

# 領域ベース CNN と Triplet Loss を用いた 移動ロボット向けオンライン転移学習型人物再同定 Person Re-identification by Online Transfer Learning for Mobile Robot using Region-based CNN and Triplet Loss

村田 祐樹<sup>\*1</sup>  
Yuki Murata

渥美 雅保<sup>\*1</sup>  
Masayasu Atsumi

<sup>\*1</sup>創価大学大学院工学研究科情報システム工学専攻  
Information Systems Science, Graduate School of Engineering, Soka University

We propose a person re-identification method for mobile robots that periodically provides services to specific groups. This method consists of a feature extractor that learns to extract person features based on the Triplet Loss from person regions detected by a region-based CNN and a person re-identifier that learns to identify persons through transfer learning of person features while moving around a room. The person re-identification incorporates adaptive transfer learning to periodically re-learn the same persons with different appearance, such as clothes etc. Performance of the proposed method is evaluated by an experiment using a public large-scale data set and an experiment using the self-made dataset periodically collected for the same group by the mobile robot.

## 1. はじめに

人物再同定とは一般に、視野を共有しない複数のカメラに映る人物から同一人物を識別する問題である[Zheng 16]. 応用例としては屋外空間に設置された防犯カメラ映像を利用した警備システムや、商業施設におけるマーケティングなどが挙げられる. 移動ロボットによる人物再同定は、移動ロボットが移動中に再会した同一人物を識別する問題となる. 本研究では、特定の集団に定期的にサービスを提供する移動ロボットのための人物再同定手法を提案する.

本手法は、領域ベース CNN が検出する人物領域から Triplet Loss[Florian 15]に基づき人物特徴表現を学習する特徴抽出器と人物特徴の転移学習を行う人物再同定器とから構成される. 人物再同定器には服装等部分的に外見が異なる人物を定期的に再学習するための転移学習が組み込まれる. 公開されている大規模データセットを用いた実験、及び移動ロボットにより同一集団に対して定期的に収集した自作データセットを用いた実験により、提案手法の性能を評価する.

## 2. 人物再同定手法

### 2.1 概要

提案手法は、図 1 に示す 3 つの CNN-人物検出 CNN、人物特徴抽出 CNN、人物再同定 CNN-から構成される. 人物検出 CNN は画像から人物領域を検出する. 人物特徴抽出 CNN と人物再同定 CNN は検出された人物領域の座標とその中間層の特徴を受け取り、それぞれ Triplet Loss を用いた人物特徴表現学習と移動しながらのオンライン転移学習に基づき人物特徴抽出と再同定を行う.

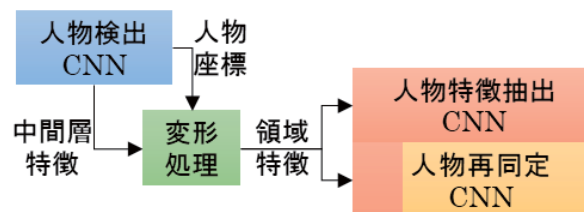


図 1 人物再同定手法

### 2.2 人物検出 CNN

人物検出 CNN は、画像から人物再同定のための人物領域を取り出す CNN である. 高精度で実時間の人物再同定を実現するためには、まず、人物検出において人物の見落としや誤検出を極力減らし、リアルタイムに検出処理を行う必要がある. 本研究では、検出器として YOLOv2[Redmon 16]を用いることで、これら課題を解決する. YOLOv2 の CNN は COCO データセットで事前に学習されたものを使用し、人物を含む 80 種類の物体を検出することができる.

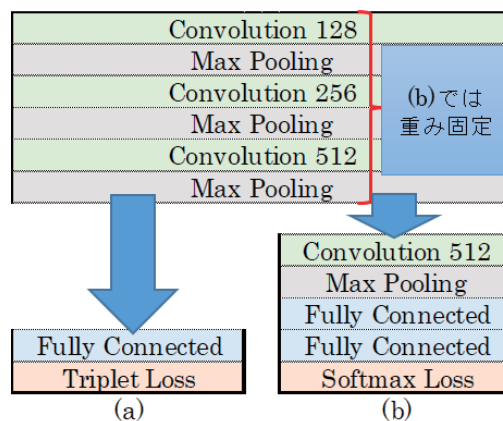


図 2 特徴抽出器(a)と再同定器(b)の CNN 構成

連絡先: 村田 祐樹, 創価大学大学院工学研究科,  
〒192-8577 東京都八王子市丹木町 1-236, Tel: 042-691-2211,  
Email: y.murata0105@gmail.com

