

修士論文要旨(2021 年度)

レイヤーとアテンションを追加したYOLO-v4による小さな目標に頑健な物体検出

Robust Small Objects Detection Using YOLO-v4 with attention and Layer

電気電子情報通信工学専攻 李 越東

1. まえがき

近年、急激に発展した画像認識技術と処理装置の性能向上を受け、深層学習を代表として、大量データ処理において機械学習も重要な役割を果たしている。多くの場合、物体の形態やテクスチャを用いたパターン認識が用いられる。しかし、現在格段に発展を遂げているパターン認識技術ではあるが、本質的な課題も存在する。例えば、災害救助や地球観測等において、広域のリモートセンシング画像内から小さな物体を検出する技術の需要は高くなる。しかし、対象物体の検出を要するため深層学習に用いる教師データを用意できないだけでなく、物体が小さく十分な形状特徴が得られないという問題点もある。この問題点を解決するために、従来の物体検出や画像認識のアルゴリズムを超えた新たなアルゴリズムの実現が期待される。

従来研究としては、ニューラルネットワークを用いた目標検出アルゴリズムは、主にR-CNN (Region-Convolutional Neural Network) シリーズに代表される2段階の目標検出アルゴリズムと、SSD (Single Shot multi box Detector) やYOLO (You Only Look Once) シリーズである[1]。

従来のモデルは最後に得られる特徴マップのみを利用していたため、1つの特徴マップが担当している領域が広がっており比較的小さい目標が検出できない課題があった。これらの理由か

ら、本実験では3つの特徴マップがあるYOLOモデルを使用することとする。

2. 提案手法

本研究において、筆者はYOLOv4に基づき、活性化関数の修正、バックボーンネットワークの修正、Attentionモジュールの追加、アンカーボックス (Anchor box) とパラメータの調整などの観点から、リモートセンシング画像の中に、小さな目標を検出するために用いる技法を提示する。

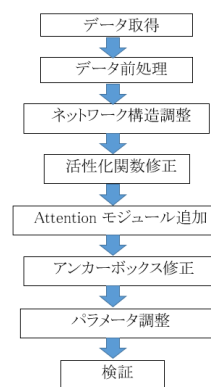


図-1: 提案手法のフロー

2.1 ネットワーク構造修正

YOLOv4[2]のアルゴリズムでは、特徴抽出にCSPDarknet-53ネットワークを使用している。このネットワークは空間分解能を捨てているため、特徴を深く畳み込んで抽出しても、小さな目標の情報は失われやすいである。

小さな目標の検出には、高解像度の特徴マップが必要であり、より広い受容野または多い情報を取得することで、目標の位置を正確に判断することができる。画像内の小目標の検出を改善するた

め、さらに高解像度化、大域な受容野のバックボーンネットワークを提案する必要がある。改良されたネットワークの構造を図-2に示す。

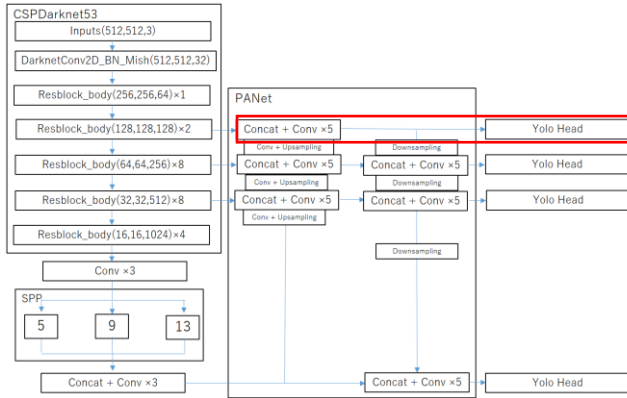


図-2: 提案するネットワーク構造

従来YOLOv4ネットワーク構造では、深いネットワーク層は小さな目標に対してより有効である。従来のYOLOv4に対して、出力の特徴マップは $52 \times 52 \times 3 \times (5 + \text{class_num})$, $6 \times 26 \times 3 \times (5 + \text{class_num})$, $13 \times 13 \times 3 \times (5 + \text{class_num})$ であり、大きな特徴図の受容野はより大きくなる。本研究では、小さな目標に対する $104 \times 104 \times 3 \times (5 + \text{class_num})$ の特徴検出層を追加することで、ネットワーク構造がより深くなり、ニューラルネットワークはさらに深い情報を学習させる。改良された特徴抽出ネットワークは、当初の三つの検出スケールを四つの検出スケールに拡張し、マルチスケール特徴融合により小目標の検出性能を向上させている。

2.2 活性化関数(Activation function)修正

ニューラルネットワークにおいて、活性化関数は、ニューラルネットワークの中にいくつかの非線形因子を追加することである。その結果、ニューラルネットワークはより複雑な問題をうまく解決することができ、モデルの表現力を上げる役割を果たす関数である。一般的な活性化関数は、主に四つがあり、それぞれはReLU, softmax, sigmoid, tanh関数である。しかし、より良い結果を取得して、上記の四つ活性化関

