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

青木 健[1]　岡部 誠[1]

概要：屋外における火災検知において，監視カメラやドローンなどによる映像を用いた検知が注目されている．しかし，固定的な形を持たない炎や煙の識別や，複雑な背景下における見逃しや誤検知が課題となっている．本研究はリアルタイム性に優れた物体検知アルゴリズム YOLOv8 をベースに，空間情報とチャネル情報を，効率的に融合する独自のアテンション機構 ESCFBlock（Efficient Spatial-Channel Fusion Block）の導入を提案する．本アテンション機構は，Coordinate Attention と Efficient Channel Attention を並列に配置し，ゲート付き残差結合により，特徴を適応的に強調する．D-fire データセットを用いた評価実験の結果，Head 部の P5 層への導入が最も効果的であった．特に，Recall の向上が確認されたことは，火災の見逃しを最小限に捉える実用的な成果である．

キーワード：火災検知，YOLOv8，深層学習

1. はじめに

火災は，人的被害のみならず，歴史的建造物や貴重な自然資源の焼失，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抑えるためには，火災の発生を早期かつ正確に検知し，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，煙感知器や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な制約が存在する．例えば，開放的な屋外施設や，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，火災検知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，画像や映像による火災検知には特有の難しさがある．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．煙は風によって拡散し，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．また，炎は照明条件や周囲の反射物の影響を強く受け，夕日や赤色の照明などと誤検知を招きやすい．特に，遠方で発生した小さな火種や，希薄な煙を精度良く検知することは，従来の画像処理技術における大きな課題である．

近年，これら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内で，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]．本研究で採用した 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．

そこで本研究では，YOLOv8 をベースに，空間情報とチャネル情報を効率的に融合する独自のアテンション機構 ESCFBlock（Efficient Spatial-Channel Fusion Block）を提案し，高精度かつリアルタイムな火災検知システムを目指す．この ESCFBlock を YOLOv8 ネットワーク内の最適な階層に組み込むことで，誤検知の抑制と，火災の検出漏れ防止を同時に達成するモデルを構築し，その有効性を比較実験により実証する．

[1] 静岡大学
Shizuoka University

2. 関連研究

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を報告している．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 SE-ResNet 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] という統合型アテンションを YOLO に導入する手法を提案している [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．

3. 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（ES-CFBlock）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．

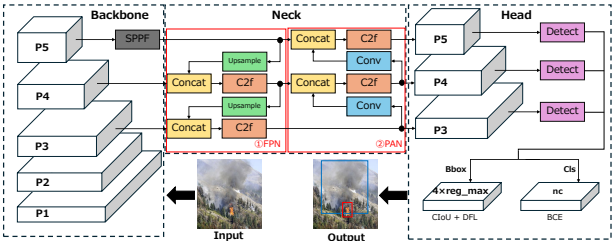

図 1: YOLOv8 の構造図

3.1 YOLOv8

本研究で採用した YOLOv8 のネットワークアーキテクチャは，主に Backbone，Neck，Head の 3 つのコンポーネントによって構成されている．**図 1** に YOLOv8 の全体構造図を示す．

■ **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，浅い層では視覚的詳細情報を，深い層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．FPN（Feature Pyramid Network）（図 1①）と PAN（Path Aggregation Network）（図 1②）を組み合わせた構造を採用しており，深い層からの意味的情報と，浅い層からの視覚的情報を，双方向に伝播させる．これにより，すべての特徴マップにおいて物体の識別能力と位置特定能力の両方を強化することを可能にしている．

■ **Head** Head は，Neck を通じて精製された特徴マップを受け取り，最終的な物体の位置特定およびクラス分類を行う．YOLOv8 では，位置特定とクラス分類を独立した経路で処理している．なお，学習時の損失関数として，位置特定には，CIoU（Complete IoU）と DFL（Distribution Focal Loss）を，クラス分類には，BCE（Binary Cross Entropy）を採用している．また，アンカーフリー方式への移行により，不定形な対象への柔軟な対応を実現している．

3.2 Coordinate Attention

一般的なチャネルアテンションでは，情報を圧縮するために 2 次元の Global Average Pooling が用いられるが，この過程で空間情報が失われるという課題がある．これに対し，Coordinate Attention（CA）は，チャネル間の依存関係と空間情報を同時に捉えるため，空間情報を保持したまま，符号化する仕組みを有する．CA は，入力特徴マップ $X \in \mathbb{R}^{C \times H \times W}$ に対して，以下のように処理を行う．また，

