

A Neural Algorithm Of Artistic Style

- Leon A. Gatys
- Alexander S. Ecker
- Matthias Bethge

芸術、特に絵画では、人間は画像のコンテンツとスタイルの複雑な相互作用を構成することで、独特の視覚体験を作り出す技能を習得しました。これまでのところ、このプロセスのアルゴリズム的な基礎は不明であり、同様の能力を有する人工システムは存在していない。しかし、物体や顔の認識のような視覚的知覚の他の重要な領域では、人間に近いパフォーマンスが近年、ディープニューラルネットワーク[1,2]と呼ばれる、生物学的に触発された視覚モデルによって実証されている。ここでは、高知覚品質の芸術的画像を生成するディープニューラルネットワークに基づく人工システムを紹介します。このシステムは、神経表現を使用して、任意の画像のコンテンツとスタイルを分離して再結合し、芸術的な画像を作成するための神経アルゴリズムを提供します。さらに、性能最適化された人工ニューラルネットワークと生物学的視覚[3-7]の間の類似性を考慮して、我々の研究は、人間が芸術的な画像を作り、知覚する方法のアルゴリズム的理解への道を進めている。

Introduction

画像処理タスクにおいて最も強力なディープニューラルネットワークのクラスは、畳み込みニューラルネットワークと呼ばれる。畳み込みニューラルネットワークは、フィードフォワード形式で階層的に視覚情報を処理する小さな計算ユニットの層で構成されています（図1）。ユニットの各層は、画像フィルタの集合として理解することができ、それぞれの画像フィルタは、入力画像から特定の特徴を抽出する。このように、与えられた層の出力は、いわゆる特徴マップ、すなわち入力画像の異なるフィルタリングされたバージョンからなる。

畳み込みニューラルネットワークが物体認識向けに訓練されると、物体情報が処理階層に沿ってますます明示されるように画像の表現を発達させる[8]。したがって、ネットワークの処理階層に沿って、入力画像は詳細なピクセル値と比較した画像の実際のコンテンツを考慮するような表現に変換されます。我々は、層の特徴マップだけから画像を再構成することによって、各層が含む入力画像に関する情報を直接視覚化することができます（図1、コンテンツの再構成、画像の再構築方法の詳細はMethodsを参照してください）。ネットワークの上位層は、入力画像における物体とその配置に関する上位レベルのコンテンツを捕捉しますが、再構成の正確なピクセル値は制約しません [9]（図1、コンテンツの再構成d、e）。対照的に、下位層からの再構成は、原画像の正確な画素値を単純に再現する（図1、コンテンツの再構成a、b、c）。従って、ネットワークの上位層における特徴応答をコンテンツ表現と呼ぶ。

入力画像のスタイル表現を得るために、本来はテクスチャ情報を捕捉するために設計された特徴空間を使用する[8]。この特徴空間は、ネットワークの各層におけるフィルタ応答の上に構築される。これは、特徴マップの空間範囲にわたる異なるフィルタ応答間の相関から構成されます（詳細はMethodsを参照）。複数の層の特徴相関を含めることにより、入力画像の静止したマルチスケール表現を取得し、それはそのテクスチャ情報を捕捉するが、全体的な配置は捕捉しない。

ここでもまた、入力画像のスタイル表現に一致する画像を構築することによって、ネットワークの異なる層上に構築されたこれらのスタイル特徴空間によって捕捉された情報を視覚化することができます（図1、スタイルの再構成）[10,11]。スタイル特徴からの実際の再構成では、色と局所化された構造に関する一般的な外観を捕捉する、入力画像のテクスチャ化されたバージョンを生成します。さらに、入力画像からの局所画像構造の大きさと複雑さは、階層に沿って増加し、その結果は受容野の大きさおよび特徴の複雑さが増すことによって説明することができる。このマルチスケール表現をスタイル表現と呼びます。

