

# 浙江财经大学

## 大作业报告

プロジェクト名 ディープラーニング——TensorFlow に基づく人工知能の実践応用

コース名 深度学习

プロジェクトタイプ 総合的

実験(実習)日 2025.6.4

クラス 人工知能 2

学生番号 220110900830

学生氏名 吳国涛

指導教員 齐峰

## 一、テーマ選定の背景

ディープラーニングの分野において、物体検出技術はその広範な応用シーンと実用性から大きな注目を集めています。今学期の学習の中で、物体検出は汎用性の高い機能であると考え、研究テーマとしました。桜には多くの品種があるため、私は桜を研究対象に選びました。桜の物体検出研究を通じて、物体検出アルゴリズムの原理と実装を深く理解し、実際のシーンでの応用を探求したいと考えています。

## 二、開発環境

Python 3.9

Anaconda

Jupyter Notebook

pytorch 2.7.1

## 三、データ処理

データソース: <https://universe.roboflow.com/rsnaru-jp-gmail-com/sakura-jklcu>

特徴: このデータセットは、位置とラベルが既にタグ付けされたデータセットです。サンプル数は 1400 以上あり、十分な量です。また、複数のフォーマットが用意されていますが、ここでは YOLO 形式のデータを選択しました。

前処理: 解凍後のデータは、各画像と情報ファイルの名前は対応していますが、名前が文字化けしているため、00001.jpg、00001.txt のような形式にリネームする必要があります。

本データセットでは、データは既に以下のように分類されています:

train/images

train/labels

test/images

test/labels

valid/images

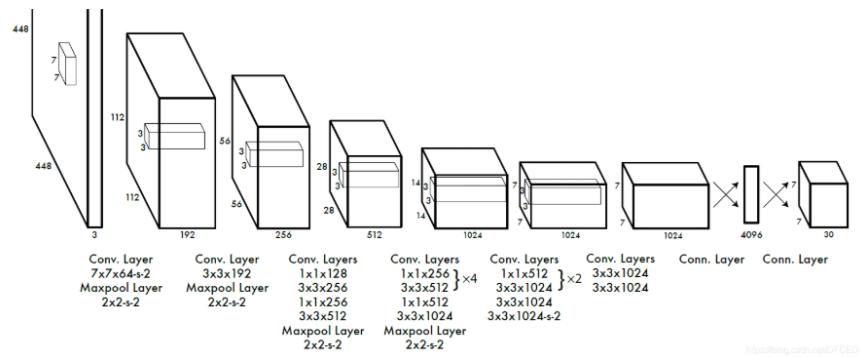
valid/labels

data.yaml (データのアドレス、種類の数、ラベル名を保存)

このデータセットは既に YOLO モデルに適応しているため、その他の処理を行う必要はありません。

## 四、モデル設計

このプロジェクトでは YOLO モデルを使用しました。



### YOLO モデル解説:

入力: 448\*448\*3 (RGB 画像)

出力:  $S \times S \times (B \times 5 + C)$

$S=7$ : 画像を  $7 \times 7$  のグリッドに分割

$C=20$ : クラス数

畳み込み層部分 (24 層)、全結合部分 (2 層)

損失関数:

座標予測誤差: 平均二乗誤差

分類予測誤差: 交差エントロピー

バウンディングボックス信頼度予測誤差: 平均二乗誤差

### YOLO アルゴリズムの思想

画像全体をグリッドの入力として利用し、出力層で位置とクラスを直接回帰させます。

手順:

