

YOLOv4 を用いた省メモリなマスク着用検出システム

覃 文標[†] 後藤 祐一[†]

埼玉大学 大学院理工学研究科 数理電子情報系専攻[†]

1. はじめに

新型コロナウイルス感染症 (COVID-19) の蔓延により, 外出時, 特に人と接する機会のある状況と場所においてマスク着用は欠かせないものとなった. さらには, 大学などが集まる場所において, COVID-19 の感染拡大を防ぐために来訪者がマスクを着用しているかどうかを監視し, 未着用な場合に適切な対応をとる必要が生じている. その結果, マスク着用の検出および監視システムが必要とされている.

実時間物体検出システム YOLO [4] を用いたマスク着用検出システムが開発されている. Degadwala ら[2]は YOLO バージョン 4 (YOLOv4) [1]を用いてマスク着用検出システムを構築している. Han ら[3]は, YOLOv4 のバックボーンの特徴抽出のモデルブロックを変更して, 検出精度の改善を試みている. Yu ら[7]は, YOLOv4 のバックボーンにある特徴抽出のための resunit, conv ブロックなどを組直して, 新たな検出モデル CBH-YOLOv4 を構築している. モデル改良した YOLOv4 学習済みモデルは検出速度と精度において改善できる点がある.

一方で, モバイル端末や組み込みコンピュータ上にマスク着用検出システムを搭載するためには, 学習済みモデルの容量は小さい方が良い. たとえば, Jun ら[8]は Mobilenet という軽量な畳み込みニューラルネットワークを用いた検出モデルを構築している. EfficientNet は Mobilenet を基礎に基づいて, Tan ら[6]が開発したモバイル端末用モデルである.

上述のYOLOv4 を用いたマスク着用検出システムは検出速度や学習済みモデルの容量において改善できる点がある. 本研究では, 省メモリ, かつ, 先行研究と同等の検出精度を持つYOLOv4 を用いたマスク着用システムの開発を行う. また, 提案した手法の有用性を実験を用いて検証する.

2. 提案手法

本研究では YOLOv4 を EfficientNet[6]を用いて改良したモデル YOLOv4-EfficientNet を提案

する. 提案モデルでは, MobilenetV2[5]の設計を参考にして, バックボーンを EfficientNet に交換し, 学習済みモデルの容量の削減と検出速度向上を試みる. 他にも, YoloBody には通常 Upsample による特徴画像拡大と Downsample による特徴画像縮小があるが, 本モデルでは Downsample を破棄するとともに Upsample を bottleneck モデルを加える. 図 1 は, 提案モデルの全体図である.

次に, YoloBody で用いられている三段階 CONV モデルブロックと五段階 CONV モデルブロック内に, Mobilenet のモデルの設計を参考に, LinearBottleneck を利用することで両 CONV モデルブロックのモデルの容量の削減を試みた. さらに両 CONV モデルブロックで用いられていた mish 関数を LeakyRelu 関数に変更した (図 3). また, 図 2 中の Yolohead は通常の CONV, BN, Mish から CONV, BN, LeakyRelu に変更した.

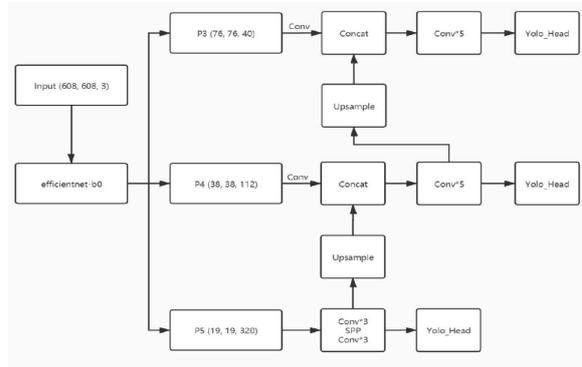


図 1: YOLOv4-EfficientNet

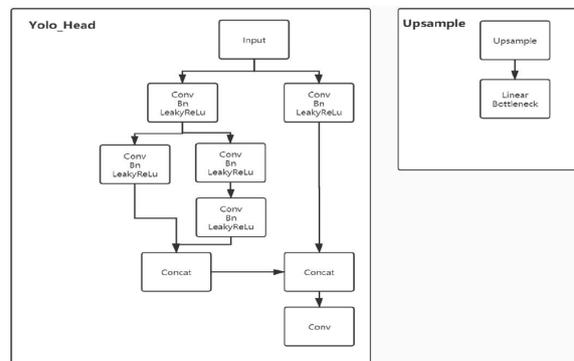


図 2: ヘッド部分の変更箇所

A Memory-Efficient Face Mask Detection System with YOLOv4

Wenbiao Qin and Yuichi Goto, Saitama University.

