FinePhys: Fine-grained Human Action Generation by Explicitly Incorporating Physical Laws for Effective Skeletal Guidance

Supplementary Material

Contents

A. Training & Dataset Details

A.1. Overview

We deploy FinePhys using PyTorch, and the training process consists of four steps: ❶ Pre-training the skeletal heatmap encoder on the HumanArt [31] dataset; ❷ Pre-training the 2D-to-3D module and the PhysNet module on Human3.6M [30] and AMASS [45] datasets; ❸ Fine-tune the 2D projection module and PhysNet module using the online detected 2D skeletons detected from FineGym [60]; ❹ Jointly fine-tuning the U-Net [56], PhysNet, and 2D projection modules on FineGym. The first three steps of training are conducted on a Linux (Ubuntu) machine with 4 Nvidia 4090 GPUs within 48 hours, while step 4 utilizes two NVIDIA L20 GPUs and completes within 12 hours.

Across all experiments, we apply a linear noise scheduler with 1,000 timesteps, linearly increasing the beta values from 0.00085 to 0.012 to progressively reduce noise during training. The U-Net backbone incorporates a motion module featuring temporal self-attention layers and positional encoding operating at resolutions $[1, 2, 4, 8]$, enabling multi-scale temporal dynamics capture. The motion module is configured with eight attention heads, a single transformer block, and dual temporal self-attention layers to effectively model temporal dependencies. To stabilize training, the module parameters are zero-initialized. We incorporate a Low-Rank Adaptation (LoRA) [27] module with a rank of 64 and a dropout rate of 0.1, facilitating efficient adaptation of the model's spatial and temporal layers while minimizing the number of trainable parameters. Training utilizes the Adam optimizer with an initial learning rate of 5×10^{-4} and a weight decay of 1×10^{-2}. Additionally, gradient checkpointing is enabled to optimize GPU memory usage during training.

A.2. HumanArt Pre-training

Initially, we train the skeletal heatmap encoder on the HumanArt dataset, a large-scale image collection containing 50K images with accurate pose and text annotations across various scenarios. We leverage the *real-human* subset, comprising 8,750 images with corresponding 2D skeleton annotations. The original COCO-format skeletons are converted to the Human3.6M format, both with 17 keypoints, and subsequently processed into limb heatmaps following the PoseConv3D approach [15]. We employ Stable Diffusion v1.5 [55] as the spatial generator and keep it frozen during training.

A.3. Human3.6M and AMASS Pre-training

To pre-train the 2D-to-3D module and PhysNet, we utilize diverse and realistic 3D human motion data from the Human3.6M and AMASS datasets. Both provide 3D pose annotations essential for skeleton modeling. We use 2D-3D skeleton pairs from Human3.6M as prompt pairs and pre-train both modules for 10 epochs.

A.4. FineGym Fine-tuning

For fine-tuning FinePhys, we use the FineGym [60] dataset, selecting three subsets with distinct motion dynamics: FX-JUMP, FX-TURN, and FX-SALTO. FX-JUMP includes 11 classes (IDs 6–16), FX-TURN comprises 7 classes (IDs 17–23), and FX-SALTO contains 17 classes (IDs 24–40), as detailed in Tab. 3. Example videos and poses are illustrated in Fig. 8.

We generate captions for each video by prompting GPT-4 [1] to transform existing textual descriptions into standardized prompts. The instruction provided to GPT-4 was: *"For each gymnastics move described in the labels below, write a detailed description as if explaining to someone who is unfamiliar with gymnastics."* For example, the label *"2 turns on one leg with free leg optional below horizontal"* is converted to *"A person executes two complete turns while balancing on one leg, allowing the lifted leg to remain below hip level or in any chosen position beneath the horizontal line throughout the turning sequence."* This augmentation

