

テーマ番号	EP165			
プロジェクト テーマ	和文	深層学習を用いた適応的な認識システムの構築に関する研究	指導教員	長田 茂美 教授
	英文	Research for constructing adaptively recognition system with Deep Learning.		
プロジェクト メンバー	4EP2-20 塩浜 健 (Takeru Shiohama)			

Abstract 人とのコミュニケーションを行う事を前提としたパーソナルロボットは、人の行動や表情だけでなく、環境変化に頑健なシステムが求められる。しかし、人とロボットの間には依然としてコミュニケーションを行うための能力の差が大きい。そこで、本研究では、概念共有を行うコミュニケーションの一つの、人からロボットへの物体教示を想定し、多様な物体を認識するために、事前に認識可能な物体数を規定せず、深層学習を用いた複数物体認識が可能なシステムの構築を目指す。リアルタイムに複数物体の学習を行うために、物体教示時にラベルとなるデータを RGB 画像にチャンネルとして付与することで、深層学習による複数物体の認識を行う手法を提案し、評価実験により複数物体の学習および、教示物体領域推定の可能性を示した。

Keywords Deep learning, Convolutional Neural Network, Image processing, HCI, Object recognition.

1. はじめに

人とのコミュニケーションを行う事を前提としたパーソナルロボットは、人とのコミュニケーションを通して人間の意図を認識し、機能や情報を提供することで人の生活をサポートするものである。しかし、実際はロボットが人の生活環境を認識できない場合、人の動作や環境の変化に適應できず、規定された動作を実行するだけに留まってしまうため、コミュニケーションは表層的なものとなる。

人とロボットがより深いコミュニケーションを行うためには、ロボットが人の生活環境の多様性に適應する必要がある。そこで、環境認識の一つとして、人から物体を教示されるシーンを想定し、事前に認識する物体を規定せずに、リアルタイムに認識可能なシステムが求められる。

事前に認識する物体を規定しない認識システムとして、川崎らによってリアルタイムに認識を行うシステムが提案されている^[1]。CNN (Convolutional Neural Network) を用いて物体を学習する際に背景も合わせて学習することで、事前に認識物体の規定や学習を必要とせず、教示物体と背景を区別することに成功している。

このことから、単一の物体を適応的に認識する手法が示されるとともに、複数の物体に適用することが次なる課題として挙げられている。また、Rosanne らによって、CNN が画像上の物体の位置を認識する際に、座標情報をチャンネルとして入力する有効性が報告がされている^[2]。

本研究では、Rosanne らの知見と川崎らの手法に基づき、CNN への入力画像にチャンネルとして情報を付加することで、事前に認識する物体を規定せずに、リアルタイムに複数物体の認識を行う手法を提案する。

2. 提案手法

川崎らの手法を基にし、背景推定および複数物体学習、認識を行うシステムを構築した。図 1 に物体学習時の処理、図 2 に物体認識時の処理の概要を示す。

2.1. 背景推定

背景推定では、教示された物体とそれ以外の背景となる領域をセグメンテーションするために、CNN を用いて背景を学習する。これにより、机や壁、床等の普遍的な背景を推定する事が可能になり、背景差分よりも正確な教示物体の抽出が可能になる。

2.2 物体学習

物体学習では、CNN を用いてリアルタイムに複数物体の学習を行うために、物体教示中のカメラ画像と背景推定で推定した背景差分を行い、教示物体を抽出する。抽出し

た教示物体画像を小矩形で切り出し走査を行う、このときに、小矩形内の教示物体領域と背景領域のピクセル数の比率を算出し、教師データとする。これを、教示物体領域を示す確信度と呼ぶ。また、複数物体の認識を可能にするために、入力データとなる切り出した小矩形画像に、物体の種類を示すラベルチャンネルを追加することで、教示物体と背景領域のセグメンテーションに加え、物体ラベルの学習を行う。

