

エッジコンピューティングのための圧縮画像認識

○高瀬 俊希†, 戸谷 響†, 西片 智広†, 山内 悠嗣†

†: 中部大学

tr22008-4099@sti.chubu.ac.jp, er19052-1939@sti.chubu.ac.jp, tr21010-0146@sti.chubu.ac.jp,
yuu@isc.chubu.ac.jp

概要: 本稿では, エッジコンピューティングを活用した画像認識システムにおけるデータ通信量を削減するために, 画像の圧縮・復元と画像認識を行う 2 つのネットワークから構成されるフレームワークを提案する. クライアント側では, データ通信量を削減するためにエンコーダにより画像を圧縮してエッジサーバへ送信する. エッジサーバでは圧縮データをデコーダにより復元した画像を認識する. 本稿では提案するフレームワークを画像分類と物体検出の 2 つのタスクに適用し, その有効性を確認する.

<キーワード> 画像圧縮, 深層学習, エッジコンピューティング

1. はじめに

エッジコンピューティングを利用した画像認識システムの普及により, インターネットに接続された低スペックな端末においても, 高機能な画像認識技術が利用できるようになった. しかしながら, 高画質・高解像度の画像のようにデータサイズの大きいファイルを送受信すると, 通信時間の増加やネットワーク帯域に負荷がかかる問題が発生する.

このような問題を解決するため, インターネットプロバイダーは通信インフラの拡大を続けているが, データ通信量も増加の一途を辿っているため根本的な解決にはなっていない. 一方で, Message Queuing Telemetry Transport (MQTT) [1] のように IoT 向け通信プロトコル・通信手法が盛んに研究されている. MQTT は IoT 機器向けの通信プロトコルであり, プロトコルヘッダを小さくすることができるため, 通信量を削減することが可能である. しかしながら, 画像のように大容量のデータを送信する場合には, データ全体に対するヘッダ部分の通信量が少ないため高い効果が見込めない.

そこで, 本稿では通信量削減を目的としたエッジコンピューティングのための画像認識フレームワークを提案する. 本研究では通信量を削減するために画像を圧縮してエッジサーバに送信する. エッジサーバでは圧縮された画像から必要な画像認識技術を適用する. 本稿では画像分類と物体検出の 2 つのタスクに適用し, 本フレームワークの有効性を確認する.

2. 関連研究

2.1. 画像の圧縮

ストレージ圧迫の防止やデータ通信時のネットワーク負荷の低減のために画像圧縮技術が使われている. 現在までに様々な圧縮技術が提案されており, 代表的なものとしては非可逆圧縮方式を採用している JPEG [2] や可逆圧縮方式を採用している PNG などがある. これらの圧縮方式は, 人間が画像を見た際に, 劣化を目視で簡単に認識できないレベルで圧縮することが可能である. しかし, 局所的な領域に着目すると, 本来は画像に含まれていないノイズの混入や, 画像に含まれている情報の欠落等の問題が発生していることが確認できる.

近年では, 深層学習の急速な発展に伴い, 畳み込みニューラルネットワークベースの画像圧縮手法の研究が進んでいる. 深層学習に基づく画像圧縮の初期の研究として, Toderici et al. が提案したリカレントニューラルネットワーク (RNN) を利用した手法 [3] [4] が挙げられる. リカレント型オートエンコーダにより画像を復元するネットワークで構成されており, エンコーダの出力を量子化することで画像を圧縮できる. 圧縮した出力をデコーダにより画像として復元できるが, 1 度の処理では画像を完全に再現できないため, 入力画像と復元画像の残差を画像として扱い, 再度リカレント型オートエンコーダに入力することで, 再現しきれなかった残差画像を出力できる. そして, 全ての出力画像の和を最終的な復元画像として出力する. この研究を発端とし, 深層学習ベースの画像圧縮手法 [5] [6] [7] [8] が提案された.

