

Context-Aware Motion In-Betweening and Editing Based on Motion Diffusion Model

이름 이현우

지도교수 유리



연구 배경 및 목표

최근 애니메이션 제작, 게임 개발, 비디오 복원 등의 산업에서의 빈 구간 모션을 채워 넣거나 편집하는 인비트윈 워닝 연구가 활발히 이뤄지고 있다. 사람이 수동으로 모션 데이터를 수정하지 않으면서도 모션 편집 및 생성을 효율적으로 진행하는 것은 중요한 도전과제이다.

Diffusion 기반 모션 생성 모델은 랜덤 모션을 생성하는 문제 및 사용자의 의도를 고려하지 않을 수 있는 문제가 있다. 누락된 구간에 대해 앞 뒤 모션 문맥을 고려하면서도 사용자 의도를 반영한 모션을 생성할 수 있다면 문맥 상 자연스러우면서도 비용 효율적인 모션 편집 및 생성이 가능할 것이다.

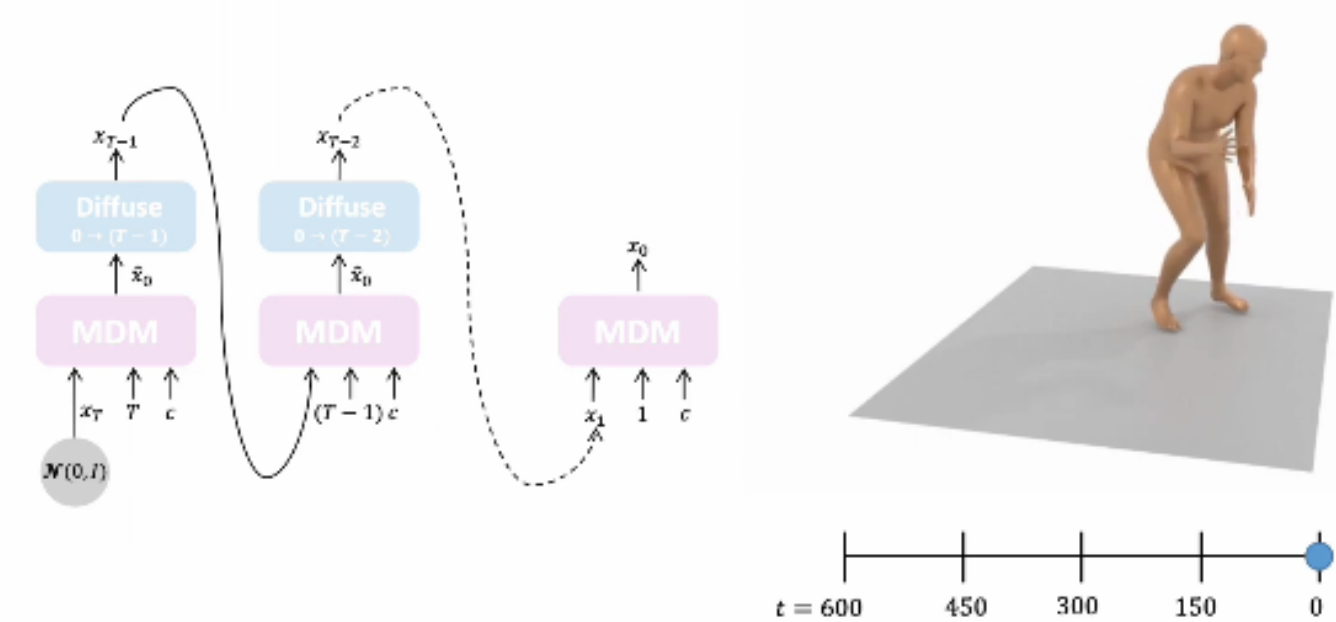
본 연구는 불완전한 모션 구간에 대해 앞 뒤 모션의 문맥을 고려한 자연스러운 모션 인비트윈 작업을 효율적으로 수행하는 것을 목표로 한다. 또한 Diffusion 기반 모델이 랜덤하게 모션을 생성하는 문제 및 사용자의 의도를 고려하지 않는 점을 보완하고자 한다. 앞 뒤 모션의 문맥을 고려한 자연스러우면서도 사용자의 의도를 UI, 텍스트 조건을 통해 반영한다면 상황에 적합한 모션을 효율적으로 생성할 수 있을 것이다.



