

特徴量差分を用いた視聴と画像認識のための階層画像符号化

Scalable Image Coding for Humans and Machines Using Feature Differences

巽 優衣[†]

Yui Tatsumi[†]

[†] 早稲田大学基幹理工学部

[†]School of Fundamental Science and
Engineering, Waseda University

進藤 高紘[‡]

Takahiro Shindo[‡]

渡部 泰樹[‡]

Taiju Watanabe[‡]

[‡] 早稲田大学大学院基幹理工学研究科

[‡]Graduate School of Fundamental Science
and Engineering, Waseda University

渡辺 裕^{†‡}

Hiroshi Watanabe^{†‡}

Abstract: With the advancement of deep learning, the demand for scalable image coding that compresses images for both human vision and image recognition tasks has been increasing. Previous research proposed a method that connects two LIC models using a feature fusion network. We improve this method by incorporating the differences of the features.

1 はじめに

画像圧縮技術は、視聴時の品質を維持しながら情報量を削減し、効率的な伝送や保存を実現するために研究されてきた。しかし、近年では深層学習の発展により、画像が画像認識のために使用されるケースが増加している。視聴と画像認識では、それぞれ求められる画像の情報量や品質が異なるため、画像認識に最適化された圧縮手法に関する研究も行われている。

特に農場や交通における監視システムなどでは、画像は主に画像認識モデルによって分析され、必要に応じて人間が確認するような運用が想定されている。このようなユースケースでは、視聴と画像認識の両方の目的に対応可能な階層型の画像圧縮技術が求められている。そこで、階層画像符号化として、2つの Learned Image Compression (LIC) モデルを用いた手法が提案されている。本稿では、特徴量差分を用いて2つの LIC モデルのより密接な連携を目指し、画像圧縮性能の改善を図る。

2 従来手法

Scalable Image Coding for Humans and Machines Using Feature Fusion Network [1] は、画像認識のための LIC モデルと、視聴用画像の再構成に必要な追加情報を圧縮する LIC モデルを、feature fusion network により結合した階層符号化手法を提案している。この論文は、視聴用に用意される画像の情報は、画像認識用に用意される画像の情報を含んでいるという考えに基づいている。そのため、画像認識を行うときは物体の輪郭情報のみを復号し、視聴用には画像認識向けの再構成画像と原画像の差を追加情報として圧縮する。Feature fusion network では、画像認識のために圧縮された特徴量に、視聴のために圧縮された追加の特徴量を加算している。これにより、画像全体の特徴量が得られ、視聴のための画像が再構成される。しかし、2つの LIC モデルは一箇所でのみ接続されており、これらの接続方法について再検討の余地がある。

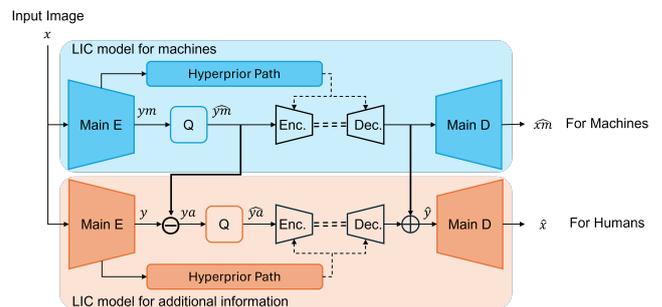


図 1: 提案手法のモデル構造。

3 提案手法

3.1 モデル構造

従来手法は、2つの LIC モデルが feature fusion network でのみ接続されている。そのため、追加情報を圧縮する LIC モデルのエンコーダーが、追加情報となる特徴量を適切に抽出する必要がある。そこで本稿では、画像全体の特徴量と画像認識のための特徴量の差分を用いた階層符号化を提案する。ネットワークに追加情報となる差分を明示的に与えることで、符号化性能の改善を目指す。

提案するモデル構造を図 1 に示す。画像認識のための LIC モデルには SA-ICM[2]、追加情報を圧縮する LIC モデルには LIC-TCM[3] を用いる。これらのモデルはどちらも Channel-wise autoregressive model[4] に基づいている。まず、入力画像は画像認識のための LIC モデルと、視聴のための追加情報を圧縮する LIC モデルのメインエンコーダーによって、それぞれ特徴量 y^m と y に変換される。これらの特徴量はそれぞれ a 個と b 個の特徴量 $\{y^m_1, y^m_2, \dots, y^m_a\}$ および $\{y_1, y_2, \dots, y_b\}$ に分割され、順に圧縮される。パラメーター b をパラメーター a より小さく設定することで、追加情報を表すための特徴量のチャンネル数を減らす。

異なるチャンネル数を持つ特徴量の差分を次の式を用いて

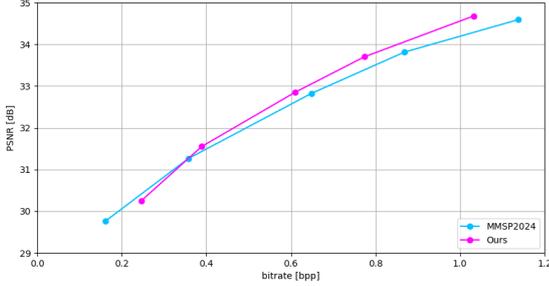


図 2: 提案手法と従来手法の画像圧縮性能.

