

2021年7月27日

**ディープラーニングで
飛躍的に進化した顔画像識別技術**

澤田雅之技術士事務所(電気電子部門)所長
元警察大学校警察情報通信研究センター所長

澤田 雅之

【 目 次 】

I 「人の目」を遥かに凌駕する顔識別技術

II 顔識別に用いるディープラーニングの基礎

II-1 ニューラルネットワークのディープラーニング

II-2 ニューラルネットワーク・アーキテクチャ 進化の系譜

II-3 ディープラーニングにおける一般的な学習方法

III CNNの全体構成と顔識別に向けたディープラーニング

III-1 顔画像の検出に向けたR-CNNのディープラーニング

III-2 顔画像の識別に向けたCNNのディープラーニング

IV 米国立標準技術研究所の顔認識技術に係るベンダーテスト(FRVT)

IV-1 2013年FRVT→2018年FRVT ディープラーニングが識別性能を飛躍的に向上

IV-2 2018年FRVT 品質の劣る顔画像(WEBカメラ画像)に対する識別特性

IV-3 2018年FRVT 顔の経年変化に対する識別特性

IV-4 2018年FRVT 真横顔に対する識別特性

IV-5 2018年FRVT 同一人物の複数ショット照合による識別精度の向上

V 顔識別技術におけるディープラーニングの効能・効果

☆ おことわり ☆

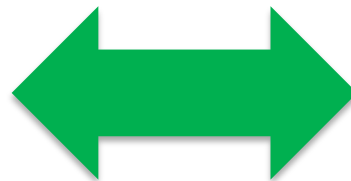
本資料の中で用いた全ての顔写真は、正当な理由・目的のもとにインターネット上に公開された顔写真を引用したものです。

I

「人の目」を遥かに凌駕する 顔識別技術

** 我が国の顔識別技術は世界のトップランナー **

「人の目」を遥かに凌駕する顔識別技術(1/3)

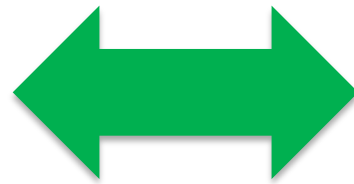


(出典)
https://search.yahoo.co.jp/image/search?p=□□□□&aq=1&ai=A45UppZZQ.qWtYro2JMNfA&ts=7671&ei=UTF-8&fr=top_ga1_sa



顔識別技術を用いて、左の画像で右の画像を、
数十万枚の中から類似度第1位に瞬時検索

「人の目」を遥かに凌駕する顔識別技術(2/3)

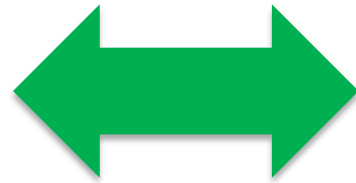


(出典)
https://search.yahoo.co.jp/image/search;_ylt=A2Riol7hIzleUj4ASCKU3uV7?p=○○○&aq=-1&oq=&ei=UTF-8#mode%3Dsearch



顔識別技術を用いて、左の画像で右の画像を、
数十万枚の中から類似度第1位に瞬時検索

「人の目」を遥かに凌駕する顔識別技術(3/3)



(出典)
https://search.yahoo.co.jp/image/search;_ylt=A2Riol7hIzleUj4ASCKU3uV7?p=○○○&aq=-1&oq=&ei=UTF-8#mode%3Dsearch



顔識別技術を用いて、左の画像で右の画像を、

数十万枚の中から類似度約百位に瞬時検索



識別精度は、顔画像の品質次第

検索速度は超高速

数百万枚の顔画像の中から類似度第1位を
瞬時に検索・抽出



顔画像同士のマッチングではなく、各顔画像から
生成した「顔特徴ベクトル」間の距離計算を行うこ
とにより、類似度を算定



***** ベクトル間の距離計算は超高速 *****

顔画像品質と識別精度との関係

顔画像の緻密さ

目間画素数(両目の中心を結ぶ直線上の撮像素子数に相当)が数十画素あれば十分 → 数百画素を確保しても、識別精度の向上には繋がらない。

顔画像の鮮明さ

ブレ、ボケ、ノイズ、低コントラストが識別精度を劣化させる。

顔の撮影角度

上下方向は30度程度まで、左右方向は45度程度までOK
→ 角度が大きいくほど、他の劣化要因への余裕度が減少

顔の表情・経年変化・整形手術・メガネ等の有無

カメラ側の工夫では対処不可能な要因については、顔識別のアルゴリズムで対処 → 識別精度劣化の主要因ではなくなっている。

顔識別技術の動作原理(ディープラーニングを用いない場合)

顔画像の検出

顔の基本的な構造を表現した「顔パターン画像(大きさ可変)」を用いて、フレーム画像等の2次元静止画像を隅から隅までスキャン → 2次元静止画像の中から「顔パターン画像」に合致する矩形領域を「顔画像」として検出

顔画像からの特徴抽出

検出した顔画像から、瞳の中心や鼻の先端等の位置を精密に求め、主成分分析等の手法で作成した「顔特徴抽出フィルタ」を適用 → 検索・照合に用いる高次元(顔特徴抽出フィルタ数に応じた次元)の「顔特徴ベクトル」を生成

顔画像の検索・照合

顔画像の検索・照合は、検索・照合したい顔画像と、検索・照合の対象となる顔画像との間の類似度を算定して行う。 → 類似度は、顔画像の「顔特徴ベクトル」間の距離計算により算出

顔識別技術の動作原理(ディープラーニングを用いた場合)

顔画像の検出(R-CNNのディープラーニングを利用)

R-CNN(後述)を用いて、フレーム画像等の2次元静止画像中に写っている「人の顔」を瞬時に全て検出 → 「人の顔」が写っている領域を四角形の枠で個々に特定し、「人の顔」らしさを表す信頼度の数値を枠の近傍に表示

顔画像からの特徴抽出(CNNのディープラーニングを利用)

CNN(後述)を用いて、R-CNNで検出した顔画像から様々な特徴を抽出し、それらを高次元の数値ベクトルに集約することにより「顔特徴ベクトル」を生成

顔画像の検索・照合

顔画像の検索・照合は、検索・照合したい顔画像と、検索・照合の対象となる顔画像との間の類似度を算定して行う。 → 類似度は、顔画像の「顔特徴ベクトル」間の距離計算により算出

II

顔識別に用いるディープラーニングの基礎

CNN(畳み込みニューラルネットワーク)と、
R-CNN(領域提案できる畳み込みニューラルネットワーク)

Ⅱ - 1

ニューラルネットワークの ディープラーニング

ニューラルネットワークとは？

AIで用いるニューラルネットワークとは、人の頭脳内部での神経回路網(ニューラルネットワーク)の仕組みと働きを、入力層ー隠れ層ー出力層として、コンピュータ上で数学的に模したものの。

具体的には、



各層には、人の脳のニューロン(神経細胞)に相当する多数のノードを配置

➡ 各ノードは、人の脳のシナプス(神経細胞間を結ぶ軸索)に相当する信号回路で結ばれ、ネットワークの結節点を構成

このため、



人の頭脳と同様にニューラルネットワークは、大量の教材を用いた学習を反復して、隠れ層内の各ノードごとの信号の伝わり具合を変えていくことにより、ニューラルネットワーク内部に暗示的に識別・判断・予測等の能力を創出できる。

ディープラーニングとは？

前ページからの再掲

人の頭脳と同様にニューラルネットワークは、大量の教材を用いた学習を反復して、隠れ層内の各ノードごとの信号の伝わり具合を変えていくことにより、ニューラルネットワーク内部に暗示的に識別・判断・予測等の能力を創出できる。

そこで、



ディープラーニングの「ディープ」とは、識別・判断・予測等の能力向上のため、隠れ層を多層化して深く学習(ディープラーニング)できるようにしたことを意味する。

また、



人には得意・不得意の分野や能力の高低があるように、ディープラーニングにも、ニューラルネットワークの構成の仕方(アーキテクチャ)の違いにより得意・不得意の分野が生じ、また、その鍛え方(学習の方法や学習用教材)の違いにより識別・判断・予測等の能力に高低が生じる。

Ⅱ - 2

ニューラルネットワーク・アーキテクチャ 進化の系譜

CNN(畳み込みニューラルネットワーク)の系譜と、
RNN(再帰型ニューラルネットワーク)の系譜に2大別

アーキテクチャの基本は、CNN と RNN

ニューラルネットワークは、その構成の仕方(アーキテクチャ)によって、

2次元静止画像を入力として事物の検出や分類などができるCNN
(Convolutional Neural Network : 畳み込みニューラルネットワーク)と、