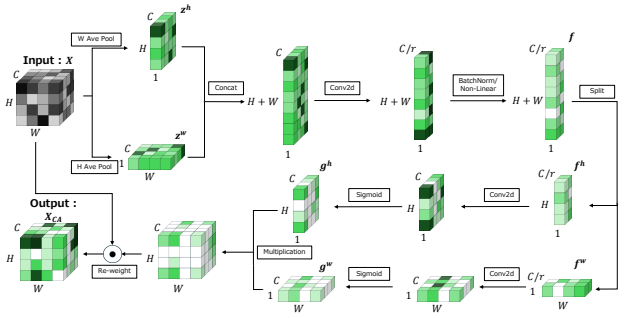

図 2: Coordinate Attention の演算プロセス概要図

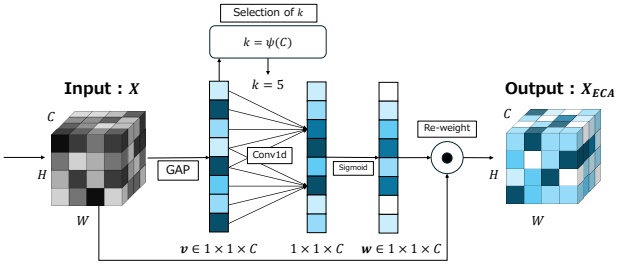

図 3: Efficient Channel Attention の演算プロセス概要図

CA の演算プロセスの概要を**図 2** に示す．

(1)　位置情報の集約
空間情報を保持するため，入力特徴マップ $X \in \mathbb{R}^{C\times H\times W}$ に対し，水平・垂直方向それぞれに独立した 1 次元平均プーリングを適用する．

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

$$z_c^w(w) = \frac{1}{H}\sum_{0\le j\le H} x_c(j,w) \tag{2}$$

これらの式に基づき，水平方向と垂直方向の情報を集約できる．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
得られた z^h と z^w を結合し，サイズが $C\times(H+W)\times 1$ の統合特徴マップを得る． 得られた特徴マップに対して，1×1 の畳み込みを行い，チャネル数を圧縮する．結果として，サイズが $(C/r)\times(H+W)\times 1$ 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す．

(3)　情報の安定化と非線形変換
畳み込みによって圧縮された特徴マップに対し，バッチ正規化と非線形活性化関数 δ を適用する．これにより，学習の安定化を図るとともに，水平・垂直方向の空間的な依存関係を，非線形にモデル化する．この処理により，中間特徴マップ $f \in \mathbb{R}^{(C/r)\times(H+W)\times 1}$ が生成される．

$$f = \delta(\mathrm{BN}(\mathrm{Conv2d}([z^h, z^w]))) \tag{3}$$

ここで，$[\cdot,\cdot]$ は結合処理を表す．

(4)　分割とアテンションウェイトの生成
続いて，中間特徴マップ f を垂直方向の特徴 $f^h \in \mathbb{R}^{(C/r)\times H\times 1}$ と，水平方向の特徴 $f^w \in \mathbb{R}^{(C/r)\times 1\times W}$ に分割する．それぞれに対して 1×1 畳み込み (F_h, F_w) とシグモイド関数 σ を適用することで，各方向のアテンションウェイト g^h, g^w を算出する．

$$g^h = \sigma(F_h(f^h)), \quad g^w = \sigma(F_w(f^w)) \tag{4}$$

(5)　出力の算出
最後に，算出された水平・垂直方向のウェイトを入力特徴マップ X に要素ごとに乗算し，位置情報を強調した出力 X_{CA} を得る．

$$\bar{x}_c(h,w) = x_c(h,w)\times g_c^h(h)\times g_c^w(w) \tag{5}$$

出力として，特定の空間情報を強調した再構成特徴マップ $X_{CA} = [\bar{x}_1, \bar{x}_2, ..., \bar{x}_C] \in \mathbb{R}^{C\times H\times W}$ を得る．

