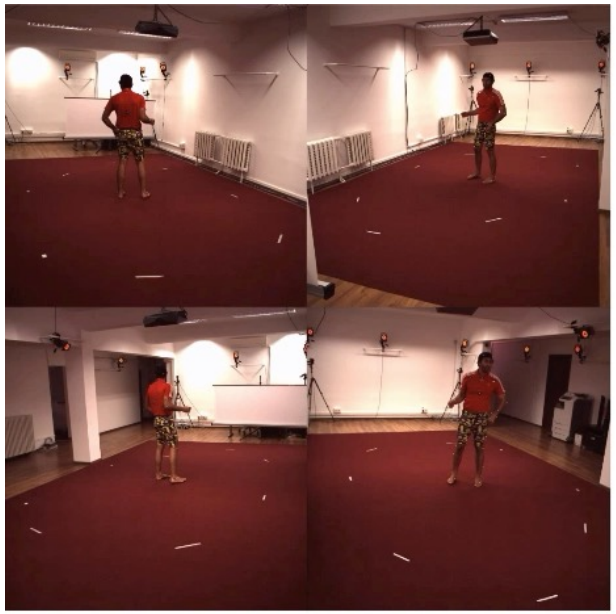
Pose Modulated Avatars from Video

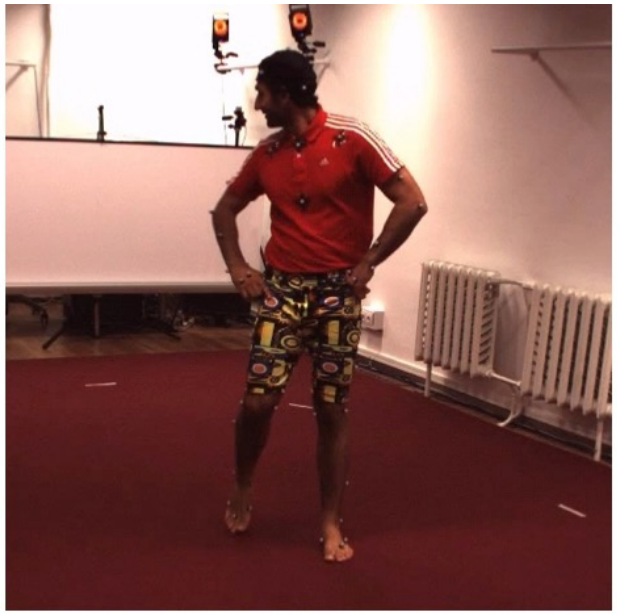
Chunjin Song[1] Bastian Wandt[3] Helge Rhodin[1,2]

[1]University of British Columbia [2]Bielefeld University [3]Linköping University