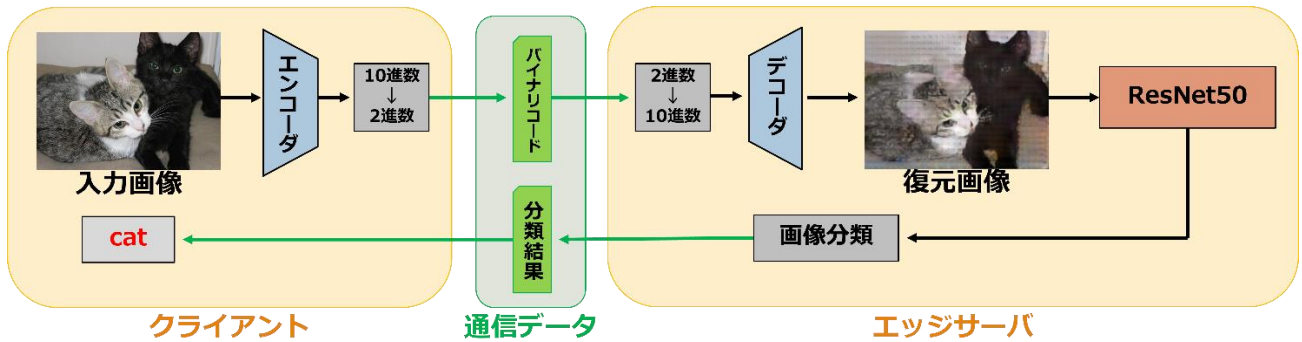


図 1. 画像分類に提案するフレームワークを適用した例.

2.2. 画像認識

2012年に開催された一般物体認識コンペティションである ImageNet Large Scale Visual Recognition Challenge (ILSVRC)において、深層学習を用いた画像分類手法 [9]がトップになって以降、深層学習ベースの画像認識に関して活発に研究されている。Krizhevsky et al. が提案した AlexNet [9]は8層で構成されたネットワークであったが、高性能化のためにより多くの層を持つネットワーク [10] [11] [12]が数多く提案された。

性能が高い深層学習を様々なタスクに応用する研究も活発に行われた。画像分類問題が深層学習により解けるようになると、より難しい問題である物体検出や画像セグメンテーションに深層学習を適用するための手法が提案された。さらには、異なるタスクへの適用だけではなく、強化学習や模範学習等の異なるフレームワークへ適用する手法が提案されている。

3. 提案手法

提案するフレームワークは、画像を圧縮・復元するネットワークと復元した画像を認識するネットワークで構築される。図 1 に提案するフレームワークを画像分類問題に適用した例を示す。提案手法は、クライアント側のエンコーダで画像の圧縮とバイナリコード化を行いエッジサーバへ送信する。エッジサーバでは受信したバイナリコードをデコーダで画像に復元し、ResNet50 で画像分類を行い、復元結果をクライアントに送信する。図 1 は画像分類問題に適用した例であるが、提案するフレームワークは様々な画像認識技術に対して適用可能である。

3.1. 深層学習に基づく画像の圧縮

画像の圧縮、復元には RNN を導入したエンコーダ、デコーダ型のネットワーク [4]を採用する。図 2 に RNN を導入したエンコーダ・デコー

ダ型のネットワークによる画像圧縮器の流れを示す。本手法は、画像からエンコーダ、バイナライザ、デシマライザ、RNN を導入したデコーダによって構成される。

クライアント側では画像をエンコーダに入力し、実数で表現されるベクトルが出力される。この実数ベクトルをバイナライザに入力し、圧縮データを表現するバイナリコードを得る。クライアントからは画像を圧縮したバイナリコードをエッジサーバに送信し、バイナリコードからデシマライザ、デコーダを介して画像を復元する。ただし、入力画像よりも少ない情報量しか保持していないバイナリコードから復元される画像は、部分的に情報が欠落するため画質が劣化する。そこで、入力画像と復元画像の差を復元しきれなかった差分画像として作成し、先ほどと同様に RNN を導入したエンコーダ、デコーダ型のネットワークに入力し、差分画像の復元画像を得る。復元画像と差分画像を合成することで復元画像を更新する。この処理を繰り返すことにより、データ通信量は増加するものの、入力画像により近い復元画像を得ることが可能なる。