数以外にも様々な代替の活性化関数が考えられてきている。その中でも最近（2019年後半～執筆時点＝2021年12月時点で）、一部でやや注目されているのがMish関数[3]である。MISH関数は、以下の関数である：

$$f(x) = \tanh(\text{softplus}(x))$$

MISH関数の考え方は、まずはLeaky ReLU[4]を試した後、さらに良い精度を求めてMish関数に置き換えて検証してみるといった考え方が検討られる。その導関数は図-3のようになり、ReLUよりも滑らかに変化し、Swish関数よりもさらにReLUに近い曲線を描くという特徴がある。

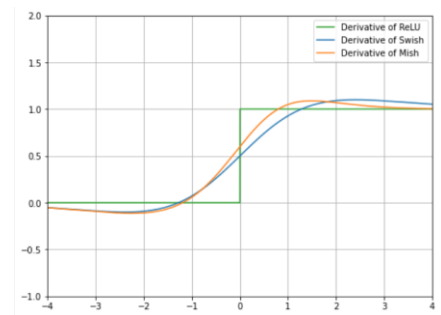


図-3: ReLU, Swish, Mish導関数

Swishは、ReLUとMishの中間に位置する可能性がある。これは、Mishのようなスムーズな機能でありながら、トレーニングに必要なメモリがはるかに少ないためである。Swishは、ほかの活性化関数と比べて、密集目標と小目標に対する特徴の抽出能力を強化し、さらにネットワークの汎化性とロバスト性も強化することができる。密集目標と小目標の検出漏れと精度低下という問題も解決することが可能になる。今回の研究はSwish非線形関数に基づき、軽量化のを実現しているのがhard-swish(h-swish)と呼ばれる新しい活性化関数を導入している。Swishとh-swish関数は、以下の関数である：

$$\text{h-swish}[x] = x \frac{\text{ReLU6}(x + 3)}{6}$$

$$\text{swish}[x] = x \times \text{Sigmoid}(x)$$

活性化関数の調整が改良アルゴリズムに与える

影響を検証するために、特徴融合層の活性化関数 (Leakly Relu) をMish関数に変更し、特徴抽出ネットワークの活性化関数はMishをHard-Swish関数に設定された。

2.3 アテンションモジュール追加

アテンションモジュールを導入したネットワークは、学習時に抽出した特徴情報に注目し、学習に有効であると判断した特徴を強調し、無関係であると判断した特徴を画素単位で抑制することができる。これにより、訓練に必要な特徴のみに焦点を当てることができ、効率の良い学習を行うことが可能になる[5]。

オリジナルYOLOv4では、ネットワーク構造が深くなると、特徴マップに含まれる小さな目標の特徴が弱くなり、特徴情報が大きく失われるため、誤検出や検出漏れなどの問題が発生する。本研究は、scSE[6]というアテンションモジュールを追加することにより、低次元の特徴量を保持することができ、各ピクセルの間に特徴を強調することも可能になった。目標検出でも非常に良い精度の向上を果たすことができる。

2.4 小目標検出の従来手法との比較

精度比較を行うために、従来研究が実施された最近の5年間にかけて各年で小目標検出に対する優れた検出精度を統計した5種類 (FasterRCNN[7], RetinaNet[8], SSD[9], YOLOv4, YOLOv5[10]) の深層学習モデルを選択した。

3. 評価実験

本研究では、画像認識アルゴリズムであるYOLOv4とそれを実装するフレームワークであるPytorchを用いて、小さな目標を検出するモデルの構築を試みた。

3.1 前処理

全体の画像データの90%を訓練データ+検証

データとし、その中に10%をランダムに選択して検証セットとする。残りの画像をテストデータセットとする。今回はCOCOセットで訓練済みの重み (Weight) を用い、エポック数全部100回の学習を行った。100回のうち、まず転移学習を行い、レイヤーやモデルを訓練対象から除外し、50エポックの「freeze (凍結)」訓練を実施した。最後にfreezeしたレイヤーやモデルを再び訓練対象として、50エポック「Unfreeze (解凍, 凍結解除)」訓練を実施した。

3.2 実験結果

目標検出の評価指標には、AP (Average Precision), mAP (Mean Average Precision) が用いられる。mAPは、目標検出の性能を総合的に評価する指標として知られており、深層学習モデルの目標検出性能を比較する際に用いられる。100エポックの中に、検出結果と正解との差を損失関数によりlossとして求め、lossが最小化したエポック数のモデルを採用する。YOLOv4の改善前と改善後の各カテゴリーのAP値を図-4に示す

