7U-07

深層学習を用いた人の行動認識と感情生成

乾 祥大 † 西出 俊 ‡ 康 鑫 ‡ 任 福継 ‡

† 徳島大学 工学部 ‡ 徳島大学 大学院社会産業理工学研究部

1. はじめに

ロボット工学の発展により，人間とインタラクションするコミュニケーションロボットの開発が盛んに行われている [1].「パロ」に代表されるセラピーロボットでは，タッチセンサやマイクから得られる情報をもとに振る舞い制御を行うことで感情を表現することができ，うつ病患者及び高齢者に対しての臨床・実証実験の結果からセラピー効果の有効性が示されている．このような背景を受け，本研究では人間の行動に付随する感情情報を読み取るシステムを構築することを目標としている．本稿ではその基礎システムとして，画像情報に基づく人間の手形状を深層学習モデルによって認識するシステムについて報告する．

2. Single Shot MultiBox Detector(SSD)

本研究では人間の手形状の認識モデルとして，高速かつ高精度に物体を検出することができる Single Shot MultiBox Detector(SSD) を用いる [2]．SSD は画像の特徴量抽出に効果的な深層学習モデルである Convolutional Neural Network (CNN) の一種であり，入力画像内に含まれる物体の種類，物体の矩形範囲であるピクセル座標，確信度を出力する．SSD は画像中の物体を単一のディープニューラルネットワークで検出することができ，GPU を使用することでさらに高速に検出することが可能である．

SSD のアーキテクチャを図 1 に示す．SSD の入力にはカメラから得られた画像列を使用する．モデル構築の前処理として，ImageNet で学習済みの既存モデルを生成する．次に，VGG で生成された各スケールの特徴マップに CNN を多段的に適用し，マルチスケールの特徴マップを生成する．SSD の出力結果には検出した手の座標がバウンディングボックスで与えられ，クラス信頼度とラベルが出力される．クラス信頼度は [0,1] の値をとり，値が 1 に近づくほど検出された手の形状が実際の手の形状と一致している確率が高い．各スケールの特徴マップに適用する CNN のカーネルサイズは一定であるため，サイズの大きい特徴マップでは小さい物体を，サイズの小さい特徴マップでは大きい物体を検出する役割がある．

3. 提案手法

本研究では図 2 に示すラッセルの感情円環モデル [3] に基づいて感情の評価を行う．ラッセルの感情円環モデルとは,「快-不快」「覚醒-眠気」の 2 軸で表される平面上に，全ての感情が円環状に並んでいる円環モデルのことである．2 次元上の各象限にはそれぞれ「喜怒哀楽」の基本感情を割り当てることができる．ラッセルの感情円環モデルでは，幸福と喜びのような類義語は円環上の近接した位置に配置され，幸福と悲しみのような反意語は円環上では対極の位置に配置される．

SSD に学習させる学習用画像データセットには，形状が判別できるように様々な角度の手の形状画像を撮影し，

Human Emotion Recognition and Action Recognition Using Deep Learning *Shota Inui (Tokushima Univ.), Shun Nishide (Tokushima Univ.), Xin Kang (Tokushima Univ.), and Fuji Ren (Tokushima Univ.)*