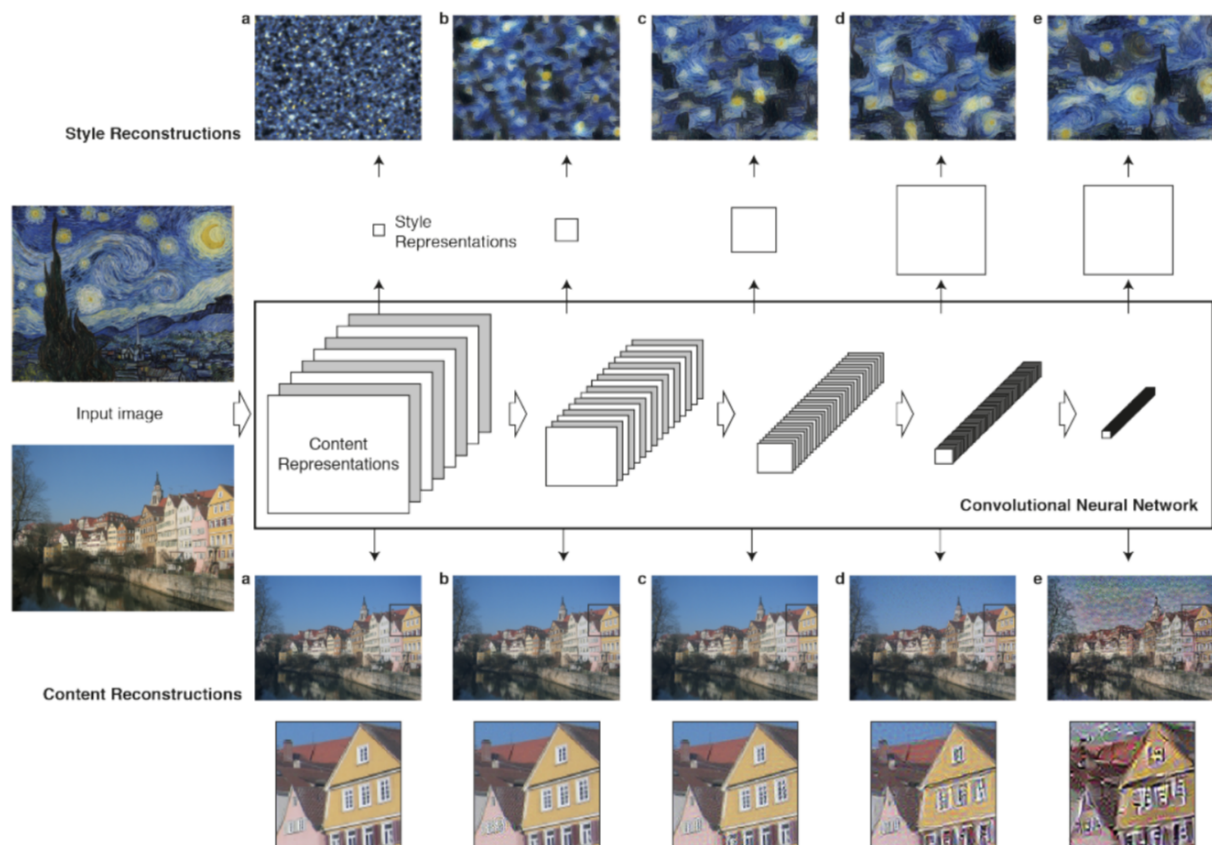


図1：畳み込みニューラルネットワーク（CNN）。与えられた入力画像は、CNNの各処理段階でフィルタリングされた画像の集合として表される。異なるフィルタの数は処理階層に沿って増加するが、フィルタリングされた画像のサイズは、いくつかのダウンサンプリング機構（例えば、最大プーリング）によって低減され、ネットワークの層当たりの総ユニット数が減少する。**コンテンツの再構成**。入力画像を特定の層内のネットワーク応答のみを知ることから再構成することで、CNNの異なる処理段階で情報を視覚化する。オリジナルのVGGネットワーク

