This supplementary material provides extended qualitative analyses to complement the main paper. Our additional experiments are organized into two categories:

- **Robustness to Noisy 2D Inputs:** We evaluate SasMamba on mildly challenging wild videos, where erroneous 2D pose detections are present. We further test the model under fixed background and virtual character/scene conditions to analyze its stability.
- **Performance on Challenging Poses:** We investigate SasMamba's ability to handle extreme human poses. Specifically, we provide qualitative results on aerial rotations and analyze its robustness under different viewpoints, including top and side views.

1. Robustness to Noisy 2D Inputs

To evaluate the robustness and generalizability of SasMamba, we conducted experiments on several moderately challenging, previously unseen real-world videos (Figure 1). In these videos, human actions evolve smoothly and scene changes occur gradually. Our results show that even when the input 2D poses exhibit noticeable noise (red arrows), SasMamba consistently produces accurate 3D pose estimates (green arrows). This demonstrates the strong resilience of the model in handling abrupt or noisy motions within 2D input sequences.

Furthermore, Figure 2 illustrates that when backgrounds are fixed or blurred, improvements in the quality of input 2D sequences lead to more stable 3D predictions. In the case of virtual characters performing high-speed motions in synthetic environments, SasMamba remains robust: despite degraded 2D pose quality, the model leverages its global structure-awareness to correct local pose estimation errors effectively.

2. Performance on Challenging Poses

To further explore the robustness limits of SasMamba, we evaluated the model on more extreme motion samples, including diving, gymnastics, and skiing. Unlike the smoother motions in earlier experiments, these actions involve rapid velocity changes and frequent high-difficulty aerial rotations. As shown in Figure 3, such irregular short-term motion patterns not only pose challenges to 2D detectors—resulting in severely degraded input quality—but also exceed the correction capacity of our model. These findings highlight that, in the 3D pose estimation pipeline, the reliability of the 2D detection stage is as critical as the lifting stage.

Additionally, in Figure 4, we evaluated extreme motions of the same subject from two side views. The results indicate that a single, consistent background and lateral viewpoints yield more reliable predictions, whereas top-down views introduce additional challenges due to occlusions and entanglement among joint positions. This further reinforces the observation that, in 3D pose estimation, the performance of the 2D pose detection stage is equally critical as that of the lifting stage.

