

修士論文概要書

Master's Thesis Summary

Date of submission: 01/26/2026 (MM/DD/YYYY)

専攻名（専門分野） Department	情報理工・ 情報通信専攻	氏 名 Name	范 洪睿	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	CD 5124F090-6		
研究題目 Title	特徴強化手法を用いた物体検出モデルの検出精度向上に関する研究 Research on Detection Accuracy Improvement of Object Detection Models Using Feature Enhancement Methods				

1. まえがき

近年、画像認識および映像解析技術は、自動運転、ドローン画像解析など多様な分野で活用されており、物体検出技術の高精度化がますます求められている。しかし、小物体は画像中に占める画素数の割合が少なく、背景情報に埋もれやすいため、検出精度の低下が課題となっている。この問題は、ドローン空撮画像のように撮影高度が高く、対象物が極めて小さく写る場面において顕著である。

既存の物体検出手法では、畳み込み演算に基づく局所的な特徴抽出が主流であるが、受容野の制約により、大域的な文脈情報や長距離依存関係を十分に捉えることが困難である。また、多段のダウンサンプリングにより、小物体に関する空間情報が失われやすいという問題も存在する。

本研究では、異なる解像度の特徴量を統合する Feature Pyramid Network[1] に Non-Local Neural Networks[2]を導入する手法と、Spatial Bias[3]を参考にチャンネル次元に着目した軽量の Channel Bias 機構を導入する手法を提案し、それぞれ異なる検出モデルおよびデータセットを用いて有効性を検証する。

2. 関連研究

2.1. Feature Pyramid Network

Feature Pyramid Network (FPN)[1]は、異なる解像度の特徴マップを統合することで、マルチスケール物体検出性能を向上させる手法である。上位層の高レベル特徴と下位層の高解像度特徴を融合することで、大きさの異なる物体に対応可能な特徴表現を生成する。一方で、FPN は畳み込み演算に基づく局所的な特徴抽出に依存しており、大域的な文脈情報や長距離依存関係を十分に捉えることが難しいという課題がある。

2.2. Non-Local Neural Networks

Non-Local Neural Networks[2]は、特徴マップ上のすべての位置間の関係性を考慮することで、長距離依存関係を明示的にモデル化する手法である。しかし、空間位置間の関係を全結合的に計算するため、計算量やメモリ使用量が大きくなりやすいという問題がある。

2.3. Spatial Bias

Spatial Bias[3]は、注意機構に比べて軽量の構成で特徴表現を補正することを目的としている。この手法では、空間次元に基づく大域的な偏りを導入し、特徴を補強する。この手法は計算コストが低く、既存ネットワークに容易に組み込める一方で、適用位置や設計によって効果が変化するという課題がある。

3. 提案手法

3.1. 概要

本研究では、小物体検出における特徴表現の不足という課題に対し、異なる観点から特徴を補強する二つの手法を検討する。一つ目は、Non-Local Neural Networks を FPN に組み込むことで、大域的な空間依存関係を明示的にモデル化する手法である。二つ目は、空間情報に依存せず、チャンネル方向の統計情報に基づいて特徴を補正する Channel Bias 機構である。

3.2. Non-Local Neural Networks を用いた特徴強化

FPN 構造に Non-Local モジュールを導入し、大域的な文脈情報を特徴表現に反映させることで、小物体に対する検出精度の向上を目指す。本手法の処理構造および導入位置を図1に示す。

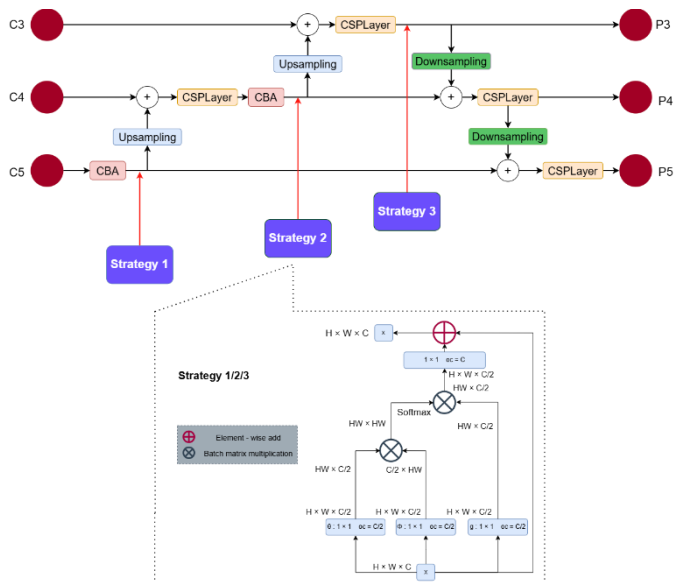


図1 Non-Local モジュール導入構成の全体図

3.3. Channel Bias を用いた特徴強化

Channel Bias は、入力特徴マップに対してチャネル方向に着目した補助情報を付加することで特徴表現を強化する手法である。まず、全チャネルの活性化度をグローバル平均プーリングにより要約し、チャネル間の傾向を抽出する。得られたチャネル統計を 1×1 畳み込みによって少数チャネルの Bias 成分へ変換し、空間方向に拡張した後、元の特徴マップに結合することで、新たな特徴表現 Y を得る。この出力は次のように表される。

$$Y = \text{Conv}_{1 \times 1}([X, CB]) \quad (1)$$

本手法の処理流れを図 2 に示す。チャネル数にのみ依存するため、軽量で既存の検出モデルに容易に組み込むことが可能である。

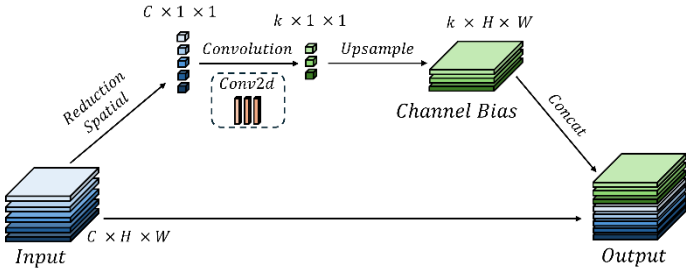


図 2 Channel Bias の概要

4. 実験

4.1. データセット

Non-Local Neural Networks 手法の評価には VOC2007 データセットを使用し、Channel Bias 手法の評価には VisDrone データセットを使用した。VOC2007 は一般物体認識を目的として構築されたデータセットであり、多様なカテゴリと撮影条件を含む自然画像で構成されている。一方、VisDrone は小物体が多数含まれるドローン空撮画像データセットであり、小物体検出性能の評価に適している。

4.2. Non-Local 手法の実験結果

表 1 の結果より、ベースライン YOLOX の FPN の各層に Non-Local モジュールを挿入した場合、精度の向上が確認された。Non-Local による大域的文脈情報の導入が有効であることが示された。一方で、導入位置によっては性能低下が見られ、モジュールの効果が構造設計に依存することも明らかとなった。

表 1 Non-Local 導入位置別の検出精度

Method	$AP_{50} \uparrow$	$AP_{75} \uparrow$	$mAP_{50-95} \uparrow$
baseline	58.3	32.1	32.2
strategy 1	58.7	33.0	32.7
strategy 2	57.4	31.6	32.1
strategy 3	55.6	29.9	30.3

小, 中, 大の物体に分類した検出性能を表 2 に示す。結果より, Non-Local モジュールの導入により, 小物体に対して顕著な精度向上が確認された。一方で, 中サイズおよび大サイズの物体に対する改善は限定的であり, 導入効果が物体スケールによって異なる傾向が見られた。この結果は, 特徴が局所的に弱くなりやすい小物体に対して, 大域的な文脈情報が有効に寄与する可能性を示唆している。

表 2 物体スケール別の検出精度

Method	All	Small	Medium	Large
baseline	32.2	11.4	22.4	36.3
strategy 1	32.6	15.2	22.2	36.7
strategy 2	32.1	19.8	21.3	36.4
strategy 3	30.3	10.8	19.8	34.2