3.3　Efficient Channel Attention

Efficient Channel Attention（ECA）は，従来のチャネルアテンションにみられる次元削減による情報の損失を回避し，極めて少ない計算コストで，チャネル間の依存関係を学習する手法である．この手法は，適応的なカーネルサイズを持つ 1 次元畳み込みを用いて，局所的なチャネル間相互作用を，直接捉えることで，パラメータ数を大幅に抑制しつつ，効果的な特徴強調を実現する．ECA は，入力特徴マップ $X \in \mathbb{R}^{C\times H\times W}$ に対して，以下のように処理を行う．また，ECA の演算プロセスの概要を**図 3** に示す．

(1)　特徴の集約
まず，入力特徴マップ X 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せる．

$$v_c = \frac{1}{WH}\sum_{0\le i\le W}\sum_{0\le j\le H} x_c(i,j) \tag{6}$$

入力特徴マップ X のチャネル数が C であるから $\boldsymbol{v}_c \in \mathbb{R}^{1\times1\times C}$ 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，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）．

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{7}$$

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)　重み付けと出力

算出された k を用いた 1 次元畳み込みと，シグモイド関数 σ により，アテンションウェイト $\boldsymbol{w}$ を生成する．

$$\boldsymbol{w} = \sigma(\mathrm{Conv1d}_k(\boldsymbol{v})) \tag{8}$$

得られた $\boldsymbol{w}$ を，入力特徴マップ X に対して，要素ごとに乗算する．$\odot$ を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．

$$X_{ECA} = \boldsymbol{w} \odot X \tag{9}$$

出力として，重要な特徴を持つチャネルが強調され，重要度の低い情報が抑制された再構成特徴マップ $X_{ECA} \in \mathbb{R}^{C\times H\times W}$ を得る．

4. 提案手法

本研究では，このトレードオフを解消するために，新たな特徴融合アテンション機構 ESCFBlock（Efficient Spatial-Channel Fusion Block）を提案する．ESCFBlock は，空間情報の符号化に特化した Coordinate Attention（CA）と，計算コストを最小限に抑えつつ，チャネル間の相関を捉える Efficient Channel Attention（ECA）の 2 つの機構を並列に統合したものである．さらに，本手法ではこれらを単に組み合わせるだけでなく，ゲート付き残差接続を導入することで，入力画像に応じて空間情報と，チャネル情報の重要度を動的かつ適応的に調整することを可能にした．これにより，YOLOv8 が本来持つ推論速度を損なうことなく，炎や煙の微細な特徴を強調し，複雑な背景ノイズと明確に分離することが可能となる．

4.1 統合のアプローチと設計思想

前章で述べたように，CA と ECA は計算コストを抑えつつ，特徴を強調する点では，共通しているが，情報の集約方法において，明確なトレードオフが存在する．

■ **CA の特性**：水平・垂直方向のプーリングにより，空間情報の保持に優れるが，チャネル間の複雑な相関関係を捉える能力は，限定的である．

■ **ECA の特性**：局所的な相互作用により，チャネル間の相関の識別能力は高いが，空間情報を圧縮することにより，正確な座標情報を保持できない．

複雑な背景が存在する実環境下の火災検知において，誤検知を抑制し高い信頼性を確保するためには，炎や煙特有のテクスチャ情報を詳細に捉えることと，背景ノイズから対象を分離するための空間情報を正確に保持することの両立が不可欠である．この観点において，空間情報の符号化に長けた CA と，チャネル間の特徴抽出に特化した ECA は，互いの弱点を補う相補関係にあるといえる．そこで，本研究では，これら 2 つの機構を並列に統合した新たなアテンション機構を構築する．本手法の目的は，YOLOv8 が有するリアルタイムな推論速度を損なうことなく，微細な特徴の識別能力と位置特定精度を向上させることにある．

4.2 提案手法：ESCFBlock

ESCFBlock(Efficient Spatial-Channel Fusion Block) は，CA と ECA の特性を最大限に引き出すため，単なる直列構造ではなく，並列構造とゲート付き残差学習を統合した設計となっている．以下に，処理過程を示す．また，**図 4** に，提案手法である ESCFBlock の概要を示す．