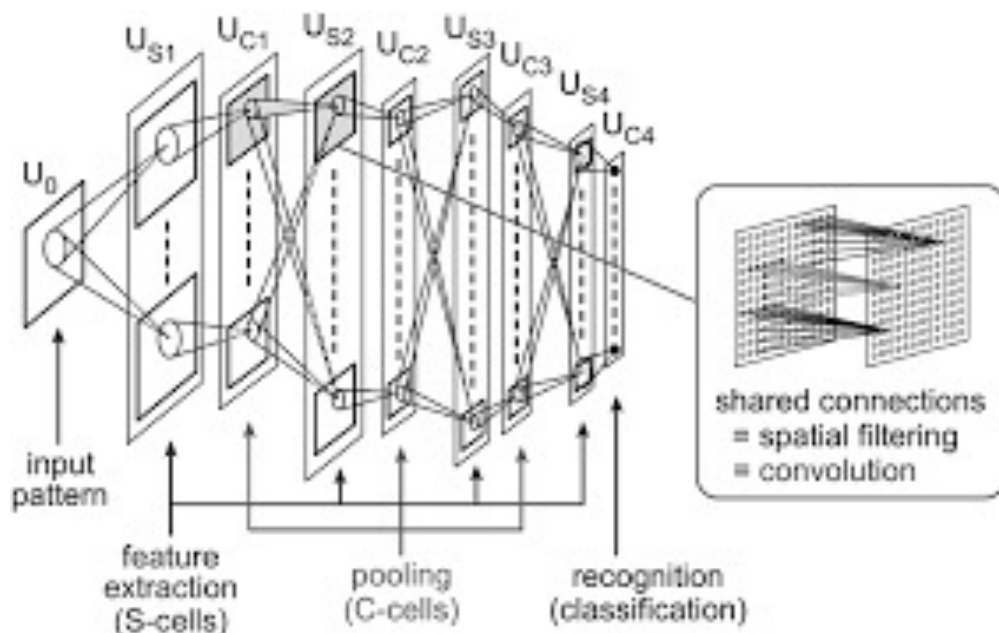
時系列のベクトルデータを入力として予測などができるRNN
(Recurrent Neural Network : 再帰型ニューラルネットワーク)に大別される。

➡ 顔識別技術に用いるのは、CNN(畳み込みニューラルネットワーク)

CNN(畳み込みニューラルネットワーク)の系譜 1/3

**** CNNの原型となったネオコグニトロンは日本人が提唱 ****

1979年、NHK放送科学基礎研究所の福島邦彦氏が、「ネオコグニトロン」の研究論文を電子通信学会論文誌に発表 → 「ネオコグニトロン」は、学習によってパターン認識能力を獲得していくニューラルネットワーク → 手書き数字の認識で高い能力を発揮できることを実証して、CNNの原型となった。



【左は「ネオコグニトロン」の構成図】

$U_{s1} \sim U_{s4}$ の単純細胞層が局所特徴の抽出を担う。 → CNNの畳み込み層に相当

$U_{c1} \sim U_{c4}$ の複雑細胞層が局所特徴の位置ずれ吸収を担う。 → CNNのプーリング層に相当

CNN(畳み込みニューラルネットワーク)の系譜 2/3

**** 2012年の画像認識コンペティションで圧勝したCNN ****

【 2012年開催の国際的な画像認識コンペティション(ILSVRC2012) 】

初登場したCNNのディープラーニングによる認識手法は、飛躍的な認識性能を発揮して、人の手で明示的に設計してプログラミングした認識手法を全て圧倒


これを契機として、



CNNのディープラーニングによる認識手法は、その有用性が瞬く間に多方面で認められ、2次元静止画像からの対象物の検出や分類などを行う強力な手段として、今日では広範な活用が進展し社会に広く深く浸透しつつある。

顔識別への活用では、

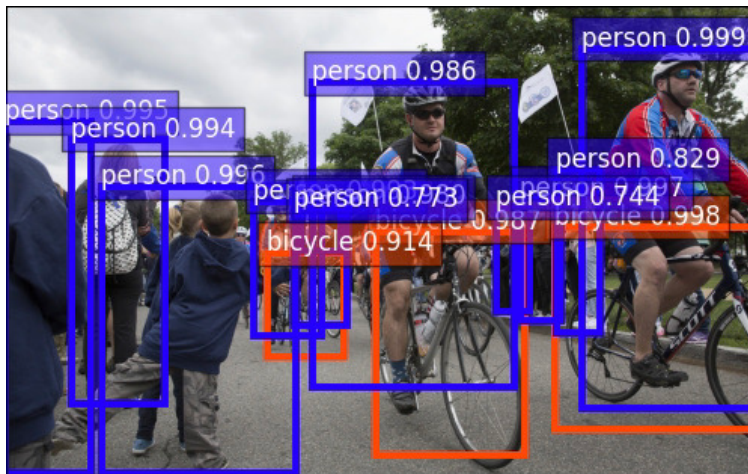


スマートフォンの顔認証によるロック解除手段や、フェイスブックの顔画像自動探知機能として、既に身近な存在  ICパスポートに記録されている顔画像とビデオカメラで捉えた顔画像との自動照合による出帰国審査の自動化ゲートなどが既に実用化

CNN(畳み込みニューラルネットワーク)の系譜 3/3

** 対象物の存在領域も示せるR-CNNへの機能拡張 **

CNNは、Region Proposal Network(領域提案ネットワーク：小規模なCNN)を組み込んで、R-CNN(Regions with Convolutional Neural Network：領域提案できる畳み込みニューラルネットワーク)に機能を拡張 ← 第3世代のFASTER R-CNNのことであり、R-CNNとFAST R-CNNでは、領域提案を外付けプログラムで処理



出典：https://cv.gluon.ai/build/examples_detection/demo_faster_rcnn.html

R-CNNでは、2次元静止画像中から「学習済みの事物」を検出 → 写っている領域を四角形の枠で特定し、分類結果を信頼度の数値と共に表示

映像(動画)は、2次元静止画像であるフレーム画像が1秒間に数十枚連続したもの

→ フレーム画像を順次にR-CNN(FASTER R-CNN)で処理すれば、映像(動画)をリアルタイムに解析できる。

RNN(再帰型ニューラルネットワーク)の系譜

**** 時系列のベクトルデータを入力して予測などができる。 ****

オリジナルのRNN(Recurrent Neural Network)

RNNは、入力層ー隠れ層ー出力層で構成されるニューラルネットワーク ➡ 「隠れ層」のフィードバックループにより、逐次入力される時系列ベクトルデータ間の関係性を保持できる。 ➡ 関係性の学習により、例えば、逐次入力されるセンサーデータに基づく「今後の予測値」が出力できる。

LSTM(Long Short-Term Memory)

RNNでは、長期間に渡る関係性の保持が困難 ➡ 入力ゲート・出力ゲート・忘却ゲートを備えた「LSTMセル」を、RNNの「隠れ層」の代わりに用いたのがLSTM ➡ 逐次入力される時系列ベクトルデータ間の関係性を長期間にわたって保持できる。 ➡ 音声認識などに活用

Transformer

Transformerは、「注意機構」を中核としたニューラルネットワーク ➡ 「注意機構」の働きにより、入力された時系列ベクトルデータの全ての関係性をベクトル空間に投影できるので、時系列ベクトルデータの入力時点に関わらず、全てのベクトルデータに対して瞬時に関係性の高低を判別できる。 ➡ 今日では、自然言語処理や自動翻訳などの精度向上の決め手となっている。

Ⅱ - 3

ディープラーニングにおける
一般的な学習方法

ディープラーニングではどのように学習するのか？

ディープラーニングでは、入力された情報とそれに対する望ましい出力結果の組み合わせをニューラルネットワークに反復して学習させ、ニューラルネットワークを次第に鍛え上げていき、実現したいアルゴリズムをネットワーク全体にわたって暗示的に生成

具体的には、



【 AIフレームワークの利用が効果的かつ効率的 】

ネットが無償入手できるAIフレームワーク(TensorFlowやPyTorchなど、ディープラーニングに必要な機能を揃えたソフトウェア群)を用いた学習が効果的かつ効率的

