修 士 論 文 概 要 書

Summary of Master's Thesis

Date of submission: 01/22/2024 (MM/DD/YYYY)

専攻名（専門分野） Department	情報理工・ 情報通信専攻	氏名 Name	飯野景	指　導 教　員 Advisor	渡辺　裕　㊞
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	5122F007-6		
研究題目 Title	補助損失の追加による画像認識のための深層画像符号化の性能改善 Improving Performance of Deep Image Coding for Image Recognition by Adding Auxiliary Loss				

1. まえがき

近年画像認識モデルの性能が飛躍的に伸びており，復号した画像や映像を人間が確認することなく認識モデルが分析するケースが増えている．これに伴い，人間の視覚ではなく，モデルの認識に最適化された画像符号化（Image Coding for Machines (ICM)）の研究が盛んになっている[1,2,3]. ICM では，符号化前と符号化後のデータが視覚的に類似している必要はなく，データサイズを最小化しつつ，認識精度の低下を抑えることが求められる．そのため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と考えられる. 本研究では ICM における学習方法に注目し，補助損失を追加することでこの機能の実現を目指す．物体検出およびセマンティックセグメンテーションタスクで実験を行い，補助損失を使用しない従来法と比較してそれぞれ 27.7%，20.3% の BD-Rate の改善を達成した．

2. 従来手法とその問題点

ICM では，認識タスクに必要な情報を識別および抽出して圧縮することが重要である．そのため ICM の研究のメインアプローチは次の二つの方法に分類される．一つは認識タスクの損失で圧縮モデルを最適化することである[1]．タスクの損失で学習を行うことで，圧縮モデルは後段の認識モデルの精度とビットレートのトレードオフを最適化することができる. もう一つは関心領域（Region of Interest (ROI)）ベースのビット割り当て方式[2,3]である．これは前景領域などに優先的にビット割り当てを行うことでタスク志向の符号化器を設計する．しかしこれらの方法はそれぞれ問題点が存在する．タスク損失での学習においては，タスクの難易度やモデルの複雑さに応じて十分な最適化が困難になる場合がある[4]．特に認識モデルのネットワークが深い場合，誤差逆伝搬の性質上最適化が困難になる[5]．ROI ベースのビット割り当ての場合，ROI を決定するための認識モデルが必要となるため，エンコーダ側に追加のオーバーヘッドが生じてしまう．また，セマンティックセグメンテーションやパノプティックセグメンテーションなどの背景のクラス分類も行うタスクの場合や，画像に対する説明を行うイメージキャプショニングタスクでは ROI を定義することが難しい．

3. 提案手法

本研究では ICM における学習方法に着目し，既存手法の問題点の解決を目指す．具体的には圧縮モデルの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）を設けて圧縮モデルと共同で最適化を行う．損失関数は以下のように定義する．

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

ここで，L_{aux}は Auxiliary branch の損失，αは重み係数である．また，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を用いる．Auxiliary branch はデコーダの直前から分岐して配

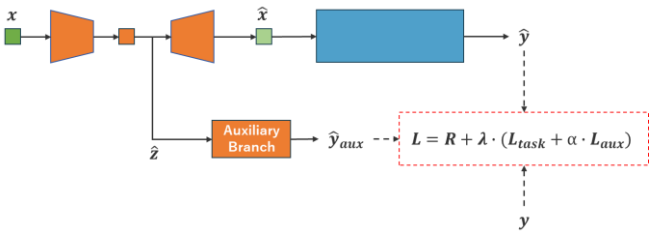

図 1　提案手法の概要

置されるため，L_{aux}で計算される誤差はエンコーダに直接伝搬される．これにより，認識モデルが大規模な場合でも，エンコーダが認識タスクに必要な情報を理解することをサポートする．また，誤差伝搬対象をエンコーダのみに絞ることで, ROI 方式のような機能をエンコーダが獲得することが期待でき，メインタスクの学習への悪影響の可能性も低減できると考える. また, 提案手法では ROI の定義を必要としないため，様々なタスクへの応用が可能と考えられる．さらに Auxiliary branch は学習時のみ使用し，推論時には使用しないため，従来の ROI 方式のようなオーバーヘッドも生じない

4. 評価実験および結果

物体検出及びセマンティックセグメンテーションのタスクで提案手法の効果を検証した．圧縮モデルには事前学習済みの cheng2020attn [6]を用い，物体検出モデルには Fater-RCNN（Resnext101-FPN），セグメンテーションモデルには DeepLabv3+（ResNet101）を使用した．データセットにはそれぞれ，COCO2017，Pascal-Context 59 をし，50 エポックの学習を行った．Auxiliary branch のネットワーク構成は各認識モデルを参考にし，バックボーンモデルを 18 の層の小さなネットワークに変更した．実験では補助損失を使用せず，タスク損失のみで学習した圧縮モデルをベースラインとして使用し，提案手法との比較を行った．また，物体検出においてはベースラインに既存手法の ROI 方式を適用したものとの比較も行った．

図 2 に物体検出，図 3 にセマンティックセグメンテーションにおける実験結果を示す．各図からわかる通り，補助損失の追加により R-D 性能が改善していることがわかる. 物体検出では平均 27.7%，セマンティックセグメンテーションでは平均 20.3%の BD-Rate を達成した．また物体検出では，既存手法の ROI 方式と比較しても提案手法が高い R-D 性能を誇っていることがわかる．

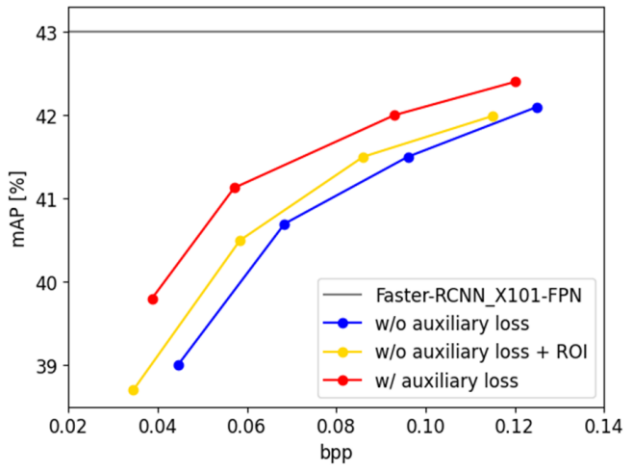

図 2　物体検出タスクにおける各手法の R-D 性能の比較

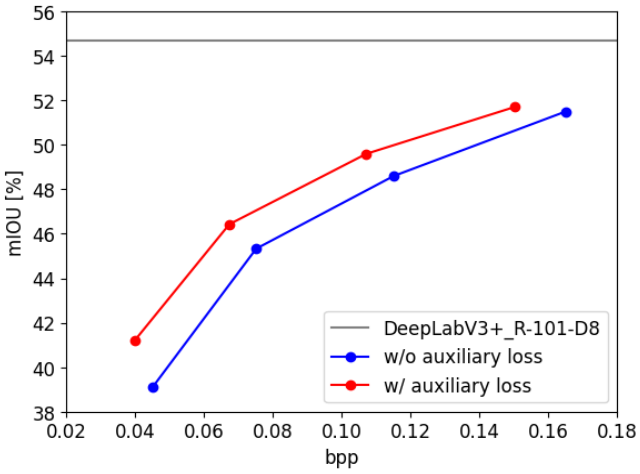

図 3　セマンティックセグメンテーションタスクにおける各手法の R-D 性能の比較

5. 結論

本研究では ICM の学習における補助損失の追加を提案した．提案手法では，エンコーダの後に Auxiliary branch を設け同時に最適化を行うことで，認識タスクに対するエンコーダの認識力獲得を補助する．補助損失を使用しない従来法と比較して，物体検出で平均 27.7%，セマンティックセグメンテーションで平均 20.3%の BD-Rate 改善を達成した．

参考文献

[1] A. Harell, et. al, “Rate-distortion theory in coding for machines and its application,” arXiv preprint arXiv:2305.17295, May. 2023.