Training

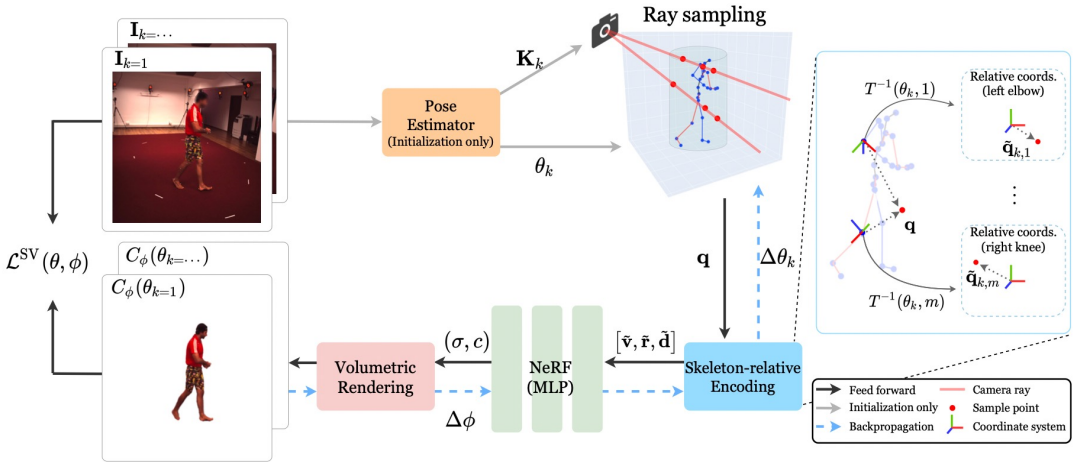


Novel view & novel pose synthesis

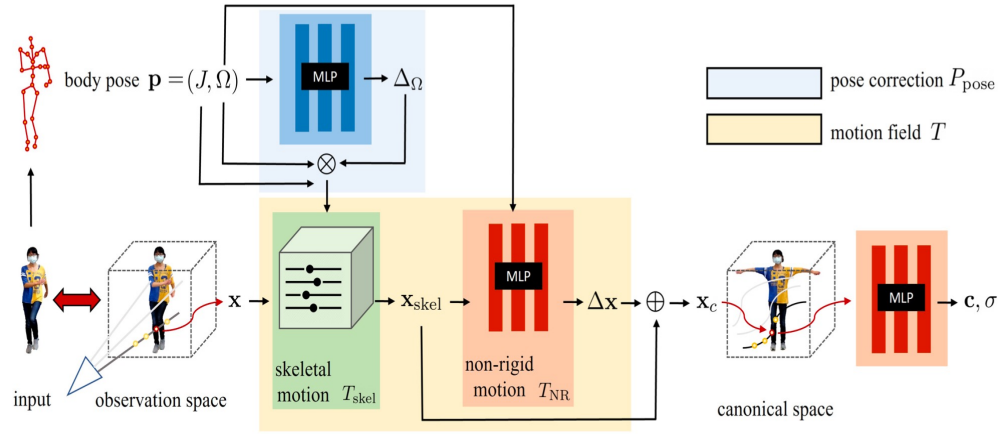


3D Geometry

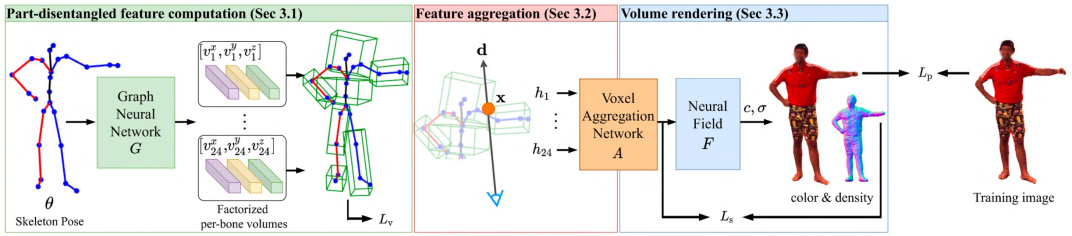
From https://zju3dv.github.io/neuralbody/



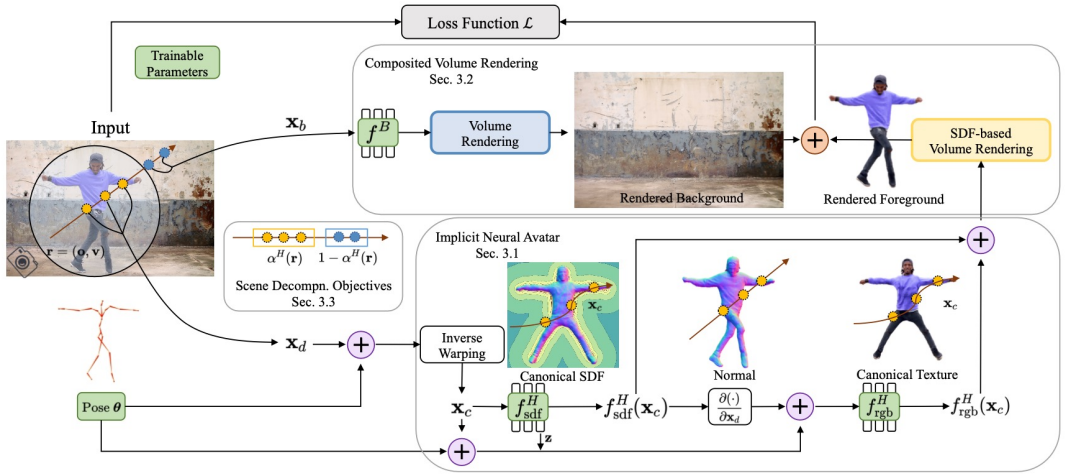
[A-NeRF, NeurIPS 2021]



[HumanNeRF, CVPR 2022]



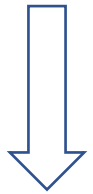
[DANBO, ECCV 2022]



[Vid2Avatar, CVPR 2023]