➡ ニューラルネットワークの構成をパラメータで定義し、事前に準備した学習データをAIフレームワークにセットして自動的に反復学習させ、テストデータで性能確認しつつ、生成されるアルゴリズムが実用レベルに達するまで学習を続ける。

➡ 大事なポイントは、構成定義段階では望ましい構成を試行錯誤で追求することと、反復学習段階では良質な学習データの大量収集や過学習に陥らないようにすること

ディープラーニングの学習が目指すのは「汎化」

**** ネット上で「本物の猫」を自動的に見つけ出すための学習を例として ****

【ディープラーニングの「学習フェーズ」と「推論フェーズ」】

学習フェーズで、学習用教材を十分にマスターすることにより、**推論フェーズ**で、教材以外にも応用が効くようにすること(汎化)が極めて重要

「学習フェーズ」

学習用教材として、「本物の猫の画像」と「猫に似ているが猫ではない画像」を多数準備

➡ 入力層に入力した「本物の猫の画像」の1枚ごとに、出力層の分類結果を「猫」に少しずつ近づけていく作業を反復。併せて、入力層に入力した「猫に似ているが猫ではない画像」の1枚ごとに、出力層の分類結果を「猫」から少しずつ遠ざけていく作業を反復 ← これが**学習フェーズ**

「推論フェーズ」

学習フェーズでニューラルネットワーク内に生成した「汎化したアルゴリズム」は、「本物の猫を見つけ出す画像フィルタ」であり、ネット投稿画像・動画などの中から「本物の猫」を自動的に見つけ出すことができる。 ← これが**推論フェーズ**

「汎化したアルゴリズム」の生成方法

*** 「本物の猫」を見つけ出すアルゴリズムを例として ***

入力層へ入力した「本物の猫の画像」の1枚ごとに、「分類結果が猫ではないとした誤差の度合い」を隠れ層内に逆伝播。また、入力層へ入力した「猫に似ているが猫ではない画像」の1枚ごとに、「分類結果が猫であるとした誤差の度合い」を隠れ層内に逆伝播

➡ 隠れ層内で、ノードへの信号入力時の重み付け値やノードからの信号出力時のバイアス値を、「分類結果の誤差の度合い」が小さくなる方向に少しずつ変化

アルゴリズムを暗示的に生成

学習を反復していく中で、隠れ層内には「普遍的な猫の特徴」を抽出して「猫という概念」に昇華する「汎化したアルゴリズム」が、ノードへの信号入力時の重み付け値やノードからの信号出力時のバイアス値の集合体として、ネットワーク全体に暗示的に生成

「過学習」に注意！

「過学習」とは、教材を何度も反復して学習させていくうちに、最終目的である「汎化」を通り越してしまい、教材以外には応用が効かなくなる現象のこと ➡ 「過学習」を避けるには、良質な大量の教材の準備と、反復学習が単調化しない工夫が肝要

III

CNNの全体構成と 顔識別に向けたディープラーニング

CNN(畳み込みニューラルネットワーク)の全体構成

CNNの基本的な構成は、入力層－隠れ層(畳み込み層－プーリング層－全結合層)－出力層であり、入力された信号は、隠れ層内の各層を含めて、入力層から出力層までのネットワーク全体をフィードフォワードで伝播し、フィードバックされることはない。



【入力層】

2次元静止画像を扱うCNNの入力層は、静止画像の画素数に応じた2次元空間であり、モノクロ画像ではチャンネル数は1で、RGB画像ではチャンネル数は3となる。

【隠れ層】

精度向上のために隠れ層内は多層化され、([複数の畳み込み層－プーリング層]－……－[複数の畳み込み層－プーリング層]－[複数の全結合層])といった具合に、100層以上で構成されることもある。

【出力層】

分類結果を求める場合は、Softmax関数を介して、出力層には分類結果が信頼度の数値とともに出力される。また、特徴ベクトルを求める場合は、全結合層の途中段階での信号出力を整序することにより、出力層には高次元の数値ベクトルが出力される。

【 CNNの隠れ層内各層の役割 】

【 畳み込み層 】

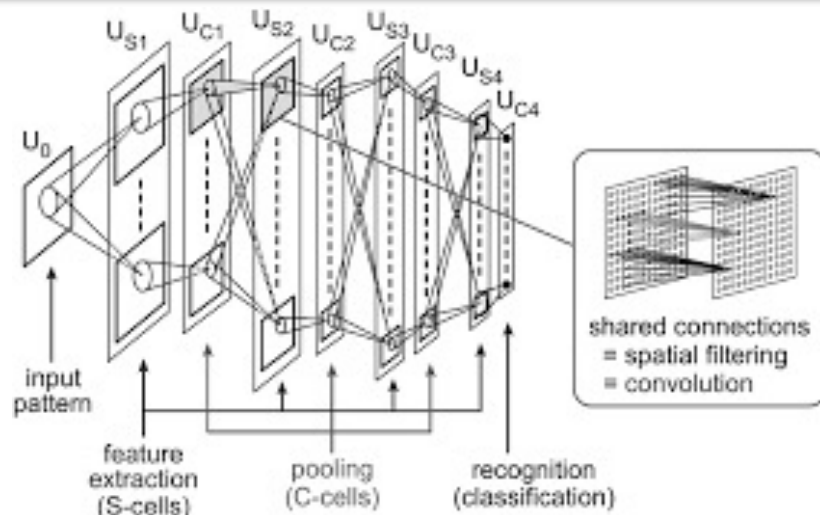
入力層などの前層のノード全体を細分化し、「部分的な特徴パターン」で一斉にスキャン
➡ 部分的な特徴を抽出してその場所をマッピングする画像フィルタとして働き、「部分的な特徴パターン」の数に応じたチャンネル数のマップを出力

【 プーリング層 】

マッピングした部分的な特徴を損なわないようにマップ全体を縮める(重要な情報を失わないように圧縮する)。 ➡ マップ内での部分的な特徴の位置変動の影響を減らすように働き、出力されるチャンネル数は変わらない。

【 全結合層 】

この層に入力された全てのチャンネルの特徴を統合するように働く。



【 左は「ネオコグニトロン」の構成図 】

U_{s1} ~ U_{s4}の 単純細胞層が局所特徴の抽出を担う。 ➡ CNNの畳み込み層に相当

U_{c1} ~ U_{c4}の 複雑細胞層が局所特徴の位置ずれ吸収を担う。 ➡ CNNのプーリング層に相当

Ⅲ-1

顔画像の検出に向けた R-CNNのディープラーニング

R-CNNによる顔画像の検出

顔画像の検出とは

顔画像の検出とは、フレーム画像等の2次元静止画像の中から「人の顔」を見つけ出して、その写っている領域を四角形の枠で囲んで特定すること。



R-CNNとは

R-CNN(領域提案できる畳み込みニューラルネットワーク)とは、CNN(畳み込みニューラルネットワーク)に、RPN(領域提案ネットワーク)を組み込んだもの。

➡ CNNが、2次元静止画像の中から「人の顔」を見つけ出して信頼度の数値を出力し、RPNが、「人の顔」の写っている領域を特定して四角形の枠で囲む。



リアルタイムな顔画像検出

動画を構成するフレーム画像に多数の「人の顔」が写っている場合でも、R-CNN(Faster R-CNN)は、瞬時に「人の顔」を全てを見つけ出して、それらが写っている領域を四角形の枠で個々に特定し、「人の顔」らしさを表す信頼度の数値を枠の近傍に表示できる。 ➡ 動画から顔画像をリアルタイムに検出可能

R-CNNで顔画像を検出するための学習フェーズ

学習用教材の準備

様々な条件下(性別、年齢、人種、表情、撮影角度、髪型、髭の有無、帽子・メガネ・マスクの有無など)で撮影された「人の顔画像」と、「人の顔と誤認する恐れがある物体の画像(人の顔がプリントされたTシャツ、お面、案山子など)」を、多数準備する。



誤差逆伝播による反復学習

「人の顔画像」については、「人の顔」らしさを表す信頼度の数値を少しずつ高めていくとともに、「人の顔」が写っている領域を四角形の枠で囲む精度を少しずつ高めていく。また、「人の顔と誤認する恐れがある物体の画像(人の顔がプリントされたTシャツ、お面、案山子など)」については、「人の顔」らしさを表す信頼度の数値を少しずつ低めていく。



