様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ５ 月 １３ 日現在

機関番号：３２６６３
研究種目：基盤研究(C)（一般）
研究期間：2019～2021
課題番号：１９Ｋ１２２８７

研究課題名（和文）深層ニューラルネットワークを用いた人物動作の多様体モデルの構築

研究課題名（英文）Human Motion Manifold using Deep Neural Networks

研究代表者
村上　真（Murakami, Makoto）

東洋大学・総合情報学部・准教授

研究者番号：８０３２９１１９

交付決定額（研究期間全体）：（直接経費）　　700,000円

研究成果の概要（和文）：本研究では，人間が動作を思い浮かべる動作生成過程と人間が動作を認識する動作推論過程は複雑で非線形だと考え，この両過程を深層ニューラルネットワークによりモデル化した．具体的には，Generative Adversarial NetworksとVariational AutoEncodersと呼ばれる2種類の異なる手法を用いて両過程をモデル化し，モーションキャプチャシステムにより収録した動作データを用いて提案モデルを学習した．学習済の生成モデルを使用することで自然で多様な動作が生成可能であることを示した．

研究成果の学術的意義や社会的意義
人間は様々な動作を思い浮かべることができる．このような人間の創造活動の一部を深層ニューラルネットワークによりモデル化できることを示したことが本研究の学術的意義である．また，3次元コンピュータグラフィックスを使用した映画やゲームには人型のキャラクタが登場することが多く，キャラクタの動作を生成・制御・編集することは重要なタスクである．本研究で提案した動作生成モデルによりキャラクタアニメーションの制作が容易になることが期待できる．

研究成果の概要（英文）：We consider that the process that people create various human motions in their minds and the process that people recognize various human motions are complicated and non-linear. And we modeled them using two different kinds of deep neural networks: generative adversarial networks and variational autoencoders. We trained the proposed models using human motion dataset captured with optical motion capture system. And we confirmed that the trained models can generate various natural human motions.

研究分野：知能情報学

キーワード：深層ニューラルネットワーク　生成モデル　GAN　VAE　キャラクタアニメーション　動作

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

3 次元コンピュータグラフィックスを使用した映画やゲームには，人型のキャラクタが登場し，人のように行動することが多い．自然な動作を生成する方法として，モーションキャプチャシステムで収録した実際の人間の動作をキャラクタに付与する方法が用いられているが，収録には手間がかかるため，これまでにコンピュータを使用して動作を制御する方法が数多く提案されてきた．近年，大量のモーションキャプチャデータから深層ニューラルネットワークを使用して学習を行い，キャラクタの動作を制御する方法が提案されている．一方，深層ニューラルネットワークを用いて例えば大量の画像データから学習を行い，多様で自然な画像を生成する生成モデルに関する研究が近年活発に行われている．

２．研究の目的

本研究では，動作データはいくつかの意味のある量から生成されていると考える．この低次元の量を潜在変数と呼び，低次元の潜在空間中で表現されている動作データを動作の多様体と呼ぶ．本研究では，動作は潜在空間に表現されている動作の多様体から生成され，この生成の過程は複雑で非線形であると考える．この逆の過程は観測された動作から意味のようなものを推論する過程となるが，この過程も複雑で非線形であると考える．本研究では，複雑で非線形な動作生成過程と動作推論過程をそれぞれ深層ニューラルネットワークとパラメトリックな確率分布を使用してモデル化し，モーションキャプチャシステムにより収録した動作データから学習を行うことで，直接観測できない動作多様体と動作生成過程及び動作推論過程を明らかにすることを目指す．また，学習済の動作生成モデルにより自然で多様な動作を生成することを目的とする．

３．研究の方法

本研究では Generative Adversarial Networks(GAN)を使用して動作の生成モデルを構築する．GAN は，多様で自然な動作データが生成できる生成モデルと，現実の動作データか生成された動作データかを識別する識別モデルから成り，生成モデルと識別モデルをそれぞれ深層ニューラルネットワークによって表現する．生成モデルは識別モデルが誤って識別するように，識別モデルは正しく識別できるように，学習データを使用してニューラルネットワークのパラメータを調整する．適切に学習が進めば，多様で自然な動作データが生成できるモデルが構築できる．
また，本研究では，Variational AutoEncoders(VAE)を使用して動作の生成モデルを構築する．VAE では，低次元の潜在空間からサンプリングされた潜在変数から非線形な確率過程を経て高次元の動作データが生成されると考える．また，動作データから潜在変数を推論する過程も非線形な確率過程であると考える．この非線形な生成過程と推論過程を深層ニューラルネットワークと正規分布のようなパラメトリックな確率分布を組み合わせて表現する．大量の動作データを用いて推論と生成を繰り返し，正しい動作データが復元されるようにニューラルネットワークのパラメータを推定する．学習が適切に進むと，生成モデルによって，多様で自然な動作が生成できるようになる．本研究では，Recurrent Neural Network(RNN)のような時間方向の依存関係を表現することのできるニューラルネットワークと VAE を組み合わせて動作生成モデルを構築する．

