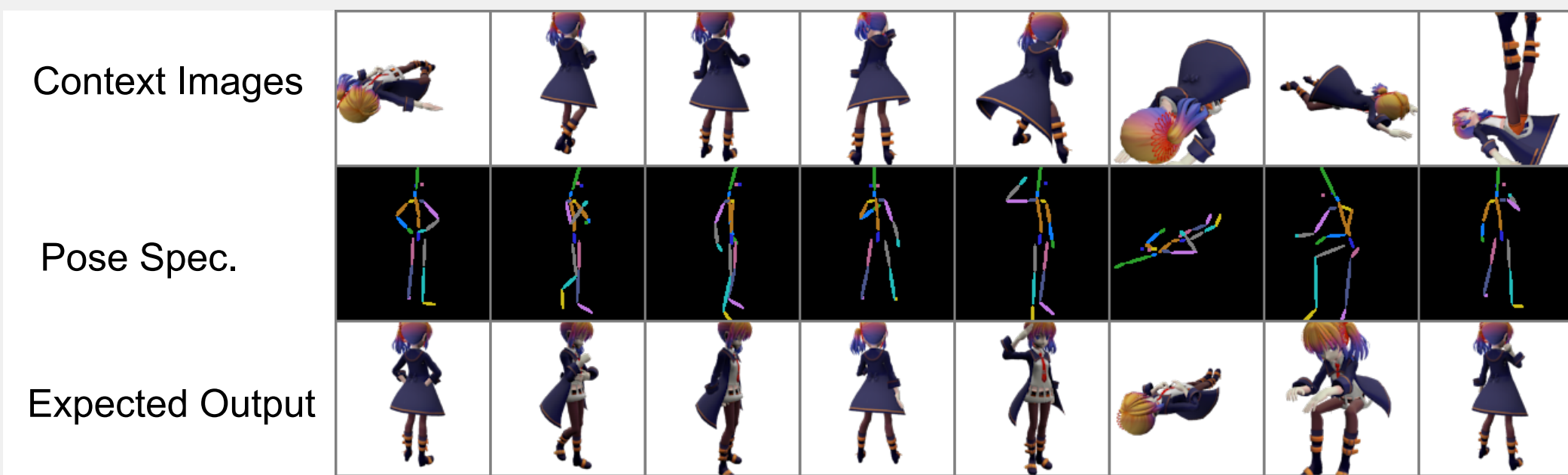


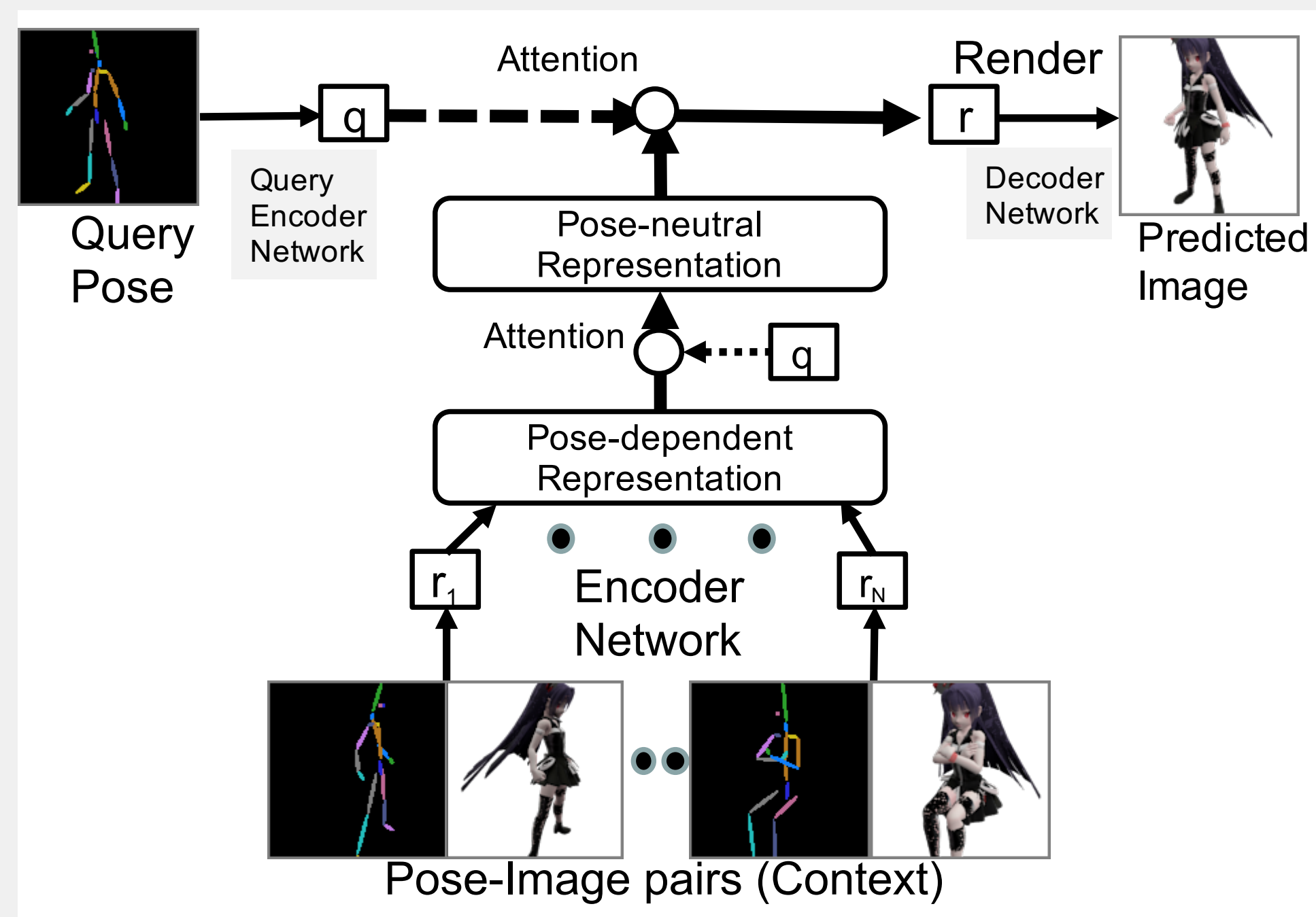
目的と成果

- ▶ 目的：任意の人物像の任意の姿勢での描画
 - ▶ 任意の人物像を異なる姿勢をとった数枚の人物画像で与える → コンテキスト画像
 - ▶ 任意の姿勢を画像として指定
 - ▶ 即時に、指定した人物の指定した姿勢の像を描画
- ▶ 成果
 - ▶ ポーズに依存しない抽象表現が得られていることを確認
 - ▶ ある程度の品質の描画を実現



