

単語情報を利用した画像の質感転送

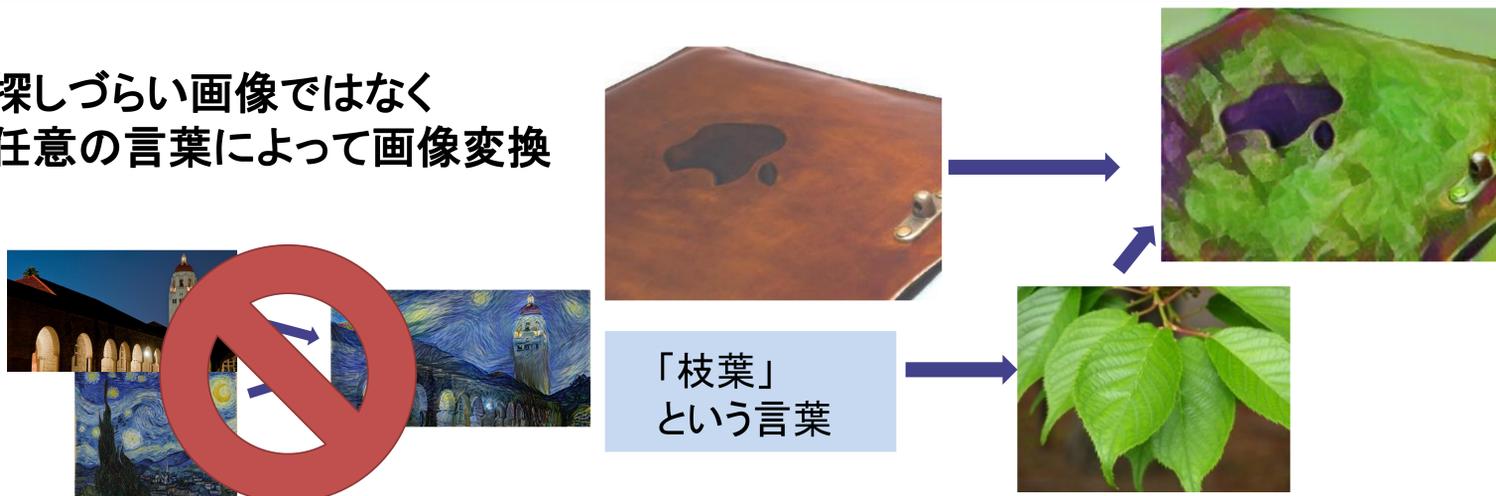
電気通信大学 大学院 情報理工学研究科 情報学専攻 杉山 優, 柳井 啓司

1. 目的

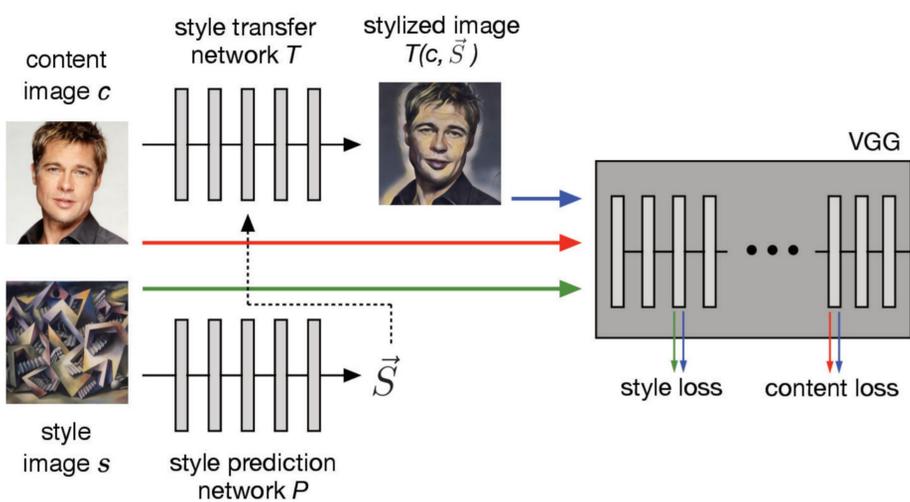
単語概念による
画像の任意スタイル変換

探しづらい画像ではなく
任意の言葉によって画像変換

既存のスタイル変換では
画像2枚から変換を行う
⇒スタイル画像を探すのは
結構大変



2. Arbitrary Style transfer[1]アルゴリズム

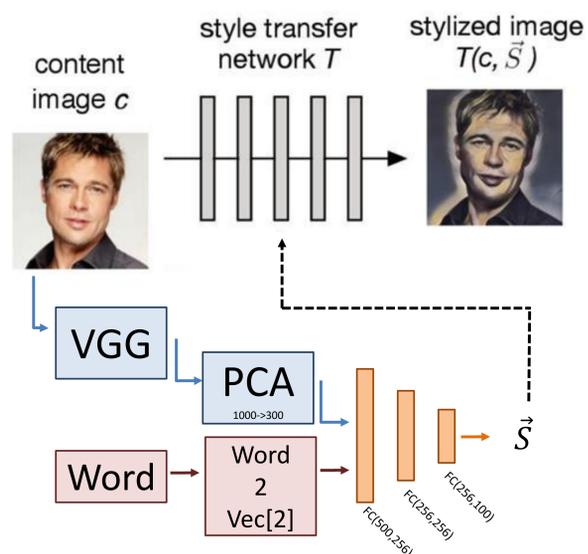


Style Prediction Networkでスタイル特徴 \vec{S} を計算し、生成ネットワークに、

- どんな画像が入力されても **style loss** と **content loss** が小さくなるように
- 任意の画像入力に対応したスタイル変換が可能

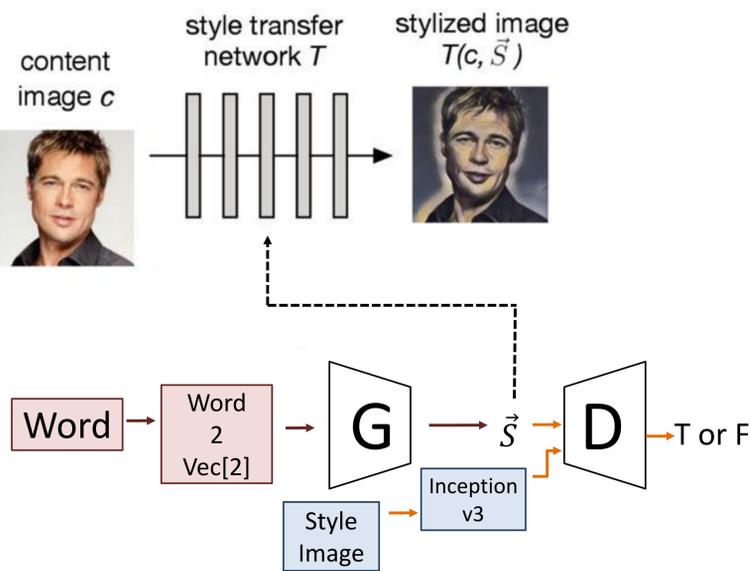
3. Style Prediction Network

Wordからスタイル特徴を生成するネットワーク



■ VGG画像特徴を入力して同じwordでもベクトルの多様性をもたせる。
あり/なしの場合を実験で比較した

4. Training with GAN[3]



→は学習時のみ通すネットワーク

6. 今後の課題

- データセットをより高品質なものにする
各タグについて類似した画像のみを抽出して学習を行うように変更する
- 各タグの代表画像を選択して画像変換を行うように変更を行う
単一画像での変換のほうがきれいに生成できるため
- より変換に適した単語の選出

5. 実験

3. の手法をUFMDで学習した

