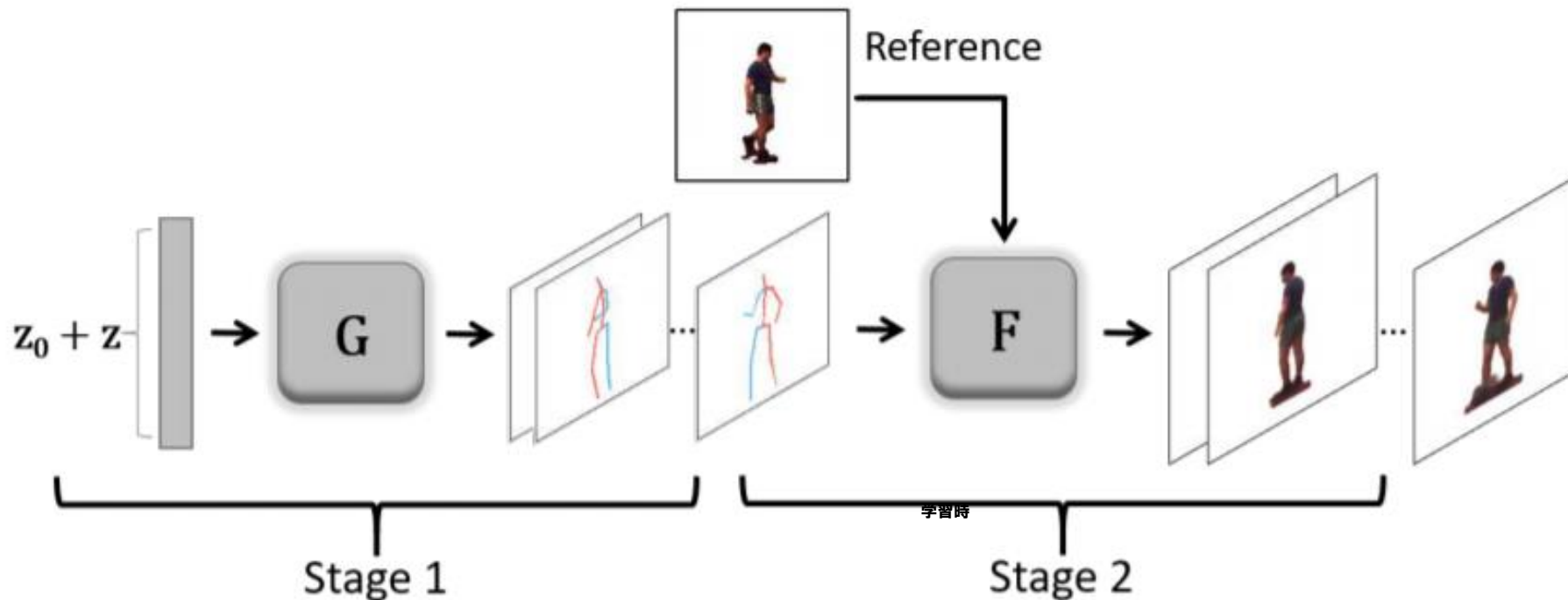


初期ポーズ生成の改良とGCNの導入による ポーズシーケンス生成モデルの拡張

電気通信大学 寺内健人 柳井啓司



研究目的: 自然なポーズシーケンスの生成



Caiら [CVPR, 2017]

