

画像認識向け深層画像圧縮における補助損失の導入

Introduction of Auxiliary Loss in Deep Image Compression for Image Recognition

飯野景¹
Kei Iino

赤松俊輔¹
Shunsuke Akamatsu

渡辺裕¹
Hiroshi Watanabe

江田毅晴²
Takeharu Eda

榎本昇平²
Shohei Enomoto

坂本啓²
Akira Sakamoto

早稲田大学大学院 基幹理工学研究科¹
Graduate School of Fundamental Science and Engineering,
Waseda University

NTT ソフトウェアイノベーションセンター²
NTT Software Innovation Center

1. まえがき

近年、深層学習モデルによる画像や映像の認識能力が大幅に向上し、様々なアプリケーションで活用されている。これに伴い、人間の視覚ではなく、深層学習モデルの認識に最適化された画像圧縮方法の需要が高まっている。本研究では画像認識向けの深層画像圧縮モデルの学習方法に着目し、補助損失の導入を提案する。提案手法は、推論時における追加のオーバーヘッド無しで、エンコーダの識別能力および圧縮性能の向上を達成することができる。

2. 提案手法

画像認識向け深層画像圧縮においては、認識タスクに必要な情報のみを識別および抽出して圧縮することが理想的である。そのためいくつかの研究では前景領域を追加のモデルで抽出して、それを元に明示的なビット割り当てを行う方法が提案されている[1, 2]。しかし、認識タスクにおいてどの領域や情報が重要かどうかは自明ではなく、タスクやモデルによって異なる。また、この方法では推論時に追加のオーバーヘッドがエンコーダ側に生じてしまう。

そこで本研究では、圧縮モデルのエンコーダに認識タスクにおける必要な情報の識別能力を獲得させるために、補助損失の導入を提案する。提案手法では、図1のようにエンコーダの直後に補助損失計算用の軽量な認識モデル (Auxiliary branch) を設けて最適化を行う。

$$L = \lambda(L_{task} + \alpha L_{aux}) + R \quad (1)$$

L_{task} は認識モデルの損失、 L_{aux} は Auxiliary branch の損失、 R は推定ビットレートを示す。また、 λ はラグランジュ乗数、 α は重み係数である。 L_{aux} は認識タスクと同じ、またはそのサブタスクの損失関数とする。 α を適切な値に調整することで、推論時における追加のオーバーヘッド無しで、エンコーダの識別能力および圧縮性能の向上を達成することができる。

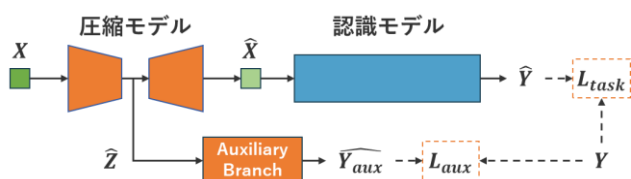


図1 補助損失を導入した学習 (X : 入力画像, \hat{X} : 復号画像, \hat{Z} : 復号潜在表現, Y : 正解ラベル, \hat{Y} : \hat{X} を入力としたときの認識モデルの出力, \hat{Y}_{aux} : \hat{Z} を入力としたときの Auxiliary branch の出力)

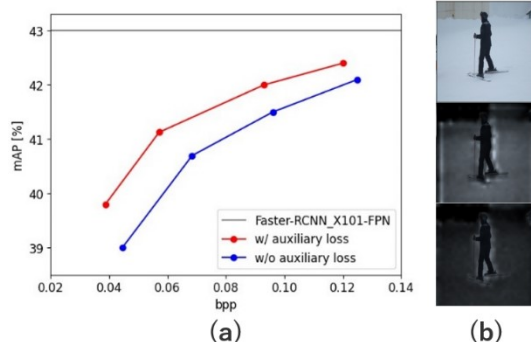


図2 (a) R-D 曲線による比較, (b) 上: 入力画像, 中: 補助損失なしのビット割り当て, 下: 補助損失ありのビット割り当て

3. 実験

物体検出のタスクで提案手法の効果を検証した。データセットとして COCO2017 を使用し、認識モデルに Faster-RCNN (X101-FPN)、圧縮モデルに Cheng2020Attention を用いた。認識モデルの重みは固定した状態で Adam optimizer を用いて圧縮モデルを 50 エポック学習した。本実験では蒸留ベースの学習を行い、未圧縮の画像を入力としたときの Faster-RCNN の出力を L_{task} , L_{aux} の正解ラベルとして用いた。また提案手法においては $\alpha = 0.1$ と設定した。

実験結果を図2(a)に示す。図2(a)より、補助損失を使用しない場合と比較して圧縮性能が向上することが確認でき、最大で約 33% の B-D レートの改善を達成した。また図2(b)にはビット割り当てを可視化した画像を示す。補助損失の導入によりエンコーダの識別能力が向上していることが確認できる。

4. むすび

本研究では、画像認識向けの深層画像圧縮モデルの学習方法に着目し、補助損失の導入を提案した。提案手法は、推論時における追加のオーバーヘッド無しで、エンコーダの識別能力および圧縮性能の向上を達成することができる。

参考文献

- [1] B. Li, L. Ye, J. Liang, Y. Wang, and J. Han, "Region-of-interest and channel attention-based joint optimization of image compression and computer vision," Neurocomputing, vol. 500, pp. 13–25, Aug. 2022.
- [2] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," IEEE J. Sel. Areas Commun., vol. 41, no. 1, pp. 55–71, Jan. 2023.