

質感画像の弱教師領域分割と その結果に基づく質感の部分的変換

2I4-OS-07a-2

Partial style transfer for texture image using weakly supervised segmentation

下田 和*¹ 松尾 真*² 柳井 啓司*³
Wataru Shimoda Shin Matsuo Keiji Yanai

*¹ *² *³電気通信大学大学院 情報理工学研究科

Department of Informatics, The University of Electro-Communications

A style transfer technique based on Convolutional Neural Network (CNN) can change appearance of an image naturally while keeping its structure. However, this algorithm changes not a style of part of an image but a style of an entire image. In this paper, we propose a partial texture style transfer method by combining a style transfer method with segmentation. We segment target object regions using weakly supervised annotation and transfer a given texture style to only the segmented regions. As results, we achieved partial style transfer for only specific object regions.

1. はじめに

2015年, Gatysら [1] によって大規模な画像データセットで事前に学習された Deep Neural Network を用いることで画像のスタイルを変換するアルゴリズムが考案された。[1] の手法は細かいパラメータを設定せずに, 物体の形状を精密に維持して画像のスタイルを変更することができる。

本研究ではこのアルゴリズムを, 素材画像のデータセットとして広く知られている Flickr Material Database(FMD)[2] の画像について適用し, 画像内物体の質感の変換を行う。ただし, スタイルの変換は画像全体に対して行うために, 背景の質感も変化してしまう。そこで, 本研究では, 領域分割により質感領域を推定することで, 質感領域のみに対してスタイルの変換を行った。また, 質感の変換された画像について, 再度領域分割を行い, 結果の変化を確認した。

画像の質感の変換が可能となれば, 画像に対する心象を意図的に変化させることができ, デザイン業界など様々な分野での応用が期待できる。

2. 手法

本手法は, 主に画像のスタイル変換, 領域分割の二つの手法を組み合わせている。まず, [1] の手法により画像全体のスタイル(質感)を変換する。その後, 領域分割により対象の物体領域を推定することで, 対象物体の領域のみの画像合成による変換結果を変換する。また, 質感を変換した画像について, 再度領域分割を行い, 領域分割結果の変化を確認した。図1に本研究の概要を示した。

2.1 画像のスタイル変換

Gatysら [1] の手法を用いて画像を合成することで, 画像のスタイルの変換を行う。変換させる画像をコンテンツ画像 x_c , スタイル画像を x_s , 合成結果画像を x_g とする。 x_c, x_s, x_g のコンテンツ表現とスタイル表現を CNN の特定の layer の活性値から求め, x_g のコンテンツ表現が x_c に, スタイル表現が x_s に近くなるように反復的に合成する。使用した CNN は VGG19[3] であり, コンテンツ表現に使用する layer は conv4_2, スタイル表現に使用する layer は

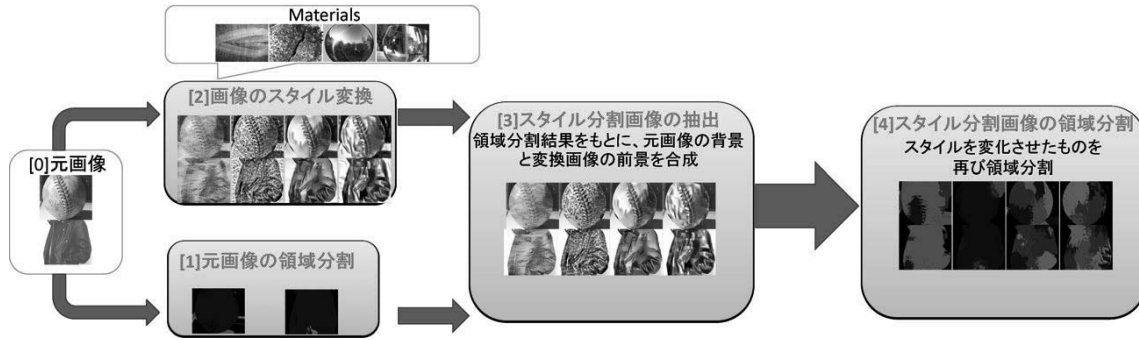


図 1: 実験の流れ

conv1_1, conv2_1, conv3_1, conv4_1, conv5_1 である。図 2 は VGG19 からコンテンツ表現とスタイル表現を抽出するレイヤーの略図である。

