

大域的整合性を考慮したカラー調整と画像補完による 自然な画像合成

秋本直郁^{†1} 林昌希^{†2} 秋月秀一^{†1} 青木義満^{†1}

概要: 本論文では、背景画像に対しオブジェクト画像を貼り付けるという合成を行う際に、簡単な操作のみで自然な合成が実現できる手法を提案する。CNN で得られる大域的な特徴を用いて、カラー調整と画像補完を同時に行う。提案するネットワーク構造と学習手順で自然な貼り付け合成ができた。

キーワード: 画像合成, ディープラーニング, GANs

Natural Image Synthesis by Color Adjustment and Image Completion Considering Global Consistency

NAOFUMI AKIMOTO^{†1} HAYASHI MASAKI^{†2}
SHUICHI AKIZUKI^{†1} YOSHIMITSU AOKI^{†1}

Abstract: In this paper, we propose a method that can realize natural synthesis only by simple operation when pasting object images to background images. We perform color adjustment and image complement at the same time using the global features obtained with CNN. Natural paste synthesis was possible with the proposed network structure and learning procedure.

Keywords: Image Generation, Deep Learning, GANs



図1 提案手法による出力結果
Figure1 Outputs of the proposed method.

1. はじめに

カメラ付きスマートフォンの普及や Instagram のような写真投稿型 SNS の普及に伴い、人々が写真を編集・加工する機会は増加している。しかしながら、デザイナーが持つようなプロフェッショナルなスキルがなければ、画像編集を自然に仕上げることは困難である。したがって、画像編

集スキルを持たない人のために、容易に高品質で自然な画像編集を実現するための補助の技術が必要になる。

本研究では、画像編集の中でも特に二枚の画像の貼り付け合成を取り上げ、背景となる画像に除去したいオブジェクトがあれば白く塗りつぶし、さらに、挿入したいオブジェクト画像を上から貼り付けるという簡単な操作のみユーザーが行い、①白塗り領域のオブジェクト除去、②挿入オブジェクトの色合いを背景に馴染むように調整、の二つの処理を Deep Neural Networks (DNN)によって実行することで自然な貼り付け合成を可能とする手法を提案する。このタ

^{†1} 慶応義塾大学
Keio University.
^{†2} Recreation Lab

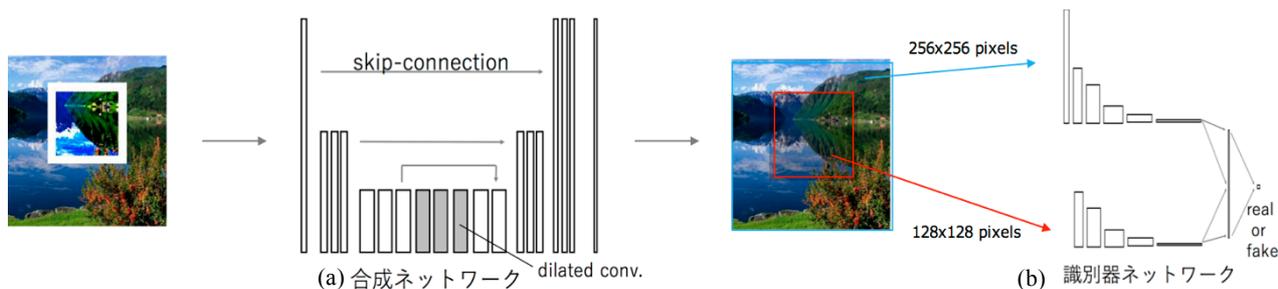


図2 ネットワーク構造
 Figure2 Network architecture.

スクを実現するために利用できる従来手法には、二画像の輝度勾配や CNN を利用する手法がある。まず、二画像の輝度勾配を用いる手法には、Poisson Image Editing [1]があり、合成したい二枚の画像の輝度勾配が滑らかに繋がるようにカラー調整を行うことで自然な合成を実現する。しかしながら、輝度勾配を用いる手法では、局所的な領域のみの情報からカラー調整を行うので貼り付ける領域に隣り合う領域の色になりやすい問題点がある。次に、CNN を利用する手法には、飯塚らによる Globally and Locally Consistent Image Completion [2]があり、CNN で抽出される大域的な特徴から欠損領域を補完する。さらに、補完した出力結果の画像全体の整合性と、補完領域を中心とした局所的な整合性を Adversarial Loss [3]を用いて評価することで、リアルスティックな画像補完ができることを示した。この手法は不要なオブジェクトを除去する手段として利用できる。しかしながら、飯塚ら[2]で提案された手法では補完領域に出現するオブジェクトを人が直接指定することはできず、CNN によって出現するオブジェクトが自動的に決定される。以上の課題をまとめると、①局所的な特徴のみを基にカラー調整が行われること、②画像補完の際に出現されるオブジェクトが指定できないこと、の二点の課題がある。