연구 진행 과정

1. 타겟 논문 및 사용 데이터셋

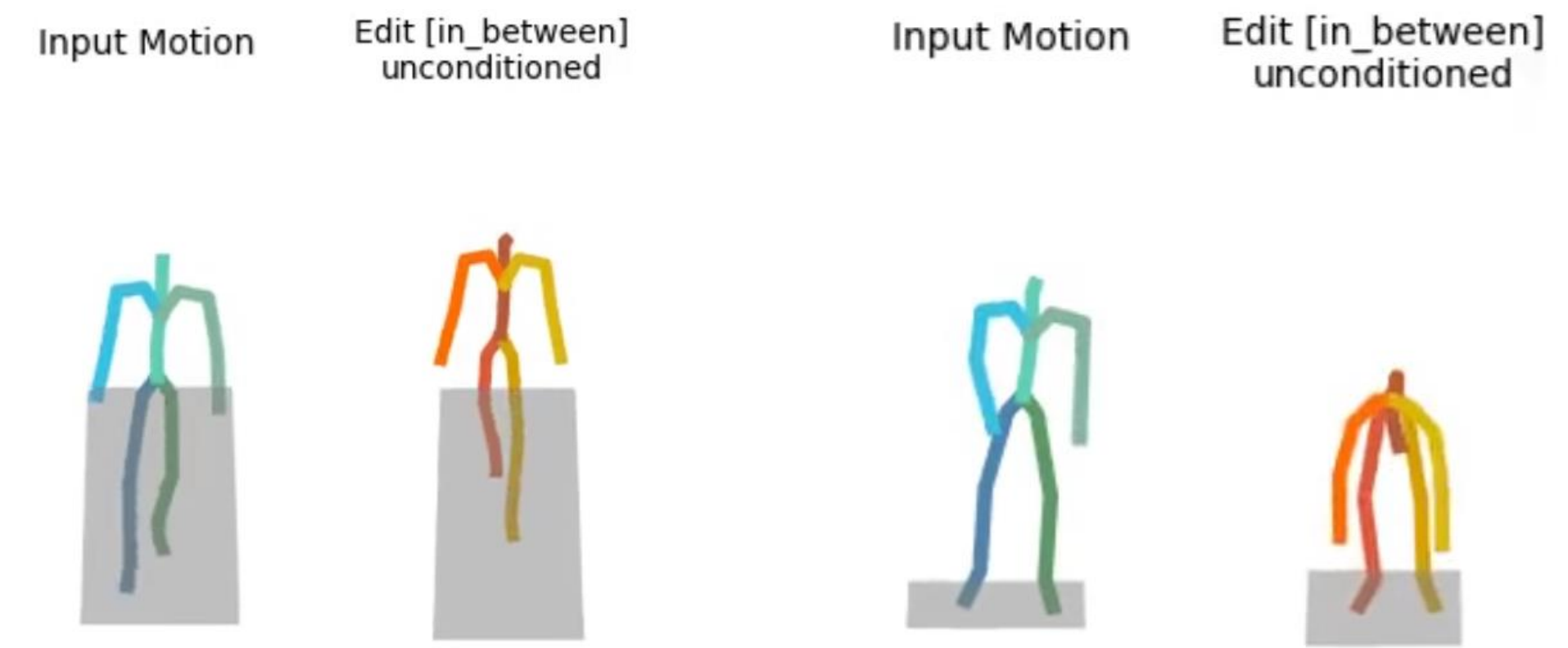
Motion Diffusion



[그림 1] MDM: Human Motion Diffusion Model 예시

Diffusion 기반 모션 생성 모델은 샘플링 다양성을 높이면서도 텍스트 조건을 기반으로 문맥에 맞는 모션을 생성하는 데 있어 효과적인 모델이다. 불완전하거나 모호한 입력에서도 유연하게 모션을 생성하는 데에 강점을 보이면서도, 일부 시퀀스에 대해 모션의 전후 문맥을 완전하게 반영하지 못하는 경우가 발생한다.

