CNNによるスタイル変換とWeb画像を用いた画像の任意質感生成

松尾 真, 柳井 啓司

電気通信大学　大学院情報理工学研究科
総合情報学専攻

はじめに

Neural Style Transfer

- Deep Neural Networkを用いた画像変換
  - 写真と絵画を入力すると，写真が絵画調に
- DNNによる画像合成分野が発足

content image

style image

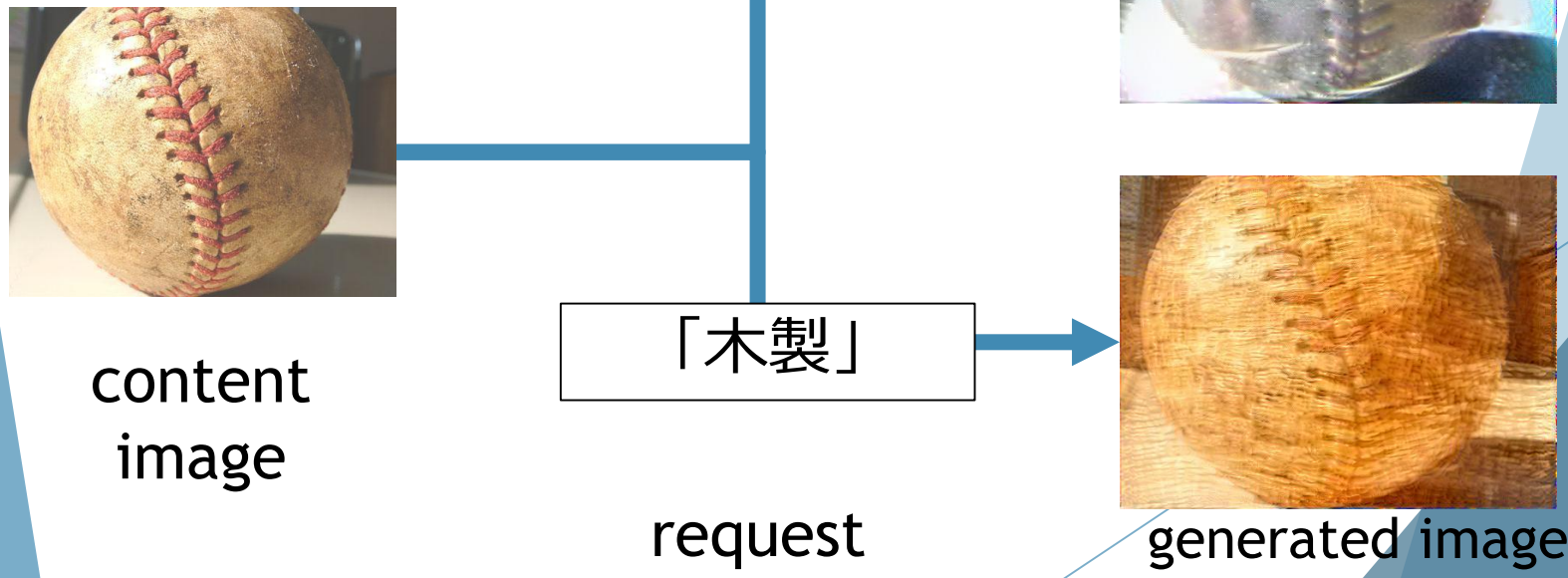
generated image

目的
単語概念からの自動スタイル生成

画像内物体の質感の任意変換への応用

- 質感別画像データの増量
- デザイン，エンターテインメント

  様々な分野への応用が期待できる

