画像内容を考慮した質感表現に基づく画像変換

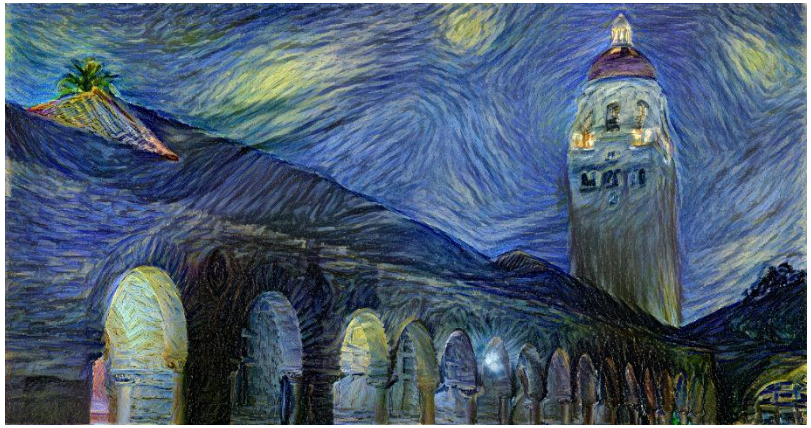
杉山 優　　柳井啓司

電気通信大学大学院

情報理工学研究科 情報学専攻

はじめに

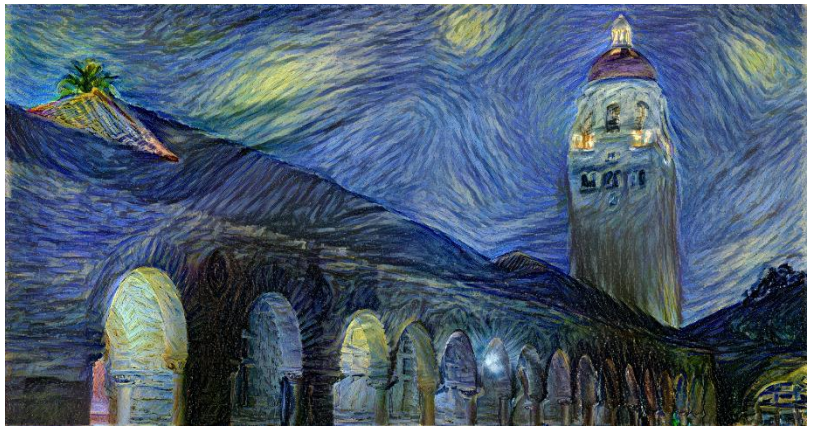
- Neural Style Transferで画像の変換が可能



はじめに

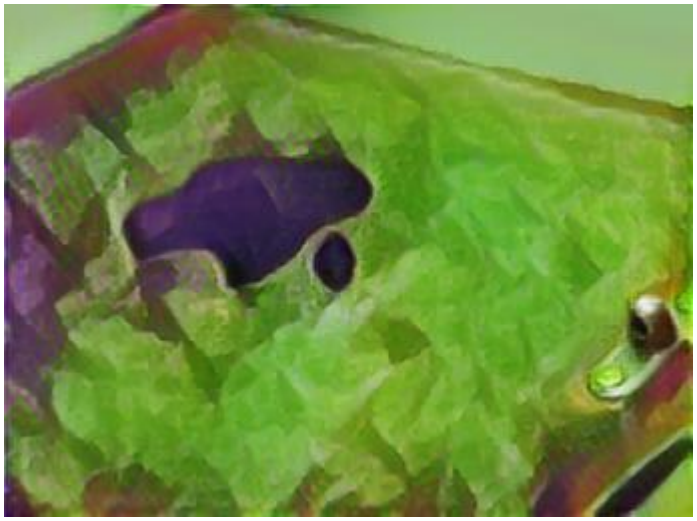
- スタイル変換では、コンテンツ画像とスタイル画像が必要

→発展させて言葉によって変換したい



目的

- 言葉を入力のひとつとすることでより手軽で汎用性の高い画像変換

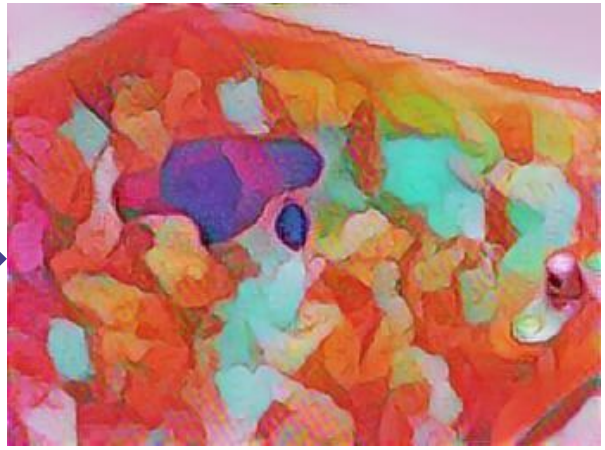