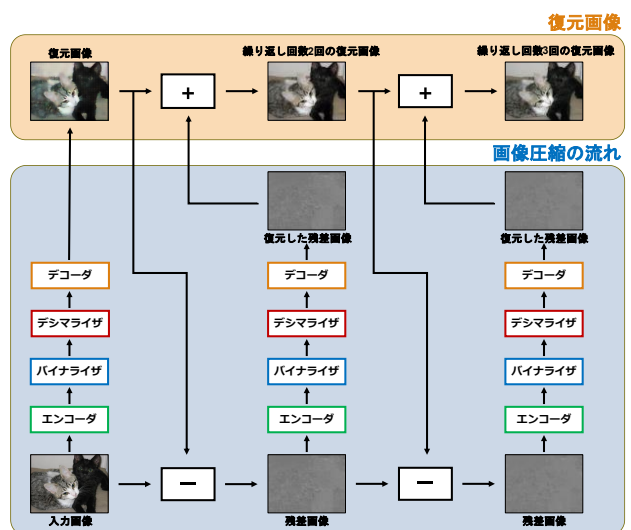


図 2. 深層学習に基づく画像圧縮器の流れ.



図 3. 各繰り返し回数の復元画像とデータサイズ.

図 3 に入力画像と繰り返し回数 1 回から 3 回までの圧縮画像, 画像の容量を示す. 図 3 より, 繰り返し回数が多くなるに従って画質が向上し, 画像容量も大きくなるのが分かる. 特に細かなテクスチャを持つ画像(図 3 の上段)は, 何度も繰り返すことにより著しく画質が向上する.

3.2. 画像分類タスクへの適用

圧縮した画像に対して画像分類を行う. 画像分類では, 入力された画像に写っている物体のクラスを推定する問題を扱う. 本研究では深層学習を用いた画像分類手法として 50 層からなる Residual Neural Network(ResNet)50 [4]を採用する. ResNet は畳み込み層の出力値に入力値を加算する残差ブロックを導入したネットワークである. これにより, ネットワークを深層化しても勾配の消失を防ぐことができるため高精度な認識が可能となったモデルである. エッジサーバでは, デコーダで復元した画像を ResNet50 に入力し, 推定したクラスをクライアントに送信する.

3.3. 物体検出タスクへの適用

圧縮した画像に対して物体検出を行う. 物体検出では, 入力された画像に写っている物体のクラスと位置を推定する問題を扱う. 本研究では深層学習を用いた物体検出手法として MobileNet-SSD [5]を採用する. MobileNet-SSD は, 高速かつ高精度な物体検出が可能な Single-Shot MultiBox Detector (SSD) [6]の畳み込み層を Depthwise Separable Convolution と呼ばれる畳み込み層に置き換えたモデルである. これにより, 通常の畳み込み処理に比べて計算量を削減することが可能である. エッジサーバでは, デコーダで復元した画

像を MobileNet-SSD に入力し, 推定した物体のクラスと位置をクライアントに送信する.

4. 評価実験

提案手法の有効性を確認するために圧縮の適用前後の平均画像容量を比較する. また, 画像分類と物体検出の 2 つの問題に適用し, 性能を確認する.

4.1. 圧縮前後の平均画像容量の比較

画像圧縮の適用前後の各データセットの平均画像容量を確認する

4.1.1. データセット

ImageNet は物体認識ソフトウェアの研究で用いるために設計された 1400 万枚を超える大規模な教師ラベル付きカラー画像データベースである. 本実験では画像枚数が多い上位 10 クラス (bird, car, cat, dog, flower, guitar, insect, person, ship, vegetable) 各クラス 50 枚, 合計で 500 枚を使用する.

