令和 7 年度 卒業論文

YOLOv8 を用いたアテンション機構導入による炎および煙の検出精度向上に関する研究

2026 年 2 月 12 日

静岡大学 工学部 数理システム工学科

岡部研究室

5021-6002 青木 健

目次

第1章　はじめに

近年，火災による被害は，社会全体の安全・安心を脅かす重大な脅威である．火災は一度発生すれば，人的被害のみならず，歴史的建造物や重大な自然資源，さらには莫大な経済的損失をもたらす．こうした被害を最小限に食い止めるためには，火災の発生を極めて早期に，かつ正確に検知し，初期消火や迅速な避難誘導へとつなげることが必要不可欠である．従来，火災検知の主役を担ってきたのは，煙粒子を光学的に捉える煙感知器や，温度上昇を感知する熱感知器といった物理センサであった．しかし，これらのセンサには物理的制約が存在する．例えば，天井の高い大規模な倉庫，開放的な屋外施設，あるいは広大な森林地帯などでは，煙や熱がセンサに到達するまでに時間を要し，検知が遅れるケースが少なくない．また，気流の影響を受けやすい環境では，発生場所の特定が困難になるという課題もある．こうした背景から，監視カメラやドローンによる映像インフラを有効活用し，画像認識によって火災を視覚的に検知する手法や，火災用のデータセットの構築が注目を集めている[1]，[2]，[3]，[4]，[5]，[6]，[7]．カメラを用いた手法は，火災が発生した瞬間の視覚的変化を捉えることができるため，物理センサよりも迅速な検知，対応が可能であり，かつ発生場所を画像上で直ちに特定できるという利点を持つ．

しかし，画像認識による火災検知には特有のむずかしさがある．炎や煙は，車両や歩行者のような固定的・定型的な形状を持たない不定形の対象物である．煙は風の流れによって拡散し，周囲の背景と混ざり合うことでコントラストが低下する．炎は証明条件や周囲の反射物の影響を強く受け，夕日や赤色の照明，あるいは動く赤いとの物体などと誤認を招きやすい．特に遠方で発生した小さな火種や，立ち上がり始めの希薄な煙を精度良く検出することは，従来の画像処理技術での課題である．近年，この課題を解決する手段として，深層学習を用いた物体検知技術が飛躍的な発展を遂げている．特に YOLO（You Only Look Once）シリーズは，単一のネットワークで物体の位置特定と分類を同時に行うアルゴリズムであり，高いリアルタイム性と検出精度の両立を実現している[8]．YOLOv8[9]は，優れたアーキテクチャにより多様な物体検知タスクで成果を挙げているが，火災検知という極めて高い信頼性が求められる領域においては，さらなる精度向上の余地が残されている．特に，複雑な背景から重要な特徴のみを抽出する能力を強化することが，誤検知の抑制と検出率の向上に直結する．

本研究では，YOLOv8 をベースモデルとし，そこにアテンション機構（Attention Mechanis）を統合することで，煙および炎の検出能力を高度化させたシステムを構築することを目的とする．アテンション機構とは，人間が視覚情報の中から特定の重要な部分に注視するように，ニューラルネットワーク（CNN）が重要な特徴量に対して動的に重み付けを行う手法である．本研究では具体的に，空間的な位置情報を保持しつつ重要な特徴を強調する Coordinate Attention（CA）と，計算コストを最小限に抑えながらチャンネル間の相関関係を適応的に調整する Efficient Channel Attention （ECA）の 2 種類に着目した．これらの機構を

YOLOv8 内に効果的に配置することで，不定形な煙の拡散パターンや，炎の輝度分布といった火災特有の視覚的特徴を，背景ノイズから分離して抽出することを目指した．本研究では，提案モデルの有効性を検証するため，多様な火災シーンを含むデータセットを用いた学習および評価実験を行い，標準的な YOLOv8（baseline）と比較して，適合率（Precision），再現率（Recall），および mAP といった主要な指標においてどの程度の改善が見られたかを詳細に分析する．

本論文の構成は以下の通りである．第 2 章では，これまでの火災検知手法の変遷から，近年の YOLO を用いた火災検知手法，およびアテンション機構に関する関連研究を整理する．第 3 章では，本研究で提案するシステムの核となる YOLOv8 の構造，および Coordinate Attention，Efficient Channel Attention の演算プロセスについて詳述する．それを踏まえて，第 4 章では，自作アテンション機構 ESCFBlock の演算プロセスと実装詳細について述べる．第 5 章では，実験環境，データセットの詳細，および提案手法による評価実験の結果を示し，baseline モデルとの比較を通してその性能を多角的に評価する．最後に第 6 章において，本研究で得られた知見をまとめ，現状の課題と今後のさらなる精度向上に向けた展望を述べる．

第2章　関連研究

本章では，YOLO を用いた火災検知についての関連研究を概説する．また，アテンション機構導入に対する成果，課題を概説する．

2.1　YOLO を用いた火災検知

これまでにも，YOLO を用いた火災検知に関する研究は 数多く行われている．例えば，Dou らは YOLOv5[10] を ベースとした火災検知モデルを提案し，多様な環境下での ロバスト性を検証した [11]．また，Gao らは YOLOv8 に対 して，双方向での特徴融合を実現するために，ネットワー クを再設計し [12]，異なるスケールの火災に対しても，高い 検知能力を示すことが報告している．さらに Ma らは，計 算資源の限られたデバイスへの実装を目的とした YOLOv8 の軽量化研究を行った [13]．しかしながら，これらの手法 において，複雑な背景下での誤検知抑制や，微小な火災の 特徴の検出には課題が残る．

2.2 アテンション機構による精度向上と課題

前節の課題に対し，特定の重要な特徴マップを強調するアテンション機構の導入が，検知精度向上のための主 要なアプローチとなっている．Wang らは ResNet[14] に Squeeze-and-Excitation（SE）[15] ブロックを統合した SEResNet を用いた手法を提案し，森林火災の検知精度を大幅 に向上させた [16]．SE ブロックに代表されるチャネルアテンションは，どのチャネルが火災検知において重要かを学習することで，背景ノイズの影響を抑制する効果がある． 一方で，SE ブロックのような単純なチャネルアテンションは，Global Average Pooling を用いて空間情報を完全に圧縮してしまうため，火災が発生している位置に関する詳 細な情報を保持できないという弱点がある．この空間情報 を補うために，Gao らは空間とチャネルの情報を考慮する CBAM（Convolutional Block Attention Module）[17] などの統合型アテンションを YOLO に導入する手法も提案 している [12]．しかし，既存の統合型のアテンション機構は，空間情報の抽出に，比較的大きなカーネルを使用することが多く，計算負荷が増大し，YOLO のリアルタイム性を損なってしまうという懸念もされている．このように，火災検知における既存のアテンション導入の研究においては，空間情報の保持能力と計算コストの抑制というトレードオフの制約を受けており，解決すべき重 要な課題となっている．

第3章　事前知識

本章では，本研究で提案する火災検知モデルの基盤となる要素技術について概説する．具体的には，まずベース モデルとして採用した，物体検知アルゴリズム YOLOv8 [9] のアーキテクチャについて述べる．続いて，提案手法 （ESCFBlock）の構成要素となる 2 つのアテンション機構，Coordinate Attention（CA）[18] および Efficient Channel Attention（ECA）[19] の基本原理と演算プロセスについて詳述する．

3.1　YOLOv8

3.1.1　YOLOv8 の概要

YOLOv8 は，2023 年に Ultralytics 社によって公開された，リアルタイム物体検出アルゴリズム YOLO（You Only Look Once）シリーズのモデルである．YOLOv8 は，広く普及している YOLOv5 等の従来モデルを継承しつつ，ネットワークアーキテクチャの抜本的な刷新により，検出精度（mAP）と推論速度の両面で大幅な向上を達成している．また，単一の Python パッケージおよび CLI（Command Line Interface）を通じた一貫性のあるプラットフォームが提供されており，モデルの実装や拡張性に優れている点も大きな特徴である．最新モデルは YOLO26，YOLO11 などがあげられる．

3.1.2　アーキテクチャの構造

YOLOv8 のネットワークは，主に Backbone，Neck，Head の 3 つのコンポーネントで構成されている．以下の図 3.1 に YOLOv8 の全体構造を示す．

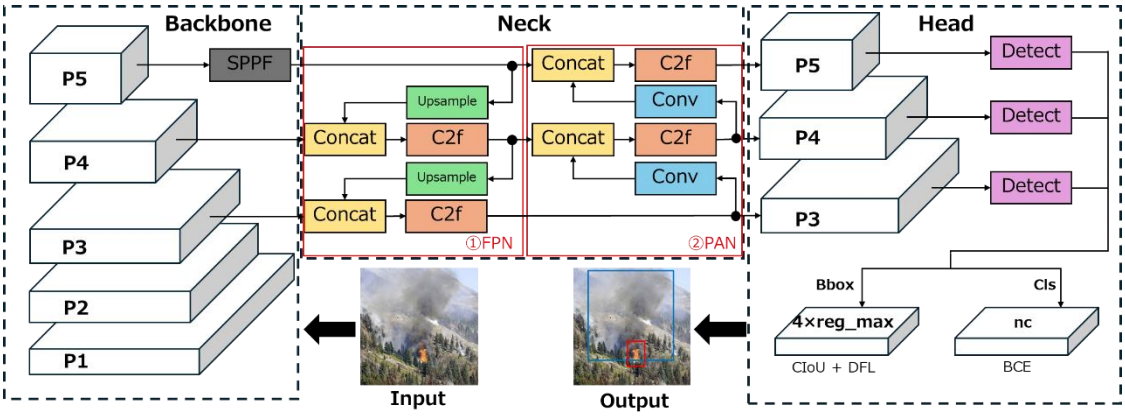

図 3.1 YOLOv8 の全体構造図

■Backbone

Backbone は，入力画像からマルチスケールな特徴抽出を行う役割を担う．CSPNet（Cross Stage Partial Network）アーキテクチャを基盤とし，ピラミッド状の階層構造を通じて，浅い層（P1 や P2）では視覚的詳細情報を，深い層（P5）では物体としての意味的情報を段階的に抽出する．

■Neck

Neck は Backbone によって抽出された異なるスケールの特徴マップを統合, 精製する役割を担う．YOLOv8 では図 3.1 の①，②で示される 2 つの経路を組み合わせることで，情報の洗練を行っている．

まず，図 3.1 中の①は FPN（Feature Pyramid Network）と呼ばれる構造である．ここでは，Backbone の深い層で抽出された意味的情報を，解像度の高い，浅い層へと伝達する．これにより解像度が高いものの，意味の理解が乏しかった浅い層において，小さな物体の識別能力が大幅に向上する．