## 2.3 人物特徴抽出 CNN

人物特徴抽出 CNN(図 2(a))は複数の畳み込み層とプーリング層で構成され、最後の全結合層で 512 次元の人物特徴を出力するネットワークである。損失関数には Triplet Loss が用いられる。Triplet は基準となるサンプル(anchor), それと同クラスのサンプル(positive), 別クラスのサンプル(negative)の 3 つの組で構成される。それぞれが重みを共有する同一の特徴抽出 CNN へ入力されて、出力される特徴量同士を比較したときに、同一クラス間では距離が小さく、別クラス間で距離が大きくなるように学習が行われる。

## 2.4 人物再同定 CNN

人物再同定 CNN(図 2(b))は、前半部分が特徴抽出 CNN と共通の層構成をとり、後半部分に再同定用の畳み込み層とプーリング層、及び全結合層を追加したネットワークで、出力層には人物クラスを分類するソフトマックス関数、損失関数には交差エントロピーを用いている。前半部分は人物特徴抽出 CNN 学習時の重みを利用し、後半で追加された層のみパラメータの更新を行う転移学習を取り入れることで、変化への適応と収束の促進を図っている。

## 3. 実験: 人物特徴抽出と再同定の性能評価

### 3.1 実験概要

特徴抽出 CNN と再同定 CNN の性能評価実験を行う。特徴抽出 CNN の実験では、出力される特徴量間の類似度を利用して識別精度を評価する。再同定 CNN の実験では異なるデータセットで学習した特徴抽出 CNN の重みを利用して転移学習を実施した際の再同定精度を評価する。事前に行った複数の実験から、各 CNN に入力する領域特徴は検出 CNN の 2 番目の畳み込み層から取得し、サイズは  $20 \times 20 \times 64$  とした。

### 3.2 データセット

実験に使用する 2 種類のデータセット構成を表 1 に示す。

表 1 データセット構成 (サンプル数 / クラス数)

|          | Train       | Gallery     | Query      |
|----------|-------------|-------------|------------|
| PRW      | 17690 / 483 | 16343 / 450 | 2057 / 450 |
|          | Train       | Test        |            |
| SOTARO-1 | 12789 / 41  | 3193 / 41   |            |
| SOTARO-2 | 12825 / 48  | 3194 / 48   |            |

#### (1) PRW データセット

公開されている人物再同定向けデータセットの多くは、人物領域が切り取られた画像で構成されている。一方、Person Re-identification in the Wild(PRW)[Zheng 16]では、画像全体と人物領域の座標情報が提供されている。提案手法では人物検出と同時に CNN の中間層から特徴を取り出すため、人物領域が切り取られる前の画像全体のデータが必要となる。そこで、今回はこのデータセットを、人物特徴表現の学習のために使用する。

#### (2) SOTARO データセット

SOTARO-1 および SOTARO-2 は、当研究室で開発を進めている授業支援ロボット SOTARO[渥美 17]で収集したデータを再構成して作成したデータセットである。2 種類のデータセットは別日に収集されており、服装等見え方の異なる同一人物のサンプルが含まれている。

## 3.3 人物特徴抽出器の評価

まず、訓練データ(表 1 の Train)からクラスラベルに基づいて Triplet を作成し、特徴抽出 CNN を学習する。次に、学習した CNN を使って全サンプルの人物特徴を抽出し、サンプル間のコサイン類似度を求めて分類を行う。PRW データセットではギャラリーデータ(表 1 の Gallery)とクエリデータ(表 1 の Query)間、SOTARO データセットでは訓練データとテストデータ(表 1 の Test)間の類似度を算出する。結果を表 2 に示す。

表 2 類似度評価の結果

| dataset  | r1   | r5   | r10  | r20  | mAP  |
|----------|------|------|------|------|------|
| PRW      | 78.7 | 89.0 | 92.2 | 94.3 | 33.6 |
| SOTARO-1 | 84.1 | 86.0 | 86.9 | 87.8 | 79.8 |
| SOTARO-2 | 78.9 | 82.2 | 83.3 | 84.5 | 74.6 |