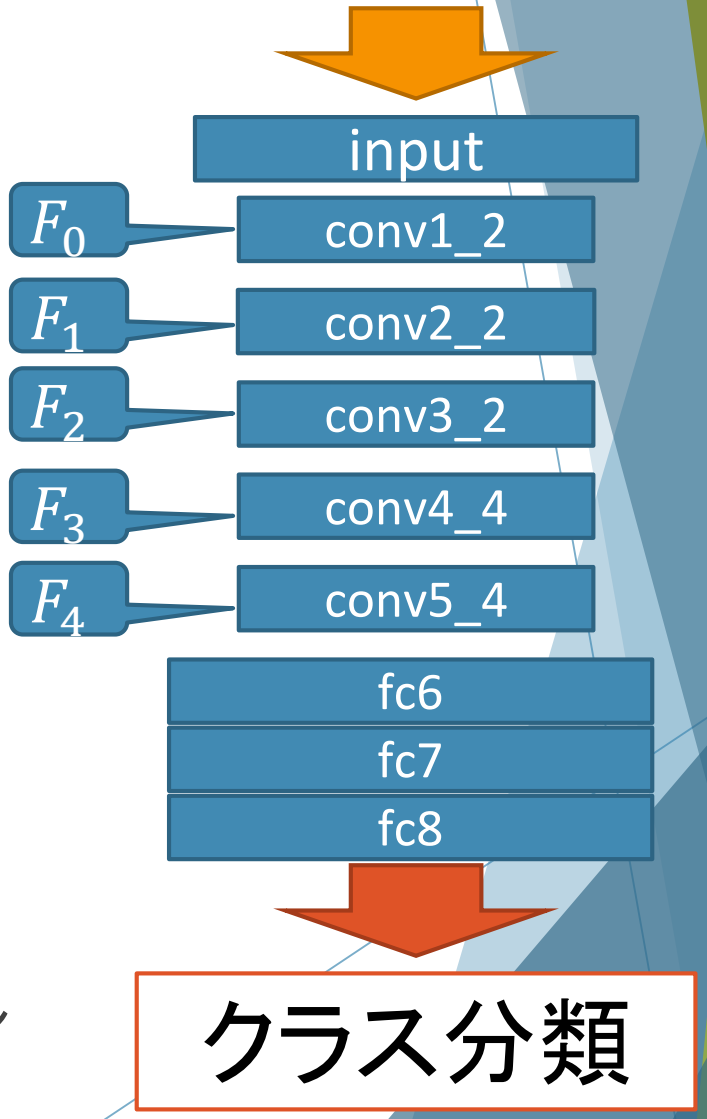


関連研究

Neural Style Transfer

L. A. Gatys et al, “A neural algorithm of artistic style,” in arXiv:1508.06576, 2015.
[Gatys et al.]

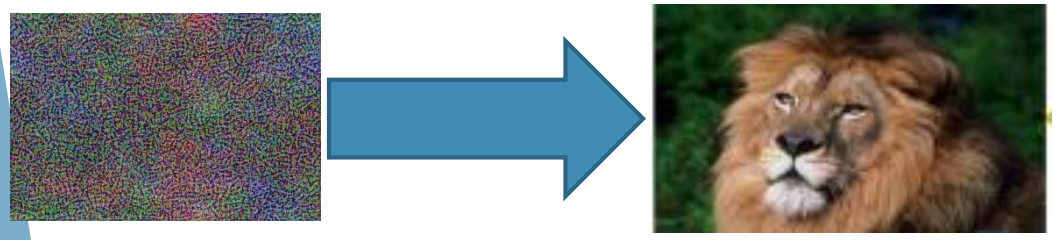
- ImageNet1000クラスの分類を学習済みのDNN
- 画像入力時に各Layer内でフィルタの活性値行列F_lが発生する



関連研究

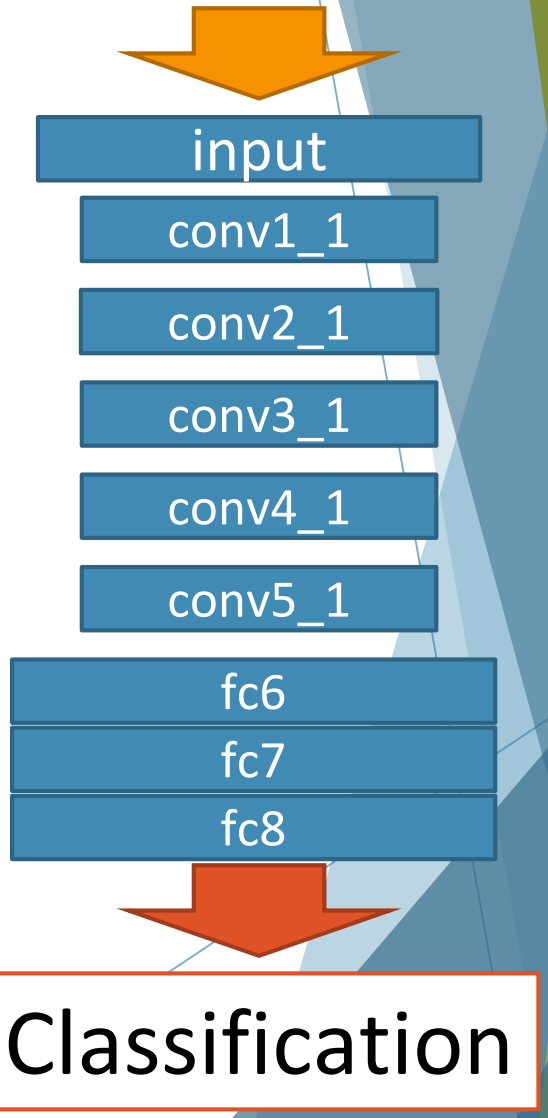
Neural Style Transfer
[Gatys et al.]