続いて，②は PAN（Path Aggregation Network）と呼ばれる構造である．ここでは，FPN とは逆に, 浅い層が持つ正確な物体の輪郭や, 色, 形状といった視覚的情報, 位置情報を再び, 深い層へと運ぶ．この浅い層から，深い層での情報のフィードバックにより，深い層においても，物体の位置を数ピクセル単位で正確に特定できるようになる．

これら①（深い層から浅い層）と②（浅い層から深い層）の双方向の情報の往復によって，全ての階層が，視覚的情報および意味的情報の両方に精通した，より豊かな特徴マップをもつことが可能になる．

■Head

Head は，Backbone と Neck によって精製された多角的な特徴マップを受け取り，最終的な物体の位置特定（バウンディングボックス（Bbox）の予測）およびクラス分類を行う役割を担う．YOLOv8 の Head には，検出性能を最大化するための以下の技術的特徴がある．

1 つ目は，Decoupled Head の採用である．図 3.1 の Head 部分に示す通り，特徴マップ P3, P4，P5 の各スケールに対して Detect モジュールが配置されている．この内部では，物体の位置特定を特定する「Bbox（位置特定）」と，物体の種類を判別する「Cls（分類）」という 2 つの処理が完全な別経路で行われている（次節で詳述する）．これにより，位置の正確さと分類の精度の両立が可能となっている．Bbox 側では，出力サイズが$4 \times \mathrm{reg_max}$となっている．これは，バウンディングボックスの 4 辺の位置を，単なる座標ではなく reg_max という一定の範囲（分布）として予測することを意味している．これに DFL（Distribution Focal Loss）を適用することで，境界のあいまいさを抑制している．また，Cls 側では，クラス数を示す nc（本研究では，炎/煙の 2 クラス）が最終的な出力サイズをなり，各カテゴリの該当確率を算出する．

2 つ目はアンカーフリー手法への移行である．YOLov8 では，あらかじめ定義された固定

枠（アンカーボックス）を使用せず．物体の中心点を直接予測し，そこから境界までの距離を算出する．これにより，炎や煙などの不定形な形状にも対応することが可能となった．これらの予測値は，学習時に CloU や DFL ， BCE といった損失関数によって最適化される．特に Bbox 側では枠の重なりだけでなく形状のゆがみまで考慮する CloU と，分布の予測のずれを補正する DFL を併用することで，極めて精密な位置特定を実現している．

3.2 Coordinate Attention

3.2.1 Coordinate Attention の概要

Coordinate Attention（CA）は，モバイルネットワークのような計算リソースが制限された環境下で，物体の識別および位置特定に関する精度を効率的に向上させるために提案されたアテンション機構である．従来のアテンション機構がチャネル間の相互作用のみを重視していたのに対し，CA はチャネル依存関係と空間情報を同時に符号化することが可能である．これによりモデルは対象物の所在を，より正確に捉えることが可能であり，複雑な背景下における検知能力が向上する．

3.2.2 演算プロセス

CA の演算は以下のステップで構成される．また，CA の演算プロセスの概要を図 3.2 に示す．

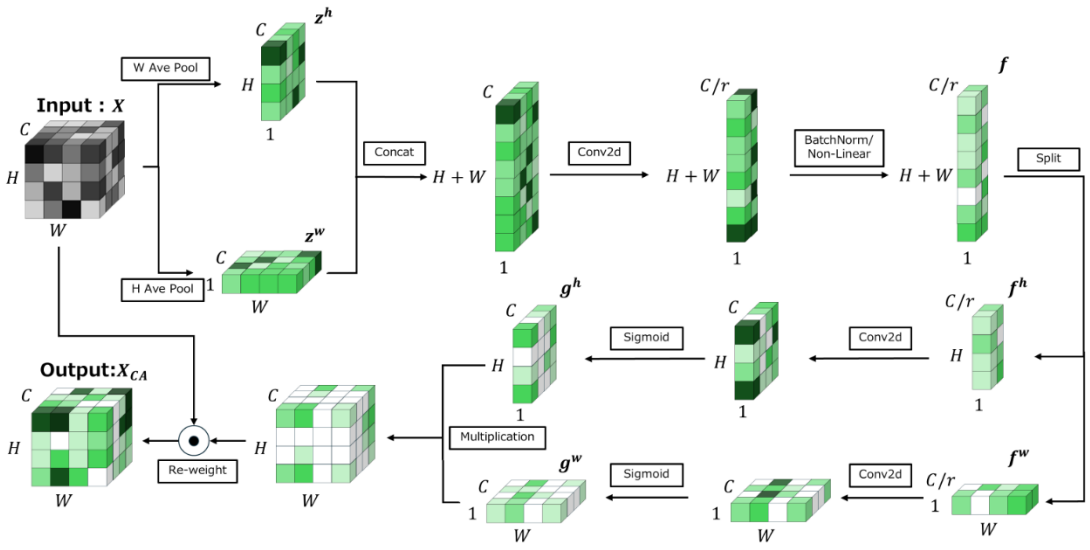

図 3.2 Coordinate Attention の演算プロセスの概要図

（1） 位置情報の集約：Global Average Pooling

通常のアテンション機構で用いられる 2 次元のグローバルプーリングは, 空間情報を 1 つのチャネル数値に圧縮するため，位置情報が失われるという欠点がある．CA ではこの問題を解決するため，図 3.2 に示すように入力特徴マップに対して，水平方向（Width）と垂直方向（Hight）のそれぞれに独立した，グローバル平均プーリング（Global Average Pooling, GAP）を適用する．それぞれ，水平方向は W Ave Pool，垂直方向は H Ave Pool である．入力特徴マップ$X = [x_1, x_2, \dots, x_C]$（サイズ$C \times H \times W$）に対し，チャネル$c$における高さ$h$の出力$z_c^h$，および幅$w$の出力$z_c^w$は次式で定義される．ただし，出力はスカラーである．

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i \le W} x_c(h, i)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j \le H} x_c(j, w)$$

これらの式に基づき，水平方向（Width）と垂直方向（Hight）の情報を集約することができた．それらの集約結果を全チャネルおよび全空間次元にわたって統合することで，垂直方向の空間依存関係を保持した特徴マップ $z^h \in \mathbb{R}^{C \times H \times 1}$ を得ることができる．同様に，水平方向の空間依存関係を保持した特徴マップ$z^w \in \mathbb{R}^{C \times 1 \times w}$を得ることができる．

（2） 空間情報の統合と次元圧縮：Concat & Conv2d

得られたz^hとz^wを結合（Concat）し，サイズ$C \times (H + W) \times 1$の統合特徴マップを得ることができる．ここで，$z^w$については，$1 \times W$を$W \times 1$に転置して扱う．さらに，得られた統合特徴マップに対して，1×1の畳み込み層（Conv2d）を用いてチャネル数を圧縮する．結果として，サイズ$(C/r) \times (H + W) \times 1$の特徴マップを得る．ここで，$r$はチャネル数の削減率を表しており，この次元圧縮により，モデルのパラメータ数および計算負荷を抑制している．また，垂直，水平方向の空間相関情報を効率的に凝縮する役割を持つ．

（3） 情報の安定化と非線形変換：BatchNorm & Non-Linear

畳み込み層によって圧縮された特徴マップに対してバッチ正規化（Batch Normalization）および，非線形関数を適用する．バッチ正規化は，各層におけるデータの分布を一定に整えることで，学習の不安定化要因となる内部共変量シフトを抑制する役割を持つ．これにより，ネットワーク全体の学習速度が向上し，安定した収束が可能となる．また，非線形関数を適用することで，ネットワークに高度な表現力を付与する．これにより，垂直方向と水平方向の相互依存関係をモデル化することが可能となる．（2），(3）の処理は以下のような式で表すことができる．

$$f = \delta(\mathrm{BN}(\mathrm{Conv2d}([z^h, z^w])))$$

ここで，$[\cdot, \cdot]$は結合処理（Concat），δは非線形関数を表している．（2），(3）によって特徴マップ$f \in \mathbb{R}^{(C/r) \times (H+W) \times 1}$を得る．

（4）分割とチャネル復元：Split & Conv2d & sigmoid

続いて，得られた特徴マップ$f \in \mathbb{R}^{(C/r)\times(H+W)\times 1}$に対し分割処理（Split）を行う．元の空間サイズに合わせて垂直方向の情報$f^h \in \mathbb{R}^{(C/r)\times H\times 1}$と水平方向の情報$f^w \in \mathbb{R}^{(C/r)\times 1\times W}$が得られる．その後，それぞれの特徴マップに対して，独立した1×1の畳み込み層（Conv2d）を適用し，チャネル数を元のCへと復元する．その後，Sigmoid 関数σを適用することで，アテンションウェイトg^h, g^wを得る．Sigmoid 関数は，各空間位置における重要度を0.0から1.0の範囲で算出する関数である．

$$g^h = \sigma(\mathrm{Conv2d_h}(f^h))\mathbb{R}^{C\times H\times 1}$$
$$g^w = \sigma\big(\mathrm{Conv2d_w}(f^w)\big) \in \mathbb{R}^{C\times 1\times W}$$

（5）出力の算出：Multiplication & Re-weight

最後に，得られた垂直方向のウェイトg^hと水平方向のウェイトg^wを要素ごとに乗算（Multiplication）する．さらに，これを入力特徴マップXに適用する（Re-weight）．

$$\bar{x}_c(h,w) = x_c(h,w) \times g_c^h(h) \times g_c^w(w)$$

出力として，特定の座標情報を強調した再構成特徴マップ$X_{CA} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_c]$（サイズ$C \times H \times W$）を得る．

この一連の処理により，モデルがどのチャネルの，どの座標に重要な情報があるかを明示的に学習することができる．

3.3 Efficient Channel Attention

3.3.1 Efficient Channel Attention の概要

Efficient Channel Attention（ECA）は，従来のチャネルアテンション機構における，チャネル次元の圧縮による情報の欠落を解消し，計算効率を維持しつつ，物体検知における識別性能を向上させるために提案されたモジュールである．ECA は次元圧縮を行わずに，1 次元畳み込み（Conv1d）を用いて各チャネルと，その近傍チャネル間の局所的な相互作用を直接学習する．これにより，パラメータ数を劇的に抑えつつ，高い性能向上が可能となる．

