

AI 特許紹介(10)
～FaceNet 特許～

2020 年 1 月 8 日
河野特許事務所
所長 弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許権者 Google

出願日 2015 年 12 月 17 日

登録日 2017 年 12 月 5 日

登録番号 US9836641

発明の名称 画像の数値埋め込み生成

641 特許は、同一人の 2 つの画像と、別人の画像との 3 つの画像のトリプレット (Triplet)を用いて、顔認識ニューラルネットワークの精度を向上させるアイデアである。

2.特許内容の説明

図 1 に示すように、数値埋め込みシステム 100 は、ニューラルネットワーク 120 及び出力ベクトル 122 を含む。入力画像 102 は、ニューラルネットワーク 120 に入力され、出力ベクトル 122 として出力される。この出力ベクトル 122 が、数値埋め込み 124 となる。

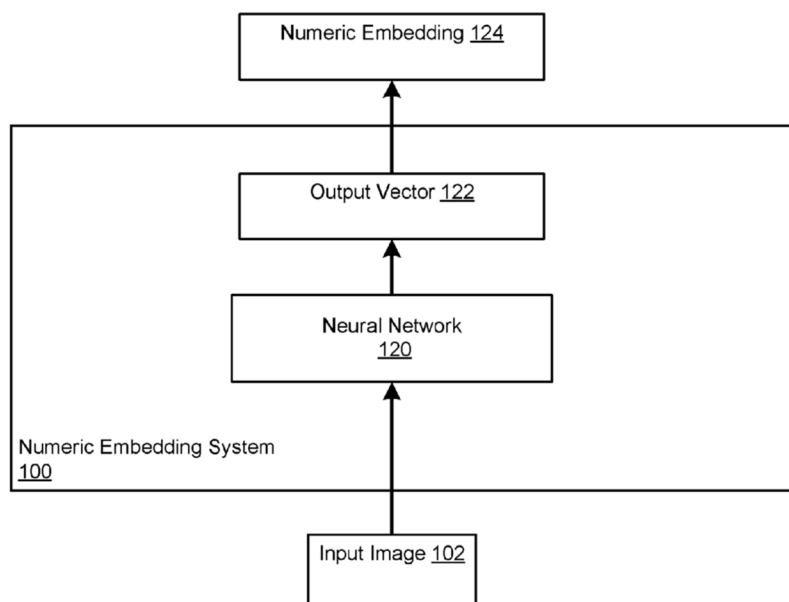
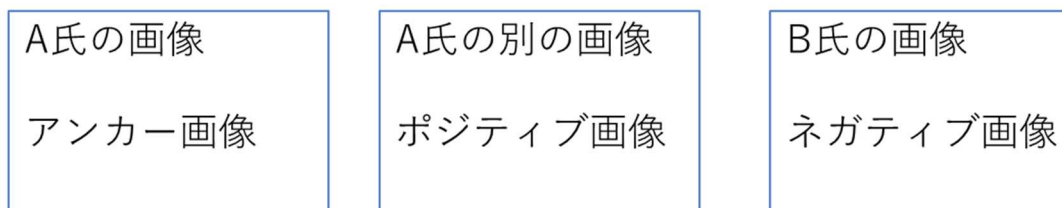
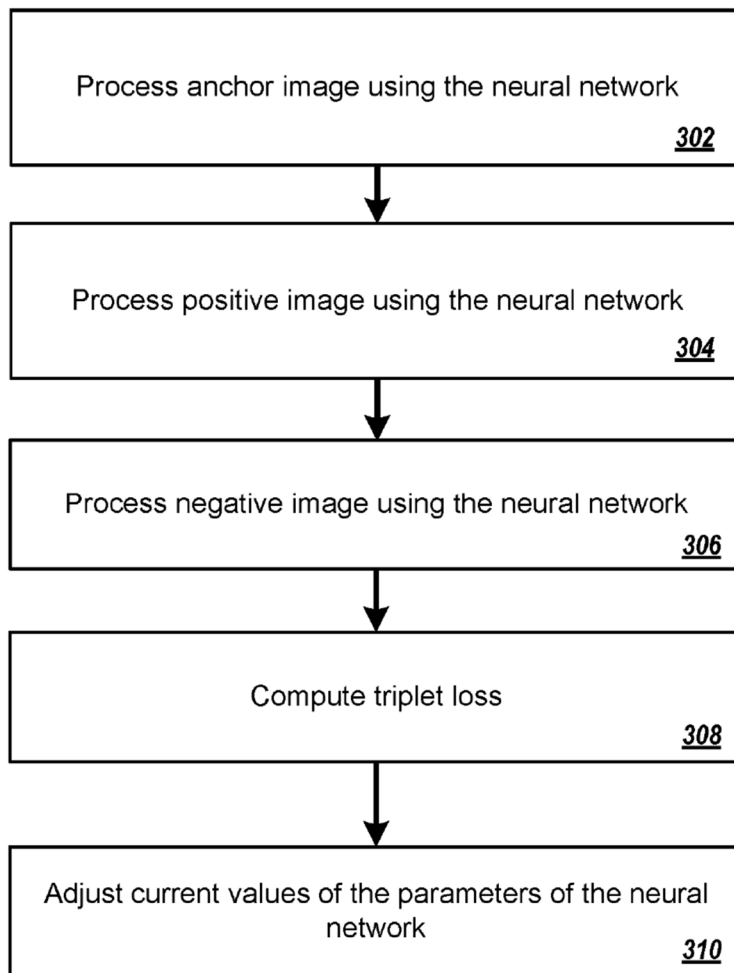


