Word-Conditioned Image Style Transfer

Yu Sugiyama and Keiji Yanai

The University of Electro-Communications, Tokyo

Introduction

- Neural Style Transfer, Image style transfer

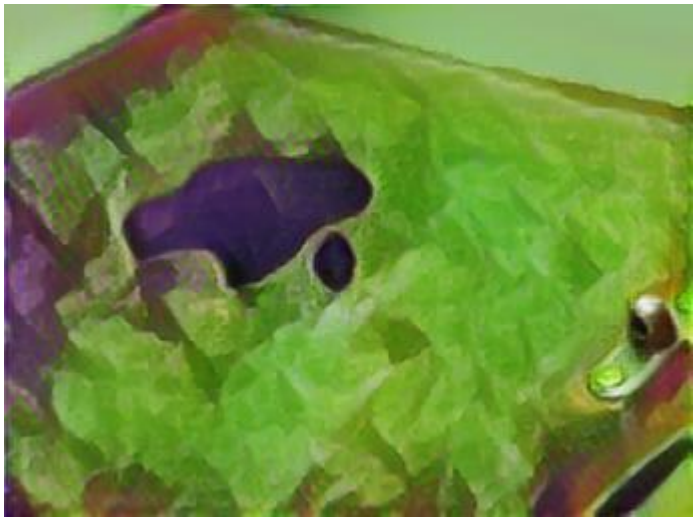
Introduction

- Style Transfer requires CONTENT and STYLE image

  → User need to find good image

Objective

- Words condition makes easy to find good style



"Leaves”

Related Work

1. A Neural Algorithm of Artistic Style

   Leon A. Gatys, CVPR, 2016

   Image style synthesis for artistic style

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

   Jastin Johnson, ECCV, 2016

   Pre-training stylization network to fast image transfer

Related Work

3. Unseen Style Transfer

Keiji Yanai, ICLR WS, 2017

improvement of fast style transfer

Stylize image with un-trained images

4. Arbitrary Style Transfer

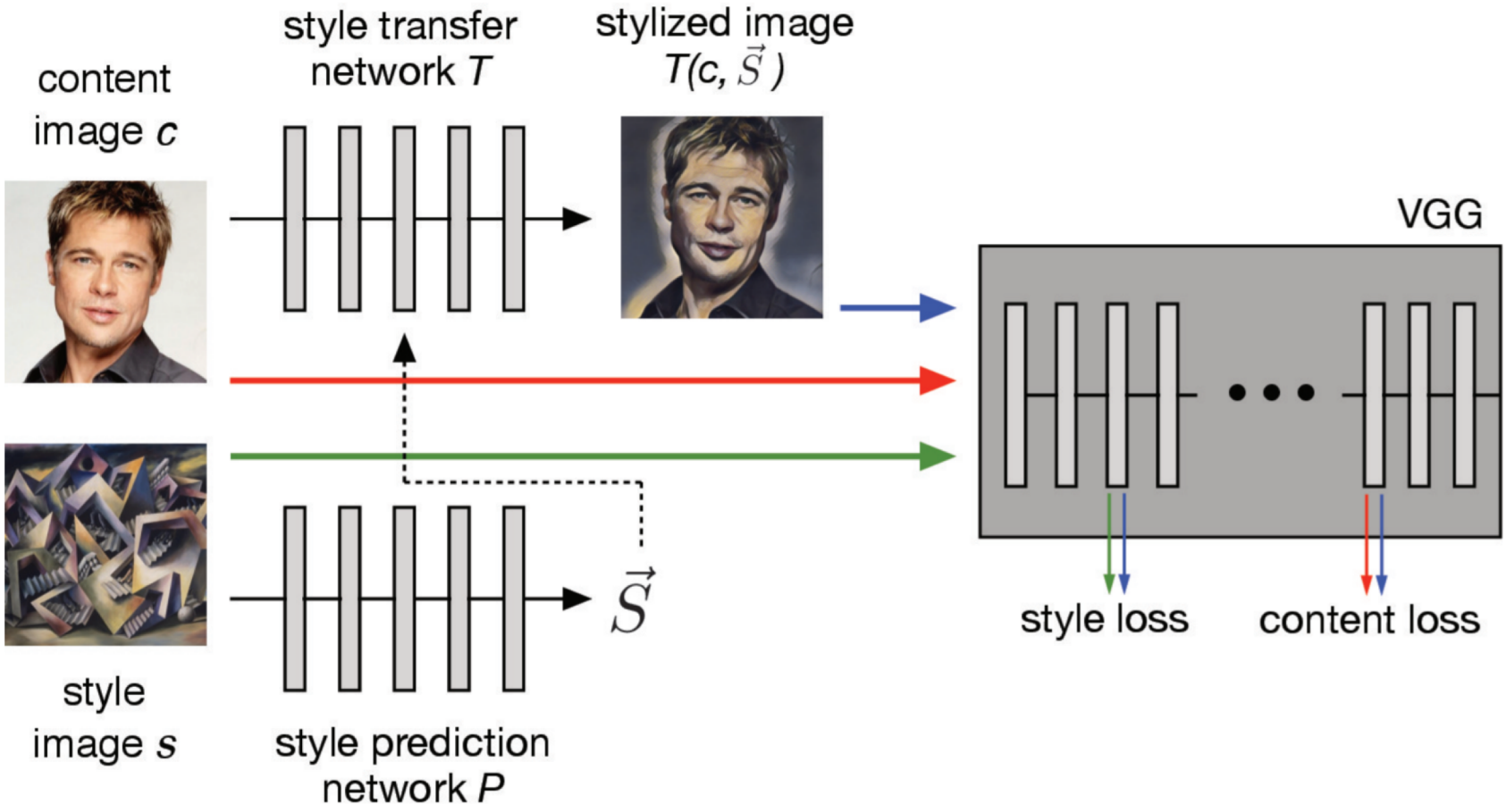
Golnaz Ghiasi, Honglak Lee, et al. BMVC, 2017