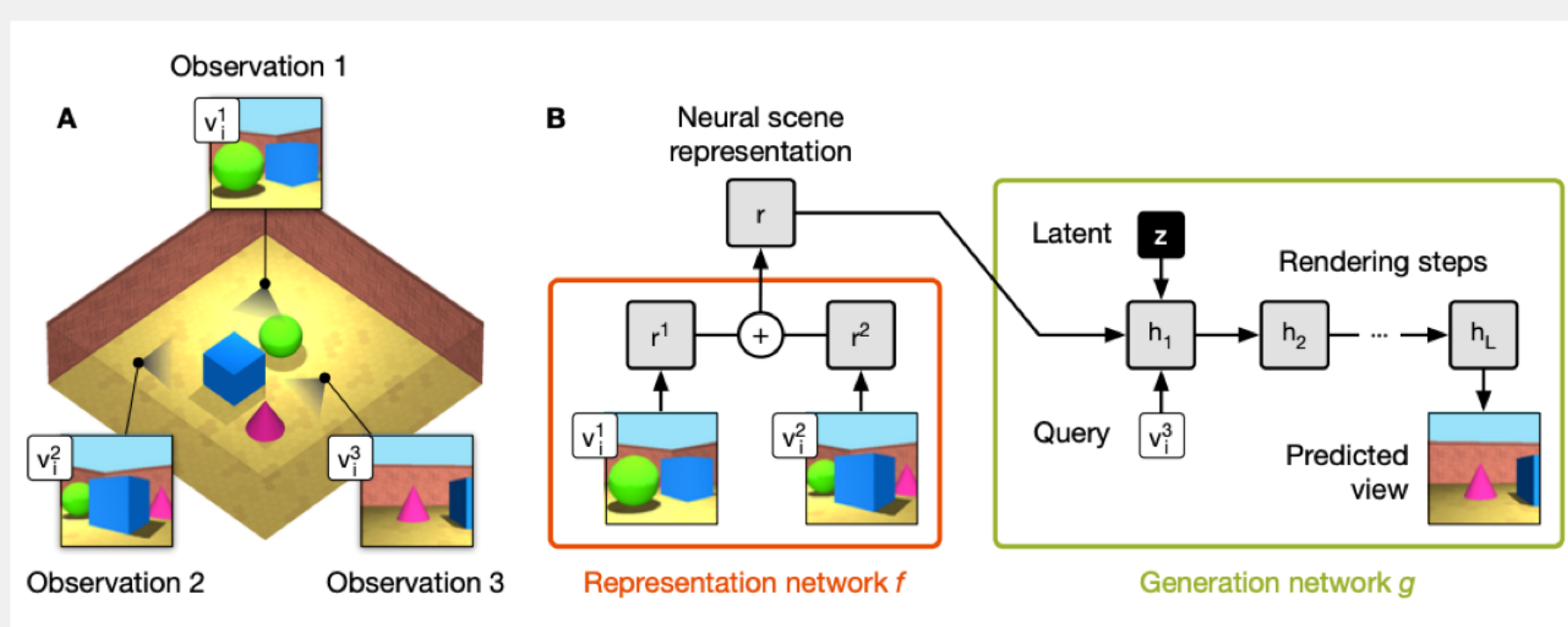
