

## 頭部矩形を用いた混雑環境に頑健な人体検出手法の提案

上村 純一† 森下 雄介† 今岡 仁‡  
 †NEC バイオメトリクス研究所 ‡NEC

### 1 はじめに

人体検出とは、画像内から人間の位置を表す外接矩形とそれに対する信頼度を推定する問題である。近年、深層学習の発展に伴い人体検出精度も向上したが、依然として混雑環境における人体検出は難しい問題となっている。混雑環境では、人間同士や他の物体との重なりによって遮蔽が発生する。遮蔽により人間を表す特徴の一部が失われ、姿勢や大きさ等の見えのバリエーションが多くなり人体検出を難しくしている。また、従来の人体検出手法で用いられる Non Maximum Suppression (NMS) と呼ばれる矩形抑制処理が、混雑環境における人体検出の精度低下の要因になっている。NMS の影響により、距離が近く重なり大きい人間同士の矩形が誤って抑制され精度低下を引き起こす。

本稿では、遮蔽の影響を受けにくい頭部から人体候補矩形を推定する手法と、推定した頭部矩形から IoU 閾値を画像毎に動的に変化させる NMS 手法を用いる混雑環境に頑健な人体検出手法を提案する。

### 2 関連研究

深層学習を用いた代表的な一般物体検出手法として、Faster R-CNN[1] がある。入力された画像から特徴マップを生成する Backbone Network, 特徴マップから物体候補領域を推定する Region Proposal Network (RPN), 推定した各物体候補領域に対し物体クラス推定と矩形の位置推定を行う R-CNN の3つの CNN で構成されている。Faster RCNN は一般物体検出の分野において著しい精度改善をもたらした深層学習による物体検出の標準的な手法となっているが、混雑環境における人体検出に應用すると、遮蔽の影響により RPN の学習が難しくなり、物体候補領域の未検出が発生する。

また、一般物体検出手法の多くは、特徴マップ上の複数のアンカーボックスに対し処理を行うため、一つの物体に対して複数の矩形が出力されることがある。そのため、NMS と呼ばれる出力矩形の抑制処理が必要不可欠である。NMS では、対象とする矩形に対し重なり度合いを表す Intersection over Union (IoU) がある閾値を超えた矩形を同一物体とみなし、信頼度が最大の矩形以外を除去する。IoU 閾値を低くすると未検出の要因となり、高くすると誤検出の要因となる。そのため、混雑度に合わせて画像毎に IoU 閾値を設定することが

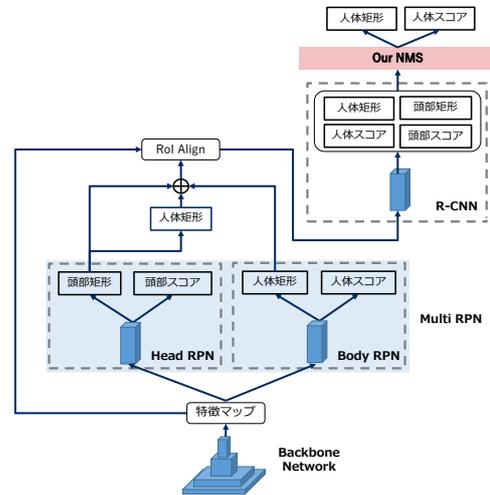


図 1. 提案手法の枠組み

好ましいが、従来の NMS では IoU 閾値が固定方式であり、なおかつ手動で設定する必要があるため、IoU 閾値を動的に設定するのが困難である。

### 3 提案手法

監視カメラを用いた実応用では同一平面上に存在する人物を上から俯瞰するように撮影することが多く、頭部に近いほど遮蔽の影響が少ない。そのため、混雑環境下では人体に比べ頭部は容易に検出でき、未検出が発生しやすい遮蔽された人体の検出の補完に、頭部矩形情報を組み合わせることが有効である。他の手法では、同時に頭部と人体を推定するものもあるが、頭部矩形と人体矩形がペアとして正解付けされている必要がある [2]。本稿では、頭部と人体を独立に推定し、推定した頭部矩形で人体候補矩形を生成、さらに頭部矩形から動的に IoU 閾値を決定する人体検出手法を提案する。提案手法の枠組みを図 1 に示す。RPN はクラス推定を行わないため、Multi RPN は頭部矩形と人体矩形を各々推定する Head RPN と Body RPN で構成されている。また、別々の人間同士の矩形の抑制を防ぐために、後処理として頭部矩形から画像ごとに動的に IoU 閾値を決定する NMS 処理 (Our NMS) を適用する。

**Multi RPN:** Head RPN は、頭部候補矩形を出力するために頭部らしさを表すスコアと頭部矩形を出力する。同様に Body RPN は、人体候補矩形を出力するために人体らしさを表すスコアと人体矩形を出力する。人体候補矩形は、Body RPN で推定した人体矩形に加え、Head RPN で推定された頭部矩形から生成した人体矩形を含

Robust Human Body Detection Method in a Crowd Using Head Boxes  
 †Jun'ichi Kamimura †Yusuke Morishita ‡Hitoshi Imaoka  
 †Biometrics Research Laboratories, NEC Corporation. ‡NEC Corporation.