[2] B. Li, et. al, “Region-of-interest and channel attention-based joint optimization of image compression and computer vision,” Neurocomputing, vol. 500, pp. 13–25, Aug. 2022.

[3] J. I. Ahonen, et. al, "Region of Interest Enabled Learned Image Coding for Machines," MMSP 2023, pp. 1-6, Sep. 2023.

[4] Y. Matsubara, et. al, ‘SC2 benchmark: Supervised compression for split computing,’ Transactions on Machine Learning Research, issn. 2835-8856, Jun. 2023.

[5] Q. Huang L. Shen, Z. Lin, “Relay backpropagation for effective learning of deep convolutional neural networks,” ECCV 2016, pp. 467-482, Oct. 2016.

[6] Z. Cheng, et. al, “Learned image compression with discretized gaussian mixture likelihoods and attention modules,” CVPR 2020, pp. 7939–7948, Jun. 2020.

2023 年度

早稲田大学大学院基幹理工学研究科　情報理工・情報通信専攻　修士論文

補助損失の追加による画像認識のための深層画像符号化の性能改善

Improving Performance of Deep Image Coding for Image Recognition by Adding Auxiliary Loss

飯野　景

（5122F007-6）

提出日：2024 年 1 月 22 日

指導教員：渡辺裕教授

研究指導名：オーディオビジュアル情報処理研究

目次

第１章　序論

１．１　研究の背景

近年画像認識モデルの性能が飛躍的に伸びており，復号した画像や映像を人間が確認することなく認識モデルが分析するケースが増えている．これに伴い，人間の視覚ではなく，モデルの認識に最適化された画像符号化（Image Coding for Machines (ICM)）の研究が盛んになっている. 実際に JPEG では JPEG-AI, MPEG では Video Coding for Machines（VCM）といった標準化活動も行われている．また，近年深層画像圧縮モデルの性能向上が目覚ましいことから，ICM の研究でも頻繁に用いられている．ICM 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り，これを達成するために様々な学習方法[1,2,3,4]，圧縮モデル構造[5,6]などが提案されている．

１．２　関連研究と問題点，および研究目的

ICM 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 ICM の研究のメインアプローチは次の二つの方法に分類される．一つはタスクの損失で圧縮モデルを最適化することである[1,2,4]．タスクの損失で学習を行うことで，圧縮モデルは後段の認識モデルの精度とビットレートのトレードオフを最適化することができる．もう一つは関心領域（Region of Interest (ROI)）ベースのビット割り当て方式[5,6]である．これは前景領域などに優先的にビット割り当てを行うことでタスク志向の符号化器を設計する．しかしこれらの方法はそれぞれ問題点が存在する．タスク損失での学習においては，タスクの難易度やモデルの複雑さに応じて十分な最適化が困難になる場合がある[7,8,9]．特に認識モデルのネットワークが深い場合，誤差逆伝搬の性質上最適化が困難になる[10]. ROI ベースのビット割り当て方式の場合，ROI を決定するための認識モデルが必要となるため，エンコーダ側に追加のオーバーヘッドが生じてしまう．また，セマンティックセグメンテーションやパノプティックセグメンテーションなどの背景のクラス分類も行うタスクの場合や，画像に対する説明を行うイメージキャプショニングタスクでは ROI を定義することが難しい． そこで本研究では ICM の学習方法に着目し，以上の問題点の解決を目指す．具体的には圧縮モデルの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）を設けてタスクの損失関数で最適化を行う．これにより，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向上を実現し，R-D 性能の改善を目指す．

１．３　本論文の構成

以下に本論文の構成を示す.

第１章　本研究における背景，目的について述べる.

第２章　本研究に関連する技術について述べる.

第３章　予備実験として, 従来の ROI ベースのビット割り当て方式の効果を物体検出およびセマンティックセグメンテーションのタスクで検証した.

第４章　本研究で提案する補助損失を追加した学習手法について述べる.

第５章　本研究の実験内容およびその結果と考察を述べる.

第６章　本研究における結論と今後の展望について述べる.

第２章　　関連研究

２．１　　まえがき

本章では，本研究における関連技術として，深層画像圧縮と ICM に関する代表的な手法について述べる．

２．２　　深層画像圧縮

深層学習モデルで画像圧縮を行う技術であり，深層画像圧縮(deep image compression)や学習型画像圧縮（learned image compression）と呼ばれる．近年，その圧縮性能が大幅に向上しており，従来の非学習型圧縮器である HEVC や VVC の性能を凌ぐことから注目を集めている[11]．圧縮モデルは，画像の再構成品質の最大化とビットレートの最小化を学習するために以下の損失関数で最適化が行われる

$$L = R + \lambda \cdot D(x, \hat{x}) \tag{1}$$

ここで，$x, \hat{x}$は入力画像および復号画像，Rはビットレート，Dは歪み（再構成誤差）を表す．また，λはラグランジュ乗数であり，値を変動させることでRとDのトレードオフを調整できる．図 2.1 に深層画像圧縮の概要図を示す．

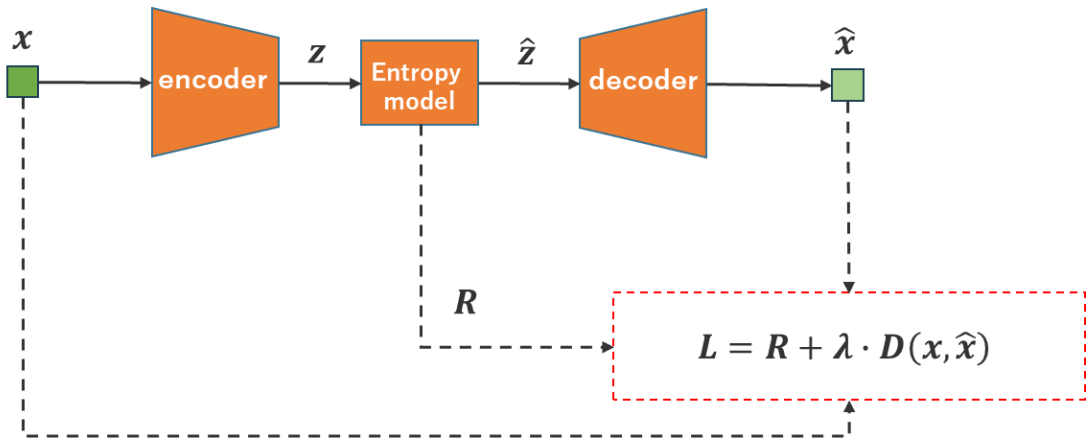

図 2.1　深層画像圧縮の概要

２．３　Image Coding for Machines (ICM)

ICM とは人間の視覚ではなく，モデルの認識に最適化された画像符号化の技術であり，大量のデータが伝送され解析される現代において，その実用化が期待されている．一般的に，復号された画像を人が確認する場合，画像全体の視覚的情報が元の画像と類似していることが求められる．一方，人ではなく認識モデルの入力として用いる場合，復号画像と元画像が視覚的に類似している必要はなく，認識精度の低下を最小化することが求められる．そのため ICM では，精度低下とビットレートの最小化を行うために，認識タスクに必要な情報を抽出して圧縮することが理想的と考えられる．多くの既存研究は，圧縮モデルの学習方法やモデル構造に注目して，この機能の実現を目指している．これらの代表的な手法として，認識モデルの損失関数を用いた最適化[1,2,4]，および ROI ベースのビット割り当て方式[5,6]について述べる．

２．３．１　認識モデルの損失関数を用いた最適化

Harell らは，ICM における R-D 理論を分割層と蒸留層の側面から分析した[1]．ICM では図 2.2 に示すように，以下の三つのアプローチに分類できる．