4.3. Channel Bias 手法の実験結果

表 3 に示す結果より, YOLOv11 および YOLOv12 に Channel Bias (CB) を導入した実験では, CB は両モデルともバックボーンの Layer3 に挿入した。その結果, 両モデルにおいて精度の改善が確認された。チャネル方向の情報補正が小物体検出に対して有効であることが示された。

表 3 Channel Bias の検出性能比較

Method	$AP_{50} \uparrow$	$AP_{75} \uparrow$	$mAP_{50-95} \uparrow$
YOLOv11	36.0	21.7	21.4
YOLOv11 + CB	36.5	21.7	21.5
YOLOv12	35.7	21.2	21.0
YOLOv12 + CB	36.4	22.0	21.6

5. むすび

本研究では, 小物体検出精度の向上を目的として, Non-Local Neural Networks を用いた大域的特徴強化手法と, Channel Bias によるチャネル方向の特徴補正手法を検討した。実験結果より, 両手法とも小物体検出に対して有効であることが確認された。Channel Bias は軽量かつ汎用的な構成であり, 既存の検出モデルに容易に導入可能である。今後は, さらなるデータセットへの適用や, 他の検出タスクへの展開について検討する予定である。

参考文献

- [1] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125, Jul. 2017.
- [2] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7794–7803, Jun. 2018.
- [3] J. Go and J. Ryu, "Spatial bias for attention-free non-local neural networks," Expert Systems with Applications, vol. 238, p. 122053, Mar. 2024.

2025 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

特徴強化手法を用いた物体検出モデルの検出精度向上に
関する研究

Research on Detection Accuracy Improvement of Object
Detection Models Using Feature Enhancement Methods

范 洪睿

(5124F090-6)

提出日：2026.1.26

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
1.1 研究背景	1
1.2 研究目的	1
1.3 本論文の構成	2
第 2 章 関連研究	3
2.1 まえがき	3
2.2 YOLO 系列による物体検出	3
2.2 Feature Pyramid Network	3
2.3 Non-Local Neural Networks	4
2.4 Spatial Bias	7
2.5 むすび	8
第 3 章 提案手法	9
3.1 まえがき	9
3.2 提案手法の概要	9
3.3 FPN に Non-Local Neural Networks を導入した手法	9
3.3.1 Non-Local Neural Networks の導入位置	9
3.3.2 FPN における特徴強化処理の流れ	10
3.4 Channel Bias を用いた特徴強化手法	11
3.4.1 Channel Bias の基本構成	11
3.4.2 チャネル方向特徴の抽出	11
3.4.3 特徴マップへの統合	12
3.5 むすび	12
第 4 章 実験結果と考察	13
4.1 まえがき	13
4.2 実験環境およびデータセット	13
4.3 評価指標	15
4.4 FPN における Non-Local Neural Networks の導入に関する検討	17
4.4.1 実験設定	17
4.4.2 全体精度の比較	17
4.4.3 対象スケール別の比較	17
4.4.4 結果の考察	18
4.5 Channel Bias 手法の評価	18
4.5.1 実験設定	18

4.5.2 全体精度の比較	19
4.5.3 結果の考察	19
4.6 むすび	20
第5章 結論と今後の展望	21
5.1 結論	21
5.2 今後の展望	21
謝辞	23
参考文献	24
表一覧	26
図一覧	27
研究業績	28

第1章 序論

1.1 研究背景

近年、深層学習を用いた映像解析技術は、自動運転、ドローン画像解析など、さまざまな分野で広く応用されており、特に物体検出技術は実環境における視覚認識の基盤技術として重要な役割を果たしている[1]。物体検出は、画像中から対象物の位置および種類を同時に推定するタスクであり、その精度は後段の認識処理や意思決定に大きな影響を与える。

物体検出技術の中でも、小物体検出は依然として困難な課題の一つである。小物体は画像内で占める画素数の割合が小さく、周囲の背景情報に埋もれやすいため、特徴量の抽出が困難となる。また、ドローン撮影画像では、撮影高度や撮影距離の影響により、対象物が極めて小さく映る場合が多く、小物体検出性能の向上が強く求められている。

近年では、YOLO 系列[2]に代表される単段検出器が、高速かつ高精度な物体検出モデルとして広く利用されている。これらのモデルは、Feature Pyramid Network (FPN)[3]などの特徴統合機構を用いることで、異なる解像度の特徴マップを活用し、小物体検出性能の向上を図っている。しかしながら、畳み込み演算を基本とする従来の手法では、局所的な特徴に依存する傾向が強く、広域的な文脈情報やチャネル間の関係を十分に活用できない場合がある。

このような背景から、特徴マップに対して空間的あるいはチャネル方向の情報を補強し、より識別性の高い特徴表現を獲得する手法が注目されている。特に、Non-Local Neural Networks[4]による空間位置間の関係性の導入や、チャネル方向の情報を集約する軽量な特徴強化手法は、小物体検出性能の向上に寄与する可能性があると考えられる。

一方で、これらの特徴強化手法は、モデル構造や適用位置によって効果が大きく異なる場合があり、検出モデルとの整合性を考慮した設計が重要となる。そのため、既存の物体検出モデルに対して、どのような特徴強化手法が有効であるかを体系的に検討することが求められている。

1.2 研究目的

本研究では、小物体検出における認識精度向上を目的として、既存の物体検出手法が抱える文脈情報の不足および小物体特徴の弱さに着目し、特徴表現の改良を行う。小物体は画像中における画素数の割合が小さいため、有効な特徴が得られにくく、周囲環境の影響を受けやすいという課題がある。このような特性を踏まえ、小物体に対して有効な情報を安定的に強調しつつ、画像全体の文脈を考慮可能な特徴表現の獲得を目指す。

上記課題に対して二つの改良手法を提案する。第一に、多尺度特徴統合を行う既存の検

出モデルに対し、画像全体の関係性を明示的に考慮できる機構を導入することで、広域的な文脈情報を特徴表現に反映させる手法を提案する。これにより、局所特徴の統合に留まりがちな従来構造を補完し、小物体に対する識別性の向上を図る。第二に、小物体では空間方向の情報が限定されやすい点に着目し、チャンネル方向の特徴選択および強調を通じて表現を補強する軽量な特徴強化手法を提案する。これにより、計算コストを抑制しつつ、小物体検出性能の向上を目指す。

ドローン空撮画像を含むデータセットおよび一般画像に含まれる小物体を対象として評価実験を行い、小物体検出精度を指標として既存手法との比較を通じて提案手法の有効性を検証する。さらに、各構成要素の寄与を分析することで、それぞれの改良が小物体検出性能に与える影響を明らかにする。

1.3 本論文の構成

本論文の構成を以下に示す。

第1章は本章であり、本論文の研究背景、研究目的について述べる。

第2章では本研究で用いる関連研究について述べる。

第3章では小物体検出に向けた特徴表現の改良手法について述べる。

第4章では実験の結果及び考察について述べる。

第5章では本論文の結論及び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，YOLO 系列に代表される物体検出手法，多尺度特徴統合に関する手法，文脈情報を考慮した Non-Local Neural Networks，および空間方向の情報を利用した Spatial Bias に関する既存研究について説明する．

2.2 YOLO 系列による物体検出

YOLO 系列は，入力画像中の物体を高速に検出する単段型の物体検出モデルである．入力画像はネットワークに通され，特徴抽出から物体位置の推定および分類までを一貫して実行する構造を持つ．一般的な YOLO の構成例を図 2.1 に示す．YOLO 系列は，特徴抽出を行う Backbone，複数段階の特徴を統合する Neck，そして検出を出力する Head の三つの要素から構成される．Backbone では浅い層で局所的な特徴，深い層で抽象的な意味情報を取得する．Neck は異なる解像度の情報をまとめ，物体サイズに依存しない特徴を形成する役割を担う．

