Repulsion Loss を導入した Single Stage Headless 構造による遠方歩行者検出

EP16041 木村秋斗　　指導教授：山下隆義

1. はじめに

歩行者検出は自動運転の実現に重要な技術の一つである．Faster R-CNN[1] の登場により，高速かつ高精度な歩行者検出が実現されている．しかし，遠方歩行者を検出する場合，取得すべき特徴が小さいため，検出精度と速度の両立が困難である．本研究では，小さい顔をリアルタイムに検出できる顔検出手法である Single Stage Headless face detector(SSH)[2] を基にした，高速な遠方歩行者検出を実現する．

2. Single Stage Headless face detector

SSH の構造を図 1 に示す．Detection モジュール 1，2，3 のスケールを物体サイズに応じて分けることで，遠方の小さい物体も含めた様々な大きさの物体を検出できる．また，特徴抽出と検出を 1 つのネットワークで同時に行うことで，高速化した手法である．Detection モジュールでは物体かどうかの分類と検出矩形の推定を行う．位置推定は，アンカーと呼ばれる事前に定義した矩形のベースサイズからのオフセットを学習する．

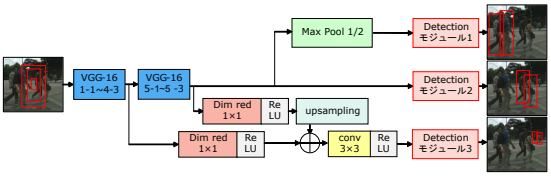

図 1：SSH の構造

3. 提案手法

本研究では，SSH により遠方歩行者を検出するために適したアンカーサイズについて調査する．また，遠方歩行者の検出領域は小さいため，誤って 1 つの領域に統合してしまうことがある．この問題を解決するために，新たな損失関数を提案する．

3.1. アンカーサイズの調査

SSH のベースアンカーサイズは 16 ピクセルである．本研究で学習，評価に使用する CityPersons データセットの歩行者の矩形の高さの分布を調査したところ，矩形の最小の高さは 8 ピクセルである．そのため，遠方歩行者を検出するためにはベースアンカーサイズを 16 ピクセルより小さくする必要がある．しかし，大幅に縮小すると誤検出が増加する．そこで本研究では，SSH のベースとなるアンカーサイズに応じた精度の比較から，遠方歩行者の検出に最適な値を調査する．

3.2. 損失関数

SSH の学習の際に，矩形の回帰損失として，異なる検出対象の矩形の位置を離すように学習する損失関数である Repulsion Loss[3] を追加する．Repulsion Loss を式 (1) に示す．

$$L_{Rep} = \alpha * L_{RepGT} + \beta * L_{RepBox} \tag{1}$$

ここで α，β は重みである．L_{RepGT}，L_{RepBox} をそれぞれ式 (2)，(3) に示す．

$$L_{RepGT} = \frac{\sum_{i \in P} l_n(IoU(b_i, g_{Rep}))}{P} \tag{2}$$

$$L_{RepBox} = \frac{\sum_{i \neq j} l_1(b_i, b_j)}{\sum_{i \neq j} \mathbb{I}[IoU(b_i, b_j) > 0] + \epsilon} \tag{3}$$

ここで，b_i は予測矩形，g_{Rep} は異なる検出対象の正解歩行者位置，P は正解と定義されたアンカー数，l_1 は smooth L1 Loss，b_j は他の検出対象の予測矩形，$\mathbb{I}(\cdot)$ は指定した値をそのまま返す恒等関数，ϵ は分母を 0 にしないための極小の値である．予測した矩形に対して L_{RepGT} は異なる検出対象の正解歩行者位置から遠ざけるように，L_{RepBox} は異なる検出対象の予測矩形から遠ざけるように学習する．

4. 評価実験

本実験では遠方歩行者の検出に適したアンカーサイズの調査を行う．また，従来の損失関数と Repulsion Loss を追加した損失関数を用いた場合の精度を比較する．

4.1. 実験概要

CityPersons データセットのうち，学習画像に 2303 枚，評価画像に 398 枚を用いる．学習回数は 45000 イタレーションとする．アンカーサイズの調査では，ベースアンカーサイズが 4，8，12，16 ピクセルの時の精度を比較する．評価はデプス画像をもとに算出した距離別に行う．また，定量的な比較に mean Average Precision(mAP) を使用する．

4.2. 実験結果

各アンカーサイズにおける距離ごとの精度を図 2 に示す．80m 以上の検出において最も精度が高かったアンカーサイズは 8 ピクセルであることがわかる．80m より手前の精度はアンカーサイズが小さいほど精度が悪くなった．これは小さいアンカーサイズでは，手前にいるサイズの大きな歩行者領域を捉えきることができないからである．

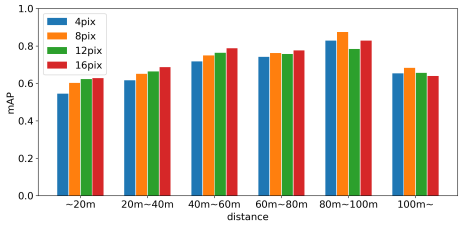

図 2：各アンカーサイズにおける mAP

従来の SSH による検出精度と Repulsion Loss を追加した 80m 以上の検出精度の結果を表 1 に示す．両手法のベースアンカーサイズはアンカーサイズの調査結果から 8 ピクセルとしている．これより，recall が 0.039 ポイント増加したことから遠方歩行者の未検出が減少したことがわかる．また，図 3 に検出例を示す．図から Repulsion Loss の導入により，重なりによって未検出であった歩行者が検出できていることがわかる．また，画像 1 枚の処理時間は入力画像サイズ 1024 × 2048 ピクセルに対して平均 0.20 秒であり，高速な歩行者検出を実現した．

表 1：損失関数変更による歩行者の検出精度

	mAP	precision	rcall
従来手法	0.648	0.797	0.659
提案手法	0.694	0.779	0.698

(a) 元画像　(b) 従来手法　(c) 提案手法

図 3：SSH による歩行者検出結果

5. おわりに

本研究では Repulsion Loss を導入した SSH 構造による歩行者検出を提案した．遠方歩行者に検出に適したアンカーサイズの調査と Repulsion Loss の導入により，従来手法より高精度な遠方歩行者検出ができることを確認した．今後は，より高速な検出の実現を目指す．

参考文献

[1] S. Ren, *et al.*, “Faster R-CNN:Towards Real-Time Object Detection with Region Proposal Networks”, IEEE, 2016.

[2] M. Najibi, *et al.*, “SSH: Single Stage Headless Face Detector”, ICCV, 2017.

[3] X. Wang, *et al.*, “Repulsion Loss: Detectting Pedestrians in a Crowd”, CVPR, 2018.