

## 1. はじめに

物体検出を組み込みデバイスで用いるには、高精度かつ高速度なモデルが必要である。Grishick らは深層学習を用いた Two-stage による物体検出モデルである R-CNN[6] を、Redmon らは One-stage の物体検出モデルである YOLO[5] を提案した。さらに、Alexey らは YOLO の特徴抽出層に、計算量を大幅削減できる CSPNet 構造と、解像度が異なる空間的な特徴量を集約する PANet 構造を用いた YOLOv4[1] を提案した。これらの物体検出モデルのネットワーク構造は複雑かつ大規模であるため、組み込みデバイスに実装することは難しい。そこで、モデルを軽量化する必要があり、精度を維持しつつモデル圧縮が可能な枝刈り手法が目ざされている。

Frankle らは枝刈り手法として、宝くじ仮説 [4] を提案した。しかし、宝くじ仮説により枝刈りしたモデルは精度が低下することが指摘されている [7]。そこで、本研究では、大規模なネットワーク出力をもとに小規模なネットワークを学習する知識蒸留と宝くじ仮説を組み合わせることで、宝くじ仮説より高精度となる学習法を提案する。

## 2. 関連研究

モデルの軽量化には、知識蒸留と枝刈りの 2 つのアプローチが用いられる。

### 2.1 知識蒸留

物体認識の知識蒸留として、Hinton らは大規模なネットワーク (教師モデル) を学習した後、教師モデルの出力情報を Soft Target として小規模なネットワーク (生徒モデル) を学習する手法 [3] を提案した。物体検出の知識蒸留として Chen らは Two-Stage の Faster R-CNN の特徴抽出層、分類損失、回帰損失を同時に知識蒸留する手法 [2] を提案した。知識蒸留を One-Stage 物体検出モデルに適用する際、教師モデルの出力に誤検出した検出結果や物体を重複して検出したか検出の結果も含まれるため、生徒モデル精度低下につながる。

### 2.2 枝刈り

深層学習モデルには、推論結果に影響しないパラメータが存在する。モデルを圧縮するには、このようなパラメータを枝刈りすることが重要である。Zhuang[8] らは BN 層の  $\gamma$  係数を L1 正則化によるスパース制約項として損失関数に追加し、出力が 0 に近い重要でないニューロンを枝刈りする手法を提案した。しかし、この手法は枝刈りに時間を要する。

Frankle ら [4] は Iterative Magnitude Pruning (IMP) 方法を用いて大規模なネットワークを枝刈りし、元のネットワークの精度と近くなる小規模なネットワーク (サブネットワーク) を見つける宝くじ仮説 [4] を提案した。この方法は、IMP 方法によって優れたスパースサブネットワークを見つけることを目的としている。IMP 法はパラメータの絶対値の大きさを、重要度を計算し、枝刈りを行う手法である。一方で、宝くじ仮説により枝刈りしたモデルはパラメータの削除率に比例して精度が低下することが指摘されている [7]。

## 3. 提案手法

本研究では、宝くじ仮説に知識蒸留を組み合わせ、精度低下を抑制する手法を提案する。

### 3.1 知識蒸留付き宝くじ仮説学習

本研究の流れを図 1 に示す。まず Frankle らが提案した宝くじ仮説により物体検出モデルを枝刈りする。次に、枝刈りする前の大規模なモデルを教師モデルとし、教師モデルから蒸留した知識を用いて枝刈りした小規模なモデルを学習する。これらの流れの詳細を以下に示す。