４．研究成果

(1) 動作データ

本研究では，光学式モーションキャプチャシステムで撮影された 2,505 個の人物動作データからなる CMU Graphics Lab Motion Capture Database[1]を使用した．元データのサンプリングレートは 120fps であるが，本研究では 30fps にダウンサンプリングしたものを使用する．元の動作データは，図 1 に示すような 19 関節の 3 軸中心の局所回転角度とルート関節（Hip）の大局平行移動量として表現されている．動作の見た目の近さを測るには，各関節の局所回転角度よりも局所的な位置のほうが適しているため，まず局所位置に変換する．具体的には，ルート関節（Hip）を床面に鉛直下向きに射影した点を原点とし，腰と左肩右肩の位置関係から体の正面方向を求め，右向きをx軸，鉛直上向きをy軸，正面方向をz軸とし，局所座標系を設定し，各関節の局所位置を求める．また，大局的な移動量はy軸中心の角速度とxz平面の速度として表現する．このようにして得られた動作データは各フレームに対して 58 次元ベクトルとなる．(2)の研究では，フレームサイズ 120（4 秒）のウィンドウを 60 フレーム（2 秒）ずつオーバーラップさせて動作を分割することで得られた 14,122 個の動作データを使用した．(3)の研究では，フレームサイズ 128（約 4 秒）のウィンドウを

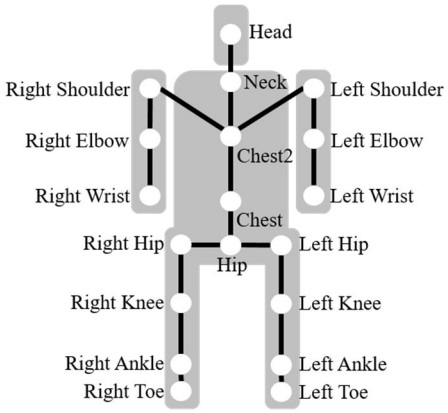

図 1 動作データの関節構造

64 フレーム（約 2 秒）ずつオーバーラップさせて動作を分割することで得られた 13,032 個の動作データを使用した．また，これらから平均を引き，標準偏差で割ることにより，データの標準化を行った．

(2) Wasserstein GAN を用いた動作生成モデルの構築

本研究ではまず Wasserstein GAN を使用して動作生成モデルを構築した．構築した GAN の識別器の構造を図 2 に示す．本研究で使用する動作データは(1)で述べたように，空間方向に 58 次元・時間方向に 120 次元のベクトルとなる．識別器の第 1 層では空間方向に 58・時間方向に 15 のサイズのフィルタを 64 種類用意し，空間方向のストライドを 58 とし，58 次元の関節位置を全て畳込んでいる．また，時間方向のストライドは 2 とし，時間方向の次元数を半分にしている．識別器の第 2 層では，時間方向に 15 のサイズのフィルタを 128 種類用意し，ストライドを 4 とし畳込みを行った．第 1 層・第 2 層ともに 25%の結合をドロップアウトし，活性化関数としてネガティブスロープが 0.2 である LeakyReLU を使用した．最後に，全結合層によりユニット数 1 のデータが出力される．

