修 士 論 文 概 要 書

Master's Thesis Summary

Date of submission: 01/23/2023 (MM/DD/YYYY)

専攻名（専門分野） Department	情報理工・ 情報通信専攻	氏 名 Name	高橋 美帆	指 導 教 員 Advisor	渡辺 裕 印 Seal
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	5121F057-1		
研究題目 Title	Context R-CNN における交通監視のためのカテゴリベースのメモリバンク設計手法 Category-based Memory Bank Design for Traffic Surveillance in Context R-CNN				

1. まえがき

近年, IoT やビッグデータを利用した解析技術の普及に伴い，自動で物体を検出する技術である物体検出が注目されている．中でも交通監視における物体検出は，車両数の計測や追跡，異常検知の精度向上のために重要なタスクである．しかし，固定カメラにおけるぶった検出では，天候や正目条件により検出が難しいといった問題がある．そこで本研究では，Context R-CNN [1]を用いて交通データの物体検出を行い，カテゴリベースのメモリバンクを設計することで検出精度の向上を目指す．

2. Context R-CNN

2.1. Context R-CNN の概要

Context R-CNN は，Faster R-CNN をベースにした二段階の物体検出モデルで，固定カメラの物体検出のために開発された手法である．Context R-CNN の構造を図 1 に示す．

ベースモデルである Faster R-CNN [2]の Stage1 と Stage2 の間に Attention block を追加することにより，中間特徴量にメモリバンクのコンテキストを付与する．これにより，これまで検出が困難であった物体においても，類似したコンテキストの情報を含むことで検出が可能となり，検出精度の向上を実現している．

メモリバンクにあらかじめ設定された時間内に同じカメラの各フレームで最も予測スコアの高い物体の中間特徴量をコンテキストとしてメモリバンクに格納することにより，長期的なコンテキストを利用可能としている．

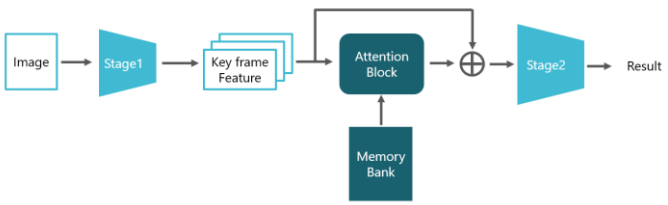

図 1 Context R-CNN の構造

2.2. 交通監視における課題

Context R-CNN では，フレームに複数のオブジェクトが含まれる場合でもフレームごとにコンテキストを格納するため，格納できるコンテキスト量に限界がある．その結果，物体数が多いフレームにおいては大量のコンテキストが利用されない，あるいは，複数のフレームにわたって存在する類似のコンテキストしか格納されない可能性がある．交通データセットである CityCam dataset [3] を用いてメモリバンクを作成した際のメモリバンクに含まれるカテゴリごとのコンテキスト数を表 1 に示す．表 1 より，カテゴリごとにコンテキスト数には大きな差が生じており，複数のカテゴリにおいてメモリバンクを活用できていないことが分かる．このように，常に多くの物体がフレーム内に存在する交通監視のデータにおいては従来手法のメモリバンク設計が適していないことが考えられる．

そこで，本研究では，カテゴリベースのアプローチにより，フレーム内のオブジェクトの数に関わらず，より多くのカテゴリのコンテキスト情報をメモリバンクに格納できるようにすることで，効果的なメモリバンクの設計を目指す．

表 1 従来手法で作成したメモリバンクの詳細

Category	The number of contexts in memory banks
taxi	1,965
black sedan	1,819
other cars	2,361
little truck	10
middle truck	11
big truck	20
van	111
middle bus	8
big bus	101
other vehicles	77
passengers	117
all	6,600

3. 提案手法

従来手法の問題点である，カテゴリごとに保存されるコンテキスト数に偏りがでる問題を解決するため，我々はコンテキストを物体のカテゴリ単位で選択する手法を提案する．

具体的な手法として，まず任意に設定した時間内の同じカメラの全てのフレームからすべての物体に対するコンテキストを取得する．その後，物体のカテゴリごとに予測スコアの高い順に，コンテキストを同数格納する．固定カメラでの精度向上を目的とするため，ベースモデルとして Context R-CNN を使用する．

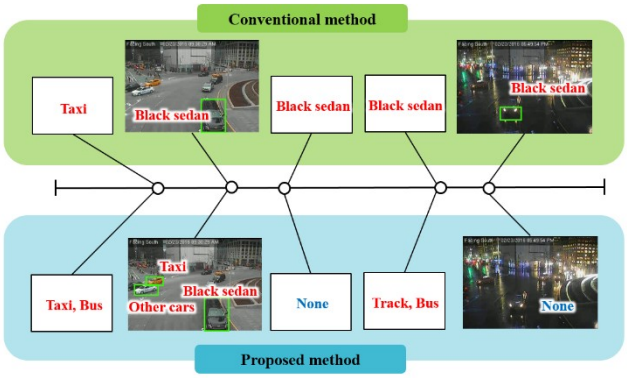

図 2 従来手法と提案手法の違い
(International Workshop on Advanced Image Technology(IWAIT2023)にて発表)

