

1. はじめに

Deep Convolutional Neural Network(DCNN) を用いた歩行者検出法は、一般的に RGB を入力としている。一方、従来の歩行者検出法では色情報だけでなく、勾配の強度や方向などのチャンネル特徴量を組み合わせた手法が高い精度を達成している。RGB から得られる情報量は少ないため、色情報以外のチャンネル特徴量を加えることで、より高精度な歩行者検出が期待できる。本研究では、Aggregated Channel Features(ACF)[1] で用いられるチャンネル特徴量を、DCNN の入力に用いることで、歩行者検出の精度の向上を図る。

2. Aggregated Channel Features

チャンネル特徴量を用いた歩行者検出法に、Integral Channel Features(ICF) がある。ICF は、RGB から複数のチャンネル特徴量を用いて検出を行うことで、歩行者の姿勢の変化や複雑な背景に対応できる手法となっている。ACF は ICF をもとにした歩行者検出法である。ICF に集約処理を導入し、Boosted Tree により検出に有効なチャンネル特徴を選択する。特徴量の補間に集約処理と Fast Feature Pyramids を導入することで、特徴量計算が削減でき、高速に処理できる。図 1 に ACF による歩行者検出の流れを示す。

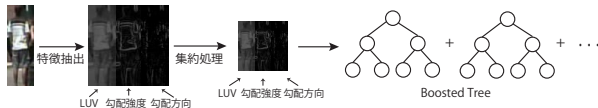


図 1：ACF による歩行者検出

3. 提案手法

本研究では、ACF を前段、DCNN を後段とした 2 段階で歩行者を検出する。DCNN の入力には ACF のチャンネル特徴量を用いる。前段の ACF では、歩行者候補領域を検出する。ACF で検出された領域から、図 2 のような LUV と勾配強度、勾配方向 6 方向の合計 10 種類のチャンネル特徴量を抽出する。抽出したチャンネル特徴量を 227 × 227 にリサイズし、後段の DCNN に入力する。入力は、RGB、LUV、勾配強度、勾配方向を組み合わせる。

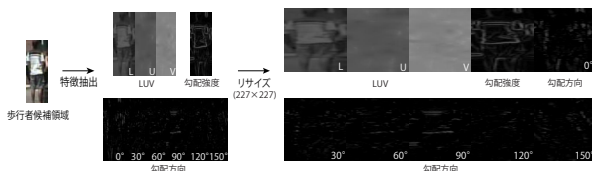


図 2：チャンネル特徴量の抽出

図 3 に、本研究で用いる DCNN のネットワーク構成を示す。ネットワーク構成は畳み込み層 5 層、全結合層 2 層からなり、畳み込み層 1 層目、2 層目、5 層目にプーリング層を適用する。

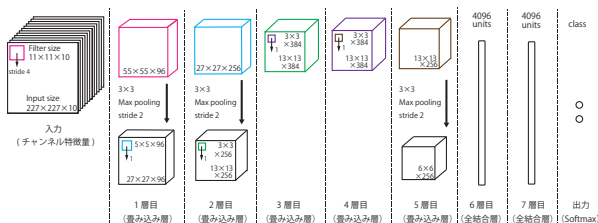


図 3：DCNN のネットワーク構成

4. 評価実験

評価実験のデータセットは、Caltech Pedestrian Dataset を用いる。ACF で検出した領域のうち歩行者をポジティブサンプル、背景をネガティブサンプルとする。歩行者と背景の分類は、アノテーションとの重なり率が 50% 以上の場合に歩行者、それ以外を背景としている。学習と評価に使用するサンプル数を表 1 に示す。

表 1：実験で用いるサンプル数

	ポジティブサンプル	ネガティブサンプル	合計
学習サンプル数	207,584	827,762	1,035,346
評価サンプル数	2,881	108,067	110,948

4.1. 検出精度の評価

評価サンプルに対する各入力の認識率を表 2 に示す。1 種類のチャンネル特徴量を入力した場合は勾配方向が、2 種類のチャンネル特徴量を組み合わせた場合は LUV と勾配方向の組み合わせが高い。3 種類のチャンネル特徴量を組み合わせた場合は、RGB と勾配強度、勾配方向の組み合わせが最も認識率が高い。このように、勾配方向を組み合わせると全体的に認識率が向上した。

表 2：評価サンプルに対する認識率

DCNN の入力	認識率 [%]	未検出数	誤検出数
RGB	99.10	78	801
LUV	99.46	42	408
勾配強度	99.33	63	682
勾配方向	99.97	5	17
LUV+勾配強度	99.65	13	266
LUV+勾配方向	<b>100</b>	0	0
RGB+勾配強度	99.65	25	294
RGB+勾配方向	99.64	22	305
勾配強度+勾配方向	99.98	2	12
LUV+勾配強度+勾配方向	99.23	0	797
RGB+勾配強度+勾配方向	99.64	19	351

ROC 曲線により検出精度を評価した結果を図 4 に示す。ROC 曲線は、DCNN が出力する歩行者クラスのスコアを閾値にして FPPI を変化させて求める。FPPI が  $10^0$  から  $10^{-2}$  の間の 9 箇所の Miss Rate の平均から算出する。前段の ACF のみの平均 Miss Rate は約 31% である。一方、RGB を入力とした 2 段階の場合は、平均 Miss Rate は約 17% と精度が向上している。チャンネル特徴量を入力した場合、RGB を入力した場合に比べ最大で約 3% 精度が向上した。しかしながら、認識率が高い LUV と勾配方向の組み合わせは、15% となっている。DCNN が出力するスコアの分布を見ると 0 または 1 に閾値を変化させても検出数が増えないため、ROC 曲線による評価では精度が低下したと考えられる。

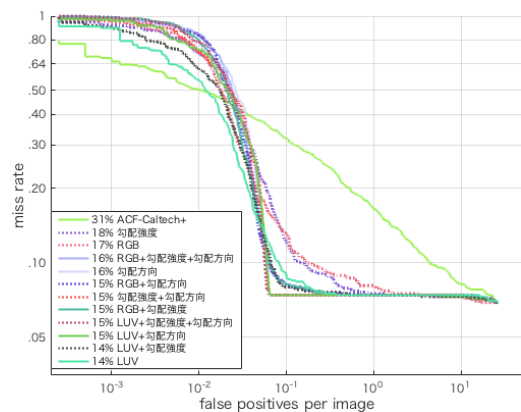


図 4：ROC 曲線による各手法の検出精度

5. おわりに

歩行者検出の精度の向上を図るために、チャンネル特徴量を組み合わせて、DCNN に入力する歩行者検出を提案した。ROC 曲線による評価では、ACF や従来の RGB を入力とした DCNN に比べ精度が向上する結果となった。今後は、DCNN のネットワーク構成や入力するチャンネル特徴量のサイズを変更し、精度が向上するか調査する。

[1] P. Dollár, et al., "Fast Feature Pyramids for Object Detection", PAMI, 2014.