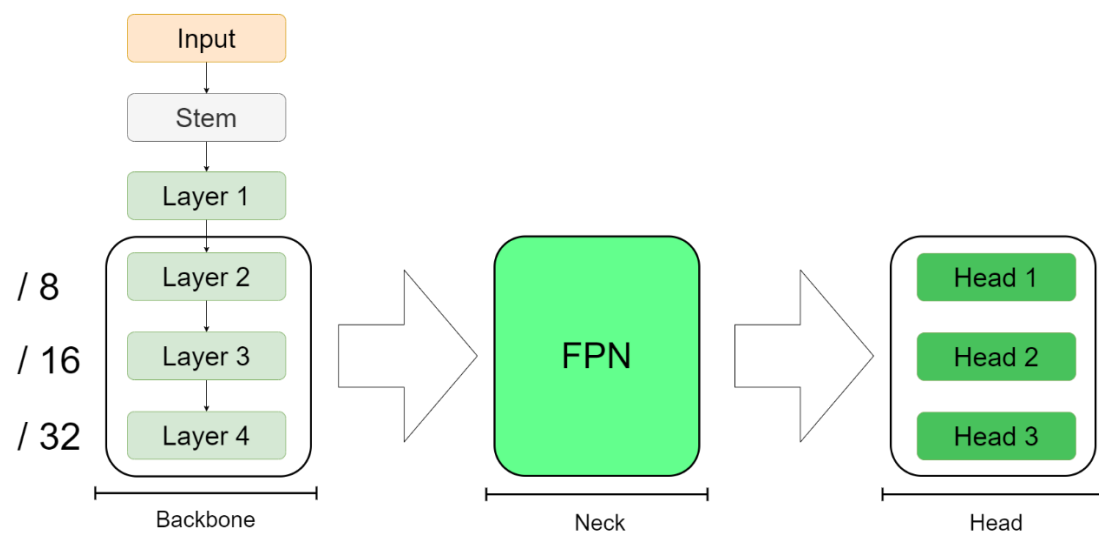


図 2.1 YOLO モデルにおけるネットワーク構成例

2.2 Feature Pyramid Network

YOLO の Neck 部に用いられる主要な構造として Feature Pyramid Networks が挙げられる．

FPN は、Backbone から得られる解像度の異なる特徴マップを活用し、多様なスケールの物体に対応するための仕組みである．図 2.2 に FPN の構成例を示す．

FPN では、高解像度の特徴マップが持つ位置情報と、低解像度の特徴マップが持つ抽象的な意味情報を、上方向および下方向の伝搬によって統合する．Up-sampling や Down-sampling を通して解像度を揃え、畳み込み層を用いて情報を融合することで、単一スケールでは得られない一貫した特徴表現を形成する．これにより、小物体から大物体まで幅広いサイズの物体に適応可能な検出特性を獲得できる．しかし、FPN は局所的な畳み込み演算を基本とするため、画像全体の文脈や離れた領域間の関係を十分に反映できない場合がある．そのため、特に遠隔領域の情報が重要になる 高密度なシーンや小物体が背景に埋もれる状況では性能が低下することが指摘されている．

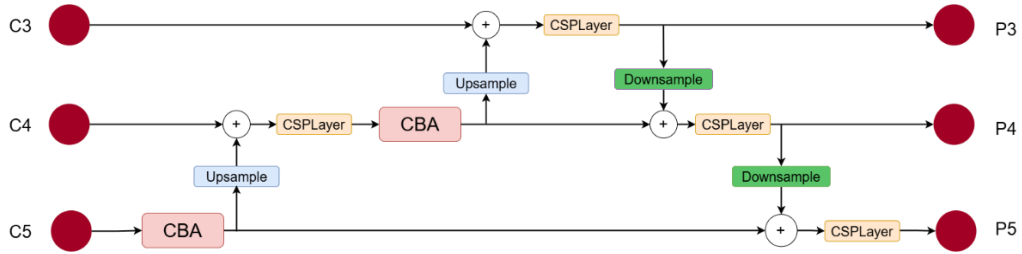


図 2.2 Feature Pyramid Network の構成

2.3 Non-Local Neural Networks

Non-Local Neural Networks は、画像や特徴マップ中の任意の位置間の関係性を考慮するために提案された手法である．従来の畳み込み演算が局所領域に基づいて特徴を抽出するのに対し、Non-Local Neural Networks では、特徴マップ全体を参照することで、離れた位置間の関係性を扱うことが可能である．図 2.3 に Non-Local Neural Networks の構成例を示す．

Non-Local Neural Networks では、入力特徴マップから複数の写像を生成し、各位置間の関係性を計算する．入力特徴、重み行列、および出力特徴は、この処理を通じて関係付けられる．

$$y_i = \frac{1}{C(x)} \sum_j f(x_i, x_j) g(x_j) \quad (1)$$

一般的には、入力特徴から生成される二つの特徴を用いて位置間の相関を算出し、その結果を重みとして用いる．この処理により、ある位置の特徴が、画像全体の他の位置の情報を反映した形で更新される．位置間の関係は、特徴ベクトル間の類似度として定義される．この類似度に基づいて重み行列を構成し、重み付き和を計算することで出力特徴が得

られる.

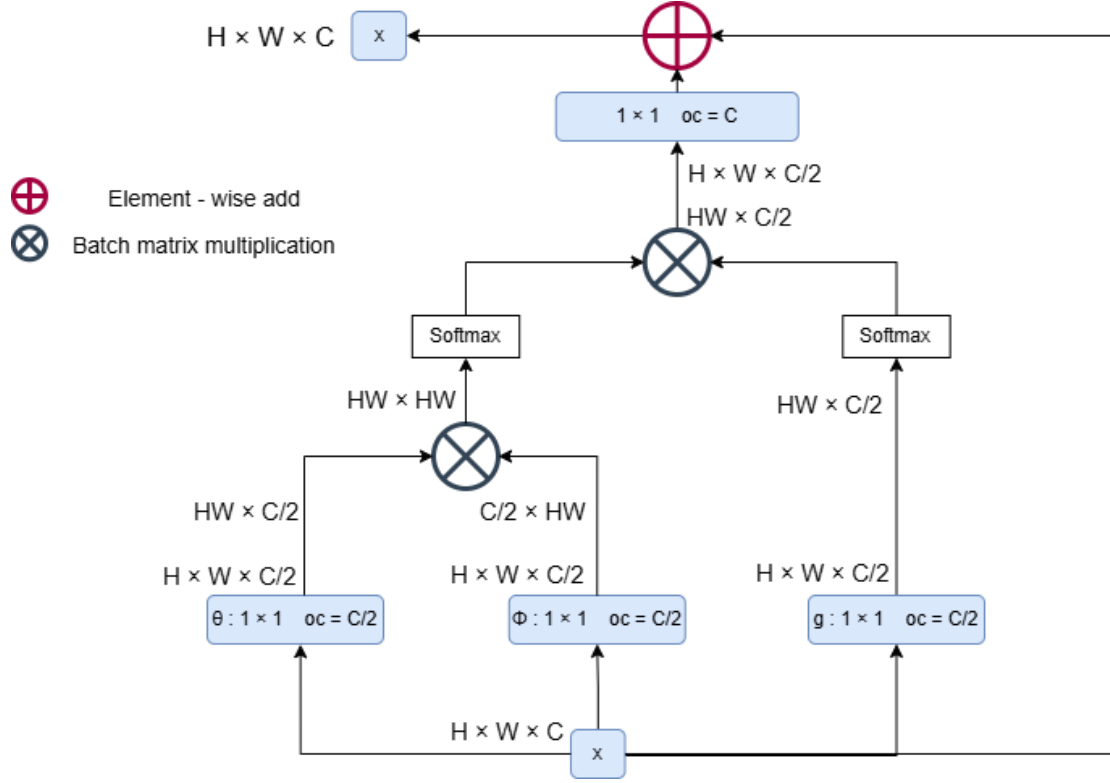


図 2.3 Non-Local Neural Networks における基本構成

図 2.4 に、Non-Local Neural Networks における位置間関係の計算例を示す．図に示すように、空間方向に展開された特徴はベクトル化され、各位置間の対応関係が計算される．この処理により、空間的に離れた領域同士の関係性を明示的に扱うことが可能となる．一方で、空間サイズを $H \times W$ とすると、位置間の関係性を表す行列は $HW \times HW$ のサイズとなる．この処理により、Non-Local Neural Networks では特徴マップ上の各位置間の相関関係が明示的にモデル化される．

