プライバシーに配慮したDeep Convolutional Neural Networkによるクラウド型顔照合システム

福井宏 †　加藤優 †　神谷龍司 †　山下隆義 †　山内悠嗣 †　藤吉弘亘 †
† 中部大学
E-mail: {fhiro@vision.cs, yuu@isc, yamashita@cs, hf@cs}.chubu.ac.jp

Abstract

障がい者や高齢者を支援するサービスを実現するために，クラウド上の端末で情報処理を担うクラウドロボティクスの考えが普及している．クラウドロボティクスにおいて顔画像を扱う場合は，プライバシーに配慮する必要がある．本稿では，クラウドロボティクスのフレームワークにおいて，プライバシーに配慮したDeep Convolutional Neural Network(DCNN) に基づく顔照合システムを提案する．DCNN は，高性能な顔照合が可能であるが，計算コストが高いため，ロボットの負担が大きくなる．提案するクラウド型顔照合システムは，DCNN の下位層をロボットで処理し，DCNN の上位層をクラウドサーバで処理する．これにより，ロボット側の計算コストを大幅に削減できる．また，クラウドサーバに顔画像ではなく，DCNN の特徴マップを送信することで，プライバシーに配慮する．本システムは，高精度かつ効率的な顔照合を行い，プライバシーに配慮した顔照合システムを実現する．

1 はじめに

少子高齢化に伴う労働力人口の減少を補うために，ロボットの活躍が期待されている．また，高齢者の増加に伴う介護士不足は大きな社会問題となっており，高齢者の生活支援や人とのコミュニケーションができるロボットの開発が必要とされている．ロボットによる生活支援や人とのコミュニケーションを行うには，カメラなどのデバイスから取得した顔画像を用いて顔照合を行い，個人に合った生活支援やコミュニケーション方法が必要である．

顔画像を用いた顔照合は，顔の局所領域や動きを特徴量として用いて照合を行う [2]．これらの特徴量は，個人で異なる顔の局所的な特徴を捉えることができる．また，顔画像から複数の属性を認識し，顔照合するアプローチも提案されている [1]．Taigman らが，Deep Convolutional Neural Network(DCNN)[3] を顔照合に適用した DeepFace を提案し，高い性能を実現したことから，Deep Learning ベースの手法が注目されている [4]．DeepFace は，2 次元と 3 次元の顔情報を利用してアライメントを行い，顔の向きに頑健な照合が可能となり，人間と同等の認識性能を実現している．また，DCNN から獲得できる特徴ベクトル同士の距離計算で顔照合をすることで，照合対象が増えた場合でも DCNN を再学習する必要がない．しかし，DCNN は高精度な顔照合が可能であるが，計算コストが非常に高い．そのため，ロボットの計算リソースを顔照合で占有する問題が発生する．

この問題を解決する方法として，2013 年に Kuffner らが提案したクラウドロボティクスがある [5]．クラウドロボティクスは，ロボットとクラウド上の認識エンジンや情報処理技術を連携させ，大規模な計算やデータベースを要求するような情報処理クラウドで担うという考え方である [5], [6], [7], [8], [9].

本研究では，DCNN による顔照合とクラウドロボティクスを組み合わせたクラウド型顔照合システムを提案する．クラウド型顔照合システムは，顔画像を用いて学習した DCNN をベースとした構成であり，DCNN の下位層をロボットの端末で処理し，DCNN の上位層の処理と顔照合をクラウドサーバで行う．このとき，クラウドサーバに顔画像は送信せず，DCNN から得られる特徴マップを送信する．DCNN の特徴マップは，DCNN の構造や重みフィルタなどのパラメータの係数がでな既知でない場合に復元が不可能である．そのため，暗号化されたデータを送受信していることと等価なため，プライバシーに配慮したシステムとなる．そして，顔照合の性能を向上させるために，DCNN の学習時に用いるデータセットの規模を検討する．顔照合処理では，k-Nearest Neighbor(NN) 法を導入することで，誤識別を抑制する．これらの処理により，本システムの顔照合性能を向上させる．本システムでは，これらの技術を導入することで，高精度かつ効率的でプライバシーに配慮した顔照合システムを実現する．