の'conv1_1' (a) 、'conv2_1' (b) 、'conv3_1' (c) 、'conv4_1' (d) 、'conv5_1' (e) 層から入力画像を再構成する。我々は、より下位の層からの再構築がほぼ完全であることを見出した

(a,b,c) 。ネットワークのより上位の層では、詳細なピクセル情報が失われるが、画像の上位レベルのコンテンツは保存されます (d,e) 。**スタイルの再構築**。オリジナルのCNN表現の上に、我々は入力画像のスタイルを捕捉する新しい特徴空間を構築した。スタイル表現は、CNNの異なる層の異なる特徴間の相関を計算します。入力画像のスタイルを、CNNの層の異なる部分集合

('conv1_1' (a) 、'conv1_1'と'conv2_1' (b) 、'conv1_1'と'conv2_1'と'conv3_1' (c) 、'conv1_1'と'conv2_1'と'conv3_1'と'conv4_1' (d) 、'conv1_1'と'conv2_1'と'conv3_1'と'conv4_1'と'conv5_1' (e)) の上に構築されたスタイル表現から再構成する。これは、スケールが増加する一方で、画像の全体的な配置情報を破棄しながら、与えられた画像のスタイルに合致する画像を生成する。

この論文の重要な発見は、畳み込みニューラルネットワークにおけるコンテンツとスタイルの表現が分離可能であることです。つまり、我々は両方の表現を独立して操作することで、新しい知覚的に意味のある画像を生成することができます。この発見を実証するために、我々は2つの異なるソース画像からのコンテンツ表現とスタイル表現を混在させた画像を生成する。特に、ドイツのTubingenの「Neckarfront」を描写した写真のコンテンツ表現と、さまざまな時代のいくつかの有名な作品のスタイル表現を一致させる（図2）。

