

卒業論文概要書

Summary of Bachelor's Thesis

Date of submission: 01/30/2024 (MM/DD/YYYY)

学科名 Department	情報通信	氏名 Name	赤松俊輔	指導員 Advisor	渡辺 裕 印
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	1W202003-5		
研究題目 Title	フレーム間差分と信頼値によるエッジ・クラウド協調型物体検出システムの構築 Development of Edge-Cloud Cooperative Object Detection System Based on Inter-Frame Differences and Confidence Values				

1. まえがき

近年、物体検出アルゴリズムの飛躍的な向上により、リアルタイムの映像処理が進歩している。これにより、様々なエッジデバイスを用いた動画に対する物体検出の需要が高まっている。そのため、限られた時間内にエッジ側で物体検出を行う必要があるが、エッジデバイス上の限られた計算資源では、処理の重い物体検出モデルを用いることは困難である。そこで本研究では、エッジ側とクラウド側で処理を分担し、動画入力に対し、伝送コストと物体検出精度を両立させる手法を提案する。このエッジ・クラウド協調型物体検出システムでは、フレーム間差分の値と、各グリッドセルにおける物体検出モデルの信頼値をクラウド伝送のオフロード基準として用いることで、クラウド側への伝送量の削減を図る。実験により、伝送量が少ない場合に、物体検出精度を維持したまま、伝送コストを削減することができることを示す。

2. 従来手法とその問題点

画像分類タスクにおける、エッジ・クラウド協調型システムとして、Edge-Cloud Net (ECNet)[1] が提案されている。このシステムでは、エッジ側に軽量の画像分類モデル、クラウド側に高性能な画像分類モデルを配置するシステムを想定し、エ

ジ側の認識モデルから出力されるクラス確率のエントロピーに応じて、エッジ側モデルの推論結果を利用するか、クラウド側へデータを伝送して、クラウド側モデルの推論結果を利用するかを定める。しかし、この研究では画像分類タスクにおける協調型システムの有効性を示すにとどまっている。自動運転技術や、監視カメラ解析などのユースケースにおいては、画像分類タスクのみでは不十分であり、特に動画における物体検出タスクに対応した協調型システムが求められる。

3. 提案手法

本研究では、動画入力のためのエッジ・クラウド協調型物体検出システムを提案する。本システムにおける物体検出モデルには、エッジ側に軽量の YOLOv3-tiny [2]、クラウド側に高性能な YOLOv3 [2]を用いる。クラウド側へのデータ伝送量を削減するため、フレーム間差分を用いた推論コントローラ、エッジ側の物体検出モデルの出力の信頼値に応じたマスク処理、マスクされたフレーム画像に対する圧縮機構を実装した。

3.1. フレーム間差分を用いた推論コントローラ

まず、入力動画に対し、各フレーム間の画素値の差分の絶対値を求める。閾値 T_f を用意し、差

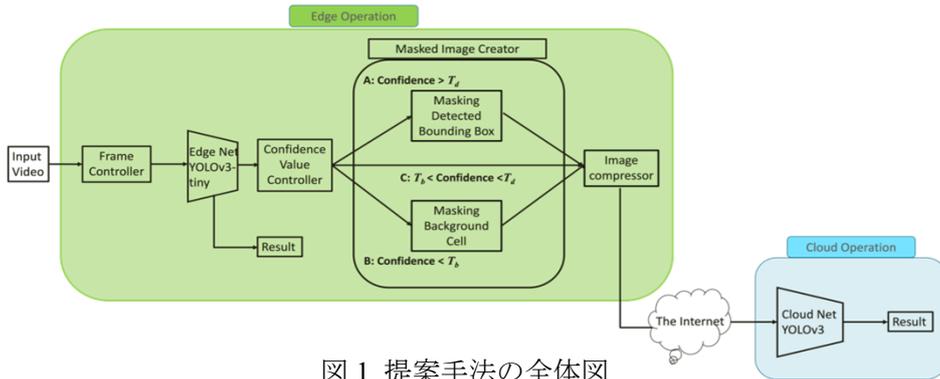


図1 提案手法の全体図

分の値が閾値以内であれば、エッジによる推論をせず、クラウド側へも伝送を行わない。これは、フレーム間差分が小さい場合、現フレームの物体は前フレームで検出された物体と同じである可能性が高いと考えられるためである。これにより、エッジ側での推論回数の削減と、クラウド側へのデータ伝送量の削減を同時に達成することができる。

3.2. 信頼値の算出とマスク画像の生成

エッジ側での推論結果に基づいて、各グリッドセルの信頼値をオフロード基準として使用する。信頼値の範囲は 0 から 1 であり、閾値パラメータとして T_d と T_b を導入する。信頼値 $> T_d$ (A) の場合、エッジ側のモデルは、そのセルの物体を正しく検出できている確率が高いので、クラウド側のモデルで再度推論する必要はない。そこで、エッジ側で検出された物体をマスクし、伝送量の削減を図る。信頼値 $< T_b$ (B) の場合、エッジ側のモデルは、そのセルが物体ではない確率が高いと判断しているため、そのセルをマスクする。これにより、エッジ側で検出された物体と背景部分がマスクされた画像が生成される。フレーム画像に検出物体マスク(A)、背景マスク(B)、両方のマスクを適用したときの一例を図2に示す。 $T_d \leq$ 信頼値 $\leq T_b$ の場合、より良い検出結果を得るために、そのセルはそのままクラウドのモデルへ送られる。

3.3. フレーム画像圧縮

エッジ側の推論結果でマスクされたフレーム画像は、圧縮率の値に応じて JPEG 形式で圧縮される。これにより、クラウド側に伝送するデータ量を削減する。

4. 実験

評価実験では、MOT17 データセット[3]から固定視点映像データを選択して使用した。全フレームにおいて、エッジ側に搭載した YOLOv3-tiny で物体検出した場合の検出精度を基準値とし、エッジ側では画像圧縮のみを行い、クラウド側へ伝送し YOLOv3 で検出した場合の検出精度を比較対象とした。各閾値の値は、実験により決定し、 $T_f = 0.003$, $T_d = 0.65$, $T_b = 0.001$ (Ours1) と、 $T_f = 0.0007$, $T_d = 0.5$, $T_b = 0.001$ (Ours2) の二組を使用した。ク