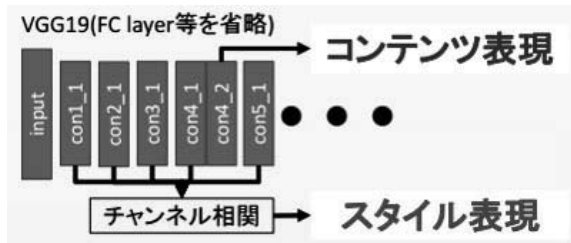


図 2: コンテンツ表現とスタイル表現の抽出レイヤーについての概要

layer l におけるコンテンツ表現はパラメータ数 N_l の活性値行列 $F(x, l)$, その誤差関数は x_c と x_g の差であり、式 1 で表される。

$$L_c(x_c, x_g) = \frac{1}{2} \sum_{i,j} (F_{i,j}(x_c, l) - F_{i,j}(x_g, l))^2 \quad (1)$$

layer l におけるスタイル表現は活性値行列の式 2 で表される相関行列 $G(x, l)$, その誤差関数は x_s と x_g の差であり、式 3 で表される。使用する layer 全体の誤差は重み w_l を用いて式 4 で表される。

$$G(x, l) = F(x, l)F^T(x, l) \quad (2)$$

$$L_{s,l}(x_s, x_g, l) = \frac{1}{4N_l^2} \sum_{i,j} (G_{i,j}(x_s, l) - G_{i,j}(x_g, l))^2 \quad (3)$$

$$L_s(x_s, x_g) = \sum_l w_l L_{s,l} \quad (4)$$

全体の誤差関数は重み w_c, w_s を用いて式 5 で表される。この式の値が最小となるように x_g を最適化する。

$$L(x_c, x_s, x_g) = w_c L_c(x_c, x_g) + w_s L_s(x_s, x_g) \quad (5)$$

2.2 弱教師あり領域分割

Simonyan ら [4] の手法を基にしてサリエンスマップを生成し、CRF[5] を適用することで、弱教師あり学習による領域分割を行う。図 3 に、領域分割の概要を示した。

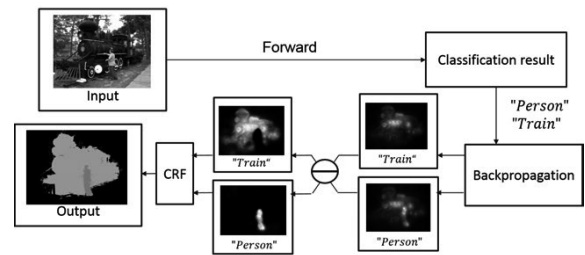


図 3: 領域分割手法の概要

2.2.1 CNN の学習

[6] はグローバルマックスプーリングを行うことで、バウンディングボックスのような詳細なアノテーションを必要とせず、高い精度でクラス分類を行った。本研究では、[6] の手法を、VGG16 [4] モデルに適応させた。

2.2.2 サリエンスマップ

[7] は CNN における学習アルゴリズムに着目し、Back propagation により得られる伝搬値が、物体の大まかな位置を反映していることを示した。本研究は、[7] の手法を以下の点について改良し、カテゴリごとに、物体の位置を表すサリエンスマップを生成した。その物体の位置を表すサリエンスマップを生成した。(1)[7] においては、画像レベルの伝搬値のみを用いて位置の推定を行ったが、本研究では中間層の伝搬値を用いることでより高い精度で位

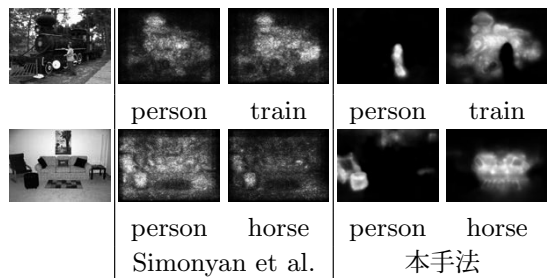


図 4: 左 (入力画像), 中 (simonayn et al.), 右 (Ours)

