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　６ 年　 ５ 月　 ８ 日現在

機関番号：３３９１０
研究種目：若手研究
研究期間：2021～2023
課題番号：２１Ｋ１７７７２

研究課題名（和文）力学系カオスに基づく時間的特徴抽出法の開発と動画像認識への応用

研究課題名（英文）Development of temporal feature extraction method based on dynamical chaos and its application to video recognition

研究代表者

平川　翼（Hirakawa, Tsubasa）

中部大学・AI数理データサイエンスセンター・講師

研究者番号：６０８４６６９０

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：本研究課題では，深層学習モデル，とりわけ動画像データのような時間的な情報遷移が必要となる深層学習モデルに対して重要となる特徴抽出ないしは特徴抽出を行うパラメータの抽出を行う手法を提案した．具体的には，従来から系列データに対して深層学習モデルで広く用いられていたLong Short-Term Memory (LSTM) および近年高い認識精度を達成し広く用いられているTransformerおよびVision Transformer (ViT) に対する有効な特徴抽出のための枝刈り手法を提案した．

研究成果の学術的意義や社会的意義
本プロジェクトにおいて開発した枝刈り技術は，深層学習モデル内の冗長なパラメータを削除することで，近年，大規模化するネットワークモデルをコンパクト化・省電力化することが可能な技術である．そのため，高性能な画像認識モデルを大規模な計算機を用いることなく様々な画像認識データに対して適用することが可能となる．

研究成果の概要（英文）：In this research project, we proposed a method for extracting important features or parameters for deep learning models, especially for deep learning models that require temporal information transitions, such as video data. Specifically, we proposed an effective feature extraction method for Long Short-Term Memory (LSTM), which has been widely used in deep learning models for series data, and for Transformer and Vision Transformer (ViT), which have achieved high recognition accuracy in recent years and are widely used. We proposed a branch-and-branch pruning method for feature extraction.

研究分野：コンピュータビジョン

キーワード：深層学習　Transformer　Network Pruning　大規模事前学習モデル　基盤モデル

１．研究開始当初の背景

（1）深層学習の発達により，画像認識精度が飛躍的に向上している．画像中の物体クラスを識別する一般物体認識や画像中の物体の位置とクラスを識別する物体検出，画像中の画素毎に物体クラスを識別するセマンティックセグメンテーション等の静止画像に対する認識問題では，既に高い精度を達成しており，すでに自動運転などへの産業応用が進められている．コンピュータビジョン分野では，より挑戦的なデータや問題へ取り組む動きがあり，その一つに動画像を扱う研究が存在する．具体的には，動画クリップからクラス推定を行う動画像認識や未来における対象の歩行者の位置を予測する経路予測などが取り組まれている．また，長尺の動画像を短くまとめる動画要約，与えられた動画像を文章で説明するキャプション生成等のより複雑な問題も存在する．

（2）これらの問題で扱われる動画像データでは静止画像の認識における空間的な特徴抽出に加えて，時間的な特徴を捉える必要がある．動画像の各フレームに相当する空間的な特徴抽出は，近年の CNN の技術的・理論的な進展により，有用な特徴抽出が実現されている．一方，時間 的な特徴抽出に関しては，再帰型ニューラルネットワーク（RNN）[Mikolov+, INTERSPEECH2010] や畳み込みに基づく処理である時間方向畳み込み [Bai+, arXiv2018] や 3 次元畳み込み（3D-CNN）[Hara+, CVPR2018] などの手法が広く用いられている(図 1).