画像を入力し、まず画像を  $S \times S$  のグリッドに分割します。

各グリッドについて、 $B$  個のバウンディングボックスを予測します。

前のステップに基づき、 $S \times S \times B$  個のターゲットウィンドウが得られるので、閾値に基づいて選別します。

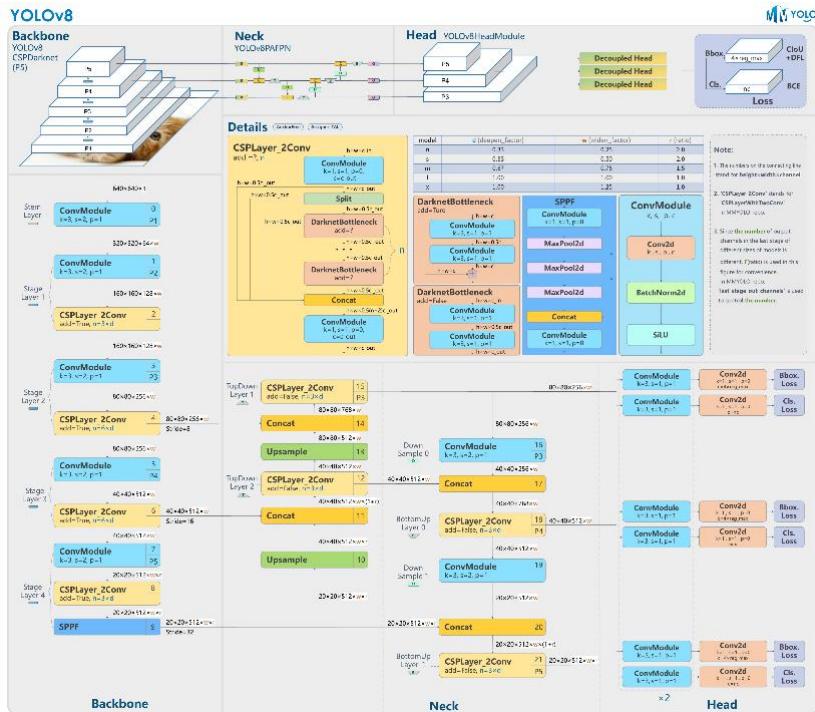
各セルは  $C$  個のクラス確率値を出力します。

$S=7, B=2, C=20$  を使用する場合、最終結果は  $7 \times 7 \times 30$  のテンソルになります。

## YOLO モデルの欠点:

近接している物体や、非常に小さな集団に対する効果が良くありません。テスト画像において、同一種類の物体が出現した場合や、新しいアスペクト比などの状況での予測効果が良くありません。

## yolo8:



## yolo8 モデル解析:

画像入力後、Conv と C2f によるそれぞれの処理を経て、ピラミッド P1 から P5 への処理を実現します。

**Conv:** 2 次元畳み込みを行い、その後 BatchNorm2d で正規化し、最後に SiLU 活性化関数を通します。

**C2f:** メインパスとバイパスを持ちます。メインパスには畳み込み層、正規化層、活性化関数があり、対してバイパスは畳み込み層が少なく、最後にメインパスとバイパスを結合 (Concat) し、特徴融合を実現します。

ピラミッド P5 に到達したデータは、アップサンプリングを経て P4 のデータと結合され、C2f、アップサンプリングを経て、P3 のデータと結合されます。再度 C2f を経た後、逆の操作（ダウンサンプリングと C2f）を行い、P4 のデータと結合、再度ダウンサンプリングと C2f を行い、P5 のデータと結合します。得られた 3 つの特徴テンソルを Detect に入力して予測を行い、位置と分類の損失関数を取得します。