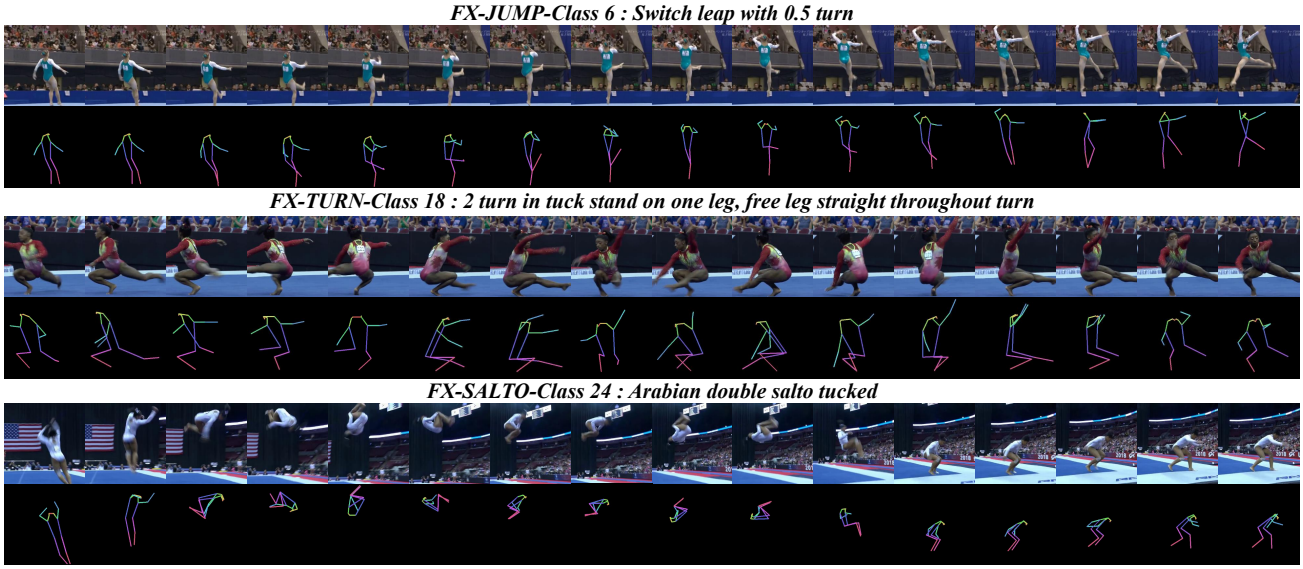

Figure 8. **Example videos from FX-JUMP, FX-TURN and FX-SALTO.** Each sample video has 16 frames, and the corresponding 2D skeleton sequence is also represented.

enhances the model's comprehension of textual prompts, facilitating subsequent video generation tasks.

With the dataset augmented by extended descriptions, we first fine-tune the PhysNet and 2D projection modules for 10,000 training steps using online-detected 2D skeletons from FineGym. Subsequently, we jointly fine-tune the U-Net, PhysNet, and 2D projection modules for an additional 8,000 training steps.

B. Elaboration on Evaluation Metrics

In this section, we elaborate on the details of evaluation metrics used in our project. First, we discuss the limitation of the original CLIP-SIM metric [51] and the corresponding improved CLIP-SIM*. Then we introduce the details of the user study as well as other metrics.

B.1. CLIP-SIM Metrics and Limitations

We analyze the CLIP-SIM metric based on three aspects: *semantic consistency*, *domain consistency*, and *temporal consistency* [22]. Below, we detail each aspect and discuss their limitations.

❶ ***Semantic Consistency*** measures the alignment between textual prompts and the generated video frames. Specifically, for a given text prompt P and a generated video $\tilde{V}$ with T frames, the semantic consistency score is computed as the average CLIP similarity between P and each frame of $\tilde{V}$:

$$\text{CLIP}_{\text{text}}(P, \tilde{V}) = \frac{1}{T}\sum_{t=1}^{T}\text{CLIP}(P, \tilde{V}(t)). \tag{27}$$

Limitations of $\text{CLIP}_{\text{text}}$: The original semantic consistency metric struggles with fine-grained action labels due to semantic ambiguity and entanglement in the CLIP embedding space. As illustrated in Fig. 9, while the metric performs adequately for coarse-grained action categories (*e.g.*, those from UCF101 [64]), it fails with FineGym labels where the embedded vectors of specific categories overlap significantly, rendering the metric ineffective for distinguishing between similar fine-grained actions.

❷ ***Domain Consistency*** assesses the similarity between generated video frames and reference images generated by an open-sourced image generation model, such as Stable Diffusion [55]. For a reference image I and a generated video $\tilde{V}$ with T frames, the domain consistency score is calculated as:

$$\text{CLIP}_{\text{domain}}(I, \tilde{V}) = \frac{1}{T}\sum_{t=1}^{T}\text{CLIP}(I, \tilde{V}(t)). \tag{28}$$

Limitations of $\text{CLIP}_{\text{domain}}$: The domain consistency metric is unreliable for fine-grained actions because reference images generated by Stable Diffusion may not accurately reflect the nuances of specific actions or their dynamics, as shown in Fig. 10. Additionally, comparing the generated results in Fig.18, higher domain scores do not necessarily correspond to better representations of fine-grained videos. For instance, T2V-Zero generates nonsensical content that still achieves a higher domain score than AnimateDiff, and VideoCrafter's highest-scoring results often contain visible artifacts and limb inaccuracies.

❸ ***Temporal Consistency*** evaluates the smoothness of transitions between frames in a generated video by computing the average CLIP similarity between randomly selected pairs of frames. Given a generated video $\tilde{V}$ and a set of N

Table 3. Categories of FX-JUMP, FX-TURN, and FX-SALTO from Gym99.

FX-JUMP from Gym99		
Class	ID	Category
6	0	Switch leap with 0.5 turn
7	1	Switch leap with 1 turn
8	2	Split leap with 1 turn
9	3	Split leap with 1.5 turn or more
10	4	Switch leap (leap forward with leg change to cross split)
11	5	Split jump with 1 turn
12	6	Split jump (leg separation 180 degree parallel to the floor)
13	7	Johnson with additional 0.5 turn
14	8	Straddle pike or side split jump with 1 turn
15	9	Switch leap to ring position
16	10	Stag jump
FX-TURN from Gym99		
Class	ID	Category
17	0	2 turn with free leg held upward in 180 split position throughout turn
18	1	2 turn in tuck stand on one leg, free leg straight throughout turn
19	2	3 turn on one leg, free leg optional below horizontal
20	3	2 turn on one leg, free leg optional below horizontal
21	4	1 turn on one leg, free leg optional below horizontal
22	5	2 turn or more with heel of free leg forward at horizontal throughout turn
23	6	1 turn with heel of free leg forward at horizontal throughout turn
FX-SALTO from Gym99		
Class	ID	Category
24	0	Arabian double salto tucked
25	1	Salto forward tucked
26	2	Aerial walkover forward
27	3	Salto forward stretched with 2 twist
28	4	Salto forward stretched with 1 twist
29	5	Salto forward stretched with 1.5 twist
30	6	Salto forward stretched, feet land together
31	7	Double salto backward stretched
32	8	Salto backward stretched with 3 twist
33	9	Salto backward stretched with 2 twist
34	10	Salto backward stretched with 2.5 twist
35	11	Salto backward stretched with 1.5 twist
36	12	Double salto backward tucked with 2 twist
37	13	Double salto backward tucked with 1 twist
38	14	Double salto backward tucked
39	15	Double salto backward piked with 1 twist
40	16	Double salto backward piked

frame pairs $\mathbb{P}$, the temporal consistency score is:

$$\text{CLIP}_{\text{smooth}}(\tilde{V}) = \frac{1}{N} \sum_{(i,j)\in\mathbb{P}} \text{CLIP}(\tilde{V}(i), \tilde{V}(j)). \quad (29)$$