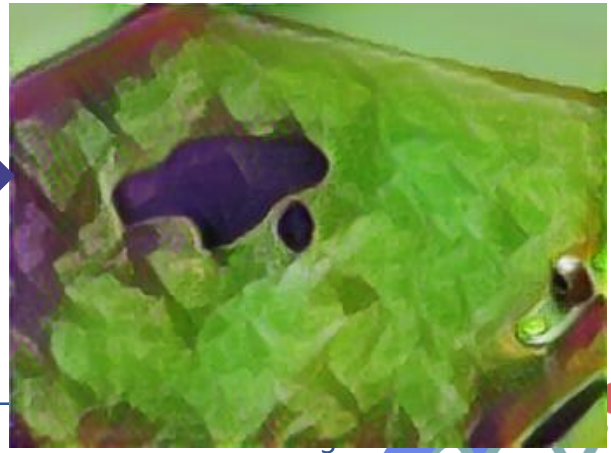
「枝葉」
という言葉

目的

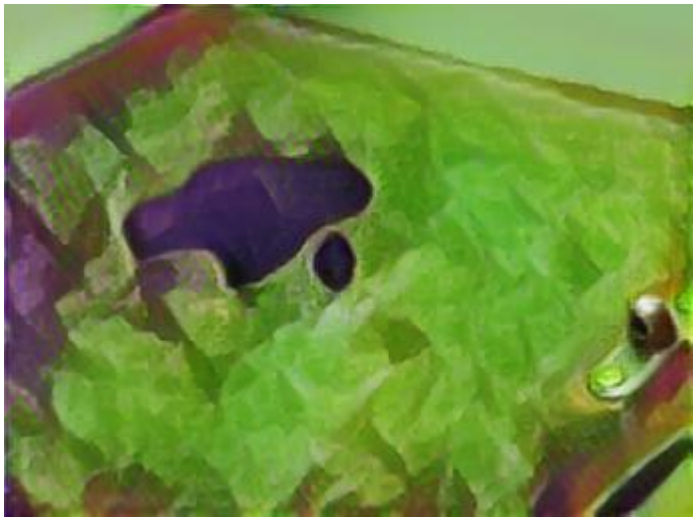
- 「枝葉」だけでは「緑の葉」
「紅葉した葉」の
区別は不可能

「枝葉」
という単語





目的



「枝葉」という
入力単語
＋
「革」という
画像認識

適した
スタイル画像

先行研究

1. A Neural Algorithm of Artistic Style

   Leon A. Gatys, CVPR, 2016

   スタイル画像とコンテンツ画像を混ぜ合わせるスタイル変換を発表