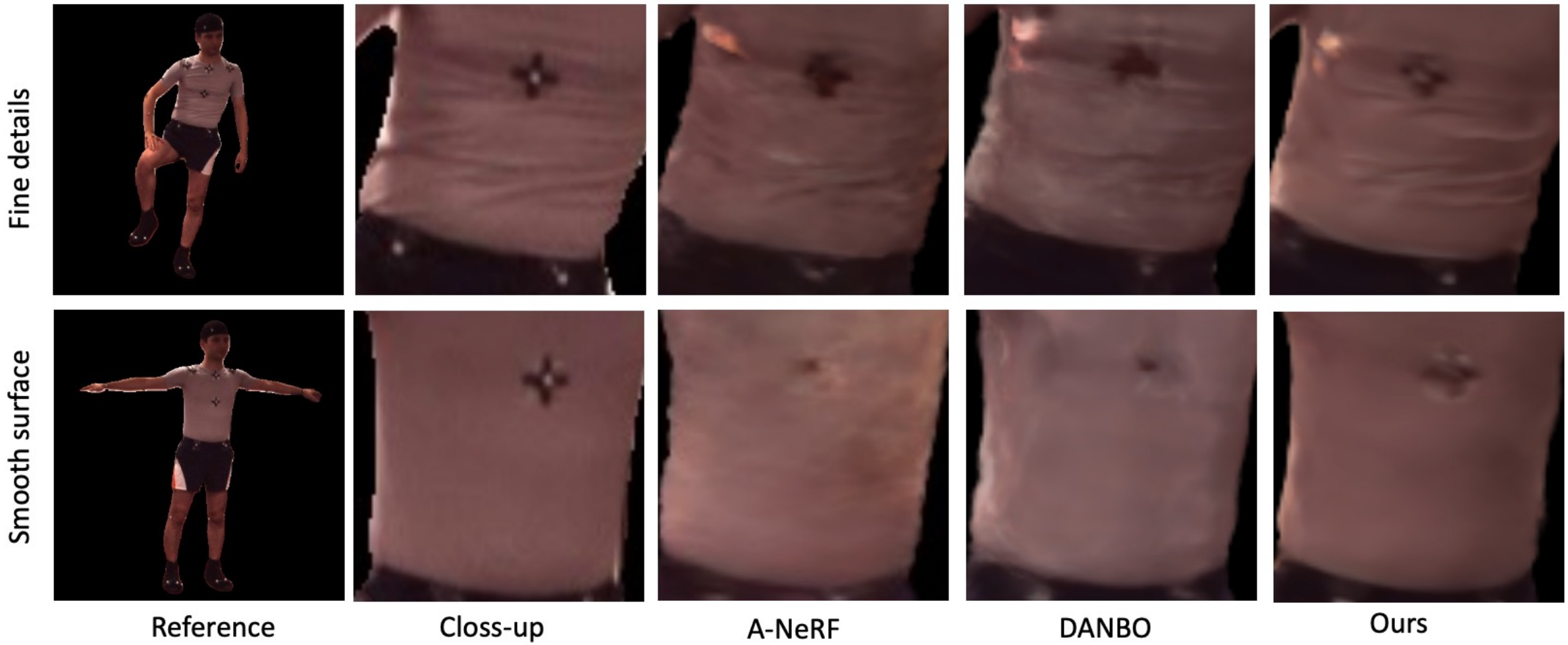
Observation

Previous models use fixed-frequency positional encoding, neglecting distinct frequency assignments for geometry and textures across different poses.



Noisy artifacts in smooth areas & blurred details in sharp regions

Observation



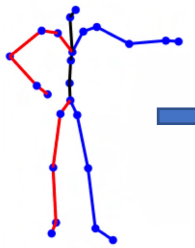
Motivation

The frequency of geometry and appearance details depend on the pose context information

Motivation

The frequency of geometry and appearance details depend on the pose context information