計算する. ただし, $k = 1, 2, \dots, b$ とする.

$$ya_k = y_k - y\hat{m}_k, \quad (1)$$

$$ya = \text{conc}(ya_1, ya_2, \dots, ya_k). \quad (2)$$

式 (1) において, y_k は追加情報を圧縮する LIC モデルのメインエンコーダーの出力を, $y\hat{m}_k$ は画像認識のための LIC モデルのメインエンコーダーの出力 ym_k の量子化後の値を表す. また, 式 (2) において, conc は特徴量を結合する関数を表す.

追加情報を圧縮する LIC モデルの Hyperprior path では, 特徴量 y から特徴量 z を抽出し, ニューラルネットワークにより hyperprior パラメーターを得る. これらのパラメーターを利用して $y\hat{a}$ を圧縮する.

デコーダー側では, 圧縮された $y\hat{a}$ と $y\hat{m}$ を feature fusion network を用いて融合することで, 画像全体の特微量 \hat{y} が復号され, 視聴向けの画像の再構成を行う.

3.2 損失関数

追加情報を圧縮する LIC モデルを以下の損失関数で学習する.

$$\mathcal{L}_a = \mathcal{R}(ya) + \mathcal{R}(z) + \lambda \cdot \text{mse}(x, \hat{x}). \quad (3)$$

式 (3) において, $\mathcal{R}(ya)$ は特微量 ya のビットレート, z は hyperprior エンコーダーの出力, $\mathcal{R}(z)$ はそのビットレートを表す. x は入力画像, \hat{x} は出力画像を表し, mse は平均二乗誤差の関数である. λ は値を変動させることでビットレートと平均二乗誤差のトレードオフを調整できる定数を表す.

4 実験および結果

提案手法の画像圧縮性能を評価する. 画像認識のための画像圧縮の性能は, SA-ICM の性能と一致する. 本稿では, 視聴のための画像圧縮性能を計測する. COCO-train データセットを学習に, COCO-val データセットを評価に用いる [5]. また, パラメーター a を 5, パラメーター b を 4 とする.



(a) 入力画像 (b) 画像認識向け (c) 視聴向け

図 3: 視聴と画像認識のための再構成画像の例.

実験結果を図 2 に示す. 図 2 より, 特に高いビットレートにおいて, 提案手法は従来手法を圧縮性能で上回ることが分かる. 定量的には BD-Rate が 4.23% 向上する. 視聴と画像認識のための再構成画像を図 3 に示す. 画像認識のために再構成された画像が, 追加情報を加えることで入力画像に近い視聴向けの画像に変換されていることが確認できる.

5 まとめ

本稿では, 特徴量差分を用いた視聴と画像認識のための階層画像符号化を提案した. 提案手法により, 特に高ビットレートにおいて画像圧縮性能が改善することを実験で確認した. 今後は, 様々な LIC モデルを組み合わせて階層画像符号化を作ることができる汎用的なフレームワークを設計する必要がある.

6 謝辞

本研究成果は, 国立研究開発法人情報通信研究機構の委託研究 (JPJ012368C05101) により得られたものです.

参考文献

- [1] T. Shindo, T. Watnabe, Y. Tatsumi, H. Watanabe, “Scalable Image Coding for Humans and Machines Using Feature Fusion Network,” IEEE International Workshop on Multimedia Signal Processing (MMSp), Oct. 2024.
- [2] T. Shindo, K. Yamada, T. Watanabe, H. Watanabe, “Image Coding for Machines with Edge Information Learning Using Segment Anything,” IEEE International Conference on Image Processing (ICIP), Nov. 2024.
- [3] J. Liu, H. Sun, and J. Katto, “Learned Image Compression with Mixed Transformer-CNN Architectures,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14388-14397, June 2023.
- [4] D. Minnen and S. Singh, “Channel-Wise Autoregressive Entropy Models for Learned Image Compression,” IEEE International Conference on Image Processing (ICIP), pp. 3339-3343, Oct. 2020.
- [5] T. Y. Lin et al., “Microsoft COCO: Common Objects in Context,” Computer Vision - ECCV 2014: 13th European Conference, Part V 13, pp 740-755, Sep. 2014.