**SPPF:** 異なるスケールのプーリング操作に使用され、異なるスケールの特徴マップを結合することで、異なるスケールのターゲットに対する検出能力を向上させます。

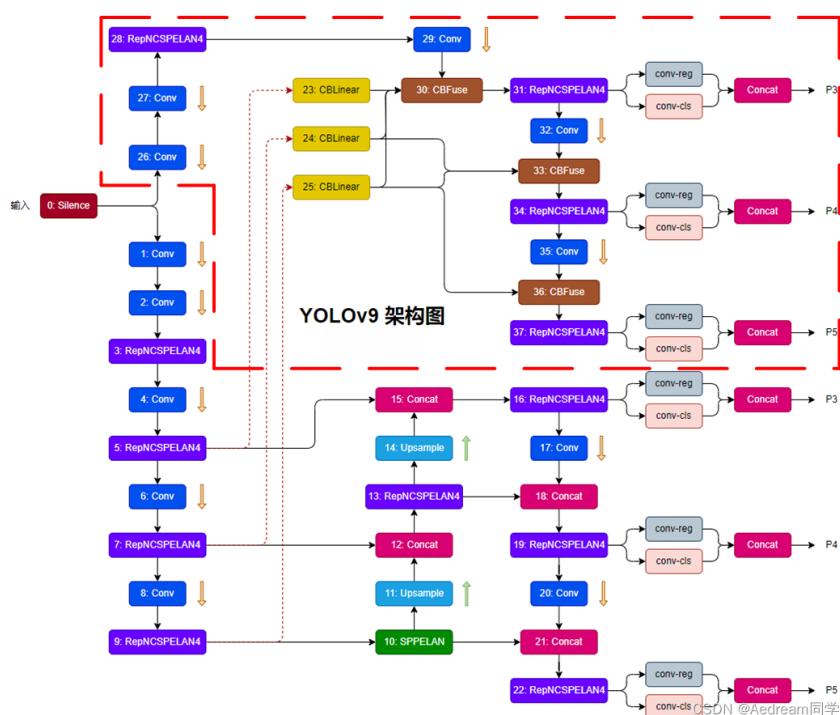
## YOLOv1 -> YOLOv8 の改良点:

固定グリッド分割を使用せず、バウンディングボックスのオフセット量を直接予測します。

固定スケールではなく、マルチスケール特徴融合を使用します。

ターゲットのサイズに応じて動的にラベルを割り当てます。

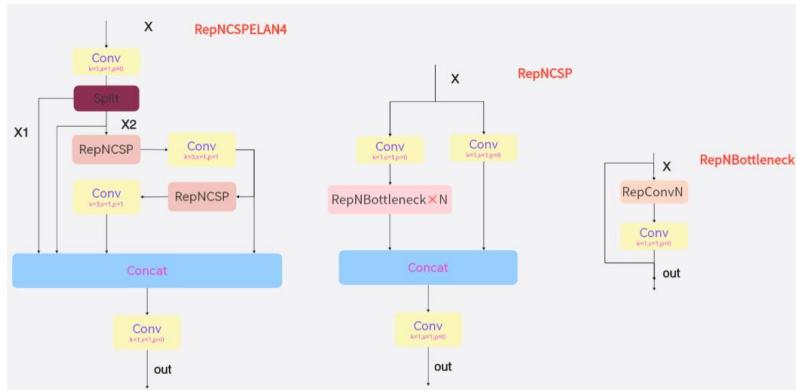
## yolo9:



## yolo9 モデル解析:

バックボーンは YOLOv8 と似ていますが、C2f モジュールが RepNCSPELAN4 モジュールに変更されました。また、元の入力に新しい分岐を追加し、バックボーンの 5、7、9 部分の特徴をその分岐に結合します。その後、CBFuse、畳み込み、RepNCSPELAN4 の操作を行い、予測を行います。30 は 3 つの特徴が結合された勾配情報（つまりモデルトレーニングの浅層情報）を受け取ることができます、その後 30 は分岐で学習した情報をバックボーンの 5、7、9 に分配し、バックボーンの学習を促進します。