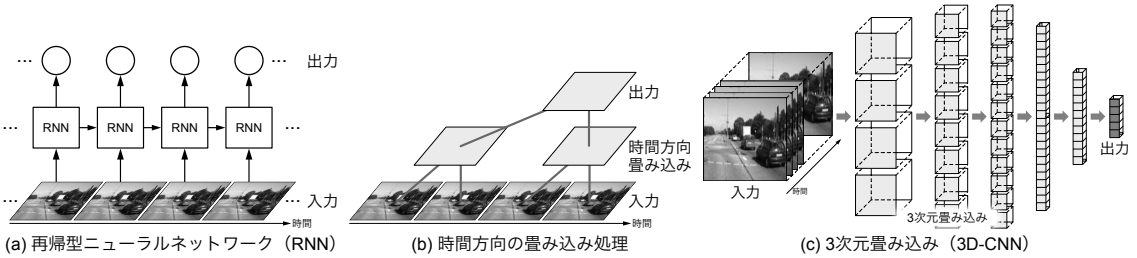


図 1. 深層学習による時間的な特徴抽出手法

２．研究の目的

（1）本研究課題では，深層学習に基づく動画像認識において，動画像データの時間遷移に対する特徴抽出のための深層学習モデルの開発を目的とする．具体的には，RNN のような再帰的な構造を持つネットワークの出力が力学系カオスの特性を有するような構造および学習方法を提案することで，空間および時間の双方に有用な特徴抽出を可能とする手法を開発し，動画像認識へと応用する．

（2）本研究課題の遂行期間中に Transformer [Vaswani+, arXiv2017] を画像認識に応用した Vision Transformer [Dosovitskiy+, ICLR2021] が発表され，現在では様々な画像認識タスクに用いられている．そのため，Vision Transformer に対する有効な特徴抽出を可能とする手法についても開発する．

３．研究の方法

（1）従来から時系列データ等に用いられている再帰型ニューラルネットワーク（RNN）に対する有効な特徴抽出法の開発（FY2021）

A)　小規模動画像データセットの作成
　大規模かつ複雑な動作クラスが設定されている既存の動画像認識データセットでは，ネットワークを十分に解析，理解することが難しい．そのため，動画像認識におけるネットワークの挙動解析を想定し，小規模な動画像認識用データセットを作成し，解析に用いる．

B)　移動エントロピーを用いた時間変化に有効な特徴抽出法の開発
　時系列データにおける情報量の尺度である移動エントロピーを活用し，各ニューロン(重みパラメータ)の時間遷移に対する重要度を判定する方法を提案する．具体的には再帰型ニューラルネットワークの一種である Long Short-Term Memory（LSTM）

の各時刻の特徴量と正解クラスの分布を用いて移動エントロピーを求める．各ニューロンの移動エントロピーを解析することにより，時系列データを考慮した特徴抽出に有用なニューロンを選択する手法を提案する．

（2）大規模事前学習モデルに対する枝刈り手法の開発（FY2022 ~ FY2023）

ここでは，近年発表された Vision Transformer (ViT) を対象として有効な特徴抽出法の開発を行う．大規模な事前学習を行った ViT は，様々なタスクで高い性能を発揮する一方で，そのモデルサイズの大きさから，特定の認識タスクである下流タスクへ再学習する際に多大な計算コストを要するという問題点がある．そのため，下流タスクに対して追加学習（ファインチューニング）を行う際に有用となる特徴量を明らかにする．具体的には，ファインチューニング前後のネットワークパラメータの変化を観測し，その傾向を調査する．その傾向を踏まえ，有効な特徴抽出を可能とする手法を提案する．

４．研究成果

（1）小規模動画像データセットの作成

前述のように，動画像認識のデータセットに関しては，大規模なデータセットのみであり，小規模なデータセットが存在しない．そこで，ネットワークの解析に有効な比較的小規模なデータセットである Action MNIST Dataset と UIUC Video Dataset を構築した（図 2）. Action MNIST Dataset は，手書き数字が動画フレーム内で移動する方向を分類するデータセットであり，UIUC Video Dataset は静止画像データセットの UIUC Dataset のテクスチャ画像を使用し，その動画フレームの動きを分類するデータセットである．

