K-004

ディープラーニングのための訓練データ自動生成
Training data synthesis for deep learning

大西 一徳† 全 へい東‡
Kazunori Ohnishi† Heiton Zen‡

1. はじめに

ディープラーニングを用いて画像認証を行う研究は近年盛んに行われており，ディープラーニングを用いた識別器はいろいろな学会において優秀な成績を収めている．

しかし，ディープラーニングにも解決しなければならない問題が 3 つ存在する．1 つ目に Hard Example（対象の光学的変化，幾何学的変化による分類が難しい画像）が存在する問題がある．例えば，顔認証における Hard Example にはカメラのフラッシュによる光源の著しい変化がある画像や，サングラスやマスクなどでオクルージョンが発生している画像などである．Hard Example に対応する訓練データがなければ識別に失敗しやすい．

2つ目，3つ目に大量の訓練データが必要になる問題，対象物体に適したディープラーニングの構造にしなければ認証率が上昇しづらい問題がある[1][2]．対象物体に適した構造であれば，訓練データの数を抑えることも可能であるが[2]，対象物体に適切なディープラーニングの構造は実験的に決定するものが多く，簡単には求めることができない．そのため，一般的にディープラーニングでは訓練データを大量に必要とする．例えば，先行研究の[1]では一人あたり約 1300 枚，訓練データセット全体では 740 万枚もの顔画像を利用している．

本研究では，1 つ目と 2 つ目の問題に対処する方法として，Hard Example を事前に作成し，訓練データに加えることを提案する．このことにより，訓練データを増やせ，また Hard Example に対して強くなると考えられる，今回は一般的な構造の Convolutional Neural Networks（CNN）[3]を利用するときの顔認証に特化した訓練データセットを作成する．起こりうる Hard Example を自動生成し，それを訓練データセットに追加する．

2. ディープラーニングにおける画像認識

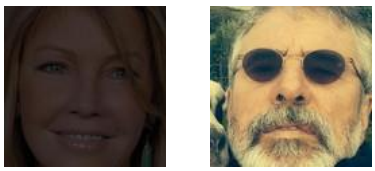
2.1 Hard Example

画像認識では，認識に失敗しやすい画像が存在する．光源が訓練データと著しく異なる場合やオクルージョンが発生した場合はうまく認識できない．このような画像を Hard Example とよぶ．

Hard Example が生成される理由は，光源の変化や物体の分光反射率の変化等の光学的変化と，オクルージョンの発生や物体の形状変化等の幾何学的変化の 2 つが挙げられる．以下の表 1 にその体系を示す．また，図 1 に顔認証における Hard Example の例を示す．同図(a)が光源の変化が大きい例であり，同図(b)がサングラスによりオクルージョンが発生している例である．

表 1 Hard Example の発生原因

光学的変化	幾何学的変化
照明変化	オクルージョン
物体の分光反射率の変化	物体の形状変化
	物体の姿勢変化



(a) (b)

図 1 顔認証における Hard Example の例

2.2 Data Augmentation

ディープラーニングでは非常に多くのデータが必要である．そのため，既存の訓練データを加工して訓練データを増加させる手法が一般的に用いられている．この過程は Data Augmentation と呼ばれ，従来手法では左右反転，変形，ノイズの付加が行われていた[1][2][4]．図 2 にその例を示す．同図(a)が左右反転の例，同図(b)がノイズ付加の例である．

(a) (b)

図 2 Data Augmentation の例

2.3 Bootstrapping Resampling（追加学習）

ディープラーニングに限らず，機械学習では追加学習によって識別率を上昇させることが可能である．特に，識別に失敗した Hard Example を学習データに追加することによって的確に識別率を上昇させることができる．先行研究として，図 3 は，歩行者を検出する CNN で Bootstrapping Resampling（追加学習）を行った例である[注)]．この CNN で Penn-Fudan Database[5]をテストデータとして認識をさせた結果，上下どちらの画像でも訓練データに存在しなかっ

† 千葉大学大学院 工学研究科，Graduate School of Engineering, Chiba University
‡ 千葉大学 統合情報センター，Insitute of Management and Information Technologies

