社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

Neural Style Transferと領域分割による画像の部分的質感操作

松尾　真† 下田　和† 柳井　啓司†

† 電気通信大学大学院 情報理工学研究科

あらまし Convolutional Neural Network(CNN) を用いた画像のスタイル変換アルゴリズムは画像の構造を精密に保ちながら，外観を変化させることができる。しかし，画像内の特定の物体のみをスタイル変換するためには対象物体のある領域を指定する必要がある。本研究では弱教師有領域分割とその結果に基づくスタイルの部分的変換を用いて画像内の物体の質感を転送する。

Partial style transfer for texture image using weakly supervised segmentation

Shin MATSUO†, Wataru SHIMODA†, and Keiji YANAI†

† Department of Informatics, The University of Electro-Communications

Abstract 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 から得られた統計量を用いることで物体の形状を精密に維持して画像のスタイルを変換するアルゴリズムが考案された。

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，領域分割により特定の質感領域を推定することで，該当する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，対称となる領域が目的の質感領域として認識されるか確認した。

画像内物体の質感の任意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。

2.1 実験の手順

以下のように実験の手順は図 1 のように行う。

（1） [1] の手法により画像全体のスタイル (質感) を変換

（2） 領域分割によりコンテンツ画像の物体領域を推定

（3） スタイル変換画像の物体領域とコンテンツ画像の背景を合成し、物体領域のみをスタイル変換した画像を生成

（4） 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評価

2.2 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

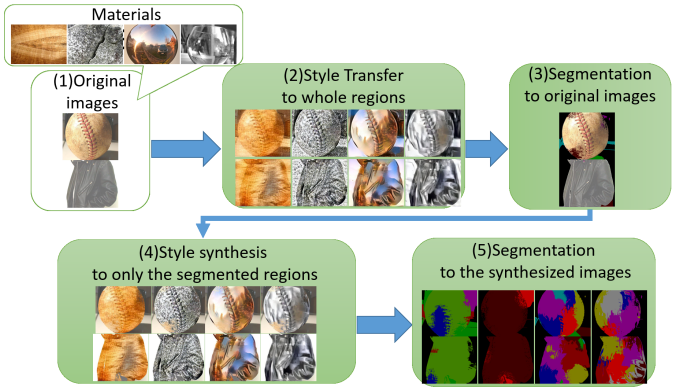

図 1　実験の流れ

使用する layer は conv4_2, スタイル表現に使用する layer は conv1_1,conv2_1,conv3_1,conv4_1,conv5_1 である。図 2 にスタイル変換アルゴリズムの概略を記す。

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 $F(x,l)$, その損失関数は x_c と x_g の差であり、式 1 で表される。

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

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 $G(x,l)$, その損失関数は 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。

$$G(x,l)=F(x,l)F^T(x,l) \tag{2}$$

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

$$Loss_s(x_s,x_g)=\sum_l w_l Loss_{s,l} \tag{4}$$

全体の損失関数は重み 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 L-BFGS 法を用いて最適化する。

$$Loss(x_c,x_s,x_g)=w_c Loss_c+w_s Loss_s \tag{5}$$

2.3　弱教師あり領域分割

Simonyan ら [4] の手法を基にしてサリエンシーマップを生成し，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。

2.3.1　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に，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。

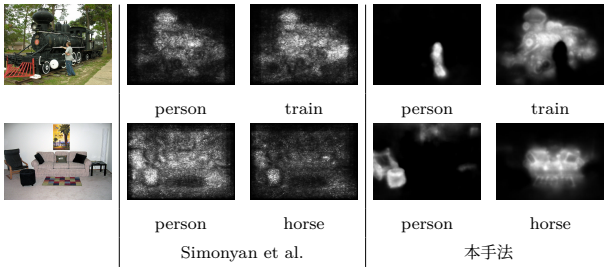

図 4　左（入力画像），中（simonayn et al.），右（Ours）

2.3.2　サリエンシーマップ

[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シーマップを生成した。その物体の位置を表すサリエンシーマップを生成した。(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シーマップを生成した。(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シーマップが生成できていることがわかる。

2.3.3　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シーマップを種として，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。

