6Q-02

# スタイル変換を用いた多様な動作合成研究

深澤 康介 [1]　Hubert P. H. Shum[2]　森島 繁生 [3]
早稲田大学 [1]　Durham University[2]　早稲田大学理工学術院総合研究所 [3]

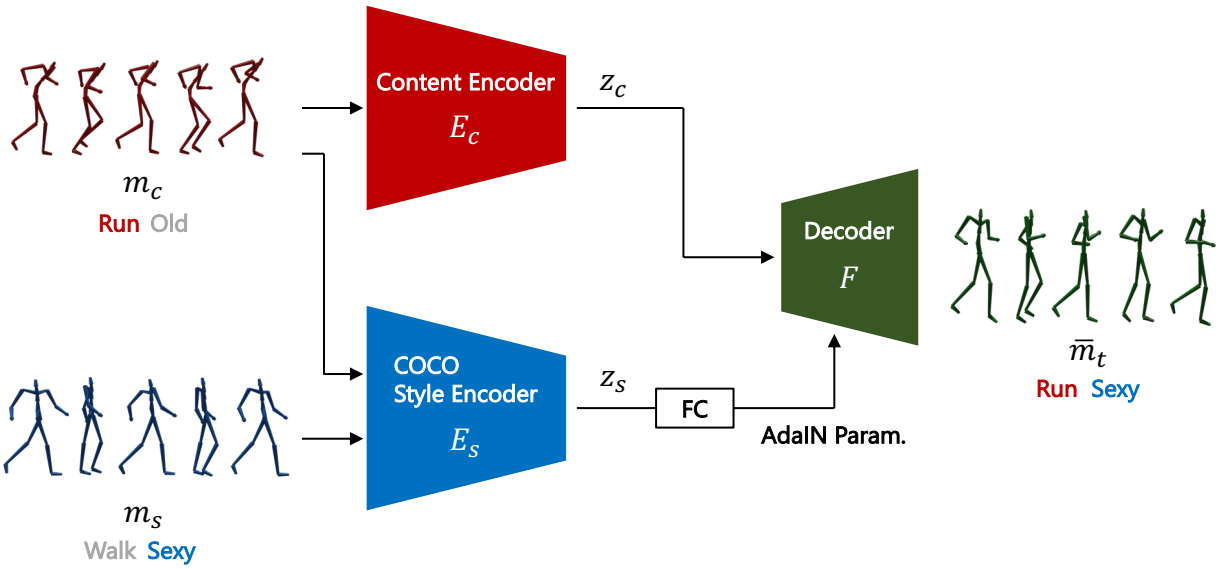


図 1: 提案手法の概要図

## 1. はじめに

昨今の AAA タイトルゲームの制作などでは，多種動作に対して多様な表現をもつアニメーション制作が求められている．キャラクターの歩行アニメーションだけでも，楽しそう，落ち込んでいる，など多様な表現が存在する．しかし，これらの表現をすべて記録するためには多大な制作労力を必要とする．

この制作労力を削減する手法として，動作の内容 (Content) を保ったまま，他の動作から動作の様子 (Style) を付与する Motion Style Transfer という技術が存在する．近年 Aberman ら [1] はペアになっていない動作間の Motion Style Transfer を実現した．しかし，彼らの手法では入力に用いる 2 つの動作が大きく離れている場合 (e.g., Punch と Kick など) に使用すると，合成される動作は壊れた内容になることが多い．この問題によって，1 つのスタイル動作から多種動作に対して Motion Style Transfer を使用することは不可能である．

本研究では，離れた動作間でも Motion Style Transfer を使用可能にすることで，多種動作にスタイル変換を施すことを目指す．これを実現するために Content-conditioned style encoder (COCO) [6] を導入し，コンテンツ入力動作に適したスタイル情報を用いることで，現実的な動作を合成する．

