3M2-1

畳み込みニューラルネットワークを用いた 画質変化や部分的な隠れに頑健な顔認証システム

Face Recognition System
Based on Convolutional Neural Network Robust to Occlusion and Low Quality Images

川合 健斗 *1　山下 宙元 *2　松尾 豊 *3
Kent Kawai　Michiharu Yamashita　Yutaka Matsuo

*1オーマ株式会社
Ohma Inc.

*2東京工業大学大学院総合理工学研究科知能システム科学専攻
Department of Computational Intelligence and Systems Science,
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

*3東京大学工学系研究科技術経営戦略学専攻
Department of Technology Management for Innovation, The University of Tokyo

Face Recognition System has wide range applications such as crime prevention and retail store analysis. In these days, with deep convolutional neural networks, we can obtain higher face-identification accuracy than human. Many face recognition models are tested with LFW or CASIA benchmark dataset, but in these dataset, there're insufficient asian faces, low-quality images, and images with occlusion, which is, important to apply Face Recognition Systems to practical use cases in asian countries. Then, we built a huge dataset which contain many asian faces. Also, our proposed model achieved 99.3% face-identification accuracy with this dataset. Furthermore, we showed this model is robust to occlusion and low quality images.

1. はじめに

近年, 画像認識技術の大幅な精度向上により, 機械が高精度で人間を認識・識別する事が可能になり, ビジネスへの適用や犯罪防止のための導入が注目されてきている. 特に,"顔"は人間を認識・識別する上で重要な情報の一つとされ, その認識・識別を行う顔認証システムは, 出入国管理や店舗での万引き防止などに適用され始めている.

ビジネスや犯罪防止に適用する上で, 実用的な精度であるかどうかの判断基準として, 人間の識別精度を超えているかどうかが 1 つの判断基準として挙げられるが, 例えば, 畳込みニューラルネットワークを用いた手法である [Schroff 15] では, 顔認証のベンチマークデータセットである LFW データセット *1 における人間の識別精度が 97.5%であるのに対し, 識別精度 99.6%を記録した. また, 他の畳込みニューラルネットワークを用いたモデルとして代表的なものとして,[Masi 16] や [Sun 14] があるが, これらも人間の識別精度を超えている.

多くの顔認証モデルは,LFW や CASIA*2 のデータセットを用いてベンチマークが実施されているが, 日本国内でのビジネスや犯罪防止においてこのようなモデルを活用するケースを想定したとき, 以下のような問題があると考えられる.

- ベンチマークデータセットには, 白人顔が多く含まれており, アジア人の割合が少ないため, アジア人に対する識別精度がうまく計測できていない. 例えば,LFW に含まれる白人の割合は,85%程度であるのに対し, アジア人の割合は,7%程度である.
- ベンチマークデータセットには, 低画質の画像はほとんど含まれないが, 実際に活用される場面では, 低画質画像と高画質画像の両方が扱える必要がある. 例えば, カメラから遠くにいる人間（低画質）を, 近くにいる人間（高画質）と同様に認識・識別する必要がある.
- ベンチマークデータセットには, 顔のほとんどが見えている画像が多く含まれるが, 実際に活用される場面では, マスク・サングラスを着用していたり, 物や人によって顔の一部が隠れていることがある.

図 1: オリジナル画像

図 2: 画質を劣化させた画像

図 3: 顔の一部分を隠した画像

そこで, これらの問題に対応するため, 本研究では, 以下の実験と貢献を行った

- アジア人を多く含むデータセットを独自に構築し, 学習データ, テストデータとして用いた. また, このデータセッ

連絡先: 川合健斗, オーマ株式会社, 東京都文京区根津 1-16-10 アーバン内田ビル 4F, kentkawai@ohma-inc.com

*1 Labeled Faces in the Wild http://vis-www.cs.umass.edu/lfw/

*2 CASIA WebFace Database http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html

トを使って, 顔認証モデルを構築し, 識別精度 99.3%を記録した.

- テストデータに含まれる全ての顔画像の画質を 1/3 程度に劣化させ, テストを行った. このデータに対して頑健なモデルを構築し, 識別精度 98.5%を記録した.
- テストデータに含まれる全ての顔画像の一部 (〜50%) をランダムに隠し, テストを行った. このデータに対して頑健なモデルを構築し, 識別精度 95.3%を記録した.