注）この実験結果は，著者らのグループが平成 26 年度千葉大学工学部卒業研究の一部として行ったものである

入力画像　　Resampling 前　　Resampling 後

図 3 Bootstrapping Resampling の例

た後ろ向きの人をうまく認識できなかった．これは，後ろ向きの人が Hard Example となっているといえる（図 3 resampling 前）．その後，後ろ向きの人を訓練データに追加したところ，後ろ向きの人も認識できるようになった（同図 Resampling 後）．このように，Hard Example を学習させる Bootstrapping Resampling（追加学習）に一定の効果があることが確認できる．

3. 提案手法

2.2 と 2.3 で紹介した Data Augmentation と Bootstrapping Resampling の 2 つの考えを拡張として．起こりうる Hard Example を事前に作成し，それを訓練データに加える手法を提案する．今回は一般的な構造の CNN を利用するときの顔認証のための訓練データセットを作成する．これにより，画像認識において Hard Example が存在するという問題と，訓練データを大量に用意しなければならないという問題の両方に対応できると考える．

3.1 Hard Example の生成

Hard Example は，2.1 で紹介したように光学的変化と幾何学的変化によって起こる．顔画像においては，カメラのフラッシュや撮影角度の変化などが考えられる．このような Hard Example を作成したい場合，2 次元画像に対して操作するのでなく，3D モデルに対して操作を行い Hard Example を作るべきである．

3.1.1 顔 3D モデルによる Hard Example の生成

一枚の正面顔画像から顔特徴点を取得する．その特徴点を利用し基本となる一般的な顔から作成されたワイヤーフレームモデルを変形させ，正面顔画像をレンダリングする．作成した顔 3D モデルに対して，光学的変化や幾何学的変化を与え，それを 2 次元画像にして Hard Example を作成する．これは人間の顔であれば，ある程度の大きさや形が決まっているために可能なことである．

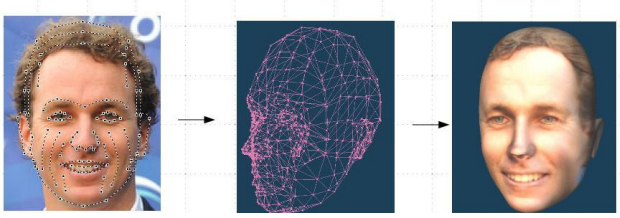


図 4 顔 3D モデルの生成例

4. 実験

本章では，3 章で提案した手法により顔画像において起こりうる Hard Example を生成し，それを訓練データセットに追加させた時の識別率を評価する．今回は生成した Hard Example を含むものと含まないものの 2 種類の訓練データセットを用意し，それぞれの訓練データセットで CNN を学習させ，その時の識別率で本手法の有効性を評価する．

4.1 訓練データ

今回の実験対象とする人物は LFW-Dataset[6]からランダムに選んだ 100 人とする．LFW-Dataset は各国の著名人の顔画像により作成されたデータセットであるが，一人あたりの顔画像の枚数は少ないため，インターネットから一人あたり 100 枚の顔画像を適当に取得する． 2 種類の訓練データセットは以下のようにする．

- A: 加工を行わない生のデータ 100 枚と左右反転を行った 100 枚の一人あたり計 200 枚の訓練データセット
- B: 1 枚の正面顔画像から作成した顔 3D モデルから 250 枚の Hard Example を生成し，A にこれらを追加した一人あたり計 450 枚の訓練データセット

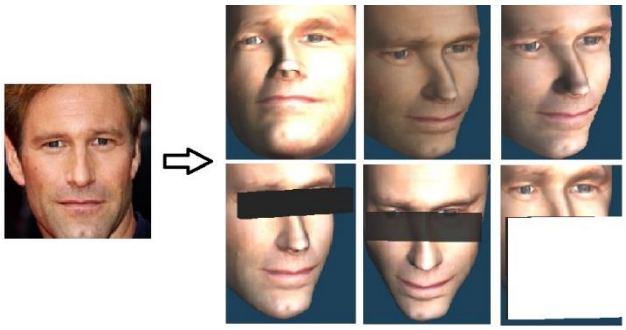
評価セットとして LFW-Dataset の画像とインターネットから取得した訓練データに使われていない顔画像とを合わせ，一人あたり 20 枚を用意した．

