

1. はじめに

物体検出は、画像中の物体の位置検出とそのクラスを識別する技術である。自動運転システムには、周辺環境を把握するために物体検出が必要不可欠である。物体検出モデルを学習するには、大量の教師ありデータを必要とする。このような大量データのアノテーションは、コストが高い。そのため、教師なしデータが活用できる半教師あり学習を物体検出に用いる研究が注目されている。従来研究では、各クラスのサンプル数が均等となるように整備されたデータセットを対象とする。車載映像を対象とした場合、オートバイや自転車といった特定のクラスのデータが十分に収集できず、サンプル数に偏りが生じやすい。このようなクラスインバランスなデータセットに対する半教師あり学習の有効性は未だに示されていない。そこで本研究では、車載映像で構成されるクラスインバランスなデータセットを対象とした際の物体検出における半教師あり学習の有効性を確認する。また、データセットのクラスインバランス問題に対応するために、クラス分類損失に対してクラスごとの重みを追加する。さらに、サンプル数とデータの多様性を拡張するために Copy-Paste Augmentation[1] を学習に採用する。

2. 関連研究

本節では、半教師あり物体検出手法である Unbiased Teacher v2 と Anchor Free の物体検出モデルである Fully Convolutional One-stage Object Detector (FCOS)、車載カメラ映像から構成されるデータセット BDD100K について述べる。

2.1. Unbiased Teacher v2

Unbiased Teacher v2[2] は、Teacher と Student の2つのモデルを用意し、それぞれ分けて学習を行う。Teacher の学習ステージでは、教師ありデータのみを用いて学習する。Student の学習ステージでは、学習済みの Teacher を用いて教師なしデータに疑似ラベルを付与し、教師ありデータと疑似ラベルを用いて Student を学習する。さらに Student の重み更新後、指数移動平均を用いて Student の重みを基に Teacher の重みを更新することで、学習した特徴を Teacher に対してフィードバックする。また、一貫性学習の考えに基づき Unbiased Teacher v2 では、Teacher に入力するデータには弱いデータ拡張、Student に入力するデータには強いデータ拡張を施す。Listen 2 Student 機構では、Teacher の予測スコアより Student の予測スコアが高い場合、Student の重みの更新を行わない。そのため、誤検出した疑似ラベルによる精度低下を抑えられる。

2.2. FCOS

FCOS[3] は、Anchor boxes を必要としない畳み込み層のみで構成する Anchor Free の物体検出モデルである。Anchor Free にすることで、ポジティブサンプルとネガティブサンプルのインバランスの改善、Anchor boxes のハイパーパラメータの削減、アスペクト比が大きく変化する物体の検出精度の向上を実現している。また、従来の Anchor Free 手法では、Bounding box からの距離を基に損失計算をするため、物体の中心から離れた位置の点もポジティブサンプルとして扱われる問題がある。そこで FCOS では、物体の中心からの左右上下の距離を基に損失設計を行うことで、この問題を回避している。

2.3. BDD100K データセット

BDD100K データセットは、車載カメラで撮影したデータで構成されたデータセットである。Train データが 70,000 枚、Validation データが 10,000 枚、Test データが 20,000 枚で構成されている。クラスとして、Person、Car といった走行に関連する 9 クラスが定義されている。BDD100K の Train データにおけるクラスごとのサンプル数を図 1 に示す。図 1 より、BDD100K データセットには Car クラス

のサンプルは多数含まれているが、Motor、Rider、Bike クラスは少数であり、クラスインバランスの状態であることがわかる。

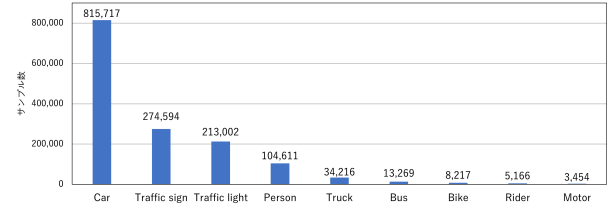


図 1: BDD100K に含まれる各クラスのサンプル数

3. 提案手法

提案手法の学習過程を図 2 に示す。提案手法では、クラスインバランス問題の対策として 2 つのアプローチを Unbiased Teacher v2 に導入する。1 つ目に、クラス損失関数である Focal Loss[4] にクラスごとの重みを導入する。Focal Loss は、物体検出タスクにおける前景領域と背景領域の不均衡が原因による精度低下に対する改善策として提案された損失関数である。この Focal Loss は、前景領域と背景領域に注目した手法のため、オブジェクトクラスのインバランスを考慮していない。そこで提案手法では、Focal Loss にクラスごとに設定した重みを適用することで、クラスインバランス問題の改善を図る。また 2 つ目のアプローチとして、弱いデータ拡張に Copy-Paste Augmentation を採用する。このとき、サンプル数の少ないクラスを優先的に利用する。サンプル数の少ないクラスの学習する機会を増やすことで、クラスインバランス問題の改善を図る。