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1,HW} \\ \vdots & \ddots & \vdots \\ a_{HW,1} & \cdots & a_{HW,HW} \end{bmatrix} \in \mathbb{R}^{HW \times HW} \quad (2)$$

図 2.5 に、この関係行列の概念図を示す．左側に示す二次元の特徴マップが一次元に展開され、各空間位置が列ベクトルとして並べられる様子を示している．この展開により、一つの空間位置に対して、他のすべての空間位置との関係が一行分の要素として表現され、同様の処理が全空間位置に対して行われることで、最終的に $HW \times HW$ の相関行列が構成される．特徴マップ全体にわたる空間的位置関係を明示的に表現する構造を有している．

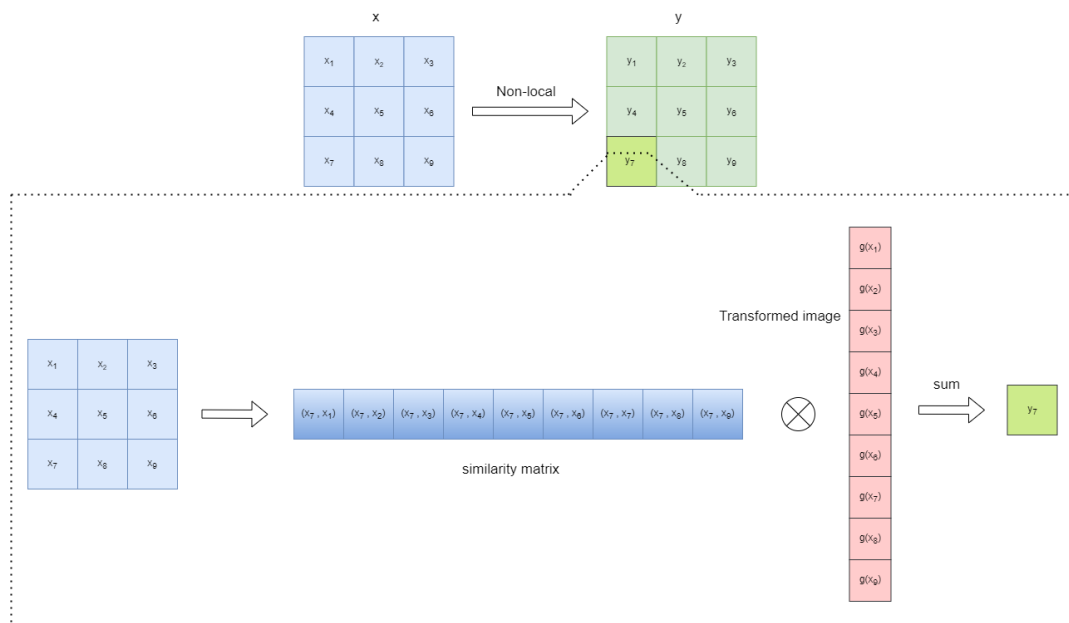


図 2.4 Non-Local 演算における空間位置間の関係の例

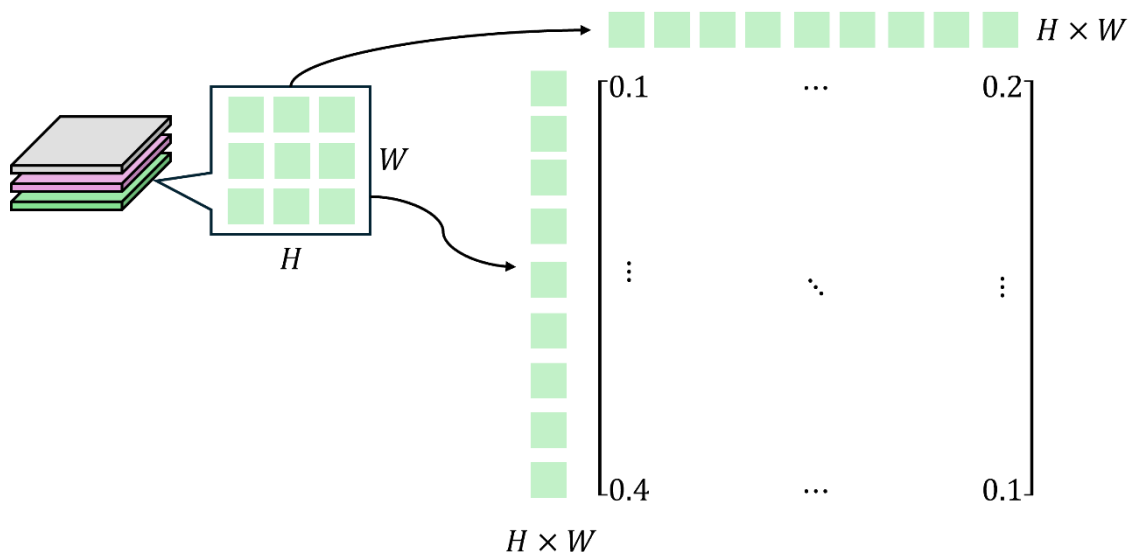


図 2.5 Non-Local Neural Networks における $HW \times HW$ 相関行列の概念図

一方で、Non-Local 演算は、上述のように特徴マップ上の全空間位置間の関係を考慮するため、 $HW \times HW$ の相関行列を明示的に構成する構造を有している．そのため、入力特徴マップの空間解像度が増加するにつれて、相関行列のサイズおよび計算量が急激に増大する．これに伴い、 θ 、 ϕ 、 g による線形変換や相関計算に必要なパラメータ数およびメモリ使用量も増加する傾向がある．その結果、高解像度の特徴マップに対して Non-Local 演算を適用する場合、計算資源の制約を受けやすい構造となる．

2.4 Spatial Bias

Spatial Bias[5]は, 畳み込み特徴に対して空間的な事前情報を付加するために提案された手法である. この手法では, 入力特徴マップから空間方向の情報を効率的に抽出し, 低い計算コストで大域的な空間パターンを扱うことを目的としている.

まず入力特徴マップをチャンネル方向に圧縮し, 低解像度の特徴マップを生成する. 次に, 空間次元を展開した特徴に対して一次元畳み込みを適用することで, 画像全体にわたる空間的な関係をモデル化する. この処理により得られた特徴は, 再び空間構造を持つマップへと整形され, アップサンプリングを通じて元の解像度に復元される. 図 2.6 に, Spatial Bias における特徴生成および結合の流れを示す. 入力特徴から生成された Spatial Bias マップは, 元の特徴マップとチャンネル方向に結合され, 後段の畳み込み処理に入力される.