3.3.2 演算プロセス

ECA の演算は以下のステップで構成され，ECA の演算プロセスの概要を図 3.3 に示す．

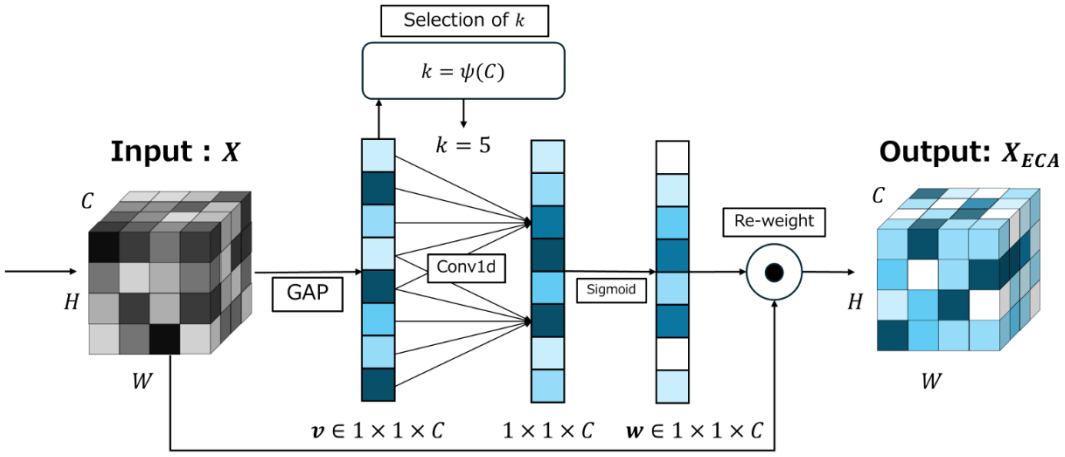

図 3.3 Efficient Channel Attention の演算プロセスの概要図

（1）Global Average Pooling （GAP）

入力特徴マップ$X = [x_1, x_2, \ldots, x_C]$（サイズ$C \times H \times W$）に対して，空間方向（高さ$H$と幅$W$）の Global Average Pooling（GAP）を行う．入力特徴マップXの成分(i, j)におけるチャネルcの値を$x_c(i, j)$とすると，GAP の処理は以下の式で表すことができる．

$$v_c = \frac{1}{WH} \sum_{0 \le i \le W} \sum_{0 \le j \le H} x_c(i, j)$$

入力特徴マップXのチャネル数がCであるから，GAP により，各チャネルのグローバルな情報を凝縮したベクトル $\boldsymbol{v} \in \mathbb{R}^{1 \times 1 \times C}$を得る．

（2）適応的なカーネルサイズkの決定

ECA の最大の特徴は，相互作用の範囲（カーネルサイズk）を特徴マップのチャネル数Cに応じて動的に決定することである（Selection of k）．

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd}$$

ここで，$|t|_{odd}$はtに最も近い奇数を表し，通常$\gamma = 2, b = 1$が用いられる．

（3）1 次元畳み込み（Conv1d）

（1）で得られたベクトル $\boldsymbol{v}$ に対し，（2）で決定したカーネルサイズ k を用いて，1 次元畳み込み（Conv1d）を行う．これにより，各チャネルとその近傍 k 個のチャネルとの間でのみ相互作用を計算し，計算コストを大幅に抑えつつ重要なチャネル情報を抽出する．

（4）Sigmoid 関数によるアテンションウェイトの算出

（3）で得られた出力に対して，Sigmoid 関数を適用する．これにより，各チャネルの重要

度が0から1の範囲のアテンションスコアとして正規化される．(3)，(4）の処理により，チャネルごとのアテンションウェイトを表すベクトル $\boldsymbol{w}$ は以下の式で表せる．

$$\boldsymbol{w} = \sigma\left(\mathrm{Conv1d}_k(\boldsymbol{v})\right) \in \mathbb{R}^{1\times1\times C}$$

ここで， σ は Sigmoid 関数，$\mathrm{Conv1d}_k$はカーネルサイズ k を用いた 1 次元畳み込みである．

（5）重み付け

得られたアテンションウェイト$\boldsymbol{w}$を，もとの入力特徴マップ X に対して，要素ごとに乗算する．⊙はチャネルごとの要素積を表すとすると，以下の式で表せる．

$$X_{ECA} = \boldsymbol{w} \odot X$$

この処理によって，重要な特徴をもつチャネルが強調され，重要度の低いチャネルの情報が抑制された再構成特徴マップ $X_{ECA} \in \mathbb{R}^{C\times H\times W}$が得られる．

第 4 章 提案手法

本章では，まず 1 節で，前章で詳述した CA モジュールと ECA モジュールの共通点，相違点，および両者のトレードオフ関係について整理し，これらの本研究において統合する意義について述べる．また，2 節で提案モジュール ESCFBlock の構造について詳述する．

4.1 CA モジュールと ESC モジュールの比較

4.1.1 共通点と相違点

両モジュールにおける最大の共通点は，コストの増大を最小限に抑えつつ，モデル（YOLOv8）の表現力を高める「軽量かつ効率的なアテンション」を目指している点である．標準的な畳み込み層は，画像内のすべての領域に対して，同一の計算を適用するため，火災に関係ない背景情報も均一に処理してしまう．これに対し，本研究で用いるアテンション機構は，抽出された特徴マップの中から，火災特有の場所や質感を選び出し，それらに高い「重み」を与えることで，重要な情報だけを適応的に強調する仕組み（Re-weight，再重み付け）を有している．

一方で，情報を集約するアプローチには明確な相違が存在する．CA には，垂直と水平の 2 方向のプーリングを用いることで「空間的な位置（どこが重要か）」を正確に符号化することに特化している．これに対し，ECA は Global Average Pooling と 1 次元畳み込み（Conv1d）を用いることで，次元圧縮による情報の損失を避けつつ「チャネル間の相関（どのような質感，色彩か）」を効率的に抽出することに特化している．

4.1.2 トレードオフと相補的関係

これら 2 つのアテンション機構には，情報の「詳細さ」と「広がり」に関係するトレードオフが存在する．

■CA の特性: 空間構造の把握には優れるが，チャネル間の複雑な依存関係を直接的に最適化する能力が ECA には及ばない．

■ECA の特性: 質感の識別能力は高いが，空間情報を完全に圧縮してしまうため，煙や炎の正確な座標を保持することができない．

火災検知においては「炎や煙特有の質感」を捉えると同時に，背景ノイズ（照明や，煙や炎を似た物体）と判別するために「ターゲットが存在する正確な位置」を把握することが不可欠である．したがって，これら 2 つのモジュールは一方が他方の弱点を補う相補的な関係にあるといえる．

4.1.3 統合の意義

本研究で提案する統合モジュールは，これら 2 つのアテンションの利点を融合させることで，既存手法では困難であった「高精度な位置特定」と「高度な質感識別」を，リアルタ

イム性を損なうことなく実現することを目的とする．単一のアテンションではとらえきれない多角的な特徴を同時に考慮することで，複雑な実環境下における火災検知の信頼性を向上させることが，両モジュールを統合することの最大の意義である．

4.2 提案モジュール：ESCFBlock

本節では，本研究が提案する特徴融合モジュール ESCFBlock（Efficient Spatial-Channel Fusion Block）の詳細なアーキテクチャと，実装上の工夫について詳述する．ESCFBlock は，前節で述べた CA と ECA の相補的な特性を最大限に引き出すため，単なる直列的な配列ではなく，並列構造と残差学習の概念を統合した設計となっている．以下の図 4.1 に ESCFBlock の構造図を示す．

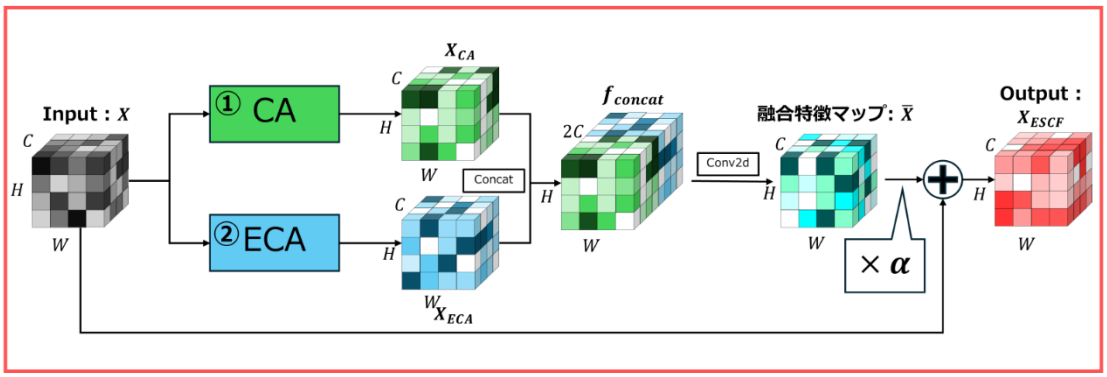

図 4.1 ESCFBlock の構造図

4.2.1 並列処理と1 × 1畳み込みによる特徴融合

ESCFBlock は入力特徴マップ $X \in \mathbb{R}^{C\times H\times W}$ を受け取り，これを CA モジュールと ECA モジュールに同時に入力する並列構造を採用している．

（1）特徴の抽出

CA ブランチ（図 5.1①）は位置情報を重視した特徴マップ X_{CA} を，ECA ブランチ（図 5.1②）はチャネル間の相関を重視した特徴マップ X_{ECA} をそれぞれ独立して抽出する．

（2）チャネル結合

（1）で抽出された 2 つの特徴マップはチャネル方向に結合され，チャネル数 $2C$ の特徴マップ $f_{concat} = [X_{CA}\ ,\ X_{ECA}] \in \mathbb{R}^{2C\times H\times W}$ を得る．

（3）情報の相互作用

結合された特徴マップ f_{fused} に対して，1×1 の畳み込み層（特徴融合層）を適用する．この処理は，単なる次元削減（$2C \to C$）にとどまらず，空間情報とチャネル情報という異なる性質をもつ特徴量間の相互作用を学習させ，1 つの融合特徴マップ $\bar{X} \in \mathbb{R}^{C\times H\times W}$ へと融合させることにある．また，計算の安定性を図るため，この層では，活性化関数を適用せず，情報