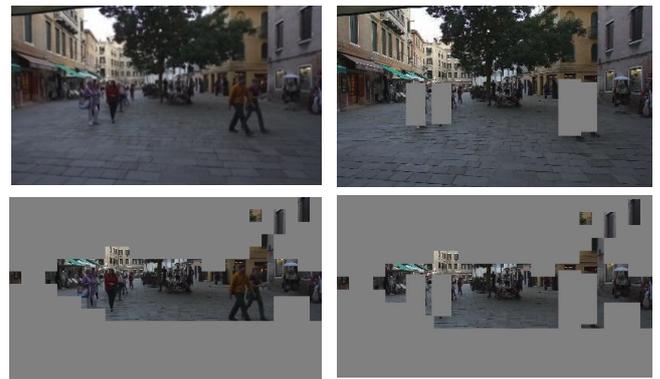


図2 提案手法によるマスク結果
(左上: 元画像, 右上: 検出物体へのマスク(A),
左下: 背景のマスク(B), 右下: 両方のマスク)

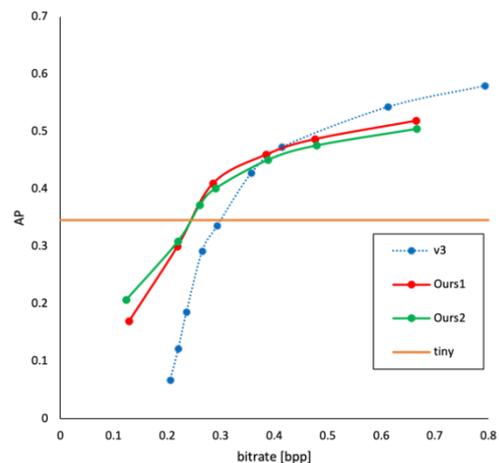


図3 クラウド側へのデータ伝送量と検出精度の関係

クラウド側に伝送するデータサイズ(bpp)と物体検出精度 AP (IoU=0.5) の関係を図3に示す。提案手法は、クラウド側へ伝送するデータ量が少なく、精度が要求される際に有効である。

5. 結論

本研究では、動画に対するエッジ・クラウド協調型物体検出システムを提案した。フレーム間の差分と各グリッドセルの信頼値を用いて伝送量を削減することで、特に低ビットレートで単に全データをクラウドに伝送するよりも有効であることを示した。

参考文献

- [1] L. Hu *et al.*, “ECNet: A Fast, Accurate, and Lightweight Edge- Cloud Network System based on Cascading Structure,” IEEE GCCE, pp. 259-262, Oct. 2020.
- [2] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv:1804.02767, Apr. 2018.
- [3] A. Milan *et al.*, “MOT16: A Benchmark for Multi-Object Tracking,” arXiv:1603.00831, Mar. 2016.

2023 年度 卒業論文

フレーム間差分と信頼値によるエッジ・クラウド協調型
物体検出方式の研究

A Study of Edge-Cloud Cooperative Object Detection Method
Based on Inter-Frame Differences and Confidence Values

指導教員 渡辺 裕 教授

提出日：2024 年 1 月 30 日

早稲田大学 基幹理工学部 情報通信学科

1W202003-5

赤松 俊輔

目次

第1章	序論.....	3
1.1	研究背景.....	3
1.2	関連研究と問題点, および研究目的.....	3
1.3	本論文の構成.....	3
第2章	関連技術.....	5
2.1	まえがき.....	5
2.2	Edge-Cloud Net (ECNet).....	5
2.3	YOLO.....	6
2.3.1	YOLOv3.....	6
2.3.2	YOLOv3-tiny.....	6
2.4	むすび.....	6
第3章	提案手法.....	7
3.1	まえがき.....	7
3.2	提案手法.....	7
3.2.1	フレーム間差分の測定.....	7
3.2.2	グリッドセル信頼値の測定とマスク画像の生成.....	8
3.2.3	フレーム画像圧縮.....	9
3.3	むすび.....	9
第4章	実験・実験結果.....	11
4.1	まえがき.....	11
4.2	データセット.....	11
4.3	評価方法.....	11
4.4	実験結果.....	11
4.4.1	信頼値によるマスクの変化.....	11
4.4.2	JPEG 圧縮による検出精度の変化.....	13
4.4.3	提案手法の性能評価.....	13
4.5	考察.....	14
4.6	むすび.....	15
第5章	結論と今後の課題.....	16

5.1	結論.....	16
5.2	今後の課題.....	16
	謝辞.....	17
	参考文献.....	18
	図一覧.....	19
	表一覧.....	20
	研究業績.....	21

第1章 序論

1.1 研究背景

近年、深層学習を用いた物体検出アルゴリズムの飛躍的な性能向上によって、リアルタイムでの映像処理が急速に進化している。これにより、監視カメラで撮影された映像を解析するなど、エッジデバイスにおける物体検出タスクの需要が高まっている。そのため、限られた非常に短い時間内にエッジ側で物体検出を行うことが求められるが、エッジデバイスの限られたリソースでは、高性能な一方で処理の重い最新の物体検出モデルを導入することは難しい。しかし、豊富なリソースのあるクラウド側のデータセンターなどで処理をする場合には、エッジ側からクラウド側へインターネットを介したデータの伝送が必要になり、伝送時の情報の欠損や遅延が生じるだけでなく、画像や動画に映る情報のプライバシーの問題も生じる。

このような問題点に対処するため、エッジ側とクラウド側の処理を組み合わせたシステムが必要とされている[1]。従来手法として、認識精度が低く、処理が軽いモデルをエッジ側に、認識精度が高い一方で処理が重いモデルをクラウド側に配置し、処理を分担することで、認識精度とデータ伝送量のバランスをとる方式が提案されている。

