

# 画像スタイル変換とWeb画像 を用いた画像の任意質感生成

松尾 真, 下田 和, 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

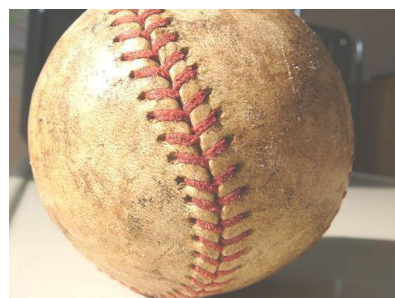


generated image

# 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



content  
image

「金属製」



「木製」



request

generated image

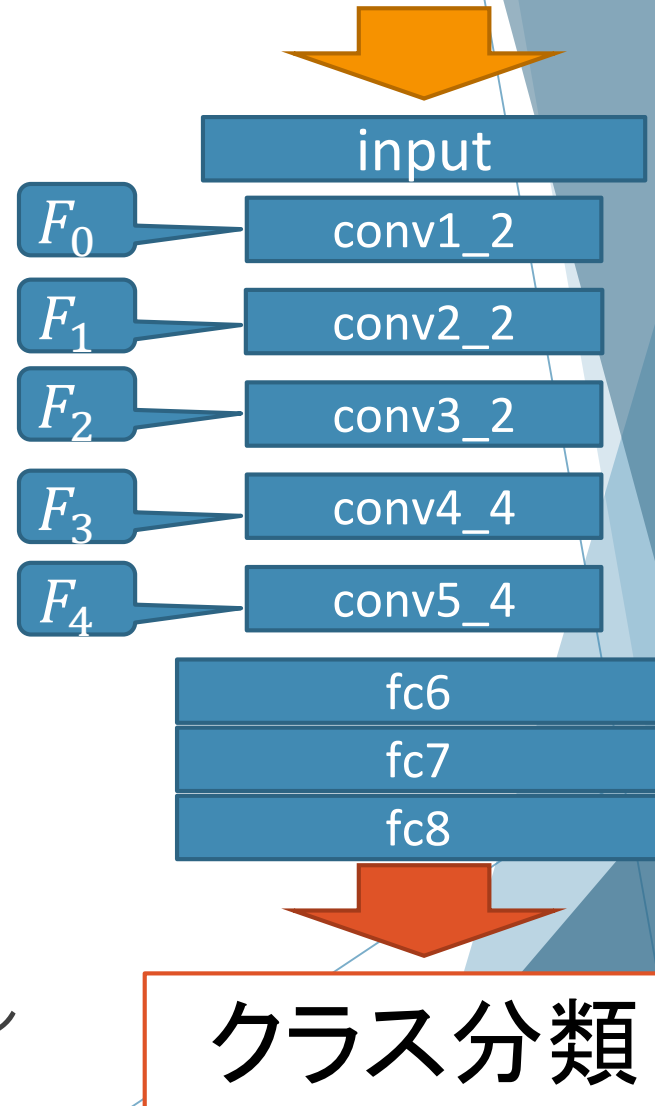
# VGG19 Network (Max pool層等省略)