テストデータによる確認

学習用教材とは別にテストデータを多数準備しておき、反復学習の進捗に伴う「顔画像の検出精度」の向上具合について、「過学習」に陥っていないかを含めて確認する。

Ⅲ-2

顔画像の識別に向けた CNNのディープラーニング

CNNによる「顔特徴ベクトル」の生成

顔画像の識別とは

識別したい顔画像と、識別対象顔画像との「類似度」を算定すること。 ➡ その算定は、顔画像から生成した「顔特徴ベクトル」を用いて、ベクトル間の距離計算により行う。

「顔特徴ベクトル」の生成方法

CNNを用いて、入力層に入力された顔画像の様々な特徴を隠れ層内の畳み込み層とプーリング層で抽出して、隠れ層内の全結合層における概念統合に向けた途中段階での信号出力を整序することにより、出力層に「顔特徴ベクトル」として出力する。

理想的な「顔特徴ベクトル」

顔画像の識別とは、識別対象顔画像から生成した「顔特徴ベクトル」が分布するベクトル空間において、識別したい顔画像から生成した「顔特徴ベクトル」との距離計算を行うこと。 ➡ 顔の表情や長期経年変化などが影響して「別人」に見える場合でも、本人の顔画像の「顔特徴ベクトル」は常に近接し、他人の顔画像の「顔特徴ベクトル」とは常に離間するのが、理想的な「顔特徴ベクトル」

理想的な「顔特徴ベクトル」を生成するための学習フェーズ

学習用教材の準備

多様（性別、年齢、人種など）で多数の人々の顔画像を準備するとともに、同一人物についても様々な条件下（表情、撮影角度、髭の有無、メガネ・マスクの有無など）で撮影された複数の顔画像を準備する。



反復学習の第1段階

任意の同一人物の任意の2枚の顔画像を選択するとともに、任意の別人の任意の1枚の顔画像を選択する。 → これらの3枚の顔画像を1つのセットとして、CNNの入力層への学習用教材として用いて、出力される3つの「顔特徴ベクトル」を比較する。



次のページへ

前のページから

反復学習の第2段階

同一人物の2つの「顔特徴ベクトル」間の距離を少し狭めるとともに、別人の「顔特徴ベクトル」との距離を少し広げるよう、「距離の開き具合」を隠れ層内に逆伝播させて、ノードへの信号入力時の重み付け値やノードからの信号出力時のバイアス値を少しずつ変化させる。

アルゴリズムの暗示的な生成

反復学習の第1段階と第2段階を繰り返すことにより、隠れ層内には、様々な異なる条件下（性別、年齢、人種、表情、撮影角度、髭の有無、メガネ・マスクの有無など）であっても、別人の顔画像の「顔特徴ベクトル」からは離間するが、同一人物の顔画像であれば近接する「顔特徴ベクトル」を生成するアルゴリズムが、ノードへの信号入力時の重み付け値やノードからの信号出力時のバイアス値の集合体として、ネットワーク全体に暗示的に生成される。

IV

米国立標準技術研究所の 顔認識技術に係るベンダーテスト(FRVT) ～ ディープラーニングの活用効果が明白 ～

【顔画像及びデータの出典】

出典1 : NIST Interagency Report 8009(2014年5月公表)

出典2 : NIST Interagency Report 8271(2019年9月公表)

米国立標準技術研究所の顔認識技術ベンダーテスト(FRVT)

米国立標準技術研究所(NIST : National Institute of Standards and Technology)では、顔認識技術に係る各種の評価試験を2000年以来実施


中でも、



2013年と2018年に実施した顔認識技術ベンダーテスト(FRVT : Face Recognition Vendor Test)は、「顔を識別するアルゴリズム」の生成にディープラーニングを活用することによる識別性能の改善効果を理解する上で重要

その理由は、



2013年のFRVTではディープラーニングを活用したベンダーは皆無。しかし、2018年のFRVTでは多くのベンダーがディープラーニングを活用し、2013年と類似したテスト環境における識別精度が桁違いに向上  2013年のFRVTの結果は、「NIST Interagency Report 8009」として2014年5月に公表。2018年のFRVTの結果は、「NIST Interagency Report 8271」として2019年9月に公表。

2013年のFRVTにおける参加企業等

NEC(日本)

清華大学(Prof. Wen) (中国)

東芝(日本)

清華大学(Prof. Su) (中国)

Ayonix Inc. (日本)

Beijing Ivsign Technology Co. Ltd. (中国)

3M/Cogent (米国)

中国科学院 Automation (中国)

HP/Virage (米国)

中国科学院 Computing Technology (中国)

Decatur Industries Inc. (米国)

Zhuhai Yisheng Electronics Tech. Co. Ltd. (中国)

Safran Morpho (フランス)

JunYu Technology Co. Ltd. (中国)

Cognitec (ドイツ)

Neurotechnology (リトアニア)

青字は、2013年と2018年の
両FRVTに参加したベンダー

2018年のFRVTにおける参加企業等(1/2)

NEC(日本)

東芝(日本)

グローリー(日本)

Ayonix Inc.(日本)

Cognitec(ドイツ)

Dermalog(ドイツ)

Smilart(ドイツ)

Idemia(フランス)(旧Safran Morpho) 中国科学院 Shenzhen Inst. Adv. Tech.(中国)

Thales(フランス)(旧3M/Cogent) TongYi Transportation Technology(中国)

Quantasoft(チェコ) Newland Computer Co. Ltd(中国)

Eyedeia Recognition(チェコ) Gorilla Technology(台湾)

Neurotechnology(リトアニア) Alchera(韓国)

Zhuhai Yisheng Electronics Tech. Co. Ltd.(中国)

ハイクビジョン(中国)

メグビー(中国)

センスタイム(中国)

ダーファ(中国)

依図テクノロジー(中国)

KanKan Ai(中国)

青字は、2013年と2018年の
両FRVTに参加したベンダー

Loginface Corp(韓国)

2018年のFRVTにおける参加企業等(2/2)

マイクロソフト(米国)

3Divi(米国)

Anke Investments (米国)

Aware (米国)

Camvi Technologies (米国)

Ever AI (米国)

Incode Technologies (米国)

Noblis (米国)

Rank One Computing (米国)

RealNetworks (米国)

Shaman Software (米国)

TigerIT Americas LLC (米国)

Vigilant Solutions (米国)

N-Tech Lab (ロシア)

Tevian (ロシア)

Vocord (ロシア)

Microfocus (英国)

Imagus Technology Pty Ltd (オーストラリア)

Innovatrics (スロバキア)

Synesis (クロアチア)

Visidon (フィンランド)

VisionLabs (オランダ)

Lookman Electroplast Industries (インド)

Alivia/Innovation Sys. (国不詳)

2013年と2018年の両FRVTで 共通して用いられた2種類の顔画像



MUGSHOT画像

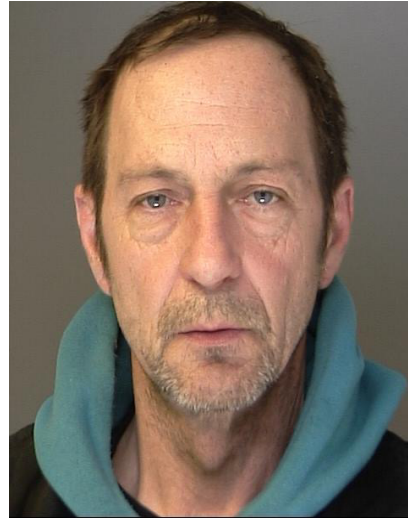
米国の警察の実際の現場で、デジタルスチルカメラにより撮影

➡ 我が国の被疑者写真に相当 (高品質な顔画像)

WEBカメラ画像

米国の国境警備隊が勾留した中南米からの不法入国者を、WEBカメラにより撮影 ➡ 国境警備隊職員の指示に基づき、WEB

カメラに顔を向けた瞬間を捉えている。 (品質が劣る顔画像)