(10カテゴリ各1,000枚, 合計10,000枚): [Fabric(布)] [Foliage(植物)] [Glass(ガラス)] [Leather(革)] [Metal(金属)] [Paper(紙)] [Plastic(プラスチック)] [Stone(石)] [Water(水)] [Wood(木)]

入力画像	布	植物	ガラス	革	金属	紙	プラスチック	石	水	木	学習しなかった単語
											アクリル 牛皮 レンガ アルミニウム
											VGGあり/なし比較
											なし あり

【参考結果】単一スタイルの転送

4. の手法をYahoo100Mで学習した

(500カテゴリ各1,000枚, 合計500,000枚): Yahoo100Mデータセットからジオタグ付き画像のタグをその画像の形容詞としてデータセットを構成。各タグにつきランダムに画像を選択。

元画像	old	new	urban	country	indoor

考察

- L2 normで学習する3. の手法では、どんな言葉を入力しても変換が大きく変化することはなかった
 - GANを用いた4. の手法ではある程度の言葉の特徴を捉えた変換を行うことができた
- 学習画像集合の重心で変換してしまう学習画像の分布に合わせてスタイルベクトルを生成することが重要

[1] Exploring the structure of a real-time, arbitrary neural artistic stylization network, BMVC, 2017.
 [2] Distributed representations of words and phrases and their compositionality, NIPS, 2013.
 [3] Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, ICLR 2016.