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

単項は，$\theta_i(c_i)=-\log(\tanh(\alpha\cdot \mathrm{M}_i^c))$ 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。

$$\mathrm{M}^b g=1-\max_{c\in\mathrm{target}} M_{x,y}^c \tag{7}$$

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

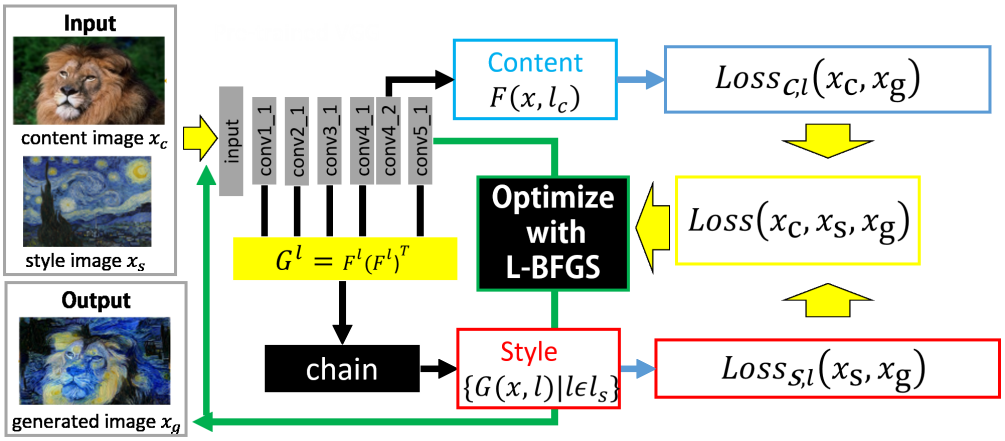

図 2　スタイル変換アルゴリズム

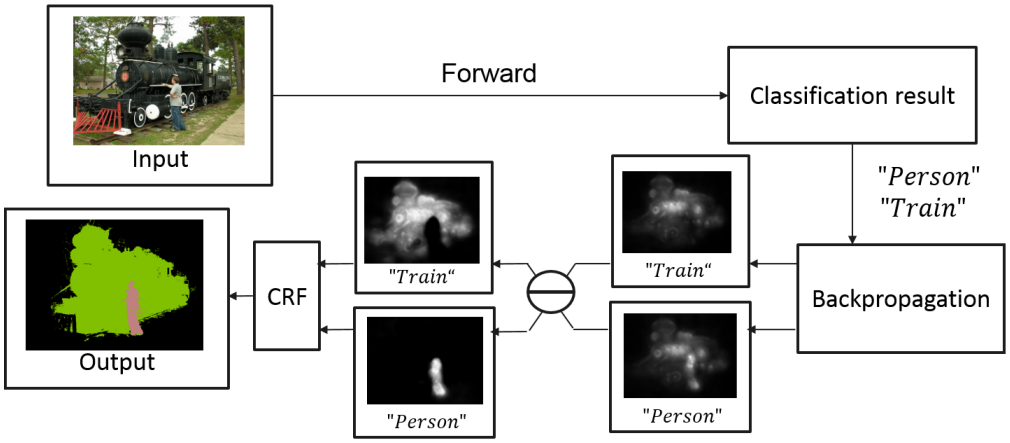

図 3　領域分割手法の概要

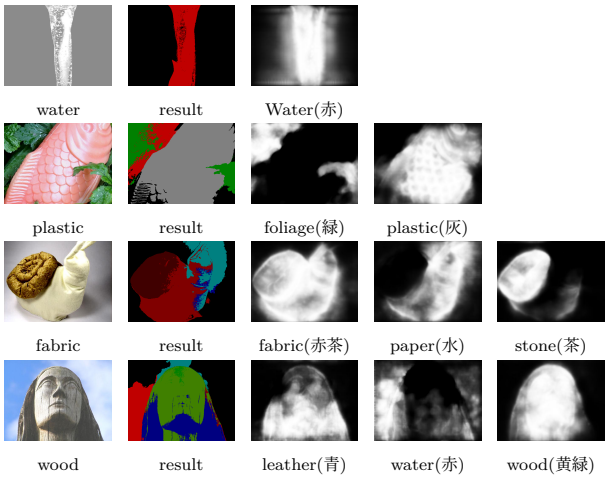

図 5　質感画像の領域分割結果とカテゴリごとのサリエンシーマップ

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 FMD から各素材 20 枚, 合計 200 枚の画像をスタイルとして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。コンテンツ画像 2 枚 (ball, jacket), スタイル画像 10 種類, 各 20 枚による合計 400 組の組み合わせについての実験を行った。