Skeleton pose



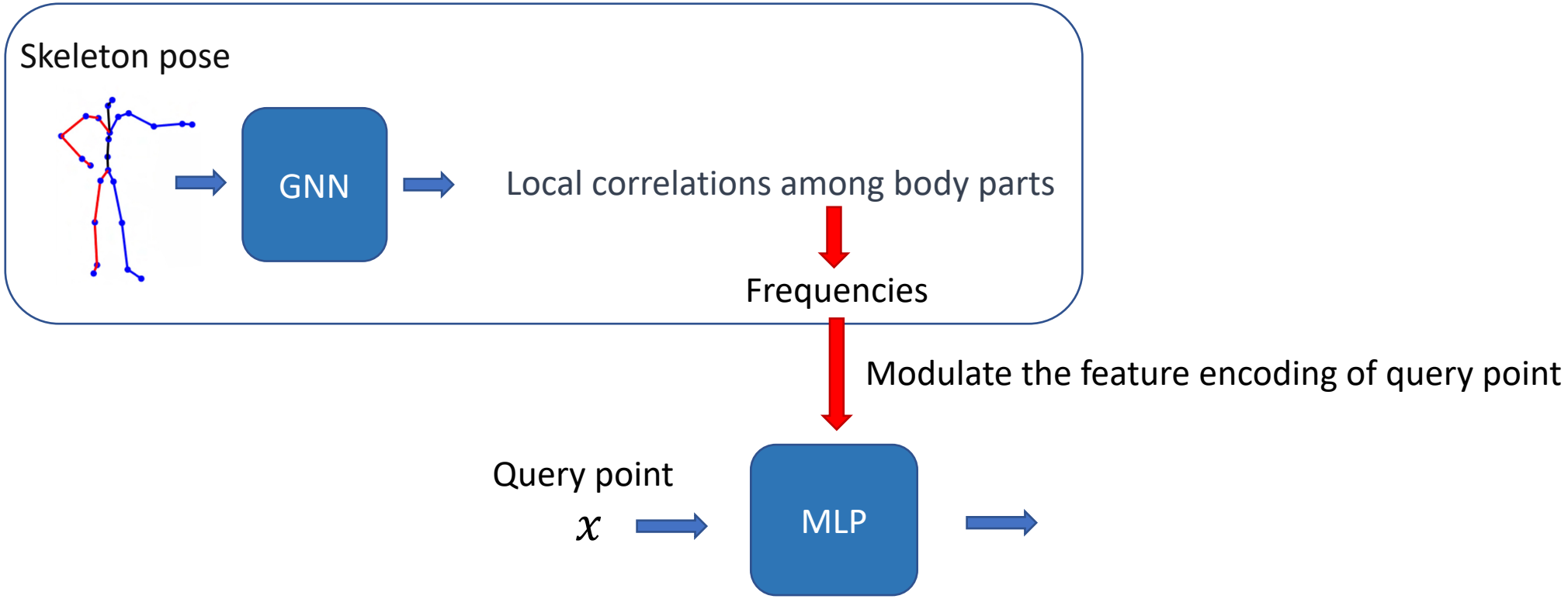
GNN

Local correlations among body parts

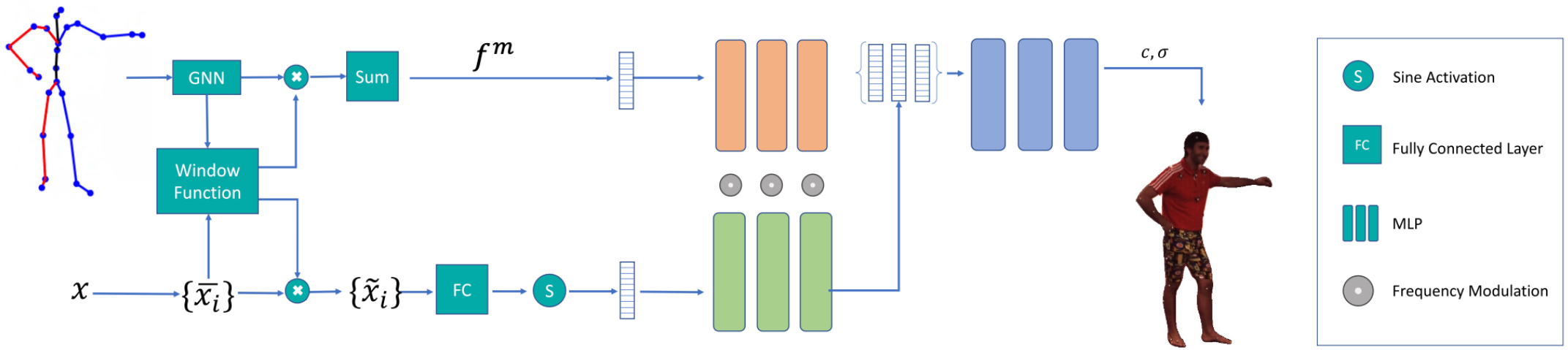
We take a graph neural network to model correlations among body parts locally

Motivation

The frequency of geometry and appearance details depend on the pose context information

