

令和 5 年 5 月 25 日現在

機関番号：12501

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K12039

研究課題名（和文）動画理解のための深層状態空間モデリング法の展開

研究課題名（英文）Deep State Space Modeling Methods for Video Understanding

研究代表者

川本 一彦（Kawamoto, Kazuhiko）

千葉大学・大学院工学研究院・教授

研究者番号：30345376

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：深層学習と状態空間モデルの統合に基づく動画理解の課題に取り組みました。まず、深層マルコフモデルを導入し、カオスダイナミクスの予測、行動認識や動画生成のための深層モデルも開発した。次に、時系列と空間データを扱う2D畳み込みニューラルマルコフモデルを開発し、動画生成や行動認識の深層モデルを開発した。そして、動画生成の制御を可能にする深層モデルの構築に取り組み、ゼロショット画像生成を開発しました。さらに、動画像に含まれる静的特徴と動的特徴を分離する逐次変分オートエンコーダを開発した。以上の研究により、本研究の有効性を実証した。

研究成果の学術的意義や社会的意義

深層学習モデルと状態空間モデルの統合により、コンピュータビジョンにおける動画理解タスクを適切にモデル化でき、行動認識、人物追跡、動画生成といったタスクがより精度高く、効率的に行えるようになる。これは、監視システム、自動運転車、ロボティクスなどの分野に貢献できる。また、動画生成技術は、エンターテインメントや広告への応用も期待できる。

研究成果の概要（英文）：This study tackled video understanding based on integrating deep learning and state-space models. First, we introduced a deep Markov model for predicting chaotic dynamics. Next, we extend the deep Markov model to a 2D convolutional neural Markov model that handles both time series and spatial data. Furthermore, we developed deep models for video generation and action recognition. Then, we worked on building a deep model that enables control of video generation and developed zero-shot image generation. Furthermore, we developed a sequential variational autoencoder that separates static and dynamic features in video images. These studies demonstrated the effectiveness of our approach.

研究分野：コンピュータビジョン

キーワード：状態空間モデル 深層学習

## 1. 研究開始当初の背景

深層学習モデルは、コンピュータビジョンやパターン認識の分野で驚異的な発展を遂げており、画像分類や物体検出などに対して、実際に解を与える工学的なアプローチを提供したことに大きな貢献がある。深層学習を活用することにより、データ駆動型サイエンスの適用可能性を大きく広げることが可能になり、その適用範囲をさらに拡大するための取り組みが求められている。一方、状態空間モデルは、時系列解析やシステム制御・同定の分野で長く応用されており、特に非線形・非ガウス型の状態空間モデルを扱えるようになっている。深層学習モデルと状態空間モデルを統合することにより、とくに時系列データを扱う動画理解タスクに威力を発揮することが期待される。現在のところ、深層学習と状態空間モデルの統合に関する研究は限定的である。コンピュータビジョン分野では、系列データを扱うための深層学習モデルとして再帰型ニューラルネットワークが存在するが、これは隠れ変数の時間ダイナミクスを表現しているわけではない。状態空間モデルと深層学習モデルの統合はまだ十分に確立されておらず、モデル統合のためのアーキテクチャ設計や学習アルゴリズムの開発が求められている。

## 2. 研究の目的

本研究の目的は、状態空間モデルと深層学習モデルを統合し、コンピュータビジョンの動画理解分野へ応用することである。状態空間モデルと深層学習モデルを結合することで、機械学習モデルの表現能力が高まり、柔軟なモデリングが可能になる。状態空間モデリングでは、マルコフ性の導入を中心とした時系列モデリングのノウハウが蓄積されており、効果的な方法を見通しよく開発できる。さらに、近年の逐次モンテカルロ法の発展により、非線形・非ガウス型の状態空間モデルが現実的に扱えるようになっている。これにより、モデリングの自由度が高まっている。状態空間モデルと深層学習モデルの統合は、演繹的なアプローチと帰納的なアプローチの統合という側面を持ち、時系列データ解析における強力なモデリング技法を提供できるということである。

## 3. 研究の方法

次の2つの動画理解タスクに対して、評価実験を通して有効性を検証する。以下の各タスクにおいて、共有できる要素技術を利用しながら、平行してモデル構築とアルゴリズム開発を進める。