従来手法と提案手法の違いの例を図 2 に示す．従来方式では，各フレームから必ず一つの物体のコンテキストを格納するため，出現頻度の高いカテゴリの物体を格納するリスクが高くなり，一部のカテゴリではコンテキストがメモリバンクに格納されない可能性がある．一方，提案手法では，すべてのフレームから得られる全ての物体から選択したコンテキストを格納するため，一つのフレームに含まれている，より多くのカテゴリからメモリバンクを作成することが可能である．

4. 実験

4.1. 実験内容

物体検出タスクにおいて，提案手法の物体ごとの選択とフレームごとの選択の性能を比較するために，mAP（mean Average Precision）を評価手法として使用する． CityCam dataset を用いて，メモリバンクに格納できるストレージの数を制限した場合の従来法と提案法の検出精度を比較することで，提案するメモリバンク設計の有効性を調査する．メモリバンクに保存するために使うフレームはランダムで選択し，データセットは撮影条件や天候をそろえるため，学習用に 13 台のカメラ，テスト用に 4 台のカメラを使用する．

4.2. 実験結果及び考察

従来手法と提案手法でメモリバンクに保存するコンテキスト数を変化させた際の比較結果を図 3 に示し，カテゴリごとの検出精度（mAP）を表 2 に示す．

図 3 に示すように，従来手法の場合は検出精度がメモリバンクの格納数に関係なく検出精度が変化している．これは，従来手法がメモリバンクに格納するコンテキストが選択するフレームによって影響を受ける為である．一方で提案手法は，フレームをランダムで選んでもカテゴリベースで格納できるため，カテゴリごとのコンテキスト数に差がなく，従来手法より検出精度の向上ができていることが分かる．

また，表 2 に示すように，提案手法では全てのカテゴリが任意の数のコンテキストを格納することにより，多くのカテゴリにおいて検出精度向上を実現している．

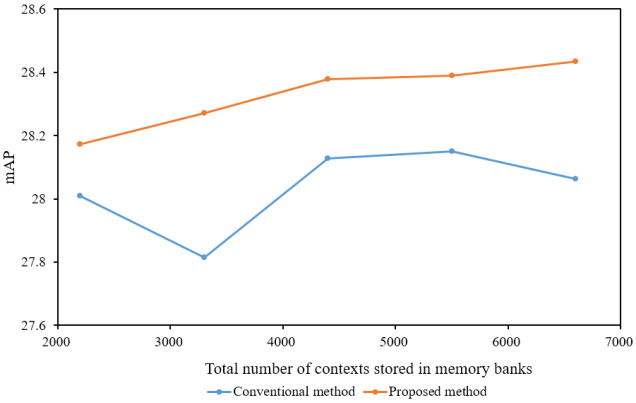

図 3 コンテキスト数を変化させた際の比較
(International Workshop on Advanced Image Technology(IWAIT2023)にて発表)

表 2 従来手法で作成したメモリバンクの詳細

Category	Conventional method	Proposed method
taxi	37.042	38.598
black sedan	34.379	34.792
other cars	34.368	34.4
little truck	12.108	10.542
middle truck	31.174	32.494
big truck	7.714	7.325
van	23.338	22.17
middle bus	40.909	41.234
big bus	51.006	50.773
other vehicles	20.147	22.316
passengers	16.505	18.129
all	28.063	28.434

5. むすび

本研究では，カテゴリベースのメモリバンク設計手法を提案した．

提案手法は，フレーム内の物体数やカテゴリごとの出現頻度に関わらず，全てのカテゴリにおいてメモリバンクを利用することができる．実験により，カテゴリベースでコンテキストを保存することで，mAP が平均 0.37 ポイント，カテゴリ別では最大 2.17%向上することを示した．

参考文献

[1] S. Beery, G. Wu, V. Rathod, R. Votel, and J. Huang: “Context R-CNN: Long Term Temporal Context for Per-Camera Object Detection,” 2020 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13072–13082, Jun. 2020.

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

[3] S. Zhang, G. Wu, J. P. Costeira, and J. M. Moura: “Understanding Traffic Density from Large-Scale Web Camera Data,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4264-4273, Jul. 2017.

2022 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

Context R-CNN における交通監視のための カテゴリベースのメモリバンク設計手法

Category-based Memory Bank Design for Traffic Surveillance in Context R-CNN

高橋　美帆
（5121F057-1）

提出日　2023.01.23
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 3

１．１　研究の背景 3
１．２　本研究の目的 3
１．３　本論文の構成 4

第２章　関連研究 5

２．１　まえがき 5
２．２　Context R-CNN 5
２．２．１　Context R-CNN とは 5
２．２．２　Attention block 6
２．２．３　交通監視における課題 7
２．３　むすび 7

第３章　予備実験 8

３．１　まえがき 8
３．２　メモリバンクの構成に関する予備実験 8
３．２．１　実験方法 8
３．２．３　実験結果 8
３．３　実験結果のまとめ，および考察 10
３．４　むすび 10