領域分割の精度は Pixel Acc と Mean IU の 2 つの尺度で評価する。Pixel Acc は画像全体のピクセルの中で、正解のラベリングをしているものの割合である。tn を正解ラベリング数,N をピクセルの総数として以下の式で求められる。

$$acc_{pixel\_acc} = tn/N \tag{8}$$

Mean IU は画像のピクセルを i, j として、以下の式により求められる。

$$t_i = \sum_i n_{i,j} \tag{9}$$

$$acc_{mean\_IU} = \sum_i n_{i,i} / \sum_i n_i \tag{10}$$

Mean IU はピクセル単位のずれの評価であり、ラベリング領域が正解とずれていれば, 広くても、狭くても精度は低下する。また，ラベリングが間違っている場合は極端に精度が低下する。

各コンテンツ・素材ごとの Pixel Acc と Mean IU をそれぞれ示した結果が図 9, 10 である。図 11 に，各素材 1 枚ずつ結果例を示した。表 1 は図 11 における質感領域の質感変換画像の領域分割の精度を示す。

4. 考　　察

図 11 から、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おり，さらに、領域分割の結果を取り入れることで，対象物体のみの質感の変換を実現することが出来ていることがわかる。

図 9, 10 の結果から、質感の変換結果を再度領域分割した結果、foliage,fabric,stone は特に高い精度で物体の領域が変換後のスタイルとして認識された。対して、特に精度が低かったのは、metal,glass,plastic だった。

foliage,fabric,stone 素材による変換結果は，肉眼で確認しても比較的よい結果であり，どの素材に変換されたのかの識別が容易である。これらの素材の画像は不規則できめの細かい構造を持つスタイル画像が多く、コンテンツに合わせての自然な変形が容易であるからだと考えられる。

図 6 はこれらの素材において、ball をスタイル変換して再領域分割を行った際に精度が低かった例である。スタイル変換の段階で失敗した例と領域分割の段階で失敗した例が見られた。前者は構造に多様性が足りないスタイル画像に対して、後者は複数の色や素材が混在したスタイル画像に対して見られた。

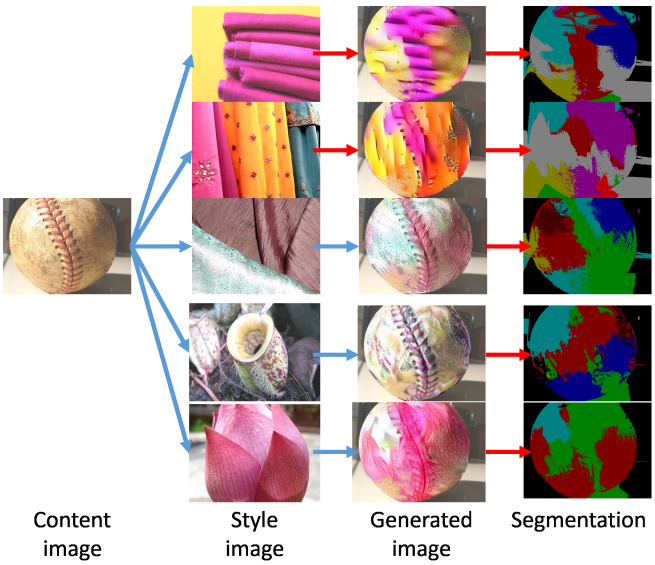

図 6　失敗例 (foliage, fabric, stone)
赤の矢印：精度低下の原因と思われるプロセス

それに対し，glass, metal, plastic 素材といった、領域分割の精度の低い素材ではどの素材に変換されたかの識別が難しい結果となっている。図 7 はこれらの素材において、ball をスタイル変換して再領域分割を行った際に精度が低かった例である。これらの例はいずれも再領域分割の結果が入り乱れており、スタイル変換画像も素材を推定しにくいものとなっている。これらの素材画像は色、形状が多彩であり、その質感がその光沢および反射される外部の環境に依存しているため、質感の学習，認識が難しかったことが精度低下の原因であると考えられる。