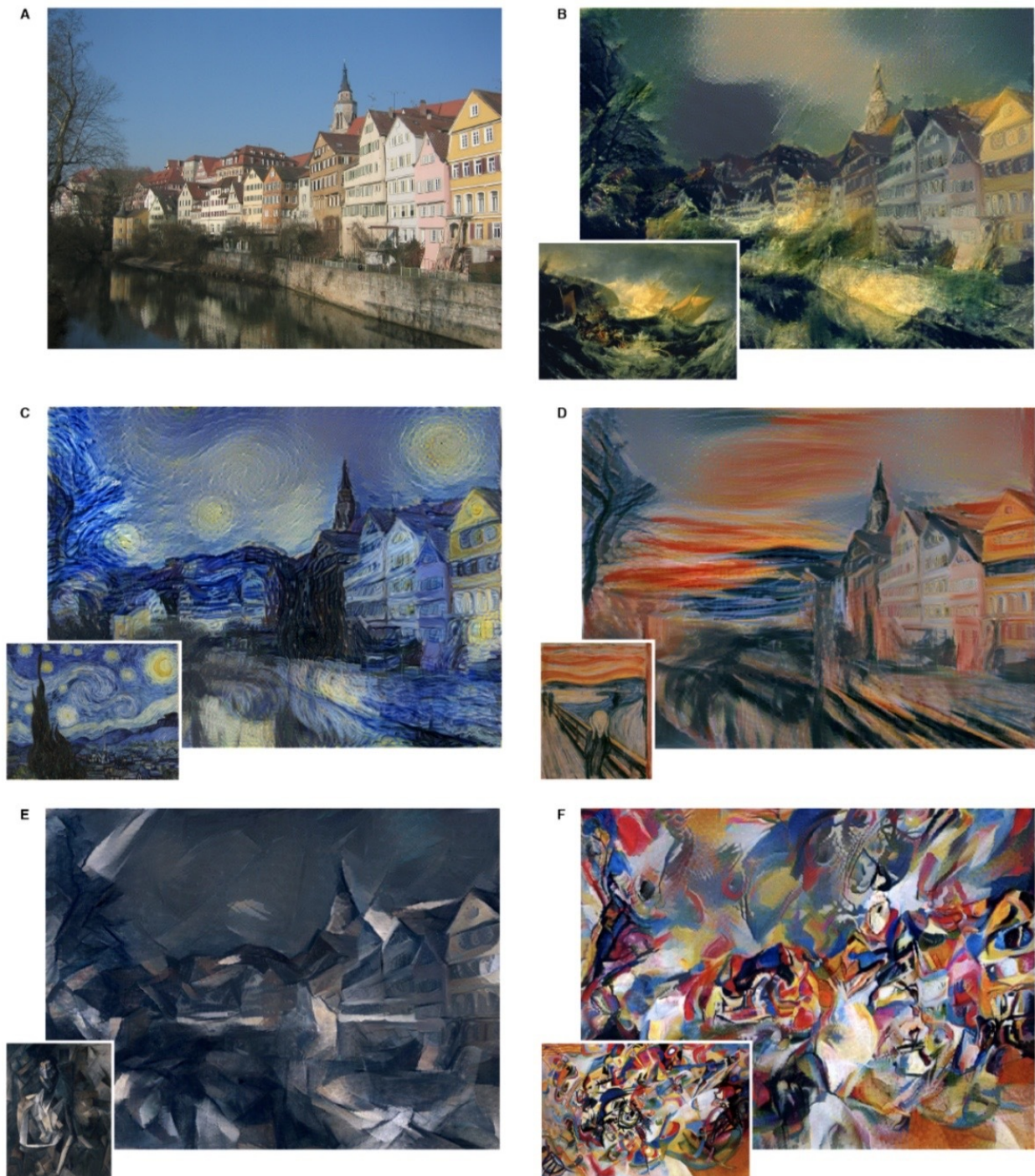


図2：写真のコンテンツといくつかの有名な作品のスタイルを組み合わせた画像。これらの画像は、写真のコンテンツ表現と作品のスタイル表現を同時に一致させる画像を見つけることによって生成される（Methodsを参照）。AはドイツのTubingenのNeckarfrontを描写したオリジナル写真（写真：Andreas Praefcke）。生成されたそれぞれの画像にスタイルを提供した絵画は各パネルの左下に表示されている。Bは1805年のJ.M.W.Turnerによる*The Shipwreck of the Minotaur*。Cは1889年のVincent van Goghによる*The Starry Night*。Dは1893年のEdvard Munchによる*Der Schrei*。Eは1910年のPablo Picassoによる*Femme nue assise*。Fは1913年のWassily Kandinskyによる

Composition VII。

画像は、写真のコンテンツ表現とそれぞれの芸術作品のスタイル表現を同時に一致させる画像を見つけることによって合成される（詳細は Methods を参照）。元の写真の全体的な配置は保存されていますが、全体の風景を構成する色や局所的な構造は芸術作品によって提供されます。効果的には、芸術作品のスタイルで写真をレンダリングするので、合成された画像の外観は写真と同じコンテンツを表示しても、芸術作品に似ます。

上で概説したように、スタイル表現は、ニューラルネットワークの複数の層を含むマルチスケール表現です。図2に示した画像では、スタイル表現にはネットワーク全体の階層からの層が含まれていました。スタイルは、より少ない数の下位層のみを含めることによって、より局所的に定義することもでき、異なる視覚体験をもたらす（図3、行に沿って）。スタイル表現をネットワークのより上位の層まで一致させると、局所画像構造がますます大きくなるスケールに一致し、よりスムーズで連続的な視覚体験が実現します。したがって、視覚的に最も魅力的な画像は、通常、スタイル表現をネットワークの最上位の層まで一致させることによって作成されます（図3、最後の行）。