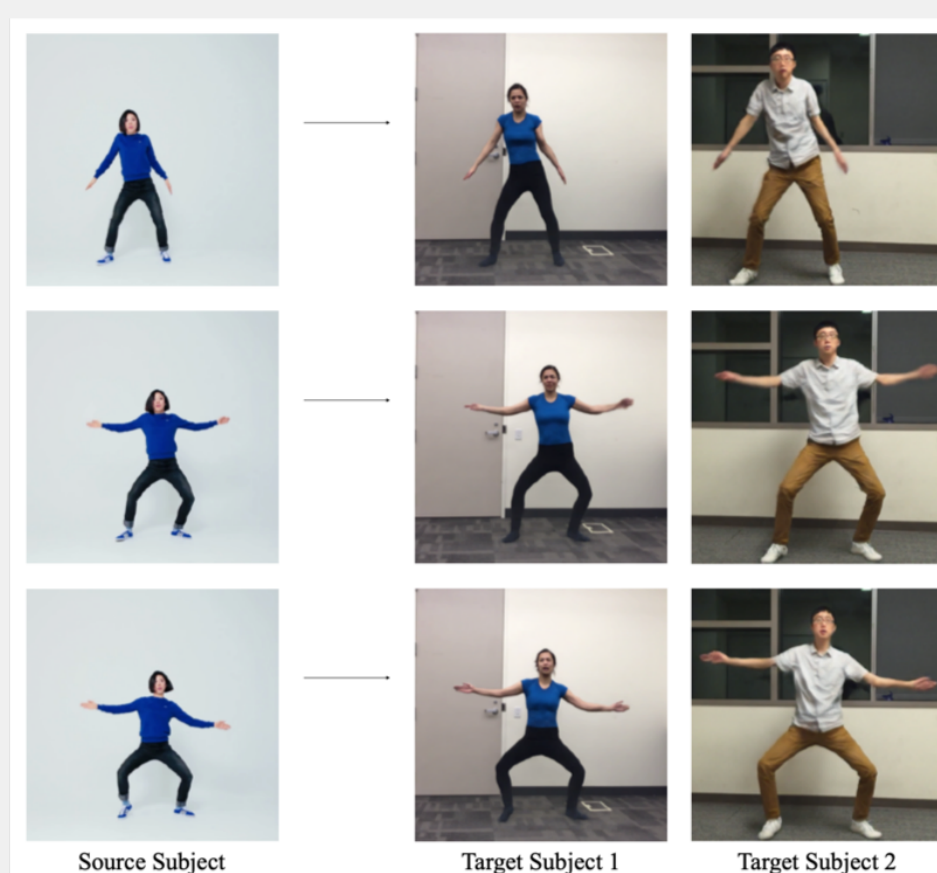
提案手法