**動画予測タスク**：過去のフレームから未来のフレームを予測するタスクである。具体的には、一連の連続した画像フレーム(動画)が与えられ、その動画の次のフレームを予測することが目標となる。このタスクは、自然な動きのパターンを理解し、それを未来に展開する能力を機械学習モデルに要求する。動画予測タスクは、自動運転車の未来のシーン予測、医療画像解析での病状進行予測、気象予報など、多くの実用的な応用が考えられる。また、動画予測タスクは、時間的な依存関係を捉える能力と、視覚的な情報を理解し処理する能力を組み合わせることを必要とするため、機械学習の研究において重要なタスクとなっている。

**動画生成タスク**：一連の静止画像(フレーム)を生成し、それらを時間的な順序に従って連結することで動画を作成するというタスクである。既存の動画や画像から学習し、それらに類似した新しい動画を生成する。新しい映像コンテンツの生成、予測とシミュレーション、欠損データの補完などに応用できる。動画生成タスクは、単純な画像生成タスクよりも一般に困難である。なぜなら、動画は時間的な連続性という追加の次元を持っており、それぞれのフレームが前後のフレームと一貫性を持つ必要があるからである。

## 4. 研究成果

動画予測タスクに関する研究成果は次の通りである。

**Deep Markov Models for Data Assimilation in Chaotic Dynamical Systems**: この研究では、深層マルコフモデルとその推論ネットワークの適用可能性をカオスな動的システムについて評価した。結果として、目標モデルの一部についての情報が既知である場合、モデルはプロセスノイズと観測ノイズの存在下でも、スムーズ化されたカルマンフィルタと同等の性能を示した。動画予測タスクに向けての予備的な成果である。

**2D Convolutional Neural Markov Models for Spatiotemporal Sequence Forecasting**: この研

究では、2D 畳み込みニューラルネットワークを含む深層マルコフモデルを用いて、目標系列の空間的側面を保持するモデルを提案した。この方法は、大きな分散を持つデータに対してロバストであり、合成データを使用した場合、長期予測期間において DNN モデルを上回る性能を示した。しかし、実世界の降水データの予測におけるモデルの制限も指摘し、これらの制限を解決するための可能なアプローチについても言及した。

Spatiotemporal Transformer for 2D Time Series Sequence Forecasting: この研究では、空間的および時間的特徴をモデル化する新たなトランスフォーマーを提案した。このモデルは、自己注意メカニズムを利用して、タイムステップに関する 2D 特徴を自己および他のタイムステップ特徴に関してモデル化する。実験では、このモデルは標準的なトランスフォーマー、LSTM、およびその seq2seq バリエーションを上回る改善を示した。また、点予測と分位数回帰の設定の両方で畳み込み LSTM モデルと比較可能な性能を示した。

動画生成タスクに関する研究成果は次の通りである。

DCVGAN: Depth Conditional Video Generation : この研究では、カラービデオだけでなく、深度ビデオを使用してビデオ生成のための GAN アーキテクチャを提案している。このアーキテクチャの生成器では、最初に深度ビデオが生成され、次に深度から色へのドメイン変換を解くことでカラービデオが生成される。深度情報に焦点を当ててシーンのダイナミクスをモデリングすることで、従来の方法よりも高品質なビデオを生成することができた。

Conditional MoCoGAN for Zero-Shot Video Generation : この研究では、ゼロショットビデオ生成のための条件付き GAN モデルを提案している。すなわち、訓練サンプルから見たことのないビデオを生成する。この目的を達成するために、GAN の潜在空間で表現を分離することを学習する。このモデルは、より良い分離表現を見つけ、良質なビデオを生成するために構築されている。

Animating Cloud Images with Flow Style Transfer : この研究では、GAN を使用して静止画像をアニメーション化する方法を提案している。ソース画像と動き転写用のビデオを入力として、ソース画像が転写用ビデオの動きに従ってアニメーション化されるビデオを生成する。この方法は、キーポイントが明示的に決定できないビデオ（例えば、動く雲の画像）のアニメーション化に対して、既存の方法よりも効果的であることを実験的に示している。

Audio-Visual Model for Generating Eating Sounds Using Food ASMR Videos : この研究では、無音の食事ビデオから食物のテクスチャの音を生成するための視聴覚モデルを提案している。このモデルは、検出された顔の視覚特徴を入力として取り、視覚ストリームに合わせた振幅スペクトログラムを出力する。このモデルの訓練のために、いくつかの食物 ASMR ビデオを含むデータセットを作成した。

Sequential Variational Autoencoder with Adversarial Classifier for Video Disentanglement : この研究では、ビデオの解離に使用できる表現学習方法として、順序変分オートエンコーダを提案している。この方法は、ビデオから静的な特徴と動的な特徴を別々に抽出することができる。他の順序変分オートエンコーダとの比較を通じて、提案方法の有効性を定性的および定量的に示している。