そこで本手法では、挿入するオブジェクト画像を明示的に指定し補完領域に出現させられる構造にした CNN を利用し画像補完を行うことで課題②を解決する。さらに、その際の畳み込みによって抽出される大域的な特徴をカラー調整にも利用することで課題①を解決する。このようにして、これらの課題を一つのネットワークで同時に解決する。

図2に示すような CNN を主とした encoder-decoder 型ネットワークを用いることで、畳み込みによって抽出される大域的な特徴を利用し、画像を再構成する。さらに、encoder-decoder 型ネットワークに skip-connection [4]を持たせることで、挿入するオブジェクトの特徴を後方の層に伝える。そのため、はじめの skip-connection により、入力されたオブジェクト画像は原型を留めたまま出力されることになるので、入力画像のエッジを大まかに保存することができる。その結果、入力画像のエッジを大まかに残しつつ画像を再構成することによって、挿入したいオブジェクト

を出力しつつ、カラー調整と補完を行うことができる。ネットワークの学習時には、生成された画像が自然であるかの判断のために、大域的な領域と局所的な領域から Adversarial Loss を算出し、学習に利用する。以上の方法を用いて、一つのネットワークで、挿入する画像が背景に馴染むためのカラー調整の実行と同時に、挿入する画像が補完領域に出現するよう直接指定できる画像補完の実行を実現する。本論文の貢献を以下にまとめる。

- ・入力画像のエッジを大まかに保存することで挿入したい画像の色空間に主な変化を加えると同時に、補完すべき領域に対しては補完を実行するための、新たなネットワーク構造を提案する。
- ・カラー調整と画像補完を同時に実行する能力を与えるためのネットワークの学習を、end-to-end で行うために工夫された学習手順を示す。
- ・本手法で大域的な特徴を基にして、二画像間の自然な貼り付け合成を達成する。

2. 提案手法

本手法では、図2に示すように合成ネットワークと識別器ネットワークの二つのネットワークが存在する。これらのネットワークは、GAN [3]の枠組みで交互に学習される。

2.1 合成ネットワーク

合成ネットワークは飯塚ら[2]の Completion Network の構成を基にしており、encoder-decoder 構造の Fully-Convolutional Network (FCN) [5]である。FCN は全ての層が畳み込み層であり、その畳み込みの繰り返しで後方へ伝えられる情報は画像の広い領域から得られた情報となる。さらに、合成するオブジェクトのエッジ情報を正確に出力側に伝えるために図2のように skip-connection 構造をもたせる。また、背景画像と合成ネットワークからの出力画像を、前景領域を表すマスクを用いて加算合成することで合成画像とする。

2.2 識別器ネットワーク

識別器ネットワークは飯塚ら[2]によって提案されたネットワークを用いている。これは2つの入力ストリームか



図3 学習時の処理手順

Figure3 Training procedure.

ら成り，一方は，合成ネットワークから出力された 256×256 pixels の画像を入力とし大域的な整合性を判断するストリームである．もう一方は，補完領域を中心として 128×128 pixels で切り出した画像を入力とし局所的な整合性を判断するストリームである．これらのストリームからの二つの出力を結合し，最終的に一つの Adversarial Loss を算出する．

2.3 学習

本手法の学習手順を図3に示す．一枚の画像から，ランダムに矩形領域を切り出し，前景画像と背景画像とする．前景画像のみに対して，輝度値を変化させ，さらに補完すべき領域としてランダムな太さの枠で白抜き領域を作成する．合成ネットワークに対して，操作した前景画像と条件画像としての背景画像を入力して，操作される前の前景画像に近くなるように再構成損失 (Reconstruction Loss)関数を定義する．出力された前景画像と背景画像を合わせることで合成画像とし，識別器ネットワークで自然であるかどうかの損失を算出し，再構成損失関数としての L2 Loss と Adversarial Loss の二つの損失関数の線形和を用いて，敵対的学習の手続きで二つのネットワークの学習を行う．一般に，これら二つの損失関数を使うことで画像補完タスクにおける敵対的学習が比較的安定することが知られている [6]．

