# 畳み込み自己符号化器を用いた対話的学習に基づく災害対応のための画像認識システム

ARNOLD Solvi[*1], 山崎 公俊 [*1]

# Convolutional Auto-Encoder based Interactive Learning of Image Understanding for Disaster Environments

Solvi ARNOLD[*1] and Kimitoshi YAMAZAKI[*1]

[*1] Department of Mechanical Systems Engineering, Shinshu University
4-17-1 Wakasato, Nagano city, Nagano 380-8553, Japan

This work aims to adapt neural network based learning of object recognition for practical use in disaster robotics applications. While object recognition has seen rapid advances in recent years, few systems are capable of quickly acquiring new object classes in unknown environments on basis of sparse data. This shortcoming is especially clear in the context of disaster robotics, as operation environments are unpredictable, relevant learning data hard to obtain, and fast response time often crucial. We develop a semi-supervised on-line learning system capable of acquiring new categories on the spot from limited human input. Heart of the system is a fully convolutional autoencoder, trained by a combination of unsupervised learning and "representation nudging", a novel training strategy for improving discrimination of categories.

***Key Words*** : Disaster robotics, Object recognition, Deep learning

## 1. はじめに

災害が起きたときの対応として，被害状況の把握や被災者救助などの目的にロボット技術を用いる動きが加速している［1］．本研究の目的は，被災現場における活動を補助するための画像認識システムの構築である．遠隔移動体にカメラを搭載し，そこから得られる画像に対して認識処理をおこない，その結果を操作者等に提示することで，環境認識支援や捜索支援を目指す．環境認識手法としては［2］［3］などが高い性能を示している．筆者らの研究グループでも，同様のアプローチを検証した経験がある［4］．しかし，従来手法では，事前の学習が不可欠である．災害対応においては，未知のカテゴリに対しても短時間で識別できるようにすることが求められる．以上を踏まえて，本稿では新たな画像認識手法を提案する．

## 2. 画像認識システムの構成

Fig. 1 に提案手法の全体像が示されている．原則として連続的に読み込まれる映像(ファイル若しくはカメラ映像)を対象にする．映像に映る物体にユーザが UI からラベルを付与することで，認識の対象となるカテゴリとラベル付きデータが得られる．以下処理の流れを説明する。

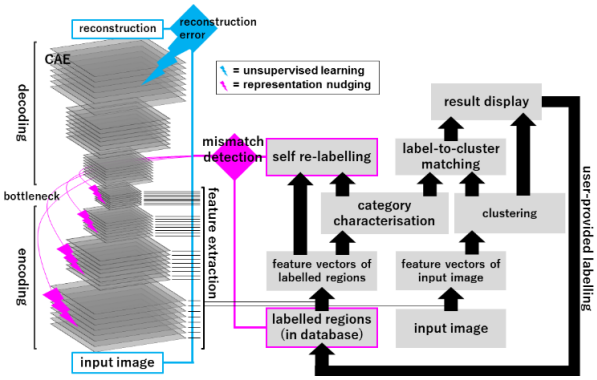


Fig. 1. Schematic system overview. The systems of black, blue and pink arrows are responsible for classification, unsupervised learning and representation nudging, respectively.

提案手法は畳み込み自己符号化器（Convolutional Auto Encoder, CAE）［5］をコアとする。この CAE の役割は入力映像の特徴量の抽出である．入力映像の毎フレームに対して教師なし学習を行う．この教師なし学習とは，従来通り，入力層に入る画像が出力層で再現されることを目標にした学習処理である．その狙いは，映像の特徴を CAE に獲得させることである．画像に対して CAE の中間層に現れる活性化模様をその画像の特徴量として利用する．自己符号化器を用いたこの特徴量抽出法は以前から有効と知られているが，ここ

での特徴量抽出方法は従来と二つの点で異なる．1）従来はネットワークの瓶首層の活性化模様のみが特徴量ベクトルとして利用されるが，ここでは認識の柔軟性を高めるために符号化部分の全隠れ層から特徴を採取する．2）従来は一枚の入力画像に対して一個の特徴量ベクトルが得られる．ここでは空間的配置を含めた同時複数物体認識(scene parsing)を目的とするため，畳み込み層にわたって空間的関係が保たれるという事実を利用して，一枚の画像から任意の数の画像領域の特徴量ベクトルを同時に抽出する方法を導入する．

