

法政大学学術機関リポジトリ

HOSEI UNIVERSITY REPOSITORY

移動ロボットのためのCNNを用いた特定人物の識別 アルゴリズムの開発

著者	福田 一輝
出版者	法政大学大学院理工学研究科
雑誌名	法政大学大学院紀要. 理工学・工学研究科編
巻	59
ページ	1-6
発行年	2018-03-31
URL	http://doi.org/10.15002/00021604

移動ロボットのためのCNNを用いた特定人物の 識別アルゴリズムの開発

DEVELOPMENT OF TARGET PERSON IDENTIFICATION ALGORITHM
FOR MOBILE ROBOT USING CNN

福田一輝

Kazuki Fukuda

指導教員 小林一行教授

法政大学大学院理工学研究科システム理工学専攻修士課程

The purpose of this study is to develop a system for detecting target persons by applying CNN. In order to achieve stable detection of the target person, huge target person images are required to learn CNN. Because of acquired target person images are limited due to experimental time and environment. To solve a limited number of acquired target person images, we apply two approach to generate images by applying increased image techniques. One is to employ the omnidirectional camera which easily generates various angle view images. The other is to generate composite images by using a diffident view of target person and background images. Validity of the proposed approach is confirmed by using actual experimental data in the Tsukuba challenge environment.

Key Words: Mobile Robot, Convolutional Neural Network, Tsukuba Challenge, Composite Image

1. はじめに

昨今、人工知能の研究が世界で再び注目されている[1]. 中でも、計算機の性能向上により現実的に実現可能となった技術である深層学習はデータの特徴を認識し高精度に判別できるものである. その深層学習の中でも画像から物体を認識することに長けた手法として、CNN(Convolutional Neural Network)[2]が挙げられる.

近年では将棋や囲碁などの娯楽から、金融や医療など実用的なものまで、様々な分野でのCNNの利用が目覚ましい. そして高精度の物体認識を必要とする移動ロボットへの適用も例外ではない. 近年自動運転などの分野は開発が盛んであるが[3], 高精度の物体認識を可能とするCNNなどの手法も盛んに取り入れられている[4][5].

自律移動ロボットの開発研究の場として、つくばチャレンジという大会がある. 詳細は次項で紹介するが、この大会では長距離(約2km)の自律走行の他に人物探索のタスクが設定されている. 2016年の大会では人物探索のタスクを達成するために53チーム中9チームが深層学習を利用して[6][7]. そこで本研究では、深層学習の一手法であり、特に画像処理に向く方法であるCNNを全天球カメラの画像に対し適用し、2017年のつくばチャレンジで人物探索を行うことを目的としてアルゴリズムとロボットの作成を行った.

2. つくばチャレンジ

つくばチャレンジとは”人間とロボットが共存する社会の実現のための先端的技術への挑戦”を掲げ、つくば市内の遊歩道などのリアルワールドを自律移動ロボットに自律走行させる技術チャレンジである[7].

技術レベルの変化に合わせて変化するが、本大会では2kmの完走だけではなく、幾つかのタスクが設定されており、そのうちの 하나가人物探索である. Fig.1につくばチャレンジ2017で実際に設定されたマップを示す. 図中のつくばセンター広場周りの斜線領域が人物探索エリアを示しており、このエリア内のどこかに複数の探索対象人物が配置されており、これらを識別し発見する必要がある. 人物探索エリア内でロボットが発見しなければならない探索対象人物をFig.2に示す. 探索対象人物は、ジャンパー(緑)にベスト(橙または青)とキャップ(橙)を着用し、椅子に座りその隣には看板(高さ約90cm)を配置している.

過去数年のつくばチャレンジでは、多くの大学研究室や、研究所等のチームでは、カメラからの色抽出による方法や、測域センサによる方法により人物探索を行ってきた[8]. しかし近年になり、GPUなどの普及と高機能センサの普及によって、深層学習によるタスククリアが現実的となってきた. 本研究ではこの人物の発見を深層

学習の一手法である CNN を用いて行き, ”いかにして高精度に特定人物(探索対象人物)を識別するか”を目的として開発を行った.