(1)　特徴抽出

まず，入力特徴マップ $X \in \mathbb{R}^{C\times H\times W}$ を得ると，CA ブランチ（図 4①）は空間情報を重視した特徴マップ X_{CA} を，ECA ブランチ（図 4②）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 $f_{concat} \in \mathbb{R}^{2C\times H\times W}$ を生成する．

(3)　相互作用の学習と次元削減

(2) で得られた特徴マップ f_{concat} に対して，1×1 畳み込み層を適用し，チャネル数を $2C$ から C へと削減する．また，空間情報とチャネル情報という異なる性質を持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 $\bar{X} \in \mathbb{R}^{C\times H\times W}$ を得る．ここで，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)　ゲート付き残差接続

本手法では，入力特徴マップ X に対して，融合特徴マップ $\bar{X}$ を加算する際，学習可能なスケーリング係数 α を介した，ゲート付き残差接続を採用する．

$$X_{ESCF} = X + \alpha \cdot \bar{X}, \quad \alpha = \sigma(\mathrm{gate}) \tag{10}$$

ここで，$\mathrm{gate} \in \mathbb{R}$ は学習可能なパラメータであり，シグモ

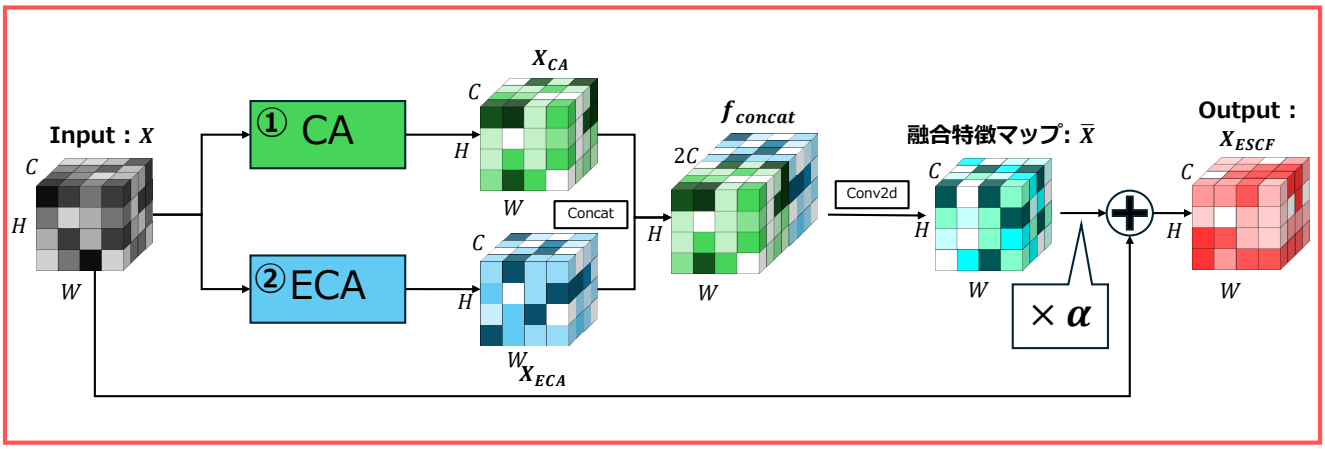

図 4: 提案手法：ESCFBlock の演算プロセス概要図

イド関数 σ を適用することで，係数 α を $(0,1)$ の範囲に正規化している．このゲート機構の導入には，以下の 2 つの重要な意義がある．

- **適応的な特徴選択**：学習可能な係数 α を導入することで，入力特徴マップから生成されたアテンションの重要度（寄与率）を最適化し，必要な特徴のみを適応的に強調することが可能である．
- **学習の安定化**：α の値を調整することで，アテンションブランチからの情報の寄与率を適応的に制御できる．特に，学習初期においては，未学習の特徴マップが，YOLOv8 のネットワークの挙動を乱すことを防ぎ，安定した勾配の伝播を保証する役割を果たす．

本手法では，YOLOv8 の事前学習済み重みがもつ汎用的な特徴抽出能力を維持しつつ，ESCFBlock を段階的に適応させるために，特徴融合層の重みとバイアスをすべて 0 で初期化している．この状態において，学習開始直後の融合特徴マップ $\bar{X}$ は 0 となり，ESCFBlock の初期出力は $X_{ESCF} = X + \alpha \cdot 0 = X$ となる．これにより，未学習の ESCFBlock が YOLOv8 のネットワーク全体に悪影響を及ぼすことを防ぐことができる．また，学習の進行とともに，α および融合層の重みが更新され，徐々に最適な特徴強調が追加される仕組みとなっている．