の線形融合にとどめている．

4.2.2 残差学習とゲート機構の導入

本モジュールの出力特徴マップX_{ESCF}は，入力特徴マップ Xに対して融合特徴マップ $\bar{X}$を加算する残差接続（Residual Connection）の形式をとる．本研究は，融合特徴マップ $\bar{X}$に対してスケーリング係数 αを乗じる「ゲート付き残差学習」を導入している．

$$X_{ESCF} = X + \alpha \cdot \bar{X}$$

ここで，αは Sigmoid 関数 σを用いて $\alpha = \sigma(gate)$として定義される．学習可能なパラメータ$gate$に対して Sigmoid 関数を適用することで，寄与率 α を 0から 1の範囲で制限している．この設計の意図は，モデル自身が「どの程度アテンションの情報を付与すべきか」を適応的に学習可能にするところにある．一般に，$\alpha = 0$から学習を開始すると，誤差逆伝播の過程で，融合ブランチ側への勾配が消失しやすく，学習の停滞を招く恐れがある．一方で，$\alpha = 1$のような大きな値から学習を開始させると，未学習のモジュール（ESCFBlock）が既存のネットワーク（YOLOv8）の出力を乱し，学習が不安定化する．今回，初期値を$\alpha = 0.1$とすることで，学習の初期段階における勾配の確保と，YOLov8 への低干渉性を両立させている．

4.2.3 事前学習済み重みの保護を目的とした初期化戦略

本研究では，YOLOv8 の事前学習済み重み（Pre-Trained Weights）がもつ優れた汎用特徴抽出能力を維持しつつ，提案モジュールを段階的に適応させるための初期化戦略を採用している．

具体的には，特徴融合層（4.2.1 節（3）参照）の重みとバイアスをすべて 0で初期化するゼロ初期化を適用している．この初期化状態において，融合特徴マップ $\bar{X}$ は常に 0 となり，提案モジュールの初期出力は $Output_{init} = X + \alpha \cdot 0 = X$，すなわち，入力特徴マップとなる．この工夫により，学習直後の時点では提案モジュールはネットワーク全体に悪影響を及ぼさず，学習が進むにつれて αおよび特徴融合層の重みが更新され，徐々に火災検知に最適な特徴強調が追加される仕組みとなっている．これは，転移学習を成功させる上でのきわめて高度な実装技術である．

第 5 章 実験

5.1 データセット

提案手法の評価を行うため，火災および煙の検知を目的として構築された公開データセットである D-fire データセット[6]，[20]を用いた．本データセットは 21,527 枚の画像で構成されており，そのアノテーションには火災（Fire）と煙（Smoke）の 2 クラスが含まれている．

5.1.1 データセットの内訳と特徴

D-fire データセットの各カテゴリにおける画像枚数の内訳を表 5.1 に，バウンディングボックス数の内訳を表 5.2 に示す．また，サンプル画像を図 5.1 に示す．

表 5.1 D-fire の画像枚数

Category	Images
Only Fire	1,164
Only Smoke	5,867
Fire and Smoke	4,658
None	9,838
合計	21,527

表 5.2 バウンディングボックスの数

Class	Bounding boxes
Fire	14,692
Smoke	11,865

図 5.1 D-fire データセットのサンプル画像

赤色の枠は Fire，青色の枠は Smoke のバウンディングボックスを示している．

本データセットは，インターネットから収集された画像のほか，監視カメラによる実画像，消防シミュレーションの記録，および遠距離の火災を模した合成画像など，多種多様なシーンが含まれているのが特徴である．これにより，色，形状，密度がことなる煙や，森林，市街地といった多様な背景，さらには雲や霧，太陽の反射光といった誤検知を招きやすいノイズ要素に頑健性の検証が可能となっている．また，本データセットには総画像の約 45%にあたる 9,838 枚のネガティブサンプル（炎，煙が移っていない画像）が含まれる．このネガティブサンプルにより，学習プロセスにおいてモデルに「火災でないもの」を正しく識別させる能力を養わせることが期待できる．

5.1.2 実験におけるデータの分割と前処理

実験においては，データセットの構成に従い，全体の 80%（17,221 枚）を訓練用，20%（4,306 枚）をテスト用とした．

■訓練データ（Train）:17,221 枚．このうち，7,883 枚は物体を含まない背景画像であり，モデルの誤検知抑制のために活用される．データの読み込みの際，26 件の破損 JPEG ファイルが検出されたが，これらは YOLOv8 に標準装備されている，自動修復プロセスを得て学習に組み込まれた．

■テストデータ（Test/Val）:4,306 枚．このうち 2,005 枚は背景画像である．精度の算出にあたり，テストデータのログを確認したところ，4 枚の画像においてアノテーションの座標値が正規化の範囲外（1.0 を超える値など）にある不正データが確認された．本研究においては，Test と Val は兼用している．

■最終有効データ数:学習の公平性を期すため，前述の不正データ 4 枚を除外した 4,302 枚を，本研究の最終的な精度評価を対象とした．

実験に使用したデータセットの詳細および有効データ数を表 5.3 に示す．

表 5.3 実験に使用したデータセットの詳細および有効データ数

項目	訓練データ (Train)	テストデータ (Test/val)	備考
初期画像枚数	17,221 枚	4,306 枚	合計 21,527 枚
None(背景画像)	7,833 枚	2,005 枚	誤検知抑制のために使用
データ不備の検出	26 件 (JPEG 破損)	4 件 (座標不正)	–
処理内容	自動修復により学習に使用	評価対象から除外	–
最終有効データ	17,221 枚	4,302 枚	–

5.2 実験設定

本節では，提案手法である ESCFBlock の YOLOv8 内の挿入位置に関する検証実験の詳細な条件について詳述する．

5.2.1 ネットワーク構成と挿入位置

ベースモデルには， YOLO シリーズの YOLOv8 を採用した．また，推論速度と，検出精度のバランスの取れた，中程度のモデルサイズである YOLOv8m を用いた．このモデルを基準（Baseline）とし，ネットワーク内の以下 7 か所に ESCFBlock をそれぞれ 1 ユニットずつ挿入した個別のモデルを作成し，比較検証を行った．以下の図 5.2 に挿入位置を示す．

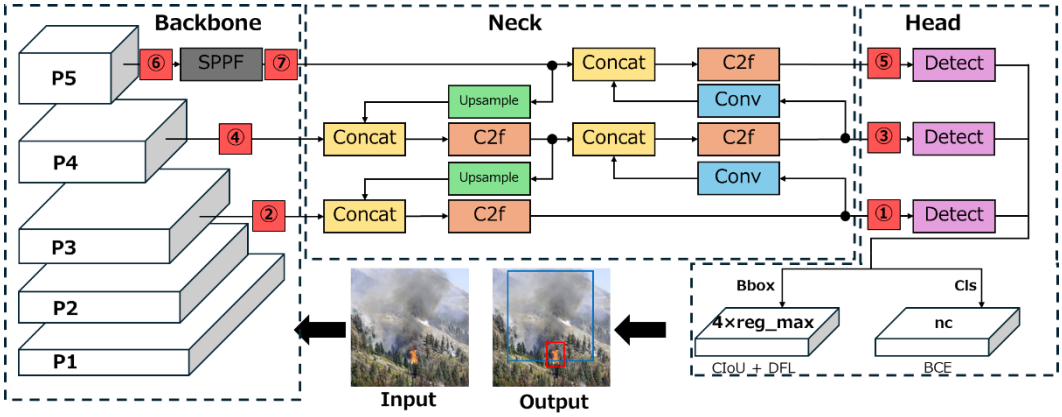

図 5.2 ESCFBlock の挿入位置（図 3.1 を一部改訂）

挿入位置は，①Head 部 P3 層，②Backbone 部 P3 層，③Head 部 P4 層，④Backbone 部 P4 層，⑤Head 部 P5 層，⑥SFFP モジュール 直前，⑦SFFP モジュール直後の 7 か所である．

5.2.2 ハイパーパラメータと学習条件

本研究におけるモデルの比較検証を公平に行うため，Baseline および ESCFBlock を挿入した全ての提案モデルにおいて，共通のハイパーパラメータを用いて学習を実施した．詳細な設定を以下の表 5.4 に示す．

表 5.4 ハイパーパラメータ設定

パラメータ	設定値
Model Size	640 × 640
Batch Size	16
Initial Learning Rate	0.01 (lr0)
Momentum / Decay	0.937 / 0.0005
Epochs	50
Patience	15
Close_mosaic	10

5.2.3 実験環境

本研究における火災検知モデルの学習および評価は，表 5.5 に示す計算資源を用いて実施した．

表 5.5 実験環境

項目	仕様
CPU	Intel(R) Core(TM) i7-7700K CPU @ 4.20GHz
GPU	NVIDIA GeForce GTX 1080 Ti (11GB VRAM)
RAM	32GB
OS	Linux (Ubuntu 24.04 系 / Kernel 6.8.0)
python	3.9.25
	Ultralytics YOLO 8.3.240
Library	PyTorch 2.7.1+cu118
	CUDA 11.8

5.2.4 評価指標

本研究では，ESCFBlock の有効性を定量的に評価するため，物体検知タスクで一般的に用いられる以下の 4 つの指標を採用した．適合率（Precision），再現率（Recall），F1 Score，および平均適合率（mAP: mean Average Precision）の 4 つの評価指標を使用する．

適合率（Precision）は，モデルがポジティブ（炎・煙）と予測したサンプルのうち，実際にポジティブであった割合を示し，誤検知の少なさを評価する指標である．一方，再現率（Recall）は，実際にポジティブである全サンプルのうち，正しく検出されたサンプルの割合を示し，検出の網羅性を評価する指標である．

$$Precison = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

記の式において，TP（True Positive）は真陽性，FP （False Positive）は偽陽性，FN（False Negative）は偽陰性をそれぞれ表す． また，適合率と再現率はトレードオフの関係にあるため，両者の調和平均である F1 Score を用いて総合的な性能バランスを評価する．

$$F1 = 2 \times \frac{Precision \times Recall}{Precison + Recall}$$

平均適合率（AP: Average Precision）は，適合率-再現率曲線（PR 曲線）の下側の面積として定義され，検出の閾値を変化させた際の平均的な適合率を表す指標である．AP が高いほど，あらゆる再現率のレベルにおいて安定して高い適合率を維持できていることを意味する．さらに， mAP（mean Average Precision）は，対象とする全クラス （本研究では炎と煙）について算出した AP の平均値であり，モデル全体の総合的な実力を示す指標となる．本研究では，IoU（Intersection over Union）の閾値が 0.5 のときの mAP@50 と，0.5 から 0.95 まで 0.05 刻みで変化させた平均値である mAP@50-95 を採用した．

また，火災検知においては，火災の発見漏れを防ぐことが重要となる．そのため本研究では，単なる精度の高さだけでなく，見逃しの少なさを示す再現率（Recall）の向上を重要な

評価基準として重視している.

