MDLab グループミーティング資料
25 年 12 月 02 日 (火)

疑似ラベルを用いた遠赤外線画像からの物体検出

B4 加藤 達也

1 研究背景および目的

- 背景： 完全自動運転の実用化に向けて技術の開発が進められており、その為に車載カメラ画像からの物体検出は重要な要素技術である。可視光画像からの物体検出は天候や時間帯によって精度が低下するので、その解決策として遠赤外線からの物体検出手法を考える。
- 課題:遠赤外線画像のデータセットは可視光画像のデータセットと比較して数が少ない。
- 目的： 遠赤外線画像を入力として低照度下でも安定的に動作する検出モデルを構築する。また、RGB 画像に適応して得た検出領域を教師とするドメイン適応を用いて、遠赤外線領域における検出モデルを構築する。それらに加えて、データ拡張と損失関数の実装によって、より精度の高い検出モデルを構築する。

図 1: RGB 画像

図 2: FIR 画像

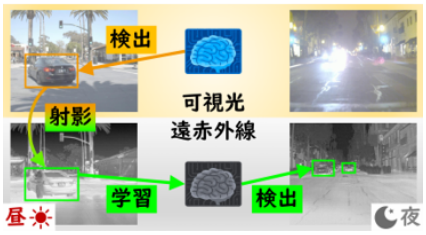
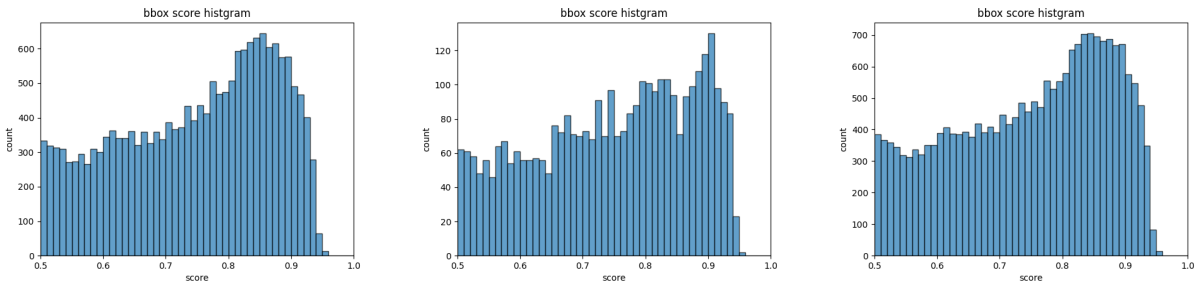


図 3: ドメイン適応の流れ

2 これまでの研究のまとめ

2.1 データセットの更新

- FLIR_ADAS_v1 に加え新たなデータセットとして FLIR_ADAS_v2 とそれらを合わせた FLIR_ADAS_v1+v2 を作成した
- v2 では解像度・視野の補正、アノテーションの変換を行っているので、v1 から画像の枚数は減ってしまったが、信頼性の高いデータが含まれている。結果として、v2 は score のグラフが右に移動している。(図 4)
- また、v1+v2 ではデータ数が増加しているので、全体的に v1 より score のグラフが右に移動している。(図 4)
- v1 と v2 では含まれる car と person のアノテーションの数やそれぞれの大きさの個数が異なるので、学習や検出結果で異なる特性を持つ。
- v1+v2 では画像のサイズが異なることから射影変換行列の切り替えが必要となる。実質現在は v1 のみがデータとして有意にはたらいている。



v1　　v2　　v1+v2

図 4: モデルによる各アノテーションへの確信度の分布

表 1: v1(sam2 整形後の annotaion の数)

category	small	medium	large	total
person	1762	2784	363	4909
car	5611	6425	1417	13453
all	7373	9209	1780	18362

表 2: v2(sam2 整形後の annotation の数)

category	small	medium	large	total
person	452	551	23	1026
car	951	1121	350	2422
all	1403	1672	373	3448