A) Full-input coding：画像を圧縮し伝送する．復号された画像が認識モデルに入力される．

B) Model-Splitting coding：認識モデルを中間層で分割し，前半の出力特徴量を圧縮し伝送する．復号された特徴量が，後半の認識モデルに入力される

C) Direct coding：入力画像を圧縮し伝送する．画像ではなく特徴量として復号され，Model-Splitting 同様に認識モデルの中間層に直接入力される．

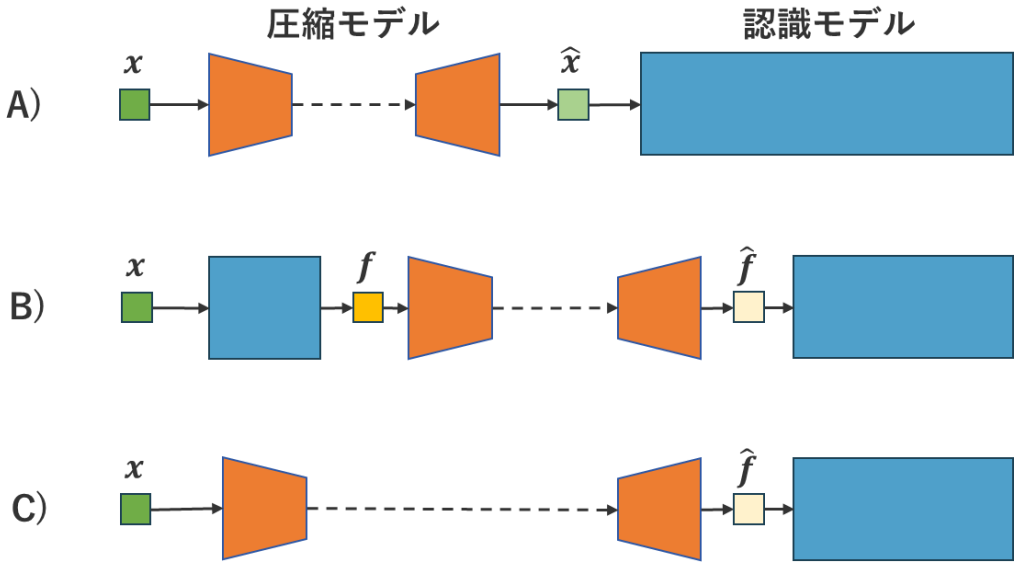

図 2.2　ICM における三つの符号化アプローチ

分割層とは図 2.3 に示す通り，Model-Splitting coding および Direct coding において認識モデルを分割する層のことである．また，ICM では認識精度の低下を抑えるために式(1)の歪みDを，認識モデルのタスク損失[1,2,4]あるいは特徴量の蒸留損失[3]に置き換えるケースが一般的である．

$$L = R + \lambda \cdot L_{task} \tag{2}$$

$$L = R + \lambda \cdot D(f_{distill}, \hat{f}_{distill}) \tag{3}$$

ここで，L_{task}は認識モデルにおけるタスク損失を表す．$f_{distill}, \hat{f}_{distill}$はともに認識モデルの特徴量を示し，$\hat{f}_{distill}$はデータ圧縮を伴う．特徴量蒸留の場合，蒸留に使用する中間層，すなわち蒸留層の選択が可能である（図 2.3）．Harell らは，理論的には分割層の選択は R-D 性能に影響せず，蒸留層の選択のみが R-D 性能に影響することを証明し，実験によりその妥当性を示した．また，蒸留層は深いほど R-D 性能が向上し，最も深い層で計算されるタスク損失が最適であると主張した．また，Yamazaki らは Model-Splitting coding における圧縮モデルの学習において，式（2）のL_{task}で使用される正解ラベルを，圧縮を伴わないモデルの出力に置き換えることを提案した[4]．実世界において学習対象のデータに対する正解ラベルが手に入ることは少ないため，この方法を採用することで，実世界においてもタスク損失による圧縮モデルの学習が可能になる．

Harell らは，理論的には分割層の選択は R-D 性能に影響しないことを示した．しかし分割層が浅い場合と深い場合とでは，学習時に誤差伝搬を行う層数が異なる．一般的に誤差伝搬時に通過する層数が増えるほど最適化が難しくなることが知られている[10]．そのため後段の認識モデルが大規模で，Full-input coding や浅い分割層での Model-Splitting/Direct-coding を用いる場合，標準的な学習方法では圧縮モデルが十分に最適化されない可能性がある[7,8,9]．

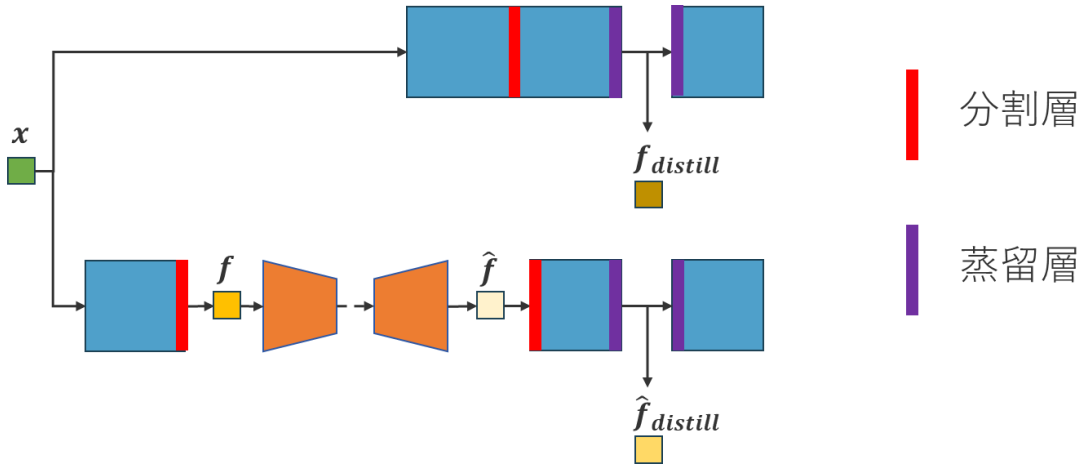

図 2.3　分割層および蒸留層の概要

２．３．２　ROI ベースのビット割り当て

ICM では，符号化前と符号化後のデータが視覚的に類似している必要はなく，データ圧縮に伴う認識精度の低下を抑えることが求められる．多くのコンピュータビジョンタスクにおいて，画像内の背景領域は前景領域と比較してタスク精度に与える影響は小さい[2,5,6]．そこで前景領域を ROI としてより多くのビットを割り当て，ROI 外である背景領域のビット量を削減するための手法が多く提案されている[5,6]．Li らは ROI マスク生成にセグメンテーションモデルを使用し，エンコーダの出力である潜在表現に対して，ROI マスクを元にビット割り当てを行う方式を提案した[5]．本手法では，0~1 の連続値である 2 次元の ROI マスクM_{hw}から，チャネル次元Cに拡張した 3 次元の二値 ROI マスクM_{chw}を生成する．具体的には，$C = 128$でピクセル値が 1 の場合，128 チャネル全てが 1 となり，0.5 の場合 64 チャネルが 1 で残りの 64 チャネルが 0 となる．このようにして生成されたM_{chw}を潜在表現zに乗算することで ROI ベースのビット割り当てを実現する．Ahonen らは ROI マスク生成に物体検出モデルを使用し，ROI 外のピクセルの分散を低減し，割り当てビット数を削減する方法を提案している[6]．まず本手法では，0~1 の連続値からなる ROI マスクM_{hw}に対して閾値処理を行い，二値 ROI マスク$\tilde{M}_{hw}$を作成する．次に潜在表現z内の非 ROI 領域に対して 1 より大きい値で除算することで，非 ROI 領域の分散を抑制でき，割り当てビット数が削減される．