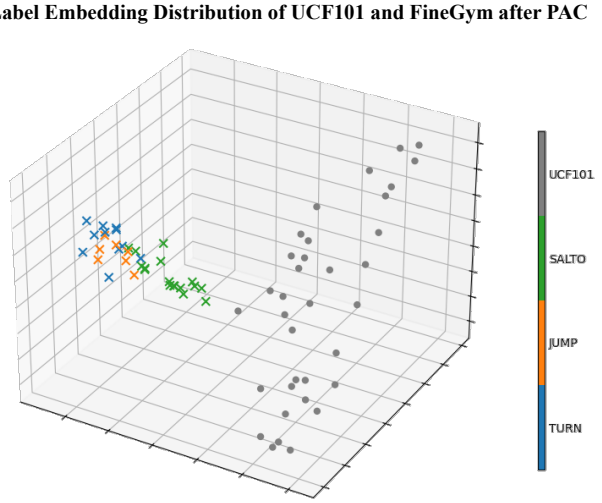

Figure 9. **Limitations of semantic consistency in original CLIP-SIM.** We utilize CLIP models to obtain the embedded textual features and Probably Approximately Correct (PAC) for dimensionality reduction. The distribution of embedded category labels from FX-JUMP, FX-TURN and FX-SALTO as well as UCF101 is shown. Label features from FineGym are entangled, while those from UCF101 are clearly seperated.

Limitations of $\text{CLIP}_{\text{smooth}}$: The original temporal consistency metric is unsuitable for fine-grained human actions, which inherently involve rapid and significant temporal changes. As demonstrated in Fig.17, models like T2I-Zero that generate predominantly static scenes paradoxically achieve the highest temporal consistency scores. This indicates that the metric fails to capture the dynamic nature of fine-grained actions, instead rewarding unnaturally smooth or static video sequences.

B.2. The Improved CLIP-SIM* Metrics

To overcome the aforementioned limitations, we propose an enhanced version of CLIP-SIM, termed CLIP-SIM*, specifically designed for evaluating fine-grained human action videos. CLIP-SIM* refines the calculations of domain consistency and temporal consistency by adopting a data-driven approach, while leaving the original semantic consistency as a minor metric.

❶ ***Improved Domain Consistency.*** Instead of relying on reference images generated by Stable Diffusion, CLIP-SIM* leverages ground-truth videos to select more relevant reference images. Specifically, we randomly choose ground-truth videos and extract three representative frames (*start, middle, end*) from each to form the reference set $\{I_j\}_{j=1}^{N}$, as shown in the right part of Fig. 10.

The domain consistency score is then computed as the average CLIP similarity between each generated frame and

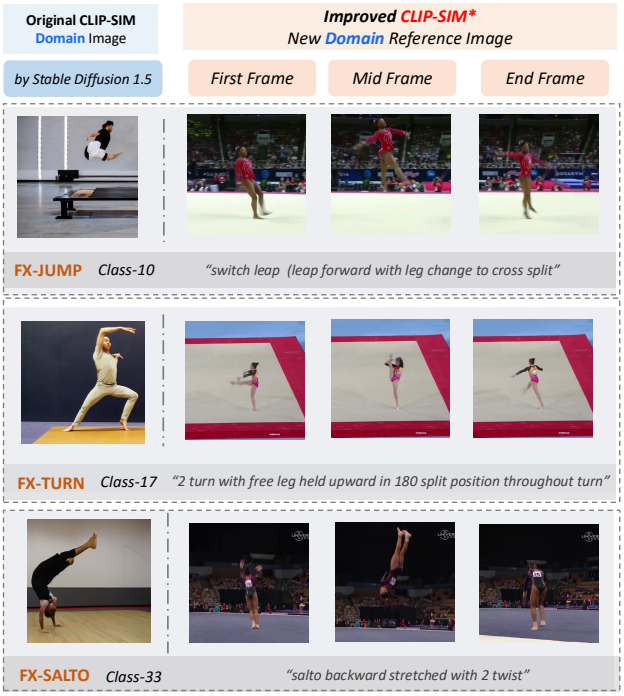

Figure 10. **Domain image of original CLIP-SIM and the improved CLIP-SIM* from FX-JUMP, FX-TURN and FX-SALTO.** Reference images generated by Stable Diffusion may not accurately reflect the nuances of specific actions or their dynamics (Original CLIP-SIM), while CLIP-SIM* randomly selects one video from the given class and extracts three representative frames (start, middle, end) to form a more reasonable reference set.

all reference images:

$$\text{CLIP}^*_{\text{text}}(\tilde{V}, \{I_j\}) = \frac{1}{N} \cdot \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{N} \text{CLIP}(\tilde{V}(t), I_j). \tag{30}$$

This approach ensures that the reference images are contextually and semantically aligned with the fine-grained actions being evaluated, thereby providing a more accurate measure of domain consistency.

