

ファインチューニング手法に基づく被災建物画像の多クラス損傷度分類の精度改善 Accuracy Improvement for Multi-Class Damage Severity Classification of Disaster-Affected Building Images Based on a Fine-Tuning Technique

山田 悠人¹⁾ 鈴木 海友¹⁾ 松田 一朗¹⁾ 多賀 祥平¹⁾
Yuto Yamada¹⁾ Kaiyu Suzuki¹⁾ Ichiro Matsuda¹⁾ Shohei Taga¹⁾
松澤 智史¹⁾ 二瓶 泰雄¹⁾ 柏田 仁¹⁾
Tomofumi Matsuzawa¹⁾ Yasuo Nihei¹⁾ Jin Kashiwada¹⁾

1 はじめに

地震や津波といった大規模自然災害の発生後、被災家屋の損傷度および損傷原因を正確に判定することは、防災・減災計画の観点から重要である。しかし現在は、被災家屋の画像を専門家が目視で判定しており、その作業には膨大な人手と時間を要する。このような背景のもと、筆者らは深層学習に基づいた 6 クラス分類 (地震大・中・小, 津波大・小, 無被害) によって損傷度と原因を推定する手法を検討した [1]。この手法では、画像分類タスクに実績のある ResNet-50[2] にファインチューニングを施して利用しているが、より高性能なアーキテクチャや学習手法を活用する余地が残されている。

一般に地震被害判定に用いる画像は、撮影条件や被害状況の多様性、データ不足、クラス不均衡、背景ノイズの多さといった課題を抱えている。さらに、従来のファインチューニング手法は全パラメータの更新を要するため、大規模なモデルにおいて計算コストやメモリ負荷の増大を招き、本研究のように訓練データが限られている場合は過学習のリスクも懸念される。そこで本稿では、分類精度の向上を目的として、より効率的なモデルやファインチューニング手法の導入について検討を加える。

2 関連研究

2.1 多クラス損傷度分類

文献 [1] では、能登半島地震後に撮影された 1,040 枚の被災家屋画像を用いて、「地震大・中・小」「津波大・小」「被害なし」の 6 クラスによる損傷度分類を行っている。分類には ImageNet で事前学習された ResNet-50 を用い、ファインチューニングに際しては、クラスごとのサンプル数の不均衡を補正するため、少数クラスに高い重みを付与する重み付きクロスエントロピー損失を採用している。

2.2 ファインチューニング手法の動向

モデルと自己教師あり学習の進展に伴い、近年 Vision Transformer を中心とした高精度な事前学習済みモデルが登場している。例えば ConvNeXt [3] は、ResNet などの従来の CNN ベースモデルを凌駕する性能を画像分類タスクにおいて達成している。

一方、従来のファインチューニング手法は、モデル全体のパラメータを更新対象とするため、訓練データが少ない状況下では過学習が生じやすいという課題があった。これに対し、Low-Rank Adaptation (LoRA) [4] は、重み行列を低ランク分解し、元のモデルパラ

メータを固定したまま低ランク分解された行列のみを更新することで、効率的な転移学習を実現する手法である。LoRA は少量の訓練データでも高い性能が得られることが報告されている。

3 検討項目と実験手順

被災家屋画像を用いた 6 クラス損傷度分類タスクに対して、複数の事前学習済みモデルを比較・検証する。具体的には、2015 年発表の ResNet-50[2] をはじめ、2021 年発表の Swin Transformer Tiny[5]、EfficientNetV2-S[6]、ViT-Base[7]、2022 年発表の ConvNeXt Tiny および Base モデル [3]、DINOv2 [8](Vit_base_patch14_dinov2) といったより新しい CNN および Vision Transformer 系のアーキテクチャを対象とする。

また、各モデルに対しては、従来の標準的なファインチューニング手法と、LoRA によるファインチューニング手法をそれぞれ適用し、その効果を比較した。標準ファインチューニングでは、各モデルの最終層を 6 クラス出力に置き換え、モデル全体のパラメータを更新する。一方 LoRA では、元のモデルパラメータを固定し、低ランク行列のみを更新することで、効率的かつ少量の訓練データでも安定した転移学習を実現している。

3.1 データセット

先行研究 [1] と同様、能登半島地震後に現地で撮影された 1040 枚の画像を「地震大」「地震中」「地震小」「津波大」「津波小」「被害なし」の 6 クラスにラベル付けし、訓練・検証用に 8:2 の比率で分割した。図 1 に、各クラスに対応する家屋画像の一例を示す。

3.2 学習方法

標準的なファインチューニングおよび LoRA を用いたファインチューニングの両者において、クラス間の不均衡を考慮した重み付きクロスエントロピー損失を用いて学習を行った。最適化には、標準ファインチューニングで Adam を、LoRA では AdamW を用いた。また、バッチサイズは 8,16,32 の範囲で探索、エポック数は 30 に固定した。これらの設定により、両手法ともに学習中にエラーを起こすことなく安定してパラメータを更新できる環境を整えた。

4 実験結果

本節では、標準ファインチューニングと LoRA ファインチューニング手法について、各モデルの検証性能を比較した結果について述べる。評価指標には Accuracy, weighted F1, macro F1 を使用した。

4.1 標準ファインチューニング

表 1 に標準ファインチューニングの結果を示す。ConvNeXt-Tiny が Accuracy 0.66 と最も高い性能を示し、ConvNeXt-Base が僅差で続いた。