2. Perceptual Losses for Real-Time Style Transfer and Super-Resolution

   Jastin Johnson, ECCV, 2016

   事前にスタイルを学習することで高速なスタイル変換を実現

先行研究

3.　Unseen Style Transfer

Keiji Yanai, ICLR WS, 2017

高速スタイル変換を改良
スタイル画像を学習に使用しなかったものからも
選択できるように改良した

4. Arbitrary Style Transfer

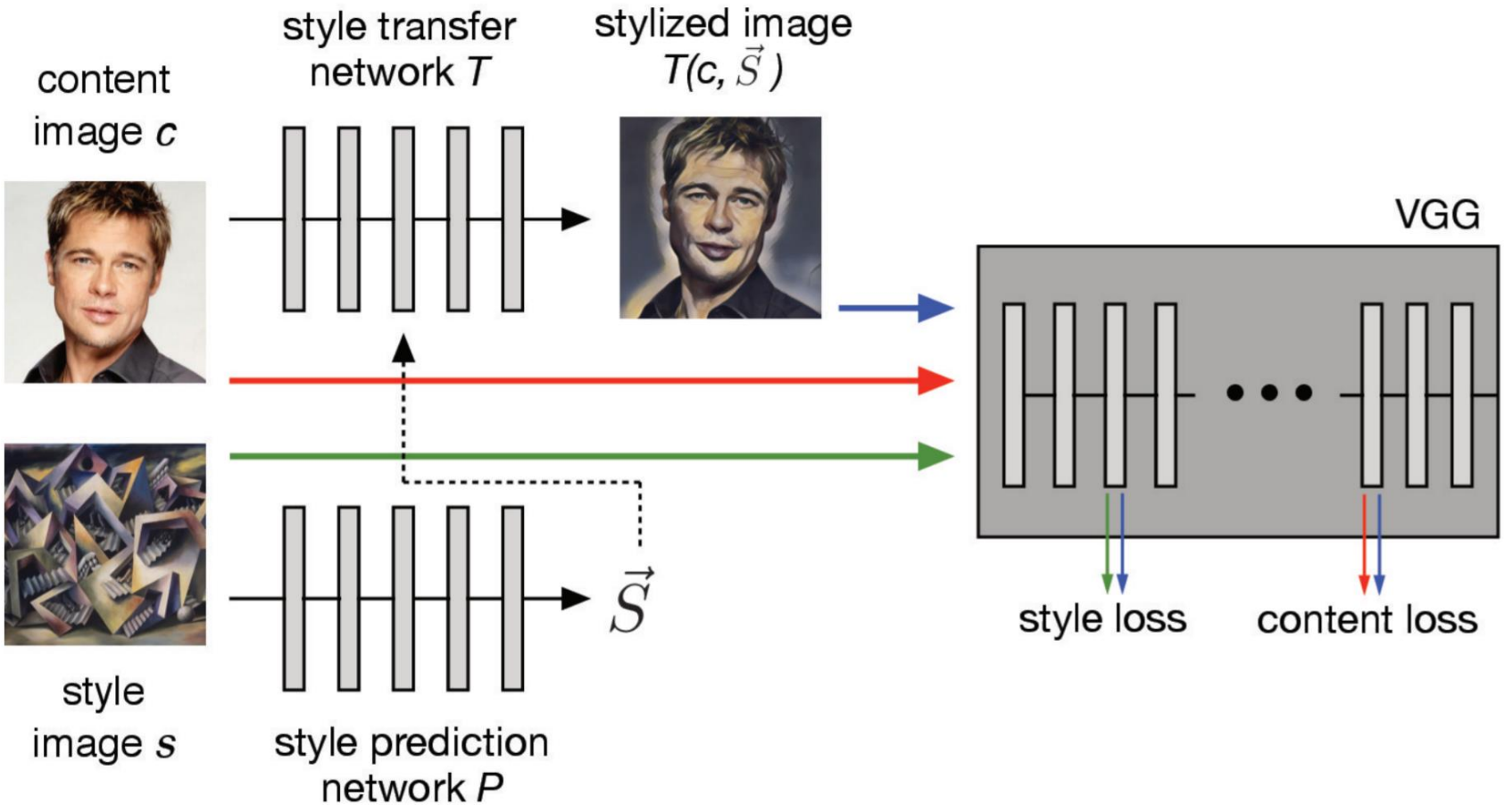
Golnaz Ghiasi, Honglak Lee, et al. BMVC, 2017