1.2 関連研究と問題点、および研究目的

コンピュータビジョンタスクに対するエッジ・クラウド協調型認識方式として、画像分類タスク向けの Edge-Cloud Net (ECNet) [2] が提案されている。これは、エッジ側に軽量の画像分類モデル、クラウド側に高性能な画像分類モデルを配置する方式を想定しており、エッジ側の軽量モデルの認識モデルから出力されるクラス確率のエントロピーに応じて、エッジ側モデルの推論結果を利用するか、クラウド側へデータを伝送して、クラウド側モデルの推論結果を利用するかを決定する。しかし、この研究では画像分類タスクにおけるエッジ・クラウド協調型システムの有効性を示すにとどまっている。一方、実際の社会において、自動運転技術や、監視カメラ解析などのユースケースにおける需要に応えるためには画像分類タスクのみでは不十分であり、特に動画における物体検出タスクに対応したシステムが求められる。

そこで、本研究では、動画を入力として、エッジ側の軽量の検出モデルと、クラウド側の高性能な検出モデルを組み合わせた二層のエッジ・クラウド協調型の物体検出方式を提案する。

1.3 本論文の構成

本論文の構成を以下に示す。

第1章 本章であり、本研究の背景、関連研究と問題点および研究目的について述べる。

- 第2章 本研究で用いる従来のエッジ・クラウド協調型認識方式および関連技術について述べる.
- 第3章 本研究の提案手法について述べる.
- 第4章 本研究における実験の方法, 結果および考察について述べる.
- 第5章 結論と今後の課題について述べる.

第 2 章 関連技術

2.1 まえがき

本章では、画像分類タスクに対するエッジ・クラウド協調型認識方式である Edge-Cloud Net (ECNet) について述べる。また、物体検出モデルである YOLO[3] から、本提案手法で用いた YOLOv3[4]、YOLOv3-tiny[5] について述べる。

2.2 Edge-Cloud Net (ECNet)

Edge-Cloud Net (ECNet) は、画像分類タスクに対し、エッジ側に処理が軽く精度の低いモデルを置き、クラウド側に処理が重い精度の高いモデルを置き、エッジとクラウドのモデルを組み合わせることで推論を行う方式である。入力に対して、その情報をエッジ側のみで処理するか、クラウド側に伝送するかどうかは、エッジ側のオフロードコントローラが判断する。この仕組みと伝送時の量子化技術を組み合わせることで、伝送量を減らしながら画像分類精度を維持することができる。先行研究では、エッジ側モデルとして YOLOv3 のバックボーンである Darknet19 を使用し、クラウド側モデルとして同様に YOLOv3 のバックボーンである Darknet53 を用いて ECNet を構築する。また、エッジ側モデルの出力から得られる入力画像のクラス確率のエントロピーをオフロード基準として採用している。ECNet の全体のシステム構造を図 2.1 に示す。

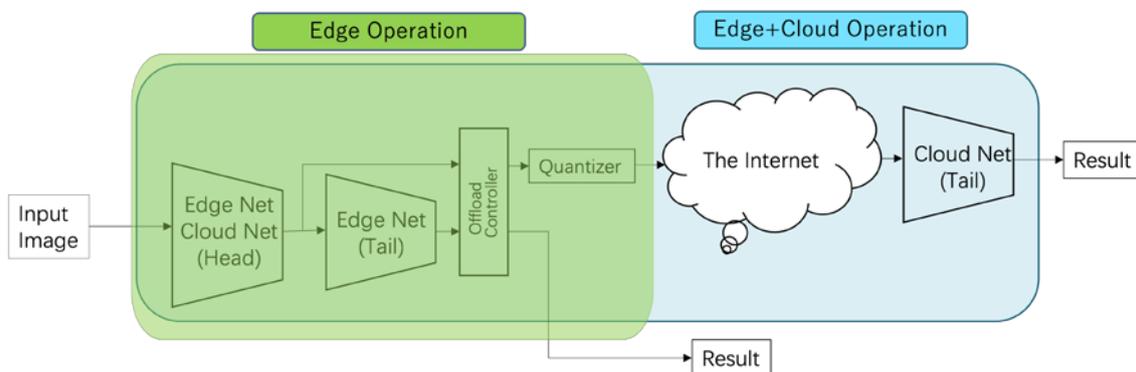


図 2.1 Edge-Cloud Net の構造

ECNet は、画像分類タスクに対する有効性を示した一方で、実際のユースケースとして考えられる動画に対する物体検出への有効性は示されていない。物体検出タスクは画像分類タスクに比べ、複数の物体の検出や、物体の位置情報が求められるため、より複雑なタスクな一方で、監視カメラ解析や自動運転などのユースケースにおいて求められるタスクである。

2.3 YOLO

YOLO (You Only Look Once)は、2016年にJ.Redmonらによって提案されたリアルタイム物体検出アルゴリズムである。物体の位置情報を示すバウンディングボックスと物体のクラス確率の予測を一度に行うエンドツーエンドのニューラルネットワークを利用することで高速な推論を可能にしている。

2.3.1 YOLOv3

YOLOv3は、YOLOシリーズの第三世代モデルであり、以前のYOLOアルゴリズムと比較して、ネットワークのレイヤ数を増やし、検出精度を向上させた物体検出アルゴリズムである。特に、前バージョンのYOLOv2[6]と比較して、YOLOv3は、ネットワークのレイヤーを19層から53層に変更することで精度を向上させただけでなく、特徴ピラミッドネットワーク[7]に似た構造を採用することで、複数のスケールでの検出を可能にする。YOLOv3では、416×416の入力画像サイズに対し、三つの特徴マップのスケール(13×13, 26×26, 52×52)が用意され、様々な大きさの物体の検出を可能にしている。最終的なモデルの出力は、バウンディングボックスの位置情報と各グリッドセルにおける信頼値からなる。

2.3.2 YOLOv3-tiny

YOLOv3-tinyは、YOLOv3に比べ、畳み込み層の数を減らして軽量化したモデルである[8]。そのため、実行速度は大幅に向上した一方で、検出精度は低下する。両モデルの性能比較を表2.1に示す。YOLOv3-tinyは、プーリング層を使用し、畳み込み層の数を減らしている。YOLOv3-tinyは、416×416の入力画像サイズに対し、二つの特徴マップのスケール(13×13, 26×26)が用意され、最終的なモデルの出力は、YOLOv3と同様に、バウンディングボックスの位置情報と各グリッドセルにおける信頼値からなる。