める。本稿では、代表的な歩行者検出データセットである CityPerson[3] の人体矩形の正解付け方法に基づき、頭部矩形の幅の3倍の幅を持つアスペクト比0.41の矩形を人体候補領域として機械的に生成する。その後、特徴マップから物体候補にあたる部分領域を固定サイズの特徴量として抽出する RoI Align[4] を行い、各物体候補の特徴量を R-CNN に入力する。R-CNN は頭部矩形、頭部スコア、人体矩形、人体スコアを推定する。

**動的な IoU 閾値決定方法:** NMS において、IoU 閾値と低くすると未検出の要因となり、高くすると誤検出の要因となる。そのため、未検出と誤検出が発生しない混雑度に合わせた適切な IoU 閾値の範囲を求める必要がある。まず、未検出が発生する IoU 閾値を考える。IoU 閾値を下げていくと出力される矩形数は減っていき、ある値 ( $TH_{lower}$ ) を下回ると異なる人体矩形同士を同一人物だと判定し未検出が発生する。ここでの  $TH_{lower}$  は、その画像内の人体矩形同士の IoU の最大値を表している。本稿は、頭部矩形から推定した人体矩形同士の IoU を計算し、最大のを  $TH_{lower}$  とする。Head RPN から推定した頭部矩形の3倍の幅を持つアスペクト比0.41の矩形を人体矩形とする。次に、誤検出が発生する IoU 閾値を考える。IoU 閾値を上げていくと矩形数は増えていき、ある値 ( $TH_{upper}$ ) を超えると同一人物に対して複数の矩形を出力し誤検出が発生する。画像中の人物数が既知なら、誤検出が発生する IoU 閾値がわかり、 $TH_{upper}$  を決定することができる。本稿は、検出器で求めた頭部矩形の個数を人体矩形と個数を同じだと仮定する。推定した人体矩形に NMS を適用し、頭部矩形と同数になる IoU 閾値を  $TH_{upper}$  とする。

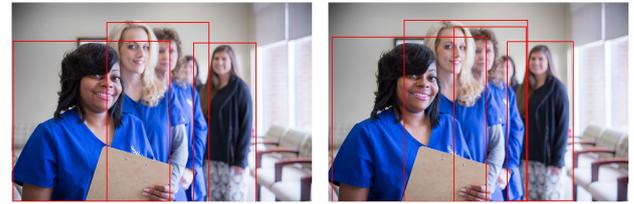
#### 4 実験

**実験設定:** 本実験では、CrowdHuman[5] の学習用と評価用サブセットを用いる。CrowdHuman は混雑環境を想定した人体検出ベンチマークであり、頭部領域、人体の可視領域、人体の領域全体の矩形が正解付けされている。学習にあたり、Backbone network として、ResNet-50[6] をベースとした Feature Pyramid Network (FPN)[7] を用いる。学習の最適化には確率的勾配降下法 (SGD) を用いる。Learning rate の初期値を0.02とし、Momentum を0.9、Weight decay を0.0005とした。学習では、NVIDIA Tesla V100 を8基用いた。IoU 閾値は  $TH_{lower}$  と  $TH_{upper}$  の平均とする。

**評価指標:** 提案手法の評価にあたり、Log-average miss rate (MR) を用いる。MR は、False Positive per Images (FPPI) が  $[10^{-2}-10^0]$  の範囲で計算される。MR が低いほど検出器の精度が高い。Reasonable (Reas.) と Heavy occlusion (HO) の2種類の評価設定を用いた。Reasonable 設定では矩形の高さ50ピクセル以上かつ可視領域が65%以上の矩形、Heavy occlusion 設定では矩形の高さ50ピクセル以上かつ可視領域が20%から65%の矩

	Multi RPN	Our NMS	Reas.	HO
ベースライン			38.32	62.17
提案手法	✓ ✓	✓	36.35 <b>35.99</b>	60.16 <b>56.96</b>

表1. ベースラインと提案手法の比較結果



(a) ベースライン (b) 提案手法

図2. 人体検出結果の画像を示す。(b) 提案手法は Multi RPN と Our NMS を適用している。

形を対象とする。

**実験結果:** Faster R-CNN をベースラインとして、提案手法との精度比較を行った。実験の結果を表1に示す。Reasonable 設定において、ベースラインの MR が38.32%に対し、提案手法では35.99%と精度向上が確認できた。Heavy occlusion 設定でも同様に精度向上が確認できた。人体検出結果の画像を図2に示す。提案手法では、重なりが大きい後方の人物も検出できていることがわかる。Multi RPN の適用による物体候補矩形の増加と、動的な IoU 閾値設定による別々の人物の矩形の抑制が抑えられたことが考えられる。

#### 5 まとめ

本稿では、遮蔽の影響を受けにくい頭部を利用した混雑環境に頑健な人体検出手法を提案した。従来手法と比較し、人体検出の精度向上が確認できた。今後は、小さい頭部と人体に対する高精度化に取り組む。

#### 参考文献

- [1] Shaoqing Ren, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99, 2015.
- [2] Kevin Zhang, et al. Double Anchor R-CNN for human detection in a crowd. *arXiv preprint arXiv:1909.09998*, 2019.
- [3] Shanshan Zhang and othres. Citypersons: A diverse dataset for pedestrian detection. In *CVPR*, 2017.
- [4] Kaiming He, et al. Mask R-CNN. In *ICCV*, Oct 2017.
- [5] Shuai Shao, et al. Crowdhuman: A benchmark for detecting human in a crowd. *arXiv preprint arXiv:1805.00123*, 2018.
- [6] Kaiming He, et al. Deep residual learning for image recognition. In *CVPR*, June 2016.
- [7] Tsung-Yi Lin, et al. Feature pyramid networks for object detection. In *CVPR*, July 2017.