第４章　提案手法 11

４．１　まえがき 11
４．２　提案手法 11
４．３　むすび 12

第５章　実験内容及び結果と考察 13

５．１　まえがき 13
５．２　実験内容 13
５．３　実験結果 13
５．４　考察 15
５．５　むすび 15

第６章　結論と今後の展望 16
６．１　結論 16
６．２　今後の展望 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　研究の背景

近年，IoT やビッグデータを利用した解析技術の普及に伴い，自動で物体を検出する技術である物体検出が注目されている．物体検出は監視カメラでの人物検知や自動運転での歩行者の検知など，我々の生活にとって身近な存在となってきている．中でも交通監視における物体検出は，車両数の計測や追跡，異常検知の精度向上のために重要なタスクである．

交通監視における物体検出の問題点として，監視カメラなどの固定カメラを長期的に利用することが挙げられる．Faster R-CNN [1]などの従来の物体検出手法では物体全体が写っていない画像や物体が鮮明に写っていない画像は物体の特徴を十分に抽出しにくいため，検出が難しくなるという傾向がある．そのため，監視カメラなどの固定カメラを長期的に利用する場合，撮影時の天候によって画面全体が曇ってしまい物体が見えなかったり，夜は物体に光が当たらず，物体が不明瞭になったりすることにより検出が難しくなる．

１．２　本研究の目的

固定カメラの物体検出に特化した手法として，他フレームの物体のコンテキストを保存し，検出に利用することで検出精度を向上させる Context R-CNN [2]が提案されている．Context RCNN はメモリバンクに同一カメラの他の画像のコンテキスト情報を保存し，Attention block を用いて補足する仕組みで，検出性能を向上させている．これは，長期的なコンテキストを重要視し，あらかじめ定義された時間軸内の同じカメラからのすべてのフレームから最もスコアの高いコンテキストを一つ保存する戦略をとっている．

しかし，交通データにおいては画像内に同時に多くの物体が現れるうえ，似ている物体が複数のフレームにわたり存在するため，メモリバンクに同じようなコンテキスト情報ばかりが格納され，出現頻度の低い物体に関するコンテキスト情報が格納されづらいという問題がある．

そこで，本研究では検出物体のカテゴリに着目したメモリバンクの設計方法を提案する．具体的には，メモリバンクの最大格納数に対して物体ごとに格納できるコンテキストの数を設定し，検出精度や出現頻度に関わらず全ての物体に対して同数のコンテキスト情報が格納されるように設定し，検出性の向上を試みる．これにより，メモリバンク内により多くのカテゴリを含むコンテキストの保存を実現し，検出精度の低い物体や出現頻度の少ない物体の検出精度向上を目指す．

１．３　本論文の構成

以下に本論文の構成を示す．

第１章　本研究の背景，およびその目的について述べる．まず研究背景について記述した後，本研究の目的について述べる．

第２章　本研究で使用する関連技術について述べる．

第３章　Context R-CNN の再現実験およびメモリバンクに含まれるコンテキスト数の分布に関する予備実験について述べる．

第４章　本研究で提案するカテゴリベースのメモリバンク設計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　本研究のまとめおよび今後の課題を述べる．

第２章　　関連研究

２．１　　まえがき

本章では，本研究で利用する関連技術として，固定カメラに特化した物体検出手法の Context R-CNN について述べる．

２．２　　Context R-CNN

２．２．１　　Context R-CNN とは

Context R-CNN は，Faster R-CNN をベースにした二段階の物体検出モデルで，固定カメラの物体検出のために開発された．Context R-CNN の構造を図 2.1 に示す．

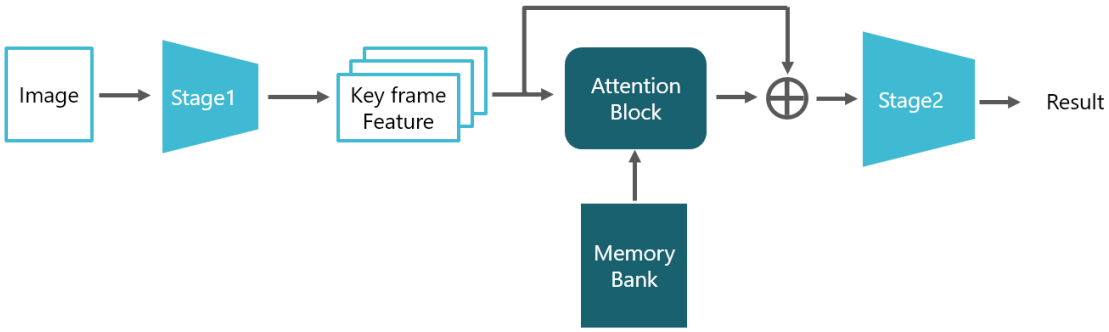

図 2.1　Context R-CNN の構造

Faster R-CNN の Stage1 と Stage2 の間に Attention block を追加することにより，中間特徴量にメモリバンクのコンテキストを付与する．これにより，これまで検出が困難であった物体においても，類似したコンテキストの情報を含むことで検出が可能となり，検出精度の向上を実現している．

メモリバンクにあらかじめ設定された時間内に同じカメラの各フレームで最も予測スコアの高い物体の中間特徴量をコンテキストとしてメモリバンクに格納することにより，長期的なコンテキストを利用可能としている．

2．2．2　Attention block

Attention block の構造を図 2.2 に示す.

