

1. はじめに

安全なモビリティの実現に向けて自動運転システムの開発が加速しており、自動運転車の実用化が近づきつつある。自動運転システムには周囲の物体を正確に検出する技術が必要である。物体検出は、カメラや LiDAR 等のセンサで取得した情報から車両周辺の物体の位置とその種類を推定する技術であり、これまでに数多くの物体検出モデルが提案されている。

高精度な物体検出モデルを構築するには、大量の教師ありデータを用いた学習が必要である。大量の画像データに対して正解となるラベルを手作業で付与するアノテーションには膨大なコストがかかる。また、各クラスのサンプル数に偏りがあることも大きな課題である。特に車載カメラで取得したデータセットの場合、車や歩行者のサンプルは多く含まれるが、電車やオートバイ等のサンプルは不足する傾向がある。画像に写る物体の大きさに着目すると、遠方の小さな物体数は多く、近方の大きな物体は少なくなる傾向があり、距離に対してサンプル数の偏りが生じる。そこで、本研究では、少量の教師ありデータと大量の教師なしデータを組み合わせた半教師あり学習において、クラスおよび大きさごとのサンプル数の偏りを考慮した物体検出法を提案する。提案手法は、クラスおよび大きさごとのサンプル数に基づいて重み付けを行うことで、半教師あり物体検出手法の精度向上を目指す。

2. 関連研究

本章では、半教師あり物体検出手法である Unbiased Teacher v2 について述べる。Unbiased Teacher v2[1] は、Teacher と Student の2つのモデルを用いて学習を行う。本手法の構造を図1に示す。Teacher の学習ステージを Burn-in Stage (B-Stage)、Student の学習ステージを Teacher-Student Mutual Learning Stage (TSML-Stage) と呼ぶ。Burn-in Stage では、教師ありデータのみを用いて Teacher を学習する。

そして、学習済みの Teacher を用いて教師なしデータに対して疑似ラベルを付与する。TSML-Stage では、教師ありデータと疑似ラベルを付与した教師なしデータおよび疑似ラベルを付与していない教師なしデータを用いて Student を学習する。Student の重み更新後、指数移動平均 (EMA) を用いて Student の重みを基に Teacher の重みを更新する。

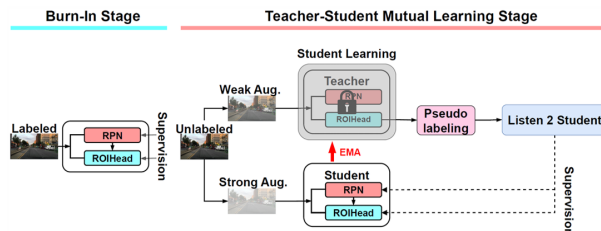


図 1: Unbiased Teacher v2 のネットワーク構造

3. 車載データセットの問題

車載カメラで収集した代表的なデータセットとして、BDD 100K データセット [2] がある。クラスは、Pedestrian や Car といった走行シーンに現れる 9 クラスが定義されている。BDD100K データセットの学習データのクラスごとのサンプル数を図2に示す。図2より、Car クラスのサンプルは多く含まれているが、Motorcycle や Rider, Bicycle といったクラスは少なく、クラス間にサンプル数の偏りがあることがわかる。

また、各クラスの物体の大きさに着目した際のサンプル数を図3に示す。各グラフの横軸は物体矩形の面積、縦軸はサンプル数を表している。多くのクラスで矩形の面積が小さいサンプル数が多く、矩形の面積が大きい近距離のサ