5. 実験

5.1 実験設定

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煙（Smoke）と炎（Fire）の 2 クラスが含まれる．このうちの 8 割（17,221 枚）を訓練用，2 割（4,306 枚）をテスト用とした．

ベースモデルには，YOLOv8 [9] 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．

表 1: 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

また，本研究において，すべての実験は，**表 1** に示す計算資源で実施した．

5.2 評価指標

本研究では，提案手法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）という 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．

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

上記の式において，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．

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

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 挿入位置の比較評価

5.3.1 概要

YOLOv8m のネットワーク内の 7 か所（①Head 部 P3 層，②Backbone 部 P3 層，③Head 部 P4 層，④Backbone 部 P4 層，⑤Head 部 P5 層，⑥SPPF モジュール直前，⑦SPPF モジュール直後）に挿入した（**図 5**）．YOLOv8m に対し，D-fire を用いて学習を行ったものを Baseline とする．また，ESCFBlock をそれぞれの位置に挿入したモデルについても同様に D-fire を用いて学習を行い，比較検証を行った．なお，モデルの評価には学習プロセスにおいて，検証データに対し，最も高い性能を示した重みファイル（best.pt）を採用した．

モデルの比較検証を公平に行うため，Baseline および ESCFBlock を挿入した全てのモデルに対して，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**表 2** に示す．ここで，Initial Learning Rate，Momentum および Decay については，Dou らによる YOLOv5 を用いた火災検知モデルの構築の研究 [11] に基づき設定した．

表 2: 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

5.3.2 結果・考察

実験結果を**表 3** に示す．これより，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ことが確認できる．以下に，各指標および挿入位置の観点から，本構成の有効性を分析する．

まず，検出の網羅性を示す Recall において，Head-P5 は Baseline の 0.731 から 0.743 へと改善（+0.012）を示した．火災検知タスクにおいて，見逃し（偽陰性）は致命的な被害に直結するため，Recall の改善は最も重視すべき成果である．これは，ESCFBlock が火災特有の微細な特徴を効果的に強調し，検出漏れの低減に寄与したことを実証している．

次に，総合的な性能バランスを示す F1 Score においても，0.757 から 0.760 への改善が確認された．一般に，Recall と適合率（Precision）はトレードオフの関係にあり，一方を向上させると他方が低下する傾向にある．しかし，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，実運用において過度な誤報を抑制しつつ，確実な検知を実現できることを意味する．

挿入位置について考察すると，Backbone 部のような低次特徴層への挿入よりも，Head 部のような高次元特徴層への挿入が有効であることが明らかとなった．Backbone 部ではエッジやテクスチャなどの局所的な特徴が支配的であるのに対し，Head 部（特に P5 層）では物体としての意味的情報が形成される．火災検知においては，単純な輝度や色情報だけでなく，意味を含めた判断が必要となるため，深層におけるアテンションによる特徴強調が，性能向上に最も寄与したと推察される．

5.3.3 アテンション寄与率 α の分析

本節では，ESCFBlock 内のゲート付き残差接続における学習可能なスケーリング係数 α の挙動を分析する．なお，事前学習済みモデルへの急激な干渉を避けるため，全条件において α の初期値を一律 0.1 に設定した．

実験の結果，α の最終値は ②Backbone-P3 で最大値（0.163）を記録した．一方，最高精度を達成した⑤Head-P5 では適度な上昇（0.130）にとどまった．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．

Backbone-P3 における α の増大は，炎や煙の質感（色やテクスチャ）といった低次特徴の強調が，損失低下に寄与したことを意味する．しかし，初期層で視覚的特徴を過度に強調することは，夕日や雲など，炎や煙に類似した背景ノイズまでも増幅させ，結果として誤検知（Precision の低下）を招く要因となった．