FIG. 1

ここで、アンカー画像、ポジティブ画像、及びネガティブ画像の3つを用意する。



A 氏の画像と B 氏の画像とが類似している場合、本来別人であるのに同一人と誤認識してしまうこととなる。641 特許はこのトリプレットの数値埋め込みを算出し、アンカー画像とポジティブ画像との数値埋め込み距離が、アンカー画像とネガティブ画像との数値距離よりも小さくなるよう、ニューラルネットワークのパラメータを学習させる。具体的には以下のフロー処理が行われる。



システムは、ニューラルネットワークのパラメータの現在の値に従ってニューラルネットワークを使用してトリプレット内のアンカー画像を処理し、アンカー画像の数値埋め込みを生成する（ステップ 302）。

システムは、ニューラルネットワークのパラメータの現在の値に従ってニューラルネットワークを使用してトリプレット内のポジティブ画像を処理し、ポジティブ画像の数値埋め込みを生成する（ステップ 304）。

システムは、ニューラルネットワークのパラメータの現在の値に従ってニューラルネットワークを使用してトリプレット内のネガティブ画像を処理し、ネガティブ画像の数値埋め込みを生成する（ステップ 306）。

システムは、アンカー画像、ポジティブ画像、ネガティブ画像の数値埋め込みからトリプレット損失を計算する（ステップ 308）。一般に、トリプレット損失は、アンカー

画像の数値埋め込みとポジティブ画像の数値埋め込みの間の距離と、アンカー画像の数値埋め込みとネガティブ画像の数値埋め込みの間の距離に依存する。トリプレット損失 L は以下の式で表すことができる。

$$L = \max(0, \|f(x^a) - f(x^p)\|_2^2 - \|f(x^a) - f(x^n)\|_2^2 + \alpha),$$

ここで $f(x^a)$ はトリプレット内のアンカー画像の数値埋め込み、 $f(x^p)$ はトリプレット内のポジティブ画像の数値埋め込み、 $f(x^n)$ は、トリプレット内のネガティブ画像の数値埋め込みであり、 α は 0 以上の所定の値である。トリプレット損失 L は、特定のオブジェクトの画像に埋め込みがある場合に最小化されるように表される。

システムは、トリプレット損失 L を使用して、ニューラルネットワークのパラメータの現在値を調整する（ステップ 310）。つまり、システムは、ニューラルネットワークのパラメータの現在の値を調整して、トリプレット損失 L を最小限に抑える。システムは、従来のニューラルネットワークトレーニング手法、たとえばバックプロパゲーションによる確率的勾配降下を使用して、ニューラルネットワークのパラメータの現在の値を調整できる。