5.3 ESCFBlock の挿入位置に関する比較実験と分析

5.3.1 挿入位置の比較評価

提案手法である ESCFBlock の最適な挿入位置を決定するため，baseline（YOLOv8m）のネットワーク構造に対し，図 5.2 に示す，①Head 部 P3 層（Head-P3），②Backbone 部 P3 層（BackboneP3），③Head 部 P4 層（Head-P4），④Backbone 部 P4 層（Backbone-P4），⑤Head 部 P5 層（Head-P5），⑥SFFP モジュール 直前（pre-SFFP），⑦SFFP モジュール直後（post-SFFP）の計 7 か所に ESCFBlock を挿入し，精度比較を行った.

なお，各モデルの評価には，学習プロセスにおいて検証データ（Validation data）に対し，最も高い性能を示した重みファイル（best.pt）を採用した． 本実験では，この総合的な選定基準に従い，各挿入位置の有効性を評価した．表 5.6 に比較実験の結果を示す.

表 5.6 ESCFBlock の各挿入位置における精度比較（太字は最高値を示す）

挿入位置	Params	Precision (P)↑	Recall (R)↑	F1 Score↑	mAP@50↑	mAP@50-95↑	α
①Head-P3	25,938,532	**0.795**	0.724	0.758	0.787	0.468	0.059
②Backbone-P3	25,938,532	0.788	0.712	0.744	0.788	0.462	0.163
③Head-P4	26,180,860	0.784	0.729	0.756	**0.794**	0.467	0.079
④Backbone-P4	26,180,860	0.774	0.725	0.749	0.784	0.462	0.102
⑤Head-P5	26,584,468	0.778	**0.743**	**0.760**	0.792	**0.472**	0.130
⑥pre-SFFP	26,584,468	0.784	0.718	0.750	0.782	0.462	0.115
⑦post-SFFP	26,584,468	0.779	0.708	0.742	0.783	0.461	0.061
Baseline	25,857,478	0.785	0.731	0.757	0.790	0.471	-

表 5.6 に示す実験結果より，ESCFBlock を⑤Head 部 P5 層（head-P5）に挿入したモデルが，mAP@50~95（0.472）およびF1 Score（0.760）において，baseline を上回る最高値を記録した．以下に，各指標の観点から Head-P5 構成の有効性を分析する.

■検出の網羅性（Recall）の向上

Head-P5 構成においてRecallは baseline の0.731から0.743へと有意な向上（+0.012）を示した. Recallは「存在する正解データをどれだけ漏れなく検出できたか」を表す指標であり，火災検知というタスクの性質上，最も重視すべき項目である．初期消火や避難誘導の遅れを防ぐためには，炎や煙の見逃し（偽陰性）を極限まで減らす必要がある． このRecallの向上は，ESCFBlock が持つ空間・チャネルアテンション機構が，火災特有の微細な特徴量を効果的に強調し，網羅的な検出を強化できたことを実証している.

■総合的な性能バランス（F1 Score）

PrecisionとRecallの調和平均であるF1 Scoreにおいても，baseline の0.757から0.760への改善が確認された．一般に，Recallを向上させようとすると誤検知（Precision の低下）が増加するトレードオフの関係にあるが，Head-P5 ではPrecision の低下を最小限（0.785 → 0.778）

に抑えつつ，総合性能であるF1 Scoreを向上させることに成功している．これは実運用において，見逃しを減らしつつ，過度な誤報も抑制できるバランスの取れた性能改善が達成されたことを意味する．

■挿入位置に関する考察

以上の結果より，ESCFBlock は Backbone のような特徴抽出の初期段階（低次特徴層）に適応するよりも，Head 部のような高次元特徴量（セマンティックな情報を持つ層）に適用する方が，検出性能の向上に大きく寄与することが明らかとなった．これは，ネットワークの深層において物体としての「意味」が形成される段階で，アテンションによる特徴強調を行うことが，火災検知において最も効果的であることを示唆している

5.3.2 全挿入位置における学習挙動の比較分析

前節の best.pt による最終的な数値的評価に加え，本節では学習過程全体における各指標の推移（学習曲線）から，ESCFBlock の挿入位置が，モデルの収束挙動や安定性に与える影響を明らかにする．図 5.3 に全挿入位置における精度指標（mAP@50，mAP@50~95），実用指標（F1 Score，Recall），および損失関数（Box Loss，Class Loss）の推移を示す．

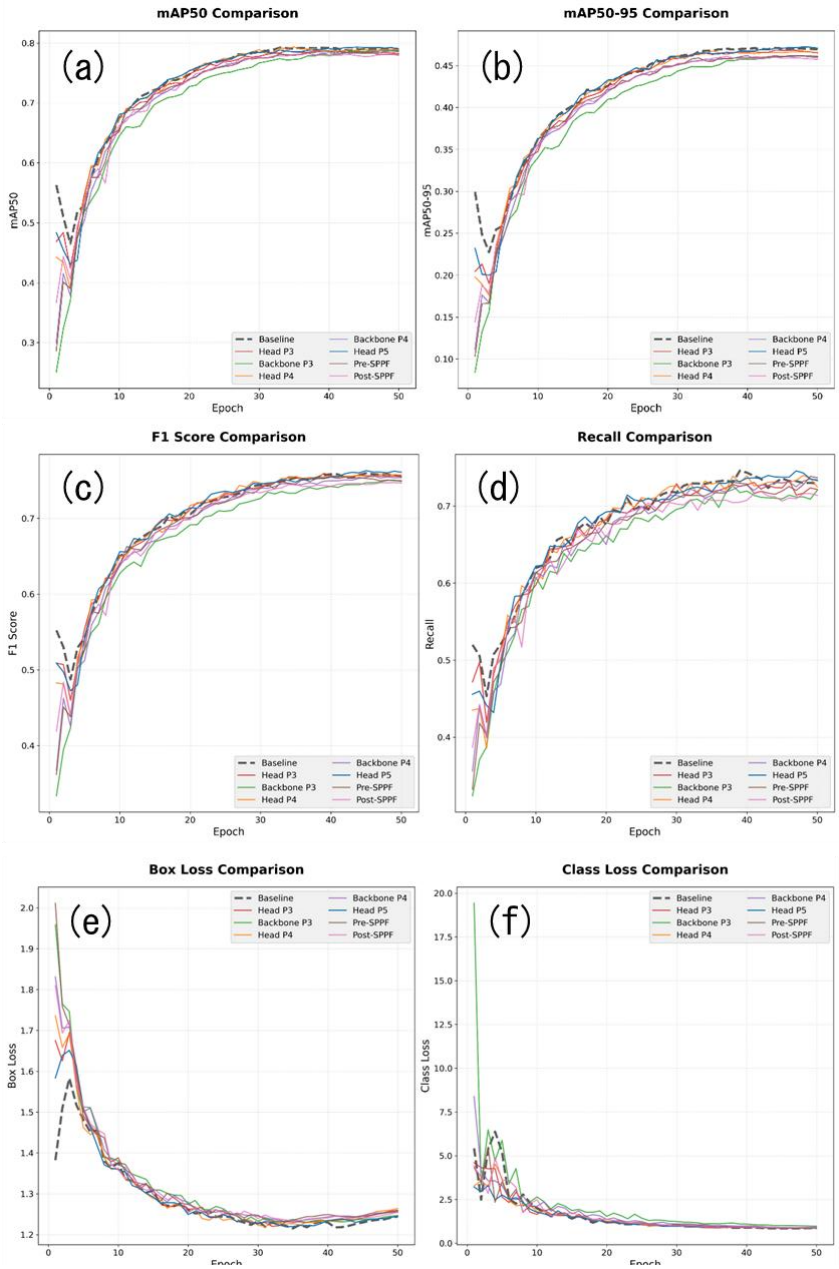

図 5.3 ESCFBlock の全挿入位置における学習曲線の比較

（a）mAP50の推移　（b）mAP50～90の推移　（c）F1 Scoreの推移　（d）Recallの推移

（e）Box Loss（位置損失）の推定　（f）Class Loss（分類損失）の推移

（1）損失関数の推移に基づく学習安定性の評価

学習の安定性を評価するため，まずは損失関数（Loss）の推移（図 5.3（e）（f））に着目する．損失関数の低下は，モデルが学習データに対して適切に適合していることを示す．

■Box Loss（位置損失）:予測したバウンディングボックスと正解データとの座標誤差を示す．

■Class Loss（分類損失）:検出物体を「炎・煙」あるいは「背景」として正しく識別できて

いるかを示す.

図 5.3（f）のClass Loss推移をみると，⑤Head-P5 に ESCFBlock を挿入したモデルは，学習の全域において baseline と同様，あるいはそれを下回る低い損失値を維持しており，分類タスクにおける学習が極めて安定していることが確認できる．一方，Backbone 部や SPPF 前後に挿入したモデルは，収束が遅く，損失値が高くなる傾向がみられた.

（2）精度指標の推移と特徴抽出への干渉 （図 5.3（a）（b））

精度指標であるmAPの推移を確認すると，挿入位置による挙動の違いが顕著に表れている．③Head-P4 および⑤Head-P5 のモデルは，学習初期からmAP@50が baseline と同等以上の高い水準で推移している．これに対し，Backbone 部（②P3，④P4）や SPPF 前後（⑥，⑦） に挿入したモデルは，学習期間を通じて精度が baseline を下回り，グラフの振動（不安定さ）も激しい．この原因として，転移学習における「事前学習済み特徴量への干渉」が考えられる．YOLOv8 の Backbone 部は，ImageNet 等による事前学習で得た汎用的な特徴抽出能力（エッジやテクスチャの検出能力）を有している．この初期段階の層に，未学習の複雑なアテンション機構（ESCFBlock）を挿入したことで，確立されていた特徴抽出プロセスにノイズが生じ，学習の不安定化を招いたと推察される.

（3）網羅性と実用性の観点からの評価（図 5.3（c）（d））

火災検知において最も重要視されるRecall（再現率）およびF1 Scoreの推移において，⑤Head-P5 の優位性は決定的である．他の提案モデル（Backbone 挿入モデル等）が学習の中盤で baseline を下回る不安定な挙動を示す中，Head-P5 モデルのみが，学習のほぼ全期間において baseline を上回る性能を維持し続けた．これは，ESCFBlock をネットワークの最深部であり，物体検出の最終判断を行う Detect モジュール直前（P5 層）に配置したことが奏功したと言える．P5 層では，画像全体の文脈を含むセマンティック（意味的）な情報が扱われるため，ここでアテンションを適用することで，火災特有の不定形な特徴を，背景ノイズに惑わされることなく頑健に強調・保持できたことが，高いRecallの維持に繋がったと結論付けられる.