## 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化値行列 $F_i$ が発生する



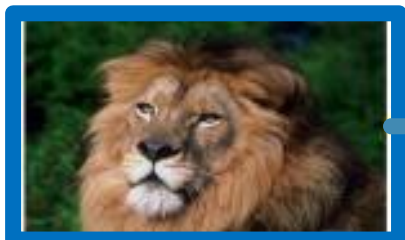
# Neural Style Transfer

## [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生成

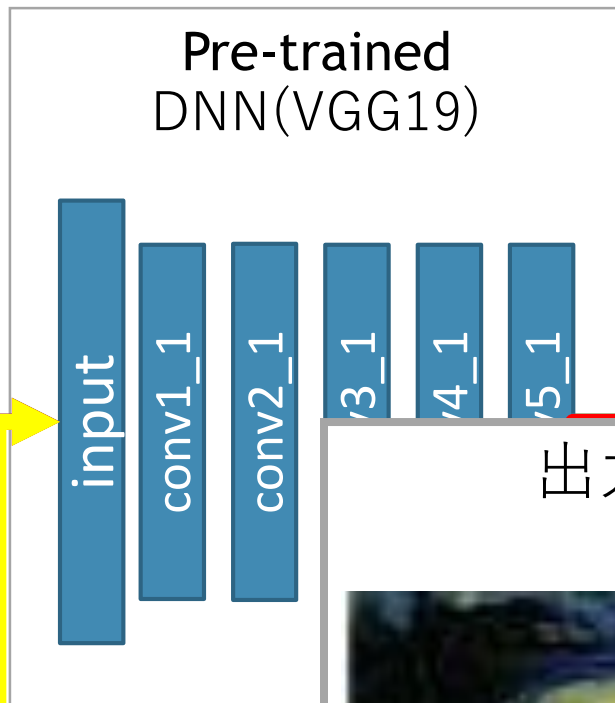
入力



コンテンツ画像  $x_c$



スタイル画像  $x_s$



$F(x_c)$

$F(x_g)$

出力 (最終状態)

出力 (初期状態)



生成画像  $x_g$



生成画像  $x_g$

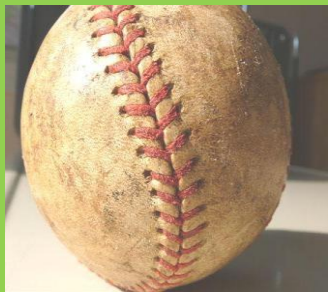
生成画像のコンテンツをコンテンツ画像

# 単語概念による画像内物体の任意 質感変換

入力

質感単語  
「ふさふさ」

コンテンツ画像



Web  
画像

スタイル表現



Neural Style  
Transfer



# オノマトペ→Web画像

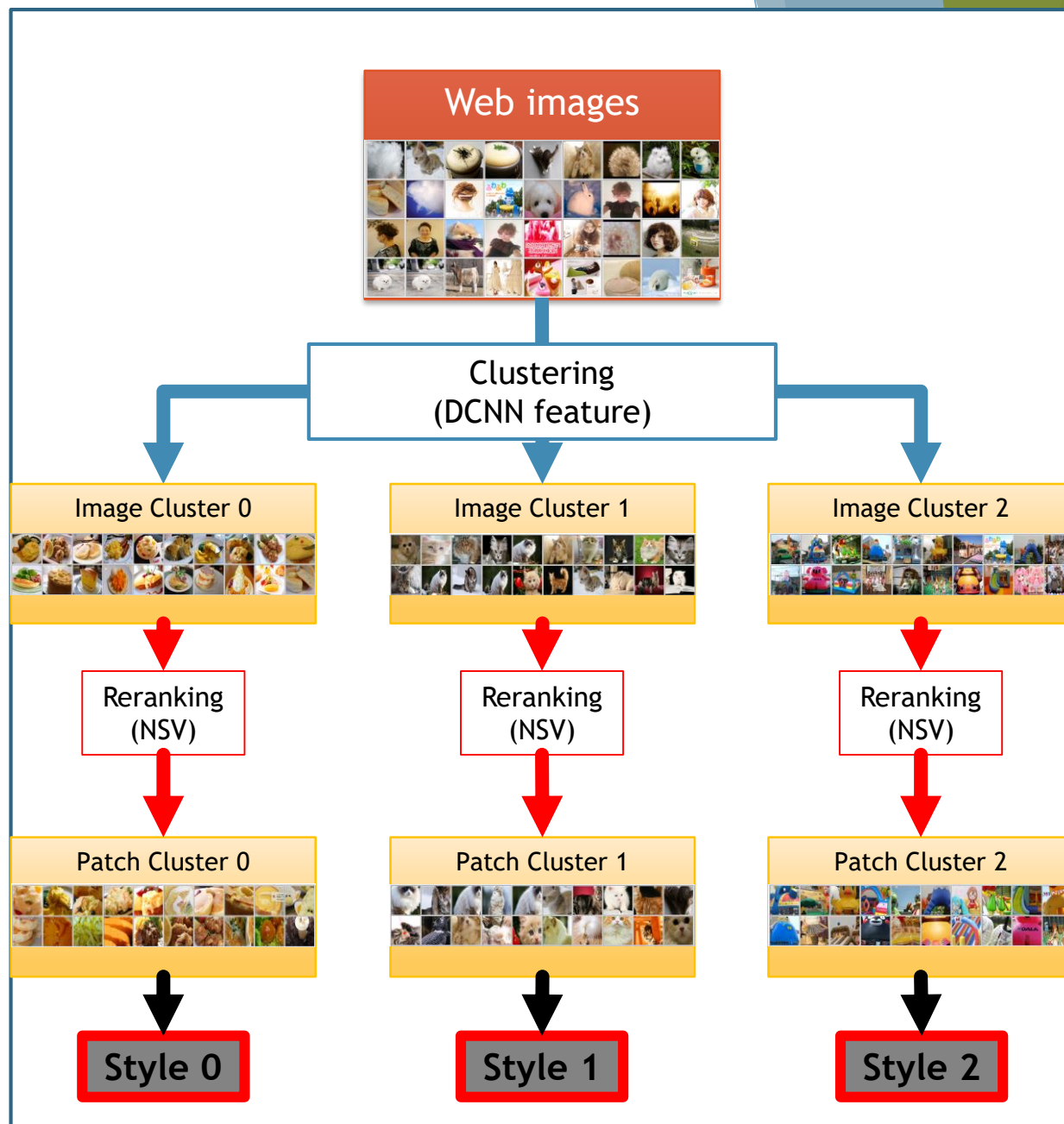
- ▶ 単語概念としてオノマトペを活用
- ▶ 複数の意味を持つオノマトペ（キャラクターの名前など）
  - ▶ クラスタリング
  - ▶ 候補を提示
- ▶ Web画像
  - ▶ Noiseが多く含まれる
  - ▶ リランキング
- ▶ スタイルの類似度を活用
  - ▶ クラスタリング、リランキング



