise Summation
Attention

Temporal Attention
(weights on tensors)

Kernel Attention
(weights on channels)
Attention Layer 0

\[ W^{(1)} \]

Attention Layer 1

\[ W^{(2)} \]

Attention Layer 2

\[ W^{(3)} \]

Attention Layer 3
\( W^{(0)} \leftarrow NCC (f_0, f_1, \ldots, f_{n-1}) \)

**Attention Layer 0**

\[ \theta^{(1)}_t \]

\[ W^{(1)} \]

\[ \theta^{(2)}_t \]

\[ W^{(2)} \]

\[ \theta^{(3)}_t \]

\[ W^{(3)} \]
The Multi-scale Dilated Convolution Structure
To increase receptive field

More layers
Attention Layer

\[ x \otimes y \otimes z \]

Level 0
\[ \theta^{(1)}_t \]

Level 1
\[ \theta^{(2)}_t \]

Level 2

Level 3

…

Output Tensors from each layer
Quantitative Evaluation:

1. Side-by-side comparison with state-of-the-art
Quantitative Evaluation:

Motion Retargeting Views
Quantitative Evaluation:

2. Joint-wise MPJPE: comparison with state-of-the-art
S11 WalkDog

Input

Ours

Pavllo et al.

Hossain et al.

Martinez et al.

Graph showing MPJE comparison with Martinez et al., Hossain et al., Pavllo et al., and our method.
Quantitative Evaluation:

3. Frame-wise MPJPE: comparison with state-of-the-art
S9 Smoking

Ours

Pavllo et al.

Hossain et al.

Martinez et al.

MPJPE

Martinez et al.

Hossain et al.

Pavllo et al.

Ours
Qualitative Evaluation:

4. Results on wild videos
Qualitative Evaluation:

5. Real-time performance using the causal model
Thank you for watching

Code is available at: (https://github.com/lrxjason/Attention3DHumanPose)