もちろん、画像のコンテンツとスタイルを完全に分離することはできません。ある画像のコンテンツと別の画像のスタイルを組み合わせた画像を合成する場合、通常は両方の制約に完全に一致する画像は存在しません。しかし、画像合成中に最小化する損失関数には、コンテンツとスタイルの2つの項がそれぞれ含まれており、それらは十分に分離されています（Methodsを参照）。したがって、コンテンツやスタイルの再構築（図3、列に沿って）のいずれかに関して、スムーズに重点を調整することができます。スタイルを強く強調すると、芸術作品の外観に合った画像になり、効果的には、テクスチャ化された画像が得られますが、写真のコンテンツはほとんど表示されません（図3、第1列）。コンテンツに重点を置くと、写真をはっきりと識別できますが、絵画のスタイルにはあまり一致していません（図3、最後の列）。ソース画像の特定のペアについて、視覚的に魅力的なイメージを作成するために、コンテンツとスタイルとの間のトレードオフを調整することができます。



図3：Wassily Kandinskyによる絵画*Composition VII*のスタイルに対する詳細な結果。行は、CNNの層の増加する部分集合のスタイル表現に一致した結果を示している（Methodsを参照）。我々は、ネットワークのより上位の層からのスタイル特徴を含めると、スタイル表現によって捕捉された局所画像構造のサイズと複雑さが増すことを発見しました。これは、ネットワークの処理階層に沿って、受容野のサイズと特徴の複雑さが増すことによって説明できます。列は、コンテンツとスタイルの再構成の間の異なる相対的な比重を示している。各列の上の数字は、写真のコンテンツと芸術作品のスタイルの強調の比率を示します（Methodsを参照）。

ここでは、画像のコンテンツとスタイルの分離を実現する人工ニューラルシステムを紹介し、他の画像のスタイルで1つの画像のコンテンツを再構築することを可能にします。我々は、いくつかの有名な絵画のスタイルと任意に選んだ写真のコンテンツを組み合わせた新しい芸術的な画像を作成することによって、これを実証します。特に、我々は、物体認識用に訓練された高性能ディープニューラルネットワークの特徴応答から、画像のコンテンツとスタイルの神経表現を導出する。私たちの知る限りでは、これは自然画像全体で、スタイルとコンテンツを分離する画像特徴の最初の実証です。スタイルとコンテンツの分離に関する以前までの作業は、異なる手書き文字や顔の画像や異なるポーズのフィギュアの画像など、はるかに少ない複雑さの感覚的な入力で評価されました [12,13]。

私たちの実証では、与えられた写真を様々な有名な芸術作品のスタイルでレンダリングしています。この問題は、通常、非写実的レンダリングと呼ばれるコンピュータビジョンの分岐からアプローチされる（最近のレビュー[14]参照）。概念的に最も密接に関連するのは、芸術的スタイルの変換を達成するためにテクスチャ変換を使用する方法である[15-19]。しかし、これらの以前のアプローチは、画像のピクセル表現を直接操作するために、主にノンパラメトリック技術に依存している。対照的に、物体認識用に訓練されたディープニューラルネットワークを用いることにより、画像の上位レベルのコンテンツを明示的に表す特徴空間における操作を実行する。

物体認識用に訓練されたディープニューラルネットワークから得られた特徴は、以前は、作成された時代によって芸術作品を分類するためのスタイル認識用に使用されていました[20]。そこで、分類器はコンテンツ表現と呼ばれる生のネットワーク活性上で訓練される。我々は、スタイル表現のような固定された特徴空間への変換が、スタイル分類におけるさらに優れた性能を達成すると推測している。