# 単語による画像 収集とクラスタ リング

スタイルの類似度から  
クラスタに分化

- Neural Style Vectorを使用

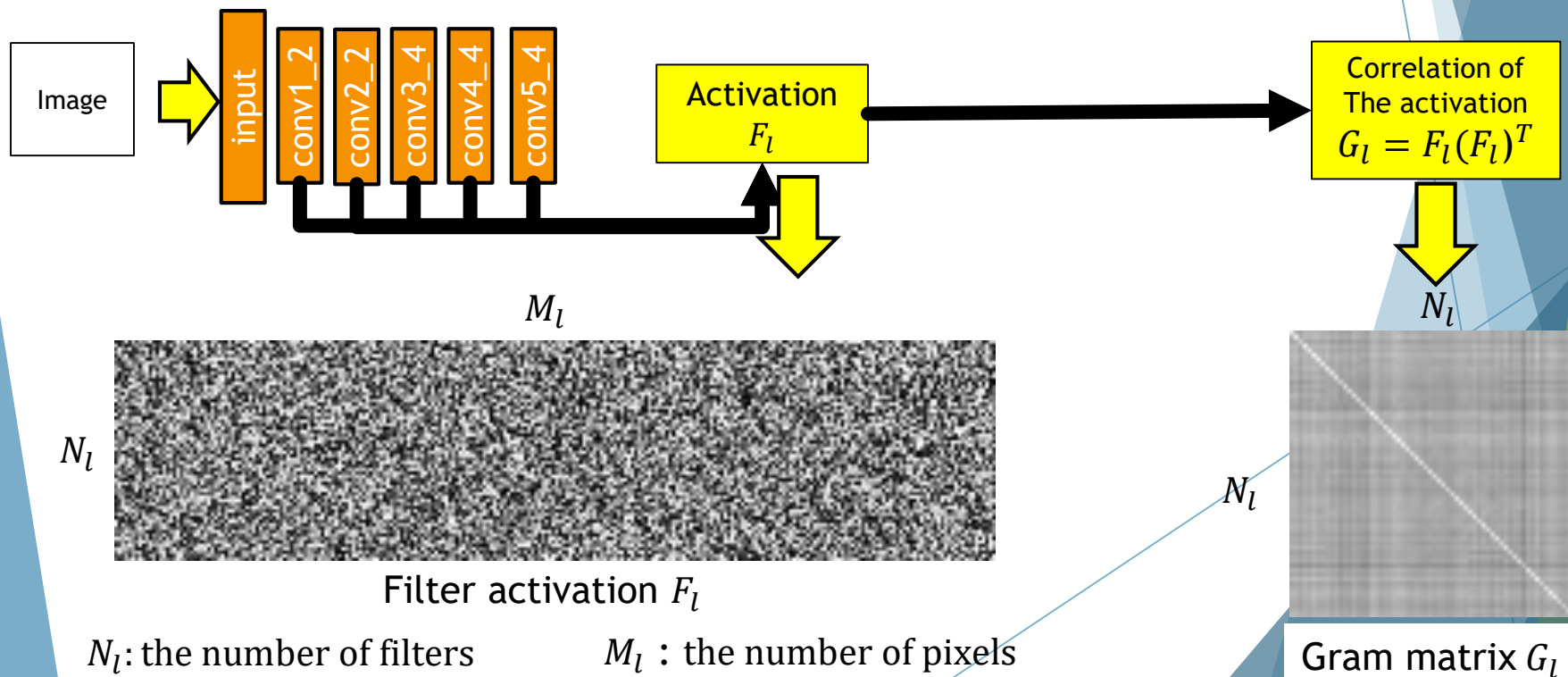


# Neural Style Vector

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
- ▶ スタイルの類似度から画像の検索・クラス分類に利用