これらのアプローチの多くが，ROI マスクを得るためにエンコーダ側に追加のネットワークを要する（図 2.4）．一般的に ICM のユースケースにおいて，エンコーダのリソースはデコーダに比べてはるかに小さいため，推論時にエンコーダ側に追加のオーバーヘッドが生じることは好ましくない．また，認識タスクやモデルによって画像内のどの領域が重要かどうかは変化するため自明ではない．特に，背景のクラス分類も行うセマンティックセグメンテーション，パノプティックセグメンテーションタスクや，画像に対する説明を行うイメージキャプショニングタスクでは ROI 領域が定義できないため，適用が困難である考えられる．

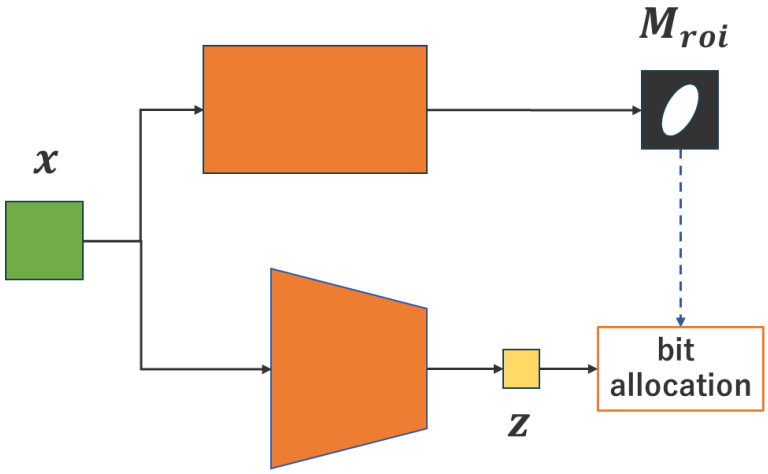

図 2.4　ROI ベースのビット割り当て方式（画像上部の認識モデルで ROI マスクを推論）

２．４　むすび

本章では，本研究における関連技術として，深層画像圧縮と ICM に関する代表的な手法について述べた．代表的な手法として，タスクモデルの損失関数を用いた最適化，および ROI ベースのビット割り当て方式を挙げ，それらの問題点について考察した．

第３章　予備実験

３．１　まえがき

本章では，第 2 章で述べた ROI 方式の問題点を確認するための予備実験を行う．

３．２　実験内容

Ahonen らの研究[6]を参考にし，物体検出およびセマンティックセグメンテーションのタスクにおいて，ROI 方式の効果を確認するための実験を行う．図 3.1 に ROI マスクと復号画像の例を示す．下段は赤枠の領域の拡大画像を示しており，ROI 外の歪みが大きくなっていることがわかる。

３．２．１　物体検出タスク

圧縮モデルには CompressAI [12]で提供されている事前学習済みの bmshj2018_factorized [13]を用い，物体検出モデルには Fater-RCNN（Resnext101-FPN）[14]を使用した. Ahonen らの方式は ROI マスクの抽出に物体検出モデルを用いたが，今回の実験では真値のインスタンスセグメンテーションマップを ROI マスクとして使用した．COCO2017 [15]の評価用データで評価を行った．

３．２．２　セマンティックセグメンテーションタスク

物体検出同様，事前学習済みの bmshj2018_factorized を圧縮モデルとして用い，セグメンテーションモデルには DeepLabv3+（Resnet101）[16]を使用した．評価用のデータセットとして Pascal-Context 59 [17]を使用した. Pascal-Context 59 は一部の僅かな領域（Ignore region）を除いて，画像内の全ての領域にクラスが割り当てられている．そのため ROI マスクとして，a) 前景クラス*の領域，b) Ignore region 以外の領域の二種類で実験を行った．

*Pascal-Context 59 のクラスのうち Pascal VOC [18]で定義されているクラスを前景クラスとした．

３．３　実験結果および考察

図 3.2 に物体検出タスク，図 3.3，3.4 にセマンティックセグメンテーションタスクにおける実験結果を示す．それぞれの図は各手法の R-D 曲線を示しており，横軸が入力画像の 1 ピクセルあたりのビット数（bpp, bits per pixel），縦軸が各認識タスクの精度を表している．青色の実線がベースラインである bmshj2018_factorized，黄色の実線が ROI 方式を適用したもの，破線が 0,1 を反転した ROI マスクを適用したものである．図 3.2 からわかる通り，物体検出では背景領域のビット量を削減することで R-D 性能が改善し，前景領域のビット量を削減することで R-D 性能が劣化している．すなわち，ROI 方式が有効であることがわかる.一方，図 3.3，3.4 からわかる通りセマンティックセグメンテーションでは a, b 共に ROI 方式の効果は確認できない．a における結果から，背景と前景共に分類クラスの対象であるため，その重要度は等価であることがわかる．b ではデータセット内の ignore region の占める割合が小さいことから，ROI 方式の効果が表れなかったためと考えられる．以上の結果から，ROI 方式は適用可能範囲が限られ，ROI 自体の定義が困難なタスクには効果を発揮しないことがわかる．

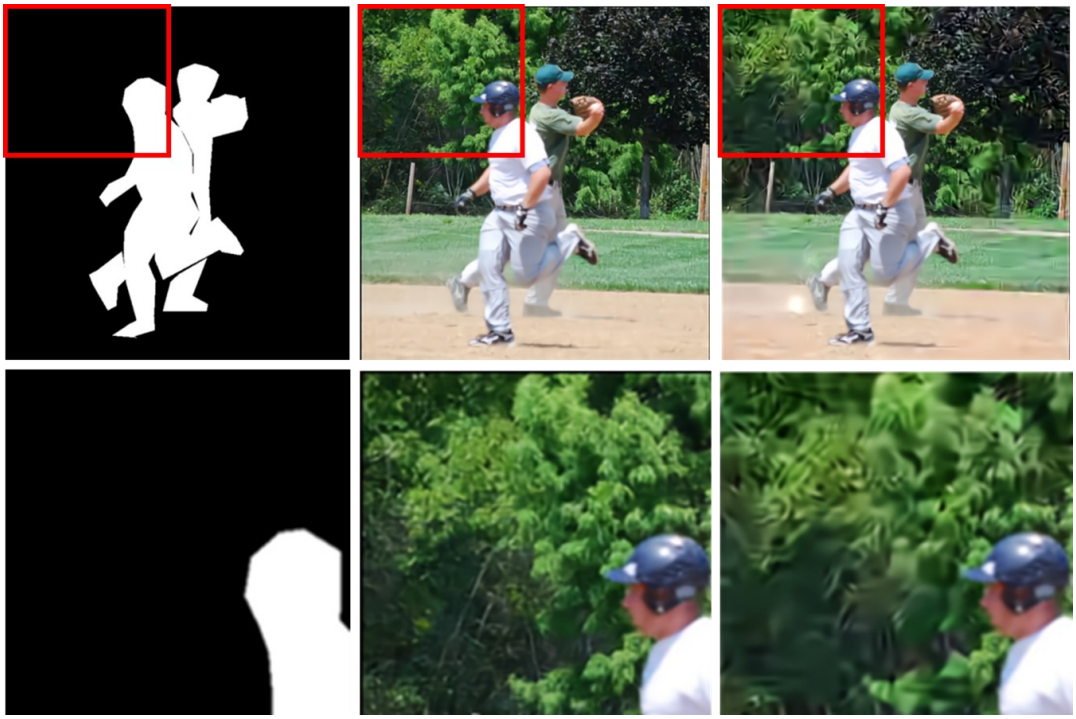

図 3.1　ROI 方式の適用例（左から ROI マスク，通常の復号画像，ROI 方式適用時の復号画像を示す．入力画像および ROI マスクは COCO2017[15]より引用.）