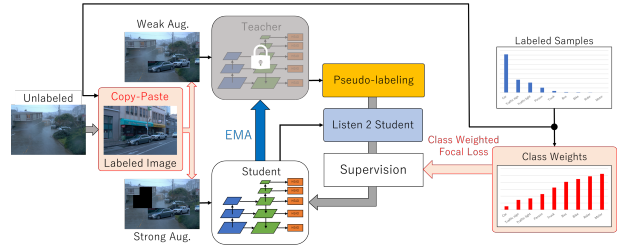


図 2: 提案手法を導入した Unbiased Teacher v2

3.1. Class Weighted Focal Loss

本研究では、Focal Loss にクラスごとに重みを加えた損失関数を、Class Weighted Focal Loss(CWFL) として式 (1) のように定義する。

$$\mathcal{L}^{CW-FL}(p_t) = -\log\left(\frac{C_{all}}{C_t}\right) \left((1-p_t)^\gamma \log(p_t)\right) \quad (1)$$

ここで p_t をクラス確率、 γ をハイパーパラメータ、 C_{all} は学習用データに含まれる全クラスのサンプルの総数、 C_t は学習用データに含まれる予測クラスのサンプルの総数とする。クラス損失に対して、サンプル数を考慮した重みを付加することでサンプル数の少ないクラスに注目した学習が可能となる。

3.2. Copy-Paste Augmentation

Copy-Paste Augmentation[1] は、画像中に存在する物体のマスク部分を切り取り、他の画像に貼り付けることで画像中のサンプル数を増加させるデータ拡張手法である。Copy-Paste Augmentation を用いることで、学習サンプルとデータの多様性を拡張することができる。しかし、物体検出に应用する場合には教師データにマスク情報が存在しないため、本研究ではマスク情報の代わりに Bounding box 情報を用いる。また、各クラスのサンプル数をもとに、サンプル数の少ないクラスを優先的に貼り付ける。

表 1: 教師あり学習と半教師あり学習の精度比較 [%]

教師ありデータの割合	Method	AP _{50:95}	Car	Traffic sign	Traffic light	Person	Truck	Bus	Bike	Rider	Motor
1%	Supervised	13.37	34.98	17.91	12.28	14.72	14.97	14.13	5.41	4.66	1.29
	Semi-Supervised	21.23	39.77	27.15	17.88	21.57	26.30	29.73	10.18	9.54	8.94
5%	Supervised	20.93	39.06	24.81	14.50	21.26	27.08	29.16	13.11	11.04	8.38
	Semi-Supervised	25.89	42.07	29.53	20.01	25.73	35.63	38.31	15.43	13.02	13.28
10%	Supervised	23.33	40.54	26.26	14.69	22.81	31.52	34.02	14.24	13.55	12.38
	Semi-Supervised	28.52	44.65	34.01	22.03	28.48	39.08	41.29	17.40	15.87	13.85

表 2: 提案手法の導入前と導入後の精度比較 [%]

教師ありデータの割合	Method	AP _{50:95}	Car	Traffic sign	Traffic light	Person	Truck	Bus	Bike	Rider	Motor
1%	Default	21.23	39.77	27.15	17.88	21.57	26.30	29.73	10.18	9.54	8.94
	CWFL	19.54	39.48	25.80	14.92	21.42	24.81	24.82	11.15	5.62	7.86
	Copy-Paste	21.70	39.58	27.38	17.99	21.43	28.76	31.63	11.47	8.07	8.99
	Proposed	20.95	38.44	24.38	16.86	22.07	26.15	27.57	13.37	10.49	9.16
5%	Default	25.89	42.07	29.53	20.01	25.73	35.63	38.31	15.43	13.02	13.28
	CWFL	26.63	42.73	31.31	20.65	26.27	34.65	38.49	17.30	15.17	13.12
	Copy-Paste	26.01	42.59	31.01	17.88	25.86	36.47	39.06	15.12	12.61	13.52
	Proposed	27.11	43.14	32.06	21.00	26.51	36.30	38.60	17.37	14.30	14.83
10%	Default	28.52	44.65	34.01	22.03	28.48	39.08	41.29	17.40	15.87	13.85
	CWFL	28.8	44.82	33.38	19.99	28.76	38.41	41.47	19.13	16.57	16.65
	Copy-Paste	28.58	45.23	34.00	22.15	28.56	40.40	41.98	17.16	14.53	13.24
	Proposed	29.00	45.27	32.73	19.86	28.80	39.56	41.78	18.99	16.71	17.25