Step1. 物体検出モデルのパラメータを初期化する。

Step2.  $k$  エポック学習したパラメータ  $w_k$  を保存する。

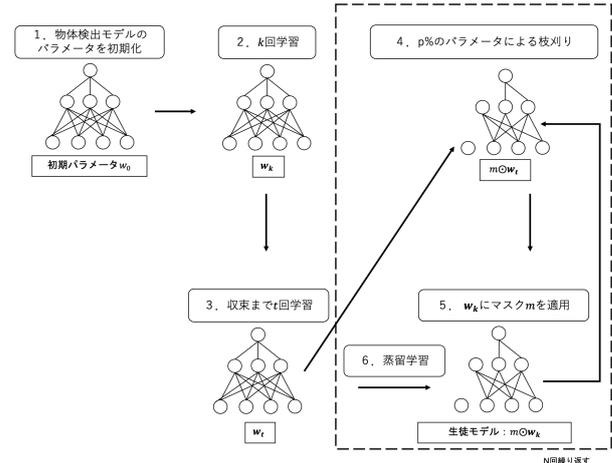


図 1: 提案手法の学習流れ

Step3. 元ネットワークを収束するまで学習し、重みを  $w_t$  を保存する。

Step4.  $w_t$  中で各値の絶対値が低いものから  $p\%$  の重みを枝刈りする。重みマスク  $m$  を保存する。

Step5.  $bmw_k$  にマスク  $m$  を適用し、生徒モデルとする。

Step6. Step3. で保存した  $w_t$  をもつ教師モデルとし、生徒モデルが収束するまで蒸留学習をする。

目標パラメータに達成するまで、Step4. から Step6. まで繰り返す。

### 3.2 損失関数

知識蒸留を物体検出に用いる場合、誤検出した検出した結果も蒸留してしまう。本研究では教師モデルの検出結果に信頼度 (Confidence) を利用して、蒸留の損失関数を設計する。誤検出の信頼度は目標物体の信頼度出力より小さいことが多いため、教師モデルと生徒モデルの分類 (Classification) の損失計算式の係数とする。これにより、蒸留した知識から誤検出の影響を抑制できる。損失関数を式 (1) に示す。

$$L_{loss} = L_{hard}(s, G) + \alpha L_{soft}(s, t) \quad (1)$$

$s$  は生徒モデル  $m \odot w_k$  の出力、 $G$  は正解ラベル、 $t$  は教師モデル  $w_t$  の出力、 $L_{hard}$  は生徒モデル出力と正解ラベルの損失、 $L_{soft}$  は生徒モデル出力と教師モデル出力の損失である。 $L_{hard}$  と  $L_{soft}$  を混合する割合をパラメータ  $\alpha$  によって調整する。本研究では  $\alpha$  を 1 とする。 $L_{hard}$  及び  $L_{soft}$  は以下の式となる。

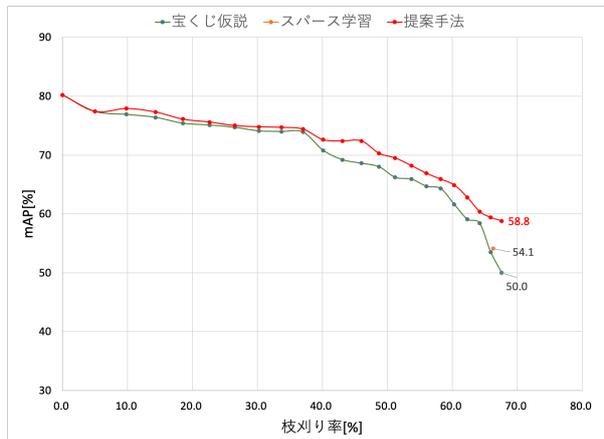
$$L_{hard}(s, G) = L_{conf}(s, G) + L_{cls}(s, G) + L_{reg}(s, G) \quad (2)$$

$$L_{soft}(s, t) = L_{conf_{kd}}(s, t) + conf_t L_{cls_{kd}}(s, t) \quad (3)$$

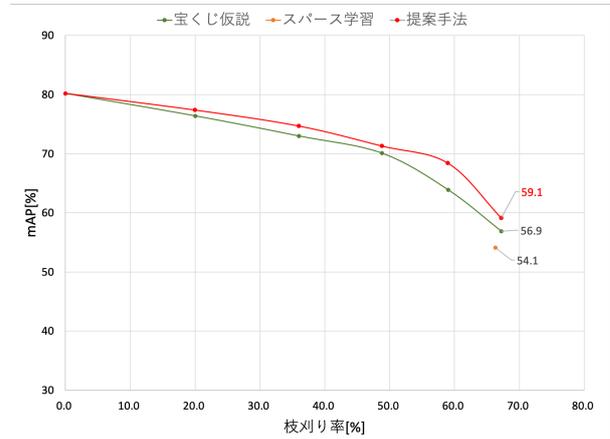
$L_{conf}$  は生徒モデルと正解ラベルの信頼度損失、 $L_{cls}$  は生徒モデルと正解ラベルの分類損失、 $L_{reg}$  は生徒モデルと正解ラベルのバウンディングボックス回帰の損失である。 $L_{conf_{kd}}$  は生徒モデルと教師モデルの信頼度損失、 $L_{cls_{kd}}$  は生徒モデルと教師モデルの分類損失である。また、 $conf_t$  は教師モデルの信頼度出力である。ここで  $L_{conf_{kd}}$ 、 $L_{cls_{kd}}$  は全て平均二乗誤差 (MSELoss) で算出する。