3. 実験

訓練データとして，places2 [7]内にある 95,000 枚の画像を利用した．このデータセットには，様々な風景画像が含まれている．以下では，貼り付け合成の実験結果を示す．

3.1 カラー調整と補完の同時実行

(1) 結果

実験の結果例を図1に示す．図1の上段は，矩形で白抜き領域を作成し，その上から木の画像を挿入する例である．この実験では，周囲に合わせて木のカラー調整を行うことと，白抜きされた領域を補完することの二つのタスクを行う．また，図1の下段は同様のタスク設定を人物画像の挿

入で行った場合である．

(2) 考察

図1の上段のように，白抜き領域は周囲に合わせて自然に補完されていることがわかる．また，木の葉の色の変化のように輝度値が周囲の葉と似るよう調整されていることがわかる．これらのことから，一つのネットワークで，カラー調整と画像補完が同時に実行されたことが言える．

図1の下段で顕著に見て取れるが，人物と補完領域の境界が滲んだような結果となっている．これを改善するために，将来的には境界に対する何らかのペナルティ項を目的関数に追加するべきだと考えている．

3.2 ユーザーに必要な操作

本手法を利用することによって，ユーザーが簡単な操作のみ行えば入力画像を作成でき，自然な貼り付け合成が行えることを示す (図4)．

元画像に対して，挿入したい画像を上から貼り付ける操作を行う (図4の(b))．次に，上から貼り付けを行なったにもかかわらず，隠しきれなかった不要なオブジェクトがあれば白く塗りつぶす操作を行う (図4の(c))．以上の手順により，ユーザーが行う入力への準備は終わり，背景を黒塗り画像に差し替えることで，ネットワークに入力する「操作した前景画像」の状態になる (図4の(d))．挿入画像に被らないように白塗りを行ったり，背景を黒塗り画像に差し替えたりする操作は，市販のペインティングツールに一般的に搭載されているレイヤー機能を使うことで簡単に実現できる．このように，画像の貼り付けと領域の選択という簡単な操作のみをユーザーが行えば良く，指によるスマートフォンの操作でも行える操作であるので，スマートフォンのアプリケーションに搭載することも可能であると考えられる．