CAE を用いて抽出される特徴量ベクトルを画像の分割と画像領域の自動的ラベル付与(認識)に利用する．先ずは現フレームの全画像領域の特徴量ベクトルに対して階層的クラステリング［6]を行う．特徴が類似する画像領域が同じクラスタに入るため，画像に映る物体や表面がそれぞれクラスタとして浮かび上がる傾向がある．クラスタ毎に色付けした画像を UI 上に表示する．表示される階層の選択が UI から可能である．対象物体とクラスタの当てはまりのよい階層をユーザが選択して，そのクラスタにラベルを付与することで，クラスタに含まれた画像領域がそのラベルと共にラベル付きデータとして保存される．必要に応じてクラスタを手動で修正することも可能である．

階層的クラステリングと，特徴量ベクトルと，ラベル付きデータを利用して物体認識を行う．ラベル・クラスタ毎に，そこに所属する画像領域の特徴量ベクトルの多変量正規分布を計算する．ラベルとクラスタのそれぞれの分布の当てはまりを定量化し，クラスタの各カテゴリに対する所属確率の尺度を得られる．これを利用して，クラステリングの全階層からカテゴリとの当てはまりのよいクラスタを収集する．UI から設定可能な所属確率閾値を超えたクラスタを一枚の画像にまとめて、ラベル付きで認識結果として表示する．ここで時間的スムージングを適用して，フリッカーを抑えて結果を滑らかにする．

カテゴリ判別性能を向上させるため、「表現ずらし」と名付けた学習処理を導入する．CAE は原則としてに教師なしで学習するため，認識の対象となるカテゴリの見分けに有効な特徴量が抽出される保証はない．有効な特徴量が抽出されない場合は，異なるカテゴリに所属するデータ点が特徴量空間の同じ領域に混在する状態が生じる．この状態を検知し，異なるカテゴリに所属する点の特徴量空間内の距離を取り入れた損失関数を用いた学習処理で，CAE にそのカテゴリの違いが特徴量に現れるように学習させる．これにより，上記の認識処理がより有効になり，性能が向上する．

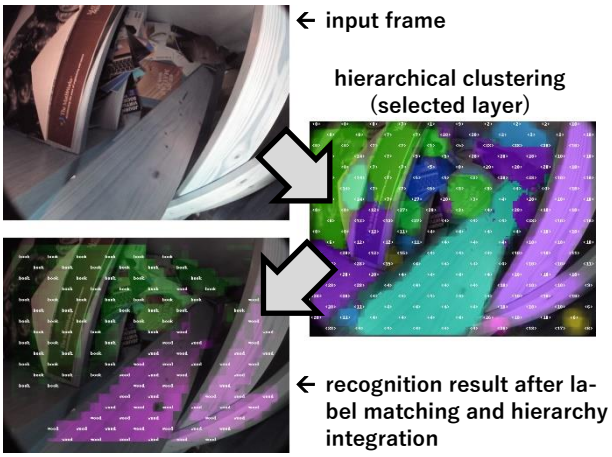


Fig. 2. Example result in simulated disaster scenario. The system was trained to identify wooden structures and books/magazines. Footage courtesy of Fukuda et al. [7].

## 3. 結果

災害現場と模擬災害現場の映像を対象にした予備実験を行った(Fig. 2)．同時に対応できるカテゴリの数や認識できるカテゴリの抽象性には限界があるが，学習データが全くない状態からでも，15 分程度で基本的な認識器の立ち上げが可能であることを確認した．

## 参 考 文 献


[1] “革新的研究開発推進プログラム ImPACT - タフ・ロボティクス・チャレンジ”, www.jst.go.jp/impact/program/07.html [アクセス日: 29 10 2016].

[2] C. Couprie, C. Farabet, L. Najman , Y. LeCun, “Indoor Semantic Segmentation using depth information,” *International Conference on Learning Representations*, 2013.

[3] P. Pinheiro , R. Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling,” *Proceedings of the 31st International Conference on Machine Learning*, 2014.

[4] S. Arnold , K. Yamazaki, “Patch-wise Object Recognition for a Mobile Robot by means of a Convolutional Neural Network,” *第33 回日本ロボット学会学術講演会*, 2015.

[5] M. Jonathan, M. Ueli, C. Dan , S. Jürgen, “Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction,” *Artificial Neural Networks and Machine Learning – ICANN 2011, Lecture Notes in Computer Science*, 2011.

[6] M. Ankerst, M. M. Breunig, H.-P. Kriegel , J. Sander, “OPTICS: Ordering Points To Identify the Clustering Structure,” *ACM SIGMOD international conference on Management of data*, 1999.

[7] J. Fukuda, M. Konyo, T. Eijiro , S. Tadokoro, “Remote Vertical Exploration by Active Scope Camera into Collapsed Buildings,” *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.