任意スタイル変換の別手法

Conditional Instance Normalizationを利用してきれいに画像を生成した

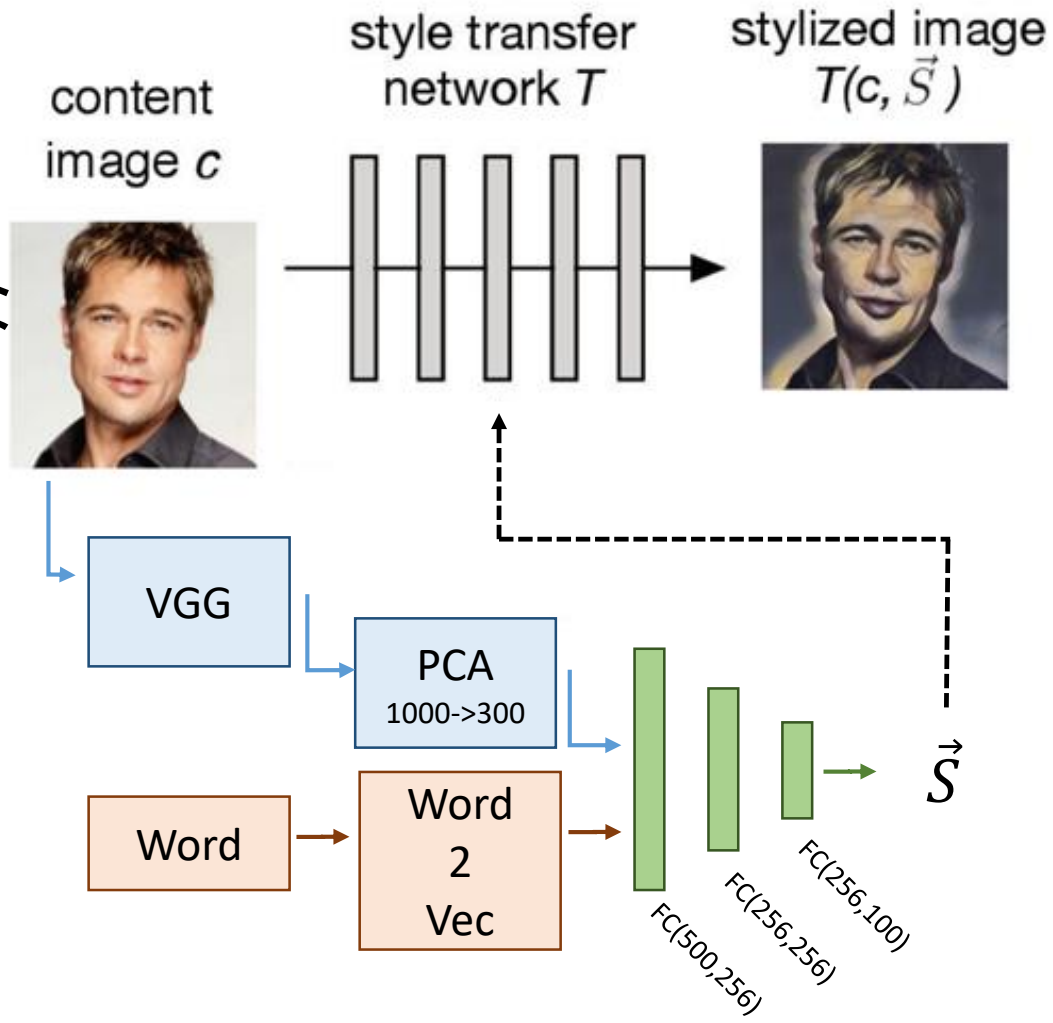
手法

- Arbitrary Style Transferのネットワーク



提案手法

- VGGを使用した画像意味特徴からコンテンツに適切な単語スタイルを学習



手法

- 入力単語のベクトル化

  -> Word2Vecを使用する

  (200次元)

- 入力コンテンツ画像のベクトル化

  -> VGG画像認識の中間出力(Fc8層)を使用する

  (1000次元)

  -> PCAで次元削減する

  (300次元)

Word2Vecの学習

- 学習データには日本語Wikipediaを使用
- Mecabを使用して形態素解析を行った
- 助詞などの頻出単語は合計出現数1000回になるようにランダムに削除

手法

- 単語から得たベクトルと画像認識から得たベクトルによって任意スタイル変換と同等のネットワークを作成

Inception v3

Style Vector

近似させる

VGG 画像 認識

画像 特徴 ベクトル

Style Select Network

Style Vector

``革''