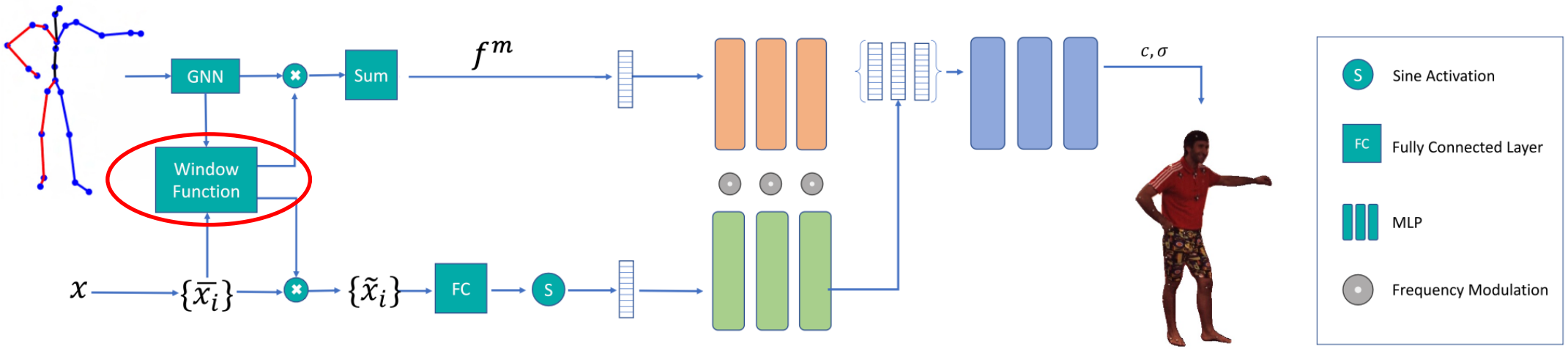


Framework



$$f_i = \sin(wf_{i-1} + b)$$

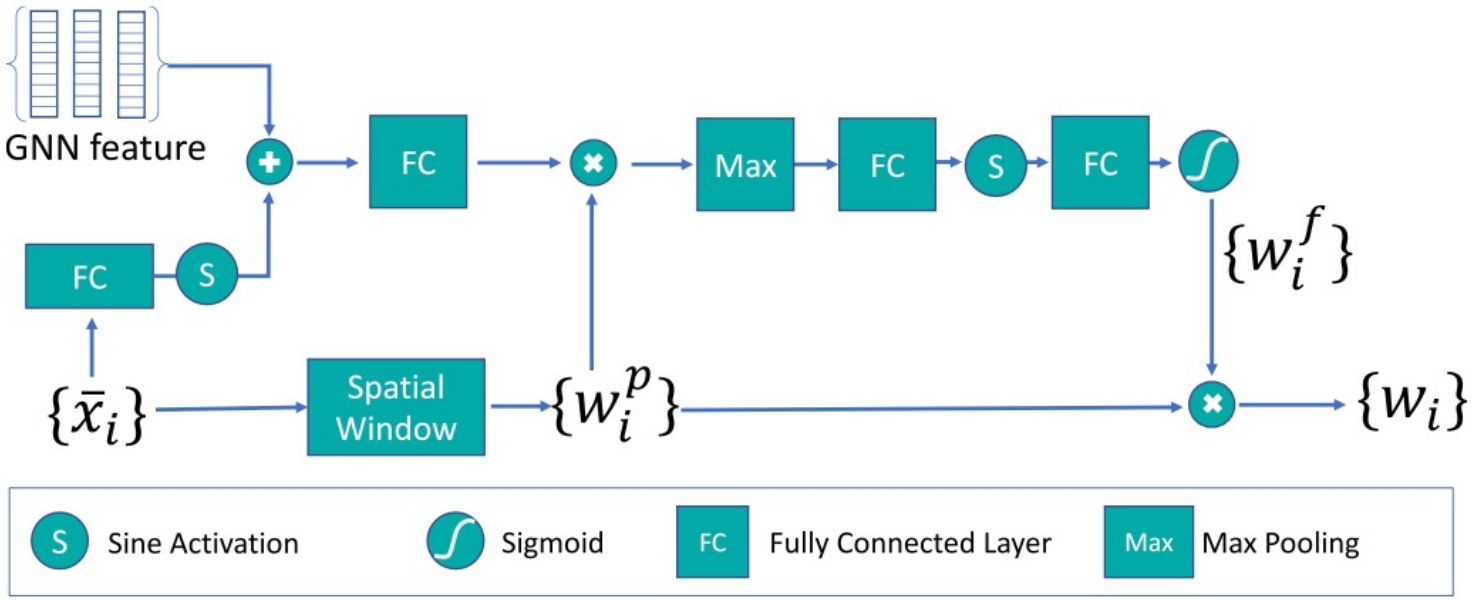
Framework



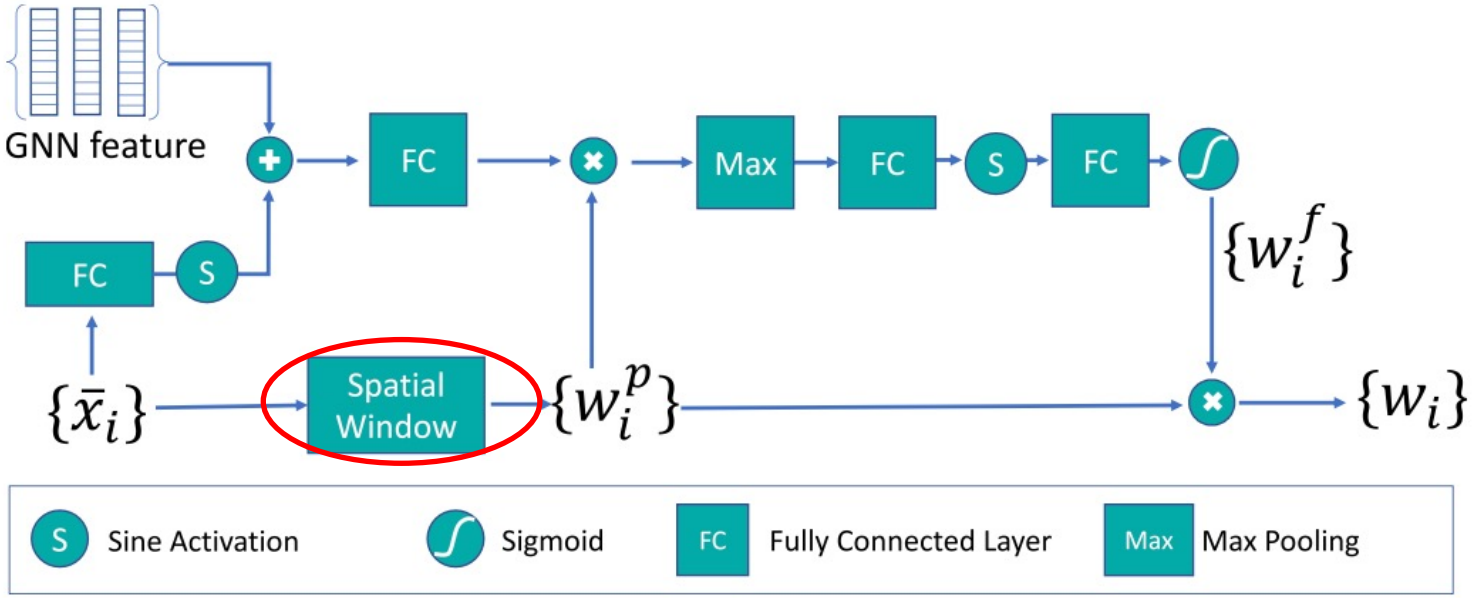
$$f_i = \sin(wf_{i-1} + b)$$