4.1.1 作成した Hard Example

今回作成する Hard Example は以下の 3 種類の変化の組み合わせから得られる画像とした．1 つ目は光学的変化として光源の変化，2 つ目は幾何学的変化として姿勢変化（ここでは顔の向きの変化）とし，3 つ目も幾何学的変化としてオクルージョンを選択した．顔 3D モデルは 1 枚の正面顔画像から作成しているため，顔の真横や，顎の下の方の情報は得ることができない．そのため，姿勢変化は完全にランダムではなく制限を設けた．左右及び下方向には 25°，上方向には 10°とした．上方向だけ制限が厳しいのは，正面顔画像からは鼻の穴の情報が得られないため，鼻の穴をあまり写さないようにするためである．また，オクルージョンは顔 3D モデルの前に直方体を置くことで作成した．その直方体の色や大きさだけでなく透過率も変化させた．表 2 に今回の Hard Example を生成するときに変化させたパラメータを示す．また，図 5 に例を示す．図 5(a)が顔 3D モデルの元になった正面顔画像であり，同図(b)が顔 3D モデルから作成した Hard Example の一例である．

表 2 Hard Example を生成するときに変化させたパラメータ

照明変化(色彩・強度)	(H_l, S_l, V_l)
照明変化（位置）	(x_l, y_l, z_l)
向き（ピッチ）	θ
向き（ヘッド）	ϕ
オクルージョン（大きさ）	(h, w, d)
オクルージョン（位置）	(x_o, y_o, z_o)
オクルージョン（色彩）	(H_o, S_o, V_o)
オクルージョン（透過率）	T



(a) (b)

図 5 作成した Hard Example の例

4.1.2 Convolutional Neural Networks（CNN）

今回利用した CNN は，一般的な CNN の構造とする．入力画像の大きさを 120×120，畳み込みフィルタの大きさは 7×7，プーリングフィルタは 3×3 の Maxpooling とした．また，層の数は畳み込み層とプーリング層ともに 3 層とした．次ページ図 6 にこれを示す．各々のパラメータの初期値は，分散 0.01，平均 0 のガウス分布に従う乱数とし，パラメータの更新は学習率 0.000001 で確率的勾配法により行った．また，活性化関数は Rectified Linear Unit (ReLU)を利用した．ReLU は CNN で一般的に用いられている活性化関数であり，式は(1)となる．

$$f(x) = \max(x, 0) \tag{1}$$

また，識別部は全結合とし，識別する際の確率の算出には Softmax 関数を利用した．Softmax 関数は出力層の値をそれぞれ確率として算出する関数であり，こちらも CNN では一般的に用いられている．式は(2)となる．なおhが出力層の値，yが実際に識別に用いる確率の値である．

$$P(y^i) = \frac{\exp(h_i)}{\sum_{j=1}^{M} \exp(h_j)} \tag{2}$$

4.2 結果

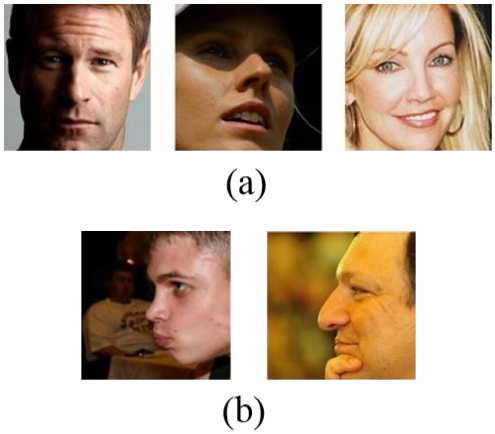
2 種類の訓練データセットでそれぞれ CNN を学習させた際の識別率を以下の表 3 に示す．

表 3　訓練データセットによる識別率の違い

訓練データセット	A	B
識別率[%]	76.2	79.9

従来手法である左右反転のみを施した A より，提案手法である自作の Hard Example を追加した B の識別率が 3.7%上昇していた．これにより Hard Example を追加した訓練データセットで学習すると，識別率が上昇することが確認できた．

