卒業論文

事前学習が Vision Transformer に与える影響

18A1066　桝田 修慎

指導教員　山口裕 助教

2022 年 2 月

福岡工業大学情報工学部情報工学科

事前学習が Vision Transformer に与える影響

概要

リザバー計算 [1, 2] を用いる.

キーワード　Vision Transformer

目次

第 1 章

序論

1.1 背景

近年，画像認識分野では，機械翻訳で脚光を浴びることになった Transformer モデル [3] をコンピュータビジョンに適応させた Vision Transformer というモデルが登場した．Vision Transformer は，層を深くし畳み込みを行う畳み込みニューラルネットワークとは違い，畳み込み演算を Attention 機構を用いて代用している．本研究では，Vision Transformer が提案された論文「An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale」を参考にし，事前学習やデータ拡張の有無が，学習及び推論に与える影響を検証した．

1.2 本研究の目的

本研究の目的を以下に示す．

1. 一定の条件下での振る舞いを従来のモデル（ResNet，VGG）比較し，ViT の優れている点・そうではない点を明らかにする．
2. 事前学習やデータ拡張が各モデルに及ぼす影響を調べる．

1.3 深層学習

深層学習とは，脳の神経回路を模したニューラルネットワークをより深くしたものを指し，入力データから有用な特徴量を自動で抽出することができる．

1.3.1　畳み込みニューラルネットワーク

1.3.2　再帰ニューラルネットワーク

1.3.3　ResNet

1.3.4　VGG

1.3.5　Transformer

Transformer[4] は，それまで機械翻訳モデルで多く使われてきた畳み込みニューラルネットワーク・再帰ニューラルネットワークのような複雑なアーキテクチャを持つネットワークとは違い Attention 機構のみを用いて構成されているエンコーダ・デコーダモデルである．

1.3.6　Vision Transformer

Vision Transformer は，機械翻訳で用いられていた Transformer をコンピュータビジョンに適応させたモデルであり，画像を複数のパッチに分割してそれぞれをベクトルとして埋め込み，平坦化して入力とする特徴がある．

1.4　論文の構成

論文の構成を書く．こんにちはおはようございますこんにちわ, 本当ですか

第 2 章

実験モデル

2.1　ネットワークモデル

ネットワーク出力 z は式 (2.1) で得られる．ResNet

$$z = W_{\text{out}} x + b \tag{2.1}$$

2.2　手順

実験の条件を表 2.1 に示す.

表 2.1. 実験の条件

条件	事前学習	データ拡張
条件 1	なし	なし
条件 2	なし	あり
条件 3	あり	あり

- 条件 1: 事前学習なし・データ拡張なし
- 条件 2: 事前学習なし・データ拡張あり
- 条件 3: 事前学習あり・データ拡張あり

実験手順を以下に示す.

1. ステップ 1
2. ステップ 2

第 3 章

実験結果

実験結果を図 3.1 に示す.

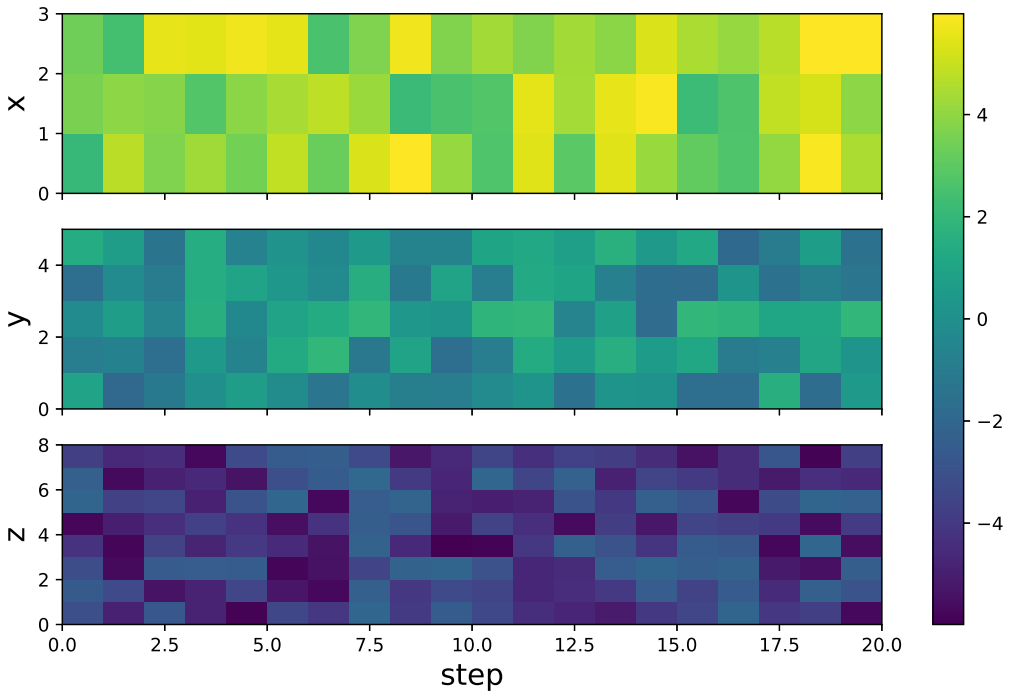

図 3.1. pcolormesh

条件ごとの結果を表 3.1 に示す.

表 3.1. 条件ごとの実験結果

条件	事前学習	データ拡張
条件 1	なし	なし
条件 2	なし	あり
条件 3	あり	あり

第 4 章

議論

議論を書く.

第 5 章

結論

結論を書く.

謝辞

謝辞を書く.

参考文献

[1] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *science*, Vol. 304, No. 5667, pp. 78–80, 2004.

[2] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, Vol. 14, No. 11, pp. 2531–2560, 2002.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

付録 A

実験結果の図

付録があればここに書く.