MUGSHOT画像

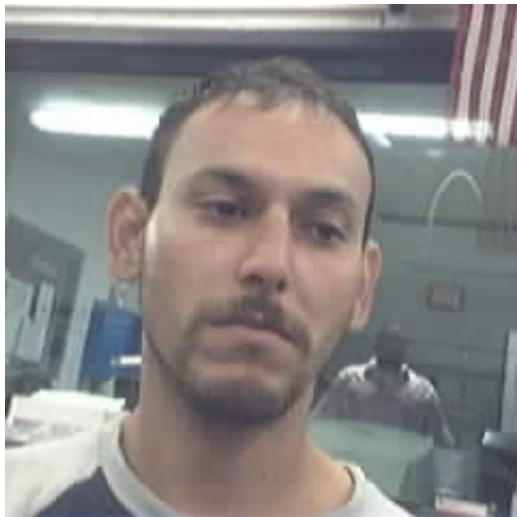


同一人物の複数のMUGSHOT画像

MUGSHOT画像の品質



- オリジナル画像サイズ :
480 × 640、240 × 240、768 × 960
- JPEG圧縮後の画像情報量 :
平均48kB
- 目間画素数 : 平均107画素で、標準偏差は40画素
- ポーズ : 顔を真正面から捉えた画像が多いが、中には、5度から10度ほど左右を向いている画像がある。また、視線がカメラに向けられていない画像が少なくない。
- 照明 : 適切な照明のもとで撮影されている画像が多い。



WEBカメラ画像

WEBカメラ画像の品質



- オリジナル画像サイズ :
240 × 240
- JPEG圧縮後の画像情報量 :
平均5.7kB
- 目間画素数 : 平均45画素で、標準偏差は12画素
- ポーズ : 顔を真正面から捉えた画像は少なく、5度から10度ほど左右を向いている画像や、下から見上げるように捉えた画像が多い。
- 照明 : 殆どが室内照明下で撮影されているため、照明が不十分で顔に影ができている画像が大半である。

2013年と2018年の両FRVTで用いられた識別精度評価方法

2013年と2018年のFRVTでは、1対1の顔画像認証(同一人物であるか否かを顔画像により確認すること)の識別精度ではなく、1対多数の顔画像照合(同一人物の顔画像を多数の顔画像の中から探し出すこと)の識別精度を、各種試験により評価

そこで、

1対多数の顔画像照合では、探し出そうとする人物の顔画像(検索用顔画像)と、探す対象となる多数の顔画像(被検索顔画像)との間の「類似度」を算出 → この算出は、検索用顔画像から生成した「顔特徴ベクトル」と、各被検索顔画像から生成した各「顔特徴ベクトル」との、ベクトル間距離を計算して行う。

このため、

1対多数の顔画像照合の結果は、検索用顔画像との「類似度」が高い順(顔特徴ベクトル間の距離が短い順)に、被検索顔画像がスコア値(顔特徴ベクトル間距離の長さに応じた1~0の正数)とともにリストアップ → 1対多数の顔画像照合での識別精度の評価方法は、スコア値に対する「閾値」を設けるか否かにより、2種類に大別される。

(1) スコア値に対する「閾値」を設けない識別精度評価方法

1対多数の顔画像照合を行った結果として、検索用顔画像と被検索顔画像との「類似度」のスコア値が高い順に被検索顔画像がリストアップされる。



【スコア値に対する「閾値」を設けない場合の識別精度評価方法】

多数の検索用顔画像を用いて1対多数の顔画像照合を行い、その結果として、検索用顔画像と同一人物の被検索顔画像が、TOP1 (リストの最上位) にリストアップできなかった割合 (「本人見逃し率」) を調べる。

ここで、



スコア値に対する「閾値」を設けなければ、検索用顔画像と同一人物の被検索顔画像が存在しない場合に、TOP1における「他人誤認率」が100%となることに注意を要する。

(2) スコア値に対する「閾値」を設ける識別精度評価方法

1対多数の顔画像照合を行った結果として、検索用顔画像と被検索顔画像との「類似度」のスコア値が高い順に被検索顔画像がリストアップされる。



【スコア値に対する「閾値」を設ける場合の識別精度評価方法】

多数の検索用顔画像を用いて1対多数の顔画像照合を行い、その結果として、検索用顔画像と同一人物の被検索顔画像が、設けた「閾値」を上回ってTOP1(リストの最上位)にリストアップできなかった割合(「本人見逃し率」)を調べる。

ここで、



【「本人見逃し率」と「他人誤認率」は、トレードオフの関係であることに注意】

「閾値」を高く設定して「他人誤認率」を小さくするほど「本人見逃し率」は大きくなり、「閾値」を低く設定して「他人誤認率」を大きくするほど「本人見逃し率」は小さくなる。 ➡
「他人誤認率」が1/10、1/100、1/1000となる「閾値」をパラメータとして設定して、各「閾値」での「本人見逃し率」を調べて識別精度を評価する。

IV- 1

2013年FRVT → 2018年FRVT ディープラーニングが識別性能を飛躍的に向上

2013年のFRVTではディープラーニングの活用は未だ見られないが、2018年のFRVTでは多くのベンダーがディープラーニングを活用して、識別性能を飛躍的に向上させている。

そこで、2013年と2018年の両FRVTに参加した6社【NEC、Thales（旧3M/Cogent）、東芝、Neurotech.、Cognitec、Idemia（旧Safran Morpho）】について、同様な試験条件下での結果を各社ごとに比較すれば、ディープラーニングの導入効果を調べることができる。

なお、ここでの同様な試験条件とは、160万人分（160万枚）の顔画像データベースに対して、高品質なMUGSHOT画像により検索した場合の「本人見逃し率」と、品質の劣るWEBカメラ画像により検索した場合の「本人見逃し率」を、それぞれ求めるものである。

2013年FRVTと2018年FRVTの検索精度(試験方法)

2013年のFRVT



検索用MUGSHOT画像

検索対象は、160万人分(160万枚)の顔画像データベース(大半はMUGSHOT画像であるが、約96千人分のWEBカメラ画像を含む。)



- ① 5万人分(5万枚)のMUGSHOT画像(データベース内の同一人物画像とは別画像)で検索し、「閾値を設けないTOP1の本人見逃し率」を算出
- ② 10,660人分(10,660枚)のWEBカメラ画像(データベース内の同一人物画像とは別画像)で検索し、「閾値を設けないTOP1の本人見逃し率」を算出

2018年のFRVT



検索用WEBカメラ画像

検索対象は、160万人分(160万枚)の顔画像データベース(全てMUGSHOT画像)



- ① 154,549人分(154,549枚)のMUGSHOT画像(データベース内の同一人物画像とは別画像)で検索し、「閾値を設けないTOP1の本人見逃し率」を算出
- ② 82,106人分(82,106枚)のWEBカメラ画像(データベース内の同一人物画像とは別画像)で検索し、「閾値を設けないTOP1の本人見逃し率」を算出

2013年FRVTと2018年FRVTの検索精度(比較上の注意点)

2013年と2018年のFRVTで用いられた顔画像データベースはいずれも160万人分(160万枚)であるが、2018年の顔画像データベースは全て高品質なMUGSHOT画像であるのに対して、2013年の顔画像データベースには品質の劣るWEBカメラ画像が全体の6%ほど(160万枚中の96,885枚)含まれている。

しかし、



このような品質の劣る顔画像の混在は、「本人見逃し率」の劣化に繋がる。

そこで、



【劣化の度合いを推測する目安として】

2013年のFRVTでは、2万人分(2万枚)の顔画像データベース(全てMUGSHOT画像)を検索対象として、2万人分(2万枚)のMUGSHOT画像(データベース内の同一人物画像とは別画像)で検索し、「閾値を設けないTOP1の本人見逃し率」を算出しているのので、その結果を次ページに青字の括弧書きで示す。

2013年FRVTと2018年FRVTの検索精度(試験結果)

閾値を設けない場合のTOP1(類似度ランク1位)の本人見逃し率

NEC

MUGSHOT : 0.041(0.028) WEBCAM : 0.113
MUGSHOT : 0.003 WEBCAM : 0.010

Thales

(IB3M/Cogent)

MUGSHOT : 0.172(0.105) WEBCAM : 0.364
MUGSHOT : 0.006 WEBCAM : 0.020

東芝

MUGSHOT : 0.107(0.060) WEBCAM : 0.237
MUGSHOT : 0.007 WEBCAM : 0.022

Neurotech.