5.3.3 最適モデル（Head-P5）の詳細分析

前節までの比較実験において，全挿入位置の中で最も優れた性能を示した ⑤Head-P5 構成について，baseline との詳細な比較・分析を行う．図 5.4 に，各指標（mAP, Recall, F1 Score）および損失関数（Loss）の推移に最大値をプロットした比較グラフを示す.

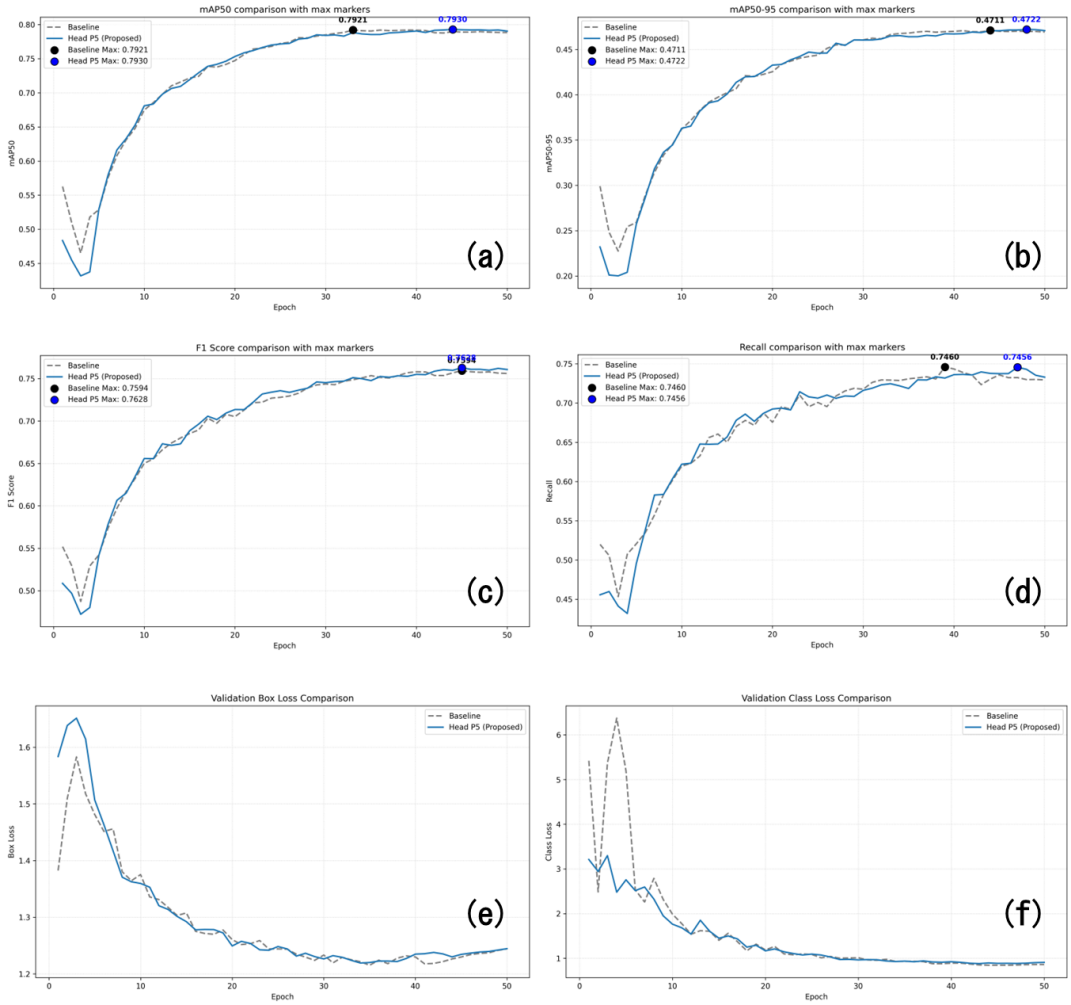

図 5.4 baseline と Head-P5 の学習曲線の比較詳細

（a）mAP50の推移　（b）mAP50～90の推移　（c）F1 Scoreの推移　（d）Recallの推移

（e）Box Loss（位置損失）の推定　（f）Class Loss（分類損失）の推移

（a）（b）（c）（d）においてマーカーは学習中の最大値を示す

（1）位置特定精度の向上（mAPの分析）

検出精度を示すmAPの推移に着目する．標準的な評価指標であるmAP@50（図 5.4（a））において，Head-P5 の最大値は baseline の0.7921に対し，0.7930 を記録し，わずかながら向上を示した．さらに，バウンディングボックスの重なり具合を厳密に評価するmAP@50~95（図 5.4 （b））においては，baseline の0.4711から 0.4722 へと確実な向上が確認された．これは，ESCFBlock の導入によって，単に物体を見つけるだけでなく，火災の発生範囲をより精密に特定する能力が強化されていることを示唆している．

（2）網羅性とバランスの評価

実用上の重要課題である「見逃しの抑制」について分析する．Recall（再現率）の推移（図

5.4（d)）を見ると，最大値に関しては baseline（0.7460）に対し Head-P5（0.7456）とわずかに及ばない結果となった．しかし，学習曲線の全体的な挙動を見ると，Head-P5 は baseline と同等の高い水準を安定して維持している．ここで，適合率（Precision）と再現率（Recall）の調和平均であるF1 Score（図 5.4（c)）に着目すると，Head-P5 は baseline の最大値0.7594を上回る0.7628を記録している．これは，Recall単体のピーク値では劣るものの，誤検知（Precisionの低下）を効果的に抑制できているため，「検出の正確さ」と「網羅性」のバランス（トレードオフ）において，Head-P5 が最も優れた最適解に到達していることを意味する．

（3）損失関数の収束と誤検知の抑制

この性能向上の要因を損失関数（Loss）の観点から考察する．図 5.4（f）のClass Loss（分類損失）において，Head-P5（青線）は baseline（点線）よりも低い値を推移して収束している．これは，アテンション機構が背景ノイズ（雲や照明など）と実際の火災特徴を識別する能力を高め，誤分類を低減させた結果であると言える．以上の詳細分析により，Head-P5 構成は，位置特定の精密化（mAP@50~95向上）と，実用的な運用に不可欠な性能バランス（F1 Score向上）の両立を実現しており，本研究における最適モデルであると結論付けられる．

5.3.4 学習によるアテンション寄与率 αの変化

本節では，ESCFBlock 内のゲート付き残差接続において，アテンションの寄与率を調整するパラメータ α（詳細は 4.2.2 節参照）の挙動を分析する．本実験では，事前学習済みモデル（YOLOv8）への急激な干渉を避けるため，全挿入位置において αの初期値を0.1に設定し，学習プロセスを通じて，モデル自身に最適な寄与率を探索させる構成を採用した．表 5.6 の右列に表すαは best.pt における寄与率であり，この値をもとに分析を行う．

（1）挿入位置による学習挙動の差異

実験の結果，②Backbone-P3 において αが最大値（0.163）を記録した．一方，精度評価で最高成績を収めた⑤Head-P5 における値は0.130であった．この「寄与率の高さ」と「最終的な検出精度」の不一致は，ネットワークの階層によってアテンション機構が果たす役割が異なることを示唆している．

（2）Backbone-P3 におけるα増大の要因：低次特徴の特徴

Backbone の初期層に近い P3 レイヤーにおいて αが最大化した事実は，学習初期の段階で「低次元・中次元の特徴強調」が損失低下に有効であったことを意味する．炎や煙といった対象は，剛体（車や人など）とは異なり明確な輪郭を持たず，色のグラデーションやテクス

チャの揺らぎといった「質感」が識別の重要な手がかりとなる．モデルは誤差逆伝播の過程で，Backbone-P3 においてこれらの局所的な視覚情報を強力に強調することが効率的であると判断し，その結果として α が上昇したと考えられる．

（3）寄与率の増大が精度向上に直結しない理由

Backbone-P3 でアテンションが強く機能したにもかかわらず，最終的な精度（mAP）が Head-P5 に及ばなかった理由として，以下の 2 点が挙げられる．

■ノイズの増幅と過適合: 初期層で視覚的特徴（赤色や高輝度な領域）を強調しすぎると，火災に類似した背景ノイズ（夕日，ネオンサイン，紅葉など）までもが同時に強調されてしまう．これが誤検知を誘発し，Precisionの低下を招いた要因である．

■意味情報の欠如: 初期層での処理はあくまで「局所的なパターンの強調」に過ぎない．画像全体の文脈（コンテキスト）を踏まえた「これが火災である」という最終的な意味的（セマンティック）な判断は，より高次な Head 部で行われる．したがって，初期層での寄与率の高さは，必ずしも最終的な識別能力の高さとは正比例しない．

（4）結論: Head-P5 における「判断の洗練」

対照的に，提案構成である Head-P5 においては，αは適度な上昇（0.130）にとどまりつつも，全条件の中で最高精度を記録した．これは，ESCFBlock が物体検出の最終判断を下す直前の階層に配置されたことで，Backbone 部で抽出された特徴量を統合し，「火災らしさ」の判断を洗練させる役割を果たした結果であると言える．以上の分析より，火災検知におけるアテンション機構の適用においては，単に特徴を強く強調する（Backbone）ことよりも，高次の意味情報に基づいて特徴を選別・統合する（Head）アプローチが，誤検知を抑制しつつ検出能力を最大化する上で最も効果的であると結論付ける．

5.4 追加実験による信頼性の評価

5.4.1 乱数シード固定による統計的な性能評価

前節までの実験で選定した Head-P5 構造の性能が，特定の初期条件に依存した偶発的なものではないことを検証するため，乱数シード（seed）を変更した 5 回の独立試行による追加実験を行った．表 5.7 に，各試行の結果および 5 回の平均値を示す．

表 5.7 乱数シード固定による試行結果の平均

Model	Precision (P)↑	Recall (R)↑	F1 Score↑	mAP@50↑	mAP@50-95↑	α
Head-P5	0.784	**0.735**	**0.759**	0.790	0.471	0.131
Baseline	**0.788**	0.729	0.757	**0.793**	**0.474**	-

総合的な性能バランスを示す F1 Score においては，HeadP5 は平均 0.759 を記録し，

Baseline と同等の高い水準を維持したまま, Recall の向上(0.729→0.735)に成功している. 以上の統計的評価により，Head-P5 構成は，見逃しの少ない確実な火災検出という本研究の目的に対し，単発の実験結果だけでなく，統計的にも信頼性の高い手法であることが証明された.