これらのデータセットを用いて時系列分類タスクにおける有効な特徴抽出の解析を行う．

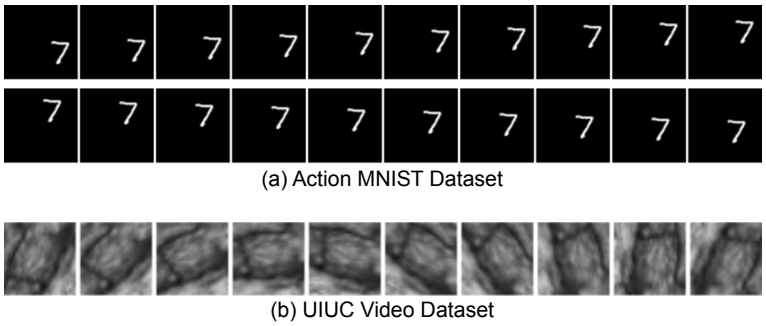


図 2　作成した動画像分類データセットのスナップショット

（2）移動エントロピーを用いた LSTM の枝刈り手法の提案

Recurrent Neural Network (RNN) は，再帰的構造によって系列性を学習する．RNN はその構造の複雑性から多くのパラメータを持つため，枝刈りによってモデルサイズを圧縮する必要がある．一方，RNN は順方向に情報を伝播するため，系列の長さに比例して枝刈りの影響も強くなる．そこで，本研究では，RNN モデルに適した移動エントロピーによる枝刈り手法を提案する．移動エントロピーは，情報理論の観点から 2 つの系列データの間の因果性を定量化する指標である．これは，2 つの確率変数感の依存度を定量化する相互情報量に，時間の概念を付与したものである．移動エントロピーにより，正解ラベルと各ユニットの因果性を定量化することで，従来手法よりも高い精度で，認識精度へ強く貢献しているユニットを識別する．

移動エントロピーを用いて RNN の各ユニットを評価する．図 3 に示すように，RNN のユニットの隠れ状態ベクトルを X，正解ラベルの出力を Y として，正解ラベルが各ユニットに与える因果性を定量化する．

図 4 に実験結果を示す．RNN モデルとして，Long Short-Term Memory (LSTM) を使用する．図 4(a)に LSTM 層（ニューロン数 : 128）の各ニューロンの相互情報量を x 軸，移動エントロピーを y 軸とし，特定のニューロンを削減した際の精度のカラーマップを示す．ここで，他のニューロンを削減した際と比べ，精度が大幅に低下するニューロン n1, n2 に着目すると n1 は相互情報量が低く，n2 は高い．一方で移動エントロピーは共に低い．したがって，LSTM では相互情報量よりも移動エントロピーの方が認識結果への寄与率が高いことがわかる．次に，図 4(b)に LSTM 層のニューロンの内，相互情報量と移動エントロピーの値の高いニューロンを削減した場合 (Low) と，低いニューロンを削除した場合 (High) の認識精度の比較を示す．共に，High と比べ，Low の精度が安定していることがわかる．これにより LSTM では情報量が低いニューロンほど重要であると言える．これは LSTM が時系列性を上手く考慮し，正解ラベルと