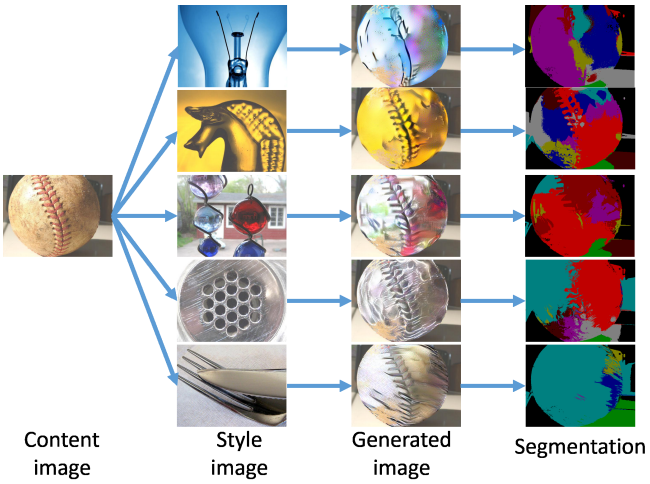

図 7　失敗例 (glass, metal, plastic)
赤の矢印：精度低下の原因と思われるプロセス

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の傾向は，素材ごとにある程度共通していることが見て取れる。したがって、変換の対称となるコンテンツに依らず, 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。このことから，質感変換に使用するスタイル画像は事前にある程度の選別が必要になると考えられる。

また、fabric 素材の再領域分割の精度は図 8 のように jacket の方が ball よりも高い精度が見られた。これは fabric 素材のデータには衣服などの、jacket と共通する部分的構造が多く含まれていたことが理由だと思われる。コンテンツ画像を初期状態としたスタイル変換ではスタイル画像に近づけるに連れて、コンテンツが崩れてしまうジレンマがあるが、共通構造があれば、コンテンツとスタイルを双方保持することが容易だからである。したがって、コンテンツ画像と類似した物体をスタイル画像とすることで適切なスタイル変換を促進できると思われる。

5. ま　と　め

本研究では、画像のスタイルを変換するアルゴリズムと弱教師有領域分割を併用することで画像のスタイルを部分的に変換し、画像内物体のみの質感を変換した。

使用した FMD の 10 種類の素材画像各 20 枚, 計 200 種類のスタイル画像と 2 種類のコンテンツ画像を用いて，400 の組み合わせで質感の変換を行った。その結果，画像内物体の質感変換については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，変換後の画像の領域分割結果は素材によりばら