VOC2007 は物体検出, セグメンテーション用に標準化された物体のクラスと位置を示すアノテーションと教師画像付きの画像データセットである. 本実験では学習用データと検証用データ合わせて 5,011 枚を使用する.

4.1.2. 平均画像容量の比較結果

各データセットの平均画像容量と繰り返し回数 1 回の圧縮画像の平均画像容量を表 1 に示す.

表 1. 平均画像容量の比較

	ImageNet	VOC2007
元画像 [kB]	65.5	87.9
圧縮画像 [kB]	3.00	22.2

表 1 より, ImageNet では 95.4[%], VOC2007 では 74.7[%]の平均画像容量を削減できていることが確認できる. ImageNet と比較して VOC2007 の平均画像容量が大きい, これは画像の複雑さに起因していると考えられる.

4.2. 画像分類タスクに適用した際の結果

提案するフレームワークが画像分類問題に対して有効であるか確認する. 圧縮器は, 4.1 で使用したモデルを使用する. 本実験には ImageNet を使用し, 分類器の学習と評価には, 学習用画像 7500 枚 (各クラス 750 枚), 評価用画像 500 枚 (各クラス 50 枚) を使用する. 本実験では, 下記の 3 つの手法のデータ容量削減率と画像分類性能を比較する.

- 従来法
未圧縮学習画像 + 未圧縮評価画像
- 提案手法 (未圧縮画像モデル)
圧縮学習画像 + 未圧縮評価画像
- 提案手法 (圧縮画像モデル)
圧縮学習画像 + 圧縮評価画像

なお, 提案手法は RNN を導入したエンコーダ・デコーダ型ネットワークの繰り返し回数を変えることで画質と画像容量をコントロールできる. 本実験では, 繰り返し回数を 1~3 回まで試行する. 画像分類性能の評価には, 正解クラスと推定クラス的一致度合いを測る Accuracy(Acc)を用いる.

4.2.1. 結果

図 4 に従来手法と提案手法の平均分類精度を示す. 図 4 より, 画像を圧縮しない従来法と比較して提案手法はデータ容量の 85[%]~95[%]を削減することができた. 次に, 画像分類性能を比較する. 画像を圧縮しない従来法の性能が最も高く, 2 つの提案手法は性能が若干低下した. RNN を導入したエンコーダ・デコーダ型ネットワークの繰り返し回数を変えることで画質と画像容量をコントロールできるが, 2 回繰り返した際の結果が最も性能が高い. 従来法よりも Acc が 3.2[%]の低下を許容できる場合には, 提案手法はデータ通信量を 91[%]削減することが可能である.

4.2.2. 考察

提案手法は従来手法と比較して Acc が若干低下している. これは画像を圧縮することにより認識に必要な情報が欠落したことや, 認識に悪影響を及ぼすノイズが混入したためだと考えられる.

図 5 に特徴空間を t-SNE により低次元化して可視化した結果を示す. 従来法と提案手法の両手

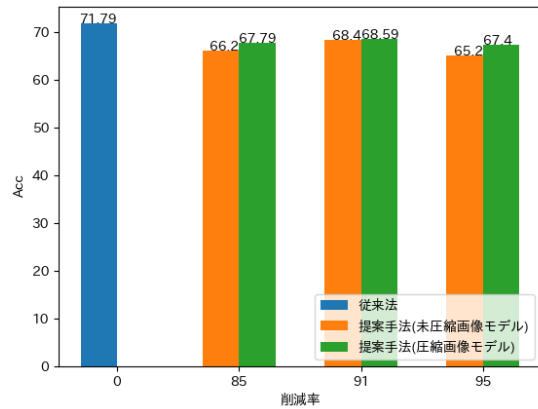


図 4. 従来法と提案手法の平均分類精度とデータ通信量削減率.