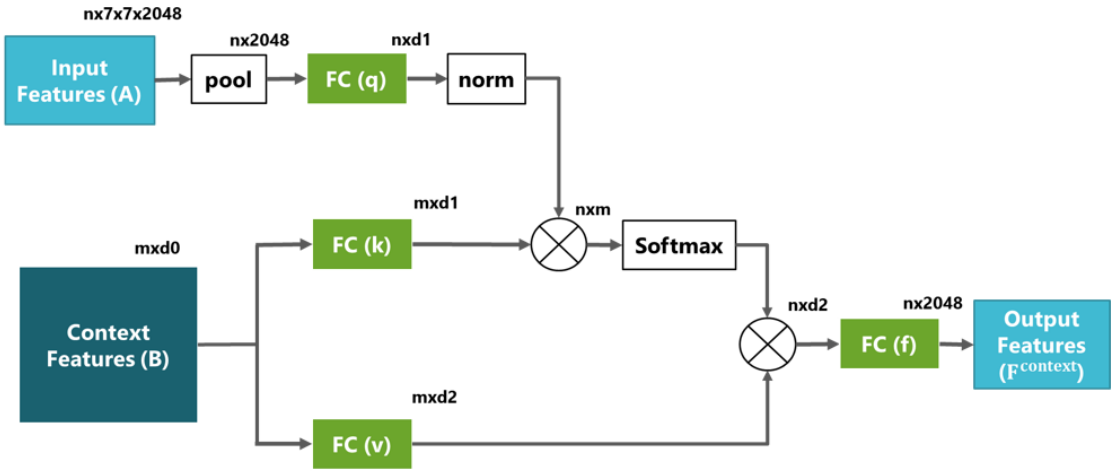

図 2.2　Attention block の構造

Attention block では，入力された中間特徴量に対するメモリバンクの全てのコンテキストの重みについて全結合層を使って算出する．入力の中間特徴量に対するコンテキストの重みは下式（2.1）で算出する．

$$w = \mathrm{Softmax}\left(\left(k\left(A^{pool};\theta\right)\cdot q(B;\theta)\right)/\left(T\sqrt{d}\right)\right) \tag{2.1}$$

ここで，w は形状 [n×m] の重み，A^{pool}は形状 [n × 2048] の空間的にプールされた入力の中間特徴量，Bは形状 [m × 2048] のメモリバンクに格納されたコンテキスト行列，$k(-;\theta)$は全結合層のキー関数，$q(-;\theta)$は全結合層のクエリ関数，d は特徴の深さ（2048），T＞0 は Softmax temperature である．次に，各ボックスについて，下式（2.2）に示すようにコンテキストの投影加重和をとることで，コンテキスト特徴 $F^{context}$ を構築する．

$$F^{context} = f(w\cdot v(B;\theta);\theta) \tag{2.2}$$

$F^{context}$ は形状が [n×2048] であり，最終的に $F^{context}$ を特徴チャネルごとのバイアスとして，元の入力特徴 A に戻すことでメモリバンク内のコンテキストを追加している．

２．２．３　交通監視における課題

Context R-CNN では，フレームに複数の物体が含まれる場合でもフレームごとにコンテキストを格納するため，格納できるコンテキスト量に限界がある．その結果，物体数が多いフレームにおいては大量のコンテキストが利用されない，あるいは，複数のフレームにわたって存在する類似のコンテキストしか格納されない可能性がある．このため，常に多くの物体がフレーム内に存在する交通監視のデータにおいて適していないことが考えられる．

そこで，本研究では，カテゴリベースのアプローチにより，フレーム内の物体数に関わらず，より多くのカテゴリのコンテキスト情報をメモリバンクに格納できるようにすることで，効果的なメモリバンクの設計を目指す．

２．３　むすび

本章では，関連研究である Context R-CNN について述べた．

第３章　予備実験

３．１　まえがき

本章では，従来のメモリバンク設計手法を用いた場合における，メモリバンクに含まれるコンテキストの分布に関する予備実験を行う．

３．２　メモリバンクの構成に関する予備実験

３．２．１　実験方法

従来のメモリバンク設計手法を用いた際の，学習データと Faster R-CNN の検出精度（mAP）に対するコンテキスト数の関係を調査する．具体的にはテストデータを用いて一つのフレームから最もスコアの高い物体に関するコンテキストをメモリバンクに格納し，メモリバンクに含まれるコンテキスト数をカテゴリごとに測定する．その際，学習データに含まれるカテゴリごとの物体数と Faster R-CNN の検出精度との相関を調べることで, 従来手法ではどのようなコンテキストがメモリバンクに保存されやすいかの考察を行う．