# Neural Style Vector

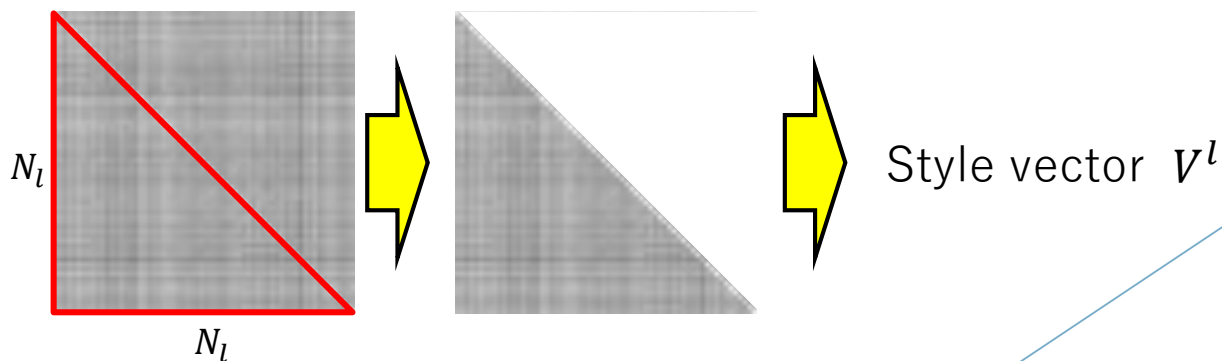
$G^l$ の対象要素と対角要素で定義

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

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

符号付平方根とL2正規化をして使用

PCAを用いて次元を圧縮



# 複数画像からスタイル表現へ

任意の単語に対応する画像集合

→スタイル表現への変換

▶ 代表画像一枚の選出

▶ ×

▶ 複数画像について平均をとる

▶ ×

▶ 複数の画像を結合



# 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

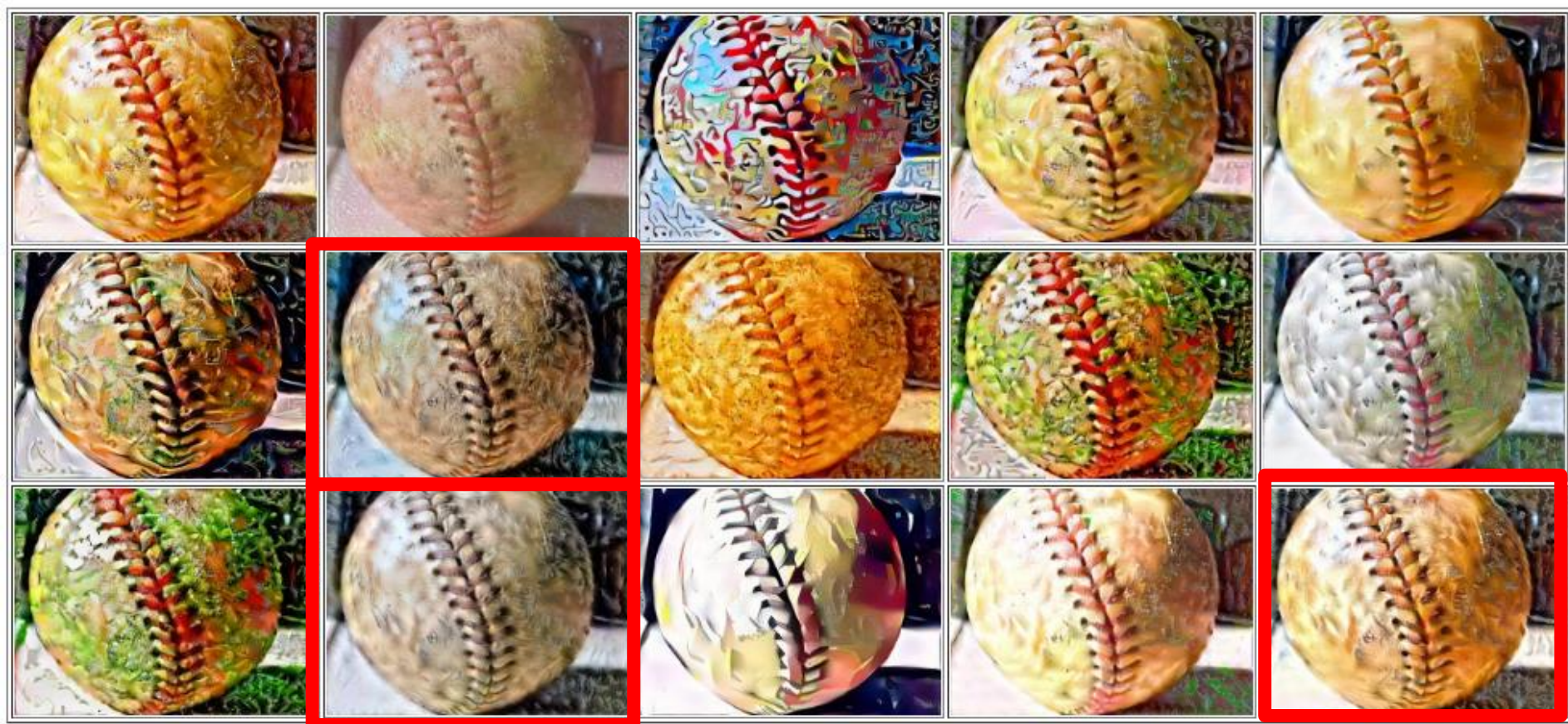