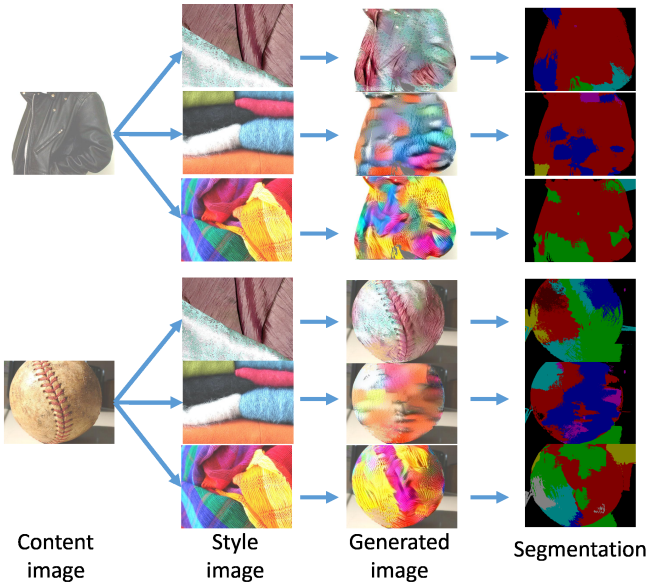

図 8　共通部位による優位性

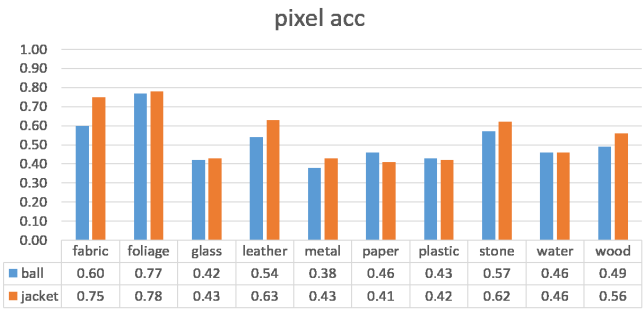

図 9　再領域分割の精度 (pixel acc)

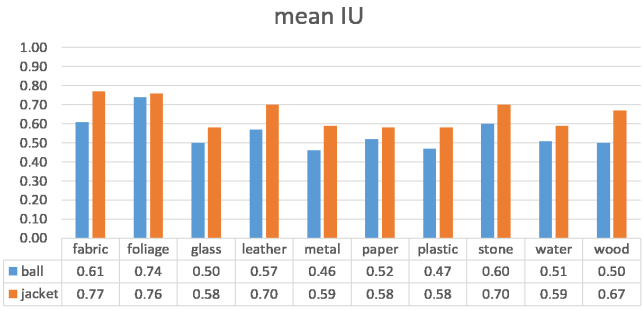

図 10　再領域分割の精度 (mean IU)

つきが見られた。中でも、metal, glass, plastic 素材は現在のスタイル変換アルゴリズムでの質感の再現が難しい素材であるということが分かった。また、変換に使用するスタイル画像はきめ細かな構造が多く、コンテンツ画像に類似する構造があるほど適切な変換がされやすいことが分かった。

したがって、今回の実験で分かったスタイル変換の傾向を活かし、今後は目的に合った良質なスタイル画像の自動選択，光沢をもつ物体に対応できるようなスタイル変換アルゴリズムの改良および領域分割の精度向上を今後の課題としたい。

class	 pixel acc.	mean IU	 pixel acc.	mean IU
fabric	0.6446	0.6132	0.8091	0.8147
foliage	0.7556	0.7309	0.8668	0.7101
glass	0.4409	0.5520	0.4301	0.5999
leather	0.5858	0.5756	0.5706	0.6897
metal	0.3086	0.4563	0.3456	0.5616
paper	0.4491	0.4983	0.3196	0.5343
plastic	0.3227	0.4212	0.1931	0.3012
stone	0.7186	0.7293	0.9431	0.9054
water	0.4683	0.5073	0.4719	0.6454
wood	0.3550	0.4567	0.2701	0.5227

表 1　図 11 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

また、今回はスタイル変換と弱教師有領域分割を用いて画像の部分的スタイル変換を行ったが、実際にはそれぞれのプロセスを別々に実行し、終了した後に得られた結果を合成しているにすぎないため、完全に連携しているとは言い難い。そのため、スタイル変換の中で、領域分割の情報を与えられていないため、一度画像全体をスタイル変換し、不要な部分を取り除くというプロセスが必要となっており、コンテンツ画像によっては背景との不自然な境界線を作ってしまうことがある。したがって、ニューラルネットワークを統合し、領域分割とスタイル変換を同時並行で行うシステムの構築も今後課題となる。

文　献

[1] “A neural algorithm of artistic style”, arXiv:1508.06576 (2015).
[2] “Exploring features in a bayesian framework for material recognition”, Proc. of IEEE Computer Vision and Pattern Recognition (2010).
[3] “Very deep convolutional networks for large-scale image recognition”, Proc. of arXiv:1409.1556 (2014).
[4] “Very deep convolutional networks for large-scale image recognition”, International Conference on Learning Representations (2015).
[5] “Efficient inference in fully connected crfs with gaussian edge potentials”, Advances in Neural Information Processing Systems (2011).
[6] “Is object localization for free? -weakly-supervised learning with convolutional neural networks”, Proc. of IEEE Computer Vision and Pattern Recognition (2015).
[7] “Deep inside convolutional networks: Visualising image classification models and saliency maps”, International Conference on Learning Representations (2014).
[8] “Striving for simplicity: The all convolutional net”, International Conference on Learning Representations (2015).