3.3 失敗例

(1) 結果

失敗例を図5に示す．図5の左列(a)は白色の椅子を挿入し，さらにその周囲に補完領域を作った例である．また，右列(b)は人物の両手を補完領域とした場合の例である．

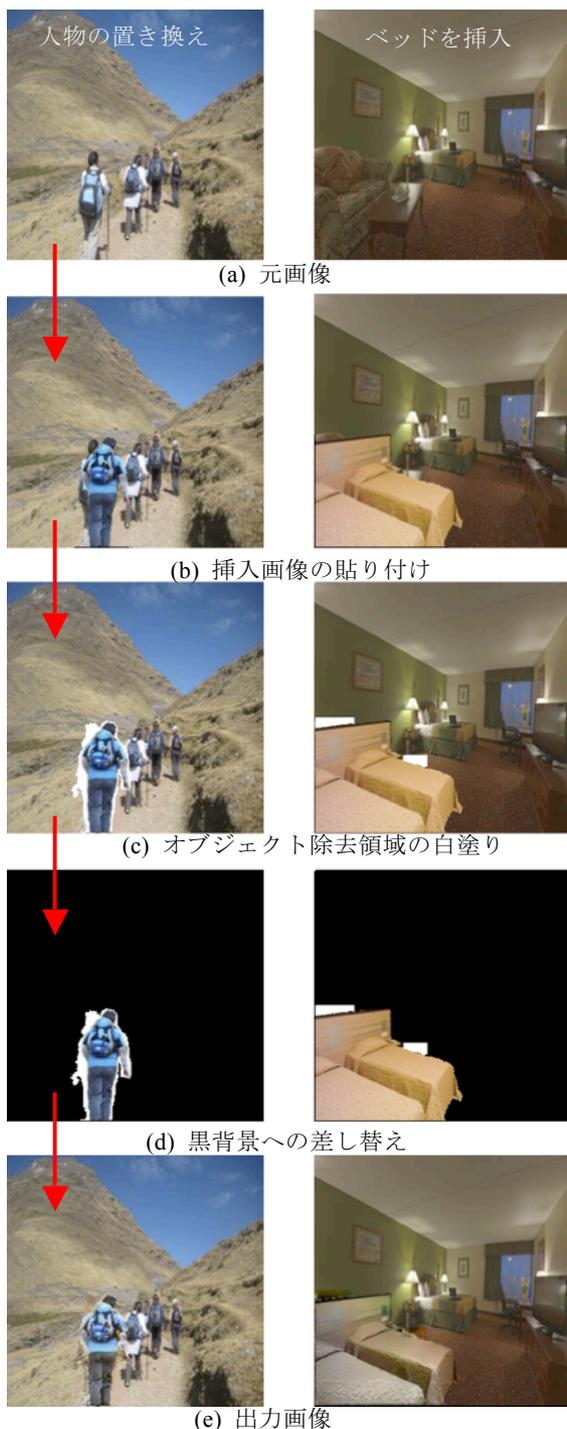
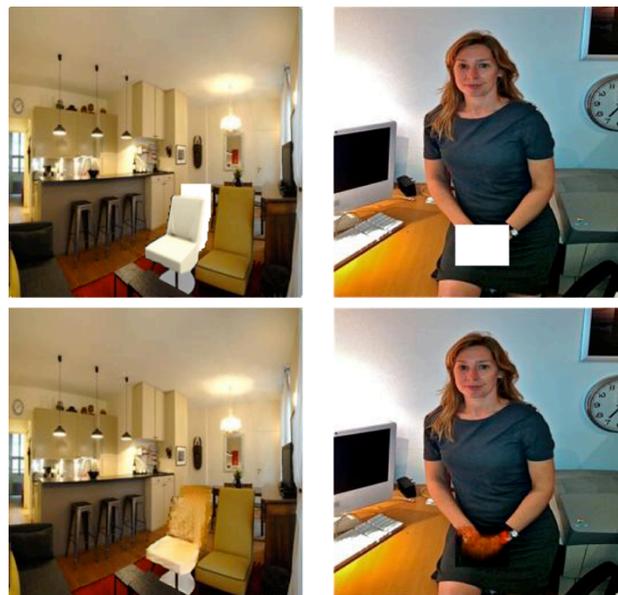


図4 アプリケーションとしての動作手順 (2例)

Figure4 Operating procedure as an application (2 cases).

(2) 考察

図5の左列(a)は白色のオブジェクトを挿入する例であるが、白色(画素値 255)で塗られた領域を補完対象としている影響のため、白色のオブジェクトと色が被ってしまい、補完領域とカラー調整の領域の区別がつかなくなったことが原因で自然な合成にならなかったと考えられる。この問題を解決する策として、カラー調整と補完領域を明示的に切り分けるために、バイナリマスク画像を入力に追加する



(a) 白色オブジェクト (b) 複雑な構造の補完の貼り付け

図5 失敗例

Figure5 Failure result.

方法を取ることが考えられる。

また、図5の右列(b)のように、人間の手や指のような複雑な構造を持つ対象を補完することは困難であることがわかる。飯塚ら[2]では顔パーツの補完を行なう実験もあり、その実験のために顔のデータセットのみで学習が行われている。手の補完の場合でも、手の画像のみで学習を行えば、複雑な構造ではあるが手の補完を行える可能性はある。

4. まとめ

本報告では、新たに提案したネットワーク構造および学習手順で、ユーザーが貼り付けと白抜き領域の選択という簡単な操作のみで自然な貼り付け合成を行うためのDNNを学習できることを示した。今後は、補完領域と挿入オブジェクトの境界を明示的に示すためのバイナリマスクも入力に加えることで、白色オブジェクトを貼り付ける場合に生じる問題を解決する予定である。

参考文献

- [1] Patrick Perez, et al. "Poisson Image Editing" In SIGGRAPH, 2003.
- [2] Satoshi Iizuka, Edgar Simo-Serra, et al. "Globally and Locally Consistent Image Completion" In SIGGRAPH, 2017.
- [3] Ian Goodfellow, et al. "Generative Adversarial Nets" In NIPS, 2014.
- [4] Olaf Ronneberger, et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation" In MICCAI, 2015.
- [5] Jonathan Long, et al. "Fully Convolutional Networks for Semantic Segmentation" In CVPR, 2015.
- [6] Deepak Pathak, et al. "Context Encoders: Feature Learning by Inpainting" In CVPR, 2016.
- [7] Bolei Zhou, et al. "Places: A 10 million Image Database for Scene Recognition" TPAMI, 2017.