一般的に、異なるソースからコンテンツとスタイルを混ぜ合わせた画像を合成する我々の方法は、アート、スタイル、およびコンテンツに依存しない画像の外観の認識と神経表現を研究する新しい魅力的なツールを提供します。我々は、2つの独立した、知覚的に意味のある変動源、すなわち画像の外観とコンテンツを導入する新たな動機を設計することができる。これは、機能的イメージングに対する精神物理学から電気生理学的神経記録までの視覚的知覚を考慮している広範な実験的研究に有用であると考えられる。実際に、私たちの研究は、ニューラル表現が画像のコンテンツとスタイルを独立して取得する方法のアルゴリズム的理解を提供しています。重要なことに、私たちのスタイル表現の数学的形式は、単一のニューロンレベルまで下げた画像の外観表現についての明確かつ試験可能な仮説を生成します。スタイル表現は、ネットワーク内の異なるタイプのニューロン間の相関を単に計算するだけである。ニューロン間の相関を抽出することは、例えば、1次視覚系（V1）のいわゆる複合細胞によって実現される、生物学的に妥当な計算である[21]。この結果は、腹側皮質視覚路に沿った、異なる処理段階での計算に似た複合細胞を成すことは、視覚的入力の外観のコンテンツに依存しない表現を得るための可能な方法である。

結局のところ、生物学的視覚の中心的な計算タスクの1つを実行するように訓練された神経システムが、画像のコンテンツとスタイルを分離することを可能にする画像表現を自動的に学習することは本当に魅力的です。その解釈は、物体認識の学習をするときに、物体の同一性を保持するすべての画像の変化に対して不変でなければならないということである。画像のコンテンツと外観の変化を要素化する表現は、このタスクにとって極めて実用的であろう。したがって、スタイルとコンテンツを抽象化する能力、さらに、芸術を作成して楽しむ能力は、主に、視覚システムの強力な推論能力の屈指の特徴です。

Methods

主なテキストに示された結果は、共通のビジュアルオブジェクト認識ベンチマークタスク[23]で人間のパフォーマンスに匹敵する畳み込みニューラルネットワークであるVGG-Network[22]に基づいて生成され、導入され、広範囲に記述されている[22]。19層のVGGネットワークの16個の畳み込み層および5つのプール層によって達成される。我々は、完全に接続された層のどれも使用していない。モデルは公に利用可能であり、caffe-frameworkで探索することができる[24]。画像合成のために、平均プールによってmax-pooling操作を置き換えると、より魅力的な結果が得られました。そのため、表示される画像は平均的なプールで生成されました。

一般に、ネットワーク内の各レイヤは、非線形フィルタバンクを定義し、ネットワーク内のレイヤの位置に折り目を付けます。したがって、与えられた入力画像 x が符号化される。その画像へのフィルタ応答によってCNNの各レイヤ内に存在する。N1個の別個のフィルタを有する層 サイズM1の各N1個の特徴マップを有する。ここで、M1は、特徴マップの幅の高さの倍数である。したがって、レイヤー l の応答は、行列 $F \in \mathbb{R}^{N1 \times M1}$ に格納することができます。ここで、 F_{ij}^l は、アクティブ化 ij でレイヤー l の位置 j にある i 番目のフィルタの階層の異なるレイヤー（図1、コンテンツ再構成）で符号化された画像情報を視覚化するために、ホワイトノイズ画像の勾配降下を実行して元の画像の特徴応答に一致する別の画像を見つける。したがって、 $p \sim x$ を原画像とし、生成された画像と P と F をレイヤ l のそれぞれの特徴表現とする。次に、2つの特徴表現の間の2乗誤差損失を定義する。

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

層 l の活性化に関するこの損失の導関数は、

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

に等しい。そこから画像 $\sim x$ に対する勾配を標準誤差逆伝播を用いて計算することができる。したがって、最初のランダム画像 $\sim x$ を、元の画像 $p \sim$ と同じCNNのある層で同じ応答を生成するまで変更することができる。図1の5つのコンテンツ再構成は、（1）、（2）、（3）、（4）のconv1 1元のVGGネットワーク。

ネットワークの各レイヤーにおけるCNN応答の上に、異なるフィルタ応答間の相関を計算するスタイル表現を作成しました。ここでは、入力画像の空間的な広がり予測をしています。これらの特徴の相関は、グラム行列 $G \in \mathbb{R}^{N1 \times N1}$ によって与えられる。ここで、 G_{ij}^l は、層 l ：