2. 関連研究

顔認証システムは,Detection,Verification の 2 ステージに分けることができる.Detection では, 映像もしくは写真内に写っている顔を検出する. 代表的な手法としては, 畳込みニューラルネットワークを用いた [Liu 15] や [Ren 15] などがある.Verification では, 検出された顔写真が他の顔写真と同一人物を指すものであるか, 別人を指すものであるかを判別する.Verification において代表的な手法としては, 畳込みニューラルネットワークを用いた [Schroff 15] がある. この手法では 22 層の畳込みニューラルネットワークが用いられており, 学習データとして, 同一人物であるか別人であるかのラベルが付与された 2 枚の顔写真のペアを用いる. このモデルでは, 顔写真を入力すると, アウトプットとして 128 次元のベクトルが得られるが, この 128 次元ベクトル空間では, 同一人物であるベクトル同士の距離が近くなり, 別人であるベクトル同士の距離が遠くなる. このベクトル間の距離に閾値を設定することにより, 同一人物であるかどうかの判定を行う.

また, 学習データに平行移動や拡大・縮小, 画質の変化などの小さな変化を加えて, データ量を増やす手法は Data Augmentation と呼ばれている.Data Augmentation は, 畳込みニューラルネットワークの学習において活用されることが多く, 小さな変化に対して頑健なモデルを構築する上で重要な手法の一つである.Data Augmentation の活用例として,[Gan 15] や [Wu 15] がある.

3. 提案手法

本研究では,Detection には Faster R-CNN をベースにしたモデル用いる. また,Verification には畳み込みニューラルネットワークをベースにし, 正面顔以外も識別可能なものに改良したモデルを用いる (提案モデル A). また, 図 2 のような画質の悪い画像に対応するため, 学習データの画質をランダムに劣化させて学習するモデル (提案モデル B) を構築し, 図 3 のような顔の一部分が隠れた画像に対応するため, 学習データの一部をランダムに隠して学習するモデル (提案モデル C) を構築する. さらに, これらのモデルを組み合わせたモデル (提案モデル A+B+C) も構築する.

学習には, 独自に構築した, 日本人を多く含むデータセットを用いる. このデータセットには,257,814 人の人物についての合計約 1,000 万枚の画像が含まれ, 約 80%の人物が日本人である.

4. 実験結果

モデルの評価を行う実験においては, データセットの一部をテストデータとして使い, 残りを学習データとする. ただし, テストデータに含まれる顔画像のペアは, 同一人物のペアの数:別人のペアの数=1:1 となるようにする. また, 比較対象のモデルとしては, 畳込みニューラルネットワークを用いた顔認証システムとして代表的なものである,Microsoft Azure の Face API を用いる.FaceNet などの既存モデルは, 大量の学習データが使われているが, その学習データが入手できず, 手法の再現ができないため, 比較対象外とした.

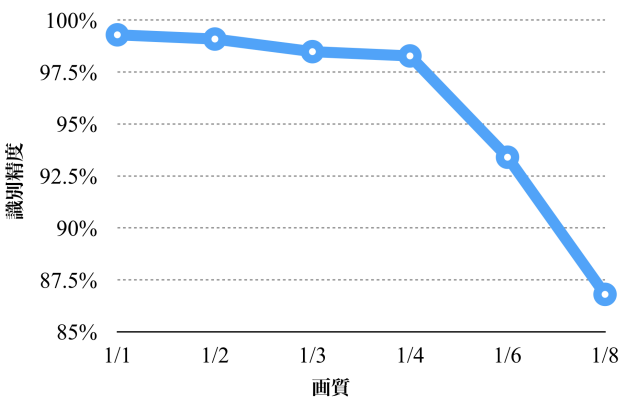

図 4: 画質を劣化させたテストデータでの提案モデル A+B+C の識別精度

モデル/データセット	Dataset1	Dataset2	Dataset3
提案モデル A	98.9%	98.1%	87.1%
提案モデル B	98.4%	98.3%	87.5%
提案モデル C	98.7%	98.2%	95.1%
提案モデル A+B+C	**99.3**%	**98.5**%	**95.3**%
Face API	95.5%	93.2%	93.2%

表 1: 各モデル, 各データセットにおける識別精度

モデルの評価は, 上記で構築したテストデータ (Dataset1), テストデータに含まれる全ての画像の画質を 1/3 に落としたもの (Dataset2), テストデータに含まれる全ての画像の一部 (〜50%) をランダムに隠したもの (Dataset3) の 3 つを用いて行った. 結果は表 1 の通りである. いずれのデータセットにおいても組み合わせモデルである提案モデル A+B+C が最高精度のモデルであり, 低画質の画像を学習データに多く含むモデルである提案モデル B は, 低画質の画像を多く含むデータセットに対して良い性能を示し, 画像の一部をランダムに隠したものを学習データとしたモデルである提案モデル C は, 画像の一部をランダムに隠したものを多く含むデータセットに対して良い性能を示していることがわかる.