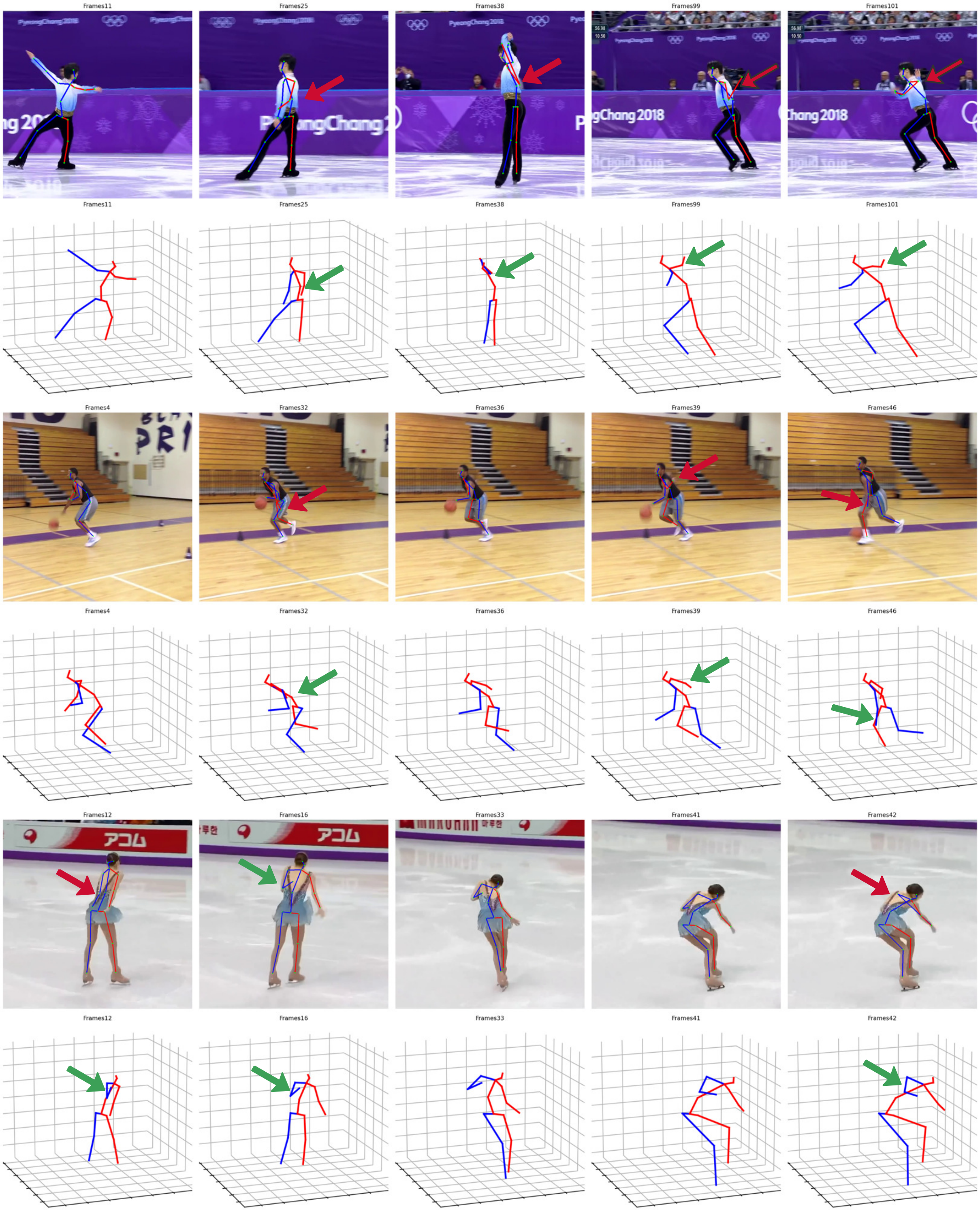

Figure 1. Qualitative Results on Mildly Challenging Wild Videos. Red arrows highlight erroneous 2D pose estimations, while green arrows indicate correct 3D predictions.