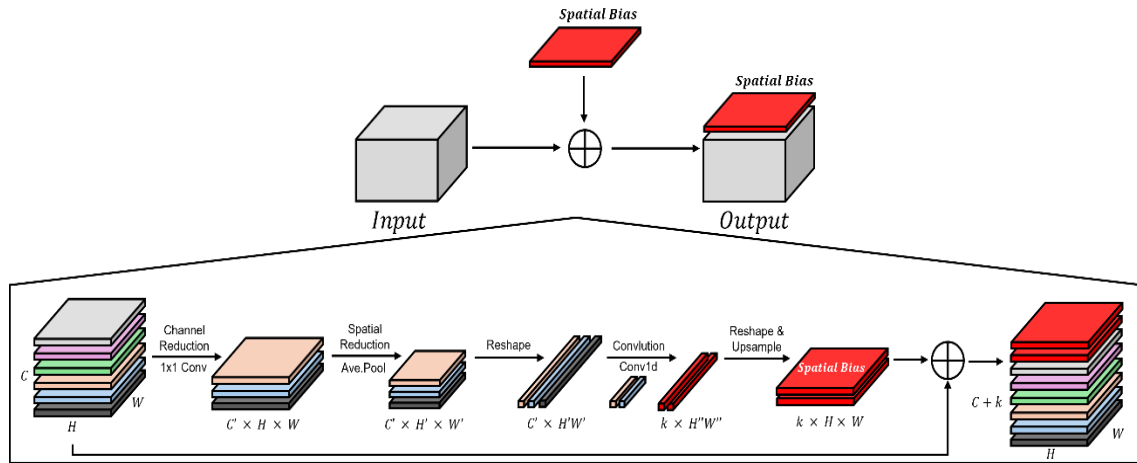


図 2.6 Spatial Bias の概要図

以上の処理により, 入力特徴に対して大域的な依存関係を考慮した特徴表現が得られる. ここで, X は入力特徴マップ, SB は X から生成された Spatial Bias マップを表す. また, $\oplus$ はチャンネル方向の結合を示し, $BN(\cdot)$ はバッチ正規化を表す.

$$Y = BN(X \oplus CB) \quad (3)$$

Spatial Bias は, 重い注意機構を用いることなく, 空間的な大域情報を特徴表現に付加できる点に特徴がある. そのため, 計算量やパラメータ数の増加を抑えた構成でありながら, 物体検出などのタスクへの適用が可能である. このような特性は, 高解像度特徴マップを扱う検出モデルにおいても導入しやすい利点といえる.

2.5 むすび

本章では, 本研究に関連する物体検出技術として, YOLO 系列に代表される単段検出器, Feature Pyramid Network による多尺度特徴統合手法, Non-Local Neural Networks による大域的文脈情報の導入手法, および軽量な空間的特徴強化手法である Spatial Bias に関する既存研究について述べた.

第3章 提案手法

3.1 まえがき

本章では，物体検出モデルにおける特徴表現の強化を目的とし，二つの提案手法について述べる．提案手法はいずれも，既存の YOLO 系列モデルを基盤とし，特徴マップに付加的な情報を導入することで，小物体検出性能の向上を図るものである．本研究では，YOLOX[6]における Feature Pyramid Network を対象とした Non-Local Neural Networks を用いた特徴強化手法と，YOLOv11[7]および YOLOv12[8]を対象とした Channel Bias を用いた特徴表現拡張手法をそれぞれ提案する．各手法の構成および適用方法について，次節以降で詳細に説明する．

3.2 提案手法の概要

本研究で扱う一つ目の手法は，特徴マップ内の離れた位置同士の間接関係を明示的に扱うことで，空間方向の特徴補完を行う方法である．具体的には，Non-Local Neural Networks を特徴抽出の段階に導入し，局所的な畳み込みが捉えにくい長距離依存を考慮可能にする構成とする．これにより，局所特徴が不足する領域に対しても広域的な特徴補強が期待できる．

二つ目の手法は，チャンネル方向に着目した特徴補強方法であり，各チャンネルの応答傾向を要約した補助情報を生成し，元の特徴マップに結合することで表現能力を拡張する Channel Bias 手法である．この方法では，チャンネル間の関係性に基づく補助特徴を導入することで，小物体のように空間情報が乏しい状況においても有用な手掛かりを提供する．

これら二つの手法は，基本的には多くの畳み込みベースの物体検出モデルに適用可能であり，本研究ではその応用例の一つとして YOLO 系統のモデルを対象とした実験により有効性を確認する．

3.3 FPN に Non-Local Neural Networks を導入した手法

3.3.1 Non-Local Neural Networks の導入位置

畳み込みベースの物体検出モデルでは，局所的な受容野に基づく特徴抽出が中心となるため，離れた領域間の関係性を十分に扱えない場合がある．この性質は特に，小物体が背景に埋もれやすい場面において問題となる．本研究では，この課題に対処するため，特徴マップ内の長距離依存関係を補う手法として Non-Local Neural Networks を導入する．

Non-Local Neural Networks は，入力特徴マップ全体を対象として，各空間位置間の相互関係を計算し，広域的な特徴情報を付加する構造を持つ．また，既存のネットワーク構造を

FPN において本処理を挿入することで，異なる解像度の特徴を統合する際に，各空間位置に存在する潜在的な関連情報を補完できる．生成された特徴マップは元の処理フローに戻され，後段の検出ヘッドに入力される．このように，本手法は既存の FPN 構造に対して Non-Local 演算を付加するのみであり，ネットワーク全体を変更せずに特徴表現を強化できる．

3.4 Channel Bias を用いた特徴強化手法

3.4.1 Channel Bias の基本構成

本節では，提案する Channel Bias の構造および処理手順について述べる．Channel Bias は，特徴マップのチャンネル方向に着目し，空間情報が不足する状況においても利用可能なグローバルな特徴傾向を補助情報として導入する手法である．図 3.5 に本手法の処理の流れを示す．

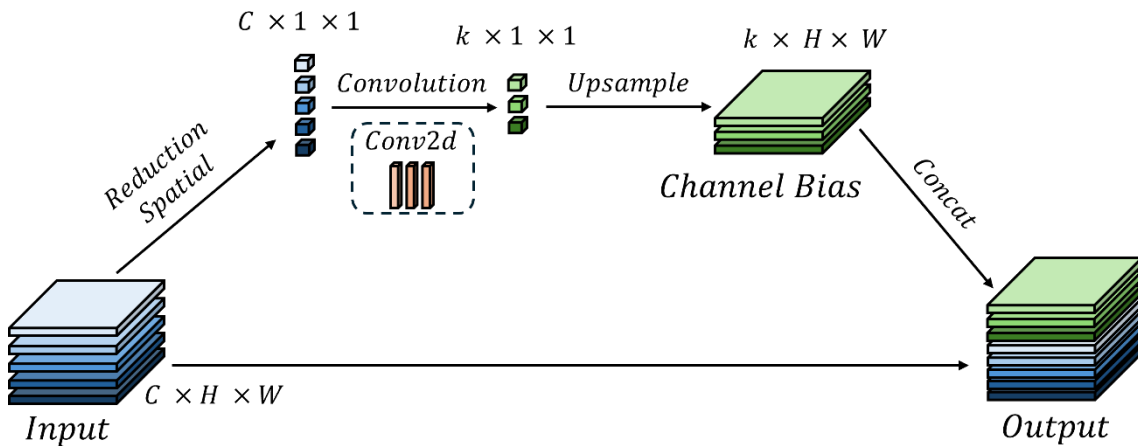


図 3.2 Channel Bias を用いた特徴表現拡張手法の概要

物体が画像中に占める領域が小さい場合，空間方向の情報が限定されるため，畳み込み処理のみでは特徴が弱まり，背景成分に埋もれやすい．本研究では，この問題に対処するため，空間方向ではなくチャンネル方向の統計的特徴に基づく補助情報を生成する Channel Bias を提案する．チャンネル方向の情報は空間に比べ安定性が高く，小物体に対しても特徴の傾向を捉えやすいという利点を持つ．

3.4.2 チャンネル方向特徴の抽出

Channel Bias は，入力特徴マップのチャンネル方向に着目した処理によって生成される．まず，各チャンネルの応答傾向を把握するために，全チャンネルに対してグローバル平均プーリ

ングを適用し、特徴マップ全体のチャンネルごとの代表値を取得する。この処理によって、空間位置に依存しない形で特徴量の分布を要約することができる。

次に、得られたチャンネル統計に対して 1×1 畳み込みを適用し、チャンネル間の関係性を圧縮しながら抽出した Bias 成分を生成する。この段階で、各チャンネルが持つ相対的な重要度や相互作用が反映されるため、小物体に対して不足しがちな空間情報を補う補助特徴として機能する。生成された Bias 成分は、元の特徴マップより少ないチャンネル数で表現され、ネットワークの計算負荷を抑えつつ、特徴傾向の要約を実現する。この補助特徴は次節の統合処理にて利用され、検出性能の向上に寄与することとなる。