表 2.1 YOLOv3, YOLOv3tiny の性能比較 [9]

Model	Flops [G] (↓)	FPS (↑)	mAP (↑)	Dataset
YOLOv3	140.70	20.00	57.80	COCO
YOLOv3-tiny	05.57	220.00	33.20	COCO

2.4 むすび

本章では、画像分類に対するエッジ・クラウド協調型画像分類システムである Edge-Cloud Net のネットワーク構造について述べた。また、関連技術である YOLOv3, YOLOv3-tiny について述べた。

第3章 提案手法

3.1 まえがき

本章では、エッジ側に YOLOv3-tiny、クラウド側に YOLOv3 を配置した、エッジ・クラウド協調型物体検出方式を提案する。また、クラウド側へのデータ伝送量を削減するため、エッジ側の処理としてフレーム間差分の算出に基づく推論回数の削減、エッジ側モデルの出力における信頼値に応じたマスキング処理、フレーム画像ごとの画像圧縮の三つの手法を提案する。

3.2 提案手法

従来の ECNet は、入力を画像とした画像分類タスクに対するエッジ・クラウド協調型認識方式であった。そこで本研究では、入力を動画としたエッジ・クラウド協調型物体検出方式を提案する。本方式における物体検出にはエッジ側に YOLOv3-tiny、クラウド側に YOLOv3 を用いる。クラウド側へのデータ伝送量を削減するため、フレーム間差分の算出による推論コントローラ、エッジ側の YOLOv3-tiny の出力における信頼値に応じたマスキング処理、マスクされたフレーム画像に対する圧縮機構を導入する。その後、圧縮されたフレーム画像は、インターネット経由でクラウドに伝送され、クラウド側の YOLOv3 によって推論される。全体の構造を図 3.1 に示す。

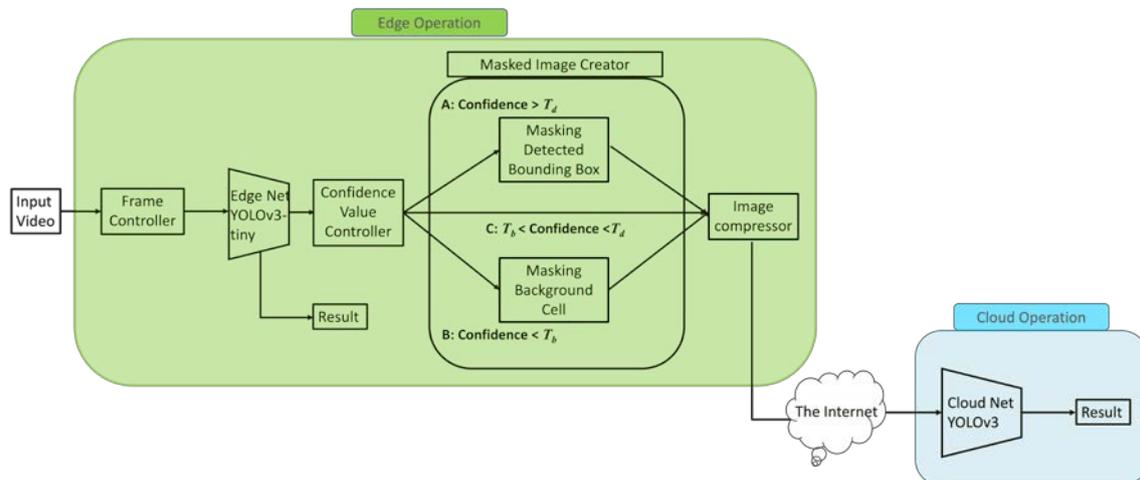


図 3.1 提案方式の全体構造

3.2.1 フレーム間差分を用いた推論コントローラ

まず、入力として受け取った動画に対し、各フレーム間の画素値の差の絶対値を求め、また、閾値 T_f を用意し、閾値以内の値については、エッジによる推論を行わず、クラウド側へも伝送を行わない。これは、差分が小さい場合、現フレームの物体は前フレームで検出された物体と同じである可能性が高いと考えられることを利用し、前フレー

ムの推論結果を現フレームの結果として用いるためである。これにより、エッジ側での推論回数の削減と、クラウド側へのデータ伝送量の削減を同時に達成することができる。

3.2.2 グリッドセル信頼値の算出とマスク画像の生成

エッジでの推論結果に基づき、各グリッドセルの信頼値をオフロード基準として使用する。信頼値は0から1の値を取り、以下の式(3.1), (3.2)で示される[10].

$$p(\text{class}_i) = p(\text{object}) \times p(\text{class}_i|\text{object}) \quad (3.1)$$

$$\text{confidence} = \max(p(\text{class}_i)) \quad (3.2)$$

ここに、 i はセル内のオブジェクトのクラスタイプを表し、 $p(\text{object})$ はセルがオブジェクトを持つ確率、 $p(\text{class}_i)$ はそのオブジェクトがクラス i である確率を表す。したがって、信頼値の値が1に近いほどそのセルに物体が存在していることを示し、0に近いほど、そのセルが背景であることを示す。

ここで、閾値パラメータとして T_d と T_b を導入する。 T_d はそのグリッドセルで物体が検出されたかどうかを判定する閾値として、 T_b はそのセルが背景かどうかを定義する閾値として使用する。また、信頼値 (confidence) が高い場合、すなわち $\text{confidence} > T_d$ の場合 (これを(A)とする)、エッジ側モデルは物体を捉えている確率が高いため、エッジ側のモデルの結果を信頼し、検出された物体に対するバウンディングボックスをマスクする。信頼度の値が低い場合、すなわち $\text{confidence} < T_b$ の場合 (これを(B)とする)、エッジ側モデルはそのセルが画像中の背景部分である可能性が高いと判断しているため、そのセルはエッジ側のモデルの結果を使用し、検出された背景部分をマスクする。そのどちらでもない場合、つまり信頼値が $T_b < \text{confidence} < T_d$ を満たす場合 (これを(C)とする)、そのセルはそのままクラウド側に送られる。これにより、エッジ側モデルで検出されたバウンディングボックスと背景部分がマスクされた画像が生成される。