Word 2 Vec

単語 ベクトル

データセットの構築

- UFMDを使用
- 学習には画像とワードのペアが必要になるため、UFMDの画像とそのラベルによって学習した
- ラベルは10種類

布, 植物, ガラス, 革, 金属, 紙, プラスチック, 石, 水, 木

- 各ラベル1,000枚、合計10,000枚

データセットの構築

革

布

石

水

木

植物

プラスチック

紙

ガラス

金属

実験

- 制作したデータセットでネットワークを学習

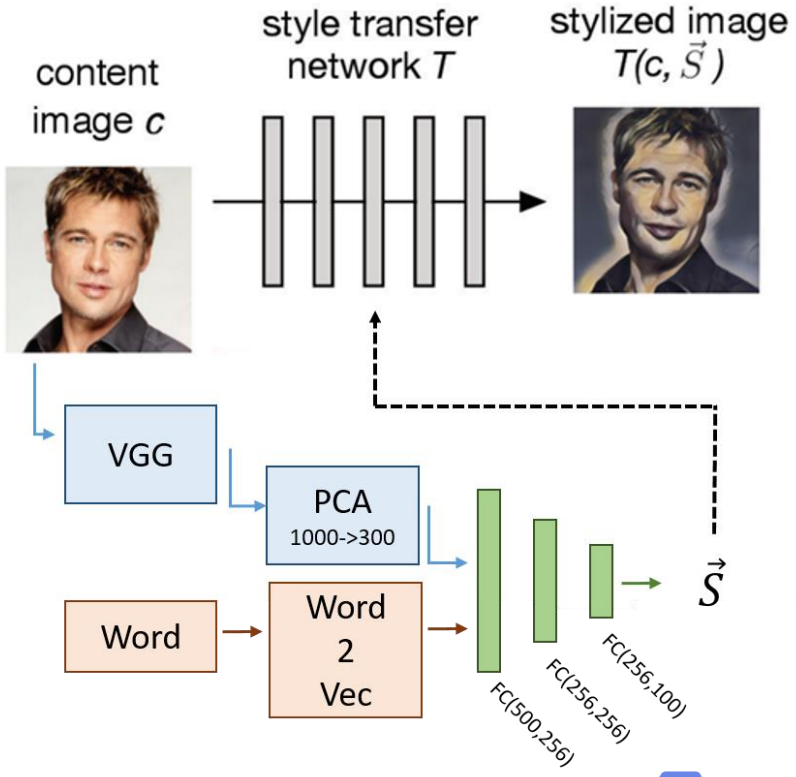
- ネットワークの実験として2項目を観察

1. コンテンツ画像の内容によっての
   変換の様子の変化

2. 入力単語の内容によっての
   変換の様子の変化

学習の条件

- 最適化関数にはAdamを利用