5.4.2 100 エポック学習における性能推移の比較

前節の 50 エポックにおける検証に加え，学習期間の延長がモデルの収束性と性能に与える影響を確認するため，100 エポックの長期学習実験を行った．本実験でも，4.4.1 節と同様に seed を固定して実験を行った．また，学習におけるパラメータは表 5.3 に示したものであり，epoch 数のみ 100 に変更して行った．以下の表 5.8 に baseline と最良構成である Head-P5 の性能を比較した．また，図 5.5 に各指標および Loss の推移を表したものを示す.

ここで，本研究ではモデルの過学習の防止として，15 エポックの猶予を設定した Early Stopping を採用している（表 5.3 参照）．本試行で baseline に関しては，80 エポックで Fitness 関数の更新が停止し，その後の 15 エポック後の 95 エポック時点で学習が早期終了した．これに対し，Head-P5 は設定された 100 エポックを完遂しており，Head-P5 が baseline と比較して，長期的な学習においても性能を向上させ続ける持続性を有していることが確認された.

表 5.8 baseline vs Head-P5 （epochs=100）

Model	Precision(P) ↑	Recall(R) ↑	F1 ↑	mAP@50 ↑	mAP@50~95 ↑	α
baseline	**0.803**	0.737	0.769	**0.799**	**0.481**	–
Head-P5	0.797	**0.745**	**0.770**	0.795	0.478	0.148

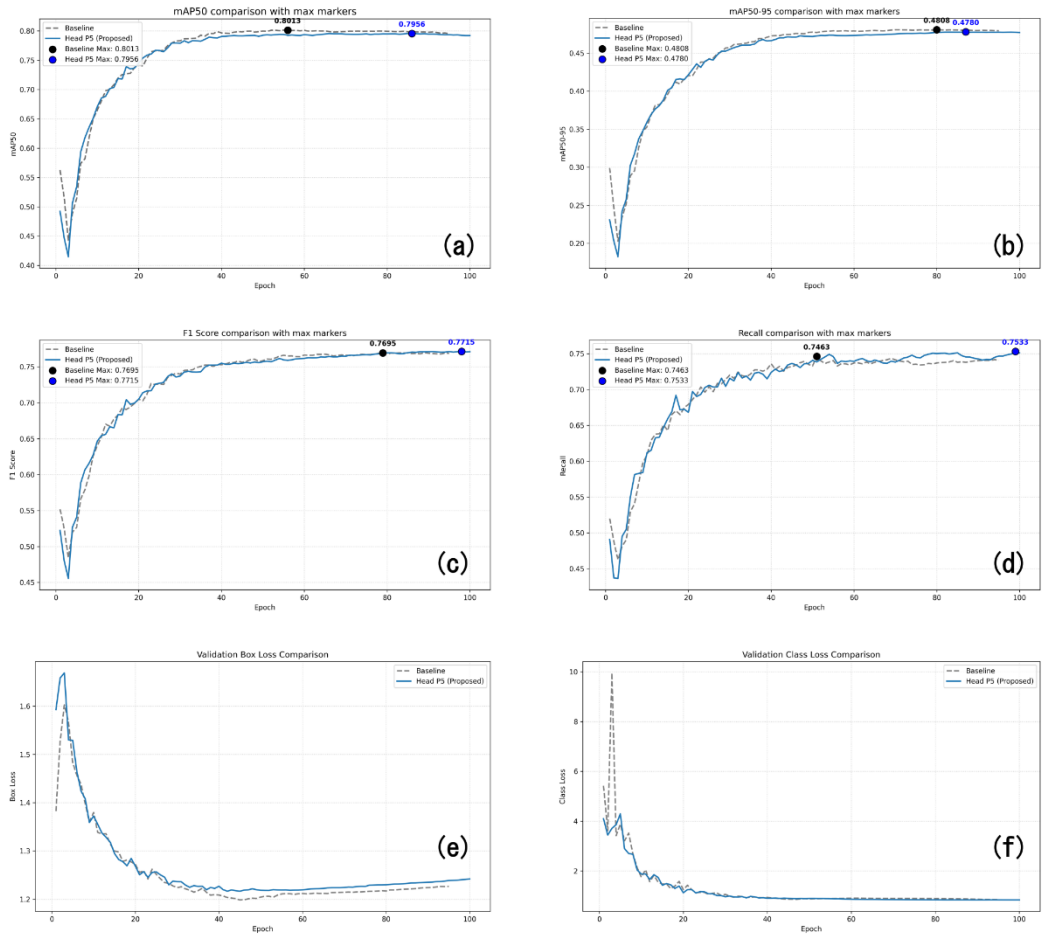

図 5.5 baseline と Head-P5 の学習曲線の比較詳細(epochs=100)

（a）mAP50の推移　（b）mAP50～90の推移　（c）F1 Scoreの推移　（d）Recallの推移

（e）Box Loss（位置損失）の推定　（f）Class Loss（分類損失）の推移

（a）（b）（c）（d）においてマーカーは学習中の最大値を示す

（1）精度指標の推移と比較

学習の全過程における各指標の推移を比較したところ，両手法ともに 40 エポック以降は概ね収束傾向であることが確認された．

■F1 ScoreおよびRecallの向上: 表 5.8 よりF1 Scoreに関して，提案手法（Head-P5）が baseline をわずかに上回った結果になった．また，Recallに関しては，Head-P5 が baseline と比較して大きく向上している．

■PrecisionとmAPの傾向: 一方で, Precision（baseline:0.803→Head-P5:0.797）およびmAP@50（baseline:0.799→Head-P5:0.795）については，baseline がわずかに上回る結果となった．これは，提案手法がより広範な検出（Recallの向上）を達成した一方で，誤検知を抑制する制度においてわずかに課題が残ることを示唆している．

（2）損失関数の推移

検証データに対する損失の推移を分析することで，学習の安定性と汎化性能を評価した．

■Class Loss: 両手法ともに，学習初期のスパイクを除き，滑らかに減少，収束している．Head-P5 は baseline と同等，あるいは初期段階において baseline よりも低い損失値で推移しており，分類タスクにおいて安定した学習が行われていることがわかる．

■Box Loss: Box lossについては 40 エポック以降，両手法ともに微増傾向がみられるが，Head-P5 は baseline と比較してわずかに高い値で推移している．これが検出精度（mAP）のわずかな低下に影響していると考えられる．

（3）小括

以上の比較実験により，提案手法（Head-P5）は baseline と比較して，Recall（再現率）を重視した検出性能の向上に寄与することが明らかとなった．F1 Scoreにおいて baseline を上回る結果を得られたことは，物体検出モデルとしての総合的なバランスが改善されたことを示している．一方で，Precisionおよび厳格なIoU閾値下でのmAPには改善の余地があり，回帰精度のさらなる最適化が今後の課題である．

5.5 アブレーションスタディ

提案手法 ESCFBlock の構成要素である各アテンション機構の寄与を明らかにするため，YOLOv8m を Baseline とし，CA のみ（YOLOv8m＋CA），ECA のみ（YOLOv8m＋ECA），および提案手法（YOLOv8m＋ESCFBlock）を搭載したモデルについて比較検証を行った．なお，すべてのモデルにおいてアテンション機構の挿入位置は Head-P5 とし，ゲート付き残差接続を適用した．また，αの初期値 は一律 0.1 ，他のハイパーパラメータは，表 5.4 の通りであ る．結果を表 5.9 に示す．

表 5.9 アブレーションスタディ結果（太字は最高値を示す）

Method	Params	Precision (P)↑	Recall (R)↑	F1 Score↑	mAP@50↑	mAP@50-95↑
YOLOv8m + ECA	26,190,412	**0.790**	0.729	0.758	0.791	0.470
YOLOv8m + CA	26,252,687	0.788	0.727	0.756	0.792	0.470
YOLOv8m + **ESCFBlock**	26,584,468	0.787	**0.740**	**0.763**	**0.793**	**0.472**
YOLOv8m(Baseline)	25,857,478	0.787	0.734	0.760	**0.793**	**0.472**

実験結果より，ESCFBlock を搭載したモデルが，Recall （0.740），F1 Score（0.763），および mAP@50-95（0.472）において，単体のアテンション機構を用いたモデル（CA のみ，ECA のみ）を上回る性能を示した．特に，Recall にお いては Baseline と比較して CA 単体では 0.727，ECA 単体では 0.729 へと低下したことに対し，ESCFBlock では 0.740 と有意な向上が確認された．この結果は，空間情報に特化した CA と，チャネル情報に特化し

た ECA が，単体では捉えきれない火災の特徴を，並列統合によって相互補完的に捉えていることを裏付けている．すなわち，提案手法における「空間・チャネル情報の融合」と「ゲート付き残差接続による適応的な強調」が，火災検知の性能向上に有効な要素であることが実証された．

5.6 Boreal Forest Fire データセットでの検証

最後に，Boreal Forest Fire データセット[21]での検証を行った．本データセットは，フィンランドの FireMan プロジェクトの一環として構築された，北方林（Boreal Forest）における森林火災を対象としたデータセットである．データは主に UAV（無人航空機）を用いて上空から撮影された映像で構成されており，火災および煙の領域に対して，物体検出用のバウンディングボックスならびにセグメンテーションマスクによる詳細なアノテーションが付与されている．本データセットには，これまで学習に使用したデータセットとは異なる植生や，上空からの視点が含まれる．そのため，本実験では，提案モデルが未知の環境下においても火災を誤検知することなく正確に認識できるか，その汎用性とロバスト性を評価することを目的として使用した．なお，本データセットのターゲットは Fire と Smoke ではなく，Smoke のみである．図 5.6 にサンプル画像を示す．

図 5.6 Boreal Forest Fire データセットのサンプル画像，[21]より引用

D-fire データセットを用いて学習を行った 2 つのモデル YOLOv8m（Baseline）と，提案モジュール ESCFBlock を Head 部 P5 層に挿入したモデル（Ours）について比較検証を行う．なお, 学習エポック数は 100 である. 5.4.2 節の実験で生成された最良の重みファイル（best.pt）を用いた．表 5.10 に結果を示す.

表 5.10 Boreal Forest Fire データセットでの検証結果

Model	Precision (P)↑	Recall (R)↑	F1 Score↑	AP@50↑	AP@50-95↑
Baseline	0.935	0.906	0.920	0.952	**0.693**
Ours	**0.955**	**0.928**	**0.941**	**0.959**	0.683

表 5.10 に実験結果を示す． 提案モデル（Ours）は，ベースラインモデル（Baseline）と比較して，Precision，Recall，F1 Score，および AP@50 の主要な指標において性能の向上が確認された．特に Recall は 0.906 から 0.928 へと向上しており，提案手法は火災の見逃しが少ないモデルであるといえる． AP@50-95 においてはわずかな低下（0.010）が見られたものの，AP@50 では 0.959 という高い精度を記録している.