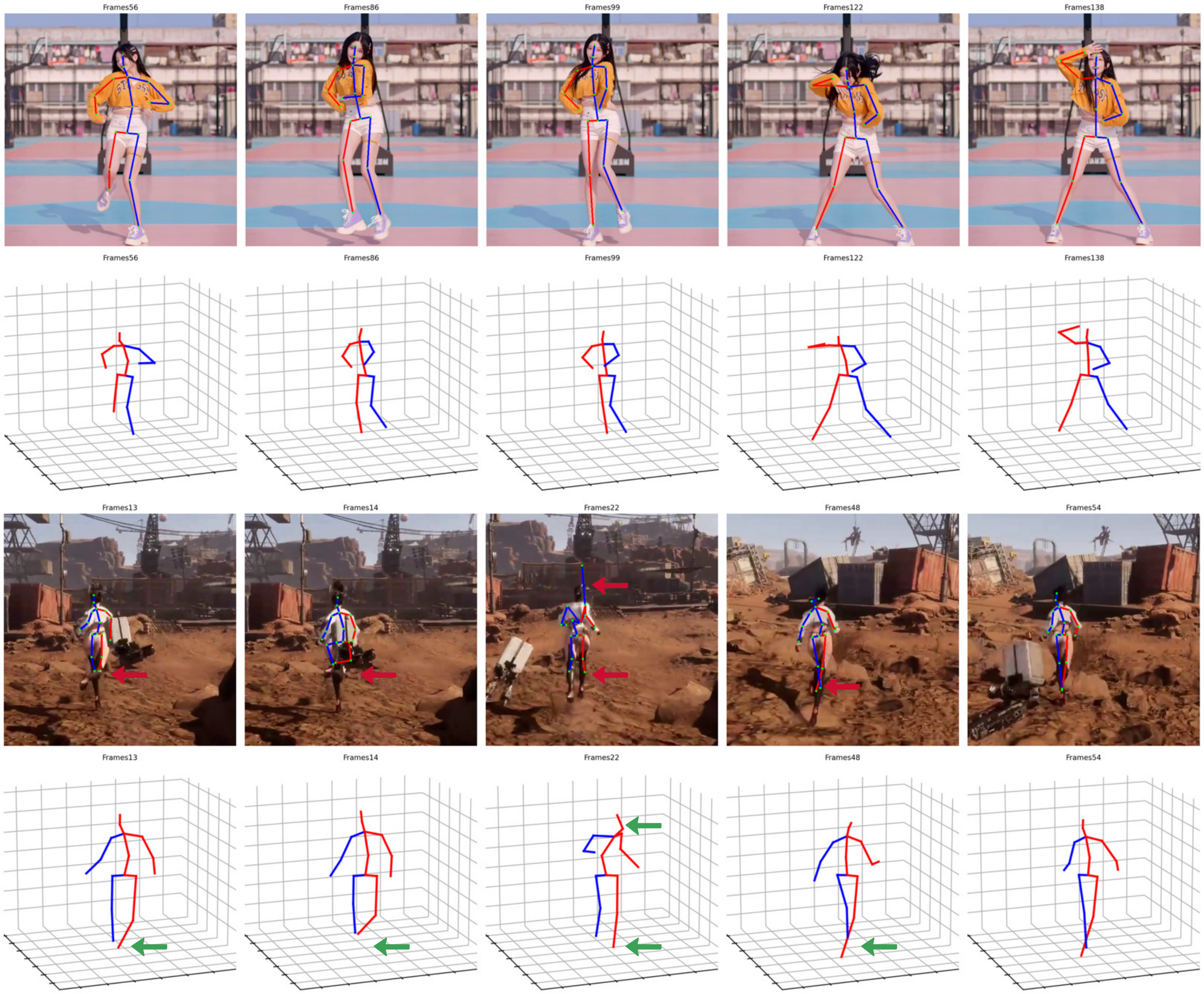

Figure 2. Qualitative Results of SasMamba under Fixed Background and Virtual Character/Scene Conditions.

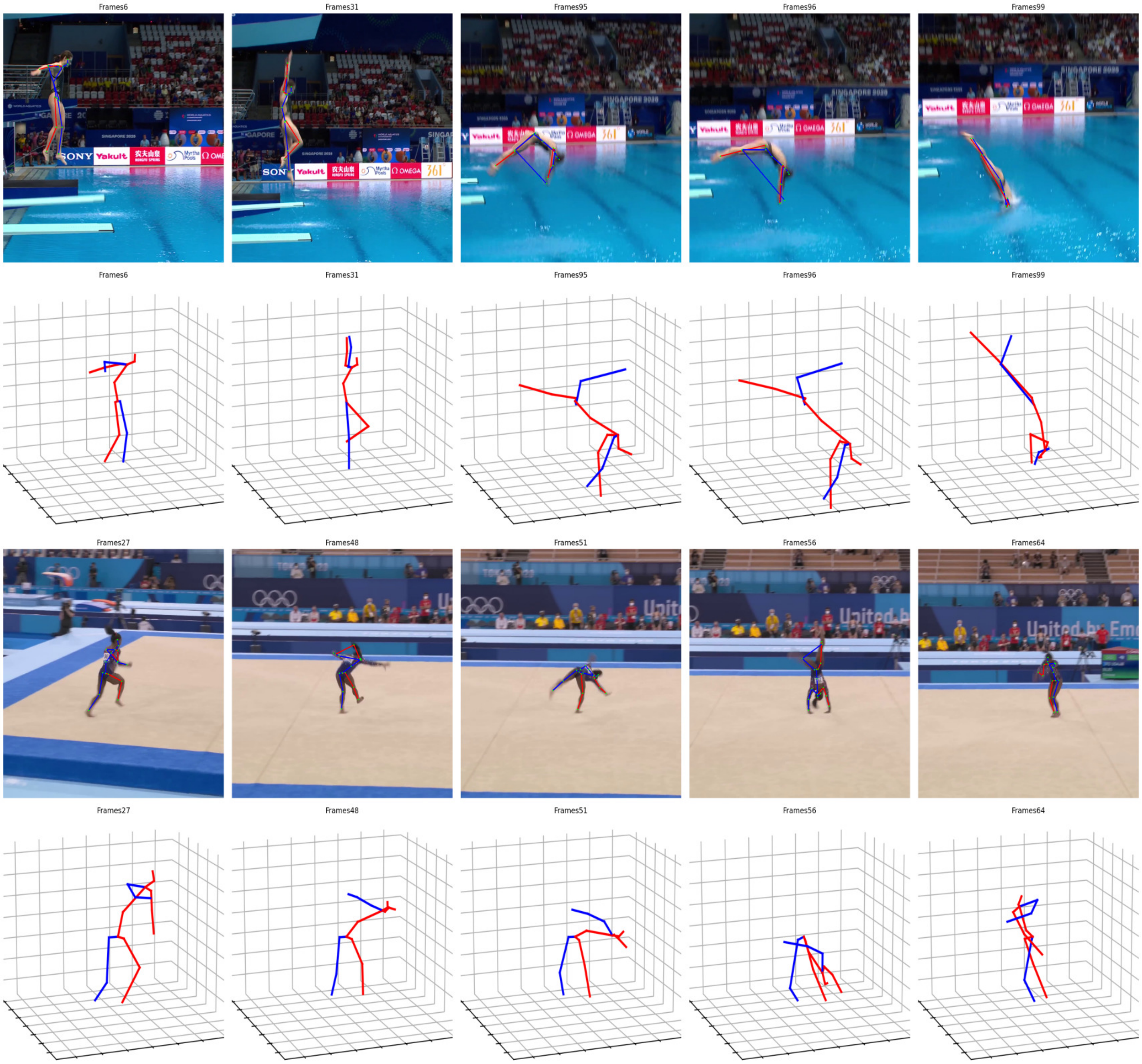

Figure 3. Qualitative Results of SasMamba on Extreme Poses (Aerial Rotation).