❷ ***Improved Temporal Consistency.*** To better assess the temporal dynamics of fine-grained actions, we propose an improved temporal consistency metric within CLIP-SIM*, which preserves the temporal changing patterns inherent to specific action classes. Instead of enforcing smoothness across all frames, CLIP-SIM* compares the generated video with multiple reference videos from the same action category. For each action label, we select M reference videos V^{Ref} and uniformly sample K_i frames from each reference video, where $K_i \in \{1, 2, 4, 8, 16\}$. The temporal consistency score is then calculated as:

$$\text{CLIP}^*_{\text{smooth}}(\tilde{V}, V^{\text{Ref}}) = \sum_{l=1}^{M} \sum_{k=1}^{K_i} \text{CLIP}(\tilde{V}(k), V_l^{\text{Ref}}(k)). \tag{31}$$

This modification allows $\text{CLIP}^*_{\text{smooth}}$ to effectively measure whether the generated video replicates the temporal dynamics of specific fine-grained actions, addressing the shortcomings of the original temporal consistency metric, as shown in Fig.19.

B.3. Details of User Study

As discussed in the main paper, we evaluate the generation results through a user study, which provides a more reliable assessment. In practice, each participant is presented with a series of text-video, image-video, and video-video pairs and asked to rate *semantic consistency*, *temporal consistency*, and *domain consistency* on a scale **from 1 to 5**. The layout of the user study interface is illustrated in Fig. 16.

Specifically, we developed a questionnaire that tested all baseline models alongside our results. Each video result was accompanied by the same textual descriptions, reference images, and reference videos. Participants were instructed to objectively evaluate the similarity of the video results to this reference information. To ensure impartiality, we omitted any details about the models used and distributed the questionnaire to 20 professionals unfamiliar with our work, thereby obtaining objective data.

B.4. Other Metrics

PickScore. PickScore [36] trains a scoring function $s(\cdot)$ based on the CLIP framework using the large-scale user preferences dataset Pick-a-Pic to score the quality of generated images. Its performance in assessing generated images surpasses that of other evaluation metrics, even outperforming expert human annotators.

Given a text prompt P and an image I as input, PickScore calculates the score of the generated image as follows:

$$s(P, I) = E_{\text{txt}}(P) \cdot E_{\text{img}}(I) \cdot \tau \tag{32}$$

where E_{txt} and E_{img} represent the text encoder and image encoder, respectively, and τ denotes the learned scalar temperature parameter of CLIP.

While PickScore was originally developed for image evaluation, we have extended it to the domain of video evaluation. Specifically, given a text prompt P and a generated video $\tilde{V}$, we compute the average PickScore across all frames of the video:

$$\text{PickScore}(P, \tilde{V}) = \frac{1}{T} \sum_{t=1}^{T} s(P, \tilde{V}(t)) \tag{33}$$

where $\tilde{V}(t)$ denotes the t-th frame of the generated video, and T is the total number of frames.

Fréchet Video Distance (FVD). FVD [67] is a widely used metric for evaluating video generation models. In the domain of temporal analysis [13, 14], it is highly correlated with the visual quality of generated samples and

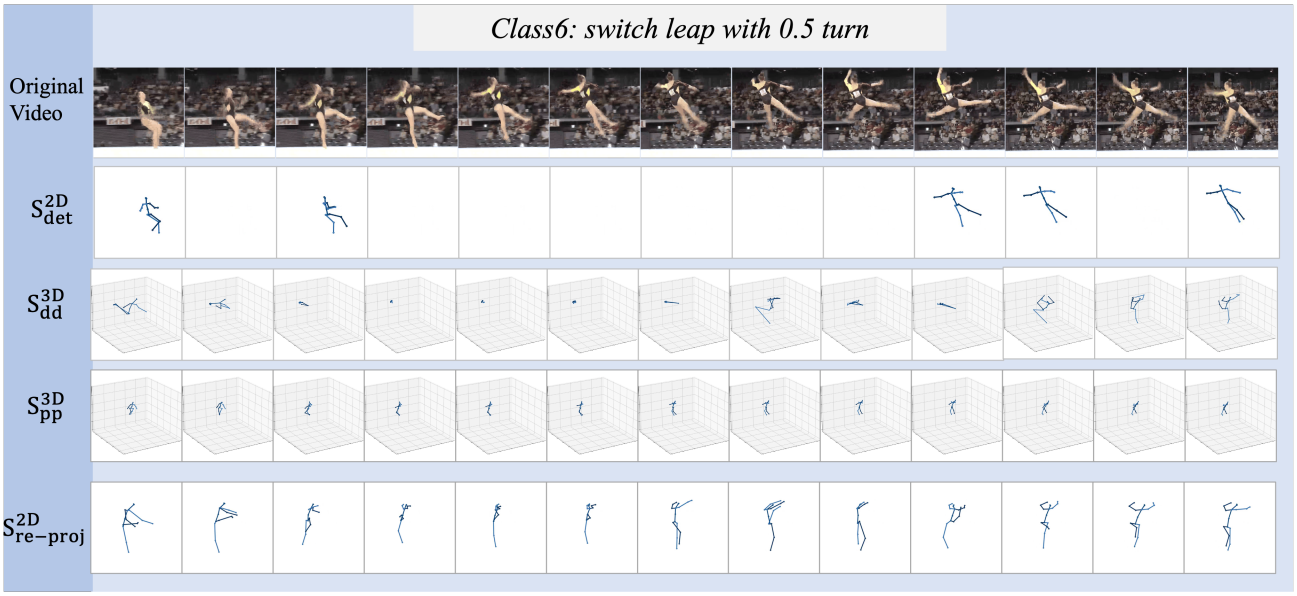

Figure 11. **Visualization of different pose sequences** on the class *"switch leap with 0.5 turn"* from the **FX-Jump** subset, demonstrating the complete transformation process within our framework.