３．２．２　データセット

データセットとして，交通カメラデータセットの CityCam dataset [3]を使用する．CityCam dataset は 17 台のカメラ画像から構成され，10 種類の車両クラスと約 90 万個の注釈付きの物体が含まれる．撮影条件や天候をそろえるため，学習用に 13 台のカメラ，テスト用に 4 台のカメラを使用する．

３．２．３　実験結果

予備実験の結果を表 3.1 に示し，学習データの物体数とメモリバンクのコンテキスト数の相関を図 3.1，Faster R-CNN での検出精度とメモリバンクのコンテキスト数の相関を図 3.2 に示す．検出精度は CityCam dataset で再学習済みの Faster R-CNN を用いてテストデータを検出した結果である．

表 3.1 学習データの物体数と検出精度に対するメモリバンクのコンテキスト数

Category	The number of instances of train data	mAP (Faster R-CNN)	The number of contexts in memory bank
taxi	96,854	40.4	1,965
black sedan	201,715	36.9	1,819
other cars	219,892	31.1	2,361
little truck	12,148	6.19	10
middle truck	20,604	9.97	11
big truck	4,173	7.87	20
van	25,757	15.0	111
middle bus	5,744	13.8	8
big bus	13,194	39.3	101
other vehicles	17,379	16.7	77
passengers	62,604	15.8	117
all	680,064	21.2	6,600

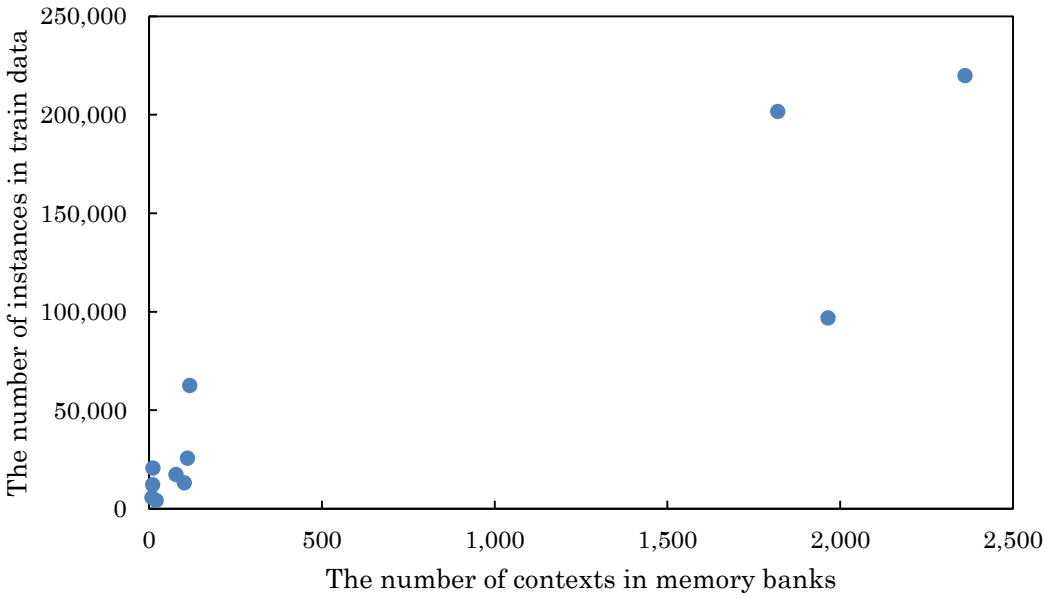

図 3.1　学習データの物体数とメモリバンクのコンテキスト数

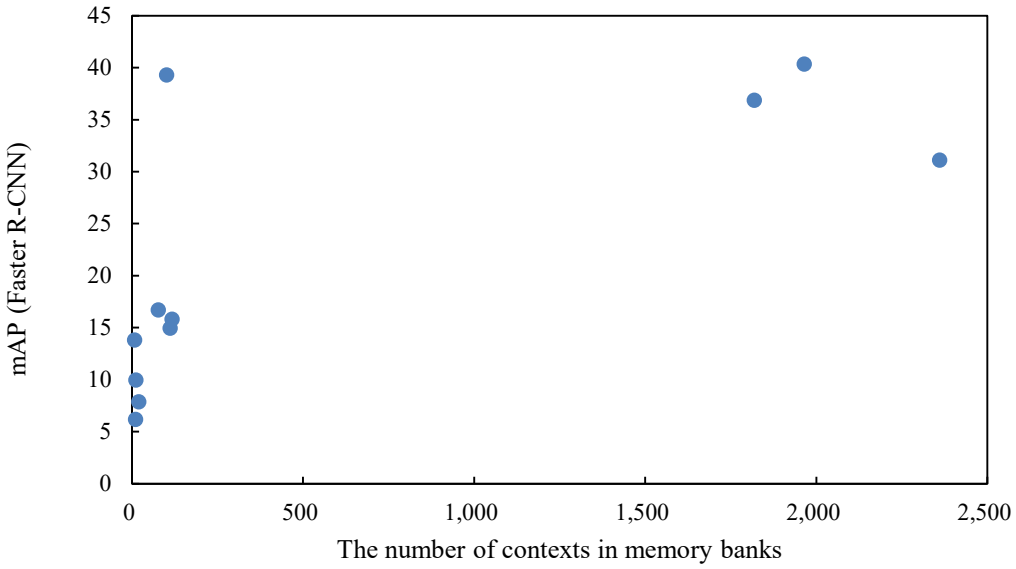

図 3.2　Faster R-CNN での検出精度とメモリバンクのコンテキスト数

３．３　実験結果のまとめ，および考察

図 3.1 より，学習データに物体が多く含まれるカテゴリは，従来手法でメモリバンクに多くのコンテキストが保存されていることが分かる．一方，学習データに含まれる物体数が少ないカテゴリはメモリバンクに保存されるコンテキスト数が少ない．これは，テストデータと学習データの分布が近く，テストデータにおいてもデータに含まれる物体数が少ないためである．