ここで r1, r5, r10, r20 は類似度の高い上位 1, 5, 10, 20 サンプル中に 1 つでも正解となるサンプルが含まれる場合に正しいと評価する指標である。また、mAP は mean Average Precision を表す。

## 3.4 人物再同定器の評価

3.3 で学習済みの特徴抽出 CNN と SOTARO データセットを用いて人物再同定 CNN の転移学習を実施し、分類性能を評価する。結果を表 3 に示す。

表 3 再同定評価の結果

| feature  | trans    | r1   | r5   | r10  | r20  | mAP  |
|----------|----------|------|------|------|------|------|
| PRW      | SOTARO-1 | 79.9 | 92.7 | 96.3 | 98.4 | 85.3 |
| PRW      | SOTARO-2 | 71.4 | 85.3 | 90.7 | 95.4 | 77.9 |
| SOTARO-2 | SOTARO-1 | 81.3 | 93.6 | 97.2 | 98.4 | 86.8 |
| SOTARO-1 | SOTARO-2 | 78.3 | 91.9 | 95.0 | 98.2 | 84.5 |

ここで、1 列目の 'feature' は特徴抽出 CNN の学習に使用したデータセットを、2 列目の 'trans' は再同定 CNN の転移学習に使用したデータセットを表す。

1・2 行目と 3・4 行目の比較から、特徴抽出 CNN の学習に PRW データセットを用いた場合より SOTARO データセットを用いた方が転移学習後の再同定性能が優れていることが分かる。この結果から、特徴表現の学習には、転移学習を行う環境と類似するデータセットを用いるべきであるといえる。

また、転移学習の収束に要した時間は GPU 環境下で 5 分ほどであり、短時間で異なるデータセットに適した再同定器の学習ができることが分かった。

## 4. オンライン転移学習型人物再同定システム

### 4.1 概要

人物検出から再同定までの各処理を同時並行的に行うシステムの構成を図 3 に示す。図 1 との違いは、データセット作成のための領域特徴・人物特徴の保存や人による GUI を用いた半自動的なラベル付け処理が追加されている点である。図中の Gallery には、人物が検出されるたびにその特徴が蓄積される。その時点で検出・取得した人物特徴と、Gallery 内に保存されている全サンプルの特徴のコサイン類似度を求め、ある条件に従って自動的にラベルを与えるか、GUI を利用したラベル付けを行うかを判断する。現在試作段階のシステム構成では、最も類似するサンプルの類似度が事前に設定した閾値以上なら自動

的にラベルを与え、そうでない場合は GUI による操作でラベルを付ける処理へ切り替えることとしている。

人物再同定 CNN の転移学習は、訓練データが集まり次第、任意のタイミングで実施する。再同定器生成後は、人物特徴の Gallery サンプルとのコサイン類似度の代わりに人物再同定結果の分類クラス確率を半教師付きラベル付けに利用する。この場合は、再同定 CNN が出力する分類クラス確率が閾値以上なら自動的にラベルを与え、そうでなければ GUI による操作へ切り替えることとしている。

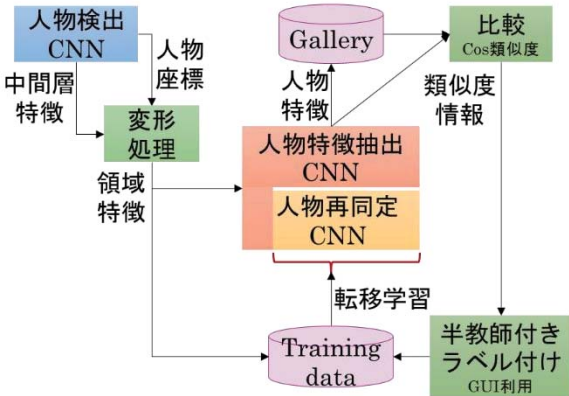


図 3 オンライン転移学習型人物再同定システム構成

4.2 実験: データセット分割による段階ごとの性能評価

本実験では、システムによる半自動ラベル付け処理が正確に行われたと仮定し、サンプル数の増加に対する再同定精度の推移を評価する。3.2 で紹介した SOTARO データセットを 6 分割し、図 4 で示すように 5 段階の再同定精度を比較した。