生成画像と元画像の F_lのずれを最小化するように生成画像を最適化



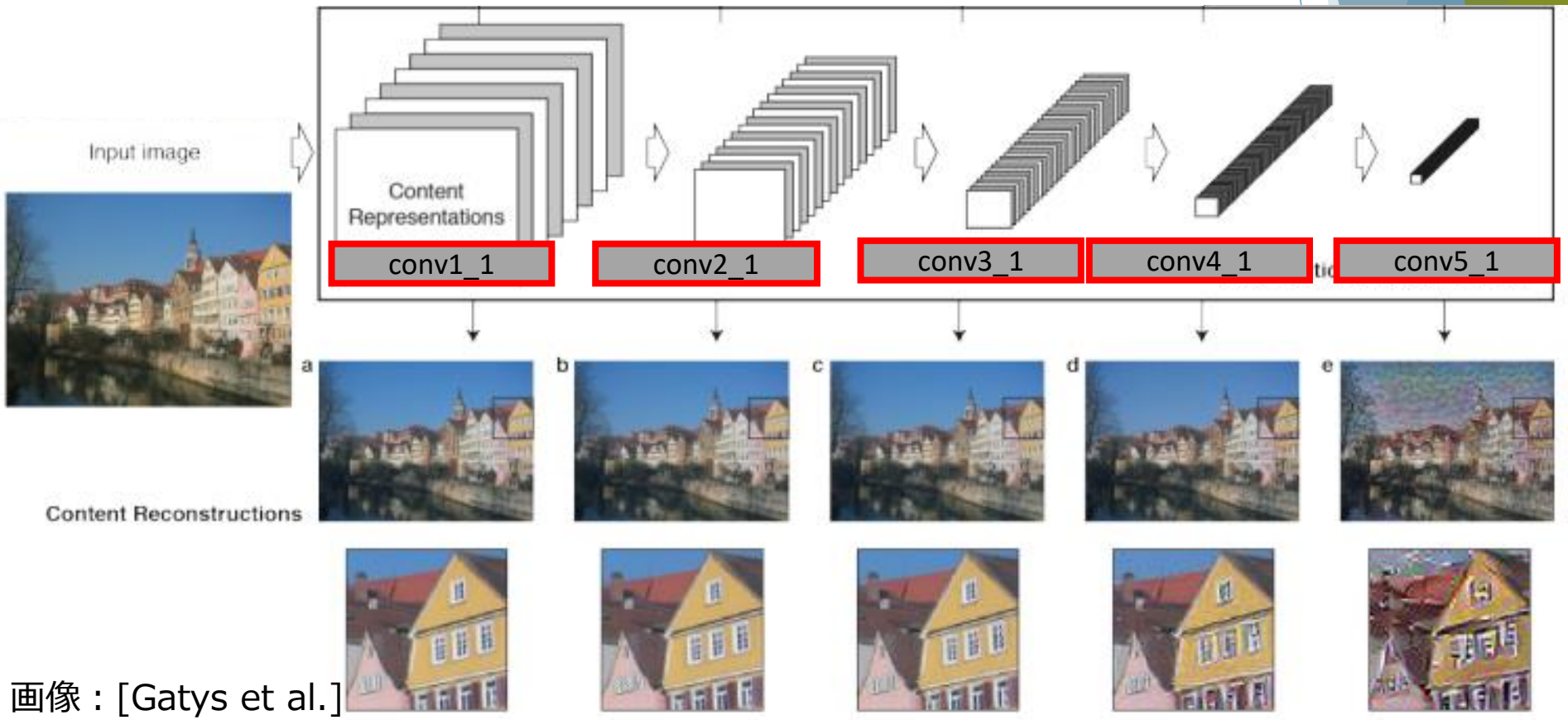
元画像を復元！

VGG19 Network
(Max pool層等省略)



Neural Style Transfer
[Gatys et al.]

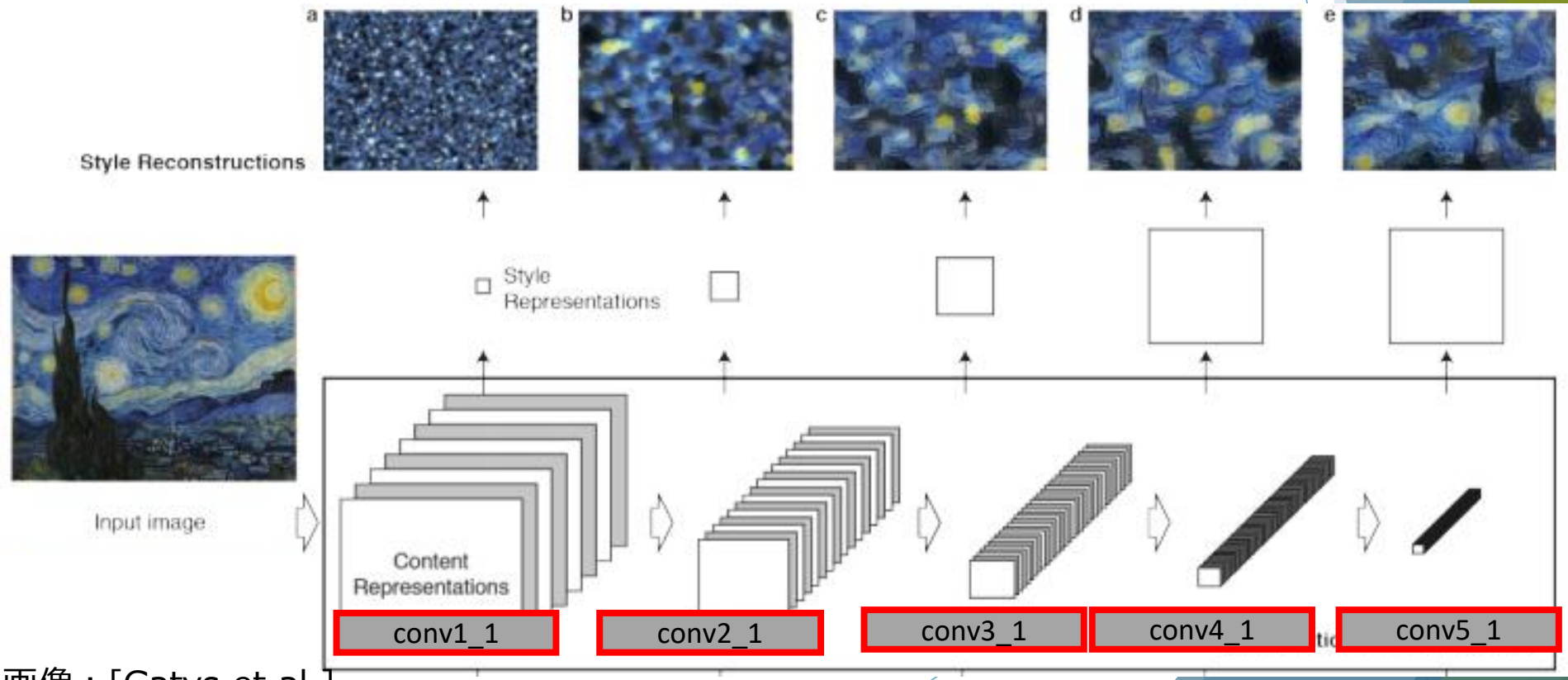
コンテンツ表現：F_lによる画像の復元



画像：[Gatys et al.]