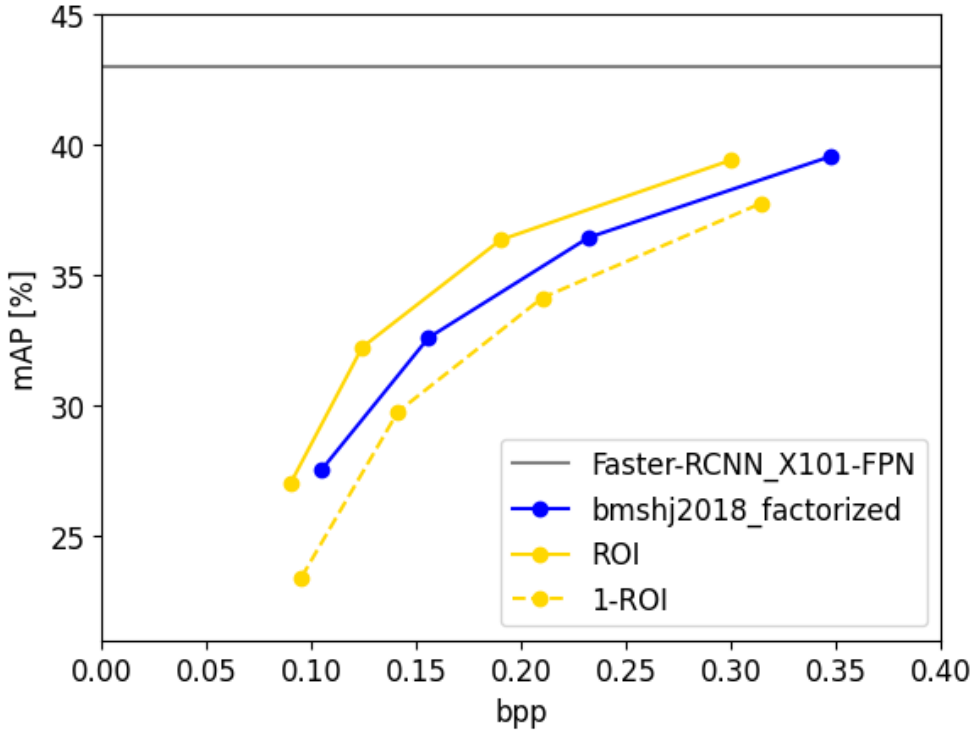

図 3.2　物体検出タスクに対する ROI 方式の適用

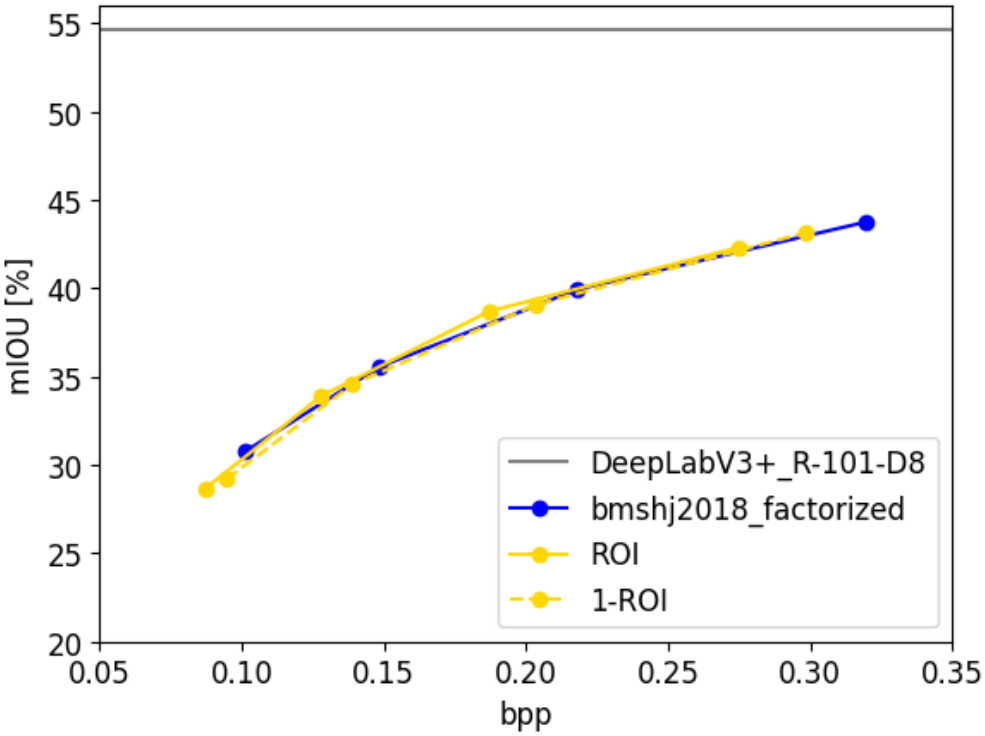

図 3.3　セマンティックセグメンテーションタスクに対する ROI 方式 a（前景クラスの領域）の適用

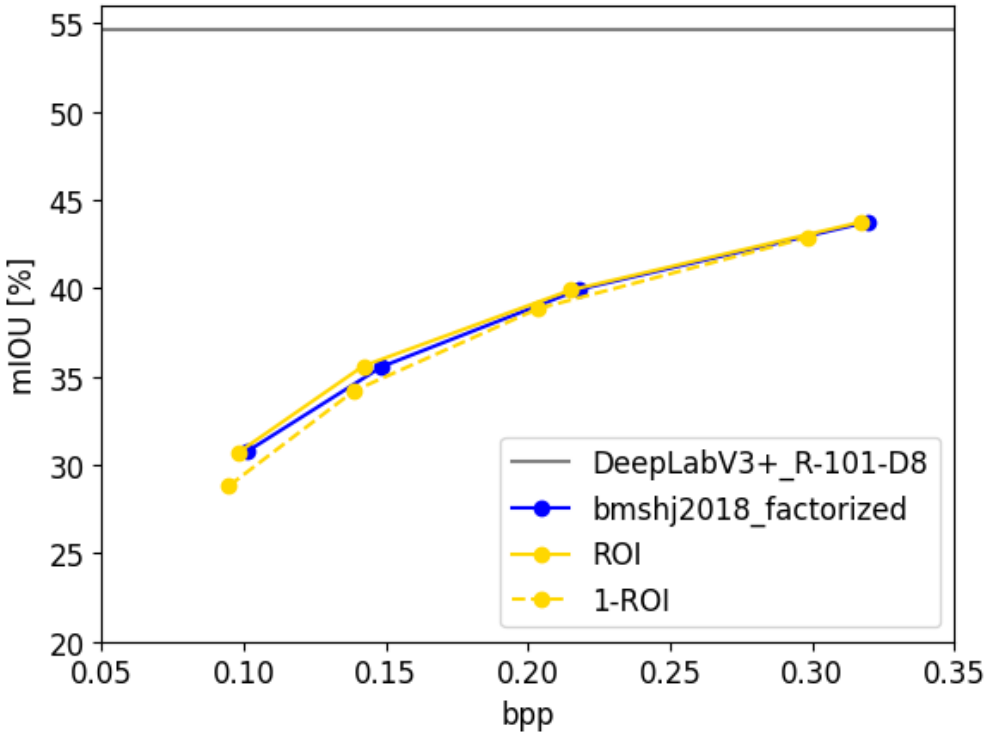

図 3.4　セマンティックセグメンテーションタスクに対する ROI 方式 b（Ignore region 以外の領域）の適用

３．４　むすび

本章では予備実験として，従来の ROI ベースのビット割り当て方式の効果を物体検出およびセマンティックセグメンテーションのタスクで検証した．実験の結果，ROI の定義が直感的にも容易な物体検出のタスクでは，ROI 方式が有効であることを確認した．一方 ROI の定義が困難なセマンティックセグメンテーションのタスクでは，ROI 方式は有効でないことを確認した．