## 評価方法

- ▶ 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

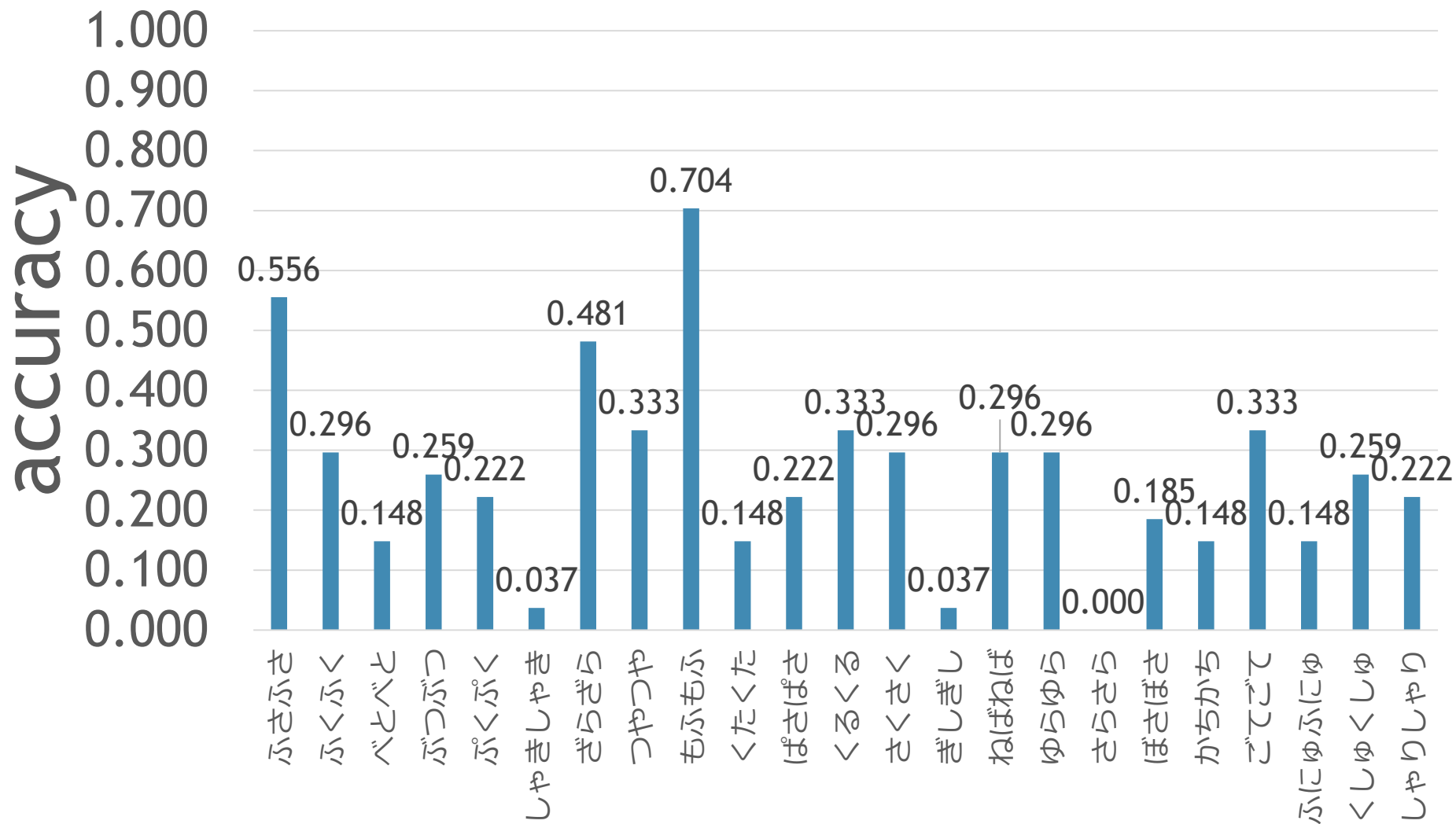
# ユーザー評価

▶ 赤枠が正解

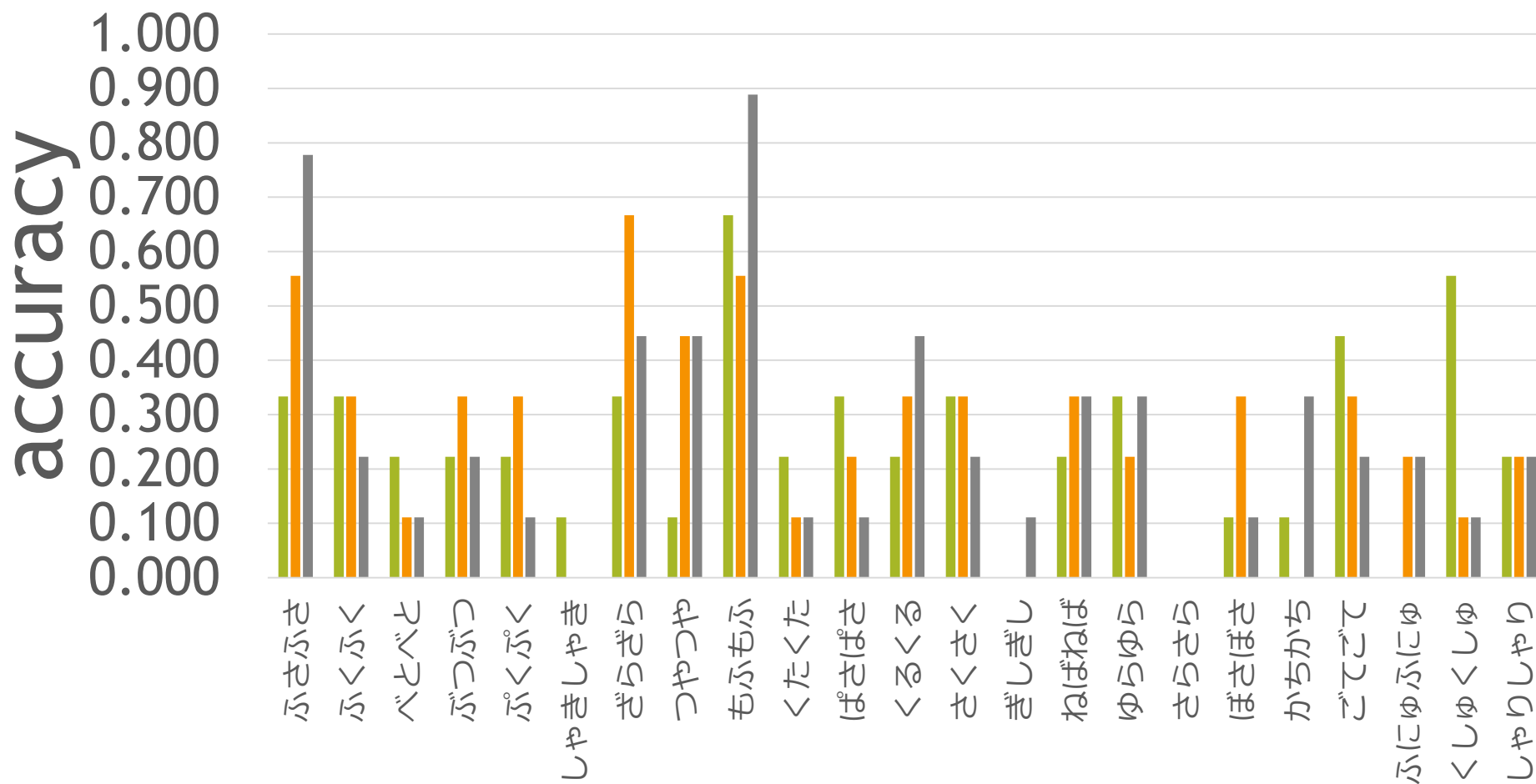
最も「ふさふさ」な画像を選んでください。0/72



# 実験結果 (総合)



# 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: 1



Vote: 0



Vote: 3



Vote: 0



Vote: 2



Vote: 5



Vote: 2



Vote: 0

# 「ふさふさ」 クラスタ



Cluster 0



Cluster 1



Cluster 2