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

内のベクトル化された特徴マップ i と j との間の内積である。与えられた画像のスタイル（図1、スタイルの再構成）に一致するテクスチャを生成するために、ホワイトノイズ画像からの勾配降下を使用して、元の画像のスタイル表現に一致する別の画像を見つける。これは、原画像からのグラム行列のエントリと生成される画像のグラム行列との間の平均二乗距離を最小にすることによって行われる。したがって、 $a \sim x$ を元のイメージとし、生成されるイメージとレイヤー l のそれぞれのスタイル表現を A と G とします。この層の総損失に対する寄与は

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

であり、総損失は

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

であり、ここで、 w_l は全損失に対する各層の寄与の重み因子である（以下の結果の w_l の具体的な値を参照）。層 l の活性化に関する E_l の導関数は、解析的に計算することができる。

$$\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_i^l M_i^l} ((F^l)^T (G^l - A^l))_{ji} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases}$$

ネットワークの下位層における活性化に対するEIの勾配は、標準誤差逆伝搬を用いて容易に計算することができる。図1の5つのスタイル再構成は、'conv1 1' (a)、'conv1 1'および'conv2 1' (b)、'conv1 1'、'conv2 1'および'conv3 1' 'conv1 1'、'conv2 1'、'conv3 1'、'conv4 1'、'conv5'、'conv1'、'conv1'、'conv1' 1' (e)。

写真のコンテンツと絵のスタイル（図2）を混在させた画像を生成するには、ネットワークの1つのレイヤー内の写真のコンテンツ表現からのホワイトノイズ画像の距離と、CNNのいくつかの層に絵を描く。だから、p〜を写真、アートワークにする。最小化する損失関数は

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

です。ここで α はコンテンツとスタイル再構成の重み係数です。図2の画像では、'conv4 2'のコンテンツ表現と'conv1 1'、'conv2 1'、'conv3 1'、'conv4 1'、'conv5 1' (w_l = それらの層では1/5、他のすべての層では $w_l = 0$)。比率は 1×10^{-3} (図2B、C、D) または 1×10^{-4} (図2E、F) のいずれかであった。図3は、コンテンツおよびスタイル再構成損失の異なる相対的重み付け (列に沿った) およびレイヤー 'conv1 1' (A)、'conv1 1'および'conv2 1' (B) のみのスタイル表現のマッチングの結果を示す。「conv1 1」、「conv2 1」、「conv3 1」、「conv4 1」 (D)、「conv1 1」、「conv2 1」、「conv1 1」、「conv2 1」、「conv3 1」、「conv4 1」、「conv5 1」 (E) である。係数 w_1 は、ゼロでない損失重み w_1 を有する活性層の数で割った1に常に等しい。

Acknowledgements

この研究は、ドイツ国立学術財団 (L.A.G.)、バーンスタイン計算神経科学センター (FKZ 01GQ1002)、および中央科学研究所 (EXC307) (M.B.、A.S.E.、L.A.G.)

References and Notes

1. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, 1097–1105 (2012). URL <http://papers.nips.cc/paper/4824-imagenet>.
2. Taigman, Y., Yang, M., Ranzato, M. & Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 1701–1708 (IEEE, 2014). URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6909616.
3. GüçlÜ, U. & Gerven, M. A. J. v. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. The Journal of Neuroscience 35, 10005–10014 (2015). URL <http://www.jneurosci.org/content/35/27/10005>.
4. Yamins, D. L. K. et al. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences 201403112 (2014). URL <http://www.pnas.org/content/early/2014/05/08/1403112111>.
5. Cadieu, C. F. et al. Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLoS Comput Biol 10, e1003963 (2014). URL <http://dx.doi.org/10.1371/journal.pcbi.1003963>.