### 4. 評価実験

提案手法の有効性を確認するため、VOC データセットを用いて従来の宝くじ仮説による学習と比較実験を行う。



(a) 1 回の枝刈り率を 5%とした時



(b) 1 回の枝刈り率を 20%とした時

図 2: mAP 変化の比較

表 1: モデルのパラメータ数と mAP の比較

手法	枝刈り回数 (枝刈り率/回)	残り パラメータ数	mAP [%]
YOLOv4		$6.404 \times 10^6$	80.4
スパース学習	1 (67%)	$2.152 \times 10^6$	54.1
宝くじ仮説	22 (5%)	$2.088 \times 10^6$	50.0
提案手法	(5%)		<b>58.8</b>
宝くじ仮説	5 (20%)	$2.101 \times 10^6$	56.9
提案手法	(20%)		<b>59.1</b>

#### 4.1 実験概要

VOC2007 及び VOC2012 を合わせて本実験の学習用データセットとして使用する。物体検出モデルにはYOLOv4[1]を用いる。学習用及びテスト用データセットの画像枚数はそれぞれ 16,551 枚と 4,9952 枚である。精度の評価指標には、mean Average Precision(mAP)を用いる。

#### 4.2 定量的評価

1 回の枝刈り率を 5%または 20%にして、枝刈りを繰り返した時の mAP を図 2 に示す。1 回の枝刈り率を 5%にした時、提案手法及び宝くじ仮説と共に枝刈り率が 40%までは精度低下が緩やかである。しかし、40%を超えると宝くじ仮説は大きく精度が低下する。一方で提案手法は宝くじ仮説よりも精度低下を抑制することができている。1 回に枝刈り率を 20%にした時も同様に提案手法の精度低下は宝くじ仮説より緩やかである。また、表 1 より提案手法のモデルは Zhuang らの手法よりパラメータ削減でき、精度が向上したことが分かる。この結果から、本提案手法で用いる知識蒸留は精度向上に貢献したといえる。

#### 4.3 定性的評価

図 3 に検出結果例を示す。図 3 の 1 行目より、提案手法は教師モデルに近い位置精度であることが確認できる。図 3 の 2 行目より、宝くじ仮説の場合、1 つの犬を重複して検出している。一方、提案手法のモデルは 1 つの物体として検出している。図 3 の 3 行目より、宝くじ仮説では誤検出しているケースに対して提案手法は正確に検出していることが確認できる。

#### 5. おわりに

本研究では、知識蒸留と宝くじ仮説手法を用いた One-stage 物体検出モデルの圧縮学習法を提案した。提案手法を用いて学習したモデルは、従来手法により同程度枝刈りしたモデルと比べ、精度低下が緩やかであり、モデルの精度を向上させることができた。

今後は、他のデータセットによる検証と異なる枝刈り率で枝刈りしたモデルの間の比較実験を行う予定である。



図 3: 出力結果の比較

#### 参考文献

- [1] A.Bochkovskiy *et al.* Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.
- [2] G.Chen *et al.* Learning efficient object detection models with knowledge distillation. *NeurIPS*, 2017.
- [3] G.Hinton *et al.* Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
- [4] J.Frankle *et al.* The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.
- [5] J.Redmon *et al.* You only look once: Unified, real-time object detection. In *CVPR*, 2016.
- [6] R.Girshick *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.
- [7] X.Ma *et al.* Sanity checks for lottery tickets: Does your winning ticket really win the jackpot? *NeurIPS*, 2021.
- [8] Z.Liu *et al.* Learning efficient convolutional networks through network slimming. In *ICCV*, 2017.

#### 研究業績

- [1] L.Liu, *et al.*, "CLASS-WISE FM-NMS FOR KNOWLEDGE DISTILLATION OF OBJECT DETECTION", IEEE ICIP, 2022.

(他 1 件)