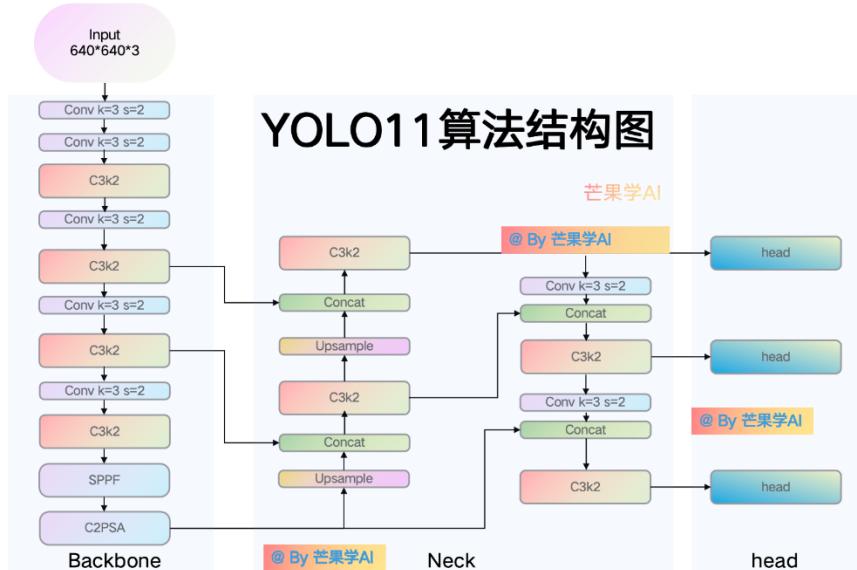
RepNCSPELAN4: C2f 構造に類似しています。



**CBFuse:** 特徴融合層であり、異なる深さの特徴マップ間で効果的な特徴統合を行うために使用されます。異なる階層の特徴マップを統一されたサイズにサンプリングしてから融合します。

**SPPELAN:** SPPF の改良構造であり、高い精度を維持しながらモデルの計算量を削減し、推論速度を向上させることを目的としています。

### yolo11:

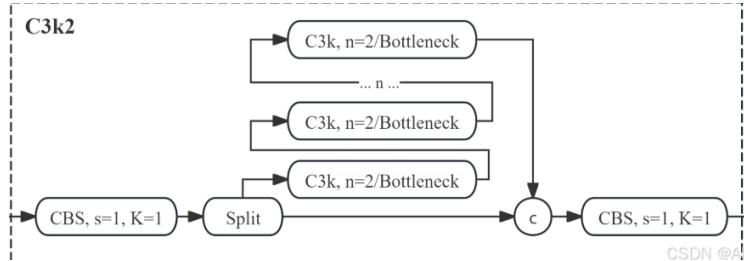


### YOLOv11 モデル解析:

全体的に YOLOv8 の構造と似ていますが、C2f が C3k2 に置き換わり、SPPF の後に C2PSA 自己注意モジュールが追加されました。

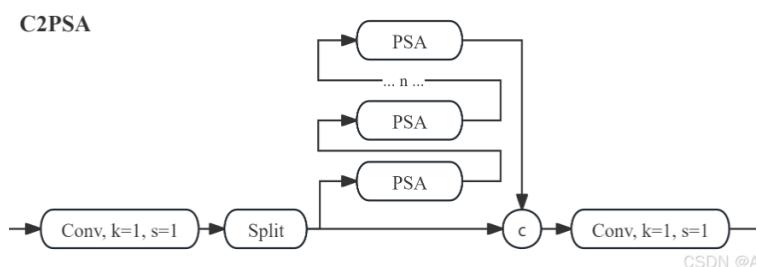
### C3K2:

マルチスケールの畳み込み核 C3K を導入しました。K は調整可能な畳み込み核のサイズ（例：3x3、5x5）であり、モデルがより広範なコンテキスト情報を捕捉できるようにします。C3k2 モジュールは通常、入力特徴を 2 つの部分に分けます。一部は通常の畳み込み操作を通じて直接伝達され、もう一部は複数の C3K または Bottleneck 構造を通じて深層特徴抽出が行われます。最終的に 2 つの部分の特徴が結合され、1x1 畳み込みを通じて融合されます。