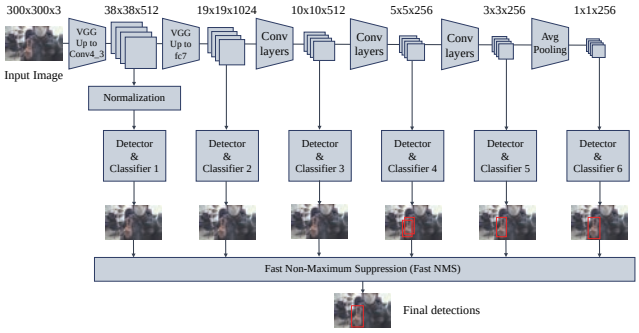

図 1: SSD のアーキテクチャ

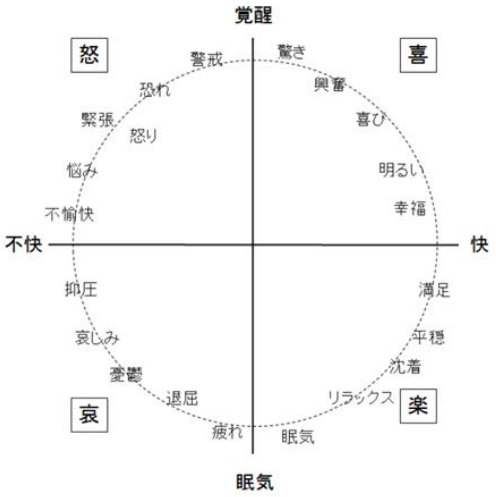

図 2: ラッセルの感情円環モデル

ラベルの付与を行った物を使用した．本研究では，図 3 に示す 4 つの手の形状のラベルごとに感情の種類を設定する．これはラッセルの感情円環モデルにおける「喜怒哀楽」の基本感情に基づいて著者らが恣意的に設定したものである．SSD によって手の形状を認識することで，その形状に付随する感情を求められる．

4. 実験設定

本実験では，深層学習フレームワークである TensorFlow と Keras を用いて提案システムの実装した．データセットは被験者一名の手の形状画像 50x4 種類と，全てのデータセットに対し左右反転したデータを用いて構築した．

モデルの学習を行う際，一つの訓練データを繰り返す回数である epoch 数を決める必要がある．epoch 数が小さすぎると学習が不十分になる一方，大きすぎると訓練精度に対し予測精度が低くなる過学習が起きるため，最適な値に設定することが重要である．予備実験として，epoch 数を 50，80，100，200，300 で検出精度を比較したところ，50〜100 の場合は誤検出数に明らかな違いがあったが，100〜300 の場合は誤検出率の差に大きな違

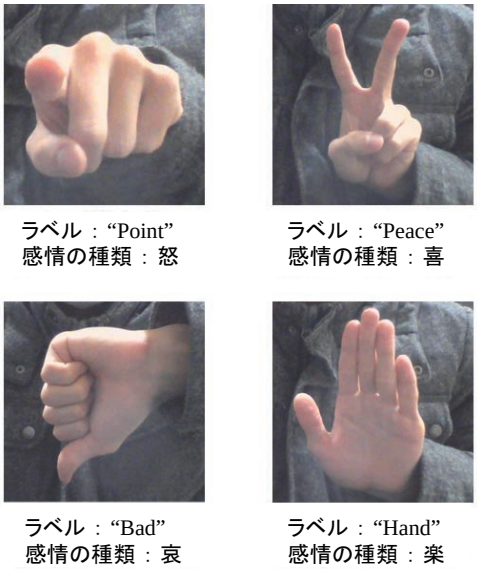

図 3: 各手の形状に対するラベルと感情の種類

いが見られなかった．予備実験の結果を受け，本実験では epoch 数を 100 とした．

本実験では USB カメラを用いて得た画像データに対して手の形状認識を行い，計算が終了するごとに次の画像データをカメラから取得するといった手法でリアルタイムな認識を行った．さらにリアルタイム性を実現するために，ASUS 社の GPU である GEFORCE GTX 1050Ti を使用し，高速化を行った．CPU を使用した場合の平均フレームレートは 2fps であったが，GPU を使用することで 15fps での実験を可能とした．

5. 実験結果

本システムは手の検出座標をバウンディングボックスで出力し，手の形状ラベルと対応する感情の種類を同時に出力する．評価実験の結果，全ての手の形状に対して認識をすることができたが，形状ラベルが誤認識される場合があった．手の形状認識に成功した出力画像例を図 4 に，“Hand”を 2 つの “Peace”と誤認識した出力画像例を図 5 に示す．図 5 の誤認識は手を広げすぎる時に起きる傾向があり，また，まれに手の形状を変更する最中に別の手の形状が認識される場合もあった．