- ▶ コンテキスト画像から人物モデルの抽象表現をアテンションを利用して作成
- ▶ ポーズ指定画像をクエリとして、人物モデルの抽象表現にアテンションをかけて、ポーズ付き抽象表現を作成
- ▶ ポーズ付き抽象表現から画像を描画
- ▶ ネットワーク全体をend-to-endで最適化。2つの手法をテスト
 - ▶ WGAN-GP: 正解画像と生成画像の分布のWasserstein距離 + VGG16の3層目出力のL1誤差を最小化
 - ▶ L1: 正解画像と生成画像のピクセルL1誤差を最小化



関連研究

- ▶ Everybody Dance Now [1]
 - ▶ 動画を入力として学習し任意の姿勢画像を生成
 - ▶ 人物ごとに再学習が必要
- ▶ Generative Query Network [2]
 - ▶ 同一のシーンを複数の視点から描画した少数の画像から任意の視点からのシーン画像を生成



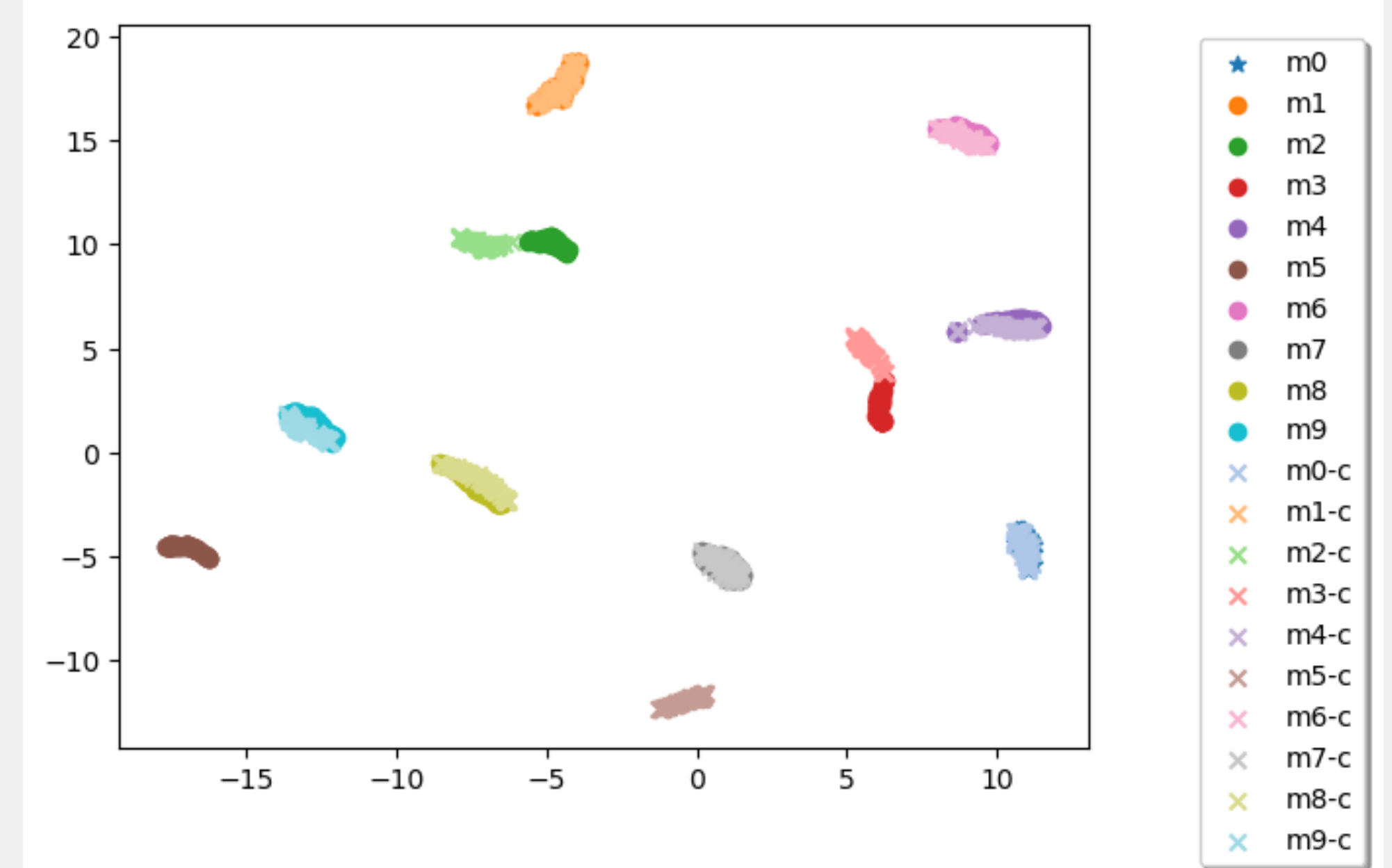
- ▶ この成果の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです
- ▶ 図中で使用したMMDモデルは下記の通りです。ありがとうございました
G-Tuneちゃん, Gちゃん (<https://www.g-tune.jp/campaign/10th/mmd>), あいえるたん (<http://www.infinitemloop.co.jp/special/iltan/3d-dl.php>), セーラー服さん (<https://bowlroll.net/file/3706>)
- ▶ 実装をお手伝いいただいた井上辰彦さんに感謝します

評価手法

- ▶ MMDの人物モデル222種、Vroidのモデル200種を使用
 - ▶ ポーズを与えてBlenderで描画してデータセットを作成
 - ▶ データ多様性のため、DeepFashionデータからポーズを抽出できたものを1割混入
- ▶ コンテキストは、人物描画像とポーズ描画像を8対で与える。ポーズはMMDポーズからランダムに選択
- ▶ 抽象表現の次元数は8x8x8x128x2