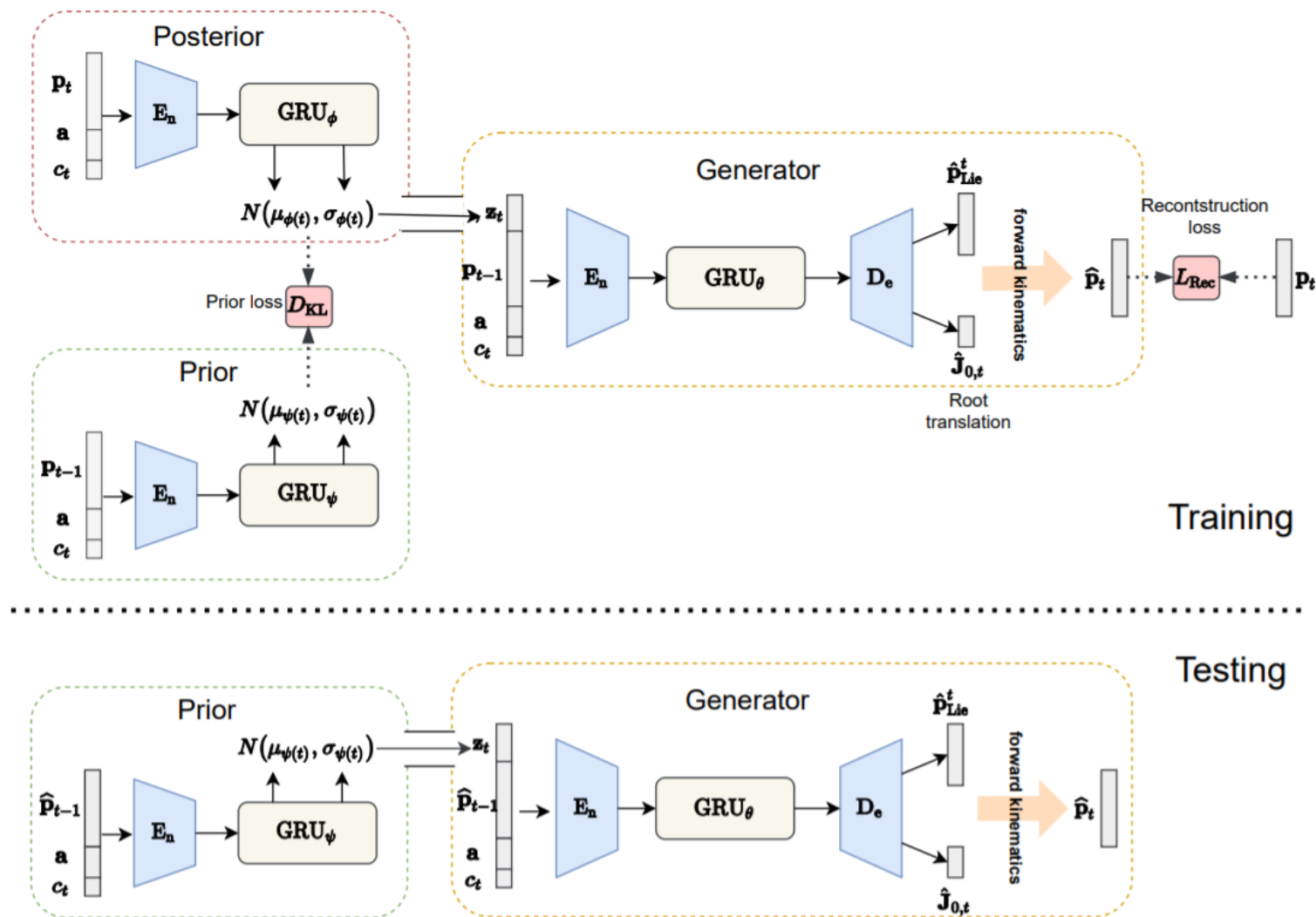
人間動作の動画生成手法ではポーズシーケンスの生成, ポーズシーケンスからの動画生成の2段階の手法が存在



ポーズ生成を自然にすることで自然な動作をする人間動作動画が生成可能



関連研究: Action2Motion [ACM Multi Media, 2020]