表 3: v1、v2 の各カテゴリごとのアノテーションスコアの平均、数、標準偏差

version	カテゴリ	平均	数	標準偏差
v1	person	0.745	4,909	0.110
	car	0.748	13,453	0.128
v2	person	0.699	1,026	0.105
	car	0.770	2,422	0.127

2.2 損失関数の変更

- 従来手法ではクラス、オブジェクトに対して BCELoss が使用されていた、bbox には IoULoss が使用されていた。(YOLOX のデフォルトの損失関数)
- person のデータ数が少ない、また車と比べて検出精度が低いことからクラス不均衡に対して効果的に作用する FocalLoss に変更、bbox は CIoULoss に変更した。

2.3 損失関数の最適化

- Optuna を使用して YOLOX のデフォルトの損失関数と、提案手法における損失関数 (FocalLoss と CIoULoss) を各データセットに合わせての最適化を行った。

3 前回の LT からの進捗

3.1 データセットについて

- 改めてデータセットの構造を理解した。
- v1 では、RGB 画像と Thermal 画像のサイズが異なり、射影変換を行う必要があったが、v2 では RGB 画像を Thermal 画像のサイズにリサイズしている。その際に作成される pair.json には RGB 画像と Thermal 画像のペア情報とリサイズ後に合わせた bbox 情報も含まれている。なので、v2 ではこれ以上緻密な射影変換を行う必要がない。
- 実際に RGB と Thermal の annotations ファイルの bbox 情報を画像の上に重ねて出力したところ、わずかなズレはあるものの、おおむね一致している。

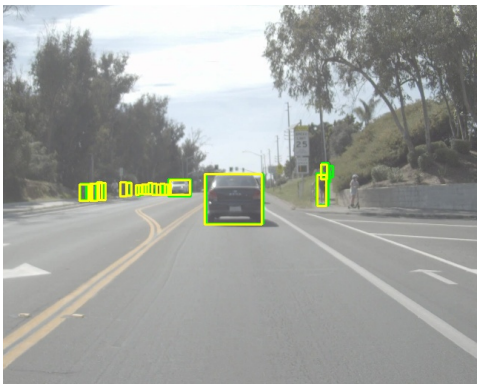


図 5: v2 の RGB と Thermal の GT の bbox の比較

3.2 疑似ラベルにおける score について

- 疑似ラベルの score について確認したところ、事前学習モデルにより画像から検出された物体に対する確信度合いであり、IoU スコアではないことがわかった。
- また疑似ラベルでは事前学習モデルによって検出された物体は、GT が想定している物体と異なるものも存在する。例えばキックボードに乗っている人を person として認識する場合があった。
- そのため、疑似ラベルの score は GT の bbox と比較した IoU スコアとは異なることに注意する必要がある。

3.3 損失関数の最適化について

- Optuna を使用して YOLOX のデフォルトの損失関数と、提案手法における損失関数 (FocalLoss と CIoULoss) を各データセットに合わせての最適化を行った。v1、v2、v1+v2 それぞれで最適なハイパーパラメータが異なることがわかった。これは各データセットでアノテーションの数や大きさの分布が異なることが影響していると考えられる。
- 各パラメータの比較一覧を作成した。

3.4 各実験環境の結果の比較について

3.5 共通の実験設定

- ベースモデル:YOLOX-s
- フレームワーク:MMDetection[8]
- 学習エポック数:300
- バッチサイズ:8
- クラス数:2(person, car)
- 最適化手法:SGD
- 評価指標:mAP(0.5:0.95)、mAP_50、mAP_70、mAP_s、mAP_m、mAP_l

3.5.1 チューニングによる検出結果への影響

表 4: v1(提案損失・チューニング前) の v1 における検出結果

category	mAP	mAP_50	mAP_70	mAP_s	mAP_m	mAP_l
person	0.036	0.135	0.008	0.034	0.100	0.172
car	0.274	0.552	0.245	0.121	0.534	0.721

表 5: v1(提案損失・チューニング後) の v1 における検出結果

category	mAP	mAP_50	mAP_70	mAP_s	mAP_m	mAP_l
person	0.077	0.247	0.026	0.066	0.168	0.285
car	0.359	0.672	0.347	0.226	0.513	0.774