また,Dataset2 において, 画質の変化を 1/3 に固定せず, 段階分けしたときの, 提案モデル A+B+C の識別精度についても計測した. その結果が図 4 の通りであり, 各段階での画像の例が図 5 である. 画質が 1/4 程度まで劣化していても, 識別精度に大きな変化がないことがわかる.

さらに,Dataset3 において画像を隠す面積を段階分けしたときの, 提案モデル A+B+C の識別精度についても計測した. ただし, 隠す位置についてはランダムとした. その結果が図 6 の通りであり, 各段階での画像の例が図 7 である. 画像の半分程度が隠れていても,90%以上の識別精度があることがわかる.

図 5: 画質を劣化させた画像の例

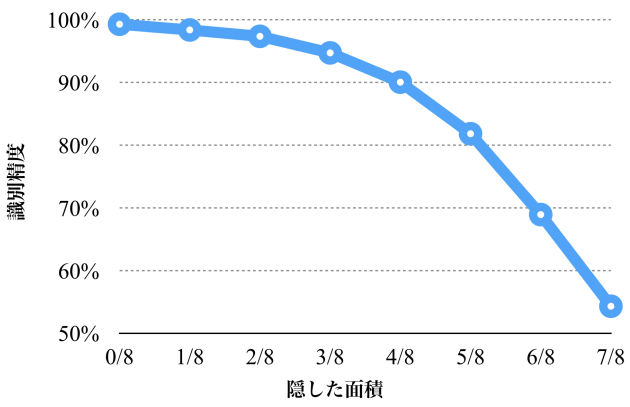

図 6: 画像の一部を隠したテストデータでの提案モデル A+B+C の識別精度

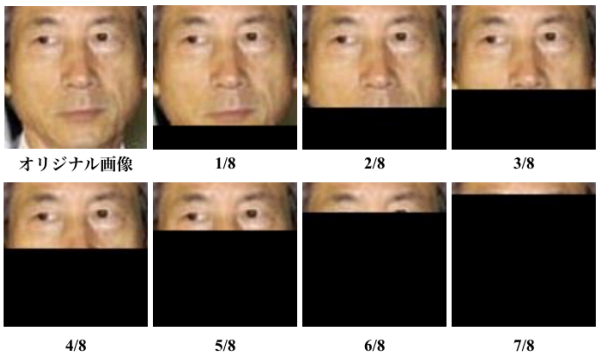

図 7: 一部を隠した画像の例

5. まとめ

本研究では, アジア人を多く含む顔画像データセットを構築し, そのデータセットを用いて顔認証モデルを構築した. また, このモデルがアジア人を多く含むテストデータにおいて高精度であることを示し, 低画質な画像や一部が隠された画像に対しても高精度であることを示した. 今後は, 実際のカメラ映像などを用いることにより, より実用的なモデルの構築も可能になると考えられる.

参考文献

[Schroff 15] Schroff, F., Kalenichenko, D., and Philbin, J.: FaceNet: A Unified Embedding for Face Recognition and Clustering., in CVPR ,2015.

[Taigman 14] Taigman, Y., Yang, M., Ranzato, M. A., and Wolf, L.:DeepFace: Closing the Gap to Human-Level Performance in Face Verification., in CVPR ,2014.

[Sun 14] Sun ,Y., Liang, D., Wang, X., and Tang, X.: DeepID3: Face Recognition with Very Deep Neural Networks., arXiv:1502.00873 ,2014.

[Masi 16] Masi, I., Tran, A. T., Leksut, J. T., Hassner, T., and Medioni, G.: Do We Really Need to Collect Millions of Faces for Effective Face Recognition?, in ECCV, 2016.

[Liu 15] Liu, W., Anguelov D., Erhan, D., Szegedy, C., and Reed, S.: SSD: Single shot multibox detector. arXiv:1512.02325, 2015.

[Ren 15] Ren, S., He, K., Girshick, R., and Sun, J.: Faster r-cnn: Towards realtime object detection with region proposal networks, in NIPS, 2015.

[Gan 15] Gan, Z., Henao, R., Carlson, D., and Carin, L.: Learning deep sigmoid belief networks with data augmentation., in AISTATS, 2015.

[Wu 15] Wu, R., Yan, S., Shan, Y., Dang, Q, and Sun., G.: Deep image: Scaling up image recognition., abs/1501.02876, 2015.