以下に各グリッドセルの信頼度に応じた処理の一例を図3.2に示す。この図の(A), (B), (C)が信頼値によるコントローラの各処理に相当する。

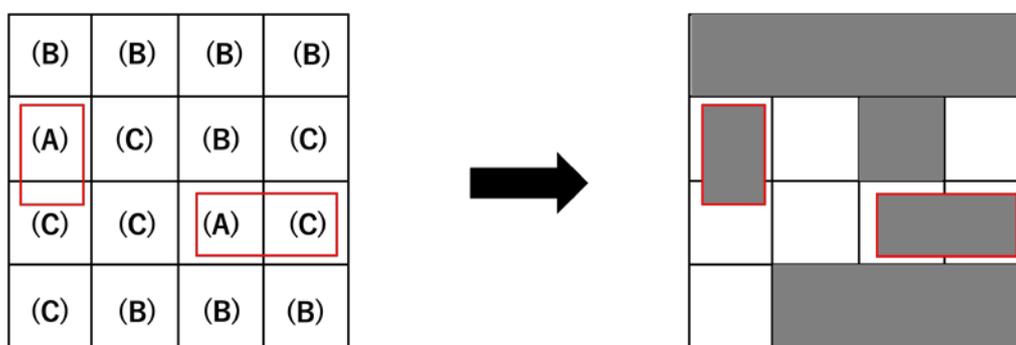


図 3.2 信頼値によるコントローラの処理の一例

3.2.3 フレーム画像圧縮

エッジ側の推論結果によってマスクされた画像に対し、JPEG [11]を用いて、圧縮率の値に応じて圧縮する。これにより、クラウド側に伝送するデータ量を減らすことができる。圧縮率は以下の式(3.3)で算出される。

$$\text{compression ratio} = \frac{I_{\text{after}}}{I_{\text{before}}} \quad (3.3)$$

ここに、 I は画像のデータサイズを表し、 I_{before} は圧縮前のデータサイズ、 I_{after} は圧縮後のデータサイズを表す。圧縮率は処理前と処理後の比率で計算され、値が小さいほどデータ量が削減される。

以下の図3.3にエッジ側のアーキテクチャによる処理の全体アルゴリズムを示す。

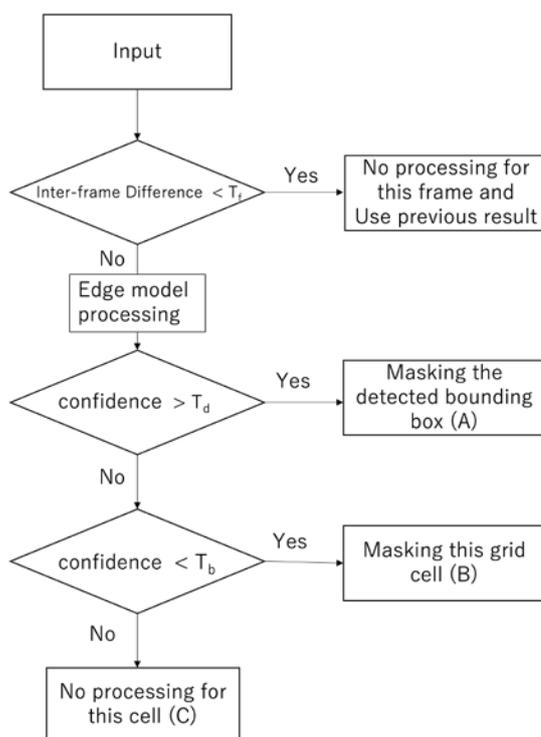


図 3.3 エッジ側処理の全体アルゴリズム

3.3 むすび

本章では、本研究の提案手法である動画に対するエッジ・クラウド協調型の物体検出方式について述べた。具体的には、エッジ側に処理の軽いYOLOv3-tiny、クラウド側に高精度なYOLOv3を置いて物体検出タスクに対する協調型モデルを構成する。また、クラウド側へのデータ伝送量を削減するため、エッジ側に三つの手法を導入する。一つめは、入力動画におけるフレーム間の差分を測定し、差分が小さい場合には、前フレームの結果を再利用することで推論自体の回数を削減する。また、二つめとして、エッジ

側モデルの出力の信頼値に応じて、フレーム画像にマスクをかけ情報量を削減する。最後に三つ目として、伝送前のマスクされたフレーム画像に対して **JPEG** を利用した画像圧縮を行う。これらの機構により、エッジ側モデルの検出結果を利用しながら、クラウド側へのデータ伝送量を減らすことが可能となる。

第4章 実験・実験結果

4.1 まえがき

本章では、本研究で用いたデータセット、実験方法、実験結果、考察について述べる。具体的には、まず本研究の評価時の入力として用いたデータセット、評価方法、提案手法の性能について述べる。その後、実際の処理についての定性的な評価と、結果に対する考察を述べる。

4.2 データセット

評価の入力として、MOT17 (Multiple Object Tracking Benchmark 2017)データセット[12]から固定視点データを選択した。このデータセットの実験での使用許可及び論文への掲載許可は Creative Commons Attribution-ShareAlike 3.0 ライセンスによる。MOT17 データセットは、複数の物体の追跡のためのデータセットであり、動画と動画内の物体のバウンディングボックスのアノテーションが含まれており、固定カメラからの映像と可動カメラからの映像がある。本研究の評価として固定カメラからの映像に限定した理由は、各フレーム間差分を正しく算出できるようにするためである。