3.4.3 特徴マップへの統合

生成された Channel Bias は、入力特徴マップとチャンネル方向に連結することで統合される。連結後の特徴マップは元の空間的な局所特徴と、チャンネル方向に着目した補助情報を同時に保持する構造となり、検出処理における表現力を拡張する。統合後は、後段の処理におけるチャンネル整合性を保つため、 1×1 畳み込みによってチャンネル数を元の構成に戻す。この構成により、ネットワーク全体のアーキテクチャを変更することなく Channel Bias を挿入でき、小物体に対して情報量の不足が生じやすい空間方向の弱点を補い、全体的な検出精度向上への寄与が期待できる。

3.5 むすび

本章では、物体検出における特徴表現の強化を目的とした二つの手法について述べた。一つ目の手法では、Non-Local Neural Networks を導入し、特徴マップ内の遠隔位置間の関係を取り込むことで、局所的な畳み込みでは捉えきれない空間的特徴を補完する構成とした。二つ目の手法では、Channel Bias を導入し、チャンネル方向の統計情報を活用して補助的な特徴を生成し、元の特徴マップに統合することで特徴表現を拡張した。

これらの手法はいずれも、既存のネットワーク構造を変更せずに導入でき、特定のモデルに依存せず汎用的に適用可能である点に特徴を持つ。次章では、これらの手法を実際の物体検出モデルに適用し、その効果を検証した結果について述べる。

第4章 実験結果と考察

4.1 まえがき

本章では, 第3章で述べた二つの提案手法の有効性を評価するために行った実験の概要, 結果, およびその考察について述べる. 本研究では, YOLOX の Feature Pyramid Network に Non-Local Neural Networks を導入した手法と, Channel Bias という手法の二つを提案しており, それぞれ異なる物体検出モデルを対象として実験を行った. 各手法について, 標準的な物体検出評価指標を用いて性能を評価し, 提案手法が検出精度に与える影響について検証する.

4.2 実験環境およびデータセット

本研究における学習および評価は, NVIDIA 製 GPU を搭載した計算環境にて実施した. 使用した環境は, NVIDIA GeForce RTX 2080 Ti を備えた PC であり, OS には Ubuntu 22.04 を用いた. また, 学習および推論の実装には PyTorch を採用した.

実験には, PASCAL VOC2007[9]および VisDrone[10]の2種類のデータセットを使用した. PASCAL VOC2007 は, 20 種類の物体カテゴリから構成されるデータセットであり, 自然画像を中心に多様な場面を含む画像が収録されている. 一方, VisDrone は, 複数の都市において UAV によって撮影された画像で構成されている. 俯瞰視点の画像が中心となっており, 画像内に占める物体の相対的な大きさが小さい点が特徴である. 各データセットに含まれる物体カテゴリの概要を表 4.1 と表 4.2 に示す.

表 4.1 VisDrone に含まれる物体カテゴリ

番号	カテゴリ名
0	pedestrian
1	people
2	bicycle
3	Car
4	van
5	truck
6	tricycle
7	awning-tricycle
8	bus
9	motor



図 4.1 VisDrone Dataset に含まれる空撮画像の一例[10]

表 4.2 PASCAL VOC2007 に含まれる物体カテゴリ

番号	カテゴリ名
0	aeroplane
1	bicycle
2	bird
3	boat
4	bottle
5	bus
6	car
7	cat
8	chair
9	cow
10	diningtable
11	dog
12	horse
13	motorbike
14	person
15	pottedplant
16	sheep
17	sofa
18	train
19	tvmonitor



図 4.2 PASCAL VOC2007 の例 (研究用途において公開可[9])

4.3 評価指標

本研究では，物体検出性能を評価する指標として Average Precision (AP)および mean Average Precision (mAP)を用いる．AP は，各クラスにおける Precision と Recall の関係から得られる Precision-Recall 曲線の下面積として定義される．Precision と Recall の算出に用いる予測結果の分類は，図 4.3 に示す混同行列に基づき整理した．

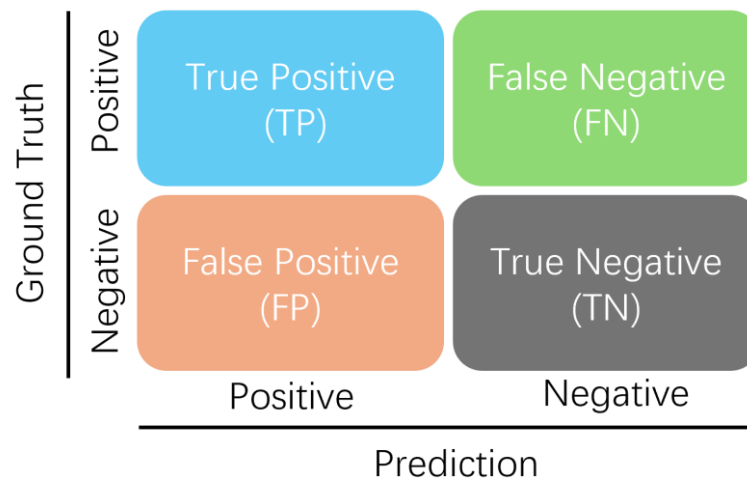


図 4.3 二値分類の混同行列

適合率 (Precision)は，Positive と判定された予測のうち，実際に Positive であった比率を示す指標である．

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

再現率 (Recall)は, 真のラベルが Positive である事例に対して, Positive と予測できた割合を表す指標である.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

平均適合率 (AP)は, Precision-Recall 曲線における下面積として定義され, Recall に対する Precision の関係を積分することで定義される.

$$AP = \int_0^1 P(R) dR \quad (6)$$

予測されたバウンディングボックスと正解領域の重なりを評価する尺度として Intersection over Union (IoU)を用いる. IoU は, 予測ボックスと正解ボックスの重なり領域の面積を, それらの和集合の面積で割った値として定義される. IoU の概念図を図 4.4 に示す.

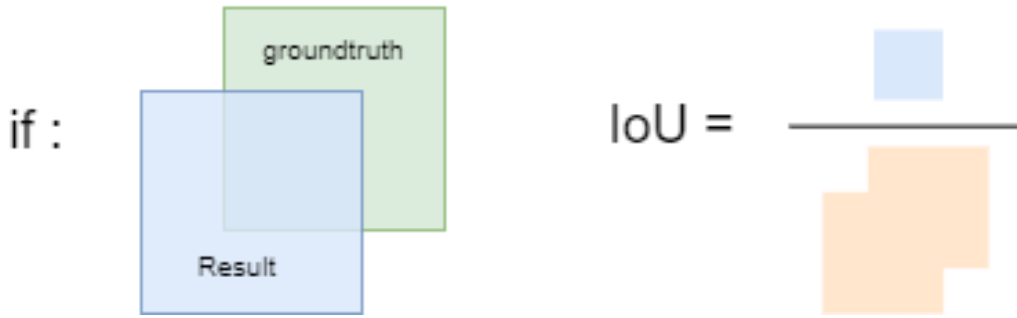


図 4.4 予測領域と正解領域における IoU の概念図

$$IoU = \frac{Result \cap Groundtruth}{Result \cup Groundtruth} \quad (7)$$

本研究では, IoU 閾値を 0.50 から 0.95 まで 0.05 刻みに変化させ, 各閾値における AP を平均した mAP を評価指標として用いる.

$$mAP_{50:95} = \frac{1}{T} \sum_{t=1}^T AP(t) \quad T = 10: \{0.50, 0.55, \dots, 0.95\} \quad (8)$$

