

# 疑似ラベルを用いた自動運転のための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。

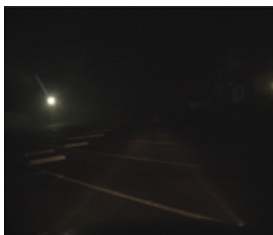


図 1: RGB 画像



図 2: FIR 画像

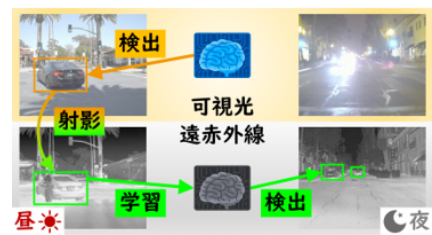


図 3: ドメイン適応の流れ

## 2 これまでの研究のまとめ

- 谷本先輩の最終的な提案手法は以下の通りである。流れとしては図 4(a) の通りである。
  1. RGB 画像に対して高精度な可視光モデルを用いて物体検出を行い、その結果を FIR 画像に変換して疑似ラベルを生成する。
  2. 可視光モデルの出力信頼度に基づき、高信頼の検出結果のみを疑似ラベルとして使用し、誤学習を防止する。
  3. FIR 画像と疑似ラベルを用いて、可視光モデルをファインチューニングし、FIR 画像専用の物体検出モデルを構築する。
- 損失関数のカスタマイズを行った。採用した損失関数は FocalLoss と L1Loss。
- クラス分類とオブジェクト性に FocalLoss、バウンディングボックス回帰に SmoothL1Loss を使用した。
- 現在、サイズが小さいオブジェクトに対しての精度が低いため、難しいサンプルに着目し、分類性能を向上させる FocalLoss を採用した。
- SmoothL1Loss に関しては YOLOX で中心座標と幅、高さの回帰に対して L1 距離を使うことが多いため、採用した。

| Category | mAP   | mAP_50 | mAP_75 | mAP_s | mAP_m | mAP_l |
|----------|-------|--------|--------|-------|-------|-------|
| person   | 0.039 | 0.194  | 0.001  | 0.039 | 0.079 | 0.087 |
| car      | 0.163 | 0.490  | 0.055  | 0.039 | 0.322 | 0.644 |

表 1: 各カテゴリの検出精度 (例)

- mAP は閾値 0.50 から 0.95 における検出精度の平均値である。

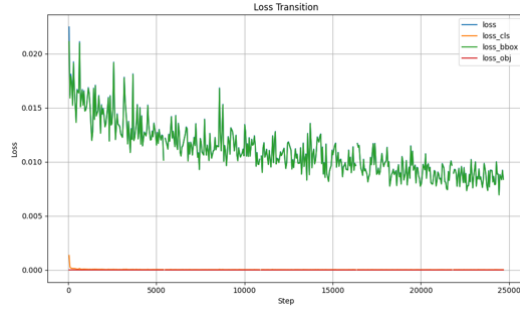


図 4: 損失の推移

### 3 前回の GM からの進捗

#### 3.1 損失関数の実装について

- config ファイルは base ファイルに上書きや追加をすることによってカスタマイズをする構造になっているが、改めて base ファイルの方を参照して見たところ、既に損失関数は実装されていた。
- bbox には IoULoss、クラス分類とオブジェクト性に CrossEntropyLoss が使用されていた。
- この 2 つとも、Yolox に最初から実装されている損失関数であったため、デフォルトの損失関数から他のカスタマイズされた損失関数に変更することによってどのように検出精度が向上するかを考えるべきであると思った。
- bbox の損失は学習初期においては大きいですが、学習末期では順調に小さくなっている。
- しかし、class や obj の損失は初期から極度に小さく、変化がほぼ見られない。
- 原因がわからない。