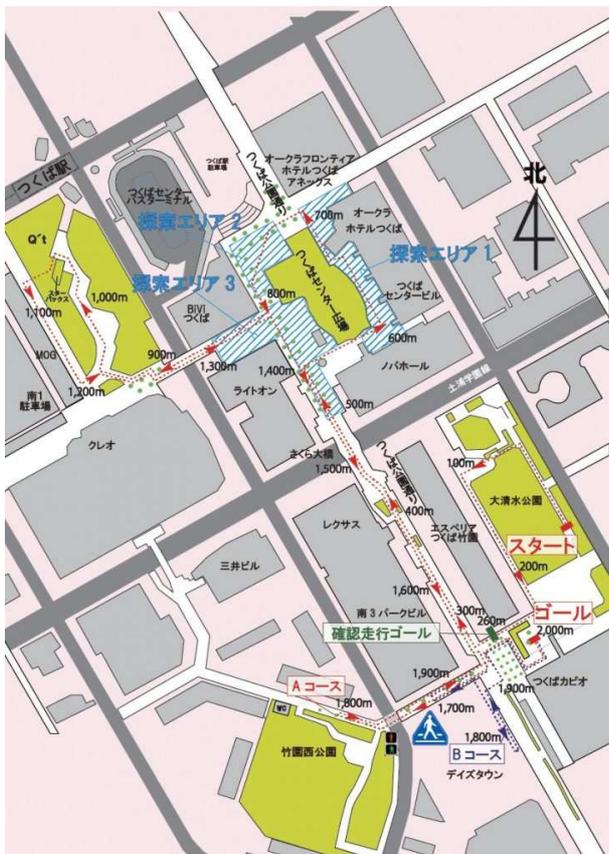


Fig.1 Tsukuba Challenge 2017



Fig.2 Mobile Robot and Target Person

3. 特定人物識別のための仮定と問題の記述

つくばチャレンジ 2017 の人物探索エリアのルール[6]にしたがい, 以下のような仮定を設ける.

(A1) Fig.3 に探索の対象となる人物を示す. 人物は, 緑色のジャンパーの上に青または橙色のベストを着用しており, 橙色のキャップを着けている.

(A2) 探索対象人物の発見の補助として探索対象人物の脇に黄色と橙色のデザインで, 高さ約 90cm の立て看板が置かれている.

以上の仮定のもと, 自律移動ロボットの人物探索のために以下の問題点を述べる.

(P) どのようにしてロボットの取得画像から特定人物を特定するか.



Fig.3 Target Person

4. 提案するシステム

(1) 開発したロボット

Fig.4 に本研究で作成した移動ロボットを示す. ロボットは対向二輪駆動ロボットを使用し, ロボット上方に Theta S を搭載している. Theta S は, 全方位を一度に撮影できるためロボット自体を回転させることなく, 周囲 360 度の画像を一度に取得できる. Fig.5 に学習時や探索時に使用した全方位画像をパノラマ展開したものを示す.

ロボットは, 自律走行系の処理をするメインコンピュータと, 人物探索を行うサブコンピュータの 2 台構成としている. Table.1 にそれぞれのコンピュータのスペックを記載する. メインコンピュータは, ロボット自体の制御, および探索対象人物の識別に直接かかわらないパノラマ画像への展開および人物識別範囲の限定処理などを行う. サブコンピュータは, CNN を得意とする GPU 演算に特化した組み込み PC である JetsonTX1 を採用し特定人物識別処理を行う.

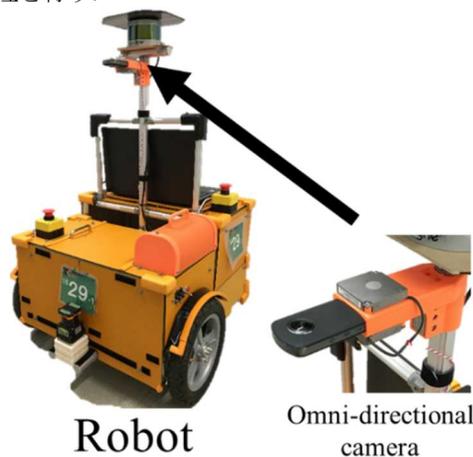


Fig.4 The Developed Mobile Robot



Fig.5 Panoramic Image

Table.1 Computer Specs

	Main Computer (FRONTIER FRNXW610/c)	Sub Computer (JetsonTX1)
OS	Ubuntu14.04	Ubuntu16.04
CPU	Intel Core i7 2.5GHz	ARM A57 1.73GHz
DRAM	16GB	4GB
GPU	None	Maxwell CUDA core 256