4.4 FPN における Non-Local Neural Networks の導入に関する検討

4.4.1 実験設定

本節では、Feature Pyramid Network (FPN) 内の異なる位置に Non-Local Neural Networks (以下 Non-Local) を導入した場合の影響を検証する。比較対象として、Non-Local を導入しない YOLOX をベースラインとし、FPN の P5, P4, P3 にそれぞれ 1 か所ずつ導入した 3 種の構成について評価を行った。評価には VOC2007 データセットを用い、AP を指標とした。

4.4.2 全体精度の比較

実験設定に基づき、Non-Local の導入位置によって検出性能がどのように変化するかを比較した。各構成における検出結果を表 4.3 に示す。本評価では、Non-Local Neural Networks を導入しないベースラインと、各 FPN 層 (P5, P4, P3) にそれぞれ導入した 3 種類の構成を比較対象とし、精度がどの程度向上あるいは低下するかを確認した。

表 4.3 VOC2007 を対象とした Non-Local 導入位置による検出精度の比較

Method	$AP_{50} \uparrow$	$AP_{75} \uparrow$	$mAP_{50-95} \uparrow$
baseline	58.3	32.1	32.2
strategy 1	58.7 (+0.4)	33.0 (+0.9)	32.7 (+0.5)
strategy 2	57.4 (-0.9)	31.6 (-0.5)	32.1 (-0.1)
strategy 3	55.6 (-2.7)	29.9 (-2.2)	30.3 (-1.9)

表 4.3 より、P5 に Non-Local を導入した構成において mAP が最も高く、ベースラインに対し改善が確認された。特に、高解像度の特徴を保持する層において非局所的な情報が補完されることで、全体的な検出性能が向上していると考えられる。一方、P4 あるいは P3 に導入した場合には、mAP が低下する結果となった。これは、浅い層では既に局所的な特徴が十分抽出されており、Non-Local による補助効果が限定的となる可能性を示唆している。これらの結果から、高次の特徴マップにおいて Non-Local が有効に働く傾向が示唆される。

4.4.3 対象スケール別の比較

対象物体の大きさに着目し、Non-Local の導入位置が検出性能に与える影響を詳細に確認する。小、中、大の物体に対する検出性能の比較結果を表 4.4 に示す。ここでは、各スケールごとに算出された mAP を比較することで、特徴量の階層や表現の違いが検出精度にどの

ように反映されるかを評価する．本比較により，Non-Local の導入が特定の物体サイズに対してどの程度有効に働くかを明確に把握することが可能となる．

表 4.4 VOC2007 における対象スケール別検出精度の比較

Method	All	Small	Medium	Large
baseline	32.2	11.4	22.4	36.3
strategy 1	32.6	15.2	22.2	36.7
strategy 2	32.1	19.8	21.3	36.4
strategy 3	30.3	10.8	19.8	34.2

表 4.4 に示されるように，小物体に対して最も改善が見られる構成は P5 への導入であり，ベースラインに対して精度向上が確認された．一方，中物体および大物体に対しては，改善量は限定的であった．この結果は，Non-Local が空間的情報が限られる領域において相対的な利点を持つことを示している．

4.4.4 結果の考察

FPN 内部の特徴マップの解像度および抽象度を踏まえると，P3 や P4 は，比較的高い空間解像度を維持しており，局所的な特徴のみで識別が可能な場合が多い．そのため，Non-Local の導入による利点が十分に発揮されず，ベースラインを下回る結果となったと考えられる．一方で，P5 はダウンサンプリングにより空間的な情報量が減少しており，局所特徴の不足を補うために広域的な文脈情報が必要となる．Non-Local の導入により長距離依存関係を補完できた結果，小物体を含む検出性能が向上したと推察される．

以上より，Non-Local は高レベル特徴が入力となる深い層に導入した場合に効果が高く，FPN 構造との併用により小物体検出に寄与する可能性が示された．

4.5 Channel Bias 手法の評価

4.5.1 実験設定

本節では，Channel Bias を Ultralytics YOLOv11[11]および YOLOv12[12]に適用した場合の検出性能を評価する．実験には，小物体が多数含まれる VisDrone データセットを使用した．Channel Bias の処理手順および挿入位置の概要を図 4.5 に示す．Channel Bias はバックボーンにおける Layer 3 に挿入し，入力特徴マップに対して 1 回の畳み込み処理を行い，出力チャネル数を 3 に変換した後，元の特徴マップと結合する構成とした．

学習条件は各モデルの標準設定に従い，Channel Bias の適用以外にネットワーク構造およ

びハイパーパラメータの変更は行っていない．また，全ての実験は単一 GPU 環境下で実施し，YOLOv11 および YOLOv12 の比較が同条件となるよう統一した．

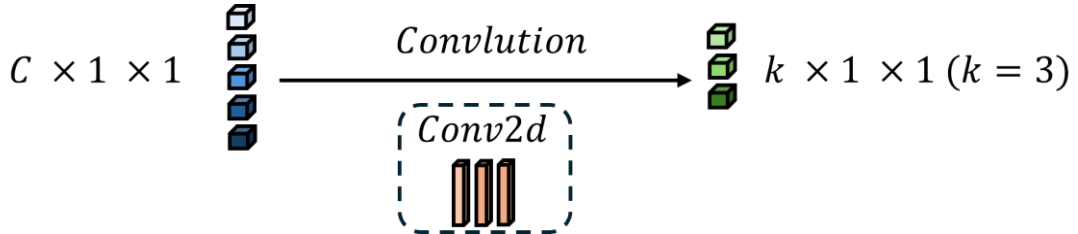


図 4.5 Channel Bias の挿入位置と構成概要

4.5.2 全体精度の比較

YOLOv11 および YOLOv12 に対し Channel Bias を導入した際の mAP の比較結果を表 4.5 に示す．

表 4.5 VisDrone における Channel Bias 導入時の検出性能比較

Method	$AP_{50} \uparrow$	$AP_{75} \uparrow$	$mAP_{50-95} \uparrow$
YOLOv11	36.0	21.7	21.4
YOLOv11 + Channel Bias	36.5	21.7	21.5
YOLOv12	35.7	21.2	21.0
YOLOv12 + Channel Bias	36.4	22.0	21.6

表 4.5 より，両モデルにおいて AP_{50-95} の改善が確認される．YOLOv11 では改善量は小さいものの一貫性があり，Channel Bias による特徴補助が検出精度に寄与していることが示唆される．一方，YOLOv12 では， AP_{50} および AP_{75} の双方で改善が確認され，YOLOv11 に比べて効果が明確に表れている．これは，より深い特徴表現を利用するモデルにおいて，Channel Bias が活用されやすい可能性を示している．

4.5.3 結果の考察

Channel Bias は，チャネル方向に付加的な情報を付与することで，元の特徴マップに対し追加の手掛かりを与える．この処理は空間的な解像度に依存しないため，YOLOv11 および YOLOv12 のいずれに対しても同一方式で導入できる点が利点である．また，モデル構造の変更や計算量の増加を伴わずに精度改善が得られることから，軽量かつ実装が容易な拡張手法として有効性が確認できる．

以上の結果より，Channel Bias は小物体検出を含む場面において有効に機能する追加モジ

ジュールとして位置づけられ、既存モデルに対して簡潔に組み込める実用性を備えていると言える。

4.6 むすび

本章では、FPN における Non-Local Neural Networks の適用と、Channel Bias による検出性能の改善について述べた。Non-Local を用いた検証では、FPN の深い層に適用した場合に精度向上が確認され、高レベル特徴を扱う段階において長距離依存関係が有効に働くことが示された。一方、浅い層では精度が低下し、局所特徴が十分に保持されている段階では追加効果が限定的であることを明らかにした。

また、Channel Bias を用いた検証においては、VisDrone データセットを対象とした評価の結果、YOLOv11 および YOLOv12 の双方で検出精度の向上が確認された。特に YOLOv12 において改善が大きく、チャネル方向の情報を補足することで小物体検出に寄与することが示された。両手法は複雑なネットワーク改変を伴わずに適用可能であり、既存のモデルに対して柔軟に組み込める手法として有効性を持つことが示された。