Neural Style Transfer
[Gatys et al.]

スタイル表現：$G_l=F_l F_l^T$による画像の復元



画像：[Gatys et al.]

Neural Style Transfer
[Gatys et al.]

二種の画像表現から損失関数を形成

- コンテンツ表現：深い層におけるF_l
  - 形状のみを保持
- スタイル表現：様々な層における$G_l=F_l F_l^T$
  - スタイルのみを保持

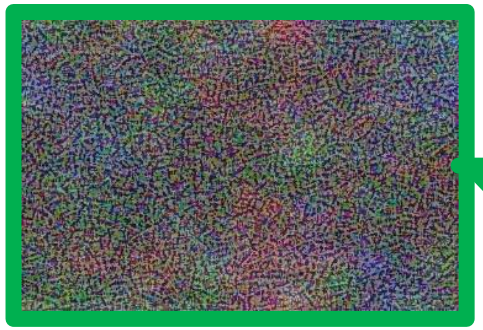
- 2枚の異なる画像の形状，スタイルを持った画像を復元

入力

コンテンツ画像x_c

スタイル画像x_s

出力（初期状態）



生成画像x_g

Pre-trained
DNN(VGG19)

input
conv1_1
conv2_1
conv3_1
1
1

$F(x_c)$
$F(x_g)$
$Loss_{content}(x_c, x_g)$

画像
最適
(L-BF

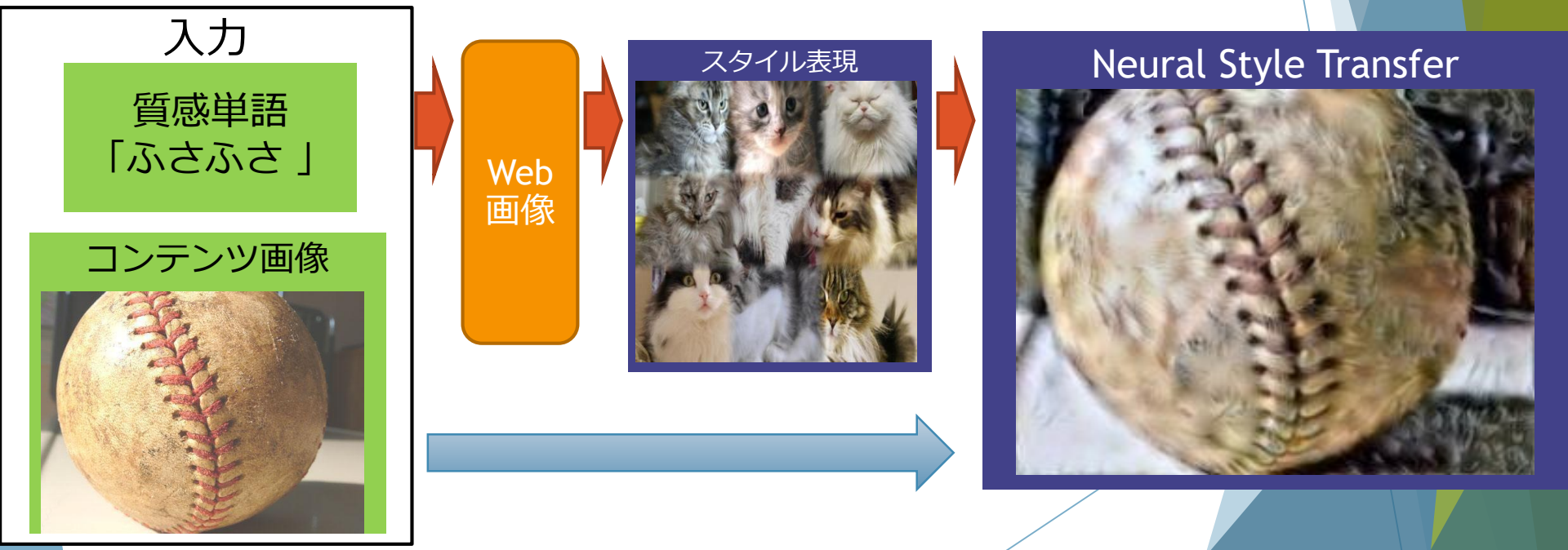
出力（最終状態）

生成画像x_g

生成画像のコンテンツをコンテンツ画像に，スタイルをスタイル画像に近づける

提案手法

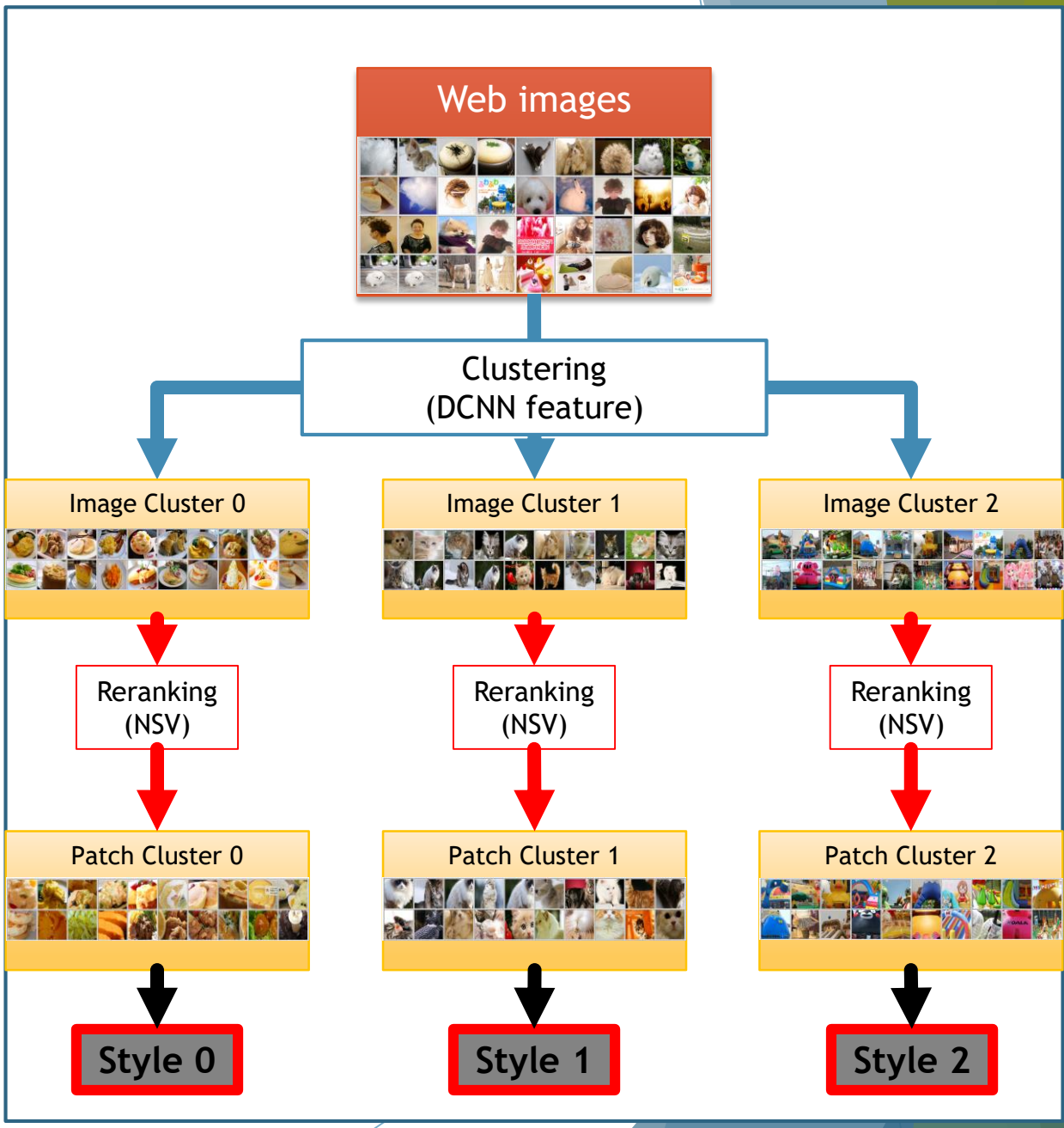
- CNNを用いた画像内物体の任意質感生成
- 質感単語からスタイル表現を自動生成



単語による画像収集とクラスタリング

Web画像をスタイルの類似したもの同士のクラスタに分化

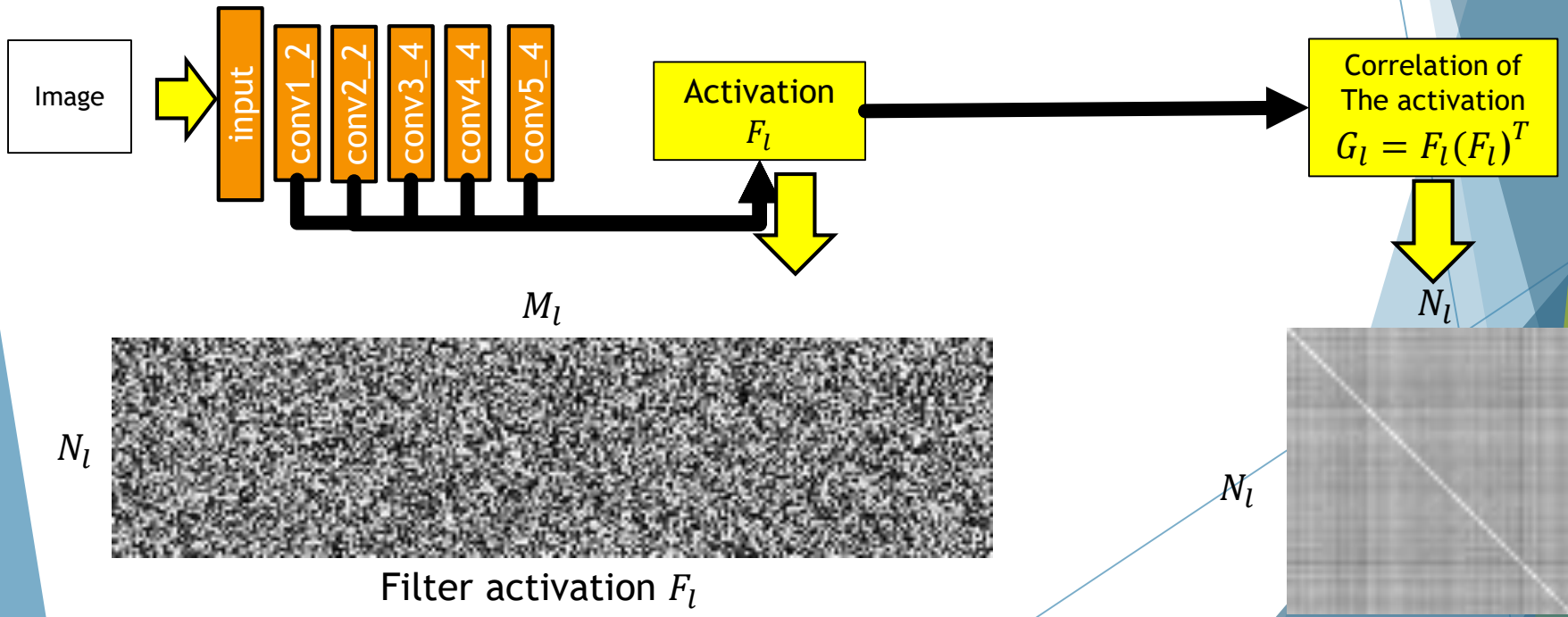
- DCNN特徴(VGG19 fc6の出力)とNeural Style Vectorを使用