## 2. 関連研究

### Motion Style Transfer

Motion Style Transfer とは，一つの動作 (Content) に対して，別の動作 (Style) からスタイル情報を付与する研究である．Holden ら [2] は事前学習されたオートエンコーダーとグラム行列による最適化を用いて，深層学習ベースの手法を提案した．Aberman ら [1] は AdaIN を用いることで，ペアでない動作に対しても使用可能な手法を提案した．Park ら [5] は Aberman ら [1] の手法に加え，Graph Convolution を用いることで関節間の空間的接続を考慮し，離れた動作間でのスタイル変換を可能にした．しかし Park らの手法では，テスト時にもスタイルラベルを必要とする制約をもち，合成される動作にはしばしばノイズが見られる．

## 3. 提案手法

本手法では，スタイル変換の入力となる 2 つの動作 (Content, Style) が離れている場合でも，現実的な動作を合成するためのフレームワークを提案する．本手法の概形は図 1 に示す．スタイルの合成方法は Aberman ら [1] の AdaIN による手法を採用する．

入力動作のうち，コンテンツ動作 $m_c$ は Content encoder $E_c$ を通して，スタイルに依存しないコンテンツ特徴量 $z_c$ を抽出する (式 (1))．Content encoder 内部では 1 次元の時間方向に対する畳込み [3] が行われる．これによって時間の連続性が保たれたまま特徴を抽出する．スタイル動作 $m_s$ はコンテンツ動作 $m_c$ とともに Content conditioned style encoder (COCO) $E_s$ を通して，コンテンツ条件付きのスタイル特徴量 $z_s$ を抽出する (式 (2))．スタイル特徴量 $z_s$ は時間方向にプーリングされ，時間に依存しない特徴となる．コンテンツ情報を条件として抽出されるスタイル特徴量 $z_s$ は，コンテンツ情報を認識可能なため，コンテンツに適した形で出力される．この作用によって，2 つの入力動作が離れている場合であっても，スタイル情報を抽出可能にする．

$$z_c = E_c\left(m_c\right) \tag{1}$$

$$z_s = E_s\left(m_s, m_c\right) \tag{2}$$

抽出されたスタイル特徴量 $z_s$ は全結合層を通して AdaIN のパラメータへと変換され，コンテンツ特徴

Diverse motion generation using motion style transfer:
*Kosuke Fukazawa[1], Hubert P. H. Shum[2], and Shigeo Morishima[3] (1. Waseda University, 2. Durham University, 3. Waseda Research Institute for Science and Engineering)*



Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

量 $z_c$ とともに Decoder $F$ に入力される．Decoder $F$ では AdaIN によってスタイル変換が行われ，合成動作 $\bar{m}_t$ を得る (式 (3))．

$$\bar{m}_t = F\left(z_c, \mathrm{FC}(z_s)\right) \tag{3}$$

## 4. 実験

### 4.1 実験環境の詳細

本手法では，学習データ，テストデータはともに Xia ら [4] のスタイルタスク別動作データセットを使用し，入出力動作は Zhou ら [7] の 6 次元関節回転表現を用いた．学習では GeForce GTX 1080 Ti を用いて，PyTorch で実装を行い，8 時間の学習時間を必要とした．

### 4.2 実験概要

提案手法の有効性を示すために，大きく異なる動作間での Motion Style Transfer という条件のもと，既存手法 [1] [5] と比較を行う．入力動作は大きく離れた動作を想定して，図 2 を使用する．スタイル変換によって，Punch の動作が保たれたまま，Angry のスタイルが反映されることが期待される．