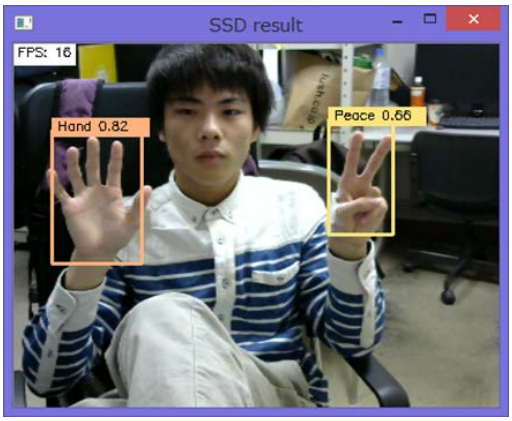

図 4: 実験結果（成功例）

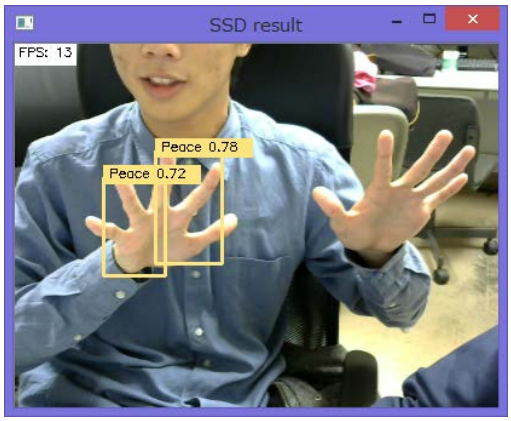

図 5: 実験結果（失敗例）

6. 考察

評価実験の結果，全ての手の形状に対して認識が行えることが確認できた．したがって，データセットの数量は十分であり，また epoch 数は適正であったと考えられる．“Hand” における誤検出は，データセット作成の際に手をどのくらい広げたかを考慮していなかったことが原因であると考えられる．データセットにあらゆる角度で撮影した画像を使用したが，それと同様にあらゆる指の間隔で撮影する必要があったと考えられる．手の形状を変更する際の誤検出は，曖昧な形状であるため一定のクラス信頼度以下は除外する処理，または誤検出が一瞬であるため，ノイズとして無視する処理を加えると改善されると考えられる．

7. おわりに

本稿では SSD を用いた認識アルゴリズムによる手の形状判別を行い，それに付随する感情を認識する手法を提案した．本手法では学習データに 4 種類の様々な角度から撮影された手の形状画像とその反転画像を使用し，epoch 数を実験的に 100 と設定したうえで学習を行った．実験では感情を設定した 4 つの手の形状に対し，手の形状・感情認識を行うことができた．

これまでは手の形状認識と同時に感情を出力するシステムを構築したが，今後は感情状態を連続的なものとして変化させる手法を導入し，ロボットに実装することで感情表現と認識が可能なシステムへと発展させていきたい．

謝辞

本研究は科学研究費補助金，若手研究 (A)(課題番号 16H05877) の支援を受けた．

参考文献

[1] Ronald C. Arkina, F. Masahiro, T. Tsuyoshi, and H. Rika, “An Ethological and Emotional Basis for Human-Robot Interaction,” *Robotics and Autonomous Systems*, Vol. 42, Issues 3-4, pp. 191-201, 2003.

[2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C-Y. Fu, and A. C. Berg, “SSD: Single Shot Multibox Detector,” *European conference on computer vision*, Springer, pp.21-37, 2016.

[3] J. A. Russell, S. Reed, C-Y. Fu, and A. C. Berg, “A Circumplex Model of Affect,” *Journal of Personality and Social Psychology*, Vol. 36, pp. 1161-1178, 1980.

[4] 前田 陽一郎, 田辺 奈々, “生物型ロボットによるインタラクティブ情動コミュニケーションの基礎研究” 計測自動制御学会論文集, Vol.42, No.4, pp.359-366, 2006.