法ともに, サンプルの分布がクラス内において密集していることがわかる. しかし, cat 等の一部のクラスにおいては, 提案手法では分布に広がりをもっており, これらのクラスにおいては分類性能が低下した. cat 等のクラスは, 被毛の細かなテクスチャを多く含むため, 画像を圧縮した際に画像が大きく劣化したと考えられる.

4.3. 物体検出タスクに適用した際の結果

提案するフレームワークが画像分類問題に対して有効であるか確認する. 本実験には VOC2007 を使用し, 学習用画像 2501 枚, 検証用画像 2510 枚, 評価用画像 4952 枚を使用する. 圧縮器は, 4.1 で使用したモデルを使用する. 従来法と提案手法の定義は 4.2 と同じである. 物体検出性能の評価には, mean Average Precision (mAP)を用いる.

4.3.1. 結果

図 6 に従来手法と提案手法の平均検出精度を示す. 図 6 より, 画像を圧縮しない従来法と比較して提案手法はデータ容量の 75[%]~77[%]を削減することができた. 次に, 物体検出性能を比較する. 画像を圧縮しない従来法が最も性能が高く, 次に圧縮画像を用いて物体検出器を学習する提案手法 (圧縮画像モデル)の性能が高いことが確認できる. RNN を導入したエンコーダ・デコーダ型ネットワークの繰り返し回数を変えることで画質と画像容量をコントロールできるが, 3 回の繰り返しの結果が最も性能が高い. 従来法よりも mAP が 3.64[%]の低下を許容できる場合には, 提案手法はデータ通信量を 75[%]削減することが可能である.

4.3.2. 考察

画像分類問題と同様に, 提案手法は従来手法と比較して mAP が低下している. これは画像を圧縮することにより認識に必要な情報が欠落したこ

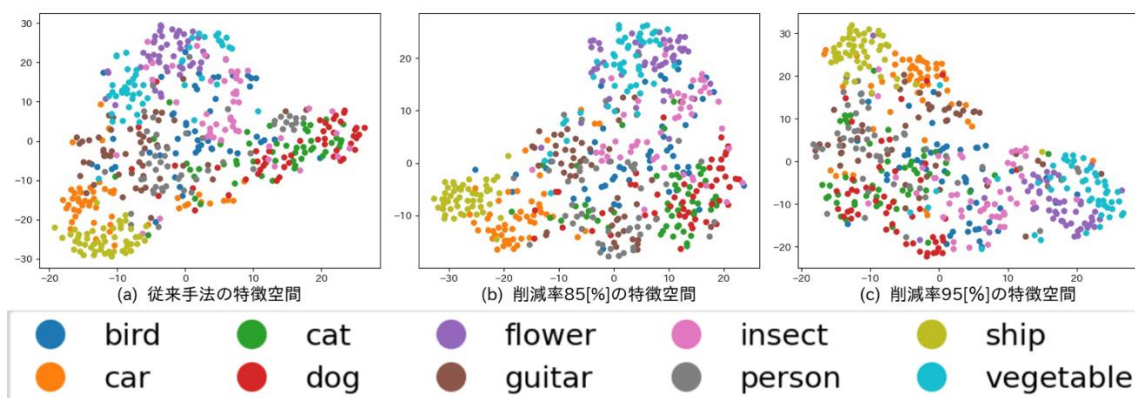


図 5. t-SNE による特徴空間の可視化.

参考文献

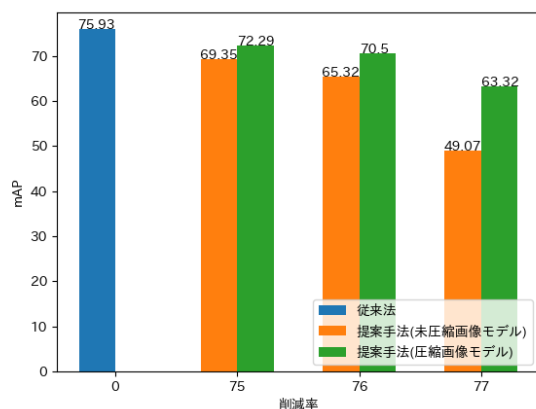


図 6. 従来手法と提案手法の平均検出精度.