ラベルチャンネルの生成方法を図 3 に示す。目標物体毎に 0 から 1 の範囲でラベルを割り当て、チャンネル内の全てのピクセルに対してその値を与えることで生成する。

2.3 物体認識

物体認識では、2.2 で学習した CNN を用いて目標物体の位置および複数物体を認識する。2.2 の学習時と同様に、カメラ画像を小矩形で切り出し、ラベルチャンネルを追加した画像を CNN へ入力し、確信度を出力する。次に、出力した確信度を小矩形毎にグレースケールへと変換し確信度画像を生成する。その後、生成された確信度画像から、閾値以上の確信度を示す領域を外接矩形で描画する。これによって、目標物体位置を認識することができる。

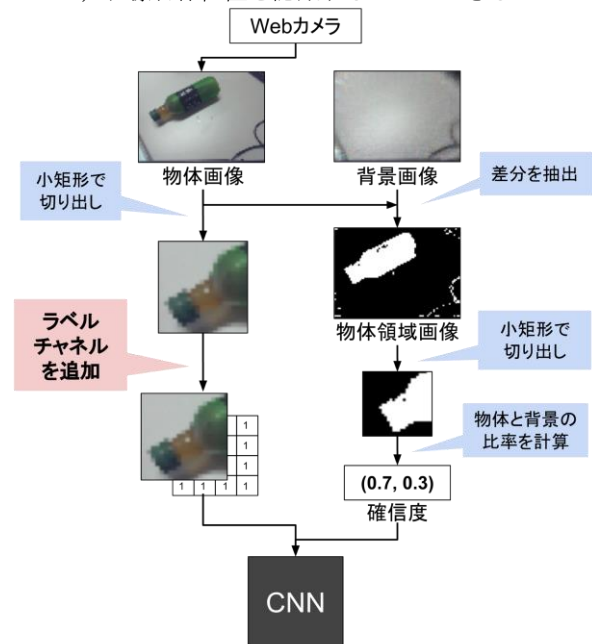


図 1 物体学習時の処理
Fig.1 Object learning processing

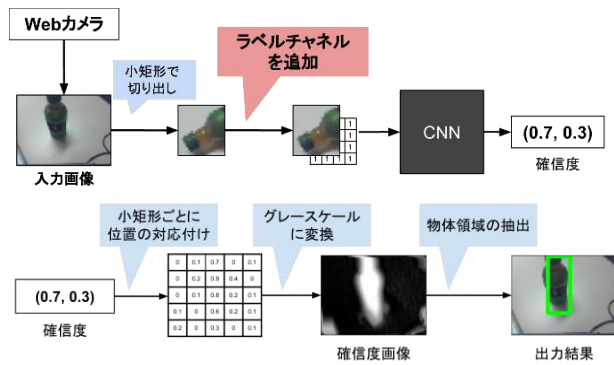


図2 物体認識時の処理
Fig.2 Object recognition processing

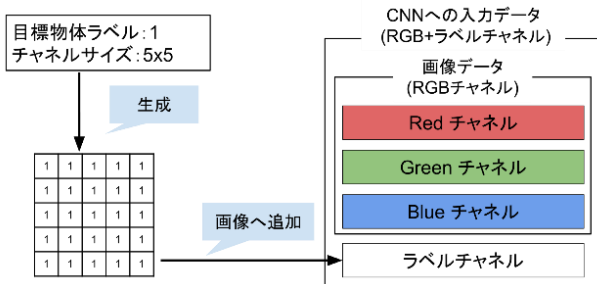


図3 ラベルチャネルの生成と追加
Fig.3 Generating and inserting of label channel

3. 評価実験

提案システムの有効性を評価するために、3種類の教示物体と Web カメラを利用し、複数物体の学習および認識の評価実験を行った。

複数物体の学習および認識が実現できているかどうかを以下の実験により検証する。

- (a) 単一物体の認識
- (b) 複数物体の認識
- (c) 教示していない物体を含む認識

実験環境は、照明があり十分に明るい室内で、机から高さ 40cm に Web カメラを設置し、教示物体は机の上に配置した。また、教示物体は単一物体認識、2種類の物体認識と、学習していない物体の3種類を用意した。使用した物体は、一般家庭での卓上を想定し、財布とウェットティッシュとマウスを使用した。