第４章　提案手法

４．１　まえがき

本章では，提案手法として ICM 学習における補助損失の追加について述べる．

４．２　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．そのため多くの研究ではタスクモデルの損失関数を用いた最適化，および ROI ベースのビット割り当て方式を採用している．しかし，タスクモデルの損失関数での学習では，認識モデルが大規模な場合，圧縮モデルの最適化が難しくなることが指摘されている．その結果，エンコーダが十分な識別能力を獲得できない可能性が考えられる．また，ROI 方式の場合，多くの研究が前景領域を ROI として扱う．しかし，認識タスクやタスクモデルにおける ROI 領域は自明ではなく，第 3 章で確認したようにセグメンテーションなどのタスクでは ROI 領域の定義が困難で，適用による効果が得られない場合がある．また，多くの ROI 方式の研究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る．

そこで本研究では，ICM における学習時に補助損失を追加することで，以上の問題点の解決を目指す．提案手法の概要図を図 4.1 に示す．提案手法では，圧縮モデルのエンコーダの後に補助損失計算用の軽量な認識モデル（Auxiliary branch）を設けて圧縮モデルと共同で最適化を行う．損失関数は以下のように定義する．

$$L = R + \lambda \cdot (L_{task} + \alpha \cdot L_{aux}) \tag{4}$$

ここで，L_{aux}は Auxiliary branch の損失，αは重み係数である．また，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を用いる．Auxiliary branch はデコーダの直前から分岐して配置されるため，L_{aux}で計算される誤差はエンコーダに直接伝搬される．これにより，認識モデルが大規模な場合でも，エンコーダが認識タスクに必要な情報を理解することをサポートする．また，誤差伝搬対象をエンコーダのみに絞ることで，ROI 方式のような機能をエンコーダが獲得することが期待でき，メインタスクの学習への悪影響の可能性も低減できると考える．また，提案手法では ROI の定義を必要としないため，様々なタスクへの応用が可能と考えられる．さらに Auxiliary branch は学習時のみ使用し，推論時には使用しないため，従来の ROI 方式のようなオーバーヘッドも生じない．

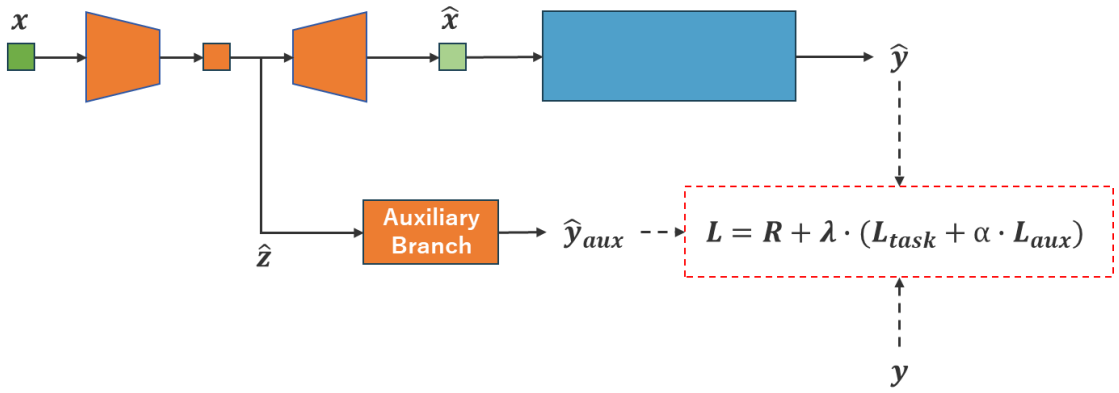

図 4.1　補助損失を追加した ICM 学習手法（$\hat{z}$はデコーダに入る前の潜在表現，yは正解ラベル，$\hat{y}$は復号画像$\hat{x}$を入力したときの認識モデルの出力，$\hat{y}_{aux}$は Auxiliary branch の出力）

４．３　むすび

本章では，提案手法として ICM 学習における補助損失の追加について述べた．

従来のタスク損失による学習では，認識モデルが大規模な場合，圧縮モデルの最適化が難しくなることが指摘されている．その結果，エンコーダが十分な識別能力を獲得できない可能性が考えられる．また，従来の ROI 方式では ROI の定義が難しいタスクには適用が困難であり，エンコーダ側の負荷も問題となる．これらの問題を解決するために，補助損失の追加を提案し，推論時の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 R-D 性能の向上を目指す．

第５章　実験内容及び結果と考察

５．１　まえがき

本章では, 物体検出及びセマンティックセグメンテーションのタスクにおいて提案手法の効果を検証する.

５．２　実験内容

５．２．１　物体検出タスク

圧縮モデルには事前学習済みの cheng2020attn [19]を用い，物体検出モデルには Fater-RCNN（Resnext101-FPN）を使用した．Yamazaki らの研究[4]同様に，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}の正解ラベルに用いて，蒸留ベースの学習を行った．データセットには COCO2017 を使用し，最適化関数には Adam を使用し 50 エポックの学習を行った．初期学習率は 0.0001 に設定し，最後の 10 エポックは 0.1 倍の減衰を行った．Auxiliary branch の全体の構成は認識モデルである Faster-RCNN を参考にし，バックボーンモデル（ResNext101）を 18 層の小さなネットワークに変更した．また式(1)の α は 0.5 に設定した.

５．２．２　セマンティックセグメンテーションタスク

圧縮モデルには事前学習済みの cheng2020attn を用い，セグメンテーションモデルには DeepLabv3+（ResNet101）を使用した．物体検出同様に，未圧縮の画像を入力としたときの DeepLabv3+の出力をL_{task}，L_{aux}の正解ラベルに用いて，蒸留ベースの学習を行った．また，学習の安定化を図るため，式(4)に式(1)における歪みDを加えて圧縮モデルの最適化を行った. データセットには Pascal-Context 59 を使用し, 最適化関数には Adam を使用し 50 エポックの学習を行った．初期学習率は 0.0001 に設定し，Polynomial decay に従い学習率の減衰を行った．Auxiliary branch の全体の構成は認識モデルである DeepLabv3+参考にし, 物体検出同様, 18 層のバックボーンモデルに変更した．また式(1)の α は 0.1 に設定した.

５．３　実験結果および考察

図 5.1 に物体検出，図 5.2 にセマンティックセグメンテーションタスクにおける R-D 性能の比較結果を示す．それぞれの図からわかる通り，補助損失の追加により R-D 性能が改善していることがわかる．物体検出では平均 27.7%，セマンティックセグメンテーションでは平均 20.3%の BD-Rate を達成した．また物体検出では，予備実験で使用した ROI 方式をベースラインに適用した．ROI 方式と比較しても提案手法が高い R-D 性能を誇っていることがわかる．

定性的な比較を行うために，圧縮された潜在表現に対する空間的なビット割り当て量を可視化し，比較する．物体検出，およびセマンティックセグメンテーションにおける可視化結果と認識結果を図 5.3，5.4 に示す．ビット割り当てマップは 0~1 に正規化してあり，明るい領域がより多くのビットが割り当てられていることを示す．図 5.3 からわかる通り，物体検出タスクでは物体領域に対して多くのビットが割り当てられている．さらに，提案手法を適用することでそれが顕著になっていることが確認できる．セグメンテーションタスクでは，図 5.4 のように認識結果が異なる場合でも，補助損失の有無によるビット割り当て量の違いは視覚的には確認できなかった．これはセグメンテーションタスクにおいて ROI 方式の適用が難しいという事実と一致する．