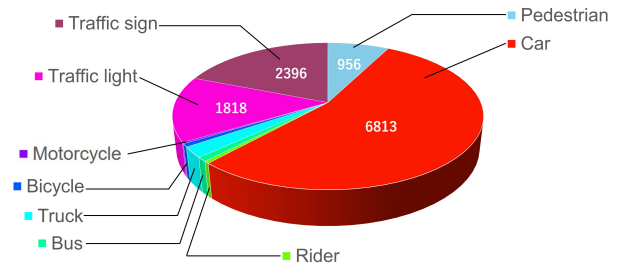


図 2: BDD100K に含まれる各クラスのサンプル数

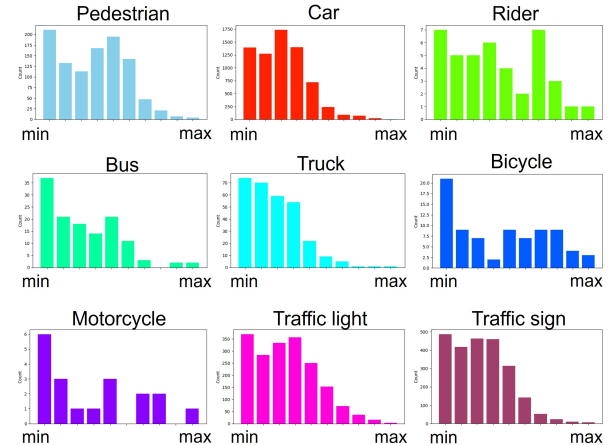


図 3: 矩形の面積に着目した各クラスのサンプル数

ンプル数が減少する。以上より、矩形の面積によりサンプル数に偏りがあることがわかる。

4. 提案手法

データセットには、クラスおよび物体の大きさごとにサンプル数の偏りがある。本研究では、半教師あり物体検出手法に対し、これら2つの偏りを考慮して、損失関数に重み付けを行い、モデルを学習する。

4.1. 損失関数

クラスおよび大きさによるサンプル数の偏りに対応して学習するために、本研究では損失関数である Focal Loss に重み W_t を導入する。重み W_t を用いた Focal loss の損失関数を式 (1) に示す。

$$FL(p_t) = -W_t \cdot ((1 - p_t)^\gamma \log(p_t)) \quad (1)$$

ここで、 p_t は正解クラスに対するクラス確率、 γ はハイパーパラメータである。 W_t は式 (2) のように定義する。

$$W_t = \frac{S_{min}}{S_t} \cdot A_t \quad (2)$$

$\frac{S_{min}}{S_t}$ によって、各クラスのサンプル数の偏りに応じた重み付けを行う。ここで、 S_t はクラス t のサンプル数、 S_{min} は最もサンプル数が少ないクラスのサンプル数である。これより、 $\frac{S_{min}}{S_t}$ はサンプル数が最も少ないクラスで 1.0 となり、サンプル数が多いほど小さくなる。

次に、 A_t によって、物体の大きさによるサンプル数の偏りに応じた重み付けを行う。各クラスのサンプルに対して矩形の面積を算出し、その最大値と最小値の区間 K_t を獲得し、 K_t を等間隔に 10 個の区間に分割する。そして、分割された各区間のサンプル数を基準に降順にソートする ($K_t \in \{k_t^1, k_t^2, \dots, k_t^{10}\}$)。次に、ソートした K_t の各区間それぞれにペアとなる係数 A_t を設定する。ここで、3 章より、矩形の面積が大きいほどサンプル数が少ない傾向がある。そのため、 $[1.0, 2.0]$ を K_t と同様に等間隔に 10 個に分割し、 K_t の各区間に割り当てて係数 ($A_t \in \{a_t^1, a_t^2, \dots, a_t^{10}\}$)

表 1: 半教師あり学習における精度比較

教師ありデータの割合	提案手法	AP	Pedestrian	Car	Rider	Bus	Truck	Bike	Motor Cycle	Traffic light	Traffic sign
1%		21.21	24.42	42.90	13.40	31.94	28.76	14.65	6.24	19.21	30.62
	✓	21.20	24.76	43.02	13.75	31.74	29.35	14.73	5.26	19.06	30.42
5%		25.03	26.84	45.39	18.88	38.15	35.00	17.53	15.20	20.82	32.52
	✓	25.21	26.82	45.22	19.08	38.38	35.75	17.93	15.50	20.91	32.47
10%		26.95	28.94	46.61	21.01	42.07	38.61	20.29	16.33	21.60	34.03
	✓	26.97	28.99	46.66	21.02	41.61	38.89	20.23	16.56	21.68	34.04

を設定する。これによって、矩形の面積が大きいほど係数の値が大きくなり、面積が小さいほど係数の値が小さくなる。

これら2つの係数を組み合わせることで、クラスおよび矩形の大きさに対するサンプル数の偏りを考慮した学習が可能となる。

4.2. モデルの学習

Unbiased Teacher v2は、教師あり学習を行う B-Stage と、疑似ラベルを用いて学習を行う TSML-Stage が存在する。本研究では、この両者に提案する損失関数を適用する。B-Stage では、教師データを用いて S_{min} と A_t を導出し、重み付けを行う。TSML-Stage では、生成された疑似ラベルを用いて損失関数に重み付けを行う。各 Epoch 終了時に、学習データに付与された疑似ラベルを用いて係数 A_t を更新する。動的に係数 A_t を更新することで、物体の大きさに適応的な学習が可能となる。一方で、 S_{min} は更新しない。これによって、初期状態の学習が安定し、効率的な学習が可能となる。