Neural Style Vector

CNN-based style vector for style image retrieval. [S. Matsuo, ACM ICMR 2016.]

- スタイル変換アルゴリズムにおけるスタイル 表現をベクトル化
- 絵画画像の検索・クラス分類に利用



N_l: the number of filters　　M_l ： the number of pixels

Neural Style Vector

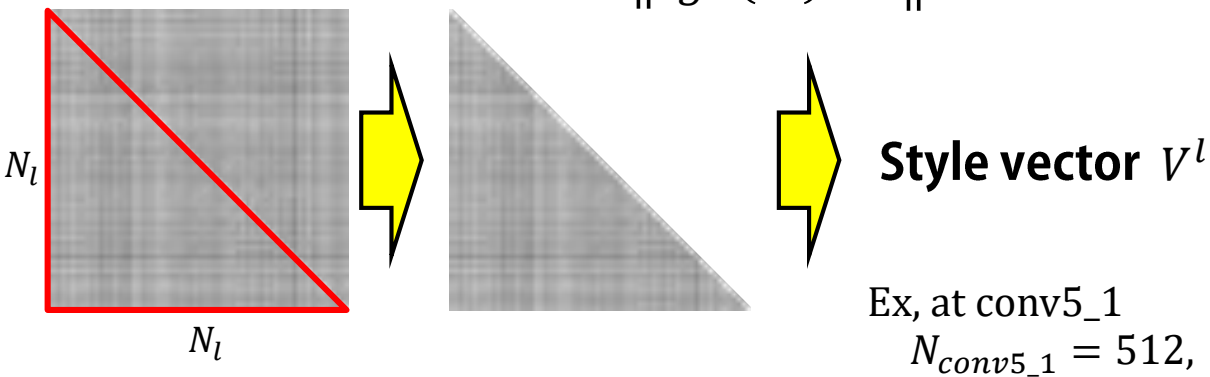
G^lの対象要素と対角要素で定義

$$V^l = [G^l_{1,1}, G^l_{2,1}, G^l_{2,2}, \dots, G^l_{N_l,1}, G^l_{N_l,2}, \dots G^l_{N_l,N_l}]$$