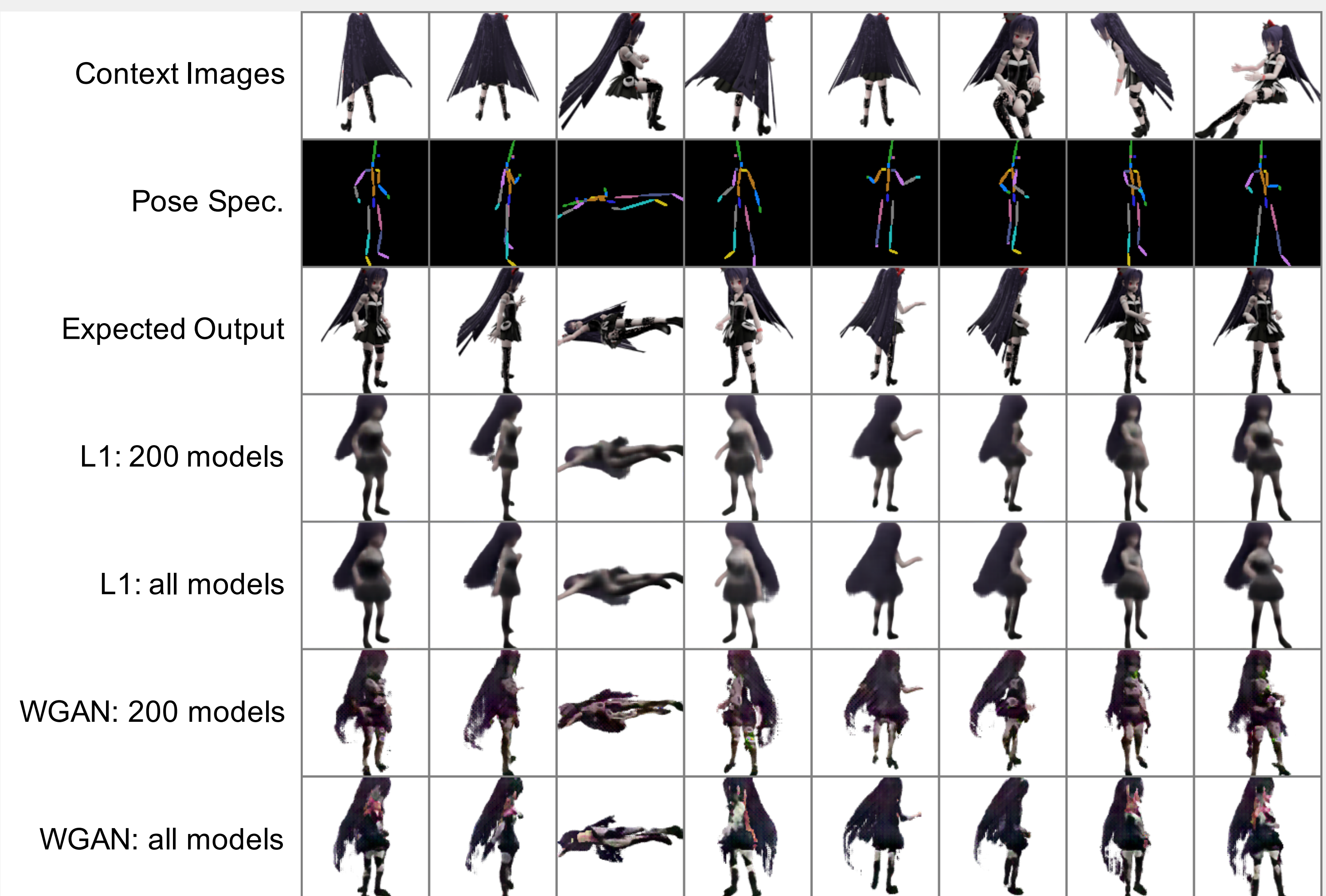
人物モデル抽象表現の評価

- ▶ 人物モデル抽象表現のT-SNEによる可視化
 - ▶ 1つの人物モデルに対して、多数のコンテキストセットを作成
 - ▶ コンテキストセットをエンコードして、抽象表現を作成
 - ▶ 人物の色相のみを変えたモデルについても同様に処理
- ▶ 人物モデル抽象表現のT-SNEによる可視化
 - ▶ 同じモデルはポーズによらず近傍にエンコードされることを確認
 - ▶ 色相のみが異なる場合も近傍にエンコードされている



生成画像

- ▶ 最適化手法：L1ロス vs. WGAN
- ▶ 学習データ：MMD200+DF vs. MMD222 + Vroid200+DF



- ▶ 評価
 - ▶ WGANのほうが一見きれい
 - ▶ 学習データ増大は効果あり
- ▶ 今後の課題
 - ▶ データの多様化 - 動画データからのデータ収集

[1] Chan, C., Ginosar, S., Zhou, T., and Efros, A. A. Everybody dance now. CoRR, 2018.

[2] Eslami, S. M. A., et al., Neural Scene Representation and Rendering Science, 360:1204–1210, 2018.