5. 評価実験

提案手法の有効性を検証するために、重み付けの有無による精度を比較する。半教師あり学習モデルとしては Unbiased Teacher v2、物体検出器には Faster R-CNN を用いる。データセットには、BDD100K データセットを用い、データセットの中から教師ありデータとして全データの1%、5%、10%を抽出し、残りのデータを教師なしデータと用いる。

評価には Average Precision (AP) を評価指標とし、矩形の大きさに応じた AP も算出する。矩形が 32×32 ピクセル未満を AP_S 、矩形が 96×96 ピクセル未満を AP_M 、 96×96 ピクセル以上を AP_L とする。ここで、学習データにおいて矩形が 32×32 ピクセル未満のデータは 46.42%、 96×96 ピクセル未満のデータは 33.29%、 96×96 ピクセル以上のデータは 20.29% である。

5.1. 定量的評価

提案する重み付けの有無の検出精度の比較を表1に示す。提案手法を導入した場合、教師ありデータの割合に関わらず、Rider や Truck の精度が従来手法よりも向上した。3章より、Rider や Truck のサンプル数が少なかったため、クラスごとの偏りに応じた重み付けを導入し、少数の物体に対する学習を強化したと考える。

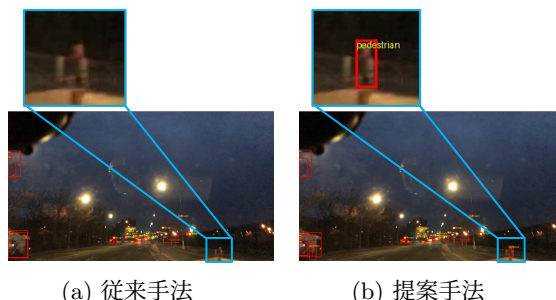
物体の大きさに着目した検出精度の比較を表2に示す。物体の大きさに着目した重み付けを導入したことにより、大きい物体における精度が教師ありデータの割合に関わらず向上した。また、教師ありデータの割合が5%以上の場合で、物体の全大きさにおいて提案手法を導入することで精度の向上が見られた。3章より、物体が大きいほどサンプル数が少ないため、矩形の大きさに応じた重み付けを導入し、偏りを軽減できたことを確認した。

5.2. 定性的評価

教師ありデータの割合が1%の場合の定性的評価を図4に示す。図4(a)の手前の歩行者は未検出であるが、図4(b)の矩形面積が小さい歩行者を検出できている。また、教師ありデータの割合が10%の場合を図5に示す。図5(b)より影で隠れているかつ小さい物体に対しても検出できている。

表 2: 物体の大きさにおける検出精度の比較

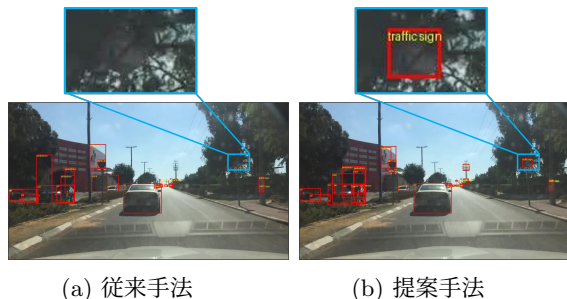
教師ありデータの割合	提案手法	AP	AP_S	AP_M	AP_L
1%		21.21	8.96	26.39	37.72
	✓	21.20	8.98	25.90	38.80
5%		25.03	10.12	29.37	45.66
	✓	25.21	10.51	29.65	46.00
10%		26.95	11.70	31.17	48.30
	✓	26.97	11.74	31.48	48.89



(a) 従来手法

(b) 提案手法

図 4: 教師ありデータの割合が1%の可視化結果



(a) 従来手法

(b) 提案手法

図 5: 教師ありデータの割合が10%の可視化結果

6. おわりに

本研究では、車載映像で構成されているデータセットで生じるクラスおよび大きさによるサンプル数の偏りに対応するために重み付けを導入した損失関数による半教師あり物体検出手法を提案した。評価実験から、提案手法を用いることで、クラスおよび大きさに対するサンプル数の偏りを考慮した学習が可能となり、全体的な精度の向上が確認できた。

今後の課題として、教師ありデータの割合が極端に少ない場合に適した提案手法の改善が挙げられる。

参考文献

- [1] Tian, *et al.*, “Unbiased Teacher v2: Semi-Supervised Object Detection for Anchor-Free and Anchor-Based Detectors”, CVPR, 2022.
- [2] F.Yu, *et al.*, “BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning”, CVPR, 2020.

研究業績

- [1] 鈴木陽太郎 等, “歩行者が交差点に存在するシーンにおける物体検出モデルの精度評価”, 自動車技術会秋季大会, 2024.