CNN を用いた顔認証システムの開発と追跡停止に対する評価

脇 一史 †　　森 駿文 ‡　　菊池浩明 †

† 明治大学総合数理学部　　‡ 明治大学大学院先端数理科学研究科

1　はじめに

顔認証カメラによって取得した顧客の行動や履歴情報防犯や商用に活用することが進んでいる．その一方，追跡を停止して欲しい顧客の削除要求や自分の登録データの開示請求などの課題が生じている．

そこで，本研究では，マスクやサングラス，帽子などの外乱によって追跡が停止出来るのかを検証するため，画像識別として主流であるディープラーニングを用いて試験システムを実装し，様々な条件下のもとで被験者の顔を識別する実験を行う．本実験では，TensorFlow 等のフレームワークを用いずに python の計算パッケージである numpy でディープラーニングを実装した．そして，実装した Convolutional Neural Network(CNN) に対して外乱ごと学習させた場合追跡が出来るのかについても明らかにし，セキュリティと生活者のプライバシーの観点で考察する．

2　実験目的

1. CNN による顔認証の精度を明らかにする．
2. 素顔を学習させた場合に，素顔及び, マスク, サングラス, 帽子, マスク＋サングラスの計 5 種類のうち最も識別精度を下げる外乱はなにかを明らかにする．
3. 外乱ごと学習することで, 被験者を追跡できるのか．

3　実験内容

3.1　顔画像データの取得

顔認証システムは明るさや表情， 髪形などの変化に対して汎用性を持つ必要性がある．そのため， CNN で用いる顔画像データを被験者 6 名に対し 1 日毎に 100 枚ずつ異なる時間に顔を撮影し， 顔の検出は，openCV を用いて 取得する範囲を画面上に表示させ， 顔がその範囲内に収まるように撮影した．使用するカメラは mac に標準で搭載されている web カメラを用いる. 取得間隔については, パソコンの前にいる被験者を 3 フレーム毎に顔を上下左右に動かしながら撮影し, 5 日間で計 500 枚取得した. さらに被験者 5 名に対し, 表 3 に示す異なる条件下で認証精度を評価するために, 上記 5 種類について別日にそれぞれ 100 フレーム分の画像を取得した．

表 1　データ拡張

手法	コントラスト調整	輝度変換	ガウシアンノイズ
種類	12% 23% 35% 47% 59% 70%	0.5 0.7 0.9 1.1 1.3 1.5 1.7	2 4
計	6	7	2

3.2　顔画像データの拡張

取得した顔画像に対し， openCV を用いて 112 × 112 にリサイズを行い， 表 1 に示す各パラメータについて 6+7+2 = 15 種類のデータに拡張した． 元画像を含め，一人一種類当たり 8,000 枚， 合計 48,000 枚の顔画像データを作成した．さらに， 画像の顔の位置によって識別さることを防ぐため， 全ての画像をランダムな位置から 96 × 96 に切り出しを行った．

3.3　CNN の構成

本稿では VGG-11[1] と呼ばれる ImageNet Large Scale Visual Recognition 2014(ILSVRC2014) の 1000 クラス識別タスクにおいて，2 位となった識別手法を参考に作成した．VGG はシンプルな構造であり，応用性が高いため多用されている手法である．さらに，本稿では [2] で掲載されている Dropout と呼ばれるニューロンをランダムに消去しながら学習することで，過学習を抑制できる手法を追加した．また，ランダムに学習データを 38400 枚，テストデータを 9600 枚に分け，学習させる回数を表す epoch は 2 回とした．学習した CNN に対し，web カメラでリアルタイムに取得した顔画像データを渡すことで結果を画面に表示する顔認証システムを実装した．実装画面を図 1 に示す．

4　実験結果

4.1　素顔で学習した CNN に対する追跡停止の評価

学習させた CNN に対し異なる条件下で評価した精度を表 2 に示す．素顔を学習させた場合，素顔の評価については 60.6%，マスクなどの外乱は 20%〜40% の再現率となった．また，被験者 5 名の 100 フレームの画像に対し，横軸を素顔，縦軸をマスクにした際に被験者毎の出力値の平均をプロットした結果を図 1 に示す．被験者 C を除いて，負の相関関係がみられた．

Evaluation and Development of face recognition system using CNN with Non-tracking Feature.