MUGSHOT : 0.205(0.142) WEBCAM : 0.702
MUGSHOT : 0.007 WEBCAM : 0.024

Cognitec

MUGSHOT : 0.136(0.085) WEBCAM : 0.576
MUGSHOT : 0.008 WEBCAM : 0.025

Idemia

(IBSafran Morpho)

MUGSHOT : 0.091(0.068) WEBCAM : 0.307
MUGSHOT : 0.009 WEBCAM : 0.032

黒字 : 2013年FRVT
赤字 : 2018年FRVT

青字の括弧内の数字
は、2013年FRVTにおいて、2万人分(2万枚)のMUGSHOT画像に対して、本人の別のMUGSHOT画像で、検索を2万回実施した結果(閾値を設けないTOP1の本人見逃し率)を示す。

2013年FRVTと2018年FRVTの検索精度(考察)

2013年と2018年の両FRVTに参加した6社【NEC、Thales(旧3M/Cogent)、東芝、Neurotech、Cognitec、Idemia(旧Safran Morpho)】は全て、高品質なMUGSHOT画像による検索時と、品質の劣るWEBカメラ画像による検索時のいずれも、2013年のFRVT(青字で括弧内に示した数値を含めて)と較べて2018年のFRVTでは、「本人見逃し率」の数値に桁違いの改善が見られる。



ディープラーニングは、2013年のFRVTでは未だ活用されず、2018年のFRVTでは活用が一気に進んだことから、識別性能が飛躍的に向上した主因はディープラーニングの活用にあると言える。

IV-2

2018年FRVT

品質の劣る顔画像（WEBカメラ画像） に対する識別特性

監視カメラや防犯カメラで捉えた顔画像は、MUGSHOT画像のような高品質な顔画像（つまり、緻密かつ鮮明な無表情顔をほぼ正面から捉えた顔画像）と較べて、緻密さや鮮明さ、撮影角度などの点で「品質」が劣る場合が多い。

このため、高品質なMUGSHOT画像による識別精度と、緻密さや鮮明さ、撮影角度の点で品質が劣るWEBカメラ画像による識別精度を詳しく比較検討すれば、監視カメラや防犯カメラで捉えた顔画像に顔識別技術を適用していく上での有用な知見を得ることができる。

MUGSHOT画像とWEBカメラ画像の検索精度 (試験方法その1)

*** MUGSHOT画像で検索する場合 ***

「他人誤認率が1/10、1/100、1/1000となる閾値」の求め方

160万人分(160万枚)の顔画像データベース(全てMUGSHOT画像)を検索対象として、331,254人分(331,254枚)のMUGSHOT画像(いずれもデータベース内に同一人物の顔画像が存在しない。)で検索し、「他人誤認率が1/10、1/100、1/1000となる閾値」を求める。



検索用MUGSHOT画像

「TOP1における本人見逃し率」の求め方

160万人分(160万枚)の顔画像データベース(全てMUGSHOT画像)を検索対象として、154,549人分(154,549枚)のMUGSHOT画像(データベース内の同一人物画像とは別画像)で検索し、「閾値を設けないTOP1の本人見逃し率」と、「他人誤認率が1/10、1/100、1/1000となる閾値を設けたTOP1の本人見逃し率」を算出する。

MUGSHOT画像とWEBカメラ画像の検索精度 (試験方法その2)

*** WEBカメラ画像で検索する場合 ***

「他人誤認率が1/10、1/100、1/1000となる閾値」の求め方

160万人分(160万枚)の顔画像データベース(全てMUGSHOT画像)を検索対象として、331,254人分(331,254枚)のWEBカメラ画像(いずれもデータベース内に同一人物の顔画像が存在しない。)で検索し、「他人誤認率が1/10、1/100、1/1000となる閾値」を求める。



検索性WEBカメラ画像

「TOP1における本人見逃し率」の求め方

160万人分(160万枚)の顔画像データベース(全てMUGSHOT画像)を検索対象として、82,106人分(82,106枚)のWEBカメラ画像(データベース内の同一人物画像とは別画像)を用いて検索し、「閾値を設けないTOP1の本人見逃し率」と、「他人誤認率が1/10、1/100、1/1000となる閾値を設けたTOP1の本人見逃し率」を算出する。

MUGSHOT画像とWEBカメラ画像の検索精度 (試験結果その1)

閾値を設けない場合のTOP1 (類似度ランク1位) の本人見逃し率

<i>NEC</i>	<u>MUGSHOT画像</u>	0.003	<u>WEBカメラ画像</u>	0.010
<i>依図</i>	<u>MUGSHOT画像</u>	0.004	<u>WEBカメラ画像</u>	0.008
<i>マイクロソフト</i>	<u>MUGSHOT画像</u>	0.003	<u>WEBカメラ画像</u>	0.011
<i>センスタイム</i>	<u>MUGSHOT画像</u>	0.005	<u>WEBカメラ画像</u>	0.016
<i>VisionLabs</i>	<u>MUGSHOT画像</u>	0.003	<u>WEBカメラ画像</u>	0.015
<i>N-Tech Lab</i>	<u>MUGSHOT画像</u>	0.006	<u>WEBカメラ画像</u>	0.017
<i>Lookman</i>	<u>MUGSHOT画像</u>	0.012	<u>WEBカメラ画像</u>	0.039
<i>Alivia</i>	<u>MUGSHOT画像</u>	0.007	<u>WEBカメラ画像</u>	0.023
<i>Neurotech.</i>	<u>MUGSHOT画像</u>	0.007	<u>WEBカメラ画像</u>	0.024
<i>東芝</i>	<u>MUGSHOT画像</u>	0.007	<u>WEBカメラ画像</u>	0.022



MUGSHOT画像とWEBカメラ画像の検索精度 (試験結果その2)

他人誤認率を一定(緑字1/10、赤字1/100、青字1/1000)としたTOP1の本人見逃し率



<i>NEC</i>	<u>MUGSHOT</u> 0.003、0.004、0.004	<u>WEBカメラ</u> 0.011、0.013、0.017
<i>依図</i>	<u>MUGSHOT</u> 0.004、0.007、0.012	<u>WEBカメラ</u> 0.011、0.017、0.027
<i>マイクロソフト</i>	<u>MUGSHOT</u> 0.004、0.008、0.014	<u>WEBカメラ</u> 0.016、0.024、0.037
<i>センスタイム</i>	<u>MUGSHOT</u> 0.007、0.012、0.023	<u>WEBカメラ</u> 0.025、0.040、0.063
<i>VisionLabs</i>	<u>MUGSHOT</u> 0.005、0.012、0.029	<u>WEBカメラ</u> 0.025、0.051、0.090
<i>N-Tech Lab</i>	<u>MUGSHOT</u> 0.010、0.021、0.039	<u>WEBカメラ</u> 0.032、0.059、0.094
<i>Lookman</i>	<u>MUGSHOT</u> 0.016、0.027、0.047	<u>WEBカメラ</u> 0.052、0.075、0.105
<i>Alivia</i>	<u>MUGSHOT</u> 0.012、0.027、0.062	<u>WEBカメラ</u> 0.039、0.068、0.107
<i>Neurotech.</i>	<u>MUGSHOT</u> 0.012、0.025、0.056	<u>WEBカメラ</u> 0.042、0.074、0.130
<i>東芝</i>	<u>MUGSHOT</u> 0.013、0.029、0.065	<u>WEBカメラ</u> 0.041、0.074、0.118

MUGSHOT画像とWEBカメラ画像の検索精度(考察)

閾値を設けない場合のTOP1における本人見逃し率

試験結果上位10社の「閾値を設けないTOP1の本人見逃し率」は、MUGSHOT画像による検索では0.003~0.012で、WEBカメラ画像による検索では0.008~0.039  高品質画像による検索時と較べて、品質が劣る画像による検索時には、10社の全てで「本人見逃し率」が数倍に劣化  顔識別技術を用いる場合には、検索用及び被検索用の顔画像のいずれも、ほぼ正面から捉えた緻密かつ鮮明な顔画像が望ましい。

閾値を設けて他人誤認率を1/1000とした場合の本人見逃し率

試験結果上位10社の「他人誤認率が1/1000となる閾値を設けた本人見逃し率」は、MUGSHOT画像による検索では0.004~0.065で、WEBカメラ画像による検索では0.017~0.130  閾値を設けない場合と較べて、閾値を設けた場合には、10社の全てで「本人見逃し率」が明らかに劣化  他人誤認率を極力低くする閾値を設ける場合には、検索用顔画像及び被検索用顔画像のいずれも、ほぼ正面から捉えた緻密かつ鮮明な顔画像であることが欠かせない。

IV-3

2018年FRVT 顔の経年変化に対する識別特性

被疑者写真データベースには、20年以上前に撮影された顔画像も多く含まれている。このため、防犯カメラ等に遺留された犯人の顔画像を元に被疑者写真データベースを検索する場合には、長期にわたる顔の経年変化で容貌が大きく変わった場合でも、同一人物か否かを高精度に判別できることが望まれる。