Select relevant part features for input query point

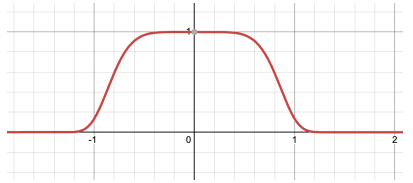
Two-stage Window Function



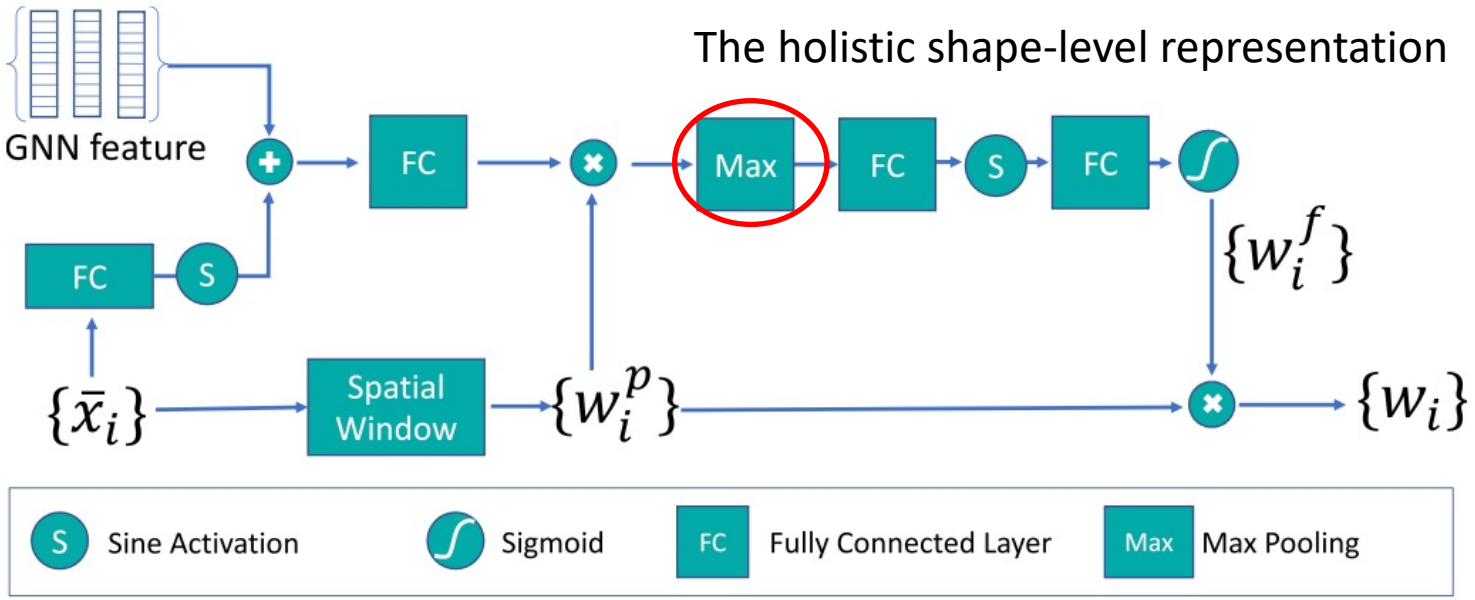
Two-stage Window Function