### C2PSA:

CSP のセグメント化された特徴処理の思想を継承しています。特徴は 1 回  $1 \times 1$  置み込みを経た後、2 つの部分に分けられます。一部は直接伝達され、もう一部は PSA 注意モジュールによる処理を経て、その後 2 つの部分の特徴が結合され、別の  $1 \times 1$  置み込みを経て元のチャンネル数に復元されます。PSA メカニズムは、多種の置み込み核を通じてマルチスケール特徴を抽出します。異なる置み込み核の置み込み操作は並列に行われ、その後特徴マップが結合され、SE モジュールを使用して特徴チャンネルに重み付けを行います。



## 五、モデルトレーニング

```
model_yolov8.train(data='data.yaml', epochs=50, batch=1, task='detect+classify')
```

ここでのトレーニングでは、ランダムバッチ、識別＋トレーニングのモードを使用しています。当初、トレーニングのエポック数を 30 に設定していましたが、出力データを観察したところ、25～30 エポックの間でも損失値が収束し続けていたため、何度か調整を行い、最終的に 50 エポックに設定しました。

## 六、モデル評価

各モデルについて、4 種類の方法で評価を行いました。

一、出力されたモデルを検証セットで評価し、以下を算出：

1. 適合率 (Precision)
2. 再現率 (Recall)
- 3.mAP 50: IoU 閾値が 0.5 の時の平均適合率の平均値。
- 4.mAP50-95 : IoU 閾値が 0.5 から 0.95 の時の平均適合率の平均値。 (IoU は予測枠と真の枠の重なり面積とそれらの和集合の面積の比率)

二、出力されたモデルをテストセットで評価し、以下を算出

- 1.適合率
- 2.再現率
- 3.mAP 50
- 4.mAP50-95

三、モデルトレーニング時に保存されたいくつかの情報を出し、モデルのトレーニング過程を観察：

1. 正規化された混同行列  
**(confusion\_matrix\_normalized.png)**
2. 信頼度閾値に伴う F1 スコアの変化曲線  
**(F1\_curve.png)**
3. 信頼度閾値に伴う適合率 (Precision) の変化曲線  
**(P\_curve.png)**
4. 信頼度閾値に伴う再現率 (Recall) の変化曲線  
**(R\_curve.png)**
5. 適合率-再現率曲線 (Precision-Recall Curve)  
**(PR\_curve.png)**
6. トレーニング過程における主要指標の変化曲線  
**(results.png)**

四、モデルをテストセットのランダムな画像でテストし、実際の効果を観察

最終的に、YOLOv11 モデルの効果が最も良いという結論に至りました。

## 七、講義の感想

## 八、参考文献

[https://blog.csdn.net/Alex\\_Tlover/article/details/142764338](https://blog.csdn.net/Alex_Tlover/article/details/142764338)

[https://blog.csdn.net/qq\\_38668236/article/details/143820170](https://blog.csdn.net/qq_38668236/article/details/143820170)

[https://www.bilibili.com/video/BV1XNUaYzEgy/?spm\\_id\\_from=333.788.top\\_right\\_bar\\_window\\_history.content.click](https://www.bilibili.com/video/BV1XNUaYzEgy/?spm_id_from=333.788.top_right_bar_window_history.content.click)

[https://www.bilibili.com/video/BV1Bx421C7qL/?p=2&spm\\_id\\_from=333.337.top\\_right\\_bar\\_window\\_history.content.click](https://www.bilibili.com/video/BV1Bx421C7qL/?p=2&spm_id_from=333.337.top_right_bar_window_history.content.click)

[https://www.bilibili.com/video/BV1DM4m1o7X3/?spm\\_id\\_from=333.788.top\\_right\\_bar\\_window\\_history.content.click](https://www.bilibili.com/video/BV1DM4m1o7X3/?spm_id_from=333.788.top_right_bar_window_history.content.click)