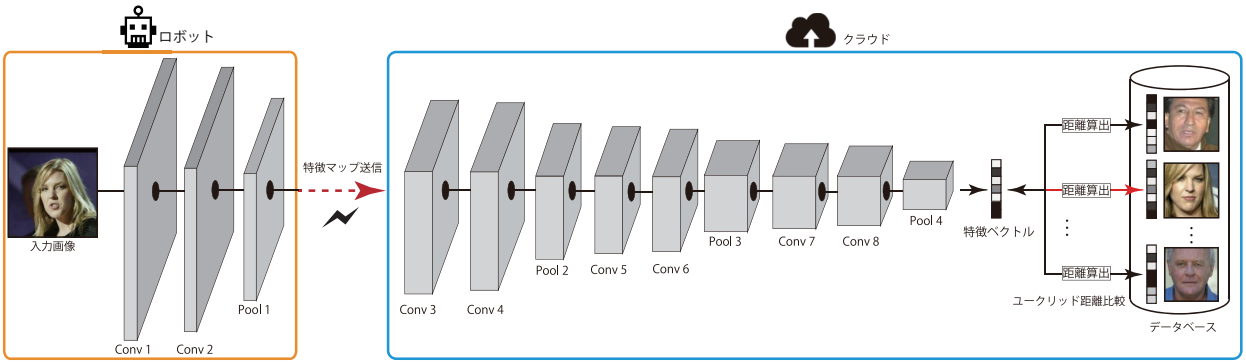


図 1　提案するクラウド型顔照合システム

2 関連研究

顔照合では，顔の局所領域を特徴量とし，個人で異なる顔の局所的な特徴量を捉えて，照合する方法が一般的に用いられる．Kumar らは，65 種類の顔属性と，笑顔認識の識別結果から顔照合している [1]．認識する属性は，人種や性別，世代，髪型，化粧の濃さなどであり，これらの属性は Support Vector Machine (SVM) を用いて認識する．笑顔認識は，目や口などの顔の局所領域を SVM へ入力して学習，認識する．また，眼球の虹彩から照合を行う方法も提案されており，Locally Adaptive Regression Kernel (LARK) 特徴量 [10] を入力として Radial Basis Function Neural Network (RBFNN) により学習，照合している [2]．

一方，DCNN が従来の一般物体認識手法の性能を大きく上回ったことで注目されている [3]．DCNN は，学習過程において識別に適した特徴量を自動的に獲得でき，家の番号認識 [11]，シーン認識 [12]，物体検出 [13] など様々なベンチマークで高い性能を達成している．Taigman らは，DCNN を顔照合に適応した DeepFace を提案し，人間の認識性能と同等の顔照合を実現している [4]．DeepFace は，4,030 人の約 440 万枚の顔画像を学習している．このとき，顔の向きに頑健な顔照合を実現するために，2 次元と 3 次元の顔情報からアライメントを行う．アライメントは，2 次元の場合に顔器官点を基準にスケーリングと平行移動を与え，3 次元の場合は顔器官点と顔の 3 次元モデルをフィッティングすることで，様々な角度の顔画像を正面向きの顔に補正する．顔照合では，顔画像を DCNN に入力した際に獲得できる特徴ベクトルとの距離計算により，照合を行う．

しかし，DeepFace のような高度な画像認識は，一般的に多くの計算量が必要となるため，ロボットの情計算リソースを占有する問題がある．また，画像認識には膨大なデータベースや学習したモデルを保存する必要があるため，ロボット側で全ての情報を保持するのは難しい．これらの問題を解決するために，クラウドロボティクスがある [5]．クラウドロボティクスは，ロボットとクラウド上の認識エンジンや情報処理技術を連携させ，大規模な計算やデータベースを要求するような情報処理を，クラウドで担うという考え方である．しかし，顔照合をする上で，ロボットで撮影した顔画像をクラウドサーバに送信すると，撮影された顔画像から個人を特定されてしまう可能性がある．