第5章 結論と今後の展望

5.1 結論

本研究では、物体検出モデルにおける小物体検出性能の向上を目的とし、二つの特徴強化手法を提案した。一つ目の手法として、YOLOX の Feature Pyramid Network に Non-Local Neural Networks を導入し、空間位置間の関係を明示的に考慮することで特徴表現を強化する手法を提案した。二つ目の手法として、YOLOv11 および YOLOv12 を対象に、Channel Bias を導入することで、チャンネル方向の情報を集約したグローバル特徴を付加し、特徴表現を拡張する手法を提案した。

実験では、YOLOX に Non-Local Neural Networks を導入した場合、FPN の導入位置を適切に選択することで、小物体検出精度の向上が確認された。一方で、導入位置によっては性能低下が生じる場合もあり、Non-Local モジュールの効果が特徴マップの解像度や情報構造に依存することが示された。これにより、Non-Local 演算は高解像度特徴マップにおいて特に有効である一方、適用には構造的な配慮が必要であることが明らかとなった。

また、YOLOv11 および YOLOv12 に Channel Bias を導入した実験では、いずれのモデルにおいても検出性能の安定した向上が確認された。特に、小物体が多数含まれるデータセットにおいて、mAP 指標の改善が見られ、Channel Bias が空間解像度に依存しない軽量な特徴強化手法として有効であることが示された。本手法は、既存のネットワーク構造を大きく変更することなく適用可能であり、計算コストの増加を抑えたまま特徴表現を補強できる点に特徴がある。

これらの結果から、本研究で提案した二つの手法はいずれも、物体検出モデルにおける特徴表現の強化に有効であり、特に小物体検出性能の向上に寄与する可能性が示された。Non-Local Neural Networks は空間的な文脈情報の獲得に有効であり、Channel Bias はチャンネル方向の情報集約による補助的な特徴付加として有効であることが確認された。本研究は、物体検出モデルの構造に応じて適切な特徴強化手法を選択・設計することの重要性を示した点に意義があるといえる。

5.2 今後の展望

本研究では、YOLO 系列の物体検出モデルを対象として、二つの異なる特徴強化手法の有効性を検証した。今後は、提案した各手法を他の物体検出モデルや異なるネットワーク構造に適用し、その汎用性について検証を進めることが課題として挙げられる。特に、Non-Local Neural Networks に関しては、バックボーンネットワーク内部への導入や、より軽量な非局所演算への置き換えを検討することで、計算コストと性能向上の両立を図ること

が期待される.

また, Channel Bias 手法については, VisDrone に加えて UAVDT [13]など, 異なる種類のデータセットや撮影条件における評価を行い, 小物体検出以外のタスクに対する有効性についても検証を行う必要がある. 加えて, 複数の特徴強化手法を組み合わせた場合の影響や, 学習過程における振る舞いの解析を通じて, より効果的な特徴表現の設計指針を明らかにすることが今後の課題である.

これらの検討を通じて, 本研究で提案した手法をさらに発展させ, 実環境における物体検出精度の向上に貢献することを目指す.

謝辞

本論文の執筆にあたり，日々のゼミ活動を通じて研究の方向性や課題についてご指導いただき，また快適な研究環境を整えてくださった渡辺裕教授に深く感謝申し上げます。

また，毎週のゼミ活動を通じて，研究の進め方や実験方法に関して多くの助言とご支援をいただきました博士後期課程在学中の劉浩元氏に，心より感謝申し上げます。

さらに，多岐にわたる研究領域を共有していただき，日頃から温かく接してくださった渡辺研究室の皆様に深く感謝申し上げます。

最後に，これまで私を育てていただき，早稲田大学での学業，サークル活動，そして研究生活を支えてくださった家族に心から感謝いたします。

参考文献

- [1] M. Nikouei, B. Baroutian, S. Nabavi, F. Taraghi, A. Aghaei, A. Sajedi, and M. E. Moghaddam, “Small object detection: A comprehensive survey on challenges, techniques and real-world applications,” *Intelligent Systems with Applications*, vol. 27, 200561, Sep. 2025, arXiv:2503.20516.
- [2] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *Computer Vision and Pattern Recognition (CVPR)*, pp.779-788, Jun. 2015.
- [3] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, Jul. 2017.
- [4] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794–7803, Jun. 2018.
- [5] J. Go and J. Ryu, "Spatial bias for attention-free non-local neural networks," *Expert Systems with Applications*, vol. 238, p. 122053, Mar. 2024.
- [6] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “Yolox: Exceeding YOLO series in 2021,” *arXiv preprint arXiv:2107.08430*, Jul. 2021.
- [7] M. Khanam and N. Hussain, “YOLO11: An overview of the key architectural enhancements,” *arXiv preprint arXiv:2410.17725*, Oct. 2024.
- [8] Y. Tian, Q. Ye, and D. Doermann, “YOLOv12: Attention-centric real-time object detectors,” *arXiv preprint arXiv:2502.12524*, Feb. 2025.
- [9] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (VOC) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
- [10] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, “Detection and tracking meet drones challenge,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7380–7399, Nov. 2022.

- [11] G. Jocher and J. Qiu, “ultralytics/yolov11,” (最終閲覧日 : 2026 年 1 月 15 日) , <https://github.com/ultralytics/ultralytics>
- [12] Y. Tian, Q. Ye, and D. Doermann, “sunsmarterjie/yolov12,” (最終閲覧日 : 2026 年 1 月 15 日) , <https://github.com/sunsmarterjie/yolov12>
- [13] D. Du, Y. Zhu, Y. Yong, Y. Duan, K. Li, G. Zhang, W. Huang, and Q. Tian, “The unmanned aerial vehicle benchmark: Object detection and tracking,” in Proceedings of the European Conference on Computer Vision (ECCV), pp. 370–386, Sep. 2018.

表一覧

表 4.1	VisDrone に含まれる物体カテゴリ	13
表 4.2	PASCAL VOC2007 に含まれる物体カテゴリ	14
表 4.3	VOC2007 を対象とした Non-Local 導入位置による検出精度の比較	17
表 4.4	VOC2007 における対象スケール別検出精度の比較	18
表 4.5	VisDrone における Channel Bias 導入時の検出性能比較	19

図一覧

図 2.1	YOLO モデルにおけるネットワーク構成例.....	3
図 2.2	Feature Pyramid Network の構成.....	4
図 2.3	Non-Local Neural Networks における基本構成.....	5
図 2.4	Non-Local 演算における空間位置間の関係の例.....	6
図 2.5	Non-Local Neural Networks における $HW \times HW$ 相関行列の概念図.....	6
図 2.6	Spatial Bias の概要図.....	7
図 3.1	YOLOX の FPN における Non-Local モジュールの導入位置.....	10
図 3.2	Channel Bias を用いた特徴表現拡張手法の概要.....	11
図 4.1	VisDrone の例.....	14
図 4.2	VOC2007 の例.....	15
図 4.3	二値分類の混同行列.....	15
図 4.4	予測領域と正解領域における IoU の概念図.....	16
図 4.5	Channel Bias の挿入位置と構成概要.....	19

研究業績

- [1] 范洪睿, 渡辺裕, “残差 MLP と注意機構を組み合わせた点群分類手法の検討,” 2024 年 電子情報通信学会総合大会, D-12B-13, Mar. 2024.

- [2] H. Fan, H. Liu, and H. Watanabe, “Enhancing Object Detection with Non-Local Modules in The Feature Pyramid Network,” IEEE Global Conference on Consumer Electronics, OS-AIP(3), pp.890-893, Oct. 2024.

- [3] H. Fan, H. Liu, and H. Watanabe, “A Lightweight Channel Bias for Small Object Detection in Aerial Images,” International Workshop on Advanced Image Technology, Jan. 2026.