assesses temporal consistency. FVD utilizes a pre-trained video recognition model to extract features from both real and generated videos, forming two sets of features, and then computes the mean and covariance matrices of these two sets. The FVD is represented as the Fréchet distance between these two distributions:

$$\text{FVD} = \|\mu - \tilde{\mu}\|^2 + \text{Tr}(\Sigma + \tilde{\Sigma} - 2(\Sigma\tilde{\Sigma})^{\frac{1}{2}}) \tag{34}$$

where μ and Σ are the mean and covariance matrix of the real video feature set, while $\tilde{\mu}$ and $\tilde{\Sigma}$ are the mean and covariance matrix of the generated video feature set. However, as observed in [37], unsatisfactory video generation results could achieve a higher FVD score, challenging its reliability.

C. Additional Illustration & Analysis

C.1. Elaboration on Euler-Lagrange Equations

In the main paper, we use the following equation to represent the process in Lagrangian Mechanics:

$$M(q)\ddot{q} = J(q,\dot{q}) - C(q,\dot{q}), \tag{35}$$

which is a common form used in robotics and dynamics, known as the equation of motion in terms of mass matrix $M(q)$, generalized forces $J(q,\dot{q})$, and Coriolis and centrifugal forces $C(q,\dot{q})$. Here we elaborate on its relation with the original Euler-Lagrange Equations, *i.e.*:

$$\frac{\partial L}{\partial q^i}(t,q(t),\dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}(t,q(t),\dot{q}(t)) = 0. \tag{36}$$

Assume the kinetic energy of the system is given by $T = \frac{1}{2}\dot{q}^T M(q)\dot{q}$, and the potential energy is typically a function of the generalized coordinates q denoted by $V = V(q)$, then the Lagrangian is defined as:

$$L = T - V = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q). \tag{37}$$

Then we calculate $\frac{\partial L}{\partial q^i}$ and $\frac{\partial L}{\partial \dot{q}^i}$:

$$\frac{\partial L}{\partial q^i} = -\frac{\partial V}{\partial q^i} + \frac{1}{2}\dot{q}^T\frac{\partial M(q)}{\partial q^i}\dot{q} \tag{38}$$

$$\frac{d}{dt}\left(M_{ij}(q)\dot{q}^j\right) = \dot{q}^j\frac{\partial M_{ij}}{\partial q^k}\dot{q}^k + M_{ij}(q)\ddot{q}^j \tag{39}$$

and substitute these results into the Euler-Lagrange equation:

$$-\frac{\partial V}{\partial q^i} + \frac{1}{2}\dot{q}^T\frac{\partial M(q)}{\partial q^i}\dot{q} - \left(\dot{q}^j\frac{\partial M_{ij}}{\partial q^k}\dot{q}^k + M_{ij}(q)\ddot{q}^j\right) = 0, \tag{40}$$

where $-\frac{\partial V}{\partial q^i}$ represents the partial derivative of the potential energy with respect to the coordinates, i.e., the generalized force, i.e., $J(q,\dot{q})$. Thus we could obtain the following formulation:

$$M(q)\ddot{q} = J(q,\dot{q}) - C(q,\dot{q}). \tag{41}$$

C.2. Visualization of the Pose Modality

Recall that our FinePhys framework fully leverages skeletal data through a sequence of specialized modules: (1) The online pose estimator generates detected 2D poses, denoted as S^{2D}_{detect}; (2) Then the in-context-learning module processes and transforms them into S^{3D}_{dd}; (3) After the PhysNet module we obtain S^{3D}_{pp}, (4) and finally we re-projected the average of S^{3D}_{dd} and S^{3D}_{pp} into 2D space to obtain $S^{2D}_{\text{re-proj}}$.

Fig.11, Fig. 12, Fig. 13 present additional visualizations of these pose sequences, illustrating the entire transformation process within our framework. Due to the large vari-