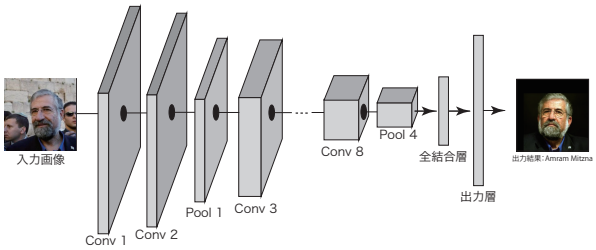


図 2　顔認識で用いる CNN の構造

3 クラウド型顔照合システム

提案するクラウド型顔照合システムは，図 1 のような DCNN をベースとした構成である．はじめに，DCNN を顔画像のデータセットで学習する．学習した DCNN を，図 1 のように DCNN の畳み込み層をロボットとクラウドサーバで処理する 2 つのネットワークに分割する．そして，顔画像を学習した DCNN に入力した際に得られる特徴ベクトルと，クラウドサーバに保存されている特徴ベクトルのユークリッド距離を求め，顔照合を行う．以下に，クラウド型顔照合システムの処理について詳細に述べる．

3.1 DCNN の構築

クラウド型顔照合システムのベースとなる DCNN を構築するために，顔認識データセットの顔画像を用いて DCNN を学習する．顔認識データセットの顔画像で学習することで，顔照合に有効な特徴量の獲得が容易となる．このとき，学習には図 2 のような顔認識で用いる構造の DCNN を用いる．

DCNN を構築する際に，層数や重みフィルタ等のネットワーク構造を決定する必要がある．図 3 に，DCNN の層数を変化させた際の顔認識精度を示す．このとき，

学習には Labeled Face in the Wild (LFW) Dataset[15] を用いる．学習サンプルには 4,031 枚の顔画像を用いており，評価サンプルは 143 枚の顔画像を用いる．実験の結果，畳み込み層を 8 層とした際の精度が最も良いことがわかる．そこで，本システムでは 8 層の畳み込み層から構成される DCNN を用いる．

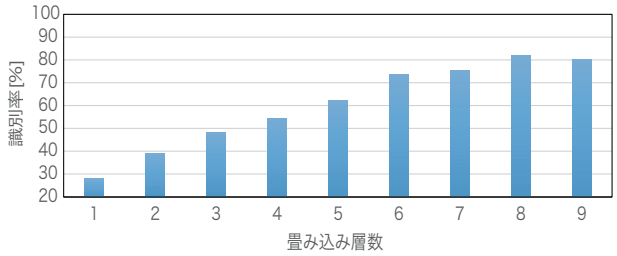


図 3　畳み込み層の層数を変化した際の精度

顔画像を学習した DCNN に対して，畳み込み層をロボットとクラウドサーバで処理する 2 つのネットワークに分割する．ロボット側は，指定した畳み込み層の特徴マップをクラウドサーバに送信する．クラウドサーバでは，受信した特徴マップを再度 DCNN へ入力し，畳み込みとプーリングの続きを行う．DCNN の構造を分割することで，ロボットの計算コストを大幅に削減できる．

3.2　DCNN による顔照合方法

入力される顔画像と登録ユーザの照合は，クラウドサーバで行う．顔照合は，DCNN の出力層で識別せず，全結合層前の特徴マップを特徴ベクトルに変換して照合する．入力された顔画像の特徴ベクトルとクラウドサーバに保存されている登録ユーザの特徴ベクトルのユークリッド距離を算出し，最も距離の短い登録ユーザを照合結果として出力する．特徴ベクトル間のユークリッド距離により照合することで，登録ユーザが増えた場合でも，DCNN を再学習してシステムを再構築する必要がない．

3.3　k-NN による顔照合の高精度化

特徴ベクトル間のユークリッド距離で顔照合する際に，他人を本人，もしくは本人を他人と誤識別する場合がある．データ分布上でユークリッド距離が最も短い登録ユーザを照合結果とするため，顔の向きなど見えの変化が生じた際に他の登録ユーザとのユークリッド距離が近くなり，誤識別する．この問題を解決するために，k-NN 法を導入する．はじめに，入力サンプルと登録サンプルのユークリッド距離を，全登録サンプルに対して求める．そして，上位 k 個の短いユークリッド距離の平均値を，評価時に使用する距離とする．k-NN 法を用いることで，見えが生じた顔画像を他の登録ユーザと誤識別した場合でも，他の登録ユーザとのユークリッド距離も算出して考慮するため，誤識別を抑制できる．