とや、認識に悪影響を及ぼすノイズが混入したためだと考えられる。ただし、RNNを導入したエンコーダ・デコーダ型のネットワークの繰り返し回数を増加させると性能も向上することがわかる。これは、物体検出が画像分類よりも難しく、画像に含まれる情報が少ないと十分な性能を発揮することができないと考えられる。

5. おわりに

本稿では通信量削減を目的としたエッジコンピューティングのための圧縮画像認識のフレームワークを提案した。画像分類と物体検出の2つのタスクにおいて、圧縮した画像を送信することで若干の性能の低下が見られるが、通信量を大幅に削減することができることを確認した。今後は画像圧縮と画像認識の2つのネットワークを End-to-End で同時に学習することを検討している。

- [1] OASIS, “MQTT Version 5.0,”: <http://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html>.
- [2] M. W. Marcellin, M. J. Gormish, A. Bilgin, M. P. Boliek, “An Overview of JPEG-2000,” Conference on Data Compression, 2000.
- [3] G. Toderici, S. M. O. Malley, S. J. Hwang, D. Vincent, “Variable Rate Image Compression with Recurrent Neural Networks,” International Conference on Learning Representations, 2016.
- [4] G. Toderici, D. Vincent, N. Johnston, S. Jin Hwang, D. Minnen, J. Shor, M. Covell, “Full resolution image compression with recurrent neural networks,” Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017.
- [5] J. Ballé, V. Laparra, E. P. Simoncelli, “End-to-end Optimized Image Compression,” International Conference on Learning Representations, 2017.
- [6] O. Rippel, B. Lubomir, “Real-time adaptive image compression,” International Conference on Machine Learning, 2017.
- [7] M. Tschannen, E. Agustsson, M. Lucic, “Deep generative models for distribution-preserving lossy compression,” Neural Information Processing Systems, 2018.
- [8] H. Ma, D. Liu, N. Yan, H. Li, F. Wu, “End-to-End Optimized Versatile Image Compression with Wavelet-Like Transform,” IEEE Transactions on

Pattern Analysis and Machine Intelligence, 2020.

- [9] A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Neural Information Processing Systems*, 2012.
- [10] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Visual Recognition,” *International Conference on Learning Representations*, 2015.
- [11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going Deeper with Convolutions,” *Computer Vision and Pattern Recognition*, 2015.
- [12] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” *Computer Vision and Pattern Recognition*, 2016.
- [13] G. a. V. D. a. J. N. a. J. H. S. a. M. D. a. S. J. a. C. M. Toderici, “Full resolution image compression with recurrent neural networks,” *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017.
- [14] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, abs/1409.1556, 2014.
- [15] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *CoRR*, abs/1704.04861, 2017.
- [16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, “SSD: Single shot multibox detector,” *European conference on computer vision*, 2016.
- [17] M. Everingham, S. Eslami, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, “The pascal visual object classes challenge: A retrospective,” *International journal of computer vision*, 2015.
- [18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” *2009 IEEE conference on computer vision and pattern recognition*, 2009.

6. 著者紹介

高瀬俊希：中部大学大学院工学研究科ロボット理工学専攻在学。現在、エッジコンピューティングのための圧縮画像認識に関する研究と深層学習を用いた Deepfake の検出の研究を進めている。

戸谷響：中部大学工学部ロボット理工学科在学。現在、エッジコンピューティングのための圧縮画像認識に関する研究を進めている。

西片智広：中部大学大学院工学研究科ロボット理工学専攻在学。現在、未来画像生成と強化学習の研究を進めている。

山内悠嗣：2012 年中部大学大学院博士後期課程修了。2010 年-2012 年 独立行政法人日本学術振興会特別研究員 DC2。2014 年中部大学助手、2017 年より同大学講師。画像認識、機械学習、知能ロボティクスの研究に従事。