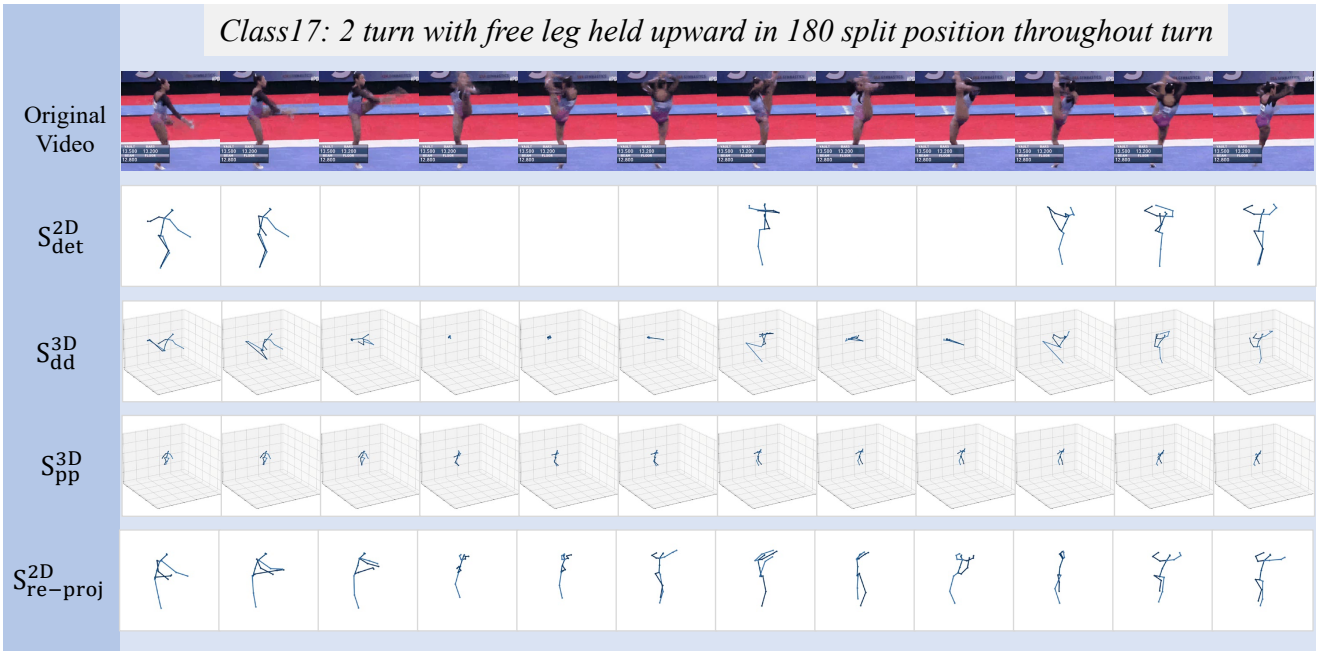

Figure 12. **Visualization of different pose sequences** on the class *"2 turn with free leg held upward in 180 split position throughout turn"* from the **FX-Turn** subset, demonstrating the complete transformation process within our framework.

ation and high complexity of fine-grained actions, the detected 2D poses (S^{2D}_{detect}) exhibit significant misidentifications across joints throughout the video. The in-context learning module improves these poses, enabling S^{3D}_{dd} to partially reconstruct missing or distorted skeletons in each frame. However, in cases of severe distortion, the data-driven approach becomes unstable, resulting in S^{3D}_{dd} being noisy and physically implausible. The PhysNet module mitigates this issue by producing S^{3D}_{pp}, which is more stable and constrained, effectively correcting deviations in S^{3D}_{dd}. Consequently, the averaged and re-projected 2D poses ($S^{2D}_{\text{re-proj}}$) show substantial improvements compared to the original detections, validating the efficacy of our approach.

C.3. More Generated Results and Comparison

In this section, we present additional qualitative results to demonstrate the effectiveness of our proposed FinePhys framework in generating fine-grained human action videos.

We compare the generated results of FinePhys with those of baseline methods across three action subsets: FX-JUMP, FX-TURN, and FX-SALTO, as illustrated in Fig. 17, Fig. 18, and Fig. 19, respectively. The key observations are as follows: ❶ Our CLIP-SIM* metric more accurately reflects the quality of video generation compared to the original CLIP-SIM metric. For example, methods such as Follow-Your-Pose and Latte achieve high scores on the original Domain Score, yet the generated actions exhibit significant inconsistencies with physical laws. Similarly, T2V-zero attains the highest score on the Smooth Score by generating continuous identical frames, which lack realistic motion dynamics. In contrast, CLIP-SIM* scores align more closely with human intuition, providing a more reliable assessment of video quality.

❷ FinePhys consistently outperforms other baseline methods across different action categories. Baseline methods that lack guidance from physical information often produce unrealistic limb movements. For instance, Latte displays multiple limb artifacts in Class 14, and VideoCraft shows unrealistic levitation in Class 20 . In contrast, FinePhys incorporates physics modeling through the PhysNet module, resulting in more natural and coherent actions that adhere to real-world physical constraints.

C.4. Limitation and Future Work.

Intractable Cases. Although FinePhys outperforms its competitors in generating results, significant challenges remain unresolved. High-speed motions and substantial body deformations pose considerable difficulties, particularly when they are intertwined, as seen in *salto* routines. Generating fine-grained actions such as *double salto backward stretched* is currently intractable, as shown in Fig. 14, let alone accurately distinguishing between actions like *"salto backward stretched with 2.5 twist"*, *"salto backward tucked with 1 twist"*, and *"double salto backward tucked with 1 twist"*. We encourage future research efforts to address these complex scenarios.