同様に，図 3.2 に示すように，検出精度が高いカテゴリはメモリバンクに保存されるコンテキスト数が多く，検出精度が低いカテゴリはコンテキスト数が少ない傾向がある．

これより，従来のメモリバンク設計では，実験データにおける出現頻度が低く，検出精度が低いカテゴリに関するコンテキストが保存されにくいと考えられる．

３．４　むすび

本章では従来手法でメモリバンクを設計した場合における，保存されやすいコンテキストの特徴に関する予備実験について述べた．

第４章　　提案手法

４．１　　まえがき

本章では，本研究の提案手法である，カテゴリベースの Context R-CNN のメモリバンク設計手法について述べる．

従来手法の問題点である，出現頻度が低く検出精度が低いカテゴリに関するコンテキストが保存されにくい問題を解決するため，我々は全てのカテゴリのコンテキストを同数保存するカテゴリベースのメモリバンク設計手法を提案する．

４．２　　提案手法

従来のメモリバンク設計では，カテゴリごとにメモリバンクに保存されるコンテキスト数に偏りが発生し，全てのカテゴリが一様にメモリバンクを活用した検出ができないため，出現頻度が少なく検出精度の低いカテゴリにおいてもメモリバンクに保存できるようにカテゴリベースのアプローチを取る．そのため我々は，メモリバンクに格納するコンテキストをフレーム単位ではなく，物体のカテゴリ単位で選択する手法を提案する．固定カメラでの精度向上を目的とするため，ベースモデルとして Context R-CNN を使用する．

具体的な手法としては，まず任意に設定した時間内の同じカメラの全てのフレームからすべての物体に対するコンテキストを取得する．その後，物体のカテゴリごとに予測スコアの高い順に，コンテキストを同数格納する．はじめに全ての物体に対するコンテキストを取得することで，出現頻度の低い物体のコンテキストもメモリバンクに格納することができる．図 4.1 に同一フレームからコンテキスト情報を取得する際の従来手法と提案手法の違いの一例を示す．

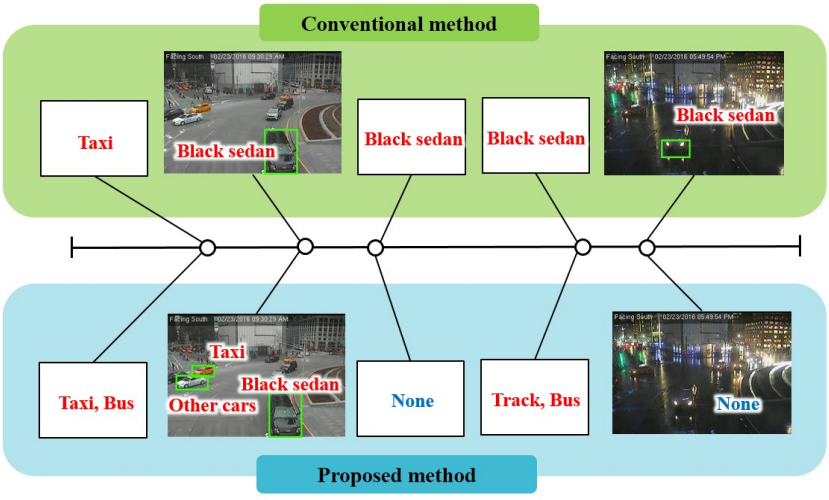

図 4.1　従来手法と提案手法の違い

（International Workshop on Advanced Image Technology (IWAIT2023)にて発表）

図 4.1 の例に示すように，従来方式では，各フレームから必ず一つの物体のコンテキストを格納するため，出現頻度の高いカテゴリの物体を格納するリスクが高くなり，一部のカテゴリではコンテキストがメモリバンクに格納されない可能性がある．一方，提案手法では, すべてのフレームから得られる全ての物体から選択したコンテキストを格納するため，一つのフレームに含まれている，より多くのカテゴリからメモリバンクを作成することが可能である．

４．３　むすび

本章では, カテゴリベースの Context R-CNN のメモリバンク設計手法について述べた．

従来手法の問題点である, 出現頻度が低く検出精度が低いカテゴリに関するコンテキストが保存されにくい問題を解決するため, 全てのカテゴリのコンテキストが保存されるようにメモリバンクを設計する．カテゴリベースのアプローチにより，従来手法では保存できるコンテキスト数が少なかったカテゴリのコンテキストも保存できるようにすることで，全てのカテゴリにおける検出精度向上を図る．