置の推定を行った。(2) 各カテゴリごとの信号から得られる伝搬値の差分をとることで, カテゴリに顕著なサリエンスマップを生成した。(3) 複数のサイズの入力画像から得られる伝搬値を統合した。(4) Reluの際に, Guided back propagation[8]を採用した。図 4 は一般画像における, 本手法と [7] の比較である。より鮮明なサリエンスマップが生成できていることがわかる。

2.2.3 Dense CRF

CRF はラベルの拡張手法であり, low level featre を用いて, 粗い領域分割結果から, スムースな領域を得るために用いることができる。本研究ではサリエンスマップを種として, Dense CRF[5]を適用し, 領域分割結果を改善した。[5]におけるエネルギー関数は以下の式に従う。

$$E(c) = \sum_i \theta_i(c_i) + \sum_{ij} \theta_{ij}(c_i, c_j) \quad (6)$$

単項は, $\theta_i(c_i) = -\log(\tanh(\alpha \cdot M_i^c))$ とした。 c はピクセルに割り当てられたラベルである。

本研究では *target* クラス + 背景クラスのラベルの領域拡張を行った。*target* は, 閾値で決定し, 背景クラスの probability は以下の式から求めた。

$$M^b g = 1 - \max_{c \in \text{target}} M_{x,y}^c \quad (7)$$

平滑化項は [5] に従った。図 5 に質感画像における領域分割結果の例を示す。

3. 実験

Flickr Material Database [2] は 10 種類 (fabric, foliage, glass, leather, metal, paper, plastic, stone, water, wood) の

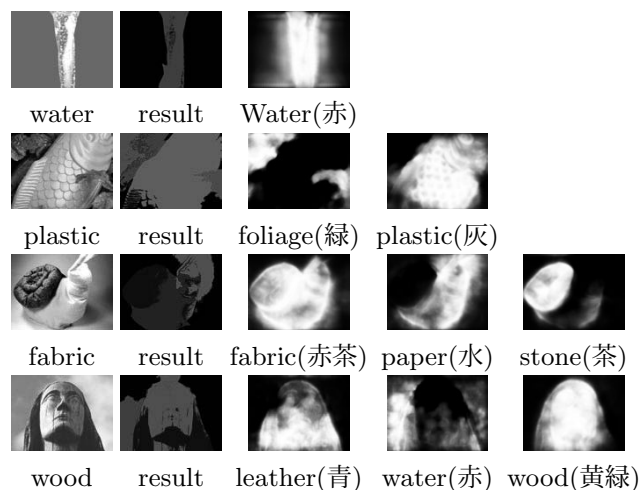


図 5: 質感画像の領域分割結果とカテゴリごとのサリエンスマップ

素材画像, 各 100 枚, 合計 1000 枚からなるデータセットである。本研究は [2] から数枚の画像を用いて実験を行った。

実験は図 1 のように, 画像全体のスタイル変換, コンテンツ画像の領域分割, 変換領域の抽出 (変換), 質感の変換結果についての領域分割を行った。図 6 に, 2 種類のコンテンツ画像について, 10 種類のスタイル画像を用意し, 合計 20 組の画像についての実験結果を示した。

多くの場合で, コンテンツ画像の物体の形状を精密に維持したまま, 質感の変換を行うことが出来ているのが見て取れる。また, 領域分割の結果を取り入れることで, より自然な質感の変換を実現することが出来た。また, 表 1 に図 6 における質感領域の質感変換画像の領域分割結果を示した。質感の変換結果を再度領域分割することで, fabric, foliage, stone, water では高い精度でスタイルの領域を推定することが出来た。これらは, 肉眼で確認しても比較的よい結果であり, よい変換を行うことが出来ていると言えるのではないかと考えることが出来る。ただし, glass, metal では低い領域分割結果の精度となった。これは, glass や metal の質感が大きく光沢に依存していることが原因として考えられる。glass や metal の質感は外部の環境と強い関係があるために, 質感の学習, 認識が難しかった可能性がある。今回は 2 種類のコンテンツ画像について, 10 種類の素