$$|V^l| = (hurf\ elements) + (diagnal\ elements) = N_l * (N_l + 1)/2$$

符号付平方根とL2正規化をして使用

$$V^{l_{sgnsqrt}} = \frac{sgn(V^l)\sqrt{V^l}}{\|sgn(V^l)\sqrt{V^l}\|}$$



Style vector V^l

Ex, at conv5_1
$N_{conv5\_1} = 512$,
$|V^{conv5\_1}|$=131,328

実験

実験データ

- コンテンツ画像：3枚
- 質感単語：オノマトペ 23 単語
- 画像収集：Bing API
- 使用クラスタ数：3

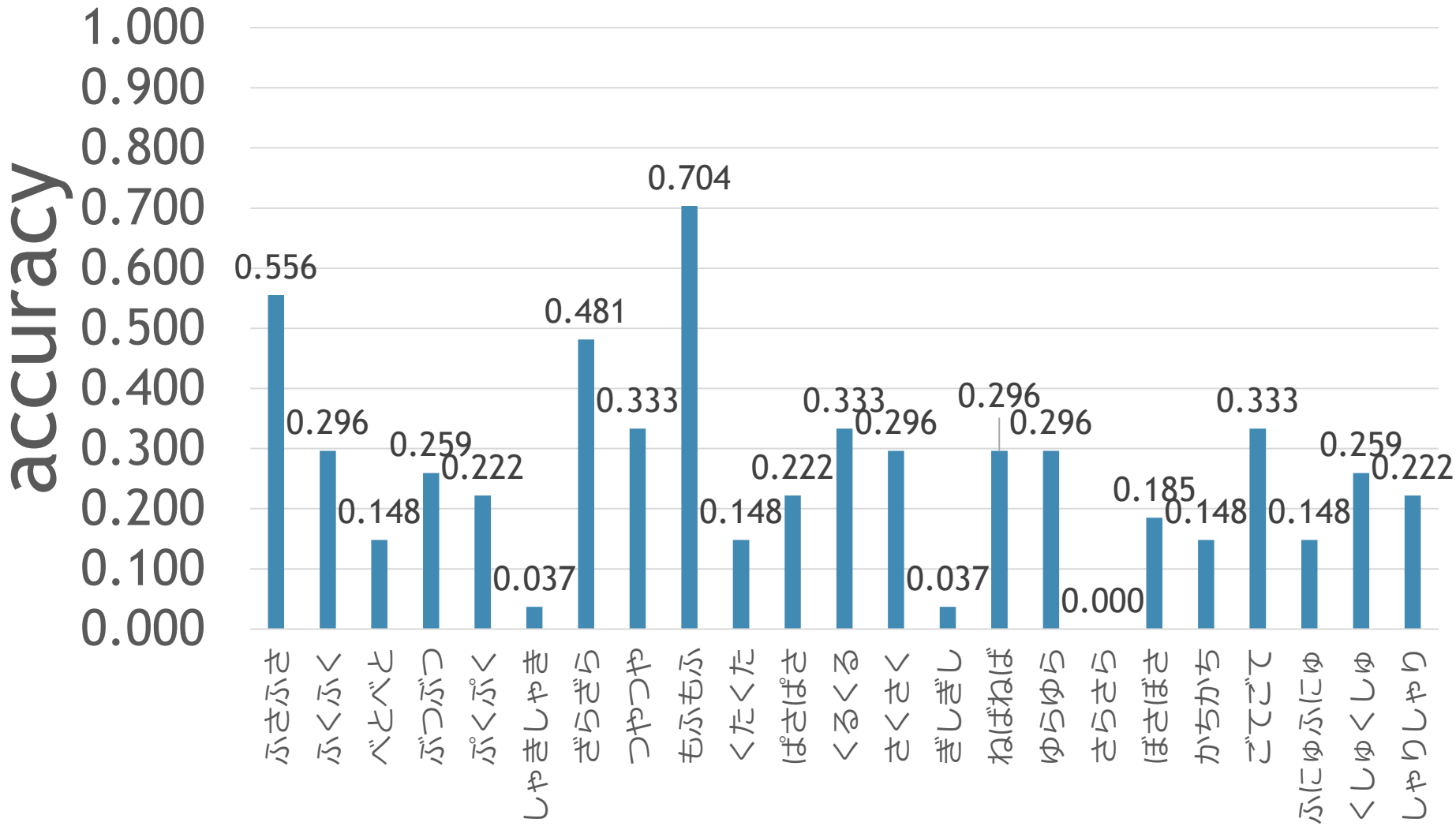
評価方法

- ユーザー評価
- 問題数はコンテンツ数 3単語数 23 語の 69 問
- 回答者数は 9 人

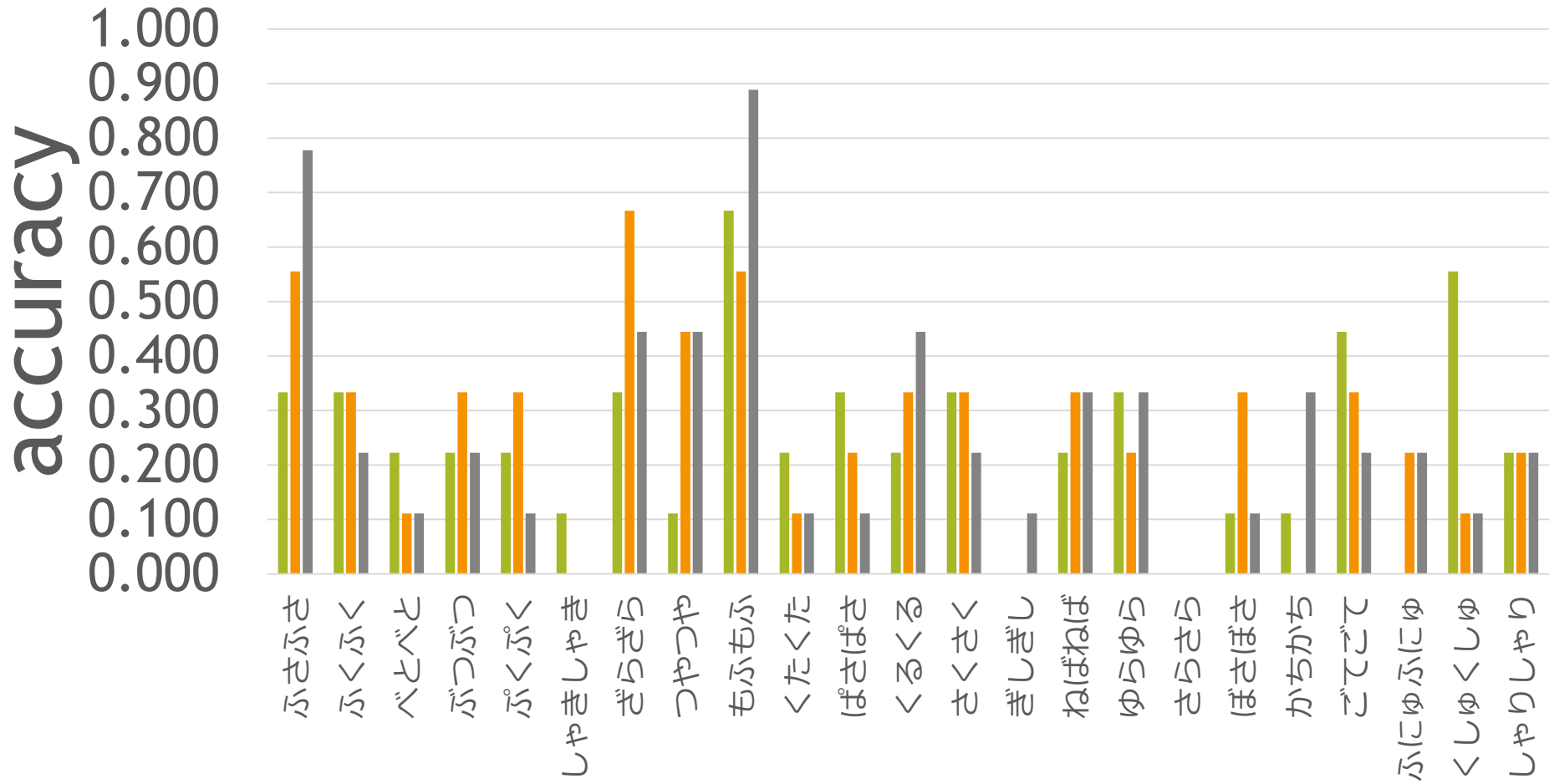
ユーザー評価

- 赤枠が正解

実験結果（総合）


accuracy
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
0.556
0.296
0.148
0.259
0.222
0.037
0.481
0.333
0.704
0.148
0.222
0.333
0.296
0.037
0.296
0.296
0.000
0.185
0.148
0.333
0.148
0.259
0.222
ふさふさ
ふくふく
べとべと
ぶつぶつ
ぷくぷく
しゃきしゃき
ざらざら
つやつや
もふもふ
くたくた
ぱさぱさ
くるくる
さくさく
ぎしぎし
ねばねば
ゆらゆら
さらさら
ぼさぼさ
かちかち
ごてごて
ふにゅふにゅ
くしゅくしゅ
しゃりしゃり

実験結果（コンテンツ別）



content0

content1

content2

結果

- 「ふさふさ」，「ざらざら」，「もふもふ」の正解率が高くなった。

- 他の単語では正解率は低くなった。