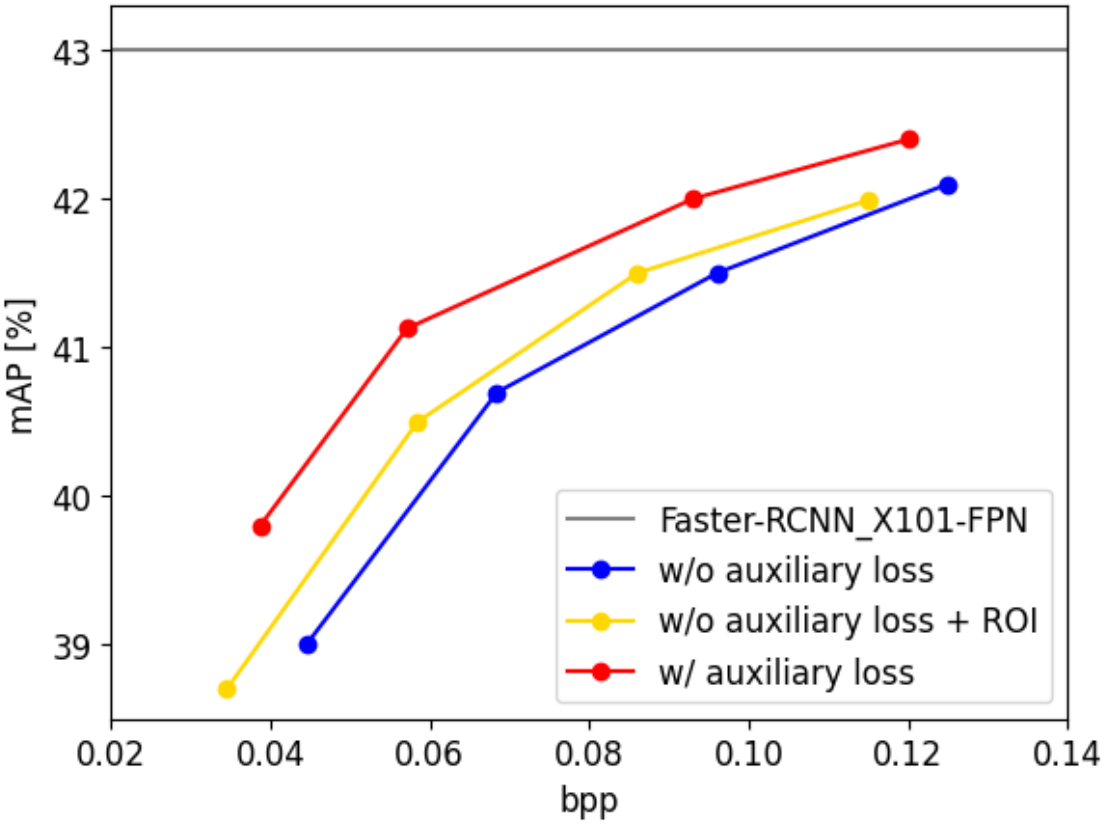

図 5.1　物体検出タスクにおける各手法の R-D 曲線の比較

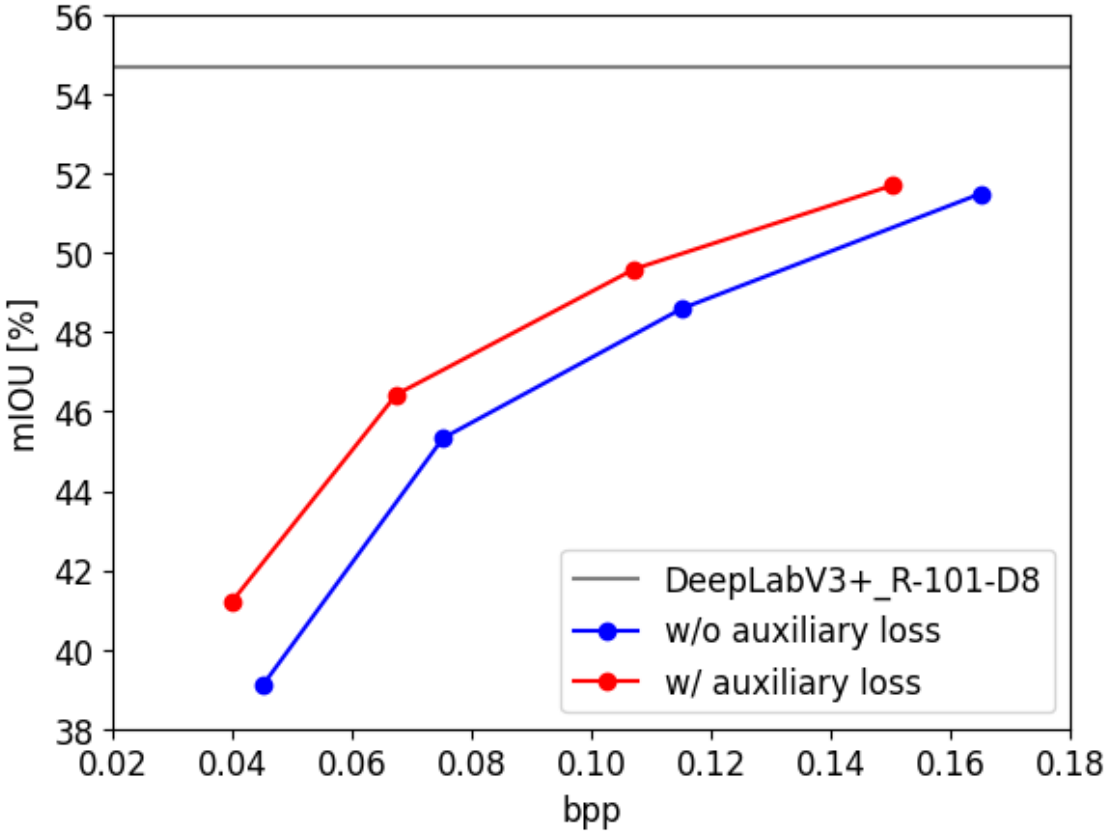

図 5.2　セマンティックセグメンテーションタスクにおける各手法の R-D 曲線の比較

図 5.3　空間的なビット割り当て量の可視化と，検出結果の比較（上段は入力画像，ビット割り当てマップ，下段は検出結果を復号画像に表示したものを示す．入力画像は COCO2017[15]より引用.）

図 5.4　空間的なビット割り当て量の可視化と，検出結果の比較（上段は入力画像，ビット割り当てマップ，下段はセグメンテーション結果を復号画像に表示したものを示す．入力画像は Pascal-Context 59[17]より引用.）

５．４　むすび

本章では，提案手法の有効性を確認するための評価実験の内容と結果，および考察について述べた．

提案手法は補助損失を追加しない従来の学習手法と比較して，高い R-D 性能を達成することを確認した．また，潜在表現に対する空間的なビット割り当て量を可視化することにより，エンコーダの識別能力が向上していることを定性的に確認した．

第６章　結論と今後の展望

６．１　結論

本研究では ICM の学習における補助損失の追加を提案した．提案手法では，エンコーダの後に Auxiliary branch を設け同時に最適化を行うことで，認識タスクに対するエンコーダの認識力獲得を補助する．物体検出，セマンティックセグメンテーションのタスクで実験を行い，各手法の評価を行った．提案手法は補助損失を使用しない従来法と比較して，物体検出で平均 27.7%，セマンティックセグメンテーションで平均 20.3%の BD-Rate の改善を達成した．また，圧縮後の潜在表現のビット割り当量を比較することで，特に物体検出におけるエンコーダの認識能力向上を定性的に確認した．

６．２　今後の展望

ICM では認識タスクに必要な情報を抽出して圧縮することが重要である．そのため多くの研究ではタスクモデルの損失関数を用いた最適化，および ROI ベースのビット割り当て方式を採用している．しかし，タスク損失のみで学習する場合，後段の認識モデルが大規模な場合，誤差逆伝搬の性質からエンコーダの識別能力獲得が難しくなる場合がある．また，ROI 方式の場合，多くの研究が前景領域を ROI として扱う．しかし，認識タスクやタスクモデルにおける ROI 領域は自明ではなく，セグメンテーションなどのタスクでは ROI 領域がそもそも定義できない場合もある．また，多くの ROI 方式の研究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るため，実社会への応用に適さない．

本研究では，圧縮モデルのエンコーダに補助損失を課すことで，上記の問題点の克服を目指した．特に提案手法では ROI の定義を必要としないため，セマンティックセグメンテーション以外でも，イメージキャプショニングなどのマルチモーダルタスへの応用も期待できる．また，既存の ROI 方式と異なり，推論時にリソースの少ないエンコーダ側に追加のオーバーヘッドが生じないため，実社会への応用にも適していると考えられる．

謝辞

本研究に際して，親切かつ的確なご指導をしてくださり，実験環境および充実した研究環境を与えてくださった渡辺教授に深く感謝申し上げます.

