J-014

Deep Convolutional Neural Network による顔画像からの表情識別

Facial Expression Identification by Deep Convolutional Neural Networks

長房 慶† 松本 和幸‡ 吉田 稔‡ 北 研二‡
Kei Nagafusa Kazuyuki Matsumoto Minoru Yoshida Kenji Kita

1. はじめに

知的なマンマシンインタフェースを設計し実現するには，顔の表情認識が重要な課題である．Ekman[1]らは基本 6 表情(怒り，嫌悪，恐怖，喜び，悲しみ，驚き)を基底とし，複雑な表情を表現することを試みた．

本稿では，Deep Convolutional Neural Network(DCNN)を用いることにより，表情識別の精度を向上させることを目的とする．動画サイトから基本 6 表情に「無表情」を加えた 7 クラスの顔画像を独自に収集し，未知の顔画像から顔表情を識別するモデルを DCNN により作成する．さらに，識別した大量の顔画像を用いて，DCNN を再学習(Fine-Tuning)させることで，新たな表情識別モデルを作成し，表情識別の精度向上を試みる．

2. 関連研究

Neagoe[2]らは，顔の表情からの感情認識のための深層学習(Deep Learning)モデルを提案している．2 つの深層神経モデルである CNN(畳み込みニューラルネットワーク)と DBN(Deep Belief Network)に焦点を当て，210 枚の少ない画像データベースをもとに学習し，CNN では平均 65.22%，DBN では平均 59.48%の感情認識率を達成し，CNN と DBN の有効性を比較している．

DCNN では学習すべきパラメータ数が膨大であるため，高精度なネットワークを構築するためには大規模な学習データが必要である．少ない学習データでは汎化能力の高いネットワークを構築することは困難である．本稿では，動画サイトから大量の顔画像を収集し，大規模な画像データベースを作成し，機械学習することで，高精度な識別モデルを作成する．また，精度向上のために識別モデルを拡張させていく．

3. 提案手法

動画サイトから 7 クラスの顔画像を収集し，収集した顔画像を識別するモデルを作成する．識別した顔画像を用いて，DCNN による高精度な識別モデルを作成する．本稿で用いる 7 クラスは「怒り」，「嫌悪」，「恐怖」，「喜び」，「悲しみ」，「驚き」，「無表情」である．

3.1 動画サイトから 7 クラスの表情を持つ顔画像の収集

動画サイトから，番組名や出演者などをキーワードとして，動画を自動収集する．収集した動画から 30 フレームごとに静止画像を切り出し，切り出した画像から 100×100 ピクセル以上の顔画像を抽出する．Deep Learning では，学習データの数が多いほど精度の高いモデルが作成できるとされているため，学習データとして約 70000 枚の顔画像を収集した．各表情を収集する際に，バラエティ番組では「喜び」，ホラー映画では「恐怖」，恋愛ドラマでは「悲しみ」などといった，各表情を多く含んでいると考えられる動画を収集した．

3.2 Caffe による表情識別モデルの作成

以下のような手順で DCNN による識別モデルを作成する．なお，DCNN の学習には Caffe[4]を用いた．

1. ベースラインモデルの作成

収集したすべての顔画像を目視による判断で各表情に振り分けることは作業量が膨大である．最初に，画像検索サイト(Google 画像検索や flickr 等を使用)から各表情につき約 100 枚ずつの顔画像をフレーズ検索および目視により収集する．このようにして集めた合計約 700 枚の顔画像からベースラインモデルを作成する．

2. 大規模な学習データの作成

収集した顔画像約 70000 枚をベースラインモデルを用いて識別し，Fine-Tuning 用の学習データを作成する．得られた学習データを新たな学習データとして機械学習させていくことで大規模な学習データにしていく．

3. 最終的なモデルの作成

作成した大規模な学習データから得られたモデルの性能を評価し，良い評価が得られなければ機械学習を繰り返すことで識別精度を向上させ，最終的に高精度な識別モデルを作成する．