4.3 評価方法

MOT17 データセットのうち固定視点データを評価用データセットとし、提案手法に対してクラウド側へのデータ伝送量と物体検出精度の関係を求める。全てエッジ側の YOLOv3-tiny による物体検出を行った場合を基準とし、エッジ側で画像圧縮のみを行い、すべてのデータをクラウド側へ伝送してクラウド側の YOLOv3 による物体検出を行った場合を比較対象とした。また、各パラメータにおいては、表 4.1 に示すように実験的に設定した。これらを考慮して、閾値は $T_f=0.003$, $T_d=0.65$, $T_b=0.001$ (Ours1), $T_f=0.0007$, $T_d=0.50$, $T_b=0.001$ (Ours2)とした。

表 4.1 各閾値の値の設定

	T_f	T_d	T_b
Ours1	0.0030	0.65	0.001
Ours2	0.0007	0.50	0.001

4.4 実験結果

4.4.1 信頼値によるマスクの変化

画像上のマスクの位置は、エッジ側モデルの出力である信頼値によって変化する。図 4.1 は、MOT17 のフレームにバウンディングボックスマスク(A)、背景マスク(B)、両

方のマスクを適用したデータである。また、様々な種類の画像フレームにマスクを適用した結果を図4.2に示す。

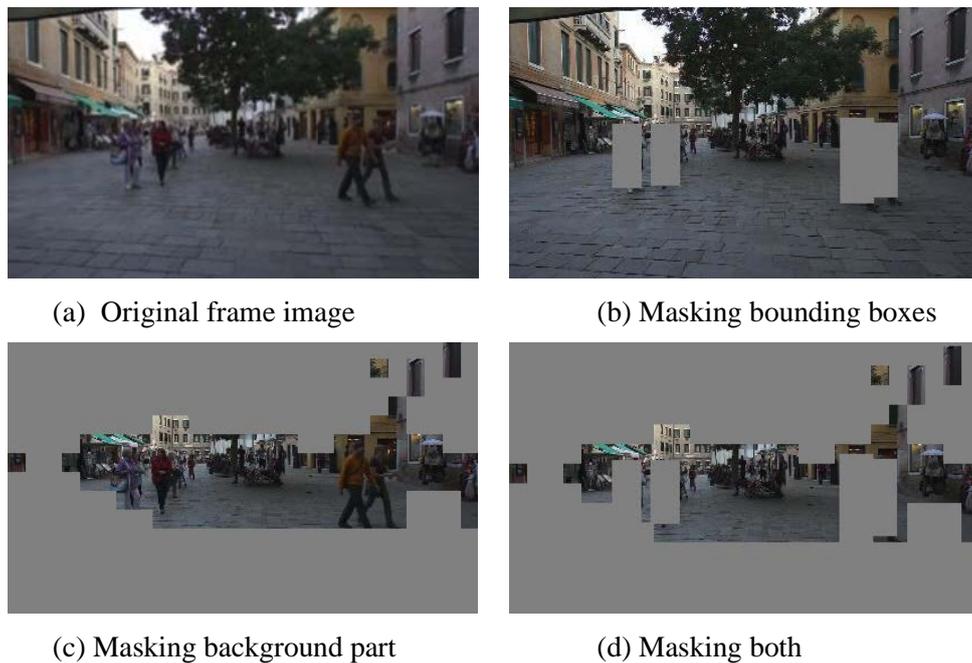


図 4.1 MOT17 データセットに対する提案手法によるマスキング結果[13]

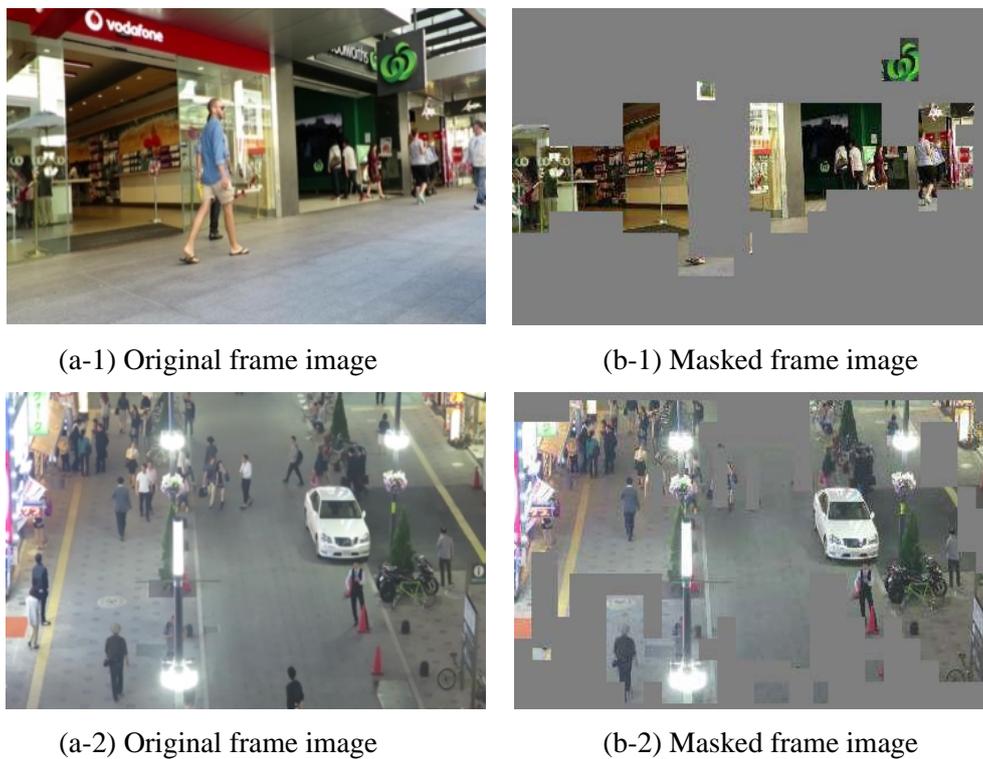


図 4.2 原フレーム画像とマスク画像の信頼度による比較結果[13]

