

1. はじめに

人物照合は、あるカメラで撮影された人物が別のカメラで撮影された人物と同一かどうか判定するタスクである。従来の人物照合 [1] は、人物領域全体から算出された類似度をもとに判定するため、姿勢変化が考慮されていない。本研究では Heterogeneous Multi-task Learning を、Deep Convolutional Neural Network(DCNN) に導入することで人物照合と骨格検出を同時に行い、人物照合の精度向上を目指す。

2. 提案手法

Heterogeneous Multi-task Learning[2] は、単一のネットワークで複数のタスクを同時に学習することができる。そのため、サブタスクと同時に学習させることでメインタスクの精度を向上させる。本研究は、Heterogeneous Multi-task Learning を導入し、人物照合と骨格検出を単一の DCNN で行う。以下に詳細を述べる。

2.1. 人物照合と骨格検出

人物領域の RGB 画像を図 1 のように DCNN へ入力する。DCNN の出力は、人物照合が 4,096 次元の特徴量、骨格検出の部位が 6×2 pixel である。人物照合は、クエリ画像と登録画像の特徴量のユークリッド距離を算出し、しきい値よりも小さければ同一人物と判定する。また骨格検出は、頭、右肩、左肩、腰、右足、左足の 6 カ所の座標 (x, y) を出力する。

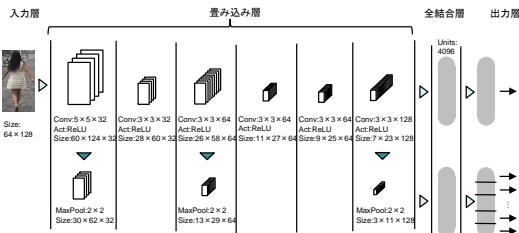


図 1：Heterogeneous Multi-task Learning を用いた DCNN

2.2. 人物照合と骨格検出の学習

本手法では、人物照合の誤差関数に Triplet loss を用いる。Triplet loss は、基準となるクエリ画像 I_i^o と、クエリ画像と同じ人物に属するポジティブ画像 I_i^+ の距離が近く、クエリ画像と異なる人物に属するネガティブ画像 I_i^- との距離が遠くなるように誤差を算出する。距離計算には、ユークリッド距離を用いる。Triplet loss を式 (1) に示す。

$$d(I_i^o, I_i^+, I_i^-, w) = \begin{aligned} & d(\phi_w(I_i^o), \phi_w(I_i^+)) \\ & - d(\phi_w(I_i^o), \phi_w(I_i^-)) \leq \tau_1 \end{aligned} \quad (1)$$

$d(\cdot)$ はユークリッド距離、 w は重み、 ϕ_w は特徴量、 $\tau_1 (< 0)$ はしきい値である。人物照合の学習誤差は式 (2) のように、Triplet loss を用いて算出する。

$$E_{person} = triplet(d(I_i^o, I_i^+, I_i^-, w), \tau_1) \quad (2)$$

骨格検出は 6 カ所の x, y 座標を出力するため、学習誤差は式 (3) のように L_2 ノルムを用いて算出する。

$$E_{pose} = \sum_{n=1}^N \|L_n - O_n\|_2^2 \quad (3)$$

このとき、 L_n は骨格検出の教師信号、 O_n は出力値、 N は部位数である。

以上より、 E_{person} と E_{pose} に重みづけして学習誤差 E を式 (4) より求め、Heterogeneous Multi-task Learning に用いる。

$$E = \alpha E_{person} + (1 - \alpha) E_{pose} \quad (4)$$

α は人物照合と骨格検出の 2 つのタスクの学習誤差に対する重みで、どちらのタスクを優先するかを決めるパラメータである。

3. 評価実験

評価実験では、単一 DCNN と提案手法を比較する。表 1 に、単一 DCNN と α を変更した提案手法との比較結果を示す。表 1 より、単一 DCNN の Accuracy は 81.84% であり、 $\alpha = 0.4$ の提案手法では 85.11% と、単一 DCNN の精度より 3.27% 向上した。

表 1： α を変更した人物照合の評価

| | Accuracy | Precision | Recall | F 値 | |
|---------|--------------------|--------------|--------------|--------------|--------------|
| 単一 DCNN | 81.84 | 94.08 | 75.57 | 95.90 | |
| 提案手法 | ($\alpha = 0.1$) | 82.27 | 92.72 | 76.69 | 95.88 |
| | ($\alpha = 0.2$) | 84.00 | 94.98 | 77.87 | 96.36 |
| | ($\alpha = 0.3$) | 84.60 | 92.19 | 80.03 | 96.30 |
| | ($\alpha = 0.4$) | 85.11 | 93.12 | 80.26 | 96.46 |
| | ($\alpha = 0.5$) | 85.08 | 91.63 | 81.02 | 96.37 |
| | ($\alpha = 0.6$) | 84.71 | 91.79 | 80.41 | 96.30 |

表 2 に、単一 DCNN と骨格を減らした提案手法 ($\alpha = 0.4$) との比較結果を示す。表 2 より、腰の部位を除いた提案手法の精度が最も低いことから、腰の部位の情報が人物照合タスクに重要であることが分かる。

表 2：骨格を減らした人物照合の評価

| | Accuracy | Precision | Recall | F 値 |
|--------|--------------|--------------|--------------|--------------|
| 骨格 6 点 | 85.11 | 93.12 | 80.26 | 96.46 |
| 頭無し | 84.56 | 92.69 | 79.73 | 96.32 |
| 肩無し | 83.90 | 93.02 | 78.67 | 96.21 |
| 腰無し | 81.01 | 94.68 | 74.35 | 95.79 |
| 足無し | 85.16 | 93.09 | 80.35 | 96.47 |

提案手法 (左足無し) で照合されたペア画像を図 2 に示す。図 2 のように、骨格検出を学習することで単一 DCNN では照合できなかった画像を照合できるようになった。



(a) 成功例 (同一人物)

(b) 成功例 (別人)

図 2：照合画像

4. おわりに

本研究では、Heterogeneous Multi-task Learning を DCNN に導入することで、人物照合の精度向上を実現した。今後は、さらなる人物照合の精度向上を目指す。

参考文献

- [1] 井尻善久 等, “サーベイ論文: 視野を共有しない複数カメラ間での人物照合”, PRMU, vol. 2011, no. 21, pp. 117-124, 2011.
- [2] X. Yang, *et al.*, “Heterogeneous Multitask Learning with Joint Sparsity Constrains”, NIPS, pp. 2151-2159, 2009.