材について質感の変換を行ったが、再領域分割結果の精度は、素材ごとに共通していることが見て取れる。これは、質感の変換が容易な素材と、難しい素材があることを示しているためであると考えられる。

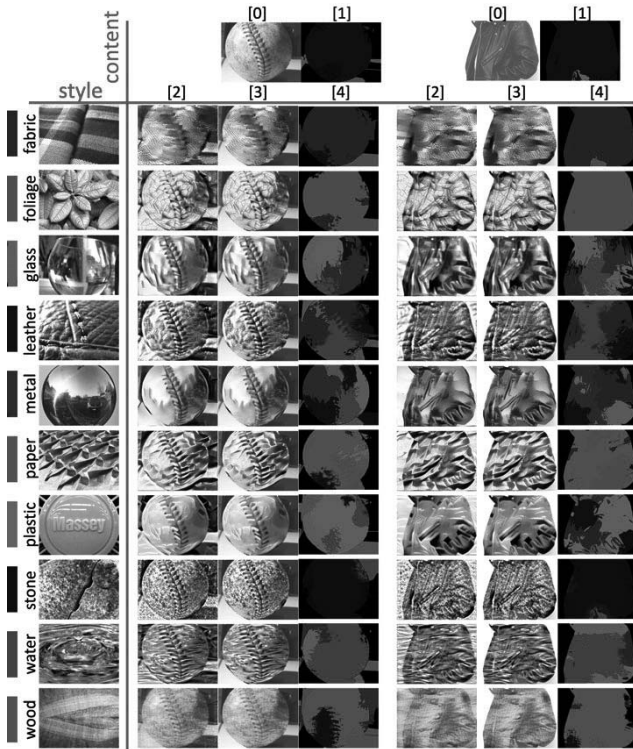


図 6: スタイル変換実験の結果。画像の上に付与された番号は実験の段階に対応している。領域の色は素材のラベルの左側の色に対応している。

4. まとめ

画像のスタイルを変換するアルゴリズムを、弱教師有領域分割を用いて画像に部分的に適用することで画像内物体の質感を変換した。使用した質感画像は FMD の 10 種類の素材画像であり、2 種類のコンテンツ画像について、質感の変換を行った。その結果、多くの例で違和感のない質感変換を行うことができた。しかし、その変換後の画像の領域分割結果は素材によりばらつきが見られた。目的に合ったよいスタイル画像の自動選択、スタイル変換アルゴリズムの改良、領域分割の精度向上を今後の課題としたい。

参考文献

[1] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. In *arXiv:1508.06576*, 2015.



class				
	pixel acc.	mean IU	pixel acc.	mean IU
fabric	0.83	0.72	0.96	0.93
foliage	0.85	0.77	0.99	0.96
glass	0.41	0.53	0.36	0.56
leather	0.24	0.33	0.27	0.52
metal	0.44	0.52	0.44	0.60
paper	0.77	0.79	0.64	0.74
plastic	0.60	0.60	0.54	0.67
stone	0.87	0.79	0.95	0.93
water	0.77	0.74	0.69	0.78
wood	0.59	0.51	0.89	0.82

表 1: 図 6 における質感領域の質感変換画像の領域分割結果の精度比較

[2] C. Liu, L. Sharan, E. Adelson, and R. Rosenholtz. Exploring features in a bayesian framework for material recognition. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2010.

[3] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proc. of arXiv:1409.1556*, 2014.

[4] K. Simonyan, A. Vedaldi, and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[5] P. Krahenbuhl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in Neural Information Processing Systems*, 2011.

[6] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Is object localization for free? -weakly-supervised learning with convolutional neural networks. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2015.

[7] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *International Conference on Learning Representations*, 2014.

[8] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for simplicity: The all convolutional net. In *International Conference on Learning Representations*, 2015.