

# 審査委員会報告書

【書式11】

令和 5年 1月 30日

申請者	フリガナ	ルイス ヘンリ パルボサ モルミレ	生年月日	1990年10月26日生
	氏名	Luiz Henrique Barbosa Mormille (男)女)	国籍又は本籍	ブラジル
	学籍番号	18D5251	専攻名	情報システム工学専攻
論文題目		Vision Transformers with Inductive Bias Introduced through Self-attention Regularization		
翻訳題目 (英文の場合のみ)		自己注意正則化により導入される 帰納的バイアスを備えたビジョントランスフォーマー		
審査 委員会 委員	(氏名) 印 (所属機関名) (役職名)			
	主査委員:	渥美 雅保	印 創価大学大学院理工学研究科	教授
	委員:	畝見 達夫	印 創価大学大学院理工学研究科	教授
	委員:	鳥居 直哉	印 創価大学大学院理工学研究科	教授
内容の要旨及び審査結果の要旨 最終試験の結果の要旨			別紙1 別紙2	
博士学位申請論文の受付			受付日: 令和 5年 1月 5日	
博士学位申請論文の受理			受理日: 令和 5年 1月 11日 <input checked="" type="radio"/> 可 ・ 不可	
論文審査の可否			実施日: 令和 5年 1月 19日 <input checked="" type="radio"/> 合 ・ 否	
最終試験の可否			実施日: 令和 5年 1月 19日 <input checked="" type="radio"/> 合 ・ 否	
審査 委員会 の結論	本論文は博士(工学)の学位論文として十分に価値を有するものであり、かつ申請者は専門分野に関して十分な学力と研究能力を有するものと判定した。			

審査委員会の審査及び最終試験の結果を受け、当該研究科委員会は以下の通り判定しました。

研究科委員会の判定	開催日: 令和 5年 2月 14日		
	出席者数 32名	可数 32名	不可数 0名

最終可否  合 ・ 否

学位記番号	博 <input checked="" type="radio"/> 甲 ・ 乙 198号	授与年月日	令和 5年 3月 18日
学位の種類	博士 (工学)	備考	

研究科長 北野 晃朗

# 内容の要旨及び審査結果の要旨

【書式11（別紙1）】

令和 5年 1月 30日

氏名（本籍）	Luiz Henrique Barbosa Mormille（ブラジル）
学位の種類	博士（工学）
学位記番号	博甲第198号
学位記の授与日	令和 5年 3月 18日
学位授与の要件	学位規則第4条第1項該当 創価大学大学院学則第31条第3項該当 創価大学学位規則第3条の3第1項該当
論文題目	Vision Transformers with Inductive Bias Introduced through Self-attention Regularization
論文審査機関	理工学研究科委員会
論文審査委員	主査委員 博士（工学） 渥美 雅保 印 委員 博士（工学） 畝見 達夫 印 委員 博士（工学） 鳥居 直哉 印

## <論文の内容の要旨>

本論文は、画像認識における基盤モデルの1つである Vision Transformer(ViT)に関して、その自己注意(self-attention)の計算に帰納的バイアス(inductive bias)を課すための自己注意正則化(self-attention regularization)手法とそれを組み込んだ ViT 縮小版モデル ARViT(Attention Regulated ViT)を提案し、それらの有効性を実験により検証した論文である。

ViT があらゆる画像認識の基盤モデルとして高い汎化性能を達成するためには、その大規模なアーキテクチャに対して大量のデータを用いて学習する必要があり、より省リソースでの効率的学習手法の開発が課題である。本論文ではこの課題に対して、画像パッチ間の「2次元空間距離に基づく自己注意正則化手法」、並びにグラム行列により計算される「スタイル類似性に基づく自己注意正則化手法」という自己注意計算に帰納的バイアスを課す2つの手法を提案している。また、これら自己注意正則化手法を組み込んだ小規模の ViT モデル ARViT を提案している。そして、ILSVRC-2012 ImageNet データセットを用いた自己教師回転評価(self-supervised rotation estimation)タスクで ARViT の事前学習を行って、5つのダウンストリームタスクの性能を従来の同規模のモデルと比較評価することを通じて、本提案手法により小規模のアーキテクチャとデータセットでも高い汎化性能を達成可能なことを示した。

本論文は7章から構成されている。第1章では、ViT の帰納的バイアスの欠如を解決するための自己注意に対する2つの正則化手法、及びそれらを組み込んだ小規模なモデル ARViT を提案し、大規模ではないデータセットで自己教師あり事前訓練されたモデルの汎化能力の向上と訓練に要する計算コストの削減を達成するという本論文の目的とその貢献に関して述べている。

第2章では、ViT における自己注意、及び自己教師あり学習について説明している。そして、機械学習における仮説空間に帰納的バイアスを課すための正則化手法の研究を概観し、本提案手法の位置づけを述べている。また、ViT の帰納的バイアスの欠如に対処する既存の手法を概観し、本研究の新規性を明らかにしている。