#### 3.2 谷本先輩の修士論文時点での検出精度と現時点での検出精度の比較

表 2: 谷本先輩の検出精度 ( $\theta_{\text{det}} = 0.50, \text{lr}=0.01$ ), best\_epoc=100

| category | mAP   | mAP_50 | mAP_75 | mAP_s | mAP_m | mAP_l |
|----------|-------|--------|--------|-------|-------|-------|
| person   | 0.024 | 0.108  | 0.001  | 0.027 | 0.072 | 0.100 |
| car      | 0.299 | 0.567  | 0.293  | 0.119 | 0.553 | 0.758 |

表 3: 私の検出精度 ( $\theta_{\text{det}} = 0.70, \text{lr}=0.00125$ ), best\_epoc=10

| category | mAP   | mAP_50 | mAP_75 | mAP_s | mAP_m | mAP_l |
|----------|-------|--------|--------|-------|-------|-------|
| person   | 0.069 | 0.236  | 0.210  | 0.058 | 0.167 | 0.118 |
| car      | 0.294 | 0.638  | 0.253  | 0.143 | 0.532 | 0.761 |

- $\theta_{\text{det}}$  は信頼度であり、信頼度以上のものが疑似ラベルとして使用される。
- 手順は全く変わっておらず、RGB 画像を元に疑似ラベルを作成→SAM2 でのアノテーション→学習・検証である。違いとしては損失関数の変更と、細かいパラメータの変更である。
- mAP が上がった原因としては、クラス分類の損失関数として設定している FocalLoss において、バランス係数 (alpha) を person と car それぞれに 0.60 と 0.40 と設定してクラス不均衡を減そうとしたことである。また、信頼度を 0.50 から 0.70 へ変更した。
- しかし、FocalLoss の特性の問題か、検出が難しい person に重みをおいているので、car の検出精度が僅かに下がっている。

$$FL(P_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

- (1) は FocalLoss の数式であり、 $\alpha$  はバランス係数であり、クラス間に不均衡がある場合に少ないクラスに対して損失の重みを大きくするために使う。[1]

- $\gamma$  は困難サンプルに集中するためのパラメータであり、判別が容易であるサンプルの影響を抑え、難しいサンプルに注目するための関数である。
- どちらも大きい数値であればあるほど影響は大きい。
- またバッチサイズが Yolo<sub>x-s</sub> と Yolo<sub>x-x</sub> でそれぞれ 16 と 8 であるが、学習率が一律で 0.01 になっていたので修正した。
- Yolo<sub>x-s</sub> は 0.0025、Yolo<sub>x-x</sub> は 0.00125 とした。
- 根拠としては YOLOX: Exceeding YOLO Series in 2021 において (2) で学習率を定めていたことからである。[2]

$$lr_{actually} = lr \times BatchSize/64 (lr = 0.01) \quad (2)$$

### 3.3 ITS ミーティングについて

- 愛知工科大学の久徳先生との毎週月曜日に行われる ITS ミーティングにおいて、ご指摘をいただいた。
- 学習率は一つの数式で決まるものではないので、繰り返し実験して自分の環境にあった学習率を模索するべき。
- 損失関数については既存の損失関数はそれぞれ想定された使い方が存在するので、それに従って適した損失関数を各所の実装する。また、数値だけではなく描画されるものも見て損失関数の傾向を掴む。
- 既存の損失関数でも良いが、谷本先輩は独自の計算での損失関数の作成を目指していた。
- 愛工大の卒業生の方が作成された v2\_dataset\_tools を使って FLIR のデータセットのバージョンアップができる。しかし、使い方がわからない。

## 4 今後の課題&スケジュール

- FLIR のバージョンアップのため、v2\_dataset\_tools について調べる。
- 損失関数に関する理解を深め、より良い検出精度を目指す。できれば独自の計算で損失関数を作成したい。

## 参考文献

- [1] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár Focal Loss for Dense Object Detection (7 Feb 2018)
- [2] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, Jian Sun YOLOX: Exceeding YOLO Series in 2021 (6 Aug 2021)