- 50エポック、バッチサイズは50で学習
- 生成ネットワークは固定
- FC層3層のみの学習なので
学習には50秒程度



実験

元画像

ガラス

革

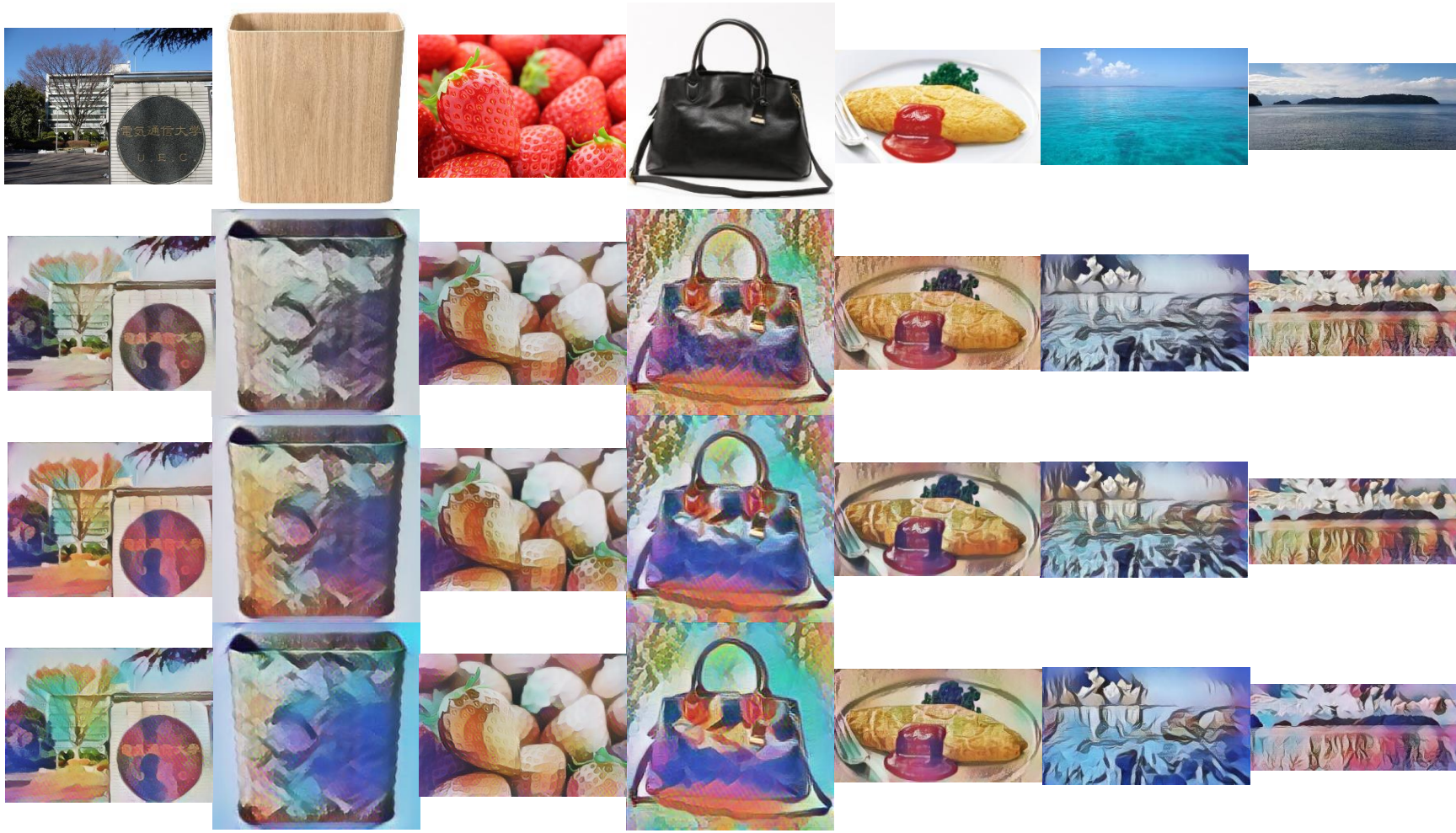
実験

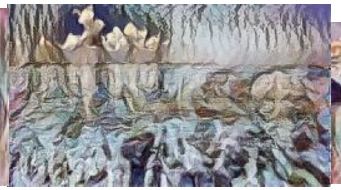
元画像

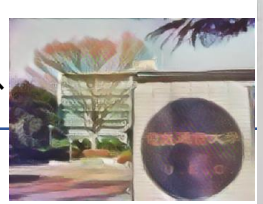
プラスチック

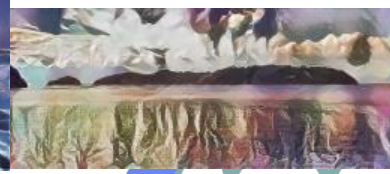
石

実験









実験

問題点

- 建物などの風景にはうまく対応できていない
- 一部画像入力によって変換が代わり映えしない学習がある
- 学習しなかった単語については意味ベクトルと画像特徴が一致するとは限らない