(2) 学習フェーズと探索フェーズ

開発したシステムは大きく分けて学習フェーズと探索フェーズの2つから構成される。Fig.6 に詳細な図を示す。

学習フェーズでは、得られた画像を用いて、GPU を搭載した PC で、CNN により探索対象人物を学習させる。

探索フェーズでは、学習フェーズで学習した分類器を移動ロボット上のサブコンピュータ (JetsonTX1) を用いて探索対象人物を識別する。探索エリア内に入ったと判断したとき、移動ロボット上の得られた画像は、メインコンピュータから送信されサブコンピュータ内の分類器により探索対象人物の有無を判断する。もし、有と判断した場合には、移動ロボットを近くまで移動させる。

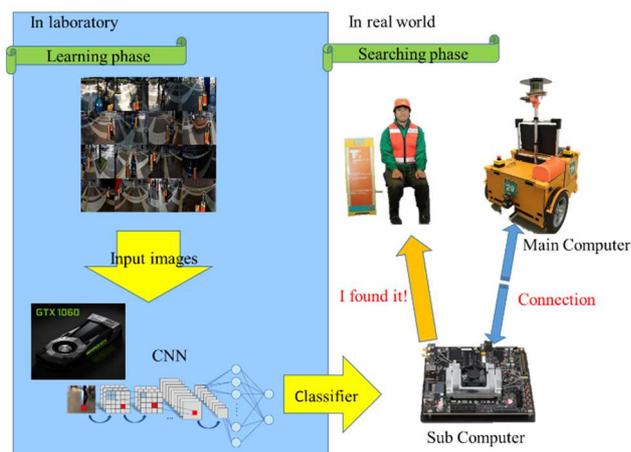


Fig.6 Learning Phase and Searching Phase

(3) パノラマ画像の人物識別範囲の限定

一般には、CNN などの深層学習において、画像から物体認識を行う場合、画像上をスライドする窓を設けて、窓ごとに切り出し、分類を行う Sliding window 法が取られる。しかし、この方法は窓の大きさによっては計算コストが膨大になる可能性が高く、リアルタイム処理には向かない。

本研究では、仮定より探索対象人物(緑・青・橙)と立て看板(橙)の色が特定されていることを利用し、パノラマ画像内での人物識別範囲の限定を行う。Theta S から得られた画像は、メインコンピュータに取り込まれ、パノラマ変換される。変換された画像上から緑・青・橙の色の有無を判断し、予め人物の場所を想定される方向の特定を行う。Fig.7 に示すように人物検出の候補となる画像サイズは、メインコンピュータの処理速度や判別精度を考慮し、130x130pixel 程まで切り取り、サイズダウンを行う。



(a)Target with Blue Vest

(b)Target with Orange Vest



(c)Cutting Image

Fig.7 Search Target Candidate and Cut Image

5. Convolutional Neural Network

本研究では深層学習の一手法である CNN(Convolutional Neural Network)を採用した。CNN は画像に様々なフィルターをかけ、最終的にニューラルネットワークで画像を分類する手法である。Fig.8 に学習に用いたモデルを示す。

より詳細には、学習しながら更新されるカーネルフィルターを画像にかける convolution 層と画像サイズを小さくすることで情報を捨て、解像度を小さくする pooling 層の二層を交互に繰り返すことで細かい特徴を抽出することができる。Fig.8 に示した今回のモデルでは3層の積み重ねを行う。最終的にニューラルネットワークに入力され、出力層でクラスに分類される。今回は、TensorFlow[9] を用いてモデルの実装を行った。ニューラルネットワークへの入力層は17x17x128層、中間層が1024層、出力層が2層とした。