†Kazushi Waki　‡Takafumi Mori　†Hiroaki Kikuchi

†School of Interdisciplinary Mathematical Sciences, Meiji University

‡Graduate School of Advanced Mathematical Sciences, Meiji University

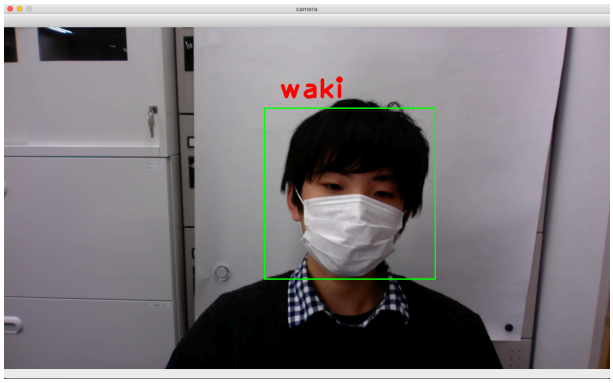

図 1　顔認証システム

表 2　学習させた CNN に対し異なる条件下で評価した場合の再現率

train \ test	素顔	帽子	マスク	サングラス	マスク＋サングラス	平均
素顔	60.6	21.4	28.8	37.0	21.2	33.8
帽子	51.4	74.8	20.0	48.8	20.0	43.0
マスク	20.2	20.0	99.4	20.0	30.0	37.9
サングラス	53.0	20.6	50.4	94.4	53.4	54.8
マスク＋サングラス	25.2	20.0	84.2	22.0	78.6	46.0
平均	42.1	31.8	56.6	44.4	40.6	43.1

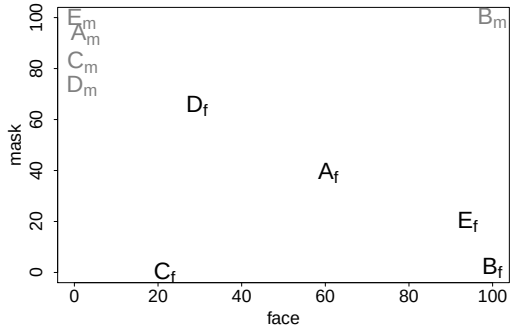

図 2　認証精度の評価
(f-素顔,m-マスク)

A さんの再現率は

$$R_A = \frac{\text{A と正しく判定した数}}{\text{本物の A さんの画像数}}$$

と定義する．平均再現率は全ユーザについての再現率の平均とする．

4.2　外乱を加えた画像で学習した時の評価

表 2 より，それぞれの外乱について以下に述べる，

- 帽子を学習... 素顔とマスク＋サングラスに関して 50% 前後の再現率となった．
- マスクを学習... 同じマスクでの評価は 99.4% となったが，一方でそれ以外については 30% 以下の再現率となった．また，図 1 より，被験者 B を除いて，素顔の出力値がほぼ 0% に対し，マスクはほぼ正しく出力されている．
- サングラスを学習... 同じサングラスでの評価は 94.4% となった．さらに，素顔，マスク，サングラスについても 50% 以上の再現率となった．
- マスク＋サングラスを学習... 同じマスク＋サングラスでの評価は 78.6% となり，それ以上にマスクが 84.2% の再現率となった．一方で，サングラスは 22.0% の再現率であった．

4.3　考察

表 2 より，学習と評価が同じ組み合わせの中で，素顔を学習させた場合の精度が最も低く，外乱を与えた場合が高くなった原因として，素顔は撮影したときの表情などによる変動が大きくなったためと考えられる．それに対し外乱を与えた場合は顔の一部が隠れているため，表情などの変化に頑強であり，露見している部分に対して各自の細かな特徴を学習したのではないかと考えられる．一方で，帽子であれば被り方，マスクやサングラスであれば付け方といったように外乱によって被験者を識別しているとも考えられる．

顔がほぼ隠れているマスク＋サングラスについて，マスクに関しても 84.2% の再現率が得られたことから，マスクを学習した場合と同じ特徴を学習していると考えられる．同じマスク＋サングラスにおいても 78.6% の再現率を得ていることから，マスクが識別する上で重要な特徴を生み出していると考えられる．

表 2 より，サングラスで学習した場合に 5 種類の評価の平均が 54.8% と最も高い数値を得た理由として，まばたきや眼球の動きなど顔の表情の中で最も変化が激しい部位である目をサングラスで隠すことによって，被験者毎の本質的な特徴を学習したため，帽子を除く外乱に対して 50% 程度の再現率が得られたと考えられる．

5　おわりに

マスクやサングラスなどにより素顔を学習させた場合に，顔認証による追跡を 73 %防止できた．しかし，外乱を持った画像を学習データとして利用されると，本人と特定される割合が高くなることが分かった．特に，マスク＋サングラスの学習データに対して，マスクのみの状態で評価した場合も素顔と比べて精度が 20% 高くなることを明らかにした．また，帽子を評価した場合の平均が 31.8% となっていることから，本実験では追跡停止に最も有効な外乱は帽子であることを示した．

参考文献

[1] Karen Simonyan and AndrewZissserman(2014):Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR,2014

[2] 齋藤康毅, “ゼロから作る Deep Learning python で学ぶディープラーニングの理論と実装”, OREILLY 2016.