考察

- 目的である「単語による画像変換」と「コンテンツ画像に適切な変換」の2つはある程度達成できた

- 物体画像では上手く変換できるが、シーン画像では上手く変換できない

  -> ネットワークに物体画像を使用したから

- 上手く変換できる単語は視覚的特徴を示す単語

  -> データセットの質が高くなりやすいから

まとめと今後の予定

まとめ

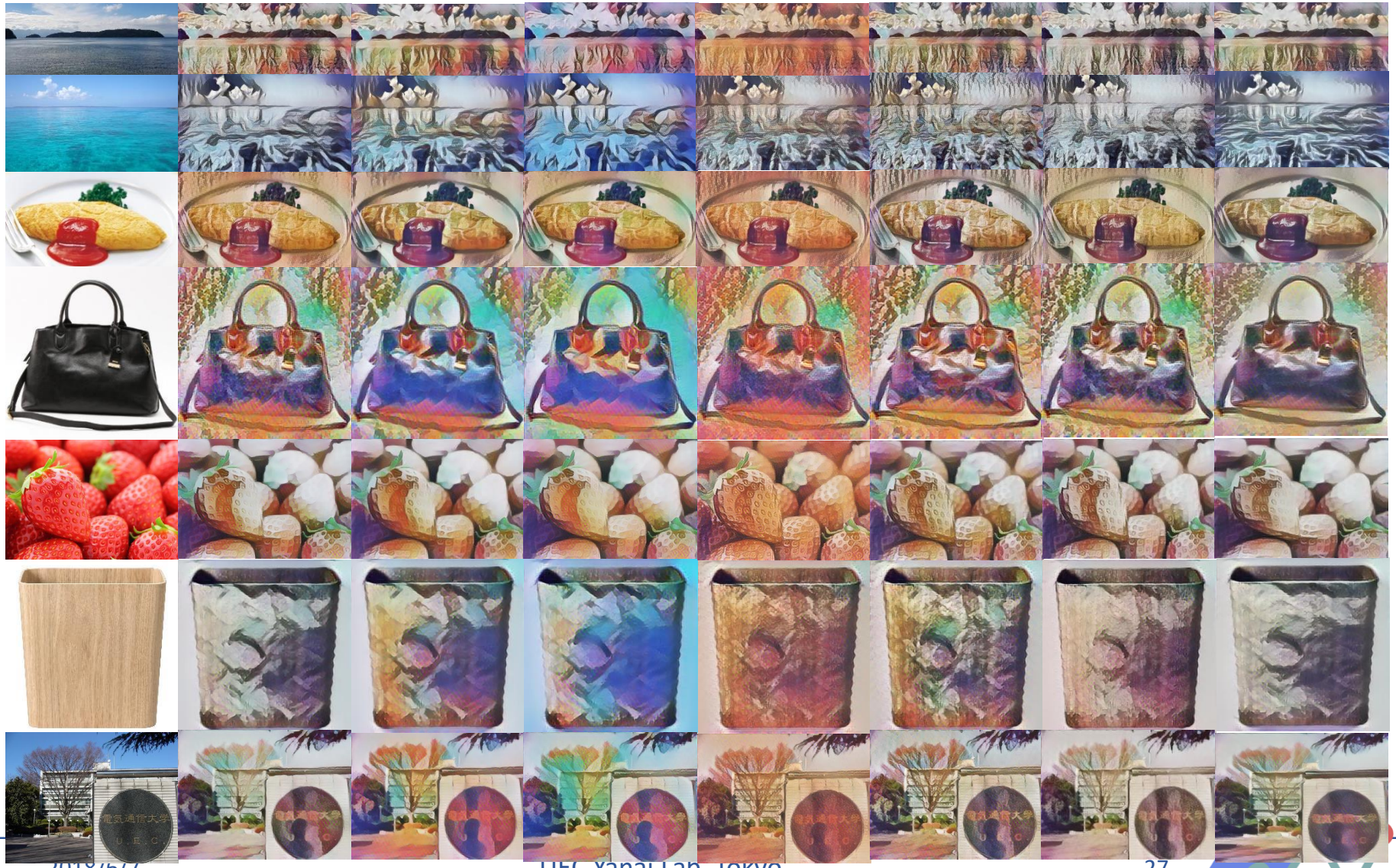
- 言葉で指定したスタイルに画像変換が行えた
- 学習に使用しなかった単語にも対応できた