以上の手順を図 1 に示す．

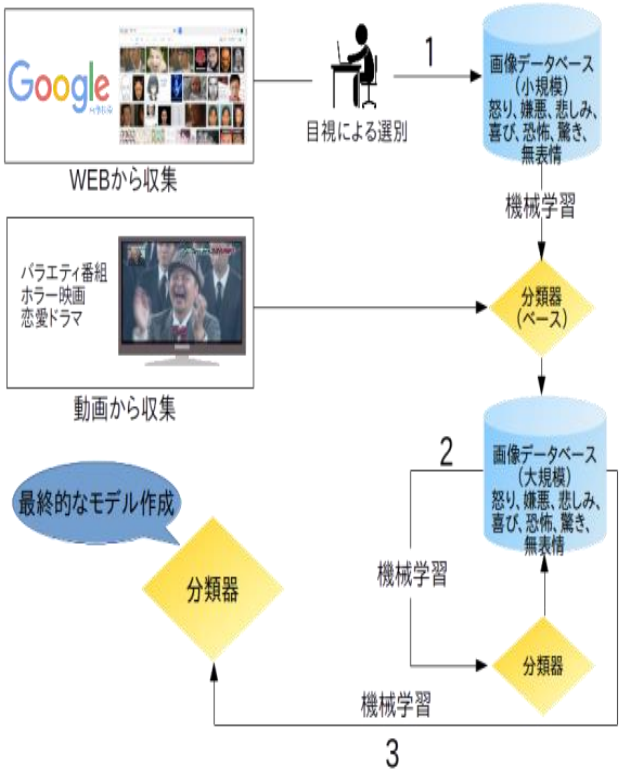

図 1:表情識別モデル作成の流れ

4. 実験

実験データとして各表情に識別された約 70000 枚の顔画像を使用する．また，評価データとして，主に顔の表情認識の研究で用いられる JAFFE 顔画像[4]という日本人女性 10 人による 213 枚の表情ごとに分類された顔画像の中から各クラス 15 枚ずつ，合計 105 枚の顔画像を使用する．図 2 に JAFFE 顔画像の一部を示す．

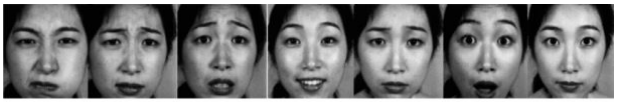

図 2:JAFFE 顔画像の 7 つの表情の例

評価方法として JAFFE 顔画像を表情識別モデルにより識別させ，学習画像の枚数を各クラス約 1000，5000，10000 枚としたときの正答率を比較する．本実験で用いる正答率の計算式を(1)に示す．

$$\text{正答率}(\%) = \frac{\text{正確に識別された画像数}}{\text{識別された画像数}} \qquad (1)$$

実験結果を図 3，表 2，表 3，表 4 に示す．

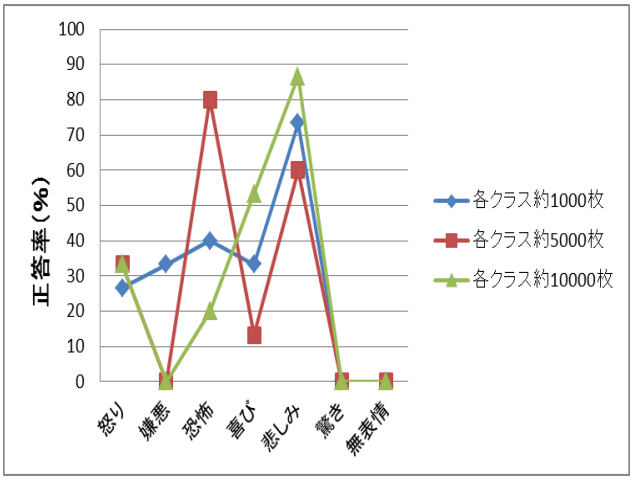

図 3:表情別の正答率の学習画像枚数ごとの比較

表 2:実験の出力結果(各クラス約 1000 枚)