第５章　実験内容及び結果と考察

５．１　まえがき

本章では，メモリバンクの最大格納数を変化させた時の従来手法と提案手法の比較実験の結果とカテゴリごとの検出精度の比較実験の結果を示し，その考察について記述する．

５．２　実験内容

物体検出タスクにおいて，提案手法の物体ごとの選択とフレームごとの選択の性能を比較するために，mAP（mean Average Precision）を評価手法として使用する．

CityCam dataset を用いて，メモリバンクに格納できるストレージの数を制限した場合の従来法と提案法の検出精度を比較することで，提案するメモリバンク設計の有効性を調査する．メモリバンクに保存するために使うフレームはランダムで選択する．また，データセットは予備実験と同様，撮影条件や天候をそろえるため，学習用に 13 台のカメラ，テスト用に 4 台のカメラを使用する．

５．３　実験結果

CityCam dataset において，従来手法と提案手法でメモリバンクに保存するコンテキスト数を変化させた場合の比較結果を図 5.1 に示す．

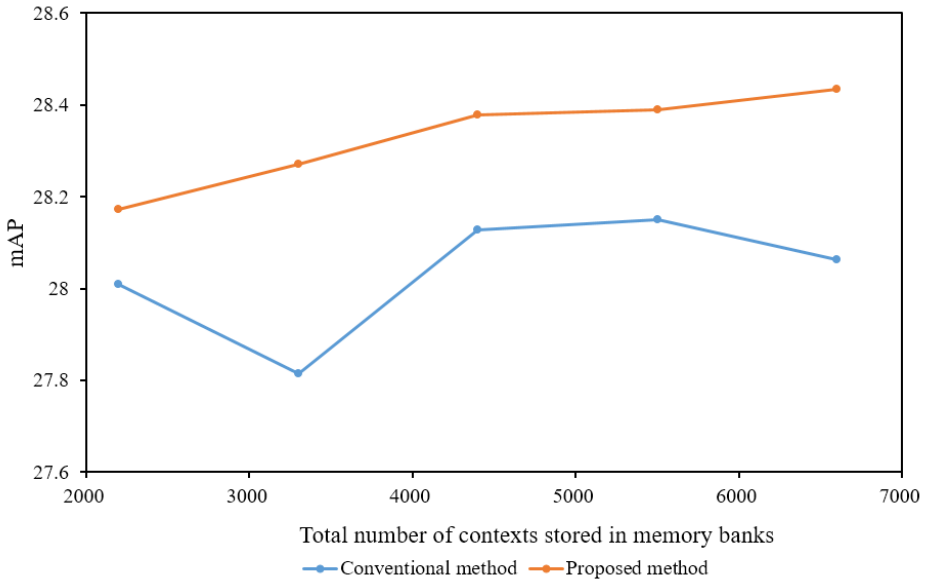

図 5.1　コンテキスト数を変化させた時の従来手法と提案手法の比較
（International Workshop on Advanced Image Technology (IWAIT2023)にて発表）

さらに, 従来手法と提案手法で同数のコンテキストを格納した場合のカテゴリごとの検出精度（mAP）を表 5.1，メモリバンクの詳細を表 5.2 に示す．従来手法と比較して提案手法では，mAP が平均 0.37 ポイント，カテゴリによっては最大 2.17%向上することが確認できた.

表 5.1 従来手法と提案手法における検出精度（mAP）の比較

Category	Conventional method	Proposed method
taxi	37.042	38.598
black sedan	34.379	34.792
other cars	34.368	34.400
little truck	12.108	10.542
middle truck	31.174	32.494
big truck	7.714	7.325
van	23.338	22.170
middle bus	40.909	41.234
big bus	51.006	50.773
other vehicles	20.147	22.316
passengers	16.505	18.129
all	28.063	28.434

表 5.2 従来手法と提案手法にメモリバンクのコンテキスト数の比較

Category	Conventional method	Proposed method
taxi	1,965	600
black sedan	1,819	600
other cars	2,361	600
little truck	10	600
middle truck	11	600
big truck	20	600
van	111	600
middle bus	8	600
big bus	101	600
other vehicles	77	600
passengers	117	600
all	6,600	6,600

５．４　考察

図 5.1 より，従来手法の場合，フレームから一つずつ物体のコンテキストを格納するため，フレームの選び方によって有用なコンテキストが含まれているかで検出精度に影響が出てしまうこと分かる．本実験では数あるフレームの中からランダムで格納数分のフレームを選択し，メモリバンクを作成しているため，図 5.1 に示すように，従来手法では検出精度はメモリバンクの格納数ではなく，選択したフレームによって検出精度が向上したり低下したりしている．

一方で，提案手法の場合は，フレームをランダムで選んでも全体としてカテゴリごとに格納される数の割合が一定のため，どのようなフレームをランダムで選択しても従来手法より検出精度の向上ができていることが分かる．