  1. Web画像が不適切
  2. オノマトペから画像を連想しにくい

- 各単語の正解率はコンテンツによって大きく異なっていた。
  - スタイル変換のクオリティはコンテンツ画像とスタイルの相性に強く依存していると考えられる。

Cluster 0

Cluster 1

Cluster 2

Vote: 2

Vote: 3

Vote: 5

Vote: 1

Vote: 0

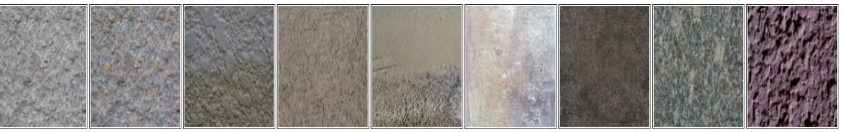
Vote: 2

Vote: 0

Vote: 2

Vote: 0

「ふさふさ」クラスタ



Cluster 0

Cluster 1

Cluster 2






Vote:3 Vote:6 Vote:4

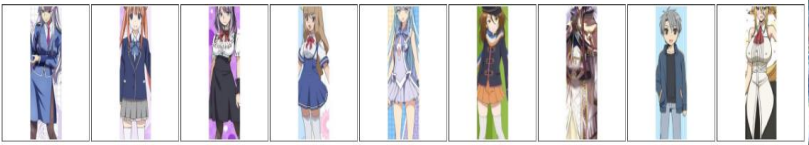


Vote:0 Vote:0 Vote:0



Vote:0 Vote:0 Vote:0

「ざらざら」クラスタ

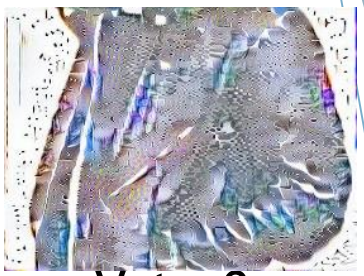


Cluster 0

Cluster 1

Cluster 2

Vote:0



Vote:0

Vote:0

Vote:1

Vote:1

Vote:1

Vote:4

Vote:4

Vote:7

「もふもふ」クラスタ

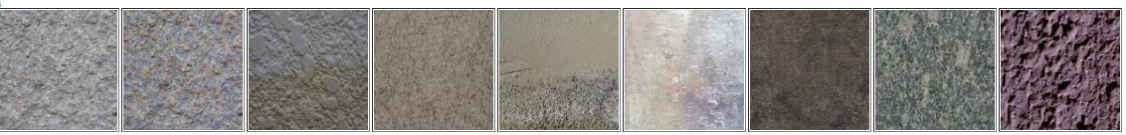
考察

コンテンツ画像と
スタイルクラスタの相性

Style

Content

ふさふさ cluster0



ざらざら cluster0

もふもふ cluster2

Vote:2 Vote:3 Vote:5

Vote:3 Vote:6 Vote:4

Vote:4 Vote:4 Vote:7

今後の課題

- 質感単語とコンテンツの関連度と印象操作への影響
- コンテンツの持つカラー傾向や局所的構造を考慮したスタイルクラスタの構築
  - コンテンツ画像をもとにしたスタイル画像のカラー変換
  - 前景領域の利用