

1. はじめに

車載カメラによる歩行者検出において、歩行者と車までの距離が重要となる。歩行者と車までの距離は、人の上端と下端の位置を用いることで算出することができる。しかし、従来の単眼カメラによる距離推定では、まず歩行者を検出し、その後、上端と下端を求める必要がある。そこで本研究では、Deep Convolutional Neural Network (DCNN) を用いて、歩行者を検出すると同時に、頭部と両足の位置を高精度に推定する回帰型 DCNN を提案する。また、Drop and Median Inference(Dn'MI) を用いることで高精度に歩行者の検出と上端と下端を推定できるかを検討する。

2. 提案手法

回帰型 DCNN では、図 1 のように人の頭部と両足の座標を推定する 6 個のユニットと歩行者を検出する 2 のユニットの計 8 個のユニットから出力層を構成する。人の頭部と両足は各座標を回帰問題として推定する。人検出も同様に、人らしさと背景らしさを回帰で行い、それぞれの信頼度を出力する。Dn'MI は、図 1 のように Dropout によりユニットの出力を 0 にしたネットワークを複数構築し、得られた各ネットワークの応答値の中央値を最終的に出力とする手法である。

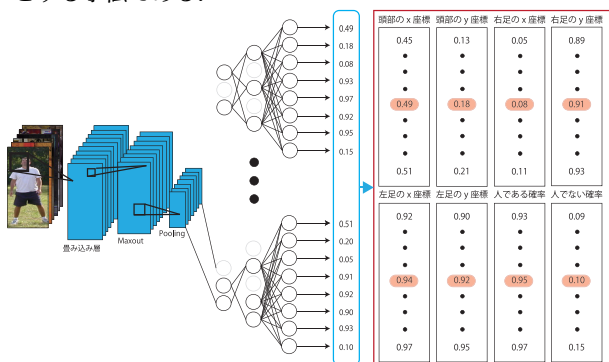


図 1: CNN の構成

3. 評価実験

提案手法の有効性を調査するために歩行者検出の精度と頭部および両足の座標の推定精度に関して実験を行う。歩行者検出では人クラスと背景クラスを判別する DCNN、座標の推定は各点の座標を回帰する DCNN と比較する。

3.1. 実験概要

本実験では、評価データとして INRIA Person Dataset を用いる。学習サンプルには、100,000 枚、評価サンプルに 10,000 枚使用する。使用する画像サイズは、128 × 64 である。評価には、False Positive Per Window(FPPW) による歩行者の検出精度と 3 点の座標の誤差を評価する。

3.2. 評価方法

3 点の位置精度は、検出した人の頭部、両足の各座標 (x, y) と教師信号の各座標 (x', y') の誤差を式 (1) により求めて評価する。ここで K は検出点の数を示す。

$$E = \frac{1}{K} \sum_{k=1}^K \sqrt{(x_k - x'_k)^2 + (y_k - y'_k)^2} \quad (1)$$

3.3. 実験結果

回帰型 DCNN と従来の DCNN との FPPW による精度の比較結果を図 2 に示す。図 2 から提案手法は、DCNN よりも検出精度が向上していることが分かる。これより、回帰型の人検出においても精度が向上させることができている。

提案手法の検出結果例を図 3 に示す。画像中の緑色の点は検出された人の頭部、両足の座標であり、赤色の点は教師信号である。図 3 よりおおよその人の各部位が推定できることが確認できる。

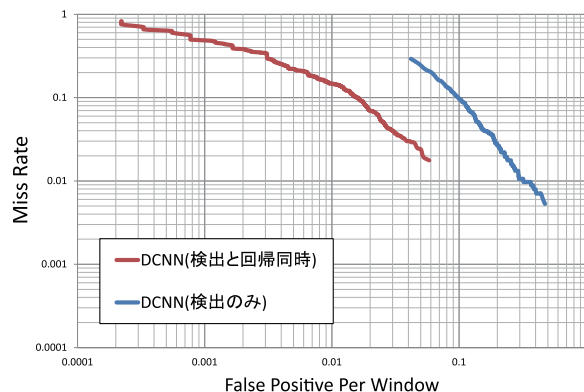


図 2: 他手法との比較



図 3: 検出結果

図 4 に各手法におけるの人の頭部の座標と教師信号との平均誤差の平均と標準偏差を示す。図 4 より検出ウィンドウの各検出部位と教師信号との平均誤差が 10pixel 以下となっている。これより、回帰推定のみ CNN と同等の精度で人検出と同時に回帰推定を実現していることがわかる。また、Dn'MI を DCNN に導入した場合、精度は向上しなかった。Dn'MI は、各ネットワークの応答値の中央値を求めるため、座標を推定する場合には有効ではないと考えられる。

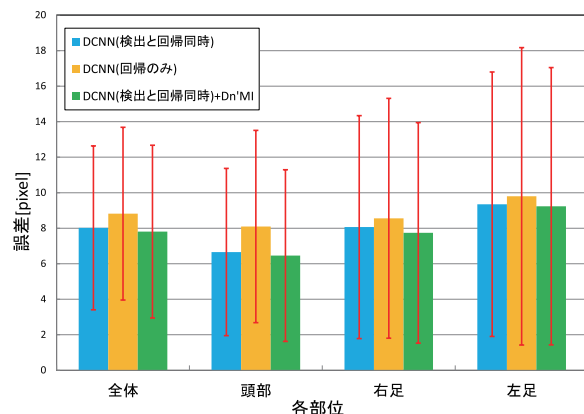


図 4: 各部位の誤差

4. おわりに

本研究では、回帰型 DCNN により人検出と同時に人の頭部、両足の位置を推定することで、歩行者検出の高精度化を実現した。また、頭部と両足の位置を高精度に推定することができた。

参考文献

- [1] Y.LeCun, et al., "Gradient-based learning applied to document recognition", Pro-ceedings of the IEEE, pp.2278-2324, 1998
- [2] 福井 宏等., "Drop and Median Inference による歩行者検出の高精度化", ビジョン技術の実利用ワークショップ, 2014.