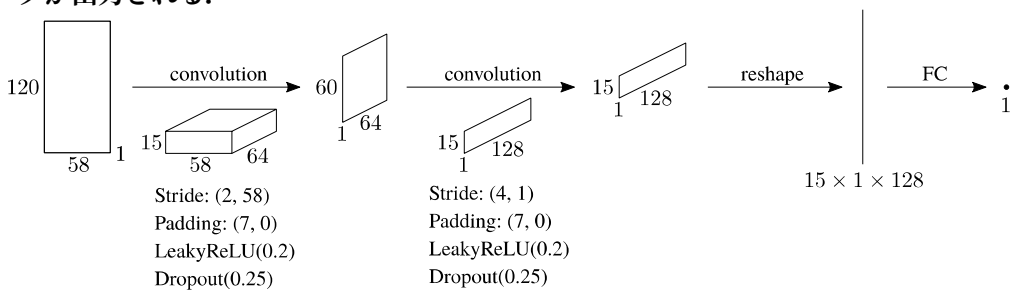

図 2 Wasserstein GAN の識別器のネットワーク構造

構築した GAN の生成器の構造を図 3 に示す．本研究では潜在空間を 512 次元とし，一様分布からサンプリングされた潜在変数から識別器と逆の変換を行うことで動作データを出力する．生成器の第 1 層は全結合層であり，512 次元の潜在変数から 15×1×128 次元の特徴量を復元する．第 2 層では，時間方向に 15 のサイズのフィルタを 64 種類用意し畳込みを行い，時間方向に 4 倍にアップサンプリングを行うことで 60×1×64 次元の特徴量を復元する．活性化関数にはネガティブスロープが 0.2 の LeakyReLU を使用し，Batch Normalization を行っている．第 3 層では，時間方向に 15 のサイズのフィルタを 58 種類用意し畳込みを行い，時間方向に 2 倍にアップサンプリングを行うことで時間方向に 120・空間方向に 58 次元の動作データを復元する．活性化関数には tanh を使用した．

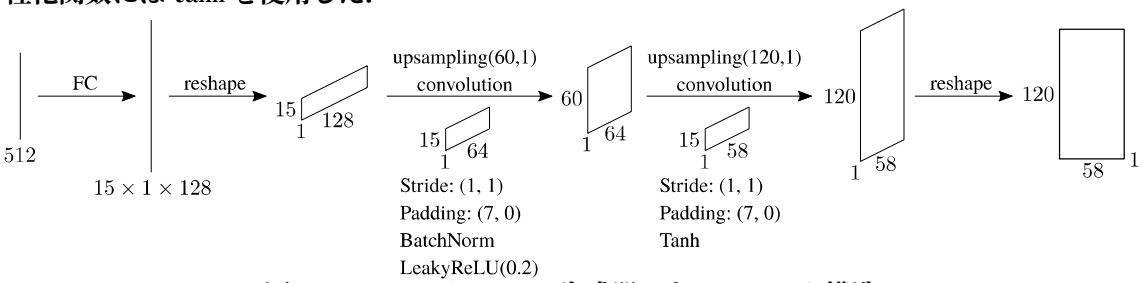

図 3 Wasserstein GAN の生成器のネットワーク構造

Goodfellow ら[2]が提案した GAN では損失関数として Logistic Loss が使用されていた．Logistic Loss には，生成器が同じようなデータを出力するように学習が進んでしまう mode collapse と呼ばれる問題があった．この問題は，識別境界から離れたデータでは損失関数の変動が小さくなり，勾配が消失することによって生じる．Logistic Loss を使用した学習では，実データの分布と生成データの分布の Jensen-Shannon divergence が小さくなるようにネットワークパラメータが最適化される．Arjovsky ら[3]は Jensen-Shannon divergence の代わりに Wasserstein divergence を損失関数とすることで mode collapse を回避する Wasserstein GAN（WGAN）を提案した．WGAN では Lipschitz 制約が必要となるため，その制約を満たすためにネットワークの重みの値を制限する weight clipping という手法が用いられる．しかし weight clipping によりモデルが簡易化されてしまうと，実データに近いデータが生成できなくなる問題がある．この問題を解決するために，Gulrajani ら[4]は gradient penalty を使用して Lipschitz 制約を満たす方法（WGAN-GP）を提案した．本研究では，WGAN-GP で使用される損失関数により，ネットワークパラメータの最適化を行った．

(1)で述べた動作データを使用し，動作生成モデルの学習を行った．具体的には，学習データ数を 12,710，評価用データ数を 1,412 として学習を行った．学習時のバッチサイズは 100，エポック数は 200 とした．また，最適化には Adam を使用した．学習済の生成器を使用してランダムに生成した 64 個の動作データを図 4 に示す．図 4 に示すように，提案モデルは自然で多様な動作データを生成できることが確認できた．