A では識別できていなかったが，Hard Example を追加した B で認識が成功した画像として図 7(a)に示すような画像が特徴的に現れた．これらは，照明変化が大きいため A では認識に失敗したと考えられる．しかし，B ではこれに対応するような Hard Example が訓練データに加えられているため，認識に成功したと考えられる．しかし，姿勢変化が原因となり A で失敗していた画像が B で認識に成功したということはあまりなかった．これは，テストセットに含まれていた画像がほぼ正面顔になっているか，同図(b)のように極端に向きが違うものしかなかったため，少し方向を変えた程度の Hard Example では学習しきれなかったと考えられる．

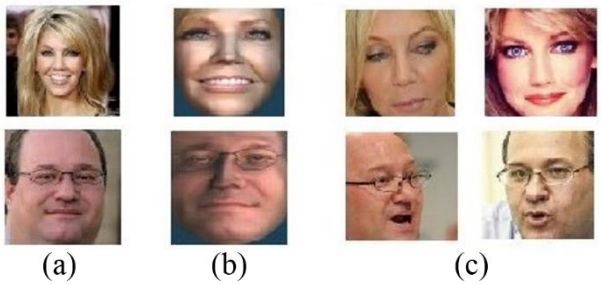


(a)

(b)

図 7 Hard Example を追加して認識に成功した例(a)と A と変わらず認識に成功しなかった例(b)

また，A では認識に成功していたが，Hard Example を加えた B で認識に失敗した画像も存在していた．図 7 にその画像を示す．図 8(a)が顔 3D モデルの元となった正面顔画像であり，同図(b)が作成した顔 3D モデル，同図(c)が Hard Example を追加した際，認識に失敗するようになった画像である．今回，Hard Example は顔 3D モデルに対して照明変化と姿勢変化，オクルージョンを加えて作成したものだけである．そのため，顔 3D モデルの元になった画像に対して表情変化が大きいものは認識できないと考えられる．更に，一枚の正面顔画像から顔 3D モデルを生成するため，鼻の高さや目の彫りの深さなどの奥行きに関する情報は得られない．そのため，基本となる一般的な顔から作成されたワイヤーフレームモデルと顔が大きく異なっている場合も，認識率が低下する．特に，アジア人の顔は基本となるワイヤーフレームモデルと大きく異なっていたため，アジア人の認証率は Hard Example を追加すると減少することもあった．Hard Example を加えたことにより，A で学習できていたことが薄まってしまったと考えられる．このことから，不用意に Hard Example を加えるのではなく，適切な枚数にしたほうが良いと考えられる．



(a) (b) (c)