3.1 結果

図4に3種類の実験結果を示す。

実験(a)の単一物体の認識は、図5(a)より財布のみを学習し、背景領域と財布の領域を区別できている。また、学習していないウェットティッシュは、背景として認識していることが、確信度画像よりわかる。

次に、実験(b)の複数物体の認識は、図5(b)より財布とウェットティッシュをそれぞれ学習し、認識時にラベルチャネルを変更することで、ウェットティッシュの認識から財布の認識に変更できている事がわかる。このことから2種類の物体を区別して学習できている事がわかる。

実験(c)では、教示していないマウスをカメラ画像内に含んでいても、3つの物体の中でウェットティッシュの確信度のみが鮮明に出力されており、学習していない物体と学習した物体を判別できている。

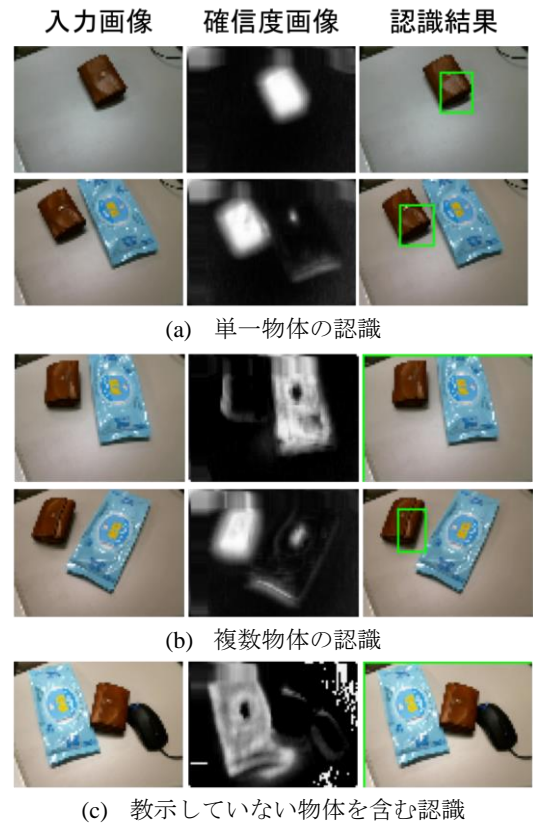


図4 実験結果

Fig.4 Experiment results

3.2 考察

実験結果から、2種類の物体を区別できる事がわかった。このことから、ラベルチャネルを変更することで、出力する確信度を変更できている、3,4種類と認識する物体を増やしていくような拡張性が期待できる。

しかし、目標物体とは別の教示物体の確信度が背景よりも高く出力されてしまう問題がある。この問題に対して、各物体の確信度画像を用いて差分処理をすることで解決を試みたが、図5(b)のように矩形抽出による認識結果の表示に失敗している。これは、差分処理によって目標物体の確信度も低くしてしまっているためだと考えられる。

この問題を解決するためには、矩形抽出に用いる閾値を確信度画像のヒストグラムから動的に決定するなどのアプローチが考えられる。

4. おわりに

事前に認識物体を規定せずに、リアルタイムで複数の教示物体の学習と認識を行う手法を提案した。評価実験によって、CNNへ入力する画像データに物体ラベルを表すチャンネルを追加することで複数の物体の確信度に差異を生み出せることを確認した。今後は、複数物体をより明確に判別するために、閾値を動的に決定することによって、より適応的なシステムへ拡張を図っていく予定である。

参考文献

[1] 川崎 邦将, 長田 茂美 “コミュニケーションロボットのための畳み込みニューラルネットワークを用いた物体認識手法の検討”, FIT2016, (2016).
[2] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, Jason Yosinski, “An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution”, NeurIPS 2018, (2018).