4.4.2 JPEG 圧縮による検出精度の変化

クラウド側へ伝送するフレーム画像を JPEG で圧縮した場合、クラウド側の YOLOv3 の検出精度は、画像に対する圧縮率によって変化する。圧縮率が大きい場合には、データ伝送量は減少する一方、伝送される画像の品質が低下するため、検出精度が下がる。また、圧縮率が小さい場合には、伝送される画像の品質は大きく低下しないため、検出精度は大きく悪化しない一方、データ伝送量は増加する。圧縮精度の評価指標として平均精度を示す Average Precision (AP)を用いており、AP 損失率 (AP loss rate)は以下の式(4.1)で示される[14]。

$$AP \text{ loss rate} = 10 \log_{10} \frac{AP_{before} - AP_{after}}{AP_{before}} \quad (4.1)$$

ここに、 AP_{before} と AP_{after} は圧縮前後の物体検出精度をそれぞれ表す。また、図4.3は JPEG による画像圧縮がクラウド側の YOLOv3 による物体検出の精度に与える影響を示す。

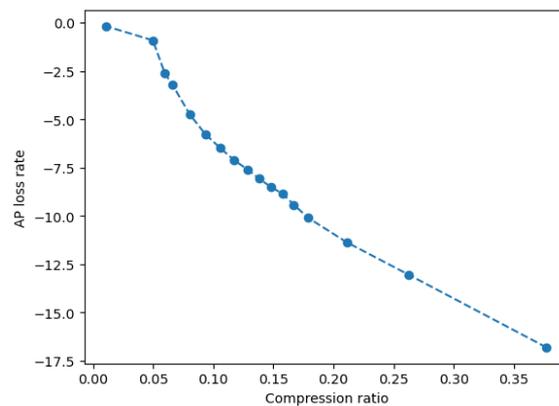


図 4.3 圧縮率の変化とクラウド側の YOLOv3 による物体検出精度への影響

4.4.3 提案手法の性能評価

評価指標として、クラウド側に伝送するデータサイズ(bpp)と物体検出の平均精度(AP) (IoU=0.5)の関係を用いた。図4.4は、二つの閾値の組み合わせからなる提案手法 (Ours1, Ours2)と、エッジ側で画像圧縮のみを行い、すべてのデータをクラウド側へ伝送してクラウド側の YOLOv3 による物体検出を行った場合の両者の関係を示している。提案手法は、特に精度が要求される一方で、クラウド側に伝送するデータ量が少ない場合に有効である。また、同一の検出精度において、単に全データをクラウド側へ伝送する場合に比べ、最大で約45%の伝送コストの削減を実現した。

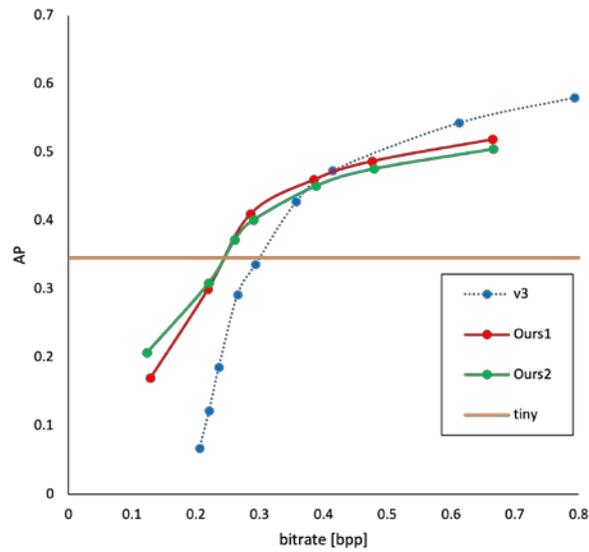
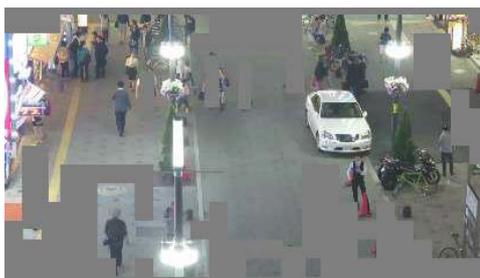


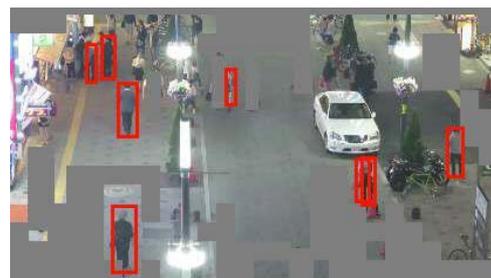
図 4.4 クラウド側へのデータ伝送量(bpp)と検出精度(AP)の関係

4.5 考察

提案手法の性能評価より，クラウド側へ伝送するデータ量が少ない場合に，提案手法が特に有効であることがわかった．これは，エッジ側のモデルで処理できる最大限の処理を行うことで，画像圧縮率が小さい場合においても，全体のデータ伝送量を削減できるためである．その結果，同じ小さな伝送量でも，より高画質な画像を伝送できると考えられる．さらに，エッジ側の YOLOv3-tiny による検出結果と，クラウド側の YOLOv3 によるマスク画像の再推論結果を図 4.5 に示す．実際に，エッジ側で検出しきれなかった物体を，クラウド側の YOLOv3 が補完して検出していることがわかる．



(a-1) Masked frame image



(b-1) Result of re-inference



(a-2) Masked frame image



(b-2) Result of re-inference



(a-3) Masked frame image

(b-3) Result of re-inference

図 4.5 エッジ側でマスクされた画像とクラウド側での再推論の結果[13]

4.6 むすび

本章では、本研究で用いた評価用データセットである MOT17 と提案手法の評価方法について述べた。また、エッジ側の機構におけるフレーム画像へのマスクングの変化や、画像圧縮が精度に与える影響について評価を行った。提案手法は、単に全データをクラウド側へ伝送する場合に比べ、特にデータ伝送量が少ない際に有効である。