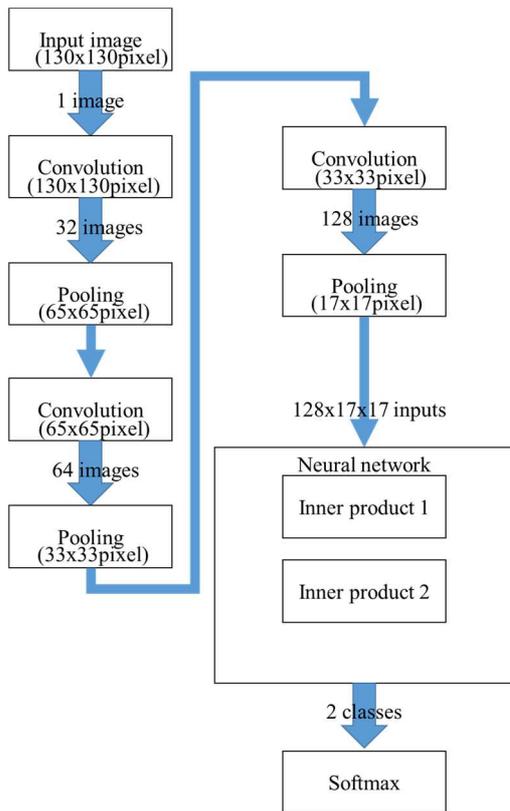


Fig.8 The Implemented CNN Model

6. データ拡張

CNNなどの深層学習において最も問題となるものは学習データの収集方法である。基本的には様々な環境下における学習データを収集することが理想的であるが、画像の撮影時間やクラス分けを行う時間が膨大となり、現実的ではない。そこでFig.9に示すとおり、スクリプトを用いて合成写真のように学習データを自動で生成し、学習データの水増しを行う手法を採用した[10]。

具体的には、つくば市内の探索区間内の探索対象人物を様々な方向から撮影した後、画像のトリミングを行い、背景を削除した探索対象人物のみの画像を用意する。これに加え、探索区間内をランダムに撮影を行い、背景画像も用意する。背景画像に対してランダムな位置設定とサイズの拡大縮小を行った探索対象人物の画像を背景画像に自動合成することによって、任意の枚数だけ、画像を生成することに成功した。この手法の利点として

- 1)学習データに使用する撮影時間が節約できる。
- 2)背景に対する過学習の抑制につながる。
- 3)学習モデルの対象の位置やサイズの変化に対するロバスト性の向上が見込める。

などが挙げられる。



Fig.9 Real Image and Composite Images

7. 実験

(1) データの生成

Table.2につくばチャレンジ2017で撮影した動画から、学習データとテストデータの画像抽出を行った。撮影した画像は約6000枚の晴れ、雨などの天候の違い、場所の違い、通行人の有無を含んでいる。この画像データ群から重複や近い状況のない約1500枚の画像を抽出し、データ拡張のベースとした。データ拡張により約1500枚から探索対象人物が写っていない6212枚の合成画像と、写っている14150枚の合成画像の合計約20000枚の学習データの生成を行った。

なおテストデータも同様に約6000枚の画像データ群から、学習のベースとなる画像と重複しない、探索対象人物が写っていない100枚と、写っている100枚の合計200枚の画像を選びだした。

Table.2 Learning Data and Test Data

(a) Learning Data	
No Target Data	6212
Target Data	14150 (All Composite Images)
Total	20362

(b) Test Data	
No Target Data	100
Target Data	100 (All Real Images)
Total	200

(2) 学習

学習フェーズでは GPU を搭載したデスクトップ PC を用いて学習を行う。Table.3 に学習に用いた PC のスペックを示す。学習はバッチ学習を用いて重みやバイアスの更新を行う。データ拡張をした約 20000 枚の画像を用い、バッチ枚数は 200 枚として CNN で学習した。今回は 500000 ステップの学習を行い、平均して 49 時間程度であった。Fig.10 に学習データの正答率の変化を示す。生データをシアンで、傾向を見るために平滑化したデータを赤の破線で示す。約 200000 ステップでほぼ 100% に収束している。

Table.3 Computer Specs

Training Computer (FRONTIER FRGXH110)	
OS	Ubuntu16.04
CPU	Intel Pentium 3.3GHz
DRAM	16GB
GPU	GeForceGTX1060 6GB