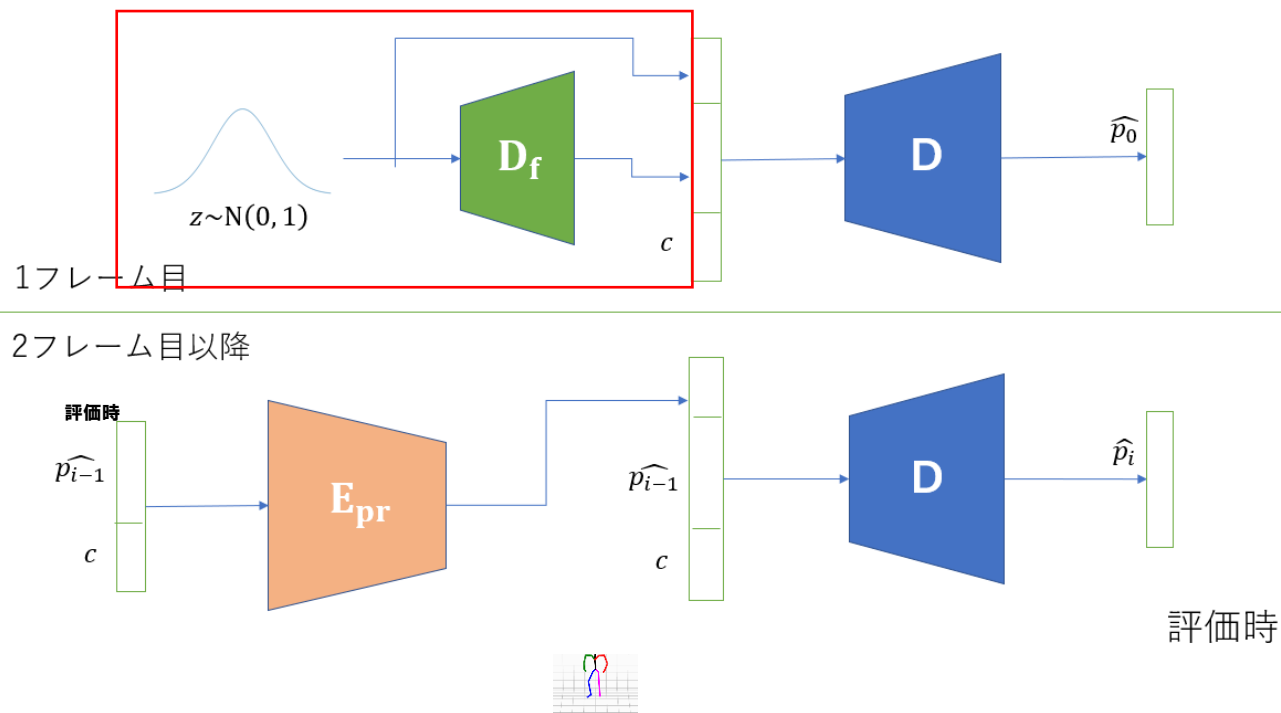