また，表 5.1 の結果から，多くのカテゴリにおいて提案手法の方が従来手法に比べて検出精度が高くなっていることが分かる．表 5.2 に示すように taxi や black sedan，other cars などは，メモリバンクに含まれるコンテキスト数が減っているが，その場合も検出精度が向上していることから，有用なコンテキストのみ格納できるようになったと考えられる．一方で，出現頻度が低く従来手法ではメモリバンクに含まれるコンテキスト数が少なかったカテゴリにおいては，提案手法によりメモリバンクにより多くのコンテキストが含まれ，多くのカテゴリにおいて検出精度が向上した，

これより，提案手法により平均して全体の検出精度を向上できることが確認できた．

５．５　むすび

本章では，提案した手法の実験方法と結果，および考察について述べた．

実験結果より，カテゴリを考慮したメモリバンク設計を行うことで，どのようなフレームを用いてメモリバンクを作成しても，全てのカテゴリのコンテキストを保存できるようになり，全体の検出精度が向上し，提案手法の有効性を確認できた．

第６章　結論と今後の展望

６．１　結論

本研究では，コンテキストをフレーム単位ではなくカテゴリ単位でメモリバンクに格納する手法を提案した．提案手法では，フレーム内の物体数やカテゴリごとの出現頻度に関わらず，全てのカテゴリにおいてメモリバンクを利用することができる．実験により，カテゴリに注目してコンテキストの保存方法を設定することで，mAP が平均 0.37 ポイント，カテゴリ別では最大 2.17%向上することを示した．

６．２　今後の展望

従来手法で提案されているメモリバング設計方法は，フレーム内に含まれる物体数が少ない場合において最も検出精度が向上するように設計されており，使用するデータセットによってフレーム内の物体数が違うため，メモリバンクを活用できない場合がある．しかし，実社会の物体検出においてはカメラを設置する環境ごとに物体検出モデルを変更することが難しいため，どのデータにおいても検出精度向上が可能な物体検出モデルであることが重要である．

本研究では，カテゴリ単位でメモリバンクに格納することでフレーム内の物体数が多くカテゴリごとに出現頻度が異なる場合でも，すべてのカテゴリがメモリバンクに格納できる手法を提案した．この手法はフレーム内の物体数が少ない場合においても同様にメモリバンクを活用できるため，フレーム内の物体数やカテゴリごとの出現頻度に関わらず，あらゆるデータに対して有効な手法であると考えられる．

謝辞

本研究に際して，様々なアドバイスや素晴らしいご指導をしてくださり，学会参加の機会や実験環境および新たな研究分野へ挑戦できる研究環境を与えてくださった，渡辺教授に心より感謝いたします.

共同で研究をさせていただき，日頃から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の方々に心より感謝いたします.

また，リモートワークが続き対面でのセッションが叶わない状況でも，様々な角度から意見をくださり，学びの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，私を自由に育ててくださり，常に明るく元気に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] S. Ren, K. He, R. Girshick, and J. Sun: “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Advances in Neural Information Processing Systems 28 (NIPS 2015), Vol. 1, pp. 91-99, Dec. 2015.

[2] S. Beery, G. Wu, V. Rathod, R. Votel, and J. Huang: “Context R-CNN: Long Term Temporal Context for Per-Camera Object Detection,” 2020 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13072–13082, Jun. 2020.

[3] S. Zhang, G. Wu, J. P. Costeira, and J. M. Moura: “Understanding Traffic Density from Large-Scale Web Camera Data,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4264-4273, Jul. 2017.

図一覧

図 2.1　Context R-CNN の構造 5
図 2.2　Attention block の構造 6
図 3.1　学習データの物体数とメモリバンクのコンテキスト数 9
図 3.2　Faster R-CNN での検出精度とメモリバンクのコンテキスト数 10
図 4.1　従来手法と提案手法の違い　（International Workshop on Advanced Image Technology (IWAIT2023)にて発表） 11
図 5.1　コンテキスト数を変化させた時の従来手法と提案手法の比較（International Workshop on Advanced Image Technology (IWAIT2023)にて発表） 13

表一覧

表 3.1 学習データの物体数と検出精度に対するメモリバンクのコンテキスト数 9
表 5.1 従来手法と提案手法における検出精度（mAP）の比較 14
表 5.2 従来手法と提案手法にメモリバンクのコンテキスト数の比較 14

研究業績

[1] 高橋美帆 :“ランドマーク情報を利用した WGAN-GP によるイラスト用顔画像の生成手法の研究（A Study on Generation Method of Face Image for Illustration by WGAN-GP Using Landmark Information）”，卒業論文，Feb．2021.

[2] M. Takahashi and H. Watanabe: “Face Image Generation for Illustration by WGAN-GP Using Landmark Information,” IEEE Global Conference on Consumer Electronics (GCCE), OS-DSC, pp.989-990, Oct. 2021.

[3] T. Shindo, T. Watanabe, R. Yano, M. Arimoto, M. Takahashi, and H. Watanabe, “Super Resolution for QR code Images,” IEEE Global Conference on Consumer Electronics (GCCE), pp. 281-284, Oct. 2022.

[4] 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), OS-ASP(1), pp.293-296, Oct. 2022

[5] 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)