第5章 結論と今後の課題

5.1 結論

本研究では、動画入力に対するエッジ・クラウド協調型の物体検出方式を提案した。動画に対する物体検出処理をエッジ側とクラウド側の二つに分割し、計算リソースが限られているエッジ側に軽量な YOLOv3-tiny、計算リソースが豊富なクラウド側に高性能な YOLOv3 を配置した。さらに、データ伝送量を削減する工夫として、エッジ側に三つの機構を導入した。まず、入力動画のフレーム間差分を測定し、前フレームとの差分が小さい場合には前フレームの結果を再利用することで、システム全体の推論回数を減らし、データ伝送量を削減した。また、エッジ側モデルの出力における各グリッドセルの信頼値に応じて、フレーム画像をマスクする手法を導入し、伝送するフレーム画像の情報量を減らした。加えて、マスクされた各フレーム画像に対し、JPEG による画像圧縮を行い、伝送量を削減した。これにより、単にすべてのデータをクラウド側に伝送して処理するよりも、提案手法は特にクラウド側へのデータ伝送量が少ない場合に有効であることを示した。

5.2 今後の課題

提案手法では、エッジ側で様々な処理をすることでクラウド側への伝送量を削減した一方で、その有効性は特にデータ伝送量が小さい場合に限られた。そのため、今後の課題として、伝送量が多い場合においても精度を維持できるような手法を考案することが考えられる。例えば、今回は JPEG を用いてマスクされた各フレーム画像を圧縮したが、エッジ側モデルの中間出力の特徴量などを伝送する仕組みを導入することで更なる伝送量の削減が期待できる。これらの手法の実現には、エッジ側とクラウド側に配置するモデルで共通の部分を作成するなどの工夫が必要になり、それぞれ適切なモデルの構築や学習が必要になると考えられる。

謝辞

本論文の執筆に当たり、丁寧かつ素晴らしいご指導をくださり、快適な研究環境を与えてくださった渡辺裕教授に心より感謝いたします。

共同研究において、多くの知識や示唆をしてくださった、NTT ソフトウェアイノベーションセンターの江田毅晴氏、榎本昇平氏、坂本啓氏、史旭氏に心より感謝いたします。

また、日頃から貴重な意見をくださり、充実した研究室環境を提供し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、私をここまで育ててくださり、常に心を支えてくださり、生活を支えてくださっている家族に感謝いたします。

参考文献

- [1] H. Choi and I. V. Bajic, "Deep Feature Compression for Collaborative Object Detection," IEEE International Conference on Image Processing (ICIP), pp. 3743-3747, Oct. 2018.
- [2] L. Hu, T. Wang, H. Watanabe, S. Enomoto, X. Shi, A. Sakamoto, and T. Eda, "ECNet: A Fast, Accurate, and Lightweight Edge- Cloud Network System based on Cascading Structure," IEEE Global Conference on Consumer Electronics (GCCE), pp. 259-262, Oct. 2020.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," IEEE Computer Vision and Pattern Recognition Conference (CVPR), pp. 779-788, Dec. 2016.
- [4] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, Apr. 2018.
- [5] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," International Conference on Advanced Computing & Communication Systems (ICACCS), pp. 687-694, Mar. 2020.
- [6] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," IEEE Computer Vision and Pattern Recognition Conference (CVPR), pp. 6517-6525, Jan. 2017.
- [7] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," IEEE Computer Vision and Pattern Recognition Conference (CVPR), pp. 2117-2125, Jan. 2017.
- [8] D.Xiao, F.Shan, Z.Li, B.T.Le, X.Liu, and X.Li, "A Target Detection Model Based on Improved Tiny-Yolov3 Under the Environment of Mining Truck," IEEE Access, vol. 7, pp. 123757-123764, Jul. 2019.
- [9] R. Huang, J. Pedoem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," in 2018 IEEE International Conference on Big Data (Big Data), pp. 2503-2510, Dec. 2018.
- [10] Z. Huang, F. Li, X. Luan, and Z. Cai, "A Weakly Supervised Method for Mud Detection in Ores Based on Deep Active Learning," Mathematical Problems in Engineering, vol. 2020, Article ID 3510313, 10 pages, May. 2020.
- [11] G. K. Wallace, "The JPEG still picture compression standard," IEEE Transactions on Consumer Electronics (TCE), vol. 38, no. 1, pp. xviii-xxxiv, Feb. 1992.
- [12] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, "MOT16: A Benchmark for Multi-Object Tracking," arXiv:1603.00831, Mar. 2016.
- [13] 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", IEEE Global Conference on Consumer Electronics (GCCE), pp364-367, Oct. 2023.
- [14] 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. 286-289, Oct. 2022.

図一覧

図 2.1	Edge-Cloud Net の構造.....	5
図 3.1	提案手法の全体構造.....	7
図 3.2	信頼値によるコントローラの処理の一例.....	8
図 3.3	エッジ側処理の全体アルゴリズム.....	9
図 4.1	MOT17 データセットに対する提案手法によるマスクング結果[13].....	12
図 4.2	原フレーム画像とマスク画像の信頼度による比較結果[13].....	12
図 4.3	圧縮率の変化とクラウド側の YOLOv3 による物体検出精度への影響.....	13
図 4.4	クラウド側へのデータ伝送量(bpp)と検出精度(AP)の関係.....	14
図 4.5	エッジ側でマスクされた画像とクラウド側での再推論の結果[13].....	15

表一覧

表 2.1 YOLOv3, YOLOv3tiny の性能比較 [9].....	6
表 4.1 各閾値の値の設定.....	11

研究業績

- [1] 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”, 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE 2023), Oct. 2023.
- [2] S. Akamatsu, K. Iino, H. Watanabe, S. Enomoto, A. Sakamoto and T. Eda “Edge-Cloud Collaborative Object Detection Model with Feature Compression”, The 8th IEEE International Conference on Image Electronics and Visual Computing (IEVC 2024), Mar. 2024, To appear.
- [3] K. Iino, S. Akamatsu, H. Watanabe, S. Enomoto, A. Sakamoto and T. Eda “Image Coding for Machines with Objectness-based Feature Distillation”, The 8th IEEE International Conference on Image Electronics and Visual Computing (IEVC 2024), Mar. 2024, To appear.
- [4] 飯野景, 赤松俊輔, 渡辺裕, 江田毅晴, 榎本昇平, 坂本啓, “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, To appear.