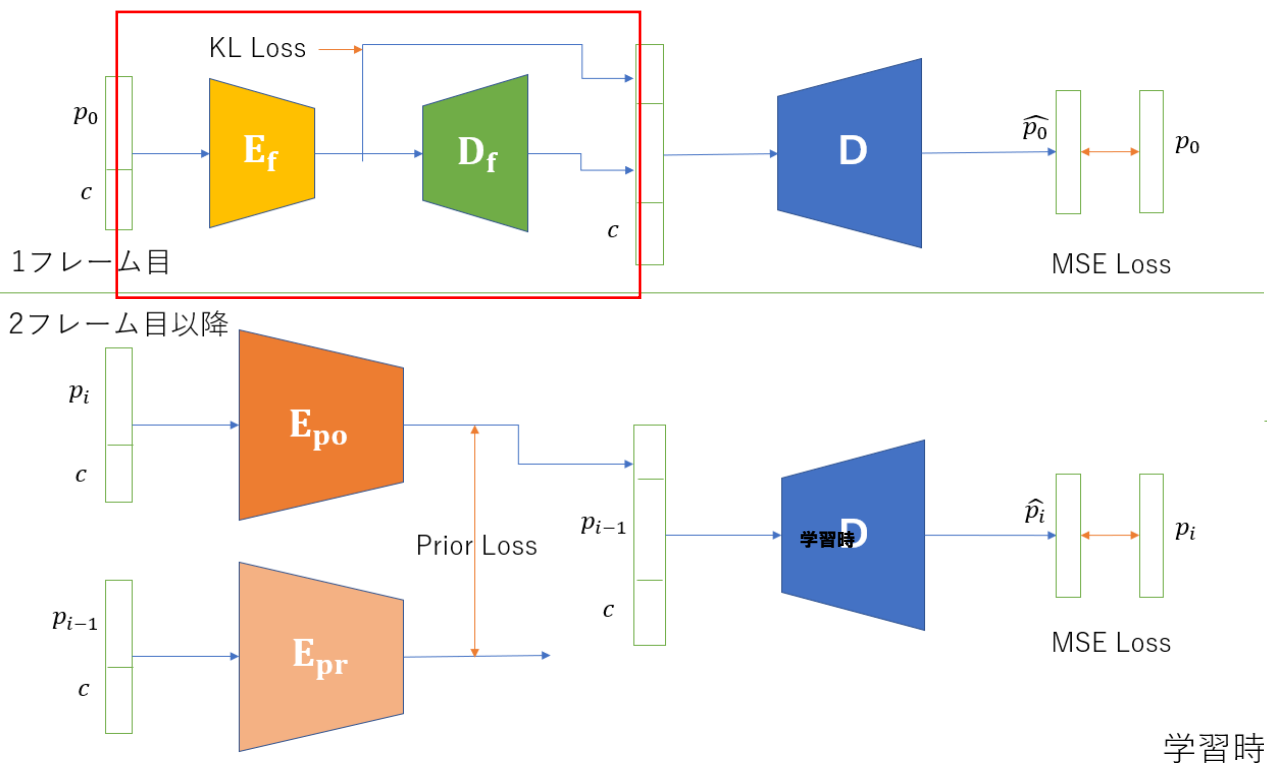
前フレームエンコーダ, 次フレームエンコーダ, デコーダ構成のモデル

