AI 特許紹介(25)

AI 特許を学ぶ！究める！
～ニューラルスタイル変換～

2021 年 2 月 10 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

特許出願人　Eberhard Karls Universitaet Tuebingen
出願日　2018 年 1 月 26 日
公開日　2018 年 6 月 7 日
公開番号 US2018/0158224
発明の名称　画像合成の方法と装置

224 特許は、オリジナル画像にスタイル画像を組み合わせて変換するニューラルスタイル変換技術に関する。

2.特許内容の説明

Fig. 1A

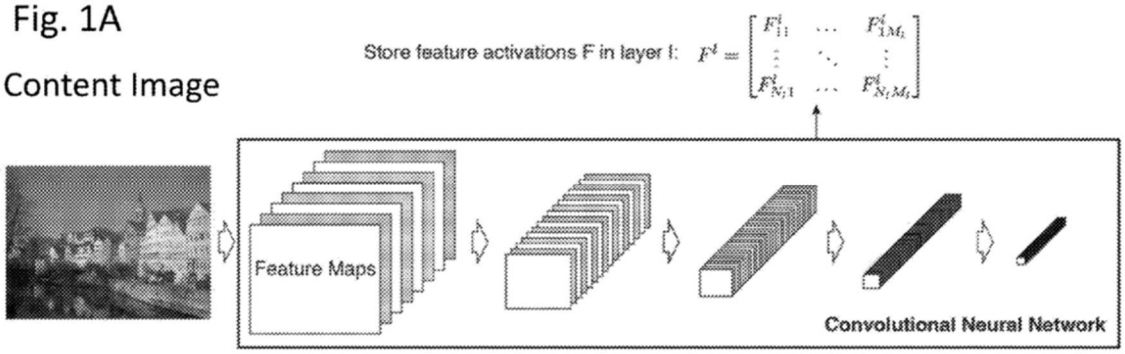


図１Ａは、コンテンツ特徴の抽出方法の概要を示す。コンテンツ画像及びソース画像の特徴は畳み込みニューラルネットワーク(CNN)により抽出される。16 個の畳み込み層と 5 個のプーリング層を有する VGG19 ネットワークによって提供される特徴空間が使用される。

N1 個の異なるフィルタを持つレイヤーには、サイズ M1 の N1 個の特徴マップがあり、ここで M1 は、特徴マップの高さと幅の積である。レイヤー1 の応答は、行列 F1 の要素 R に格納できる。ここで、Fij は、レイヤー1 の位置 j での i 番目のフィルタのアクティベーションである。

Fig. 1B

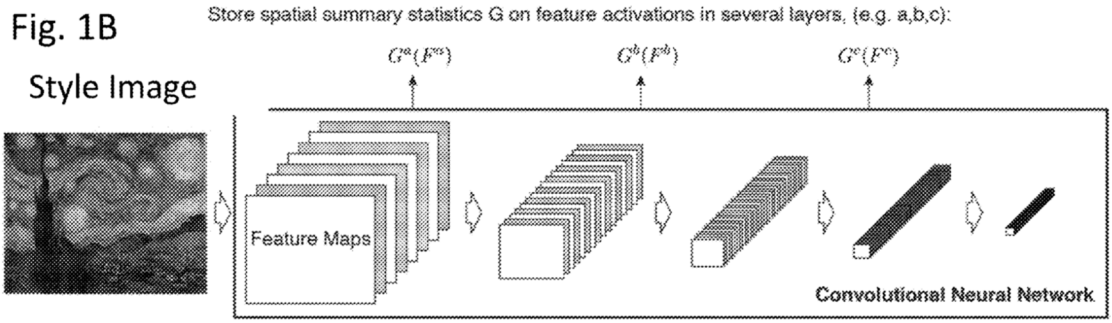


図 1B は、異なるフィルタ応答間の相関を計算することによって、ネットワークのすべての層における CNN の応答に関して、スタイル表現がどのように構築されるかを概略的に示している。ここで、期待値は、入力画像の空間範囲にわたって取られる。この特徴相関は、この場合、グラム行列 G によって与えられる。ここで、G は、レイヤー1 のベクトル化された特徴マップ i と j の間の内積である。

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}$$

複数のレイヤーの特徴相関を追加することにより、ソース画像の静止マルチスケール表現が取得される。これは、画像のテクスチャ情報をキャプチャするが、グローバルな

配置はキャプチャしない。

ネットワークのレイヤーから 2 つの特徴空間が形成され、特定のソース画像のコンテンツとスタイルに関する情報が保持される。

まず、ニューラルネットワークの上位層のユニットをアクティブ化すると、詳細なピクセル情報をキャプチャせずに、主にソース画像のコンテンツをキャプチャする。次に、ネットワーク内のいくつかのレイヤーの異なるフィルタ応答間の相関関係により、特定のソース画像のスタイル情報がキャプチャされる。このスタイルまたはテクスチャ表現は、ソース画像のグローバル構成を無視するが、色とローカル画像構造の観点から全体的な外観を保持する。