$$w_i^p = \exp(-\alpha(\|\bar{x}_i\|_2^\beta))$$



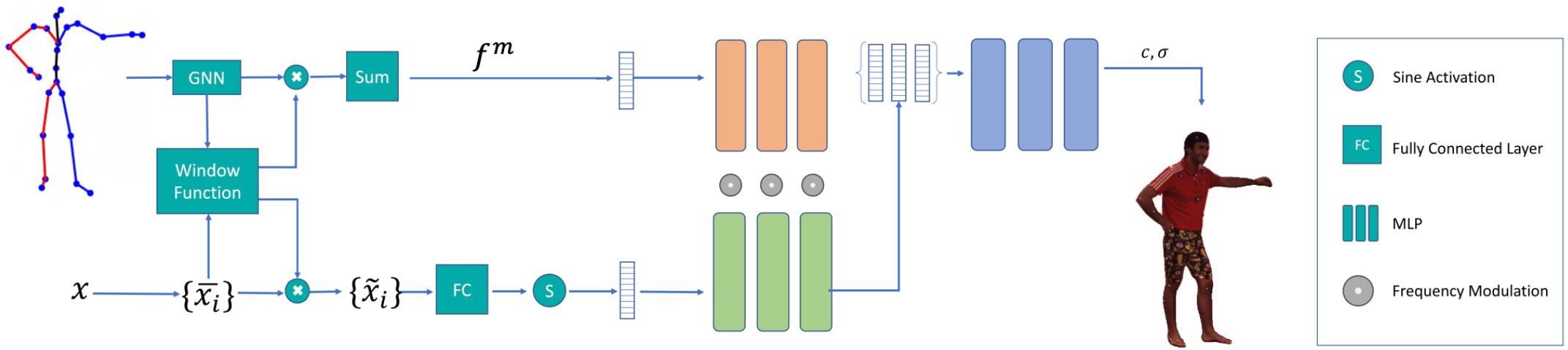
Two-stage Window Function



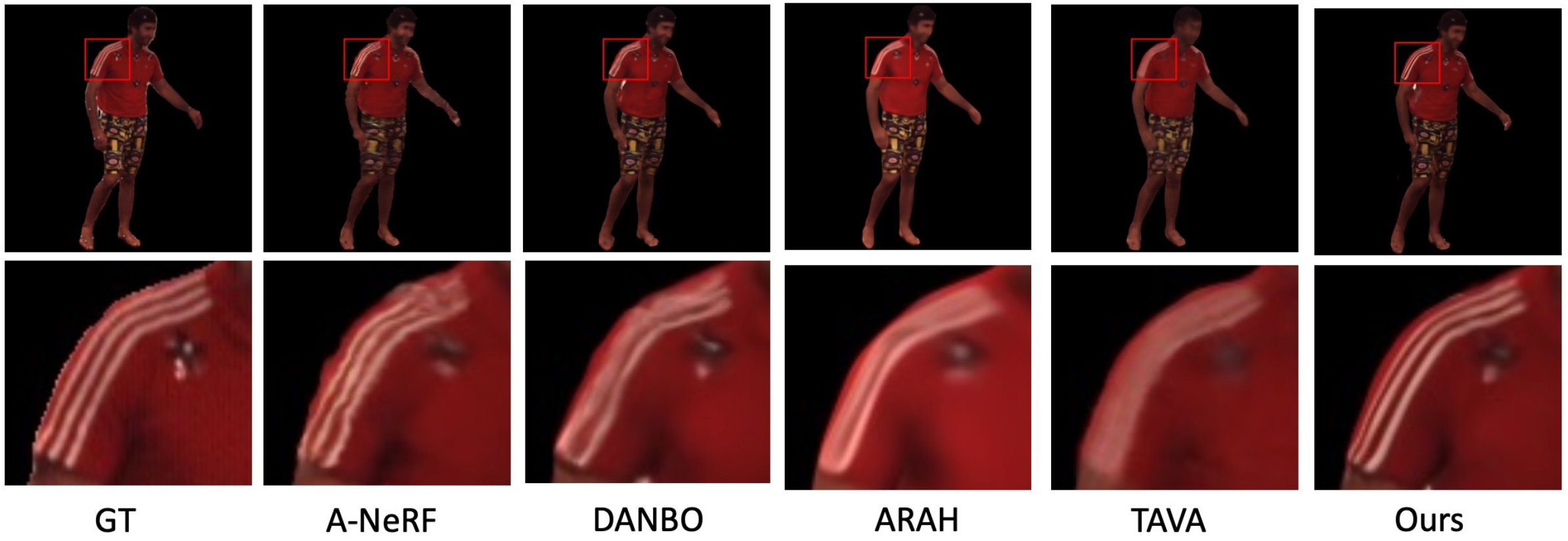
Feature prioritization problem when multiple parts' features overlap