前フレームエンコーダの出力と次フレームエンコーダの出力の分布を近付けるPrior Lossを導入, 再構成ロスと合わせて学習

次フレームエンコーダと出力が近づいた前フレームエンコーダを用いて次フレームを逐次的に生成



提案手法: 変更 (1) 初期フレーム考慮のモジュール追加



1フレーム目をエンコードする初期フレームエンコーダ,
1フレーム目のデコーダ条件を生成する初期フレーム条件デコーダを
追加

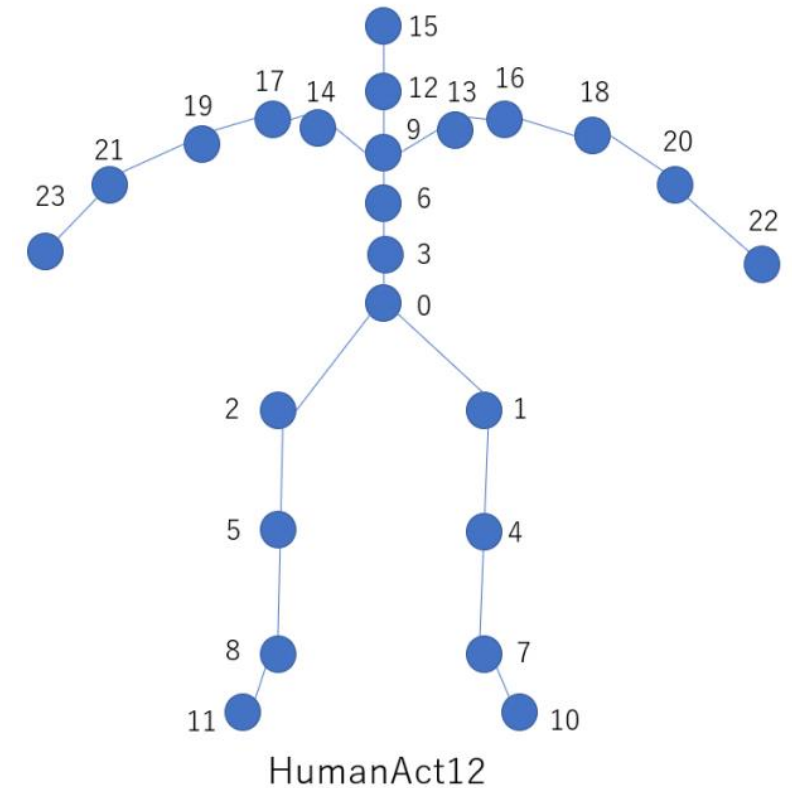
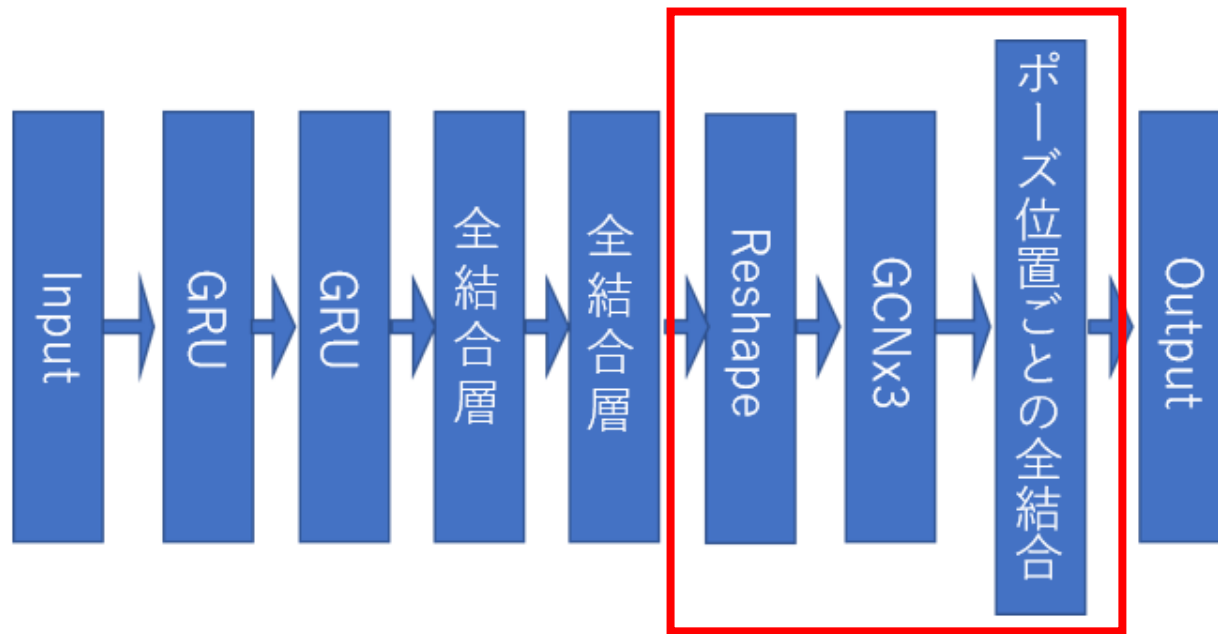
生成時は, 1フレーム目は正規分布のノイズから
初期フレーム条件デコーダを用いて条件生成