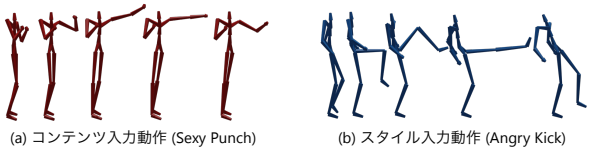


図 2: 実験に使用する入力動作

### 4.3 実験結果

図 2 を入力として，提案手法の動作合成プロセスで得られた結果と，既存手法との比較を図 3 に示す．

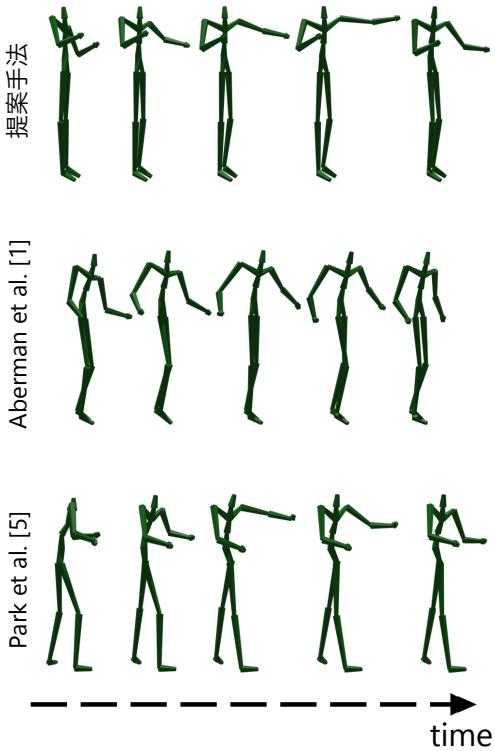


図 3: 提案手法と既存手法の比較

図 3 上段の提案手法では，コンテンツを保ちながら動作を合成することが可能である．また本手法では，図 1 に示すような近い動作間のスタイル変換にも適用可能である．中段の Aberman ら [1] の手法では，Kick の動作に影響される不自然な足の動きや，本来保持しておくべき Punch 動作が壊れている様子が見られる．下段の Park ら [5] の手法では本来ほとんど動作しない右手の不自然な動きや，ノイズの影響による膝の不自然な曲がり方が見られる．なお，Park ら [5] の手法のみ，実験時にスタイルラベルを与えている．

主観評価実験などを通してのスタイル変換精度の検証は，今後の課題とされたい．

## 5. おわりに

本研究では，離れた動作間でも使用可能な Motion Style Transfer の手法を提案し，その合成動作を示した．COCO を導入することによって，離れた動作間でもコンテンツ情報を保ったままスタイル変換が行われ，既存手法のように動作の内容が破壊されてしまうことを防ぐ．しかし，本手法の限界として，Aberman ら [1] が行っている動画を入力としてスタイル特徴量を得ることができず，これを達成する事ができなかった．また，一部の入力に対して，コンテンツ入力動作とほとんど変わらない動作を出力する例があった．

今後の課題として，様々な動作に適用した際のスタイル変換精度保証や，出力に確率的プロセスを用いることでいくつかの出力候補を合成することが挙げられる．

## 謝辞

この研究は、JST 未来社会創造事業 (JPMJMI19B2) および JSPS 科研費 (19H01129, 19H04137, 21H05054) の補助を受けた．

## 参考文献

[1] Kfir Aberman, Yijia Weng, Dani Lischinski, Daniel Cohen-Or, and Baoquan Chen. Unpaired motion style transfer from video to animation. *ACM Transactions on Graphics (TOG)*, Vol. 39, No. 4, p. 64, 2020.

[2] Daniel Holden, Jun Saito, and Taku Komura. A deep learning framework for character motion synthesis and editing. In *SIGGRAPH 2016*, 2016.

[3] Daniel Holden, Jun Saito, Taku Komura, and Thomas Joyce. Learning motion manifolds with convolutional autoencoders. In *SIGGRAPH Asia 2015 Technical Briefs*, 2015.

[4] Shi hong Xia, Congyi Wang, Jinxiang Chai, and Jessica K. Hodgins. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Transactions on Graphics (TOG)*, Vol. 34, pp. 1 – 10, 2015.

[5] Soomin Park, Deok-Kyeong Jang, and Sung-Hee Lee. Diverse motion stylization for multiple style domains via spatial-temporal graph-based generative model. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, Vol. 4, No. 3, pp. 1–17, 2021.

[6] Kuniaki Saito, Kate Saenko, and Ming-Yu Liu. Cocofunit: Few-shot unsupervised image translation with a content conditioned style encoder. *arXiv preprint arXiv:2007.07431*, 2020.

[7] Yi Zhou, Connelly Barnes, Lu Jingwan, Yang Jimei, and Li Hao. On the continuity of rotation representations in neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.



Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.