- v1 データセットにおいて、損失関数のチューニングを行うことで、全体的に mAP が向上した。 特に person クラスにおいて顕著な改善が見られた。 これは FocalLoss がクラス不均衡に対して効果的に作用したためと考えられる。car に関しては mAP_m が低下している。

3.5.2 損失関数の違いによる検出結果への影響

表 6: v1(デフォルト損失・チューニング後) の v1 における検出結果

category	mAP	mAP_50	mAP_70	mAP_s	mAP_m	mAP_l
person	0.04	0.149	0.007	0.038	0.125	0.174
car	0.32	0.61	0.292	0.134	0.56	0.778

表 7: v1(提案損失・チューニング後) の v1 における検出結果

category	mAP	mAP_50	mAP_70	mAP_s	mAP_m	mAP_l
person	0.077	0.247	0.026	0.066	0.168	0.285
car	0.359	0.672	0.347	0.226	0.513	0.774

- v1 データセットにおいて、提案損失関数を用いた場合、デフォルト損失関数と比較して全体的に mAP が向上した。 特に person クラスにおいて顕著な改善が見られた。 これは FocalLoss がクラス不均衡に対して効果的に作用したためと考えられる。
- ただ、car に関しては mAP_m、mAP_l が低下している。これは v1 データセットにおいて car の中でも medium が多かったので、FocalLoss の効果が限定的であった可能性がある。

3.5.3 データセットの違いによる検出結果への影響

表 8: v1(提案損失・チューニング後) の v1 における検出結果

category	mAP	mAP_50	mAP_70	mAP_s	mAP_m	mAP_l
person	0.077	0.247	0.026	0.066	0.168	0.285
car	0.359	0.672	0.347	0.226	0.513	0.774

表 9: v2(提案損失・チューニング後) の v1 における検出結果

category	mAP	mAP_50	mAP_70	mAP_s	mAP_m	mAP_l
person	0.076	0.234	0.035	0.066	0.222	0.112
car	0.404	0.753	0.374	0.291	0.542	0.76

- 上記の比較から分かることとしては、v2 で学習したモデルでは person の large の検出精度が低下してしまう。2.1 の表で示した通り、v2 データセットでは large の person のアノテーション数が非常に少ないため、学習が十分に行われなかった可能性がある。person_large の割合として v1 は約 7.4%、v2 は約 2.2%である。
- 一方で car に関しては全体的に mAP が向上している。これについては同じく 2.1 の car のアノテーションスコアの平均が v1 から上がっていることが検出精度の向上に寄与していると考える。

4 今後の課題&スケジュール

- 現在、v1 と v2 は全てのデータセットにおいてのテストデータを出している。v1+v2 用損失関数のチューニングが終了次第、結果の比較を行う。

参考文献

[1] 谷本 樹希「セグメンテーションによる疑似ラベル補正を用いたドメイン適応型遠赤外線物体検出」 2024 年修士論文

[2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár Focal Loss for Dense Object Detection (7 Feb 2018)

[3] Zhaohui Zheng, Ping Wang, Wei Liu, Jingdong Wang, Junjie Yan, and Dong Chen Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression (6 Apr 2020)

[4] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO Series in 2021[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8370-8379.

[5] Takuya Akiba, Shotaro Sano, Takeru Yanase, Toshihiko Ohta, Masanori Koyama Optuna: A Next-generation Hyperparameter Optimization Framework (20 Feb 2019)

[6] FLIR ADAS and Thermal Dataset. https://www.flir.com/oem/adas/adas-dataset-form/(accessed: 2025-11-22)

[7] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. SAM 2: Segment Anything in Images and Videos. arXiv preprint arXiv:2408.00714, 2024.

[8] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng,Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X.Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C.C. Loy, and D.Lin, " MMDetection: Open MMLab detection toolbox and benchmark, " arXivpreprint, vol.arXiv:1906.07155, June 2019.