以上の結果から，提案手法は学習に使用していない未知の環境（Boreal Forest Fire データセット）においても，Baseline より高い検出能力を発揮し，優れた汎用性を有していることが示された.

第 6 章 まとめ

本研究では，YOLOv8 をベースとした高精度な火災検知 モデルの構築を目的とし，空間情報とチャネル情報を適応的に統合する新たなアテンション機構 ESCFBlock（Efficient Spatial-Channel Fusion Block）を提案した. 本手法 は, Coordinate Attention と Efficient Channel Attention を並列配置し，ゲート付き残差接続を導入することで，計算コストの増大を最小限に抑えつつ，火災特有の微細な特徴を強調するアーキテクチャである． D-fire データセットを用いた評価実験の結果，ESCFBlock を Head 部の P5 層（Head-P5）に導入したモデルが最も高い性能を示した． 特に，火災検知において重要視される Recall において，Baseline と比較して有意な向上が確認された．これは，提案手法が複雑な背景ノイズの中から 炎や煙の兆候を網羅的に捉え，実運用における「見逃しリスク」を低減できることを示唆している．また，アブレーションスタディにより，空間情報とチャネル情報を相互補完的に統合することの有効性が実証された.

今後の展望として，より多様な環境下（悪天候や夜間など）におけるロバスト性の検証や，バウンディングボックスの回帰精度のさらなる向上が挙げられる．また，発生初期の極小な火種や煙に対しても確実に検知できる感度の追及が不可欠である.

謝辞

本研究及び論文の作成にあたり，研究の着想や論文執筆等，多くのご指導，ご助言を頂きました，静岡大学工学部の岡部誠准教授に心から感謝申し上げます．また，ご助力頂いた修士課程学生および学部生の皆様に深く感謝いたします．

参考文献

[1] Saydirasulovich, Saydirasulov Norkobil and Mukhiddinov, Mukhriddin and Djuraev, Oybek and Abdusalomov, Akmalbek and Cho, Young-Im: An improved wildfire smoke detection based on YOLOv8 and UAV images, Sensors, Vol. 23, No. 20, p. 8374 (2023).

[2] Pesonen, Julius and Hakala, Teemu and Karjalainen, V¨ain¨o and Koivum¨aki, Niko and Markelin, Lauri and Raita-Hakola, Anna-Maria and Suomalainen, Juha and P¨ol¨onen, Ilkka and Honkavaara, Eija: Detecting Wildfires on UAVs with Real-time Segmentation Trained by Larger Teacher Models, 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE, pp. 5166–5176 (2025).

[3] Pesonen, Julius and Raita-Hakola, Anna-Maria and Joutsalainen, Jukka and Hakala, Teemu and Akhtar, Waleed and Koivum¨aki, Niko and Markelin, Lauri and Suomalainen, Juha and Alves de Oliveira, Raquel and P¨ol¨onen, Ilkka and others: Boreal Forest Fire: UAVcollected wildfire detection and smoke segmentation dataset, Scientific Data, Vol. 12, No. 1, p. 1419 (2025)

[4] Avazov, Kuldoshbay and Hyun, An Eui and Alabdulwahab, Abrar Sami S. and Khaitov, Azizbek and Abdusalomov, Akmalbek Bobomirzaevich and Cho, Young Im: Forest fire detection and notification method based on AI and IoT approaches, Future Internet, Vol. 15, No. 2, p. 61 (2023).

[5] Titu, Md Fahim Shahoriar and Pavel, Mahir Afser and Goh, Kah Ong Michael and Babar, Hisham and Aman, Umama and Khan, Riasat: Real-time fire detection: Integrating lightweight deep learning models on drones with edge computing, Drones, Vol. 8, No. 9, p. 483 (2024).

[6] De Venancio, Pedro Vinicius AB and Lisboa, Adriano C and Barbosa, Adriano V: An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices, Neural Computing and Applications, Vol. 34, No. 18, pp. 15349– 15368 (2022).

[7] Zhang, Qi-xing and Lin, Gao-hua and Zhang, Yong-ming and Xu, Gao and Wang, Jin-jun: Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images, Procedia Engineering, Vol. 211, pp. 441– 446 (2018).

[8] Muhammad Yaseen: What is YOLOv8: An In-Depth Exploration of the Internal Features of the NextGeneration Object Detector (2024).

[9] Glenn Jocher and Ayush Chaurasia and Jing Qiu: Ultralytics YOLOv8 (2023).

[10] Glenn Jocher: Ultralytics YOLOv5 (2020).

[11] Dou, Zhan and Zhou, Hang and Liu, Zhe and Hu, Yuanhao and Wang, Pengchao and Zhang, Jianwen and Wang, Qianlin and Chen, Liangchao and Diao, Xu and Li, Jinghai: An improved YOLOv5s fire detection model, Fire Technology, Vol. 60, No. 1, pp. 135–166 (2024).

[12] Gao, Pengcheng: A Fire and Smoke Detection Model Based on YOLOv8 Improvement, International Journal of Advanced Computer Science & Applications, Vol. 15, No. 3 (2024).

[13] Ma, Shuangbao and Li, Wennan and Wan, Li and Zhang, Guoqin: A lightweight fire detection algorithm based on the improved YOLOv8 model, Applied Sciences, Vol. 14, No. 16, p. 6878 (2024).

[14] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian: Deep Residual Learning for Image Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).

[15] Hu, Jie and Shen, Li and Sun, Gang: Squeeze-andExcitation Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018).

[16] Wang, Xin and Wang, Jinxin and Chen, Linlin and Zhang, Yinan: Improving Computer Vision-Based Wildfire Smoke Detection by Combining SE-ResNet with SVM, Processes, Vol. 12, No. 4, p. 747 (2024).

[17] Woo, Sanghyun and Park, Jongchan and Lee, JoonYoung and Kweon, In So: CBAM: Convolutional Block Attention Module, Proceedings of the European Conference on Computer Vision (ECCV) (2018).

[18] Hou, Qibin and Zhou, Daquan and Feng, Jiashi: Coordinate Attention for Efficient Mobile Network Design, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021).

[19] Wang, Qilong and Wu, Banggu and Zhu, Pengfei and Li, Peihua and Zuo, Wangmeng and Hu, Qinghua: ECANet: Efficient Channel Attention for Deep Convolutional Neural Networks, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020).

[20] Gaia Solutions on Demand: DFireDataset, GitHub (online), available from https://github.com/gaia-solutions-on-demand/DFireDataset (accessed 2026-02-09).

[21] Pesonen, Julius, et al. "Boreal Forest Fire: UAV-collected wildfire detection and smoke segmentation dataset." *Scientific Data* 12.1 (2025): 1419.

付録 A 定性評価

提案手法の有効性を視覚的に 検証するため, Baseline（YOLOv8m）と提案手法（Ours: Head-P5 への ESCFBlock の挿入モデル）による推論結果の比較を行った．図 A.1 に推論結果の比較を示す．図 A.2 には YOLO から自動出力される，推論結果を載せる．

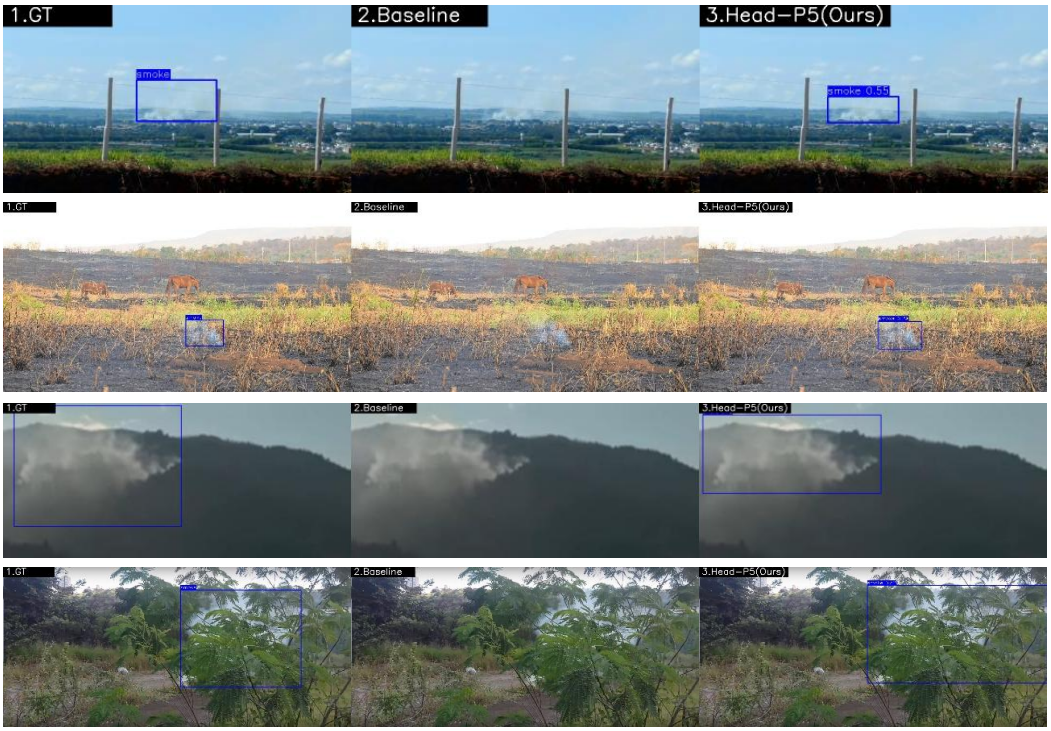

図 A.1 推論結果

左：GT（正解データ），中央：Baseline，右：Ours

図 A.2 Ours による推論結果

付録 B Smoke および Fire の項目別評価

5.4.2 節で実施した 100 エポックの長期学習での結果の詳細である．具体的には，ターゲットである Smoke および Fire に対する各カテゴリの評価を示す（表 B）．

表 B Smoke および Fire の項目別評価

Model	Category	Precision (P)↑	Recall (R)↑	F1 Score↑	AP@50↑	AP@50-95↑
Baseline	all	0.803	0.737	0.769	0.799	0.481
	Smoke	0.844	0.814	0.829	0.863	0.558
	Fire	0.762	0.659	0.707	0.735	0.403
Ours	all	0.797	0.745	0.770	0.795	0.478
	Smoke	0.844	0.826	0.835	0.861	0.559
	Fire	0.750	0.665	0.705	0.729	0.397

両者ともに，Smoke に対しては精度よく出ていることがわかる．今後の展望として，Fire に対する精度向上があげられる．