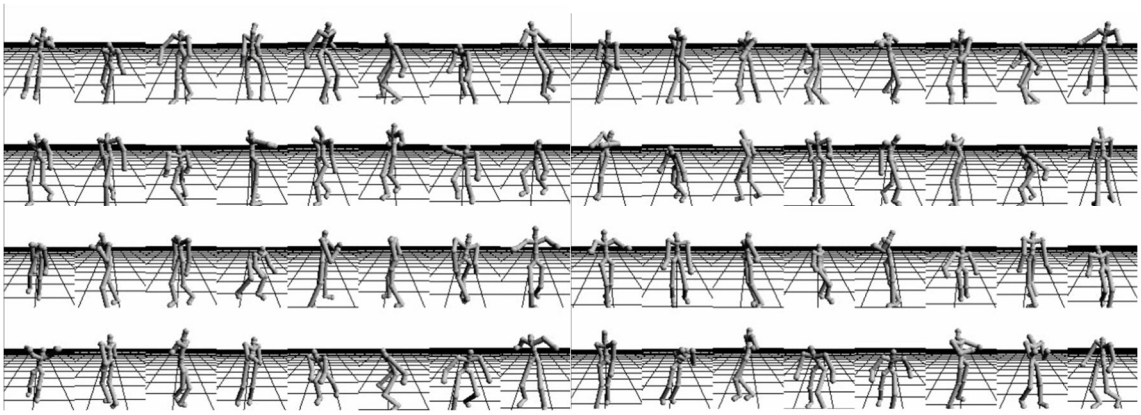

図 4 Wasserstein GAN の生成器により生成された動作データ

(3) Variational Recurrent Neural Network を用いた動作生成モデルの構築

本研究では，図 5 に示すように，多階層のニューラルネットワーク(Multi-Layer Networks: MLN)により動作特徴量を抽出し，RNN により過去の動作との依存関係を表現するとともに，VAE を使用して低次元の潜在空間に動作特徴を確率分布として表現することができるモデルを構築した．

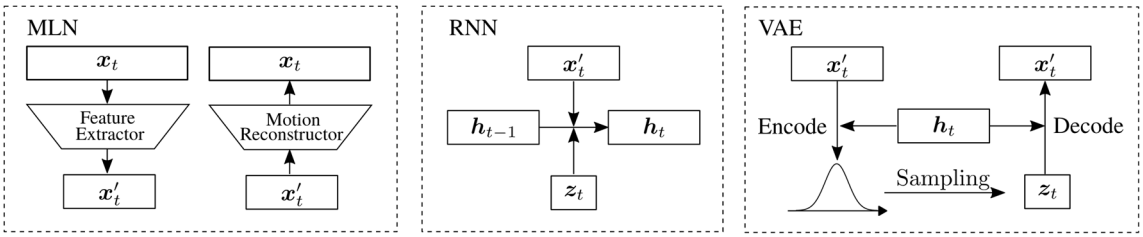

図 5 Variational Recurrent Neural Network を用いた動作生成モデルの概要

本研究で構築したモデルでは，58 次元の動作データ$\boldsymbol{x}_t$を 8 次元の潜在変数$\boldsymbol{z}_t$で表現する．$\boldsymbol{x}_t$を$\boldsymbol{z}_t$に変換するエンコーダネットワークの構造を図 6 に示す．提案モデルでは 2 層の特徴抽出器f_1, f_2により動作データから特徴抽出を行っている．f_1, f_2は全結合層(FC)と活性化関数 Rectified Linear Unit(ReLU)により構成されており，f_1, f_2によって抽出される特徴量の数はそれぞれ 56，32 とした．また，図 7 に示すように RNNρには Long Short-Term Memory(LSTM)を使用し，動作特徴量$\boldsymbol{x}''_t$と潜在変数$\boldsymbol{z}_t$から隠れ状態ベクトル$\boldsymbol{h}_t$を更新する．隠れ状態ベクトル$\boldsymbol{h}_t$の次元数は 32 とした．エンコーダネットワークは，図 6 に示すように動作特徴量$\boldsymbol{x}''_t$と隠れ状態ベクトル$\boldsymbol{h}_{t-1}$から 1 層の全結合層$\varphi_\mu, \varphi_\sigma$により正規分布の平均と標準偏差を出力することで，確率モデルとして表現されている．標準偏差を出力する層φ_σには sigmoid 関数をかけている．デコーダネットワークは図 8 に示すように 3 層の全結合層ψ, f'_1, f'_2により構成されている．1 層目のψが潜在変数$\boldsymbol{z}_t$から動作特徴量$\boldsymbol{x}''_t$への変換であり，残りの 2 層f'_1, f'_2が動作特徴量$\boldsymbol{x}''_t$から動作データ$\boldsymbol{x}_t$への変換となっている．ψ, f'_1では ReLU 活性化関数をかけている．事前確率ネットワークでは，図 9 に示すように 1 層の全結合層ϕ_μ, ϕ_σにより正規分布の平均と標準偏差を出力することで，事前確率分布を表現している．標準偏差を出力する層ϕ_σでは sigmoid 関数をかけている．