そこで、2018年のFRVTでは、高品質なMUGSHOT画像を用いて、最大で18年に及ぶ「顔の経年変化」が識別精度に及ぼす影響を調べている。

顔の経年変化に対する識別特性(試験方法)

「他人誤認率が1/1000となる閾値」の求め方

300万人分(300万枚)の顔画像データベース(全てMUGSHOT画像)を検索対象として、331,254人分(331,254枚)のMUGSHOT画像(いずれもデータベース内に同一人物の顔画像が存在しない。)で検索し、「他人誤認率が1/1000となる閾値」を求める。

「TOP1における本人見逃し率」の求め方

1人につき複数枚のMUGSHOT画像がある3,068,801人の累犯者について、最も古いMUGSHOT画像のみを3,068,801枚集めて検索対象の顔画像データベースとし、それ以外のMUGSHOT画像(2,853,221人分の10,951,064枚)で10,951,064回検索する。

そして、最も古いMUGSHOT画像と検索に用いたMUGSHOT画像との間の経過年数(例えば、【2~4年の隔たり】や【10~12年の隔たり】などをパラメータとして、「閾値を設けないTOP1の本人見逃し率」と、「他人誤認率が1/1000となる閾値を設けたTOP1の本人見逃し率」を算出する。

顔の経年変化に対する識別特性(試験方法)



検索用MUGSHOT画像



検索用MUGSHOT画像



検索用MUGSHOT画像

1人につき複数枚のMUGSHOT画像がある場合に、最も古いMUGSHOT画像のみを3,068,801枚集めたものを検索対象として、それ以外のMUGSHOT画像(2,853,221人分の10,951,064枚)で検索



最も古いMUGSHOT画像

顔の経年変化に対する識別特性(試験結果その1)

閾値を設けない場合のTOP1(類似度ランク1位)の本人見逃し率

【0~2年前】、【2~4年前】、【4~6年前】、【6~8年前】、【8~10年前】、【10~12年前】、【12~14年前】、【14~18年前】

<i>NEC</i>	0.003	0.004	0.004	0.004	0.004	0.005	0.006	0.004
<i>依図</i>	0.006	0.008	0.008	0.008	0.009	0.011	0.015	0.021
<i>マイクロソフト</i>	0.003	0.005	0.006	0.007	0.009	0.010	0.013	0.016
<i>VisionLabs</i>	0.004	0.006	0.008	0.009	0.011	0.013	0.015	0.019
<i>N-Tech Lab</i>	0.005	0.009	0.015	0.022	0.030	0.040	0.057	0.080
<i>Lookman</i>	0.012	0.014	0.016	0.017	0.017	0.018	0.021	0.026
<i>Alivia</i>	0.009	0.012	0.014	0.016	0.018	0.020	0.023	0.031
<i>Neurotech.</i>	0.009	0.012	0.014	0.015	0.017	0.019	0.023	0.031


顔の経年変化に対する識別特性(試験結果その2)

他人誤認率が1/1000となる閾値を設けたTOP1の本人見逃し率

【0~2年前】、【2~4年前】、【4~6年前】、【6~8年前】、【8~10年前】、【10~12年前】、【12~14年前】、【14~18年前】

<i>NEC</i>	0.007	0.009	0.011	0.013	0.015	0.017	0.021	0.027
<i>依図</i>	0.012	0.020	0.031	0.047	0.067	0.096	0.14	0.20
<i>マイクロソフト</i>	0.027	0.047	0.072	0.10	0.13	0.16	0.20	0.26
<i>VisionLabs</i>	0.048	0.080	0.13	0.17	0.21	0.24	0.30	0.36
<i>N-Tech Lab</i>	0.035	0.063	0.10	0.15	0.20	0.26	0.34	0.42
<i>Lookman</i>	0.043	0.069	0.098	0.13	0.16	0.19	0.23	0.28
<i>Alivia</i>	0.064	0.11	0.14	0.19	0.24	0.28	0.34	0.40
<i>Neurotech.</i>	0.065	0.10	0.14	0.18	0.23	0.27	0.33	0.40

顔の経年変化に対する識別特性(考察)

NEC社の【14～18年の隔たり】における「閾値を設けないTOP1の本人見逃し率」の数値(0.004)は、他の上位7社の数値(0.016～0.080)と較べて桁違いに優秀。同社の【14～18年の隔たり】における「他人誤認率が1/1000となる閾値を設けたTOP1の本人見逃し率」の数値(0.027)も、他社の数値(0.20～0.42)と較べて桁違いに優秀  数十年前に撮影した被疑者写真も多く含まれる被疑者写真データベースの検索・照合において、【14～18年の隔たり】における識別性能の桁違いの優秀さは大きな意義がある。



【顔の長期経年変化に対する識別精度を更に向上させるには】

「顔を識別するアルゴリズム」をディープラーニングにより生成する「学習フェーズ」で、より多くの「同一人物の昔の顔写真と今の顔写真のセット」を教材として用いて、反復学習することが効果的

IV-4

2018年FRVT 真横顔に対する識別特性

2013年の時点では、真横顔画像を正面顔画像と照合して同一人物であるか否かを見分けられる顔識別技術はどこにも存在せず、顔の正面から見て30度から45度ほどの斜め横方向から撮影した顔画像が、正面顔画像との照合により同一人物であるか否かを見分けられる限界であった。

しかし、2018年の時点では、真横顔画像を正面顔画像と照合して同一人物であるか否かを見分けられる顔識別技術が出現している。

真横顔に対する識別特性(試験方法)

「他人誤認率が1/10、1/100、1/1000となる閾値」の求め方

160万人分(160万枚)の顔画像データベース(全て正面顔のMUGSHOT画像)を検索対象として、10万人分(10万枚)の真横顔のMUGSHOT画像(いずれもデータベース内に同一人物の顔画像が存在しない。)で検索し、「他人誤認率が1/10、1/100、1/1000となる閾値」を求める。

「TOP1における本人見逃し率」の求め方

160万人分(160万枚)の顔画像データベース(全て正面顔のMUGSHOT画像)を検索対象として、10万人分(10万枚)の真横顔のMUGSHOT画像(いずれもデータベース内に撮影時点が異なる同一人物の正面顔のMUGSHOT画像が存在する。)で検索し、「閾値を設けないTOP1の本人見逃し率」と、「他人誤認率が1/10、1/100、1/1000となる閾値を設けたTOP1の本人見逃し率」を算出する。

真横顔に対する識別特性(試験方法)



160万人分(160万枚)のMUGSHOT画像(正面顔)に対して、10万人分のMUGSHOT画像(真横顔)で10万回検索



最新のMUGSHOT画像(検索用)

最新のMUGSHOT画像(真横顔)で、最新から2番目のMUGSHOT画像(正面)を検索



最新から2番目のMUGSHOT画像(被検索用)

真横顔に対する識別特性(試験結果)

他人誤認率一定(緑字1/10、赤字1/100、青字1/1000)としたTOP1の本人見逃し率
紫字は、閾値を設定しないTOP1の本人見逃し率

マイクロソフト 0.089、0.109、0.148、0.203

VisionLabs 0.130、0.198、0.322、0.461

N-Tech Lab 0.208、0.317、0.443、0.566



NEC 0.272、0.340、0.479、0.664

Tevian 0.329、0.483、0.661、0.910

Dermalog 0.517、0.642、0.856、0.948

依図 0.831、0.845、0.875、0.902

真横顔に対する識別特性(考察)

最上位4社の「閾値を設けないTOP1の本人見逃し率」の数値(0.089~0.272)は、160万人分(160万枚)の顔画像データベース(全て高品質な正面顔のMUGSHOT画像)を検索対象として、高品質な真横顔のMUGSHOT画像で検索を行った結果  被疑者写真データベースの検索・照合において大きな意義  犯人の遺留顔画像の撮影角度が真横に近かったとしても高品質(顔画像が緻密で鮮明)であれば、被疑者写真データベースを検索して得られるTOP5000(類似度の上位5000位以内)に、犯人の被疑者写真をリストアップし得ることを示している。



【 真横顔に対する識別精度を更に向上させるには 】

「顔を識別するアルゴリズム」をディープラーニングにより生成する「学習フェーズ」で、より多くの「真横顔と正面顔のセット」を教材として用いて、反復学習することが効果的