また，共に研究をして下さり，貴重な知見と示唆を与えていただいた NTT 株式会社の江田毅晴様，榎本昇平様，坂本啓様，史旭様，森永一路様（2022 年当時）に心からの謝意を表します.

さらに，日常的に有益な意見を与えていただき，研究室での心地よい雰囲気を作ってくださった渡辺研究室の皆様に感謝いたします.

最後に，日頃から体調面でも精神面でも支えていただいた家族に対して，深い感謝の意を表します.

参考文献

[1] A. Harell, Y. Foroutan, N. Ahuja, P. Datta, B. Kanzariya, V. S. Somayaulu, O. Tickoo, A. de Andrade, and I. V. Bajic, “Rate-distortion theory in coding for machines and its application,” arXiv preprint arXiv:2305.17295, May. 2023.

[2] N. Le, H. Zhang, F. Cricri, R. Ghaznavi-Youvalari, and E. Rahtu, “Image coding for machines: an end-toend learned approach,” 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1590–1594, Jun. 2021.

[3] A. Harell, A. De Andrade, and I. V. Bajic, “Rate-distortion in image ´ coding for machines,” 2022 Picture Coding Symposium (PCS), pp. 199–203, Dec. 2022.

[4] M. Yamazaki, Y. Kora, T. Nakao, X. Lei and K. Yokoo, "Deep Feature Compression using Rate-Distortion Optimization Guided Autoencoder," 2022 IEEE International Conference on Image Processing (ICIP), pp. 1216-1220, Oct. 2022.

[5] 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.

[6] J. I. Ahonen, N. Le, H. Zhang, F. Cricri and E. Rahtu, "Region of Interest Enabled Learned Image Coding for Machines," 2023 IEEE International Workshop on Multimedia Signal Processing (MMSP), pp. 1-6, Sep. 2023.

[7] Y. Matsubara, R. Yang, M. Levorato, and S. Mandt, ‘SC2 benchmark: Supervised compression for split computing,’ Transactions on Machine Learning Research, issn. 2835-8856, Jun. 2023.

[8] Y. Matsubara, D. Callegaro, S. Baidya, M. Levorato, and S. Singh, “Head Network Distillation: Splitting Distilled Deep Neural Networks for Resource-Constrained Edge Computing Systems,” IEEE Access, vol. 8, pp. 212177-212193, Nov. 2020.

[9] Y. Matsubara, D. Callegaro, S. Singh, M. Levorato, and F. Restuccia, “BottleFit: Learning Compressed Representations in Deep Neural Networks for Effective and Efficient Split Computing,” 2022 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 337–346, Jun. 2022.

[10] Q. Huang L. Shen, Z. Lin, “Relay backpropagation for effective learning of deep convolutional neural networks,” 2016 European Conference on Computer Vision (ECCV), pp. 467-482, Oct. 2016.

[11] J. Liu, H. Sun, and J. Katto, “Learned Image Compression with Mixed Transformer-CNN Architectures,” 2023 The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14388-14397, Jun. 2023.

[12] J. Begaint, F. Racape, S. Feltman, and A. Pushparaja, “CompressAI: a pytorch library and evaluation platform for end-to-end compression research,” arXiv preprint arXiv:2011.03029, Nov. 2020.

[13] J. Balle, D. Minnen, S. Singh, S. Hwang, and N. Johnston, “Variational image compression with a scale hyperprior,” 2018 The International Conference on Learning Representations (ICLR), Apr. 2018

[14] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” 2015 Advances in Neural Information Processing Systems 28 (NIPS), Vol. 1, pp. 91-99, Dec. 2015.

[15] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, L. Zitnick, and P. Dollar, “Microsoft COCO: Common Objects in Contex t,” 2014 European Conference on Computer Vision (ECCV), pp. 740-755, Sep. 201 4.

[16] L. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” arXiv preprint arXiv:1706.05587, Jun. 2017.

[17] R. Mottaghi, X. Chen, X. Liu, N. Cho, S. Lee, S. Fidler, R. Urtasun and A. Yuille, “T he Role of Context for Object Detection and Semantic Segmentation in the Wild,” 201 4 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 891-898, Jun. 2014.

[18] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn and A. Zisserman, “The P ASCAL Visual Object Classes Challenge 2012 (VOC2012) Results,” http://www.pasc alnetwork.org/challenges/VOC/voc2012/workshop/index.html.

[19] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, “Learned image compression with di scretized gaussian mixture likelihoods and attention modules,” 2020 The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7939–7948, J un. 2020.

図一覧

研究業績

[1] K. Iino, M. Takahashi, H. Watanabe, I. Morinaga, S. Enomoto, X. Shi, A. Sakamoto, and T. Eda: “Inter-Feature-Map Differential Coding of Surveillance Video,” IEEE Global Conference on Consumer Electronics (GCCE), pp.293-296, Oct. 2022

[2] M. Takahashi, K. Iino, H. Watanabe, I. Morinaga, S. Enomoto, X. Shi, A. Sakamoto, T. Eda: “Category-based Memory Bank Design for Traffic Surveillance in Context R-CNN,” International Workshop on Advanced Image Technology (IWAIT2023), No.42, pp.1-4, Jan. 2023. (Best Paper Award)

[3] 飯野景, 高橋美穂, 渡辺裕, 江田毅晴, 榎本昇平, 坂本啓, 史旭, 森永一路, “特徴量圧縮モデルへの注意機構導入の検討,” 電子情報通信学会総合大会，Mar. 2023.

[4] S. Akamatsu, K. Iino, H. Watanabe, S. Enomoto, X. Shi, A. Sakamoto and T. Eda, “A Video Object Detection Method of ECNet Based on Frame Difference and Grid Cell Confidence,” 2022 IEEE Global Conference on Consumer Electronics (GCCE), pp.364-367, Oct. 2023

[5] K. Iino, S. Akamatsu, H. Watanabe, S. Enomoto, A. Sakamoto and T. Eda, “Image Coding for Machines with Objectness-based Feature Distillation,” 2024 IIEEJ International Conference on Image Electronics and Visual Computing (IEVC), Mar. 2024. (2024 年 3 月掲載予定)

[6] S. Akamatsu, K. Iino, H. Watanabe, S. Enomoto, A. Sakamoto and T. Eda, “Edge-Cloud Collaborative Object Detection Model with Feature Compression,” 2024 IIEEJ International Conference on Image Electronics and Visual Computing (IEVC), Mar. 2024. (2024 年 3 月掲載予定)

[7] 飯野景, 高橋美穂, 渡辺裕, 江田毅晴, 榎本昇平, 坂本啓, 史旭, 森永一路, “画像認識向け深層画像圧縮における補助損失の導入,” 電子情報通信学会総合大会，Mar. 2024. (2024 年 3 月掲載予定)