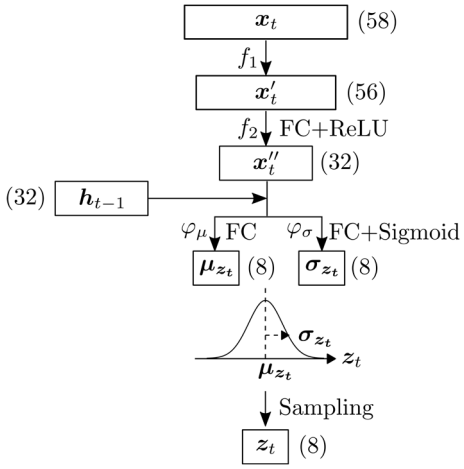

図 6 エンコーダネットワーク

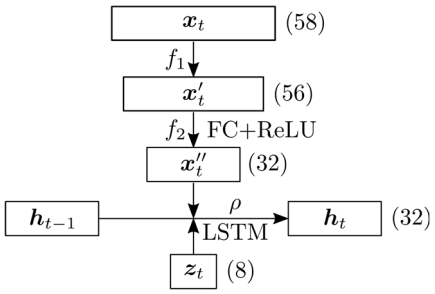

図 7 リカレントネットワーク

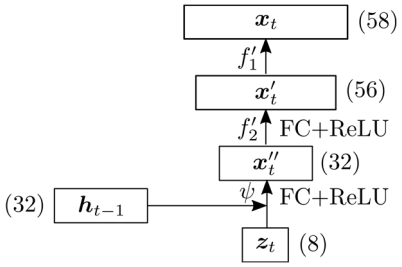

図 8 デコーダネットワーク

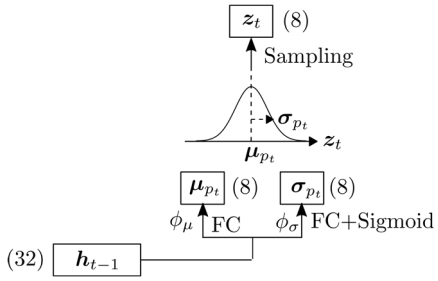

図 9 事前確率ネットワーク

(1)で述べた動作データを使用し，動作生成モデルの学習を行った．具体的には，学習データ数を 10,426, 評価用データ数とテスト用データ数をそれぞれ 1,303, バッチサイズを 1,600 として学習を行った．学習のエポック数は 2,000 とした．また，最適化には Adam を使用した．構築した動作生成モデルを評価するための実験を行った．具体的には，事前確率ネットワークにより推定された事前確率分布から潜在変数z_tをサンプリングし，それをデコーダネッワークによりデコードすることで，動作データx_tを生成した．ランダムにサンプリングした 36 個のz_0から生成した動作データを図 10 に示す．図 10 に示すように，提案モデルは多様な動作を生成することができ，各フレームの姿勢は WGAN を使用したモデルよりも自然であることが多かった．