Take the pose context into account

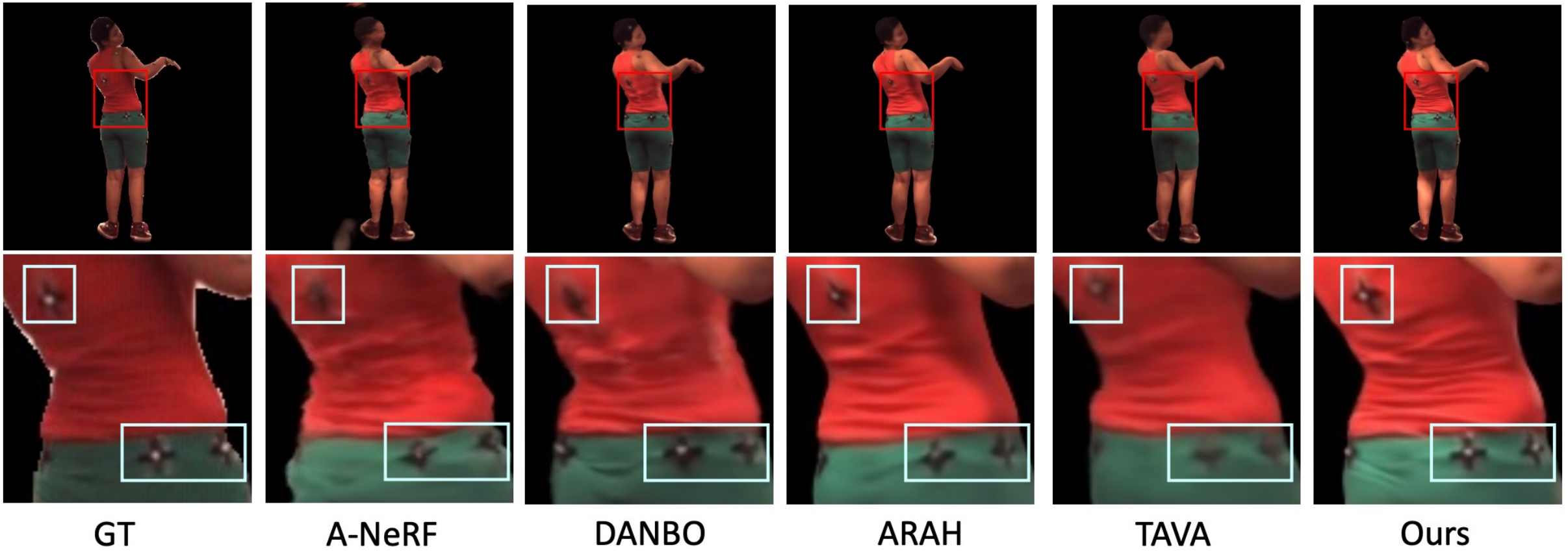
Framework


GNN
Sum
f^m
Window Function
x
$\{\bar{x}_i\}$
$\{\tilde{x}_i\}$
FC
S
c, σ
Sine Activation
Fully Connected Layer
MLP
Frequency Modulation

Novel View Comparison



Novel Pose Comparison



Geometry Comparison

Pseudo GT

Humannerf

Monohuman

Vid2Avatar

Vid2Avatar_mask

Ours

Out of Distribution Pose Rendering

Reference

Our method demonstrates robust time consistency even when subjected to extreme novel pose extracted from the sequence in another dataset.

Contribution

1. Introducing a novel two-branch neural network for high-fidelity human video representation via frequency modulation.

2. Utilizing a simple part feature aggregation function for high-frequency detail synthesis and artifact reduction near overlapping joints.

3. Comprehensive evaluation and ablation studies highlight the significance of window functions and frequency modulations, showcasing state-of-the-art results.

Pose Modulated Avatars from Video

Chunjin Song[1] Bastian Wandt[3] Helge Rhodin[1,2]

[1]University of British Columbia [2]Bielefeld University [3]Linköping University