4. 評価実験

評価実験では、半教師あり学習モデルとして Unbiased Teacher v2, 検出器には FCOS, バックボーンには事前学習済みの ResNet-50 を用いる。データセットには, BDD100K を用いる。データ全体の 1%, 5%, 10% を教師あり集合としてランダムに抽出し, 残りのデータをラベルなしデータとする。学習回数を Teacher は 30,000 iteration, Student は 150,000 iteration とし, バッチサイズを教師ありデータは 8, 教師なしデータは 8 として学習する。また, 評価指標には mean Average Precision (mAP) を用いる。

4.1. 実験結果

教師あり学習と半教師あり学習による物体検出精度を表 1 に示す。全ての条件において半教師あり学習の方が高い精度となっており, 車載カメラ映像を対象とする場合にも半教師あり学習は有効であることがわかる。特に, 教師ありデータが少ない程精ほど差が大きくなることから, 教師ありデータが少ない場合及び教師なしデータが大量に使用できる場合に半教師あり学習の効果が大きいことがわかる。

提案手法導入前と導入後の比較を表 2 に示す。ここで, Default は従来の Unbiased Teacher v2, CWFL は Class Weighted Focal Loss のみを取り入れた場合, Copy-Paste は Copy-Paste Augmentation のみを取り入れた場合, Proposed はどちらも取り入れた場合である。全体精度の比較から, 教師ありデータの割合が 1% の場合を除いて提案手法が最も高精度となっている。1% の場合に CWFL を用いることで精度が低下しているのは, 教師ありデータに含まれるサンプル数に偏りが生まれやすく, サンプルの少ないクラスの重みが大きくなりすぎるためだと考えられる。また, クラスごとの精度の比較では CWFL を取り入れることでサンプル数が少ないクラスである Rider, Motor, Bike の精度が向上しているため, クラスインバランスの改善が確認できる。サンプル数が多いクラスである Car や Pedestrian の精度も向上していることから, 小さな重みを付加することで過学習の抑制ができています。しかし, Bus, Truck, Traffic light, Traffic sign といったサンプル数が中間に位置するクラスは, 精度が低下する傾向がある。これは, クラス重みの影響を強く受けるサンプル数が少ないクラスが重視されることで, 重みの影響をほとんど受けないクラスの学習が抑制されていると考えられる。以上のことから, サンプル数の多いクラスとサンプル数の少ないクラスに対して提案手法は有効であるといえる。

4.2. 定性的評価

各手法による検出結果例を図 3 に示す。画像の左側に駐車されている Motor に注目すると, Default では Person と誤認識し, CWFL では Car と誤認識している。一方で, Copy-Paste と Proposed では, Motor と正しく認識している。このことから, Copy-Paste Augmentation を取り入れることで Motor クラスの精度向上が確認できる。また, 画像右上に写る Traffic light に注目すると Default では 2 つ写る Traffic light を 1 つが未検出である。一方, 提



図 3: 各手法による検出結果例

案手法を導入することで 2 つとも正しく検出している。

5. おわりに

本研究では, 車載映像で構成されるクラスインバランスなデータセットを対象とした物体検出における半教師あり学習の有効性の検証と提案手法の評価実験, 定性的評価を行った。評価実験より, 提案手法を用いることでクラスインバランスを考慮した学習が可能になり, 全体的に精度が向上することを確認できた。しかし, 教師ありデータが極端に少ない場合には, クラス間のサンプル数の差が大きくなりすぎることで, クラス間の重みの差が大きくなり学習を抑制してしまう結果となった。定性的評価より, 提案手法と Copy-Paste Augmentation によるクラスインバランスを考慮が確認できた。今後の課題として, CWFL の重みの上限値を設けることで重みの差が大きくなりすぎる問題の改善, 特定クラスにのみ Copy-Paste Augmentation の適用が挙げられる。

参考文献

- [1] G.Ghisai, *et al.*, “Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation”, CVPR, 2021.
- [2] Tian, *et al.*, “Unbiased Teacher v2: Semi-Supervised Object Detection for Anchor-Free and Anchor-Based Detectors”, CVPR, 2022.
- [3] Tian, *et al.*, “FCOS: Fully Convolutional One-Stage Object Detection”, ICCV, 2019.
- [4] T.Lin, *et al.*, “Focal Loss for Dense Object Detection”, ICCV, 2017.

研究業績

- [1] 法華津伸一 等, “車載カメラ映像からの物体検出における半教師あり学習の有効性の検証”, 自動車技術会 春季大会, 2023.
- [2] S.Hoketsu, *et al.*, “Class Weighted Focal Loss for Improving Class Imbalance in Semi-supervised Object Detection”, VISAPP, 2024. にて発表予定