3.4　プライバシーへの配慮

クラウドロボティクスで顔画像を取り扱う際に最も問題となるのが，プライバシーへの配慮である．顔画像は個人を特定できる情報であり，顔照合を行う際に顔画像の漏洩等には細心の注意を払う必要がある．本システムでは，ロボット側の DCNN で算出した特徴マップのみを送信し，クラウド側に顔画像を保存しないため，プライバシを配慮したシステムとなる．DCNN から得られる特徴マップは，重みフィルタ等のパラメータの係数が既知でない場合，元画像への復元が不可能なため，安全性の高いシステムとなる．

4　評価実験

クラウド型顔照合システムの有効性を調査するために，評価実験を行う．本実験では，クラウド型顔照合システムの顔照合性能を確認し，DCNN の構造を分割した際のデータ通信量及び計算量を調査する．そして，本システムがどの程度プライバシーに配慮しているのかを，17 人の被験者を用いて定量的に評価する．実験で使用する DCNN の構造を，表 1 に示す．実験では，畳み込み層を 1 層から 9 層まで増やし，各構成における照合精度を比較する．各畳み込み層の活性化関数には，Maxout を使用し，プーリング層は畳み込み層の 2 層目，4 層目，6 層目及び 8 層目に配置する．DCNN の学習及び，顔認識の評価には，CASIA-WebFace Dataset [14] を用いる．CASIA-WebFace Dataset は，10,575 人の 494,414 枚の顔画像から構成されている大規模な顔画像データセットである．顔照合の評価に使用するサンプルは，LFW Dataset から 143 枚の顔画像を使用する．

はじめに，学習時に使用する顔画像の人数と Data Augmentation の有無の調査を行い，k-NN 法を導入した際の顔照合精度を確認する．ここで，Data Augmentation で与える幾何変化は，平行移動を ±5pixel，回転を ±10°，スケーリングを 0.8 から 1.2 倍とする．本実験で用いる顔画像の人数は，105 人の場合と 10,552 人であり，それぞれ 41,831 枚と 439,224 枚の顔画像を有している．評価方法は，Probability Distribution Function(PDF) を算出し，算出した PDF から Receiver Operating Characteristics(ROC) カーブを作成して False Acceptance Rate(FAR) と False Rejection Rate(FRR) により評価する．

DCNN の顔照合精度を確認した後に，DCNN の第何層で分割するとロボット及びクラウドサーバの負荷が少ないかを検証する．使用するロボットとクラウドサーバのマシンスペックを，表 2 に示す．そして，顔照合精度とロボット及びクラウドサーバの負荷を定量的に評価した後に，本システムのプライバシーへの配慮に

表 1　実験で使用する DCNN の構造

入力層	画像サイズ	100 × 100
1 層目	重みフィルタ	3 × 3 × 16
2 層目	重みフィルタ	3 × 3 × 16
	Maxout	2
	Max pooling	2×2
3 層目	重みフィルタ	3 × 3 × 32
4 層目	重みフィルタ	3 × 3 × 32
	Maxout	2
	Max pooling	2×2
5 層目	重みフィルタ	3 × 3 × 64
6 層目	重みフィルタ	3 × 3 × 64
	Maxout	2
	Max pooling	2×2
7 層目	重みフィルタ	3 × 3 × 128
8 層目	重みフィルタ	3 × 3 × 128
	Maxout	2
	Max pooling	2×2
9 層目	重みフィルタ	3 × 3 × 256

表 2　使用する各端末のスペック

端末	CPU	メモリ	GPU
ロボット	Intel(R) Corei7-3520	8GB	–
クラウドサーバ	Intel(R) Corei7-4790	8GB	GeForce GTX980