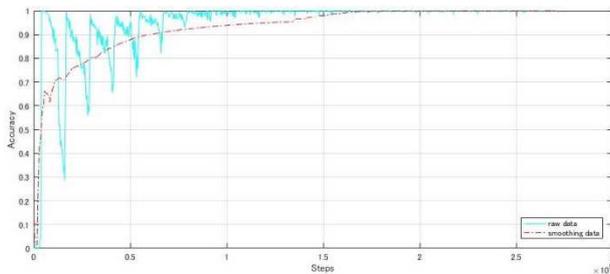


Fig.10 Accuracy of CNN

(3) 検証

学習した分類器は、画像を入力すると探索対象人物がいる場合、精度(0~100%)の数値で結果が出力される。合成画像において探索対象人物の大きさが分類器の精度に大きな影響を与える。本検証では、テストデータ 200 枚を用いて、実際に学習した 3 つの分類器がどの程度の精度で探索対象人物を判別できるのか検証を行った。

Fig.11 にテストデータを分類器に入れた場合の出力結果を示す。積立グラフは、縦軸がテストデータ数、横軸は、左から探索対象人物の画像を背景画像に合成する際の縮小率 40%、25%、5%の順での精度を表す。赤色で示したバーは、探索対象人物がいる画像 100 枚の結果であり、いない場合の結果を青色で示す。精度 100%が多いほどよく、グラフ形状は右上がりの単峰性を持つことが望ましい。

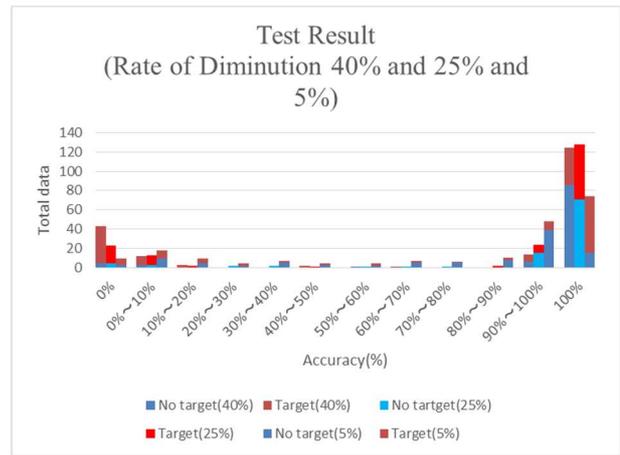


Fig.11 Test Result

(Rate of Diminution 40% and 25% and 5%)

Fig.11 より、3 種類のグラフは右上がりの傾向を持つ。不正解である精度 0%のところを見ると、縮小率 40%では 200 枚中 43 枚、縮小率 25%では 23 枚、縮小率 5%では、9 枚が不正解となった。一方、正解である精度 100%のところは、縮小率 40%では 200 枚中 124 枚、縮小率 25%では 128 枚、縮小率 5%では 74 枚であった。

縮小率 40%では、不正解である精度 0%の結果が、200 枚中 43 枚であり、他 2 パターンに比べ多い。縮小率 5%では不正解である精度 0%の結果が 200 枚中 9 枚と少ないが、正解を示す精度 100%の結果を見ると、200 枚中 74 枚と余り多くない。そのため縮小率 40%、25%、5%の 3 つの中では、縮小率 25%がより高い判別ができる。

分類器からの出力は、精度 0~100%の数値で出力されるため、探索対象人物が有と判別するか否かは、適切な閾値を設ける必要がある。縮小率 25%を採用した場合、精度 20%から不正解が多くなるため、精度 20%を閾値として、探索対象人物が有と判別することができる。このときの正解率は 81.0%であった。

(4) 考察

探索対象人物が存在するテストデータで、合成写真の縮小率 25%で学習した分類器が精度 100%で特定できた例は 57 枚あった。Fig.12 に、精度 100%で判別できた特徴のある 4 枚を示す。No.1 と No.4 は、探索対象人物のみであり、ペアとなる立て看板が写っていないが判別できた。また、No.2 と No.4 は、探索対象人物が見切れているが判別できた。これはデータ拡張で生成した学習データの中に、見切れのある画像を使用し学習したからであると考えられる。また No.3 は、探索対象人物およびペアとなる立て看板も全て写っているが、日影にいたため、人間の目でも誤認しやすい画像であると考えられるが、判別できた。これは様々な条件下での探索対象人物の画像を用意したことためと考えられる。



Fig.12 Success Example on Accuracy 100%

一方で、探索対象人物が存在するテストデータで、学習した分類器が精度 0%で人物が確実にいないと判断したものは 19 枚あった。Fig.13 にその中の特徴のある 4 枚を示す。