これにより、互いに分離された画像のコンテンツおよびスタイルを表現することが可能となる。また、コンテンツとスタイルを個別に操作することもでき、特に、写真の内容とさまざまな芸術作品の外観を組み合わせた新しい画像を生成できる。

Fig. 2

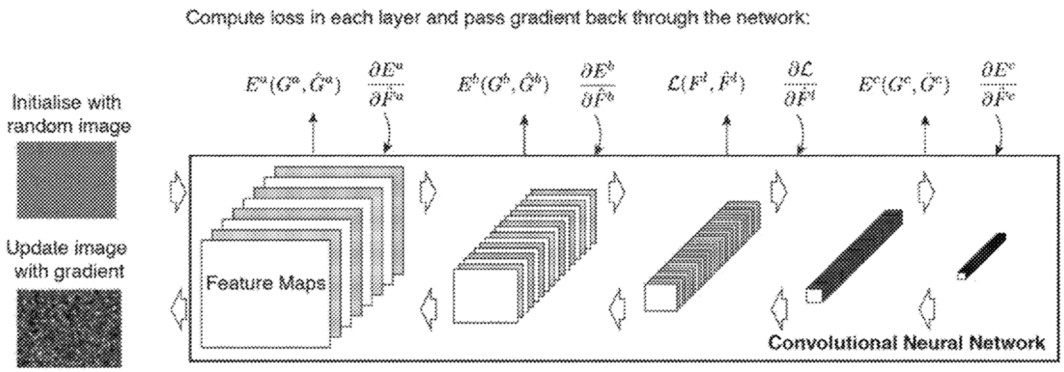


図 2 は、本発明の実施形態による画像を生成するための方法の概要を示す。

写真などのターゲット画像のコンテンツと、ペイントされた画像などのソース画像のスタイルを混合した画像を生成するために、適切な開始画像で初期化された画像検索が実行される。たとえば、初期画像としては、ホワイトノイズに応じて輝度値が分散されたランダム画像である。

これにより、ネットワークの層内のターゲット画像のコンテンツ表現およびニューラルネットワークのいくつかの層内のソース画像のスタイル表現からのコンテンツおよび初期画像のスタイル表現の距離が一緒に最小化される。

元の画像とターゲットまたはソース画像のコンテンツとスタイル特性の間のそれぞれの距離は、適切な損失関数 $L_{content}$ 及び L_{style} によって表すことができる。

写真が p で、アートワークが a の場合、最小化される全損失関数は以下のとおりとなる。

$$L_{total}(\vec{p},\vec{a},\vec{x})=\alpha L_{content}(\vec{p},\vec{x})+\beta L_{style}(\vec{a},\vec{x})$$

ここで、α と β はそれぞれ重み係数である。重み係数は、コントローラを介して、連続的に調整可能である。

スタイルをより強調すると、ターゲット画像、つまり写真の本質的な内容を表示せずに、アートワークの外観に対応する画像が得られる。コンテンツを重視することで、写真をより明確に識別できるが、スタイルはソース画像のスタイルに対応している。