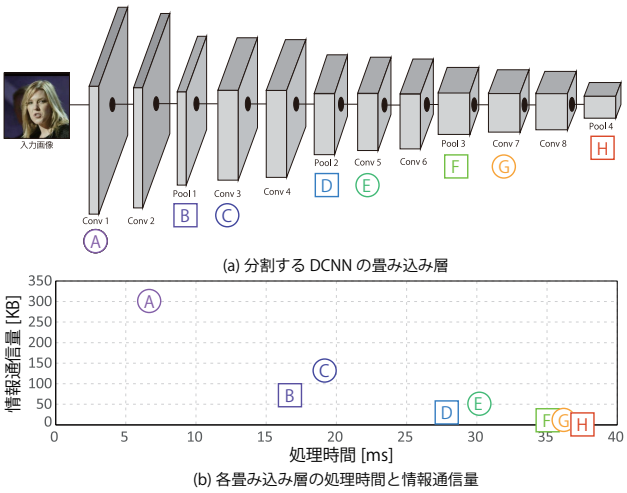


図 5　ロボットから送信するデータ送信量と処理時間

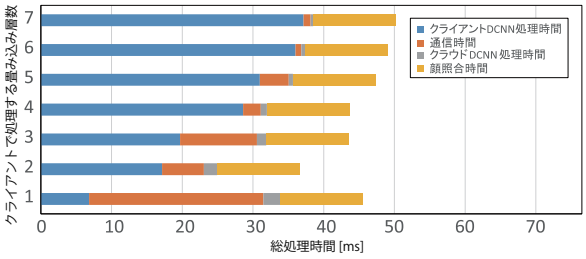


図 6　分割する畳み込み層の変化による総処理時間の変化

ついて検証する．

4.1　顔照合に使用する CNN の性能評価

図 4(a)，(b) に，ROC カーブによる顔照合の性能を比較した結果を示す．図 4(a) では，DCNN を学習する際に使用する顔画像の人数を変更した際の性能を比較している．図 4(a) より，Data Augmentation を導入せず，10,522 人の顔画像を使用したほうが高精度に顔照合が可能であることがわかる．

図 4(b) では，k-NN 法を導入した際の性能を比較している．ここで，$k = 1$ の性能が，従来の顔照合方法である．図 4(b) より，k を増加することで顔照合の性能が向上していることがわかる．しかし，k が 6 より大きくなった場合に，顔照合の精度が低下している．FAR=0.05 を基準としたとき，$k = 6$ のときに約 96%の精度で顔照合が可能である．

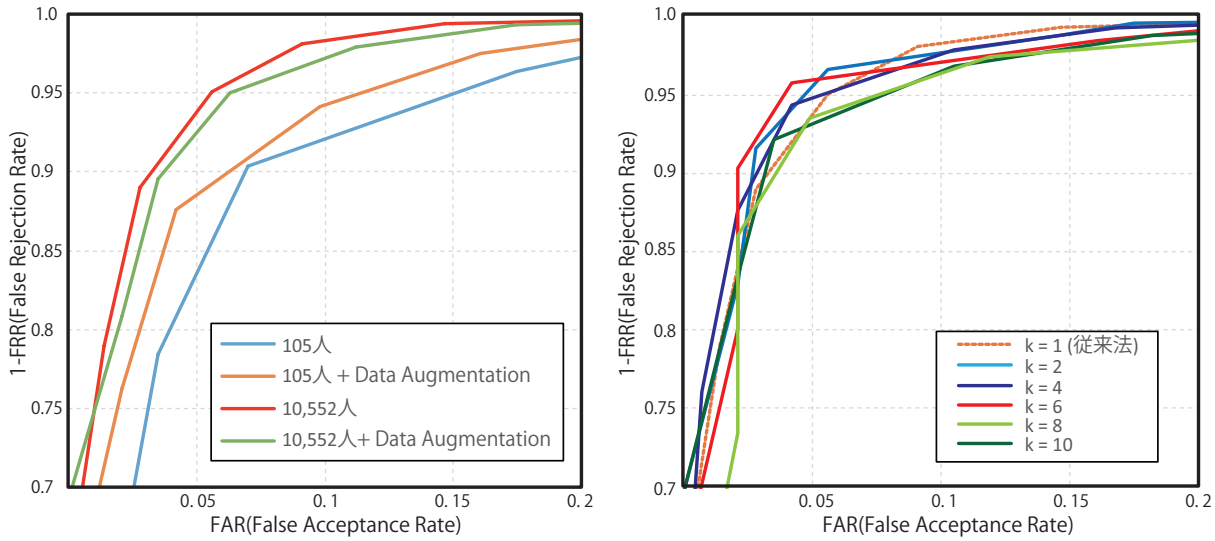
4.2　計算コストの比較

図 5 に，ロボットで処理する畳み込み数を変化させた際のロボット側の処理時間とデータ通信量を示す．ロボットはクラウドサーバと比べて計算能力が低いため，ロボットで処理する層が多い場合に総処理時間が長くなる．しかし，ロボットで処理する層数を浅くした場合，データ通信量が膨大になるため，通信時間が長くなる．図 5 より，2 層目の畳み込み層 (Conv2+Pool1) で分割した際に，ロボットの処理時間が短く，かつデータ送信量が少ないことがわかる．

