画像の圧縮復元認識フレームワークの軽量化と高速化

Lightweight and Acceleration of Image Compression, Restoration, and Recognition Framework

柴田蓮 [1] 大原渉偉 [1] 山内悠嗣 [1]
Ren Shibata Shoi Ohara Yuji Yamauchi

中部大学 [1]
Chubu University

1 はじめに

Internet of Things(IoT) の普及に伴いエッジコンピューティングを活用したサービスが拡大し，高性能でリアルタイムな処理が求められている．そこで，IoT デバイス上で圧縮した情報をサーバーへ伝送し，サーバー上で情報を復元し認識モデルで推論する方法が提案されている．さらに，高精度化を目的として，画像の圧縮・復元モデルと画像認識モデルを End-to-End に学習する方法が提案された [1]．本研究では，上記の手法 [1] を軽量化と高速化するため，圧縮・復元モデルと認識モデルから復元モデルを取り除く．これにより，GPU の使用メモリ量と計算量を削減することが可能となる．

2 提案手法

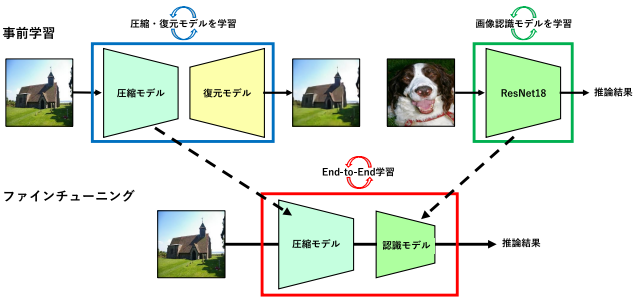

図 1 提案手法による事前学習とファインチューニング.

提案手法の学習の流れを図 1 に示す．圧縮モデルに RNN ベースの再帰型オートエンコーダ [2] を採用する．事前学習の際は，画像の圧縮・復元モデルとして学習する．ファインチューニングの際は，復元モデルを取り除き，圧縮モデルのみを使用する．これにより，復元器による処理を省略できるため，GPU の使用メモリ量と計算量を削減することが可能となる．

認識モデルは ResNet18[3] をベースとし，入力層から幾つかの残差ブロックまでの層を削除する．圧縮器から出力される特徴量を直接認識モデルに入力できるようにするため，圧縮された特徴量の空間的な情報をサンプリングする畳み込み層を追加する．これにより圧縮した情報を復元することなく認識処理を行うことが可能となる．圧縮モデルと圧縮した特徴量を入力とする認識モデルは End-to-End に学習する．

3 評価実験

提案手法の有効性を確認するための評価実験を行う．提案手法と圧縮・復元モデルと認識モデルを End-to-End に学習する従来手法 [1] の性能を比較する．提案手法は復元モデルと入力層から 1 つ目の残差ブロックまでを削除してファインチューニングした提案手法 (layer1) と復元モデルと入力層から 2 つ目までの残差ブロックを削除してファインチューニングした提案手法 (layer2)，復元モデルと入力層から 3 つ目までの残差ブロックを削除してファインチューニングした提案手法 (layer3) を評価する．データセットには ImageNet のサブセットである ImageNet-100 を使用する．

表 1 各手法の処理時間とパラメータ数及び認識精度.

	処理時間 [ms]	パラメータ数	画像認識精度 [%]
従来手法	42.8	20,293,224	64.94
提案手法 (layer1)	23.8	14,867,364	66.40
提案手法 (layer2)	11.4	14,581,796	65.14
提案手法 (layer3)	9.0	13,404,452	54.16

表 1 に各手法におけるモデルのパラメータ数と画像 1 枚当たりの処理時間，認識精度を示す．提案手法 (layer1) は従来手法に対して処理時間を約 43%，パラメータ数を約 27%削減しながら，分類精度を約 1.5%向上させることができた．提案手法 (layer1) よりも残差ブロックを 1 層多くを取り除いた提案手法 (layer2) は，従来手法と同等の認識精度を保持しながら，処理時間を約 74%短縮した．また，提案手法 (layer3) は，従来手法に対して最大となる約 79%の処理時間を短縮した一方で，従来手法に対して認識精度が約 10%下回った．これは認識モデルの残差ブロックを大幅に削ったことにより，特徴表現能力が不足したことが原因だと考えられる．

4 おわりに

本稿では，画像の圧縮復元認識フレームワークの軽量化と高速化を提案した．圧縮した情報を入力として扱えるよう再設計した画像認識モデルと，圧縮・復元モデルとして学習した圧縮モデル部分のみを抽出し 1 つのフレームワークとして End-to-End に学習した．提案手法は，従来手法と同等の認識精度を保持しながら，モデルのパラメータ数と処理時間を削減することができた．

参考文献

[1] 柴田 蓮，山内悠嗣，“画像の圧縮復元モデルと認識モデルの end-to-end 学習”，動的画像処理実利用化ワークショップ，2024.

[2] G. Toderici, D. Vincent, N. Johnston, S. Jin Hwang, D. Minnen, J. Shor, and M. Covell, “Full resolution image compression with recurrent neural networks”, *Conference on Computer Vision and Pattern Recognition*, pp. 5306–5314, 2017.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, *Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.