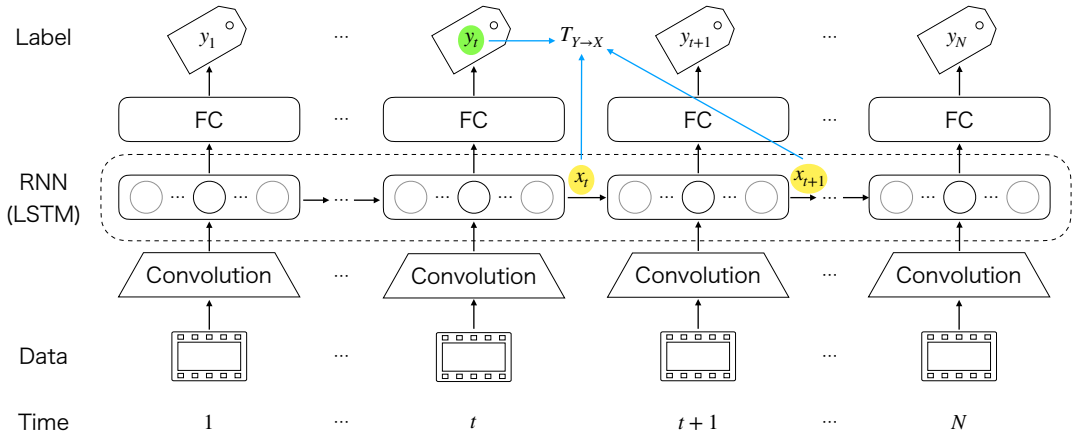
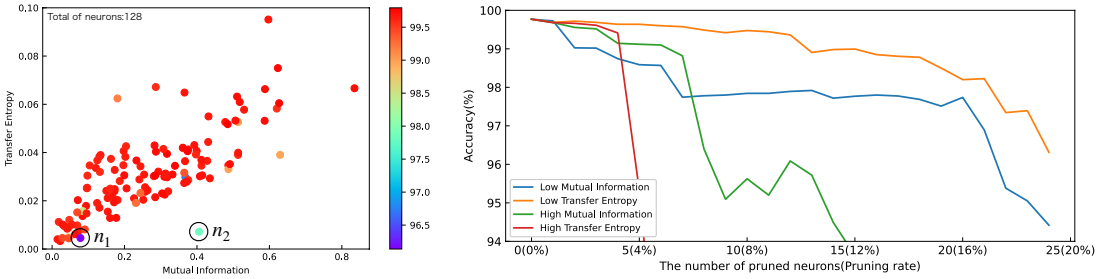


図 3. 再帰型ニューラルネットワークに対する移動エントロピーの算出



(a) LSTMの各ニューロンの
相互情報量と移動エントロピー

(b) LSTMの枝刈りによる精度の比較

図 4. LSTM に対する移動エントロピーによる枝刈りの結果

の曖昧さの少ないニューロンが重要なためだと考えられる．また，相互情報量よりも移動エントロピーを用いた方が高い認識精度を維持している．これにより LSTM では相互情報量では得られない重要度を移動エントロピーが獲得していると言える．

(3) 事前学習を考慮した非構造枝刈り手法の提案

大規模な事前学習を行った Transformer モデルは，様々なタスクで高い性能を発揮する一方で，そのモデルサイズの大きさから下流タスクへ再学習する際に多大な計算コストを要する．本研究では，大規模なデータセットを用いて事前学習したモデルを下流タスクへ最適化する際，パラメータの値がほぼ変化しない傾向があることを予備実験により示し，その傾向を考慮して，パラメータの値の大きさを直接評価する事前学習済みモデルに対するシングルショット非構造枝刈り手法を提案する．

図 5 に ViT-B/16，及び Mixer-B/16 の学習結果を示す．図 5(a)，(d) より，ImageNet-21k のような大規模データセットを用いて事前学習を行った場合，再学習の前後において，パラメータの値はほぼ変化していない．一方，図 5(b)，(e)，(c)，(f) より，事前学習に使用したデータセットの規模に比例して，再学習の前後でパラメータの値は不規則に変化している．これらの結果は，ImageNet-21k の事前学習によって得られた知識が，CIFAR-10 のような下流タスクで十分に活用されていることを意味する．

この傾向を踏まえ，事前学習を考慮した枝刈り手法を提案する．大規模なデータセットにより，十分な事前学習が行われていた場合または高次元なパラメータ空間を持つネットワークを用いている場合において，その勾配は極端に小さくなる．従来の勾配ベースの枝刈り手法は，このようなパラメータがほぼ変化しない状況において，その効果を十分に発揮することができない．そこで，本研究では勾配ベースの枝刈り手法に対してマグニチュードベースの枝刈り手法を導入することで事前学習済みモデルに適したパラメータの評価方法を提案する．

$$S(\theta_q) = \left| \frac{\partial \mathcal{L}}{\partial \theta_q} \theta_q \right| + \alpha \theta_q^2$$