各段階におけるサンプル数とクラス数(人数)、そして再同定精度をまとめた結果を表 4・表 5 に示す。表 4(分割実験 1)は特徴抽出 CNN の学習に SOTARO-2 を、再同定 CNN の転移学習に SOTARO-1 を使用した結果、表 5(分割実験 2)はその逆の条件で実験を行った結果を示している。分割実験 1 では学習に使用するサンプル数が増えるたびに再同定精度も向上する結果となった。一方、分割実験 2 では段階 3 において精度が若干低下した。これは段階を踏むごとにデータ量とともにクラス数も増加しており、分類問題としての難易度が上がっているためと考えられる。2 つの実験から、データ量を増やすことが精度向上に繋がり、早い段階から多くの人数のサンプルを集めることが重要であることが分かった。

4.3 ラベル付け GUI 設計

現在開発中の GUI を図 5 に示す。この GUI は、ラベル選択部とラベル調節部で構成される。画面上部のラベル選択部では、左端に表示されたサンプルと比較して、類似度の高い 5 つのラベル候補サンプルからユーザーが手動でラベルを付けることができる。また、それまで出会ったことのない人物を見つけた場合、新たなラベルを追加することができる。画面下部のラベル調節部では、各ラベルに含まれるサンプル数やその代表となる画像を表示するほか、ラベルの修正が可能である。

5. むすび

本論では、領域ベース CNN の中間層特徴から Triplet Loss に基づく特徴学習と転移学習により人物再同定を行う移動ロボ

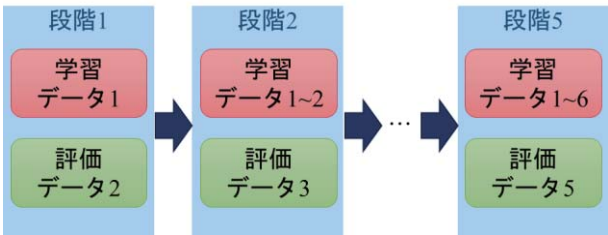


図 4 再同定精度の段階的評価方法

表 4 データセット分割実験 1

|       | 段階1  | 段階2  | 段階3  | 段階4   | 段階5   |
|-------|------|------|------|-------|-------|
| サンプル数 | 2663 | 5327 | 7991 | 10654 | 13318 |
| クラス数  | 35   | 40   | 40   | 40    | 40    |
| 精度    | 19.5 | 27.4 | 56.8 | 59.7  | 81.4  |

表 5 データセット分割実験 2

|       | 段階1  | 段階2  | 段階3  | 段階4   | 段階5   |
|-------|------|------|------|-------|-------|
| サンプル数 | 2668 | 5336 | 8004 | 10672 | 13340 |
| クラス数  | 28   | 42   | 45   | 48    | 48    |
| 精度    | 26.2 | 35.0 | 26.3 | 52.4  | 74.9  |



図 5 ラベル付け GUI

ット向けの手法を提案した。そして、大規模データセットと自作のデータセットに対して行った実験により、その有用性を確かめた。また、データセットを分割して行った追加の実験から、再同定精度の向上とサンプルとなるデータ量の関係が明らかとなった。今後の課題としては、高精度に人物再同定を行うために早い段階でデータ量を増やす工夫や、再同定結果を移動ロボットのサービス提供に活かすことが挙げられる。

参考文献

[Zheng 16] L. Zheng, et al.: Person Re-identification in the Wild, arXiv preprint arXiv:1604.02531, 2016.  
[Florian 15] S. Florian, et al.: FaceNet: A Unified Embedding for Face Recognition and Clustering. arXiv preprint arXiv:1503.03832, 2015.

- [Redmon 16] J. Redmon, and A. Farhadi.: YOLO9000: Better, Faster, Stronger. arXiv preprint arXiv:1612.08242, 2016.
- [渥美 17] 渥美雅保, et al.: SOTARO:オープン チャットとロボットの連携による演習授業支援システム—Human-in-the-loop 型機械学習によるアプローチ. 情報処理学会第 79 回全国大会, 2017.