본 연구는 기존 Motion Diffusion Model 논문의 기학습 모델 및 HumanML3D 데이터셋을 사용하였다. 위 데이터셋은 다양한 모션에 상응하는 자연어 설명으로 구성되어 있는 강점을 가진다. 본 연구는 해당 연구의 다양한 샘플링과 텍스트 조건 기반 생성 기능 및 데이터셋을 활용하면서도, 문맥에 맞는 인비트윈 모션 생성을 수행하는 것을 목표로 한다. 구동 환경 설정 후 관찰 결과 Ground-truth 모션과 문맥상 일치하지 않은 경우가 있었으며, 보완하기 위한 샘플링 과정을 수정하였다.

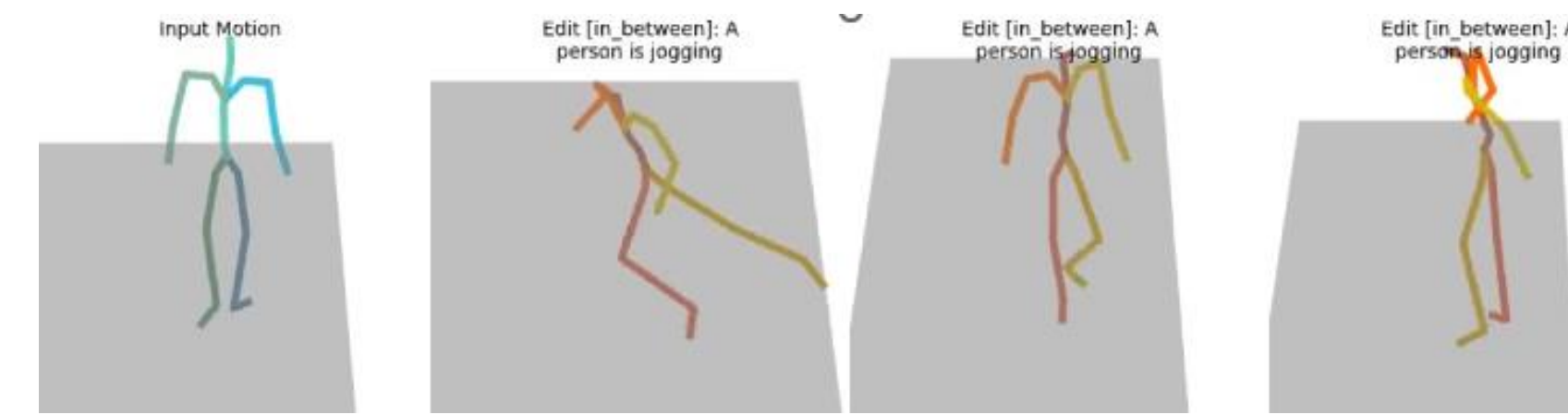


[그림 2,3] 문맥상 불일치하게 생성된 인비트윈 모션 예시. 각각 입력과 반대 방향으로 걷거나 속이고 있다.