これにより，下流タスクに最適化されるパラメータと，事前学習によって得られたほぼ変化しないパラメータの両方を評価することが可能となる．

提案手法の有効性を検証するために，様々な Transformer モデルを用いて評価実験を行う．本実験では，CIFAR-10, CIFAR-100，および ImageNet-1K を用いた分類精度の比較によって提案手法の有効性を検証する．事前学習には ImageNet-21K を使用する．表 1 に Transformer モ

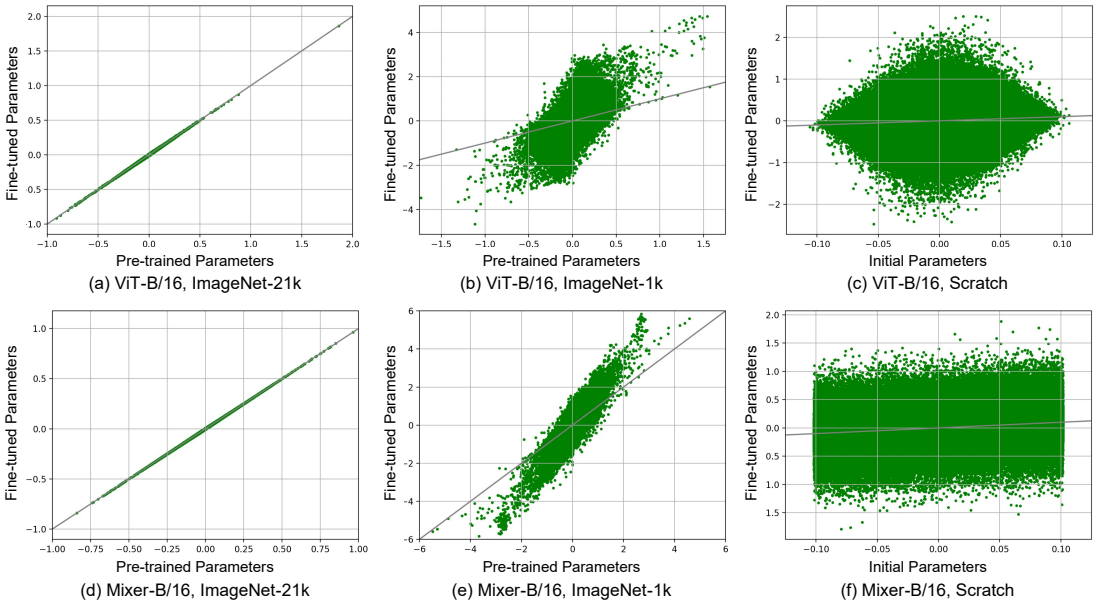


図 5. CIFAR-10 で学習した際のパラメータの変化．横軸は事前学習後の値（または初期値），縦軸は CIFAR-10 による学習後の値を示している．

Model	Sparsity [%]	# Params.	CIFAR-10 Acc. [%]	CIFAR-100 Acc. [%]
ViT-B/16	90.0	84.9M→ 8.5M	95.92 (85.70)	81.09 (61.24)
	95.0	84.9M→ 4.2M	94.37 (83.45)	77.05 (57.22)
	98.0	84.9M→ 1.7M	90.92 (81.96)	68.35 (55.59)
ViT-L/16 [5]	90.0	302.0M→ 30.2M	96.54 (87.96)	81.58 (64.94)
	95.0	302.0M→ 15.1M	95.82 (85.91)	79.56 (60.52)
	98.0	302.0M→ 6.0M	93.70 (83.41)	74.74 (56.70)
Mixer-B/16	90.0	58.4M→ 5.8M	95.17 (80.48)	78.62 (55.51)
	95.0	58.4M→ 2.9M	92.34 (79.34)	72.74 (54.77)
	98.0	58.4M→ 1.2M	87.05 (79.25)	63.26 (53.44)
Mixer-L/16 [25]	90.0	206.1M→ 20.6M	93.93 (79.63)	74.10 (53.25)
	95.0	206.1M→ 10.3M	90.86 (78.52)	68.51 (53.41)
	98.0	206.1M→ 4.1M	88.82 (78.43)	66.11 (52.90)
Pool-M36 [32]	90.0	55.3M→ 5.5M	96.29 (84.69)	82.84 (61.41)
	95.0	55.3M→ 2.8M	94.30 (79.06)	78.17 (53.72)
	98.0	55.3M→ 1.1M	92.53 (64.25)	74.40 (39.34)
Pool-M48 [32]	90.0	72.5M→ 7.3M	96.72 (86.38)	83.67 (62.87)
	95.0	72.5M→ 3.6M	95.10 (79.05)	79.51 (54.28)
	98.0	72.5M→ 1.5M	93.46 (64.91)	76.23 (37.76)

表 1. 提案手法を用いて Transformer モデルを枝刈りした際の分類精度．括弧内の数値は，ランダムに枝刈りした際の分類精度を示している．

デルを用いて枝刈りをした際の分類精度を示す．提案手法による枝刈りはランダムに枝刈りしたモデルと比較して分類精度が向上していることがわかる．また，ViT や MLP-Mixer と比較して，PoolFormer の分類精度が高くなる傾向があることがわかった．これは枝刈りによって ViT や MLP-Mixer のトークンミキサーがパッチ間の特徴量のほとんどを失っている一方で，PoolFormer のトークンミキサーは枝刈りの影響を受けず，プーリング層のみで安定したパッチ間の情報の混合を行うことができているためだと考えられる．