## 5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 0件/うちオープンアクセス 4件）

1. 著者名 Haga Takeshi, Kera Hiroshi, Kawamoto Kazuhiko	4. 巻 23
2. 論文標題 Sequential Variational Autoencoder with Adversarial Classifier for Video Disentanglement	5. 発行年 2023年
3. 雑誌名 Sensors	6. 最初と最後の頁 2515 ~ 2515
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/s23052515	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Wu Nan, Kawamoto Kazuhiko	4. 巻 21
2. 論文標題 Zero-Shot Action Recognition with Three-Stream Graph Convolutional Networks	5. 発行年 2021年
3. 雑誌名 Sensors	6. 最初と最後の頁 3793 ~ 3793
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/s21113793	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Uchiyama Kodai and Kawamoto Kazuhiko	4. 巻 9
2. 論文標題 Audio-Visual Model for Generating Eating Sounds Using Food ASMR Videos	5. 発行年 2021年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 50106 ~ 50111
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ACCESS.2021.3069267	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Kurisaki Kazuma and Kawamoto Kazuhiko	4. 巻 9
2. 論文標題 Animating Cloud Images With Flow Style Transfer	5. 発行年 2021年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 3269 ~ 3277
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ACCESS.2020.3048160	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Halim Calvin Janitra, Kawamoto Kazuhiko	4. 巻 -
2. 論文標題 Deep Markov Models for Data Assimilation in Chaotic Dynamical Systems	5. 発行年 2020年
3. 雑誌名 Advances in Artificial Intelligence	6. 最初と最後の頁 37 ~ 44
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-39878-1_4	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Nakahira Yuki, Kawamoto Kazuhiko	4. 巻 -
2. 論文標題 DCVGAN: Depth Conditional Video Generation	5. 発行年 2019年
3. 雑誌名 IEEE International Conference on Image Processing	6. 最初と最後の頁 749 ~ 753
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICIP.2019.8803764	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計10件 (うち招待講演 0件 / うち国際学会 1件)

1. 発表者名 芳賀壮, 計良宥志, 川本一彦
2. 発表標題 動画像の解きほぐしに向けた敵対的補助分類器の効果検証
3. 学会等名 電子情報通信学会パターン認識・メディア理解研究会
4. 発表年 2022年

1. 発表者名 Shun Kimura and Kazuhiko Kawamoto
2. 発表標題 Conditional Motion and Content Decomposed GAN for Zero-Short Video Generation
3. 学会等名 the 7th International Workshop on Advanced Computational Intelligence and Intelligent Informatics (国際学会)
4. 発表年 2021年

1. 発表者名 芳賀壮, 計良宥志, 川本一彦
2. 発表標題 変分オートエンコーダを用いた動画像の解きほぐし
3. 学会等名 第24回画像の認識・理解シンポジウム
4. 発表年 2021年

1. 発表者名 木内拓実, 計良宥志, 川本一彦
2. 発表標題 メイン適応を用いた一人称行動認識
3. 学会等名 第24回画像の認識・理解シンポジウム
4. 発表年 2021年

1. 発表者名 内山光大, 川本一彦
2. 発表標題 Food ASMR動画を用いたマルチモーダル深層学習による食感音の生成
3. 学会等名 第23回画像の認識・理解シンポジウム
4. 発表年 2020年

1. 発表者名 木村駿, 川本一彦
2. 発表標題 ゼロショット動画生成のための条件付きMoCoGAN
3. 学会等名 第34回人工知能学会全国大会
4. 発表年 2020年

1. 発表者名 小林 春平, 川本 一彦
2. 発表標題 Shared 2D-convolutions netを用いた深層マルチタスク学習による一人称行動認識
3. 学会等名 第22回画像の認識・理解シンポジウム講演
4. 発表年 2019年

1. 発表者名 中平有樹, 川本一彦
2. 発表標題 デプスからカラーへのドメイン変換を用いたGANによる動画生成
3. 学会等名 計測自動制御学会システム・情報部門学術講演会
4. 発表年 2019年

1. 発表者名 白濱 淳也, 川本 一彦
2. 発表標題 Residual Networksに対する確率的正則化の提案: Shake-ResDropとShake-SENet
3. 学会等名 第33回人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 栗崎一真, 川本一彦
2. 発表標題 ダイナミックスタイル変換を用いた自然風景画像の動画化
3. 学会等名 情報処理学会研究報告
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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
---------	---------