

1. はじめに

Deep Convolutional Neural Network (DCNN) は姿勢推定や人物検出において高い精度を実現している。姿勢推定と運転手検出を個々の DCNN で構築すると、計算コストが増加するため、実用化の面で大きな課題となる。そこで、本研究では複数のタスクを同時に学習できる Heterogeneous Learning を DCNN に導入することで、高精度な姿勢推定と運転手検出を 1 つの DCNN で実現する。

2. 提案手法

本研究では、Heterogeneous Learning を導入することで、姿勢推定と運転手検出を 1 つの DCNN で行う。

2.1. 姿勢推定と運転手検出

運転時に起きる明暗変化に対して頑健に識別するために、入力には IR 画像と Depth 画像を用いる。対象となる運転席領域を切り出した  $96 \times 120$  の画像を図 1 のように DCNN へ入力する。DCNN の出力ユニットは、姿勢推定の部位位置の数  $8 \times 2$  と、運転手検出 2 クラスの 18 ユニットである。推定する運転手の部位位置は、頭、首、右肩、右肘、右手、左肩、左肘、左手の 8ヶ所であり、その  $x$  座標および  $y$  座標を出力する。また、運転手検出では、背景クラスと運転手クラスの確率を出力する。

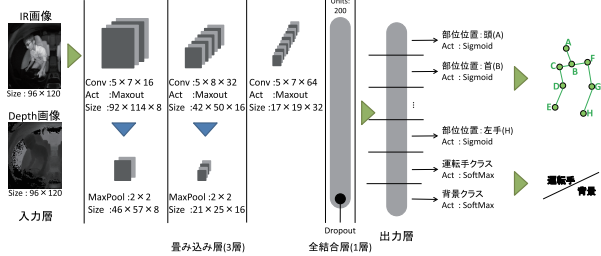


図 1：Heterogeneous Learning を導入した DCNN

2.2. Heterogeneous Learning の導入

複数の識別タスクを実現する場合、タスクの数に比例して識別器を構築する必要がある。Heterogeneous Learning は、単一のネットワークで複数のタスクを学習する。そのため、推定時にタスクが増加しても計算コストは大きく増加しないというメリットがある。

本研究では、出力層で姿勢推定は回帰、運転手検出は認識の値を出力する。姿勢推定は部位位置 8ヶ所の  $x$  座標と  $y$  座標が出力されるため、学習誤差  $E^{skelton}$  は式 (1) のように L2 ノルム二乗誤差を用いる。

$$E^{skelton} = \sum_{n=1}^N \|L_n^{skelton} - O_n^{skelton}\|_2^2 \quad (1)$$

このとき、 $L_n^{skelton}$  は姿勢推定の教師信号、 $O_n^{skelton}$  は出力値、 $N$  は部位位置の数である。運転手検出は入力画像に対する各クラスの確率を出力するため、学習誤差  $E^{human}$  は式 (2) のようにクロスエントロピー誤差を用いる。

$$E^{human} = - \sum_{c=1}^C (L_c^{human} \log O_c^{human}) \quad (2)$$

このとき、 $L_c^{human}$  は運転手検出の教師信号、 $O_c^{human}$  は出力値、 $C$  はクラス数を表す。Heterogeneous Learning における学習誤差  $E$  は式 (3) となる。

$$E = \frac{1}{M} \sum_{m=1}^M (\alpha E_m^{skelton} + (1 - \alpha) E_m^{human}) \quad (3)$$

$\alpha$  は姿勢推定と運転手検出の学習誤差に対する重みであり、どちらのタスクを優先するかを決める。Heterogeneous Learning では  $\alpha$  を 0.5 とし、2 つのタスクを平等に学習している。 $M$  はミニバッチである。運転手検出のタスクは、背景画像に推定する姿勢が存在しないため、式 (3) では背景画像の姿勢を学習するという問題がある。そこで、提案手法では、式 (4) に示すように運転手クラスの教師信号  $L_1^{human}$

を用いて背景クラスにおける姿勢推定誤差  $E^{skelton}$  を消去する。運転手クラスの教師信号  $L_1^{human}$  は運転手クラスの場合 1、背景クラスの場合 0 となる。このとき、 $\alpha$  は実験により最適値を決める。

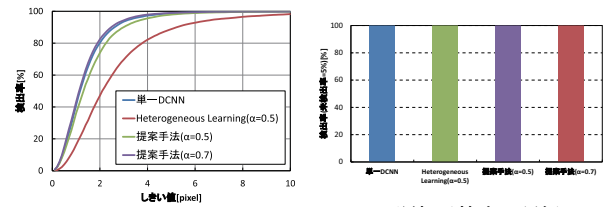
$$E' = \frac{1}{M} \sum_{m=1}^M (\alpha L_{1,m}^{human} E_m^{skelton} + (1 - \alpha L_{1,m}^{human}) E_m^{human}) \quad (4)$$

3. 評価実験

提案手法の有効性を評価実験により確認する。本実験で用いる学習サンプル枚数は 25,000 枚、評価サンプル枚数は 2,000 枚である。

3.1. 実験結果

図 2 に姿勢推定の評価結果を示す。図 2(a) より、 $\alpha = 0.5$  のとき、Heterogeneous Learning より提案手法の方が精度は良いことが分かる。また、しきい値を 2pixel としたとき提案手法 ( $\alpha = 0.7$ ) が 83% であり、最も精度が良い。図 2(b) より、運転手検出では全ての手法において検出率 100% であった。



(a) 姿勢推定の評価

(b) 運転手検出の評価

図 2：評価結果

図 3 に運転手の姿勢推定例を示す。赤が教師信号、緑が提案手法による推定結果である。図 3 より、簡単な姿勢から一部の部位が遮蔽されている姿勢でも、姿勢を正確に推定することが可能であることが分かる。

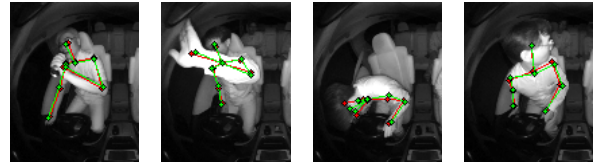


図 3：姿勢推定例

図 4 に処理速度の比較結果を示す。GPU は Geforce GTX1080 を使用する。図 4 より、提案手法は単一 DCNN と比べて約 1.5 倍 (1.7msec) の処理時間の短縮を実現した。また、モデルパラメータのメモリ使用量は単一の DCNN で  $7.4+7.5=14.9$ MB で、提案手法は 7.4MB であるため、約 2 倍のメモリ使用量の削減を実現した。

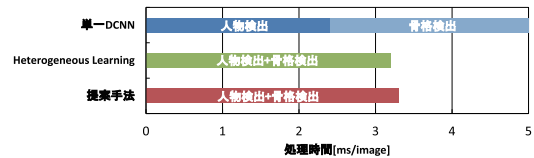


図 4：処理速度

4. おわりに

Heterogeneous Learning を導入することで、2 つのタスクを同時処理し、姿勢推定では単一 DCNN より高精度に、運転手検出では同等の精度で推定できた。今後は、多タスクを組み合わせた Heterogeneous Learning の高精度化について検討する。

参考文献

[1] X. Yang, et al., "Heterogeneous Multi-task Learning with Sparsity Constrains", NIPS, 2009.