Reliance on Initial Pose Detection. FinePhys fully utilizes the pose modality; however, the initial step of the

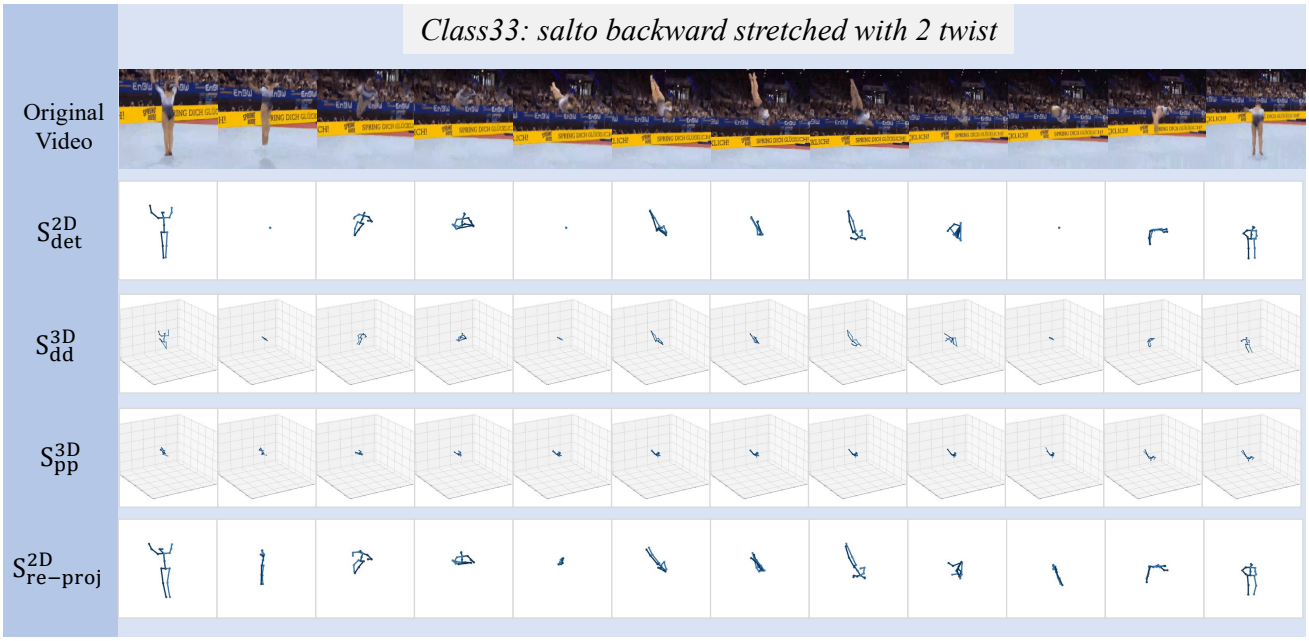

Figure 13. **Visualization of different pose sequences** on the class *"salto backward stretched with 2 twist"* from the **FX-Salto** subset, demonstrating the complete transformation process within our framework.

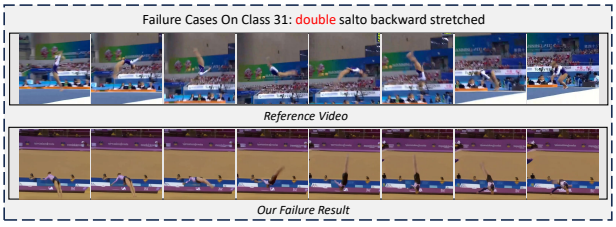

Figure 14. **Limitations in intractable cases.** For class 31: *double salto backward stretched*, FinePhys fails to generate a double salto, resulting in only a single flip being observed.

pipeline involves online 2D pose estimation. Due to the complexity of fine-grained human actions, we observed that the online pose estimator can occasionally fail completely, resulting in no detected 2D poses, as shown in Fig. 15. In such cases, the initial poses rely entirely on the pose prior used in the in-context learning module. Even if we can restore the human structure spatially, no motion is present. In future work, we will consider selecting appropriate scenarios to evaluate our current FinePhys implementations and explore additional modalities (e.g., optical flow) to address this issue.

Focus on Fine-grained Human Actions. Although video generation techniques have been extensively explored and improved, applying these methods to the specific and challenging domain of fine-grained human actions can reveal the limitations of current approaches and inspire future advancements [8, 59, 61]. In this work, we select three fine-grained human action subsets, each encoding distinct mo-

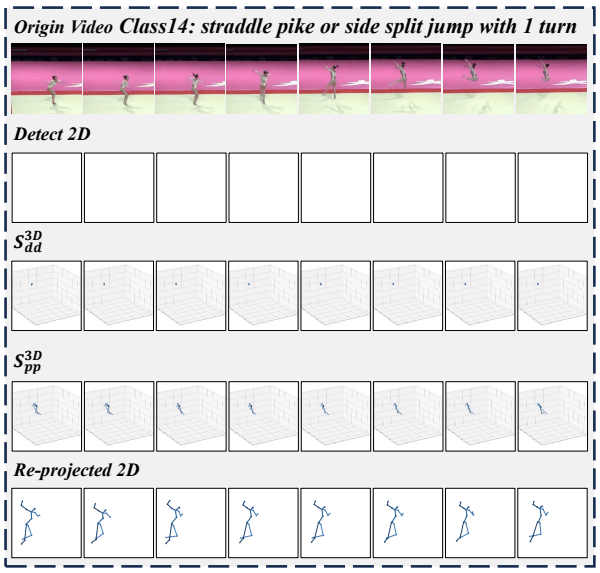

Figure 15. **Negative Impact of Initial Pose Detection.**. Current online pose estimators may fail completely due to the complexity of fine-grained human actions, which affects subsequent processing stages in the FinePhys framework. Even when the physical structure of the human body is spatially restored, the intricate motion dynamics cannot be accurately reconstructed, resulting in unrealistic or static video outputs.

tion dynamics: ❶ *Turning* Focuses on precise rotational movements; ❷*Jumping* emphasizes rapid vertical motion combined with moderate rotations; ❸ While *Salto* involves complex aerial maneuvers with multiple twists and flips, and is the most challenging. By conducting comprehensive quantitative comparisons alongside qualitative analyses, we aim to draw greater attention to the challenges inherent in generating fine-grained human actions. This focused evaluation not only highlights the strengths and weaknesses of existing methods but also provides valuable insights for future research and development in this domain.