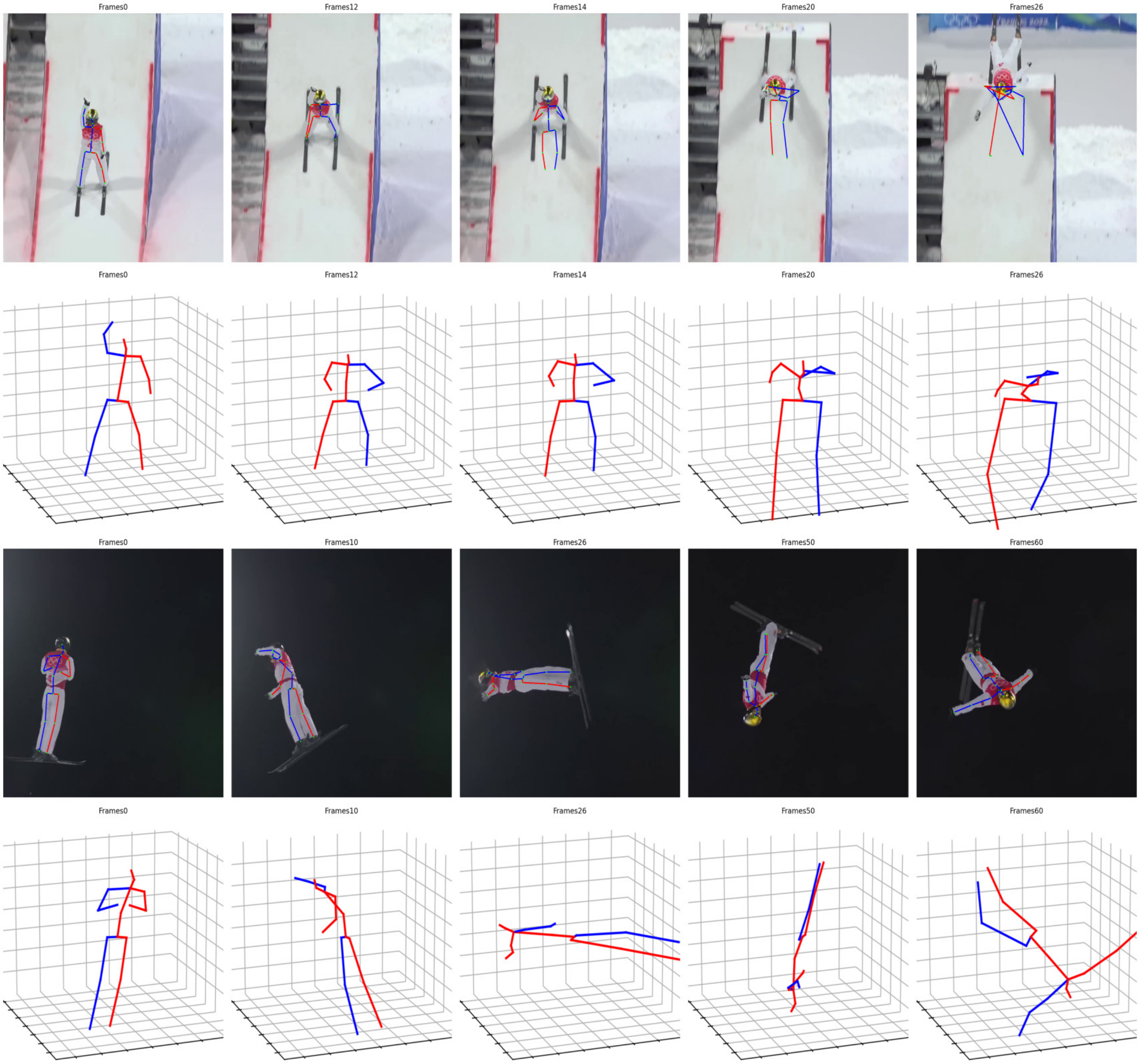

Figure 4. Qualitative Results of SasMamba on Aerial Rotations from Top and Side Views.