Method	Ours	Magnitude	SNIP	GraSP
ViT-B/16	**79.07**	78.62	**79.07**	77.23
ViT-L/16	**80.79**	80.60	80.61	78.53

表 2. 98%枝刈りした Transformer モデルを ImageNet-1K で再学習した際の精度

次に，表 2 に 98%枝刈りをした Transformer モデルを ImageNet-1K で再学習した際の精度を示す．結果より，ほぼ全ての実験で提案手法は従来手法を上回る分類精度を獲得している．以上の結果より，提案する枝刈りの指標を用いることで，事前学習を考慮した枝刈りを行うことが可能となり，精度を維持したまま大幅にパラメータ数を削減することを可能とした．

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計3件（うち招待講演　0件／うち国際学会　1件）

１．発表者名 小濱大和，平川翼，山下隆義，藤吉弘亘
２．発表標題 Recurrent Neural Network における移動エントロピーを用いた枝刈り
３．学会等名 第25回 画像の認識理解シンポジウム
４．発表年 2022年

１．発表者名 小濱大和，箕浦大晃，平川翼，山下隆義，藤吉弘亘
２．発表標題 事前学習を考慮した シングルショット非構造枝刈り手法の提案
３．学会等名 第26回 画像の認識理解シンポジウム
４．発表年 2023年

１．発表者名 Hirokazu Kohama, Tsubasa Hirakawa, Takayoshi Yamashita, Hironobu Fujiyoshi
２．発表標題 Single-Shot Pruning for Pre-trained Models: Rethinking the Importance of Magnitude Pruning
３．学会等名 International Conference on Computer Vision Workshop（国際学会）
４．発表年 2023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
研究協力者	藤吉　弘亘 (Fujiyoshi Hironobu) (20333172)	中部大学・理工学部AIロボティクス学科・教授 (33910)	

6．研究組織（つづき）

	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
研究協力者	山下　隆義 (Yamashita Takayoshi) (60564721)	中部大学・工学部情報工学科・教授 (33910)	
研究協力者	小濱　大和 (Kohama Hirokazu)	中部大学・工学研究科情報工学専攻・学生（修士） (33910)	

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1件

国際研究集会	開催年
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2021	2021年～2021年

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