3. クレーム

641 特許のクレーム 1 は以下の通りである。

1. 方法において、

特定のオブジェクトタイプのオブジェクトの画像として分類されている複数のトレーニング画像を取得し、

トレーニング画像の複数のトリプレットを生成し、トリプレットのそれぞれは、それぞれのアンカー画像、それぞれのポジティブ画像、およびそれぞれのネガティブ画像を含み、各トリプレットについて、アンカー画像とポジティブ画像は共に特定のオブジェクトタイプの同じオブジェクトの画像として分類されており、ネガティブイメージは、特定のオブジェクトタイプの異なるオブジェクトの画像として分類されており、

ニューラルネットワークの複数のパラメータのトレーニング値を決定するために、各トリプレットでニューラルネットワークをトレーニングし、ニューラルネットワークは、特定のオブジェクトタイプのオブジェクトの入力画像を受け取り、入力画像の数値埋め込みを生成すべく入力画像を処理するよう構成されており、

ここで、ニューラルネットワークのトレーニングは、トリプレットのそれぞれについて以下を含む：

アンカー画像の数値埋め込みを生成するために、ニューラルネットワークのパラメータの現在値に従ってニューラルネットワークを使用してトリプレット内のアンカー画像を処理し、

ポジティブ画像の数値埋め込みを生成するために、ニューラルネットワークのパラメータの現在値に従ってニューラルネットワークを使用してトリプレットのポジティブ画像を処理し、

ネガティブ画像の数値埋め込みを生成するために、ニューラルネットワークのパラメータの現在値に従ってニューラルネットワークを使用してトリプレットのネガティブ画像を処理し、

アンカー画像、ポジティブ画像、及びネガティブ画像の数値埋め込みからトリプレット損失を計算し、

オブジェクトの入力画像の数値埋め込みと同じオブジェクトの他の画像のそれぞれの数値埋め込みとの間の距離が、オブジェクトの入力画像の数値埋め込みと他のオブジェクトの他の画像の数値埋め込みとの間の距離よりも小さくなるように、トリプレット損失を最小化すべく、ニューラルネットワークのパラメータの現在値を調整し、

トレーニング画像の複数のトリプレットを生成することは、トリプレットのそれぞれについて、以下のようにトリプレットを生成することを含む、

(i) トリプレットのポジティブ画像の数値埋め込みは、ポジティブ画像とアンカー画像と同じオブジェクトの他の画像の数値埋め込みよりも、トリプレットのアンカー画像の数値埋め込みから遠く離れ、

(ii) トリプレット内のネガティブ画像の数値埋め込みは、ポジティブイメージとアンカーイメージとは異なるオブジェクトの他のイメージの数値埋め込みよりもアンカーイメージの数値埋め込みに近くなる。

4. FaceNet に関する論文

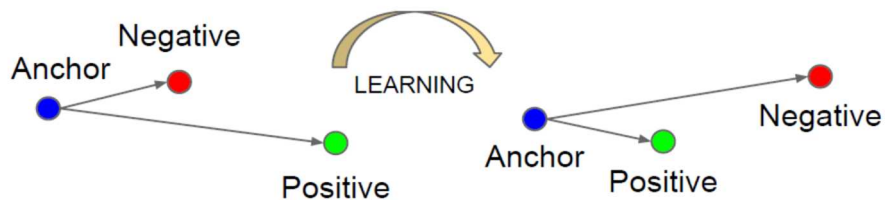
トリプレットを用いた FaceNet は、Florian Schroff らの論文¹で公表されている。



上記図はネットワーク構成を示す。ネットワークはトリプレット損失を出力すべくバ

¹ Florian Schroff, Dmitry Kalenichenko, James Philbin “FaceNet: A Unified Embedding for Face Recognition and Clustering”

ッチ入力層、ディープ CNN 及び L2 正規化層により構成される。



上図に示すように、トリプレット損失は、同じアイデンティティを持つアンカーとポジティブ間の距離を最小化し、アンカーと異なるアイデンティティのネガティブ間の距離を最大化する。

YouTube の FaceDB を用いた実験では、 $95.12\% \pm 0.39$ の分類精度が得られた。下記図は、複数の写真から同一人の写真をグループ化したものである。障害物、ライティング、ポーズ、年齢が異なっても適切に分類されている。



以上

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI研究所 AI コース修了。

[AI 特許コンサルティング](#)の他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門 2.0](#)」がある。