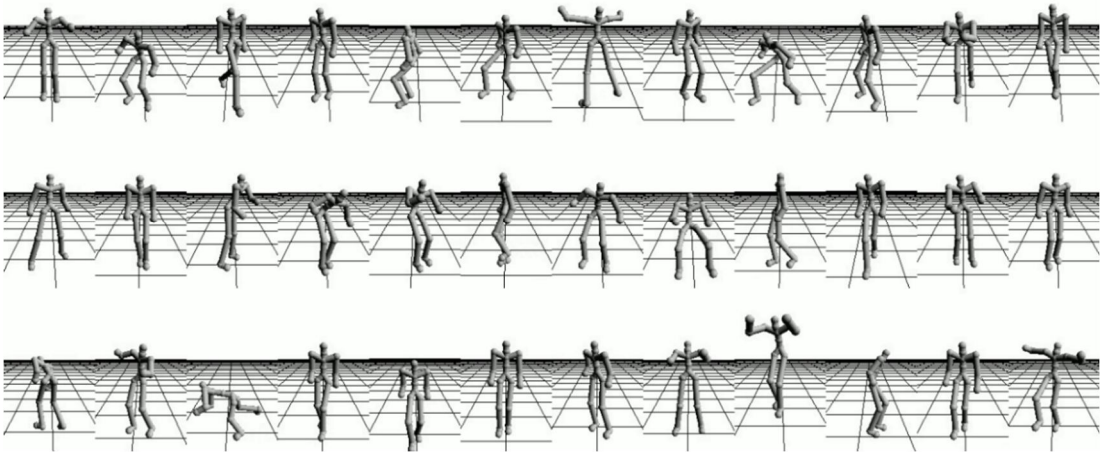

図 10 VRNN を用いた動作生成モデルにより生成された動作データ

(4) 得られた成果の位置づけと今後の展望

本研究では，主に画像生成に使用されてきた GAN や VAE を動作生成の問題に適用し，自然で多様な動作を生成することに成功した．特に，VAE を用いた生成モデルでは，RNN により過去の動作との依存関係を表現すると同時に，多階層のニューラルネットワークにより動作特徴量を抽出し，低次元の潜在空間に確率分布として動作特徴を表現することができた．

本研究で提案した動作生成モデルは，入力を伴わず，確率的に多様で自然な動作を生成することができるモデルである．一方，何かを入力すると動作が出力されるモデルも提案されており，このモデルを使用すると，例えば，床の上の移動の軌跡を指定することにより，その軌跡に沿って移動する動作を生成することができる[5]．しかし，これらのモデルではユーザが思い描いた通りの動作を全て生成できるとは限らない．今後は，本研究で構築した動作生成モデルの前段に，単語や単語列の分散表現に使用される深層ニューラルネットワークを配置することで，映画監督が俳優に言語で演技指導を行うように，ユーザの言語による指示によって意図した通りの動作が出力されるシステムを構築する予定である．

＜引用文献＞

[1] CMU: Carnegie Mellon University -CMU Graphics Lab- motion capture library, http://mocap.cs.cmu.edu.

[2] I. Goodfellow, et al.,“Generative Adversarial Nets,”Advances in Neural Information Processing Systems, 27, pp.2672-2680, 2014.

[3] M. Arjovsky, et al.,“Wasserstein GAN,”Proceedings of the 34th International Conference on Machine Learning, pp.214-223, 2017.

[4] I. Gulrajani, et al.,“Improved training of wasserstein GANs,”Advances in Neural Information Processing Systems, pp.5768-5778, 2017.

[5] D. Holden, et al.,“A Deep Learning Framework for Character Motion Synthesis and Editing,”ACM Transactions on Graphics, vol.35, issue 4, no.138, pp.1-11, 2016.

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計5件（うち招待講演　0件／うち国際学会　2件）

１．発表者名 Makoto Murakami, Takahiro Ikezawa
２．発表標題 Human Motion Generation using Variational Recurrent Neural Network
３．学会等名 6th International Conference on Digital Signal Processing（国際学会）
４．発表年 2022年

１．発表者名 Ayumi Shiobara,Makoto Murakami
２．発表標題 Human Motion Generation using Wasserstein GAN
３．学会等名 International Conference on Computer Graphics and Virtuality（国際学会）
４．発表年 2021年

１．発表者名 村上真,生澤隆広
２．発表標題 Variational Recurrent Neural Networkを用いた人物動作生成モデルの構築
３．学会等名 情報処理学会 コンピュータグラフィックスとビジュアル情報学研究会
４．発表年 2021年

１．発表者名 Ayumi Shiobara, Makoto Murakami
２．発表標題 Human Motion Generative Model using Wasserstein GAN
３．学会等名 情報処理学会 コンピュータグラフィックスとビジュアル情報学研究会
４．発表年 2020年

１．発表者名
塩原歩，村上真

２．発表標題
Generative Adversarial Networksを用いた人物動作生成モデルの構築

３．学会等名
電子情報通信学会 人工知能と知識処理研究会

４．発表年
2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

研究プロジェクト｜メディア情報研究室｜村上真研究室｜東洋大学総合情報学部
http://www.makotomurakami.com/projects.html

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関