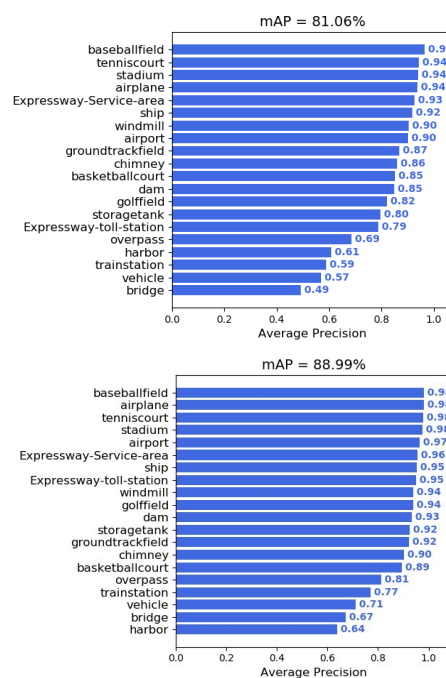


図-4: 改善前と改善後の各カテゴリーのAP値

図-4からわかるように、改良されたYOLOv4のmAP値は約7.93%増加しており、対応する各カテゴリーのAP値もある程度増加している。例えば、

対象としたカテゴリーBaseballfiledのAP値は，アルゴリズムの改善前後であまり変化しなかった，これは画像に占める割合が大きいため，サイズの縮小によって精度が低下しにくいことが主な理由である．小目標のVeichelのAP値が多く増加し，約14%を増加となり．そこで，今回の提案を使うことで小目標の検出精度を向上させることがわかった．精度を比較するため，学習モデルは6つで評価する(表-1)．

表-1: 従来手法との比較

	結果 (mAP)	AP (Veichel)
Faster-RCNN	62.68	13.30
RetinaNet	64.12	18.83
SSD	76.00	38.09
YOLOv5	84.16	50.77
オリジナルYOLOv4	81.06	56.94
提案したYOLOv4	88.99	71.11

4. おわりに

本研究では，YOLOv4を用い，小さな目標の検出を行った．主に特徴検出レイヤー追加やAttentionの追加や活性化関数の修正などを行い，またk-meansによるアンカーボックスの取得，入力解像度調整などを用いることにより，全体のmAPが0.889と高い精度での検出が可能となった．小さな目標の検出の結果として，従来手法と同じく目標の誤検出が発生しない性能を維持しつつ，自動車やタンクなど小目標については未検出数を減少させ，検出精度を向上させることができた(図-5)．また，従来手法との比較を行い，提案手法の有効性を示した．

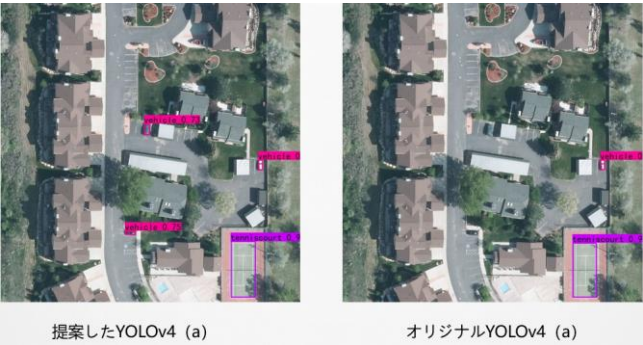


図-5: 改善前と改善後の検出結果一例

今後は別のデータセットを用いて実験を行うことで，提案手法の有効性を検証したいと考えられる．また，本実験で用いたYOLOモデルの一種であるYOLOv4よりさらに改良が加えられたYOLOXの研究も進みたいと考えられる．

参考文献

- [1] 目標検出手法の歴史:YOLOの紹介,
<https://qiita.com/mdo4nt6n/items/68dcda71e90321574a2b>.
- [2] Alexey Bochkovskiy, Chien-YaoWang, Hong-YuanMarkLiao, “YOLOv4:Optimal Speed and Accuracy of Object Detection ” ,
arXivpreprintarXiv:2004. 10934.
- [3] “Mish: A Self Regularized Non-Monotonic Activation Function ” ,
<https://arxiv.org/pdf/1908.08681v3.pdf> .
- [4] Andrew L. Maas, Awni Y. Hannun, Andrew Y. Ng. “Rectifier Nonlinearities Improve Neural Network Acoustic Models” . 2014.
- [5] 平松侑樹, 堀田一弘, “Gated Attention moduleを用いたセマンティックセグメンテーション” ,
PRMU2018-28 SP2018-8.
- [6] Abhijit Guha Roy, Nassir Navab, Christian Wachinger. “Concurrent Spatial and Channel Squeeze & Excitation in Fully Convolutional Networks ” , MICCAI 2018,
<https://arxiv.org/abs/1803.02579>.
- [7] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks ” ,
in Proc. of NIPS, 2015.
- [8] Lin, T., Goyal, P., Girshick, R.B., He, K., & Dollár, P. (2017). “Focal Loss for Dense Object Detection ” , 2017 IEEE ICCV, 2999–3007.
- [9] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). “Ssd: Single shot multibox detector ” , In European conference on computer vision (pp. 21-37).
- [10] 久保 栞, 全 邦釘, 伊藤 克雄, “YOLOv5を用いた導水路トンネルにおけるチョーキング箇所の検出 ” , AI・データサイエンス論文集 2(J2) 87-96 2021年11月.