今後の予定

- 画像認識とシーン認識を合わせて利用する
- データセットに対応した単語を増やすことでより精度を向上させる


UEC
TOKYO


元画像
ガラス
プラスチック
革
石
植物
紙


元画像
ガラス
アクリル
プラスチック
革
牛皮
石
レンガ

手法

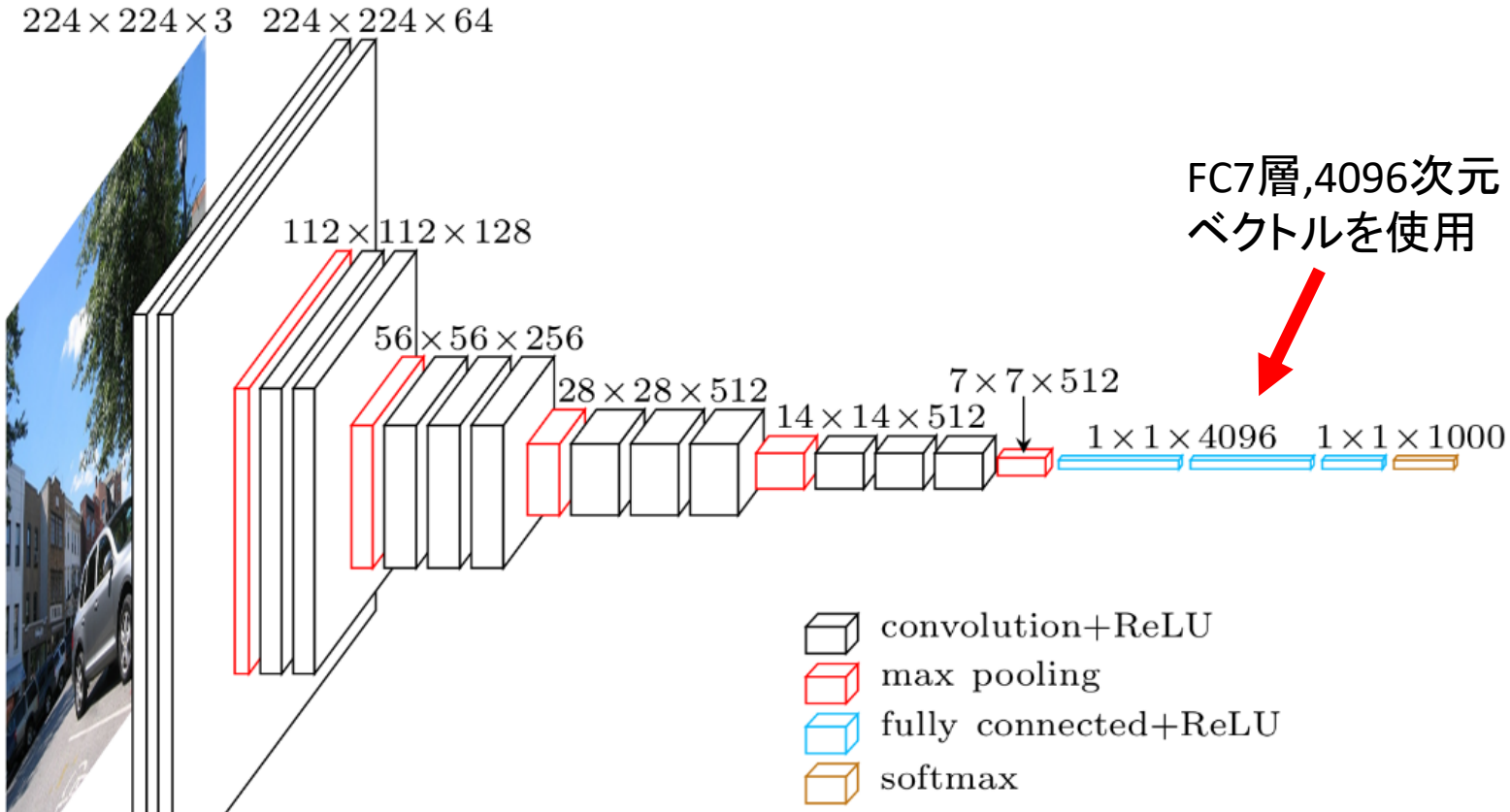
- 言葉をベクトルによって管理するためにWord2Vecを使用する

端末

```
sugiya-y@gp15> python test2.py
pos>> 魚
v1.shape:
(500,)
v1:
[ -1.65409409e-03   6.16507716e-02   4.18421300e-03  -1.84675641e-02
   1.33760497e-02  -2.22606841e-03   8.93203542e-02  -2.14479417e-02
   3.52570340e-02  -1.03908237e-02  -3.84614468e-02   5.67243733e-02
  -7.19319955e-02  -5.98911988e-03  -1.24127371e-02   4.85736765e-02
   6.33691177e-02   7.05734715e-02   2.34751888e-02  -4.74605225e-02
  -1.79290511e-02   1.35685671e-02  -5.31266779e-02  -4.99152355e-02
   1.15473811e-02   1.01831049e-01  -3.31960842e-02  -2.48594191e-02
  -3.51100974e-02  -5.92484288e-02   6.52251840e-02   7.77722448e-02
   2.22318303e-02  -9.80505347e-02  -5.36268838e-02   8.61750990e-02
  -3.28920297e-02   8.82115662e-02  -2.43228450e-02  -5.96812256e-02
   5.67847453e-02  -1.76143311e-02  -8.21785536e-03   3.96078154e-02
   1.75542012e-02   5.36593720e-02  -2.54095227e-01   3.73274386e-02
   1.00360982e-01   1.30009567e-02  -1.07335240e-01   2.04007365e-02
   1.48245329e-02   4.68749702e-02  -2.41931267e-02   3.68897878e-02
  -3.00001563e-03   1.56983770e-02  -1.66391395e-02   3.37712094e-02
  -8.84477198e-02   1.53843844e-02   1.86057705e-02  -8.53718165e-03
  -3.75012867e-02  -1.43703977e-02  -7.60126784e-02   2.07564328e-02
   1.40071847e-03  -3.50439064e-02  -2.10920023e-03   2.64538676e-02
```

手法

- 画像特徴を得るためにVGGネットワークを使用する



まとめ

- 単語とコンテンツ画像の内容によってスタイル変換を行うことができた

- 任意スタイル変換の応用によって入力をユーザーにとってわかりやすい形に変更することができた