感情	怒り	嫌悪	恐怖	喜び	悲しみ	驚き	無表情	正答率(%)
恐怖	0	0	3	4	8	0	0	20
喜び	0	0	0	8	7	0	0	53.33
悲しみ	0	0	2	13	0	0	0	86.67

表 3:実験の出力結果(各クラス約 5000 枚)

感情	怒り	嫌悪	恐怖	喜び	悲しみ	驚き	無表情	正答率(%)
恐怖	0	0	12	0	3	0	0	80
喜び	0	0	6	2	7	0	0	13.33
悲しみ	0	0	3	3	9	0	0	60

表 4:実験の出力結果(各クラス約 10000 枚)

感情	怒り	嫌悪	恐怖	喜び	悲しみ	驚き	無表情	正答率(%)
恐怖	0	0	6	0	9	0	0	40
喜び	0	1	1	5	8	0	0	33.33
悲しみ	0	0	1	1	11	0	2	73.33

結果として，「恐怖」，「喜び」，「悲しみ」に着目し，図 3，表 2，表 3，表 4 を見ると，「悲しみ」の表情では，約 5000 枚のときより約 1000 枚のときの方が正答率は高くなり，約 10000 枚のときに最も高い正答率が得られた．「喜び」の表情では，約 5000 枚のときより約 1000 枚のときの方が正答率は高く，約 10000 枚のときに最も高い正答率が得られた．「恐怖」の表情では，約 5000 枚のときには正答率が約 1000 枚のときより上がったが，約 10000 枚のときに最も正答率が低くなった．全体的に高い正答率が得られた表情もあったが，枚数を増やすと「悲しみ」の表情に偏って識別されるという傾向があった．

5. 考察

顔画像が「悲しみ」の表情に識別されやすいという結果から「悲しみ」の表情は他の表情と少しずつ類似しており，曖昧さが生じているのではないかと考える．さらに約 70000 枚の顔画像を識別した際に，各表情の枚数にばらつきあることや間違った表情に識別された学習データを用いて，モデル作成を行ったことが原因と考えられる．このことから，より詳しく原因を特定するため作成した学習データを目視などで評価することが必要である．今回扱った顔画像は顔が正面に向いていないものが多く含まれていることから，動画から顔を切り出して収集した顔画像を frontalization[5]を使用し，正面に向けて切り出すことで精度向上が見られるのではないかと考える．学習に用いた顔画像のサイズを統一していないため，前処理によりリサイズされたことで表情が識別できないほど画質が粗くなってしまった顔画像が含まれていることから有効な顔画像のサイズの検討をすることも重要である．

6. おわりに

本稿では，動画サイトから 7 クラスの表情を持つ顔画像を収集し，収集した顔画像を識別するベースラインモデルを人手による判断でラベル付けした学習画像をもとに作成した．さらに，このベースラインモデルにより識別された顔画像を学習データとして識別モデルを拡張させることで，より高精度な識別モデル作成を提案した．今後は，顔画像を正面に向くように切り出し，精度向上が見られるかどうか実験により検証する．

謝辞

本研究の一部は，JSPS 科研費 15K00425, 15K00309, 15K16077 の助成を受けたものです．

参考文献

[1] P.エクマン, W.フリーセン, “表情分析入門 表情に隠された意味をさぐる”, 誠信書房, 2000.
[2] Victor-emil Neagoe, Andrei-petru Brar, Nicusebe, Paul Robitu, “A Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions”, Recent Advances in Image, Audio and Signal Processing, 2013.
[3] Caffe, http://caffe.berkeleyvision.org/
[4] The Japanese Female Facial Expression (JAFFE) Database, http://www.kasrl.org/jaffe.html
[5] Tal Hassner, Shai Harel, Eran Paz, Roee Enbar, “Effective Face Frontalization in Unconstrained Images”, IEEE Conf, 2015.