図 6(b) に，分割する畳み込み層を変えた時の総処理時間を示す．ここで，DCNN の畳み込み層は図 6(a) のように分割する．総処理時間は，ロボット及びクラウドサーバにおける DCNN の処理時間，クラウドサーバにおける顔照合時間，通信時間を総和したものである．顔照合時間は，クラウドサーバに保存されている 10 人分のデータと照合するための時間である．通信時間は，通信速度を 100Mbps(12.5KB/ms) と仮定した際の値である．図 6(b) より，2 層目の畳み込み層で分割した場合に，総処理時間が最も短い．クラウドサーバで行う顔照合時間は，どの層で分割しても変化が少ないため，ロボットの処理時間と通信時間が大きく影響していることが確認できる．

4.3　プライバシーの評価

顔画像は個人を特定できる情報であるため，インターネットを介した通信には注意が必要である．図 7 に，各畳み込み層の特徴マップを可視化した結果を示す．図 7 より，DCNN の畳み込み層が上位層になるにつれて，



(a) 学習で使用する人数を変化させたときの性能　(b) k を変化させたときの性能

図 4　ROC カーブによる顔照合の評価

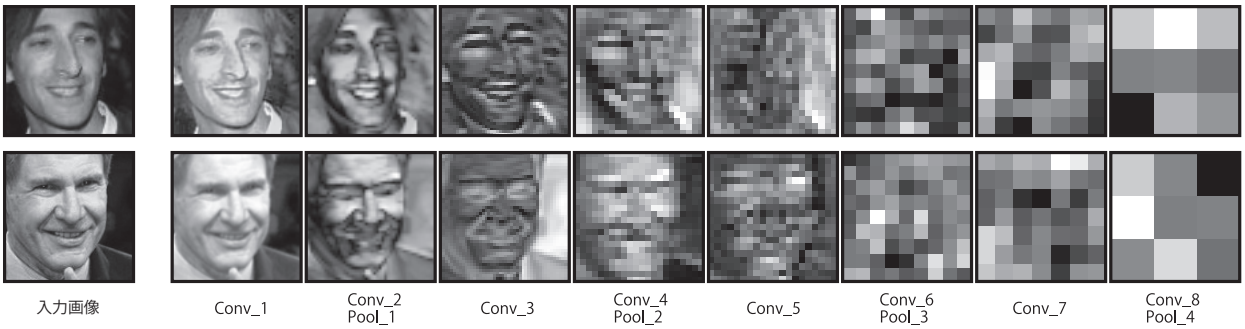


図 7　各畳み込み層における特徴マップの可視化

目視による個人の特定が困難になることが確認できる．

DCNN の分割する畳み込み層とプライバシーへの配慮の関係を定量的に評価する．そのために，17 人の被験者に各畳み込み層の特徴マップを目視したとき，顔照合が可能であるかのアンケートを採る．顔照合には，顔照合の評価で使用している LFW Dataset の 143 枚の顔画像と各層の特徴マップを目視で比較する．このとき，LFW Dataset からランダムに選択した顔画像を，DCNN に入力した際に得られる特徴マップを使用する．被験者は、同一人物，他人，分からないの 3 つから回答する．図 8 に，被験者の顔照合精度とプライバシー率を示す．図 8 の棒グラフが被験者による顔照合率を示しており，折れ線グラフがプライバシー率を示している．ここで，プライバシー率は，被験者が特徴マップを目視した際に，顔照合が不可能と判断した割合を示している．図 8 より，分割する畳み込み層が上位層になるにつれてプライバシー率が向上し，被験者の顔照合精度が低下していることが確認できる．特に，切り離す畳み込み層が 4 層目のときに，プライバシー率が大幅に向上し，顔照合精度が大幅に低下していることがわかる．