第3章では、本論文で提案する第1の自己注意正則化手法である「2次元空間距離に基づく自己注意正則化手法」について述べている。本手法では、画像の2つのパッチ間の Manhattan 距離が大きいほど自己注意に課せられるペナルティが大きくなるという帰納的バイアス仮説のもとで「距離ロス」を定義して、これを正則化項としてモデルの訓練のためのロス関数に追加することで訓練の効率化を図る。距離ロスの計算では、まず、画像をパッチに分割し、各パッチ間の距離行列を計算する。次に、距離行列からペナルティ行列を計算し、ペナルティ行列から距離ロス関数を用いて距離ロスを計算する。そして、元の分類ロスと距離ロスの重み付き和を最小化する訓

練により、帰納的バイアス仮説のもとでのモデルを求める。

第4章では、本論文で提案する第2の自己注意正則化手法である「スタイル類似性に基づく自己注意正則化手法」について述べている。本手法では、類似のスタイル表現を持つ画像領域に含まれるパッチ間には小さなペナルティが、異なるスタイル表現を持つ画像領域に含まれるパッチ間には大きなペナルティが自己注意に課せられるという帰納的バイアス仮説のもとで「類似性ロス」を定義して、これを正則化項としてモデルの訓練のためのロス関数に追加することで訓練の効率化を図る。類似性ロスの計算では、まず、画像を複数のパッチを含む領域に分割し、各領域のスタイルを表現するグラム行列を計算する。次に、グラム行列間の平均二乗誤差により領域の類似性を表す類似性行列をペナルティ行列として計算し、これと自己注意から計算される自己注意行列から類似性ロス関数を用いて類似性ロスを計算する。そして、元の分類ロスと距離ロスの重み付き和を最小化する訓練により、帰納的バイアス仮説のもとでのモデルを求める。

第5章では、2つの自己注意正則化手法を組み込むためのViT縮小版モデルARViTについて述べている。ARViTは、(1)画像パッチの単一畳み込み演算による埋め込みと位置埋込の正弦位置埋込への変更、(2)Transformerエンコーダ層数の削減、(3)各Transformerエンコーダの隠れ次元数の削減と自己注意変換用のMLPの単一全結合層への置換、(4)分類ヘッドへの入力ークラストークンに対する特徴ベクトルから全画像パッチの特徴ベクトルへの変更、によりViTの訓練パラメータ数を大幅に削減したモデルである。

第6章では、ARViTを用いた2つの自己注意正則化手法の実験評価について述べている。ILSVRC-2012 ImageNetデータセットを用いた自己教師回転評価タスクにより、2次元空間距離に基づく自己注意正則化を組み込んだARViT-ARViT-L、スタイル類似性に基づく自己注意正則化を組み込んだARViT-ARViT-R、自己注意正則化を用いないARViT-ARViT-Base、及び既存モデルの事前学習を行い、5つのベンチマークデータセット(CIFAR-10、CIFAR-100、Flowers、Imagenette、Imagewoof)を用いた分類ダウンストリームタスクでTop-1アキュラシーにより汎化性能を比較評価した。その結果、(1)ARViT-Baseが同規模の既存モデルViT-Tinyを平均11.59%上回ること、(2)ARViT-LとARViT-RがARViT-Baseをそれぞれ平均1.44%、7.36%上回ることを確認し、本論文で提案するARViT、及び2つの自己注意正則化手法の有効性を示した。

第7章では、本研究で達成された成果のまとめに基づいて本論文の総括的な結論、及び今後の研究の発展への展望が述べられている。

#### <論文審査結果の要旨>

Vision Transformerの自己注意機構に対する自己注意正則化は、事前学習における元のモデルのロス関数に、自己注意正則化において計算される注意ロスー距離ロス、類似性ロスーを追加して学習に用いることで実現できる。これより、自己注意正則化手法は、任意のビジョン系のTransformerモデルに対して、それらのアーキテクチャを変更することなく追加的に事前学習に適用することが可能な汎用的な手法であり高い価値を有する。

以上のことから、本論文は、博士(工学)の学位論文として十分な価値を有するものと認められる。

なお、本論文の内容の一部は、下記の信頼できる査読制度を有する権威ある学術雑誌に2編が掲載されており、本論文の研究成果が国際的にも評価されていることを示している。

- 1) L. H. Mormille, C. Broni-Bediako and M. Atsumi, "Regularizing Self-attention Transformers with 2D Spatial Distance Loss", *Artificial Life and Robotics*, Vol.27, pp.586-593, 2022.
- 2) L. H. Mormille, C. Broni-Bediako and M. Atsumi, "Introducing Inductive Bias on Vision Transformers through Gram Matrix Similarity-based Regularization", *Artificial Life and Robotics*, (accepted on Nov. 30, 2022).

# 最終試験の結果の要旨

【書式11（別紙2）】

令和 5年 1月 30日

フリガナ 申請者氏名	ルイス ヘンリ バルボサ モルミレ Luiz Henrique Barbosa Mormille	専攻名	情報システム工学専攻
審査委員会委員	主査委員	渥美 雅保	印
	委員	畝見 達夫	印
	委員	鳥居 直哉	印

## 要旨

審査委員 3 名により、学位論文の内容、及び関連する学問分野に関して口頭で試問を行った。その結果、申請者が博士（工学）としての十分な学力と研究能力を有することを確認した。よって、審査委員会は最終試験の結果を合格と判定した。