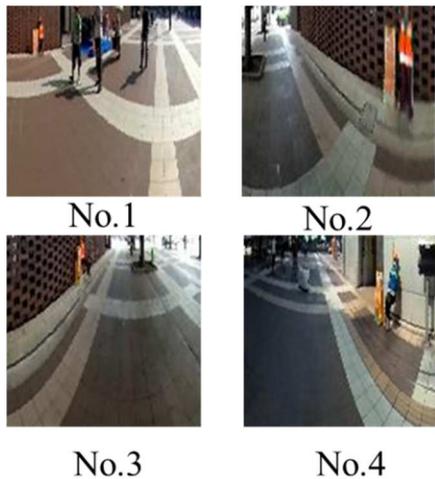


Fig.13 Failure Example on Accuracy 100%

No.1 と No.3 は画像サイズに対して探索対象人物の大きさが 25%未満であったため、判別に失敗したと考えられる。また、No.2 は撮影ブレの影響であると考えられる。移動ロボットは、移動しながら撮影するため、画像は、頻繁にブレが発生する。同様に精度 0%と判別した例は、19 枚中 4 枚存在しており、これは CNN における畳み込みの際に形状が変化してしまうため、判別できなかったものと考えられる。また No.4 はごく少数であったが、画像内に明暗が存在しているため、失敗したのと考えられる。

失敗例と成功例において、本手法は見切れに強く、また特定位置に対する過学習が見受けられないことから、探索対象人物の位置に対して十分に許容的であることがわかる。

8. 結論

本研究ではつくばチャレンジにおける人物発見タスクの CNN による特定人物識別アルゴリズムについて述べた。開発したシステムでは、学習フェーズと探索フェーズにわけ、メインコンピュータとサブコンピュータに処理を分割することで、計算コストを下げ、リアルタイム処理が可能であることを実験により検証した。合成画像を用いることで、実際の画像を撮影する時間を節約した。学習した分類器は、未学習のテストデータを用いた場合でも、ある程度、特定人物の有無を判別可能であることを確認した。現状での学習した分類器の正解率は、閾値を精度 20%にした場合、81.0%であるが、ファインチューニングなどの一般的な精度向上のための手法を適用することでさらなる精度向上が見込める。

9. 今後の展望

本研究は発展の余地は十分にあり、既存のシステムに容易に組み込めることから、有用性は十分にある。今後は分類器の精度向上を行い、加えてロボットの性能を上げ、さらなる識別の速度向上にも取り組んでいきたい。

参考文献

- 1) 松尾 豊：人工知能は人間を超えるか ディープラーニングの先にあるもの, KADOKAWA/中経出版, 2015
- 2) Wang Zhiqiang and Liu Jun : A Review of Object Detection Based on Convolutional Neural Network, Proceedings of the 36th Chinese Control Conference, pp.11104-11109,2017
- 3) Self-Driving Car Project <https://waymo.com/> 2018/02/11 アクセス
- 4) M. Paolanti, M. Sturari, A. Mancini, P. Zingaretti, and E. Frontoni: Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning, in Mobile Robots(ECMR), 2017 European Conference on. IEEE, 2017.
- 5) Yiyu Sun, Tonghua Su and Zhiying Tu : Faster R-CNN Based Autonomous Navigation for Vehicles in Warehouse, 2017 IEEE International Conference on Advanced Intelligent Mechatronics, pp.1639-1644, 2017
- 6) つくばチャレンジ 2017 <http://www.tsukubachallenge.jp/> 2018/02/09 アクセス
- 7) TSUKUBA CHALLENGE 2016 参加レポート集, 巻末資料,2016
- 8) Fukuda Yoshitaka, Kobayashi Kazuyuki, Watanabe Kajiro, Kinoshita Tetsuo: Target Object Classification Based on a Fusion of LIDAR Range and Intensity Data, SICE Annual Conference 2014,September 9-12, Hokkaido University, Sapporo, Japan,pp.1312-1317, 2014
- 9) TensorFlow <https://www.tensorflow.org/> 2018/02/09 アクセス
- 10) TSUKUBA CHALLENGE 2016 参加レポート集, pp2-8,2016