2. 입력 모션의 특성을 추출한 후 샘플링에 반영 작업 수행

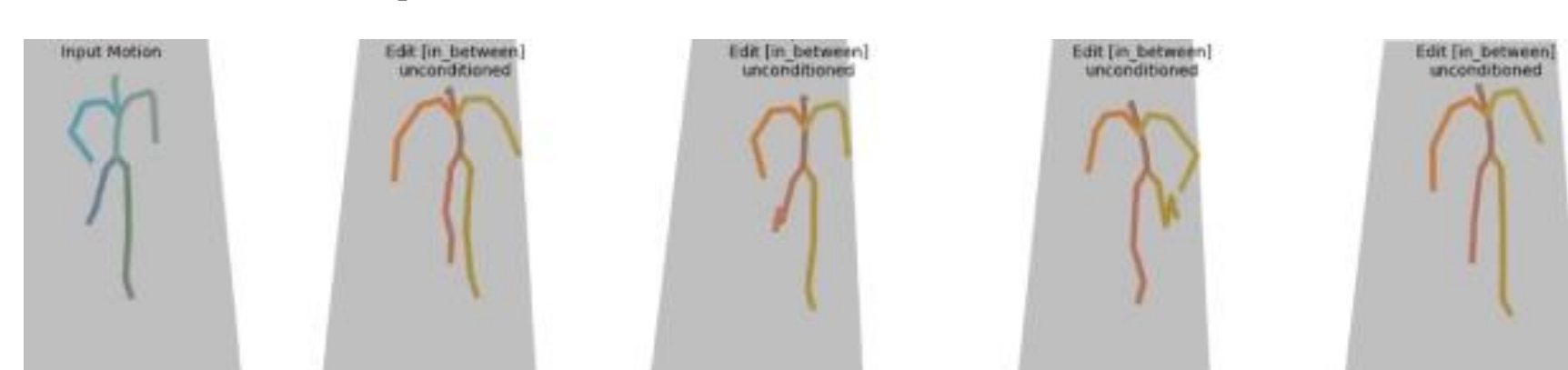
본 연구는 Ground-Truth 입력 모션이 앞 뒤 일정 비율로 주어지고, 중간 빈 구간에 해당하는 모션을 생성(인비트윈)하는 상황을 가정하였다.

입력된 모션의 관절 위치 데이터로부터 회전 및 속도 정보를 얻은 후, 임의의 기준과 개수로 샘플링을 조절할 수 있도록 수정했다. 우선 입력 모션의 속도와 대칭성이 비슷한 것을 중심으로 샘플링을 수행하였다. 이렇게 샘플링 조건에 변화를 주어 학습한 결과 모션의 페이스와 대칭성이 비슷한 것으로 샘플링 할 수 있었다. 그러나 [그림 4]와 같이 입력 모션과 동떨어진 방향으로 움직이거나 속도만 일치하면 비슷하다고 판단하는 문제점이 있었다.



[그림 4] 속도와 대칭성은 일치하나 입력 모션과의 궤적 또는 문맥이 일치하지 않은 예시

위 [그림 4]와 같은 문제점을 해결하기 위해 입력 모션의 주기성과 궤적과 비슷한 모션을 샘플링 하도록 했다. 주기성과 궤적은 각각 각 관절의 신호를 고속 푸리에 변환으로 분해하거나 모션의 전체 각도 차이 합으로 구했다. 또한 모션 중 특정 관절이 고정되어 움직이는 경우 (무언가 한 손으로 들고 걷는 모션)를 고려하여 특정 관절을 지정하여 해당 관절 주기만 비슷하게 샘플링 가능하도록 수행했다. 그 결과 [그림 5]와 같이 뒷걸음 치는 모션에 대해 이전까지와 달리 같은 방향으로 뒷걸음치기만 하는 문맥의 모션을 샘플링 할 수 있었다.