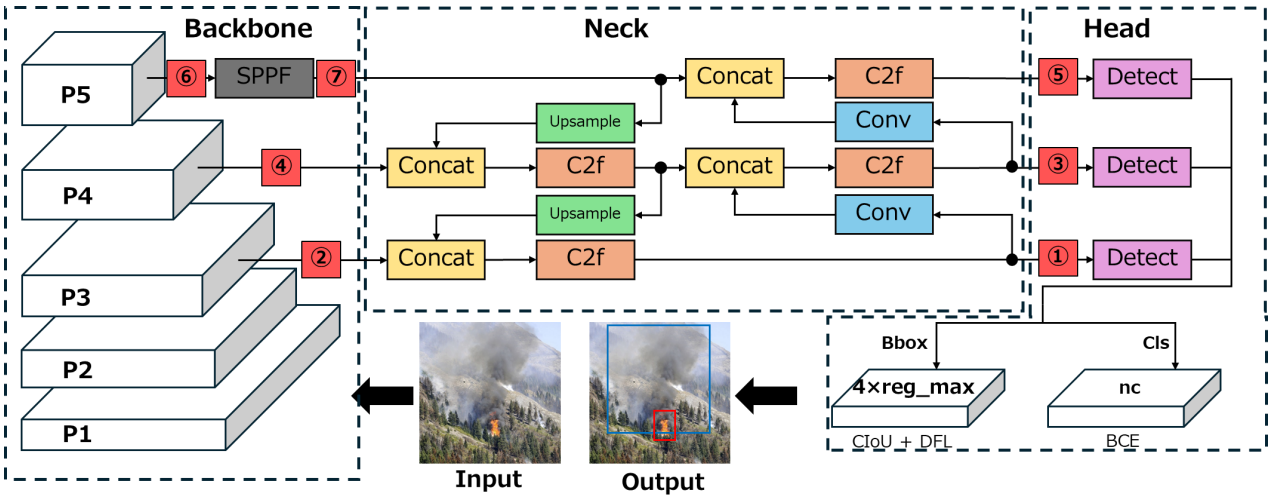

図 5: ESCFBlock の挿入位置

表 3: 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-SPPF	26,584,468	0.784	0.718	0.750	0.782	0.462	0.115
⑦post-SPPF	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 4: 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**	-

表 5: アブレーションスタディ（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**

対照的に，Head-P5 は物体検出の最終判断直前の階層に位置し，より高次な意味的情報を扱う．そのため，単なる特徴の強調ではなく，画像全体の文脈に基づいた判断の洗練に，アテンションが機能したと考えられる．

以上の結果より，火災検知においては，低次特徴を単純に強調するよりも，高次の意味情報に基づいて特徴を選別・統合するアプローチが，誤検知の抑制と検出能力の最大化に有効であると結論付けられる．

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

前節の実験で最も優れた性能を示した Head-P5 構造が，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) を固定した 5 回の独立試行による追加実験を行った．比較対象は Baseline とし，ESCFBlock が火災検知において有効なアテンション機構であるかを分析する．なお，本実験においても，検証データに対して最も高い性能を示した重みファイル（best.pt）を用いて比較を行った．**表 4** に両モデル（Baseline と Head-P5 構造）の 5 回の独立試行を行った結果の，各指標の平均値

図 6: 定性評価
左：GT（正解データ），中央：Baeline，右：Ours による推論結果

を示す．

総合的な性能バランスを示す F1 Score においては，Head-P5 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 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 2 の通りである．結果を**表 5** に示す．

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 と低下したことに対し，ESCFBlock では 0.740 と改善が確認された．

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．

5.5 定性評価

最後に，定性評価を行う．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 6** に，結果を示す．左列は正解データ（Ground Truth），中央列は Baseline による推論結果，右列は提案手法による推論結果である．推論対象は，D-fire データセットの画像であり，Baseline では検知できなかった煙が，提案手法では，検知できていることがわかる．

6. 結論

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 と比較して改善が確認された．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求が不可欠である．

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, Väinö and Koivumäki, Niko and Markelin, Lauri and Raita-Hakola, Anna-Maria and Suomalainen, Juha and Pölönen, 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 Koivumäki, Niko and Markelin, Lauri and Suomalainen, Juha and Alves de Oliveira, Raquel and Pölönen, Ilkka and others: Boreal Forest Fire: UAV-collected 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 Next-Generation 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-and-Excitation 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, Joon-Young 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: ECA-Net: 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?tab=readme-ov-file⟩ (accessed 2026-02-09).