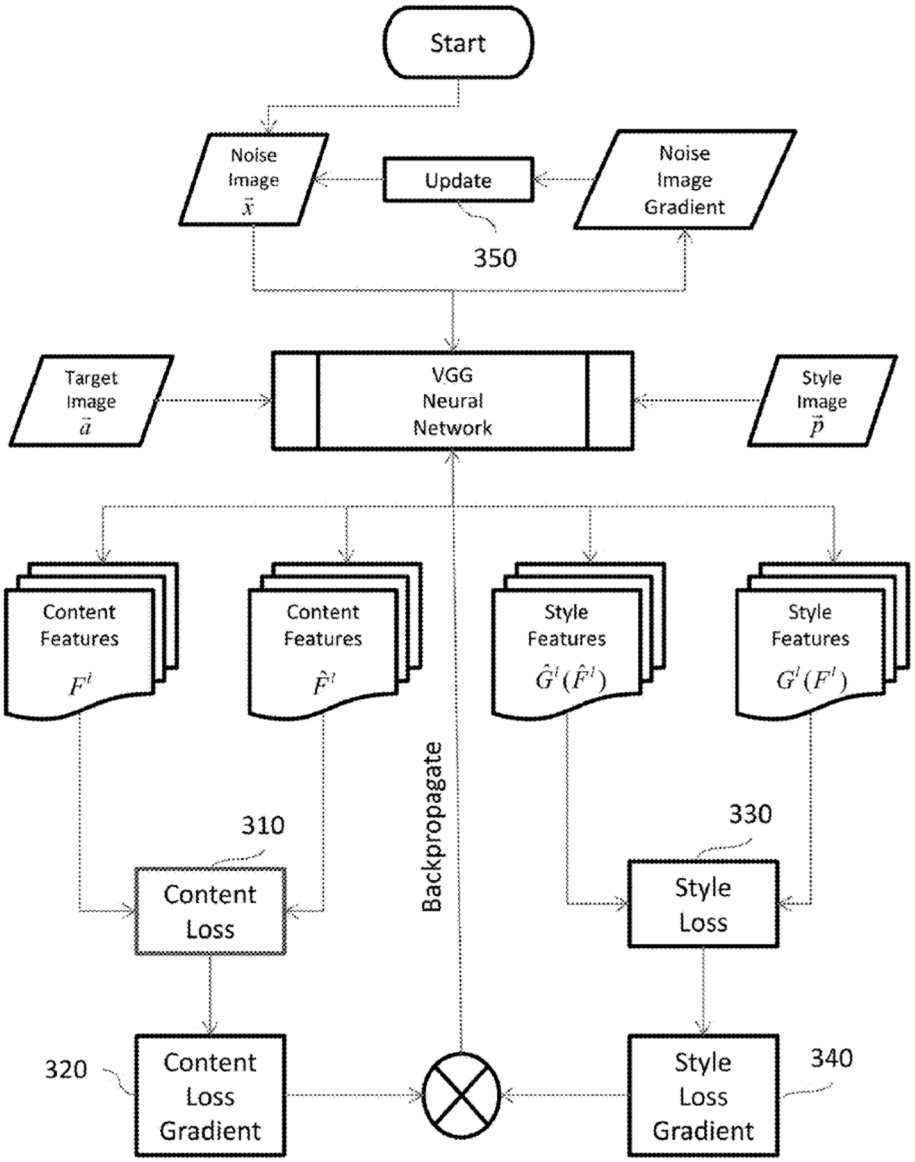


図３は、抽出特徴に基づいて画像を合成するための方法の概略図を示す。

ホワイトノイズに応じて輝度値が分布しているランダム画像をニューラルネットワークの入力として使用し、L、a、b、c 層の特徴アクティベーション F ^を取得する。次に、要約統計量 Ĝ が層 a、b、および c について計算される。次のステップでは、損失関数 L が層 L、a、b、および c について計算される。層 l のターゲット画像の損失は次の形式となる。

$$L_{content}(\hat{F}^l, F^l) = \frac{1}{2}\sum_{i,j}\left(\hat{F}^l_{ij} - F^l_{ij}\right)^2.$$

層 a、b、c でのソース画像の損失は次の形式となる。

$$E^a(\hat{G}^a, G^a) = \frac{1}{4N_a^2 M_a^2} \sum_{i,j} \left(\hat{G}_{ij}^a - G_{ij}^a\right)^2.$$

その後、損失の勾配は、この層の特徴アクティベーション F^に関して各層で計算される。層 1 のターゲット画像の勾配は次の形式となる。

$$\frac{\partial L_{content}}{\partial F_{ij}^l} = \begin{cases} (\hat{F}^l - F^l)_{ij} & \text{if } \hat{F}_{ij}^l > 0 \\ 0 & \text{if } \hat{F}_{ij}^l < 0 \end{cases}.$$

層 a、b、c のソース画像の勾配は次の形式となる。

$$\frac{\partial E_a}{\partial \hat{F}_{ij}^a} = \begin{cases} \frac{1}{N_a^2 M_a^2} ((\hat{F}^a)^T (\hat{G}^a - G^a))_{ji} & \text{if } \hat{F}_{ij}^a > 0 \\ 0 & \text{if } \hat{F}_{ij}^a < 0 \end{cases}$$

次に、勾配はネットワークを介してエラーバックプロパゲーションによって伝播され、ホワイトノイズ画像に関する勾配が計算される。その後、ホワイトノイズ画像を調整して、レイヤー1、a、b の損失を最小限に抑える。このプロセスは、損失が適切な終了基準を満たすまで、調整された画像で続行される。あるいは、この方法は、ソース画像またはターゲット画像を初期画像として使用することができる。