6. Ku'immerer, M., Theis, L. & Bethge, M. Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet. In ICLR Workshop (2015). URL </media/publications/1411.1045v4.pdf>.
7. Khaligh-Razavi, S.-M. & Kriegeskorte, N. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Comput Biol 10, e1003915 (2014). URL <http://dx.doi.org/10.1371/journal.pcbi.1003915>.
8. Gatys, L. A., Ecker, A. S. & Bethge, M. Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. arXiv:1505.07376 [cs, q-bio] (2015). URL <http://arxiv.org/abs/1505.07376>. ArXiv: 1505.07376.
9. Mahendran, A. & Vedaldi, A. Understanding Deep Image Representations by Inverting Them. arXiv:1412.0035 [cs] (2014). URL <http://arxiv.org/abs/1412.0035>. ArXiv: 1412.0035.
10. Heeger, D. J. & Bergen, J. R. Pyramid-based Texture Analysis/Synthesis. In Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '95, 229–238 (ACM, New York, NY, USA, 1995). URL <http://doi.acm.org/10.1145/218380.218446>.
11. Portilla, J. & Simoncelli, E. P. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. International Journal of Computer Vision 40, 49–70 (2000). URL <http://link.springer.com/article/10.1023/A%3A1026553619983>.
12. Tenenbaum, J. B. & Freeman, W. T. Separating style and content with bilinear models. Neural computation 12, 1247–1283 (2000). URL <http://www.mitpressjournals.org/doi/abs/10.1162/089976600300015349>.
13. Elgammal, A. & Lee, C.-S. Separating style and content on a nonlinear manifold. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, vol. 1, 1–478 (IEEE, 2004). URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1315070.
14. Kyprianidis, J. E., Collomosse, J., Wang, T. & Isenberg, T. State of the "Art": A Taxonomy of Artistic Stylization Techniques for Images and Video. Visualization and Computer Graphics, IEEE Transactions on 19, 866–885 (2013). URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6243138.
15. Hertzmann, A., Jacobs, C. E., Oliver, N., Curless, B. & Salesin, D. H. Image analogies. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 327–340 (ACM, 2001). URL <http://dl.acm.org/citation.cfm?id=383295>.
16. Ashikhmin, N. Fast texture transfer. IEEE Computer Graphics and Applications 23, 38–43 (2003).
17. Efros, A. A. & Freeman, W. T. Image quilting for texture synthesis and transfer. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 341–346 (ACM, 2001). URL <http://dl.acm.org/citation.cfm?id=383296>.
18. Lee, H., Seo, S., Ryoo, S. & Yoon, K. Directional Texture Transfer. In Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering, NPAR '10, 43–48 (ACM, New York, NY, USA, 2010). URL <http://doi.acm.org/10.1145/1809939.1809945>.

19. Xie, X., Tian, F. & Seah, H. S. Feature Guided Texture Synthesis (FGTS) for Artistic StyleTransfer. In Proceedings of the 2Nd International Conference on Digital Interactive Media in Entertainment and Arts, DIMEA '07, 44–49 (ACM, New York, NY, USA, 2007). URL <http://doi.acm.org/10.1145/1306813.1306830>.
20. Karayev, S. et al. Recognizing image style. arXiv preprint arXiv:1311.3715 (2013). URL <http://arxiv.org/abs/1311.3715>.
21. Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. JOSA A 2, 284–299 (1985). URL <http://www.opticsinfobase.org/josaa/fulltext.cfm?uri=josaa-2-2-284>.
22. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs] (2014). URL <http://arxiv.org/abs/1409.1556>. ArXiv: 1409.1556.
23. Russakovsky, O. et al. ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575 [cs] (2014). URL <http://arxiv.org/abs/1409.0575>. ArXiv:1409.0575.
24. Jia, Y. et al. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the ACM International Conference on Multimedia, 675–678 (ACM, 2014). URL <http://dl.acm.org/citation.cfm?id=2654889>.