Further Exploration on Physics. In future work, we aim to enhance the integration of physics modeling in video generation from diverse perspectives, such as collision dynamics, fluid interactions, etc. Currently, generating fine-grained human actions restricts the model's ability to focus solely on motion dynamics, as it must also account for the spatial structure of the human body [9, 65, 73, 74]. To address this complexity, we plan to simplify scenes by utilizing basic geometric shapes for environmental interactions, thereby reducing model complexity while maintaining a robust incorporation of physical principles. Additionally, we will investigate the incorporation of physical laws into video generation, which may involve developing new algorithms or refining existing techniques to more accurately simulate real-world physical behaviors.

Figure 16. **Display of the interface of User Study.**

Figure 17. **Qualitative Results on FX-JUMP.** FX-JUMP focuses on the motion continuity of the gymnastics' body. Compared with other baselines, our method demonstrates superior performance in understanding physical consistency.

Figure 18. **Qualitative Results on FX-TURN.** FX-TURN focuses on the minor difference of the gymnastics' body. Compared with other baselines, our method demonstrates superior performance in understanding complex and fine-grained semantics, keeping the consistency of bio-physical characteristics, and adhering to the physical principles.

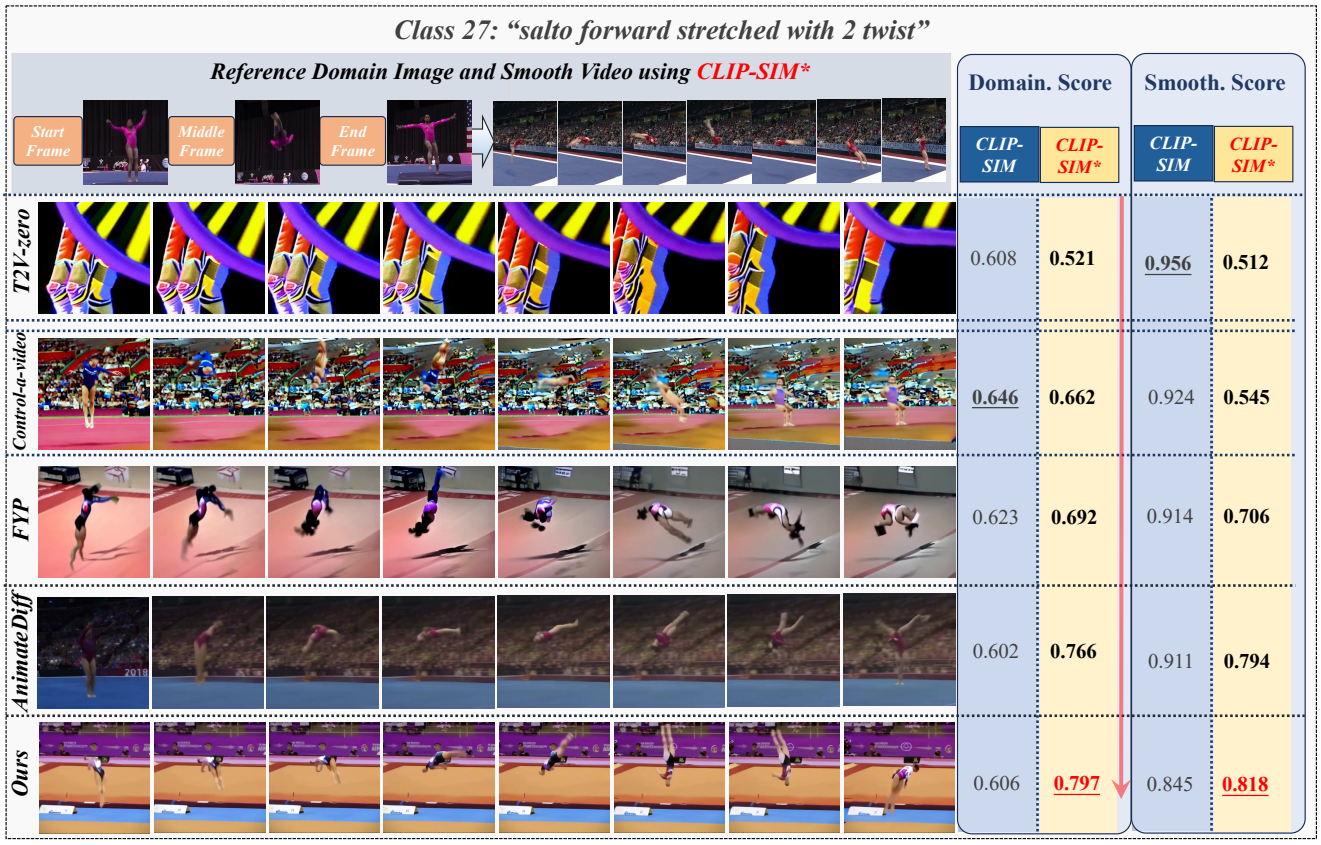

Figure 19. **Qualitative Results on FX-SALTO.** FX-SALTO demands gymnastics's body rotates 360° around a horizontal axis with the feet passing over the head, which is the most difficult in all of three sub-datasets in FineGym. Compared with other baselines, results in our methods maintain better temporal consistency, more adhering to the bio-physical rules.