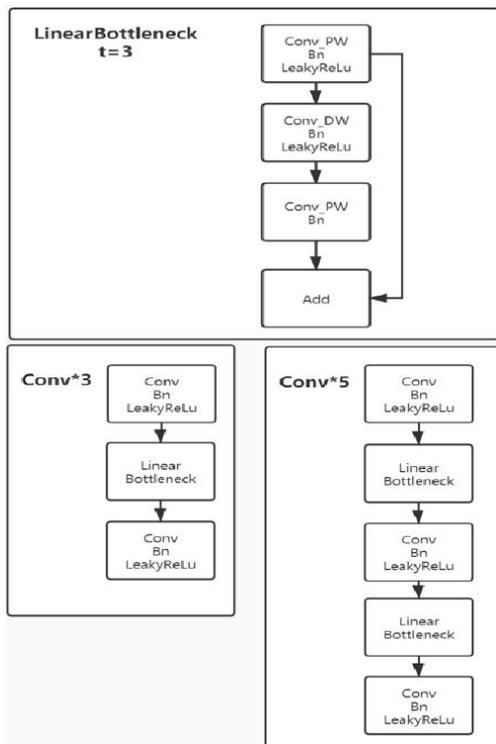


図 3: CONV モデルブロックの変更箇所

3. 実験

3.1 実験方法

実験ではおよそ 4,000 枚の人物画像（マスクなし 2,000 枚，マスクあり 2,000 枚）を用いて，訓練および検出を行った．比較対象としては，YOLOv4[1]，Yu らの開発した CBH-YoloV4[7]，MobilenetV1-YoloV4，MobilenetV2-YoloV4，MobilenetV3-YoloV4 を用いた．

3.2 実験結果

動的検出においては検出精度を表す mAP (mean Average Precision) と検出速度を表す FPS (Frame Per Second) が重要である．そこで，2 つの指標と学習済みモデルの容量を比較する．

表 1: モデル性能比較

	mAP	FPS	Size
YOLOv4	91.1%	8.0	245MB
CBH-YOLOv4	84.2%	12.0	400MB
MobilenetV1-YOLOv4	90.1%	21.0	52MB
MobilenetV2-YOLOv4	89.6%	17.0	47MB
MobilenetV3-YOLOv4	89.8%	15.0	55MB
YOLOv4-EfficientNet	89.1%	17.0	30MB

表 1 は実験結果 である．表 1 から，YOLOv4-EfficientNet は他のモデルと同等の mAP と

FPS の値であるにも関わらず学習済みモデルの容量は 30MB と最も少ない値を示している．このことから，YOLOv4-EfficientNet はマスク着用検出において有望であると考ええる．

4. おわりに

本研究で提案した YOLOv4-EfficientNet は Efficient の使用に加え，Mobilenet のブロックモデルを使用したモデルである．実験結果より，YOLOv4-EfficientNet の学習済みモデル容量は少なく，かつ，実行時には高速に動作し，その検出精度も他の手法と同等である．この利点はスマートフォンやウェブなどに実装することで便利に使用できると期待できる．

今後の研究方針は，モデル削減モデル，例えば pruning, GAN ネットの利用を考えている．

参考文献

- [1] A. Bochkovskiy et al.: YOLOv4: Optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934, April 2020.
- [2] S. Degadwala et al.: Yolo-v4 Deep Learning Model for Medical Face Mask Detection, Proc. ICAIS2021, pp. 209-213, March 2021.
- [3] W. Han et al.: A Mask Detection Method for Shoppers Under the Threat of COVID-19 Coronavirus, Proc. CVIDL2020, pp. 442 - 447, 2020.
- [4] J. Redmon et al.: You only look once: Unified, real-time object detection. Proc. CVPR2016, pp. 779-788, 2016.
- [5] M. Sandler et al.: MobileNetV2: Inverted Residuals and Linear Bottlenecks, Proc. CVPR, pp. 4510-4520, June 2018.
- [6] M. Tan et al.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, Proc. ICML 2019, pp. 6105 - 6114, Jun 2019.
- [7] J. Yu et al.: Face Mask Wearing Detection Algorithm Based on Improved YOLOv4, Sensors 21(9), e3263, April 2021.
- [8] Z. Jun et al.: Small Sample Underwater Target Recognition Based on Mobilenet YOLOV4 Algorithm, CONVERTER, pp. 359 - 372, July 2021.