IV-5

2018年FRVT

同一人物の複数ショット照合による 識別精度の向上

被疑者写真データベースでは、累犯者には複数の被疑者写真が存在する。このため、防犯カメラの録画映像に遺留された犯人の顔画像に基づき、被疑者写真データベースを検索する際には、累犯者の被疑者写真の扱い方について2通りの方法がある。つまり、最新の被疑者写真のみを検索対象とする方法と、全ての被疑者写真を検索対象とする方法である。どちらの方法が識別精度に優れるのか、興味深いところである。

同一人物の複数ショット照合による識別精度の向上（試験方法）

「他人誤認率」が1/1000となる「閾値」の求め方

160万人分（160万枚）の顔画像データベース（全てMUGSHOT画像）を検索対象として、331,254人分（331,254枚）のMUGSHOT画像（いずれもデータベース内に同一人物の顔画像が存在しない。）で検索し、「他人誤認率が1/1000となる閾値」を求める。

複数ショットを含まない場合の「TOP1における本人見逃し率」の求め方

160万人分（160万枚）の顔画像データベース（全てMUGSHOT画像で、同一人物の複数ショットを含まない。）を検索対象として、154,549人分（154,549枚）のMUGSHOT画像（データベース内の同一人物画像とは別画像）で検索し、「他人誤認率が1/1000となる閾値を設けたTOP1の本人見逃し率」を算出する。

複数ショットを含む場合の「TOP1における本人見逃し率」の求め方

160万人分（3,351,206枚）の顔画像データベース（全てMUGSHOT画像で、同一人物の複数ショットを全て含む。）を検索対象として、154,549人分（154,549枚）のMUGSHOT画像（データベース内の同一人物画像とは別画像）で検索し、「他人誤認率が1/1000となる閾値を設けたTOP1の本人見逃し率」を算出する。

同一人物の複数ショット照合による識別精度の向上(試験方法)

160万人分(160万枚)のMUGSHOT画像(最新から2番目の画像)に対して、最新のMUGSHOT画像(154,549人分)で154,549回検索

160万人分(3,351,206枚)のMUGSHOT画像(最新から2番目以前の全ての画像)に対して、最新のMUGSHOT画像(154,549人分)で154,549回検索



同一人物の最新以外の
複数のMUGSHOT画像
(被検索用)



全てのMUGSHOT
画像を検索



最新のMUGSHOT画像
(検索用)

最新から2番目の
MUGSHOT画像のみを検索

同一人物の複数ショット照合による識別精度の向上(試験結果)

他人誤認率が1/1000となる閾値を設けたTOP1の本人見逃し率

赤字は、検索対象が160万人分(160万枚)のMUGSHOT画像の場合

青字は、検索対象が160万人分(3,351,206枚)のMUGSHOT画像の場合 ← 同一人物の複数ショット照合

<i>NEC</i>	<u>0.0044</u>	<u>0.0021</u>
<i>依図</i>	<u>0.0123</u>	<u>0.0074</u>
<i>マイクロソフト</i>	<u>0.0141</u>	<u>0.0080</u>
<i>センスタイム</i>	<u>0.0234</u>	<u>0.0165</u>
<i>VisionLabs</i>	<u>0.0289</u>	<u>0.0185</u>
<i>N-Tech Lab</i>	<u>0.0391</u>	<u>0.0301</u>
<i>Lookman</i>	<u>0.0463</u>	<u>0.0425</u>
<i>Alivia</i>	<u>0.0620</u>	<u>0.0402</u>
<i>Neurotech.</i>	<u>0.0564</u>	<u>0.0527</u>
<i>東芝</i>	<u>0.0648</u>	<u>0.0529</u>

160万人分(3,351,206枚)の内訳

1人1枚	80.1%
1人2枚	13.4%
1人3枚	3.7%
1人4枚	1.4%
1人5枚	0.6%
1人6枚	0.3%
1人7枚以上	0.2% (最大33枚)

同一人物の複数ショット照合による識別精度の向上(考察)

160万人分(3,351,206枚)のMUGSHOT画像では、160万人中の約32万人(全体の約2割)が、撮影期日の異なる複数枚のMUGSHOT画像を有するため、160万人分(160万枚)のMUGSHOT画像と較べて検索対象枚数が甚だしく増加(160万枚から約335万枚に倍増)している。しかし、検索結果については、検索対象枚数の倍増にも関わらず、いずれのベンダーについても識別精度に顕著な改善が見られる。

このことから、



顔画像データベース内に1人につき複数枚の顔画像が存在する場合には、最新の1枚のみに対して照合するよりも、最古も含めた複数枚全てに対して照合する方が、一定の閾値の下で他人誤認率を悪化させることなく、本人見逃し率を低減できる。

視点を変えれば、



ライブ映像の中から目的とする人物(ターゲット)を見つけ出す場合には、映像内の1ショットのみの顔画像でターゲットデータベースを照合するよりも、映像内の複数ショットの顔画像を用いてターゲットデータベースを照合する方が、一定の閾値の下で他人誤認率を悪化させることなく、本人見逃し率を低減できる。

V

顔識別技術における ディープラーニングの効能・効果

顔識別技術におけるディープラーニングの効能・効果

2013年の時点では、真横顔画像を正面顔画像と照合して同一人物であるか否かを見分けられる顔識別技術は皆無 → 顔の正面から見て30度から45度ほどの斜め横方向から撮影した顔画像が、正面顔画像との照合により同一人物であるか否かを見分けられる限界 ← 真横顔画像と正面顔画像との照合は、「人の目」でも困難

しかし、

2018年の時点では、識別精度が高いとは未だ言えないが、真横顔画像を正面顔画像と照合して同一人物であるか否かを見分けられる顔識別技術が出現 ← ディープラーニングの効能・効果を如実に示している。

このことから、

【 顔識別技術の精度向上は、ディープラーニングの適用による工夫次第 】

真横顔画像と正面顔画像との照合時の識別精度の向上は、「顔を識別するアルゴリズム」をディープラーニングにより生成する「学習フェーズ」において、より多くの「真横顔と正面顔のセット」を教材として用いて、効果的かつ効率的な反復学習を行うことに尽きる。

顔識別技術における人種バイアスの問題

米国では、犯罪捜査に顔識別技術を用いた際の人種バイアス、つまり、白人の顔に比べてアフリカ系やアジア系の有色人種の顔に対する他人誤認率が高くなることが、人種的偏見の助長に繋がりがねないとして、今日の大きな社会問題化

ところで、

顔識別技術に用いられるディープラーニングは、人の頭脳内部での神経回路網の仕組みと働きを、コンピュータ上で数学的に模したもの → 人種バイアスの原因や対策を検討するには、「人の目」における人種バイアスについて検討することが有益

「人の目」では、

人は、相手の顔を見て美醜などの特徴を瞬時に見分けている。この見分ける基となっているものは、誕生以来、見たり接したりしてきた無数の顔から生み出された「頭の中の平均顔」と考えられる。見たり接したりしてきた顔の大半が日本人であれば、日本人の平均顔が頭の中に生み出されるであろう。日本人には、白人の顔、あるいは黒人の顔が、日本人の顔ほどには見分けが付けづらいが、日本人の平均顔を基に顔の特徴を見分けているとすれば、白人の顔、あるいは黒人の顔が、日本人には皆同じように見えてしまうのも無理はない。

【警察政策第17巻(2015)「顔画像識別における人の目の特性と機械の目の特性」からの引用】

ここから類推すれば、

次のページへ

前のページから

ここから類推すれば、

大半を日本人の顔で学習した顔識別技術では、白人や黒人の顔に対する識別精度が日本人の顔ほどには上がらない。 ➡ 米国で犯罪捜査に顔識別技術を用いた場合における人種バイアスの問題は、白人の顔に比べてアフリカ系やアジア系の有色人種の顔に対する学習不足が最大の原因

解決策は、

「顔を識別するアルゴリズム」をディープラーニングで生成する「学習フェーズ」での学習用教材として、アフリカ系やアジア系の有色人種の顔画像を更に多く用いて、白人の顔画像と同等の識別精度が達成できるまで、効果的かつ効率的な反復学習を行うこと

つまり、

ディープラーニングを活用した顔識別技術における識別精度の向上には、反復学習の充実強化が最も重要なファクター ← 真横顔画像と正面顔画像との照合時の識別精度向上方策と同じ

2021年7月27日

ディープラーニングで
飛躍的に進化した顔画像識別技術

終

澤田雅之技術士事務所(電気電子部門)所長
元警察大学校警察情報通信研究センター所長

澤田 雅之