立教大学学術推進特別重点資金（立教ＳＦＲ）
大学院学生研究
２０２３年度研究成果報告書

研究科名	立教大学大学院　人工知能科学研究科　人工知能科学専攻		
研究代表者（2024年3月現在のものを記入）	在籍課程・学年 □博士前期課程　年 ■博士後期課程２年	氏　名 立浪　祐貴	
指導教員	所属部局・職名 人工知能科学研究科　准教授	氏　名 瀧　雅人	
自然・人文・社会の別	自然	**個人・共同の別**	個人
研究課題	画像認識のための再帰型ニュートラルネットワークの高速化と理解		
研究組織（研究代表者・共同研究者）※2024年3月現在のものを記入	在籍研究科・専攻・課程・学年 研究代表者 立教大学人工知能科学研究科　博士後期課程２年	氏　名 立浪　祐貴	
研究期間	２０２３　年度		
研究経費（1円単位）	（支出金額）　247,500円　／（採択金額）　250,000円		

研究の概要（200～300字で記入、図・グラフ等は使用しないこと。）

近年，Vision Transformer(ViT)やその派生の手法が画像認識分野では主流となっている．しかしながら，ViT が唯一の有力なアーキテクチャかというと議論がある．特にRNNベースのSequencerがViTを脅かす存在である．しかしながら，Sequencerにはいくつかの懸念がある．特に速度に関する懸念である．この問題は物体認識のような高解像度の画像を扱う場合ほど顕著である．この問題を解消するために RNN ベースのRotational Sequencerなる新しいアーキテクチャを考案した．

キーワード（研究内容をよく表しているものを３項目以内で記入。）

〔　深層学習　〕〔　画像認識　〕〔　RNN　〕

研究成果の概要（図・グラフ等は使用しないこと。）

コンピュータビジョン分野においては、2010 年代は畳み込みニューラルネットワーク(CNN)の時代だった。ところが、ここ数年、自然言語処理で発展を遂げた Transformer アーキテクチャがコンピュータビジョン分野にも影響を及ぼしている。Vision Transformer(ViT)は、画像認識に Transformer を適用した例として広く知られている。ViT に続いて、様々なタスクに Transformer アーキテクチャを応用した研究がされてきた。

画像認識に ViT 以外に考えられないと思われるが、実は必ずしもそうではない。ViT の鍵となる self-attention 層を使用しなくても、よい精度が達成できるアーキテクチャが反証として提案されているのだ。例えば MLP-Mixer は self-attention を使用せす、大部分が MLP を使用したアーキテクチャである。MLP-Mixer は ViT に匹敵する性能を達成している。その他にも様々な MLP ベースのアーキテクチャが提案されている。

最近では、RNN ベースのアーキテクチャである Sequencer を研究代表者らが提案している。これは、Transformer の self-attention 層の代わりに、LSTM などの再帰ニューラルネットワーク(RNN)を採用したアーキテクチャである。Sequencer もまた、MLP-Mixer 同様、ViT の絶対性に対するアンチテーゼである。Sequencer の核となるアイデアは、self-attention 層でトークン間の処理する代わりに、ゲート付き RNN によって長距離相互作用を実現することである。実際、Sequencer はうまくいき、これまでの CNN、ViT、MLP ベースのアーキテクチャにも匹敵する結果を達成した。さらに、ViT や MLP ベースのアーキテクチャと違って特別な工夫なく Sequencer は様々な画像解像度に対応しており、扱いやすさがある。また、ViT よりも推論に必要なメモリが少なくて済む利点がある。

Sequencer はトークン間の相互作用を RNN で表現する。したがってトークンが多くなればなるほど、それらのトークンを逐次処理することになるため、スループットに影響する。画像においてはトークンが多いというのは、解像度が高いことに対応する。特に物体検出やセマンティックセグメンテーションのようなタスクは、画像分類に比べて高解像度の画像を扱うことが多い。とくにこのようなタスクでは、スループットの問題は深刻になる。

この課題を対処するために、Rotational RNN（RotRNN）と呼ばれる RNN の新しいモジュールと、トークン集約プーリング（TAP）を提案する。RotRNN は回転によって一括で処理することによって、TAP はトークンの数を一時的に減らすことによって、従来よりも高速な処理が実現されることが期待される。また、これらの提案モジュールを活用した新しいアーキテクチャを Rotational Sequencer（RoS)を提案する。

RoS について、画像分類での Sequencer 以上の精度を確認し、さらにモデルやデータのスケールに応じて精度が向上することを確認した。また、セマンティックセグメンテーション、インスタンスセグメンテーション、物体検出、２次元姿勢推定などの様々なダウンストリームタスクのバックボーンとして RoS を採用した際に、RoS がパフォーマンスを発揮することを確認している。スループットの比較を精細に実施している途中である。

本期間の研究において、実は RoS が物体検出に向いたアーキテクチャであることがわかってきた。従来の Sequencer は一般的な物体検出モデルのバックボーンで精度を満たさないことを過去の研究で報告していた。これは RNN のせいではないかと推測していたが、実はそうではないということがこの事実によって明らかにされた。この結果は情報処理学会第 86 回全国大会で報告している。現在、これらの結果をもとに論文を執筆中であり、国際雑誌への投稿も予定している 。

研究成果の概要（つづき）

※この（様式２）に記入の成果の公表を見合わせる必要がある場合は、その理由及び差控え期間等を記入した調書（Ａ４縦型横書き１枚・自由様式）を添付すること。

研究発表（研究によって得られた研究成果を発表した①～④について、該当するものを記入してください。該当するものが多い場合は主要なものを抜粋してください。なお、成果発表を確認できる資料を合わせて研究成果報告書提出フォームより提出してください（紙媒体等、研究成果報告書提出フォームから提出できない場合は、別途リサーチ・イニシアティブセンターへ提出してください）。

①雑誌論文（著者名、論文標題、雑誌名、巻号、発行年、ページ）
②図書（著者名、出版社、書名、発行年、総ページ数）
③シンポジウム・公開講演会等の開催（会名、開催日、開催場所）
④その他（学会発表、研究報告書の印刷等）
※修士論文・博士論文は含みません。

③ シンポジウム・公開講演会等の開催

1. 立浪祐貴，瀧雅人，“画像分類のための深層 LSTM”第 29 回画像センシングシンポジウム，パシフィコ横浜，June 14，2023.
2. 立浪祐貴，瀧雅人，“物体検出のためのバックボーンとしての再帰ニューラルネットワーク”情報処理学会第 86 回全国大会，神奈川大学横浜キャンパス，March 17，2024.

④ その他　査読付き国際会議

1. Yuki Tatsunami, Masato Taki, ″FFT-Based Dynamic Token Mixer for Vision″, The 38th Annual AAAI Conference on Artificial Intelligence on Vancouver, Poster presentation, February 2024.