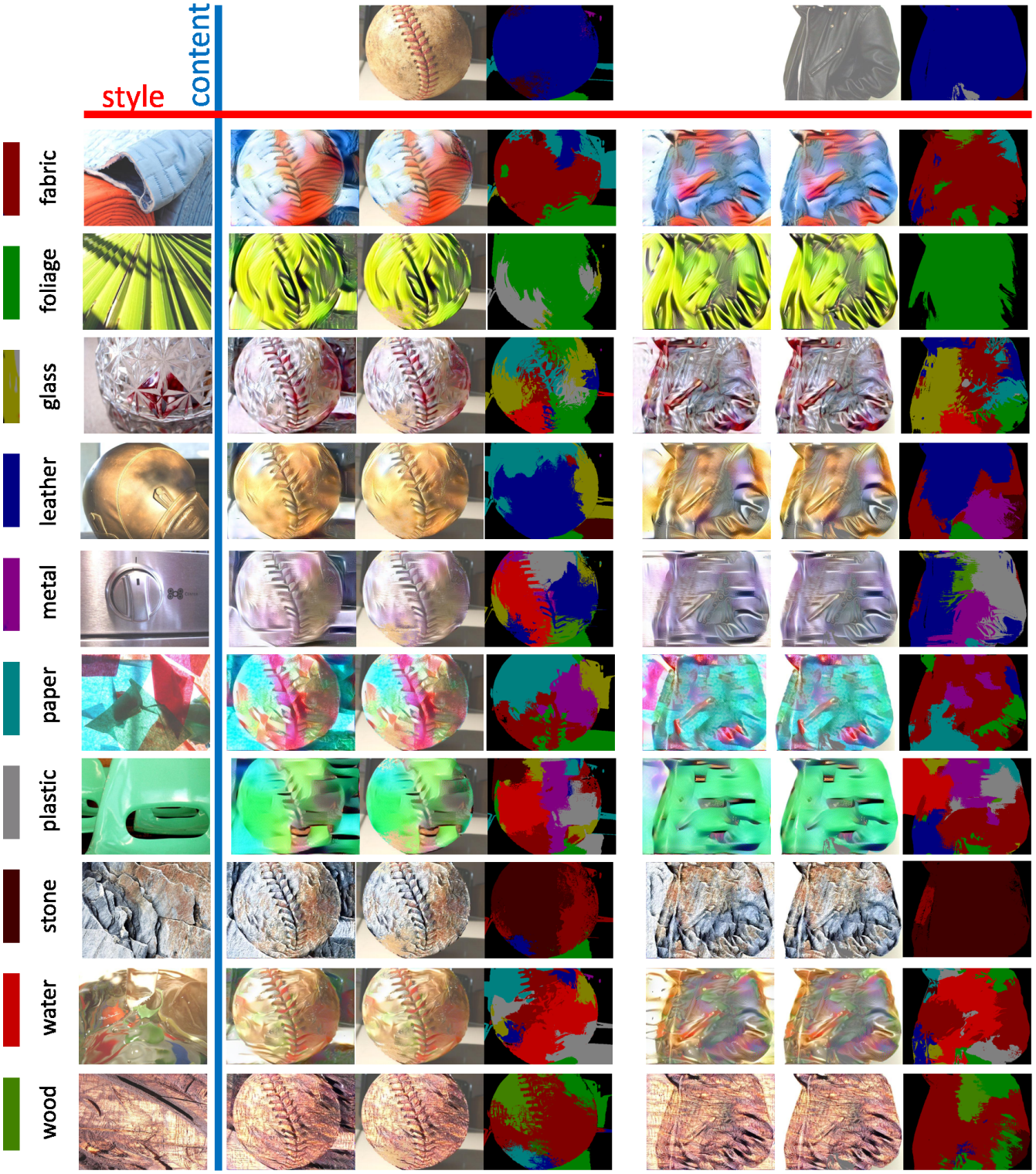

図 11　スタイル変換実験の結果。領域の色は素材のラベルの左側の色に対応している。