3.クレーム

224 特許のクレーム 1 は以下の通りである。

1.少なくとも 1 つのソース画像からターゲット画像にスタイル特徴を転写するためのコンピュータ実装の方法において、

(A) ソース画像とターゲット画像に基づいて結果画像を生成し、

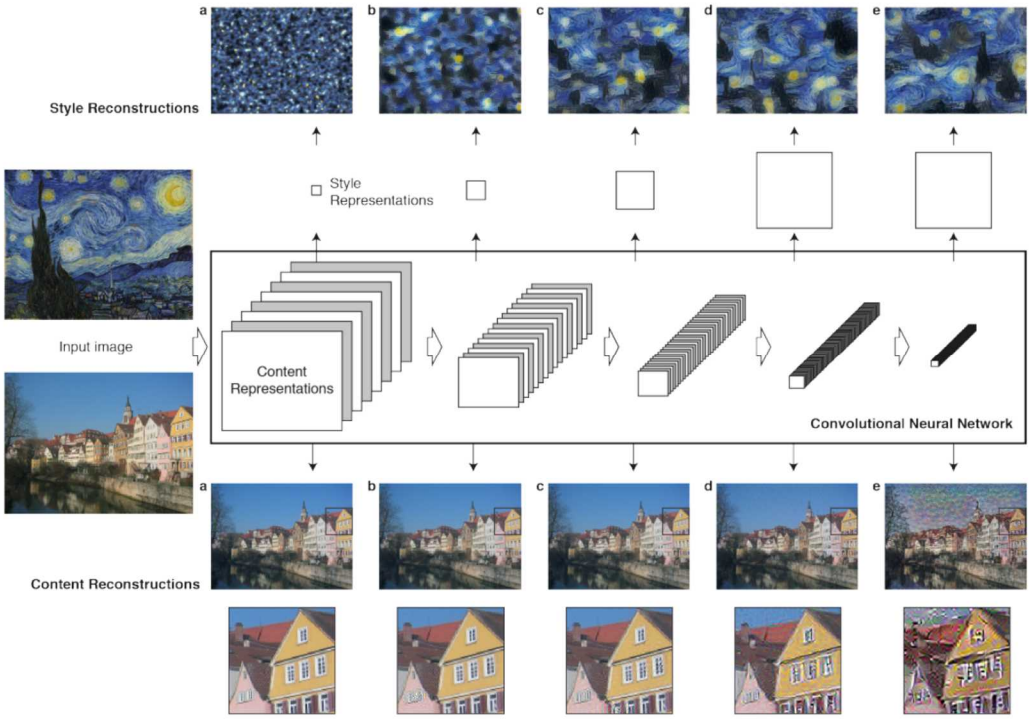
ここで、結果画像の 1 つまたは複数の空間的変化特徴は、ターゲット画像の 1 つまたは複数の空間的変化特徴に対応し、結果画像のテクスチャは、ソース画像のテクスチャに対応し、

(B) 結果画像を出力し、

テクスチャは、ソース画像の空間的変化特徴の要約統計量に対応する。

4. ニューラルスタイル変換に関する論文

本特許に関連する論文[1]が Leon A. Gatys,氏らにより発表されている。



上記図は、CNN、コンテンツ画像及びスタイル画像を示す説明図である。ニューラルスタイル変換はコンテンツ損失、スタイル損失、及び全変動損失を最小化するものである。

コンテンツ損失に関しては、ベース画像と同じコンテンツを含むよう最小化される。層 l のコンテンツ画像の損失は次の形式となる。

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} \left(F_{ij}^{l} - P_{ij}^{l}\right)^2$$

スタイル損失に関しては、全般的にスタイル画像と同じ画風となるよう最小化される。トータルのスタイル損失はグラム行列を用いて、以下の式で表すことができる。

$$\mathcal{L}_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l$$

[1] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge “A Neural Algorithm of Artistic Style” arXiv:1508.06576v2 [cs.CV] 2 Sep 2015

全変動損失は全体的に滑らかになるよう最小化される。最終的に最小化すべき損失関数は以下で表すことができる。なおαはコンテンツの重み、ベータはスタイルの重みである。

$$\mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x})$$

誤差逆伝播法を用いて損失関数を最小化するよう画像の画素値が更新される。

上記図は写真画像に左下枠にある著名な絵画を組み合わせたものである。

上記図は、ワシリーカンディンスキーによるコンポジション VII の絵画のスタイルの詳細な結果を示す。

行方向は、CNN レイヤーを増加させて（Conv1-Conv5）サブセットのスタイル表現を一致させた結果を示している。ネットワークの上位層からのスタイル機能を含めると、スタイル表現によってキャプチャされたローカル画像構造のサイズと複雑さが増すこととなる。

一方、列方向は、コンテンツとスタイルとの再構築間の様々な相対的な重みを示している。各列の上の数字は、写真の内容とアートワークのスタイルを一致させることに重点を置く比率 α/β を示す。スタイルの重み β を増減させることによりスタイル画像の影響を変化させることができる。

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、医療 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「世界のソフトウエア特許(共著)」、「FinTech 特許入門」、「AI/IoT 特許入門 2.0」、「ブロックチェーン 3.0(共著)」がある。