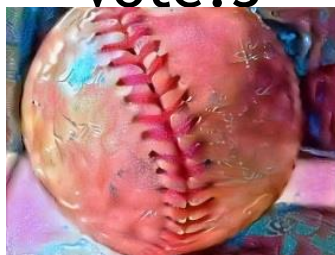
Vote:3



Vote:6



Vote:4



Vote:0



Vote:0



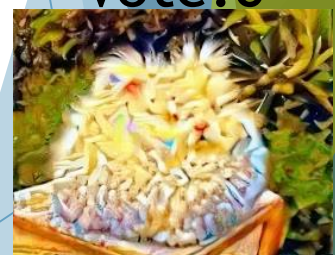
Vote:0



Vote:0



Vote:0



Vote:0

「ざらざら」 クラスタ



Cluster 0



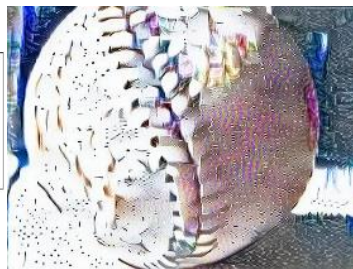
Vote:0



Vote:0



Vote:0



Vote:0



Vote:0



Vote:0



Cluster 1



Vote:1



Vote:1



Vote:1



Cluster 2



Vote:4



Vote:4



Vote:7

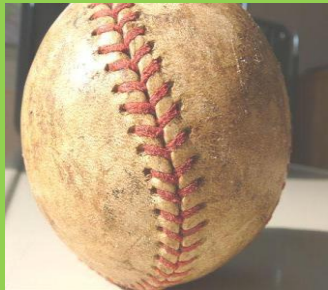
# 「もふもふ」 クラスタ

# 単語概念による画像内物体の任意 質感変換

入力

質感単語  
「ふさふさ」

コンテンツ画像



Web  
画像

スタイル表現



Neural Style  
Transfer



一度、CNNで画像生成モデル  
を学習して、スタイル表現を  
生成する方法も実験中。  
(DCGAN, Generative  
Adversarial Net)

# 結果

- ▶ オノマトペ「ふさふさ」について、Web画像を用いて直接/DCGANで生成したスタイル画像を適用



単純結合



DCGAN

# 今後の課題

- ▶ より洗練された単語からのスタイル表現の獲得
- ▶ スタイルとコンテンツの関連度と印象操作への影響
  - ▶ 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