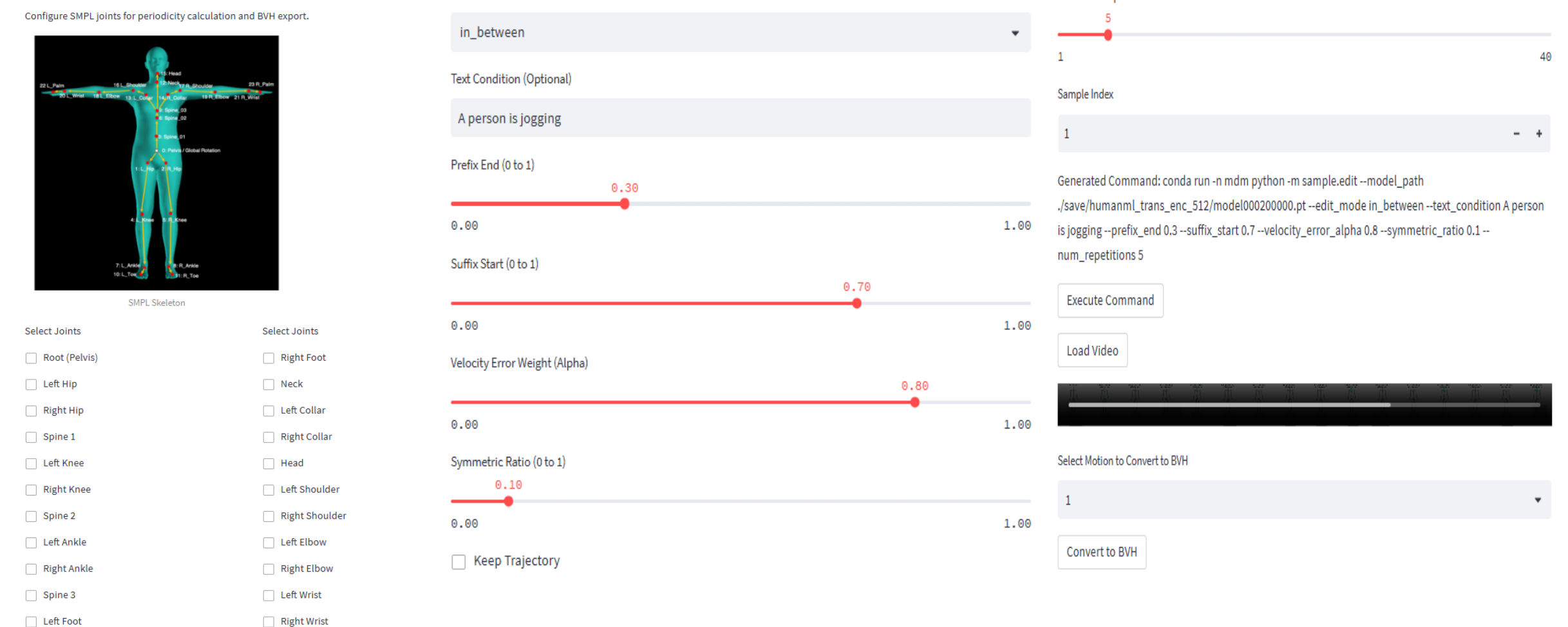


[그림 5] 뒷걸음 치는 문맥에 맞게 뒷걸음만 치는 모션만 생성된 모습. 그러나 입력 모션과 비슷한 문맥만 생성한다는 단점 또한 지닌다.



결과 및 분석

Motion In-betweening Model Interface with SMPL Joints and BVH Conversion



사용자 친화적으로 원하는 인비트윈 모션 옵션을 조절하고 애니메이션 파일로 변환하는 UI 추가를 수행했다. 생성하고자 하는 구간 길이, 특정 기준(속도, 주기, 대칭성, 궤적 유지 여부 등) 및 관절 지정, 샘플링 횟수 등의 옵션을 조절할 수 있다. 또한 영상을 불러와 원하는 샘플을 지정하여 애니메이션 파일(.bvh)로 변환할 수 있다. 그러나 현재 HumanML3D 기반 데이터셋의 일반성 한계와 입력 모션에 비슷해 지는 경향을 보이는 한계를 개선해야 할 것으로 보인다.