1) 東京理科大学 創域理工学部

Faculty of Science and Technology, Tokyo University of Science

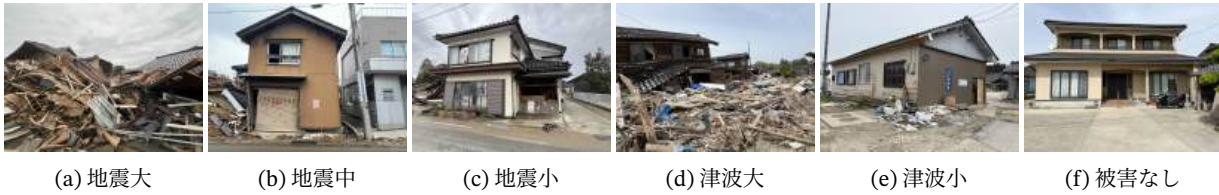


図1: ラベルごとの家屋画像例

表1: 標準ファインチューニングのモデル性能

モデル名	Accuracy	weighted F1	macro F1
ResNet-50	0.51	0.53	0.46
Swin-Tiny	0.62	0.59	0.50
EfficientNetV2-S	0.40	0.42	0.38
ViT-Base	0.63	0.63	0.54
ConvNeXt-Tiny	0.66	0.63	0.54
ConvNeXt-Base	0.64	0.63	0.53
DINOv2	0.45	0.43	0.37

表2: LoRA 適用時のモデル性能

モデル名	Accuracy	weighted F1	macro F1
ResNet-50	0.45	0.45	0.37
Swin-Tiny	0.47	0.50	0.46
EfficientNetV2-S	0.18	0.19	0.16
ViT-Base	0.49	0.51	0.44
ConvNeXt-Tiny	0.64	0.64	0.58
ConvNeXt-Base	0.66	0.64	0.56
DINOv2	0.80	0.80	0.77

4.2 LoRA ファインチューニング

LoRA を適用した結果を表2に示す。DINOv2 が Accuracy と weighted F1 で 0.80 を記録し、全モデル中、最高性能となった。ConvNeXt-Base/Tiny も引き続き高い水準を保った一方で、EfficientNetV2-S は標準ファインチューニングよりも精度が低下した。

ConvNeXt 系は標準ファインチューニングによって比較的高い精度が得られている。一方、LoRA 導入により Vision Transformer 系、とくに DINOv2 の性能が大幅に向上したことから、これらのモデルと低ランク適応の相性の良さが示唆された。これに対し、ResNet-50 の改善量は小さく、既存 CNN では LoRA の利得が限定的である可能性が高い。

4.3 考察

上記実験の対象は、能登半島地震による被災家屋画像を用いた 6 クラス損傷度分類という特定の地震データセットである。この特性を踏まえ、ConvNeXt 系モデルは標準的なファインチューニングで高い精度を示し、CNN アーキテクチャが本タスクの視覚的特徴を効果的に捉えていることを示した。一方、Vision Transformer 系モデルは LoRA の適用により性能が大幅に向上し、特に DINOv2 は Accuracy 0.80 を達成した。これは、DINOv2 が大規模自己教師あり学習により堅牢な視覚表現を獲得し、LoRA が被災家屋に特化した特徴を効率的に付与できたためと考えられる。また、ResNet-50 では LoRA の効果が限定的であった。これは、モデル容量や

既存の特徴が本タスクに対して十分汎用的である一方、LoRA による調整の幅が狭いためと推察される。さらに、ConvNeXt-Tiny が Base より精度が劣る理由として、モデル容量不足が考えられる。

5 おわりに

本稿では、能登半島地震で撮影された被災家屋画像を用いた 6 クラス損傷度分類タスクにおいて、複数の事前学習済みモデルの性能比較および LoRA 適用による性能改善効果を調査した。特に、LoRA を適用した DINOv2 モデルが優れた結果を示し、地震被害という特殊なドメインにおいても自己教師あり学習による Vision Transformer モデルと低ランク適応の相性の良さが明らかになった。今後は、Transformer 系モデルの可視化手法を導入精査し、地震被害判別の判断根拠の解明を進めると共に、建物領域抽出による入力前処理の有効性を実運用データで検証する予定である。

謝辞

本研究は、東京理科大学マルチハザード都市防災研究拠点からのご支援のもとで実施されました。ここに心より感謝の意を表します。

参考文献

- [1] 鈴木 海友 他, “被災家屋画像から被害度や種類を分類するシステムの構築”, ソフトウェア・シンポジウム論文集, Vol. 2025, pp.39–45, Jun. 2025.
- [2] K. He *et al.*, “Deep Residual Learning for Image Recognition”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2016), pp.770–778, Jun. 2016.
- [3] Z. Liu *et al.*, “A ConvNet for the 2020s”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2022), pp.11966–11976, Jun. 2022.
- [4] E. Hu *et al.*, “LoRA: Low-Rank Adaptation of Large Language Models”, Proc. 10th Int. Conf. on Learning Representations (ICLR 2022), Apr. 2022.
- [5] Z. Liu *et al.*, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows”, Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV 2021), pp.10012–10022, Oct. 2021.
- [6] M. Tan *et al.*, “EfficientNetV2: Smaller Models and Faster Training”, Proc. 38th Int. Conf. on Machine Learning (ICML 2021), pp.10096–10106, Jul. 2021.
- [7] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, Proc. 9th Int. Conf. on Learning Representations (ICLR 2021), May 2021.
- [8] M. Oquab *et al.*, “DINOv2: Learning Robust Visual Features without Supervision”, Trans. on Machine Learning Research, 2024.