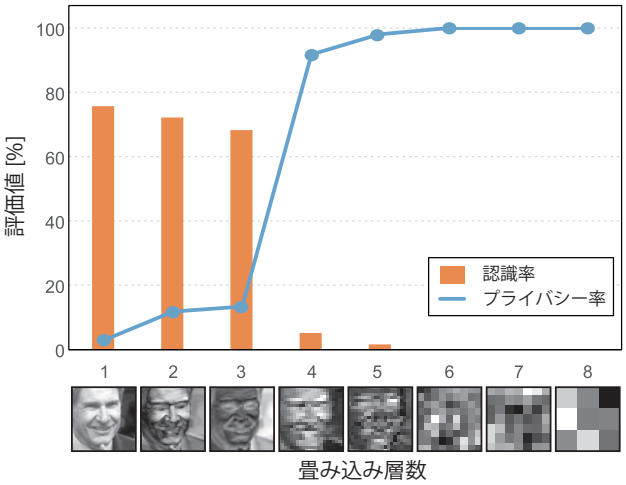


図 8　各畳み込み層の正解率とプライバシー率

5　おわりに

本稿では，プライバシーに配慮した DCNN によるクラウド型顔照合システムを提案した．本システムでは，学習サンプルに大量の人物の顔画像を DCNN の学習に用いることで高精度な顔照合を実現した．さらに，特徴ベクトル間のユークリッド距離を用いて照合する際

に，k-NN 法を導入することでさらに高精度な顔照合を実現した．また，DCNN の下位層をロボットで処理し，DCNN の上位層の処理と顔照合をクラウドサーバで行うことで，効率的な顔照合を実現した．そして，クラウドサーバに人物の顔画像を送信するのでなく，DCNN の特徴マップを送信することで，暗号化されたデータを送受信していることと等価になり，プライバシーに配慮したシステムを実現した．

参考文献

[1] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and Simile Classifiers for Face Verification", IEEE International Conference on Computer Vision, 2009.

[2] Y. Wang, T. Tan, and A. K. Jain, "Combining Face and Iris Biometrics for Identity Verification" , Audio-and Video-based Biometric Person Authentication, No.9, pp.805-813, 2003.

[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition" , Proceedings of the IEEE, 1998.

[4] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, " DeepFace: Closing the Gap to Human-Level Performance in Face Verification" , Computer Vision and Pattern Recognition, pp. 1701-1708, 2014.

[5] B. Kehoe, A. Matsukawa, S. Candido, J. Kuffner, and K. Goldberg, "Cloud-Based Robot Grasping with the Google Object Recognition Engine" , IEEE International Conference on Robotics and Automation, 2013.

[6] O. Zweigle, R. van de Molengraft, R. d'Andrea, and K. Haussermann, "Roboearth: connecting robots worldwide", ACM International Conference on Interaction Sciences: Information Technology, Culture and Human, pp. 184-191, 2009.

[7] M. Waibel, M. Beetz, J. Civera, R. Andrea, J. Elfring, D. Galvez-Lopez, K. Haussermann, R. Janssen, J.M.M. Montiel, A. Perzylo, B. Schiessle, M. Tenorth, O. Zweigle and M.J.G. (Rene) Van de Molengraft, "RoboEarth - A World Wide Web for Robots" Robotics and Automation Magazine, vol 18, no 2, pp 69-82, June 2011.

[8] W. J. Beksi, J. Spruth and N. Papanikolopoulos, "CORE: A Cloud-based Object Recognition Engine for Robotics", International Conference on Intelligent Robots and Systems, 2015.

[9] K. Sugiura and K. Zettsu, "Rospeex, A Cloud Robotics Platform for Human-Robot Spoken Dialogues", International Conference on Intelligent Robots and Systems, 2015.

[10] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction", IEEE Transaction Image Process, No. 16, pp. 349-366, 2007.

[11] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnold, and V. Shet, "Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks" ,CoRR, Vol. abs/1312.6082, 2013.

[12] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning Hierarchical Features for Scene Labeling", IEEE Transacations on Pattern Analysis and Machine Intelligence, 2012.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation" , Computer Vision and Pattern Recognition, 2014.

[14] Y. Dong, L. Zhen, L. Shengcai, and L. Z. Stan, "Learning Face Representation from Scratch". arXiv preprint arXiv:1411.7923, 2014.

[15] G. B. Huang, M. Ramesh, T. Berg, and E. L. Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments" , University of Massachusetts, Amherst, 2007.