2フレーム目以降はAction2Motionと変わらず, 前フレームエンコーダ
と次フレームエンコーダの出力を近づけるように学習

2フレーム目以降は前フレームエンコーダで前フレームを
エンコードすることで逐次的に次フレーム生成



提案手法: 変更 (2) デコーダーにGCNを導入



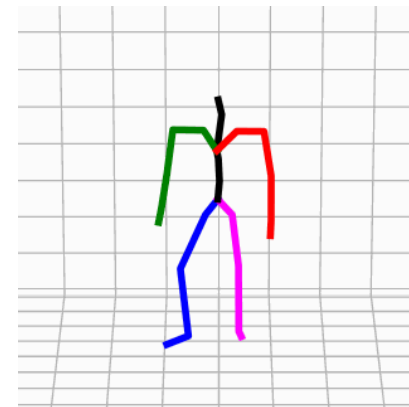
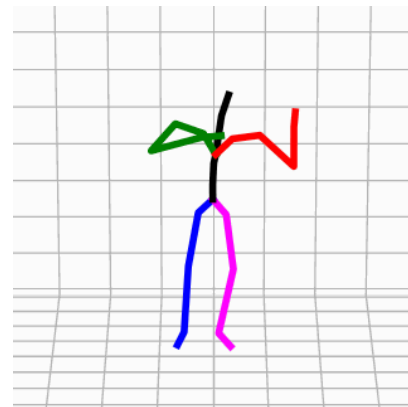
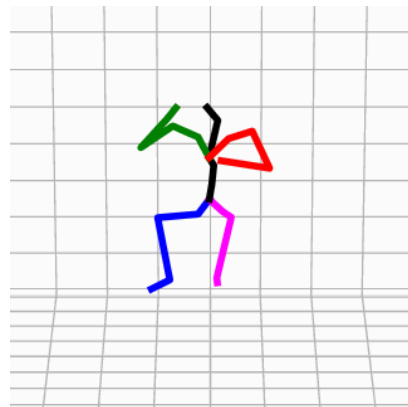
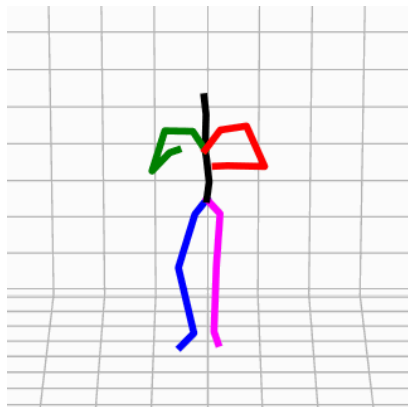
デコーダにGCNを導入することを考え, Action2MotionのGRUの後が続く全結合層の後にグラフ表現に埋め込み, 3層のGCNで畳み込む

畳み込み後にそれぞれのポーズ位置ごとの意味的な一貫性を保つために位置ごとに全結合

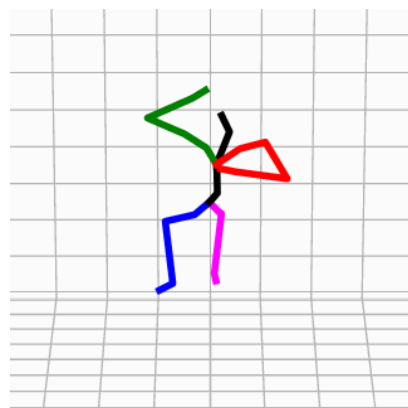
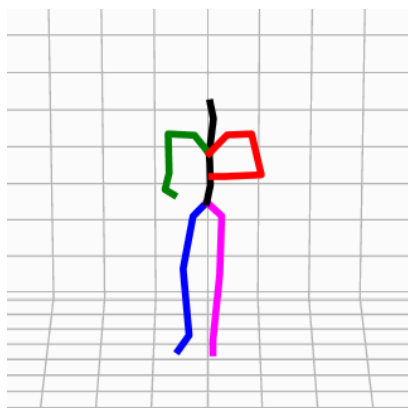


定性評価: 生成したポーズシーケンスの例

Action2
Motion



提案手法



走る

電話をする

投げる

座る



定量評価

Action2Motionと同等のAccuracyを達成しつつ, FIDを改善した

(1) のみだとFIDは改善するが, Accuracyは悪化し, (2) のみの変更だと生成に失敗した

| 手法 | Accuracy ↑ | FID ↓ | Diversity → | Multimodality → |
|---------------|------------|--------|-------------|-----------------|
| real | 0.997 | 0.092 | 6.857 | 2.449 |
| Action2Motion | 0.923 | 2.458 | 7.032 | 2.870 |
| 提案モデル (1) | 0.864 | 1.979 | 6.924 | 3.388 |
| 提案モデル (2) | 0.542 | 13.599 | 5.933 | 3.309 |
| 提案モデル | 0.924 | 2.252 | 6.962 | 2.861 |

Accuracy: アクション条件で生成したポーズシーケンスを分類モデルで分類したときのアクション条件との一致率

FID: RNNベースのアクション分類モデルで抽出した特徴空間での分布間距離

Diversity: 全カテゴリクラスにまたがった生成の多様性

Multimodality: 各カテゴリクラスでの生成の多様性