図 8 Hard Example を追加して認識ができなくなった例

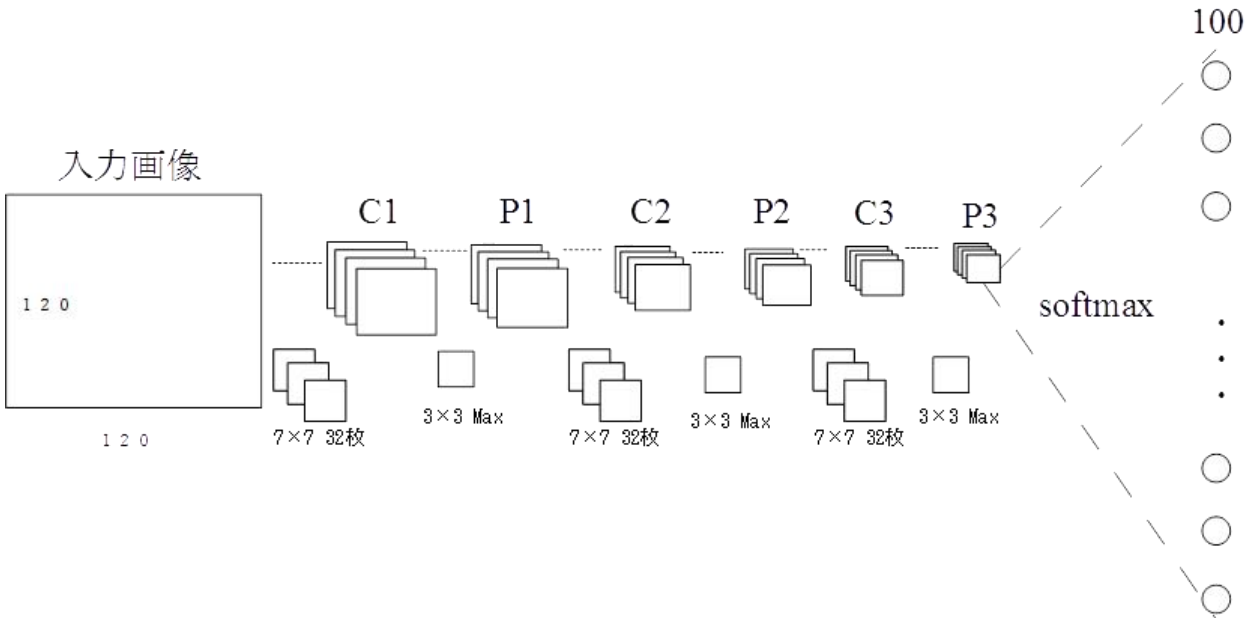


図 6 今回使用した CNN の構造の概略図

5. おわりに

ディープラーニングでは，Hard Example が存在する，訓練データを大量に必要がある，適切な構造を見つける必要があるという問題が存在した．そこで，本研究では，Hard Example を自作し，それを訓練データに加えることにより，以上の問題を解決する手法を提案し，実験を行った．

実験では，先行研究で使われていた Data Augmentation のみ利用して枚数を増加させた訓練データセットと自作した Hard Example を加えた訓練データセットのそれぞれで，一般的な CNN を学習させた際の識別率を評価した．作成した Hard Example は光学的変化として光源の変化，幾何学的変化として姿勢変化を与えたものとオクルージョンを発生させたものとした．Hard Example を追加させたとき識別率は 3.7%上昇していた．特に光源が大きく変化している画像の認識に成功できるようになった．

今回は，光源の変化と姿勢変化，オクルージョンを与えた Hard Example しか利用しなかったが，それ以外が原因となる Hard Example も作成したい．特に顔画像では分光反射率の変化が重要になると考えられる，例えば化粧や，老化，汗や脂などにより認識が失敗することがあるため，それに対応できると考えられる．

今回の実験により，Hard Example を加えすぎることによる障害も確認できた．どれだけ自作した Hard Example を加えてよいか調査するのも今後の課題となる．また，これらの問題は，顔認証以外のディープラーニングにおいてももちろん起こるため，顔認証以外の認識に実験をし，評価を行うことも課題としたい．今回の Hard Example は，基本となるワイヤーフレームモデルを個人の顔にあわせて変形させ，顔 3D モデルを作成し，それを利用して作成した．この方法で可能なのは人間の顔はある程度の大きさや形が決まっているためであった．しかし，顔に限らずに様々な対象で 3D モデルを利用した Hard Example の生成する場合，どのように 3D モデルを作成するのかが問題となる．現在は，3 次元情報を取得できる特殊な機材の利用等を考えている．

参考文献

[1] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", Computer Vision and Pattern Recognition, pp.1701-1708(2014).

[2] Yi Sun, "Deep Learning Face Representation from Predicting 10，000 Classes", Computer Vision and Pattern Recognition, pp.1891-1898 (2014).

[3] Krizhevsky, Alex, Ilya Sutskever, GeoffreyE. Hinton, "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems(2012)

[4] Patrice Y. Simard, Dave Steinkraus, John C. Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis", International Conference on Document Analysis and Recognition, pp.958-963(2003).

[5] Liming Wang, Jianbo Shi, Gang Song, I-fan Shen, "Object Detection Combining Recognition and Segmentation", Asian Conference on Computer Vison, pp.189-199(2007)

[6] Gary B. Huang, Manu Ramesh, Tamara Berg, Erik Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments", University of Massachusetts, Amherst, Technical Report 07-49(2007)