Different method of Unseen Style Transfer

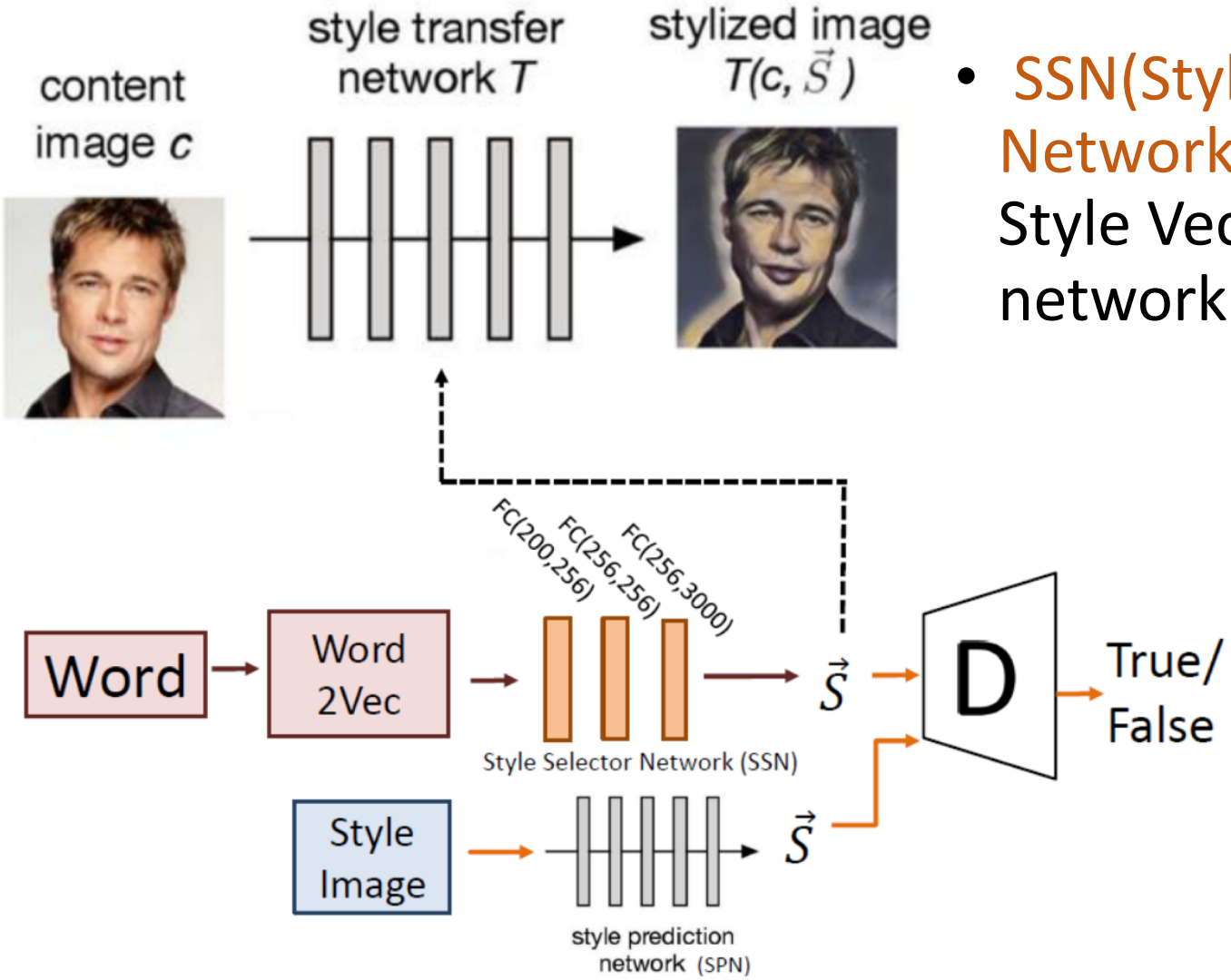
Synthesis images with Conditional Instance Normalization

Method

- Arbitrary Style Transfer Network



Method



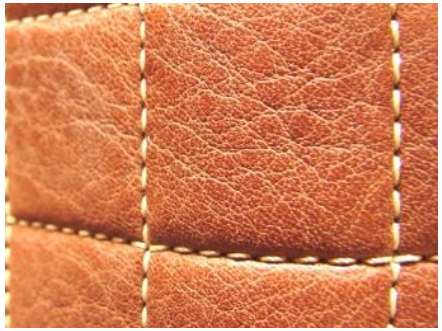
- SSN(Style Selector Network) predicts Style Vector for transfer network

Train Word2Vec

- Generate 200-dim vector from words
- English Wikipedia Corpus
- Reduce row corpus
  - Remove all Stopping words, low frequency word(under 5 times)
  - Random remove High frequency word(over 1000 times)

Method

- SSN makes Style vector same as Inception-v3 feature extractor



Inception v3 (Style Precision Network)

Style Vector

Approximate

Style Vector

``Leather''

Word 2 Vec

Embedding vector

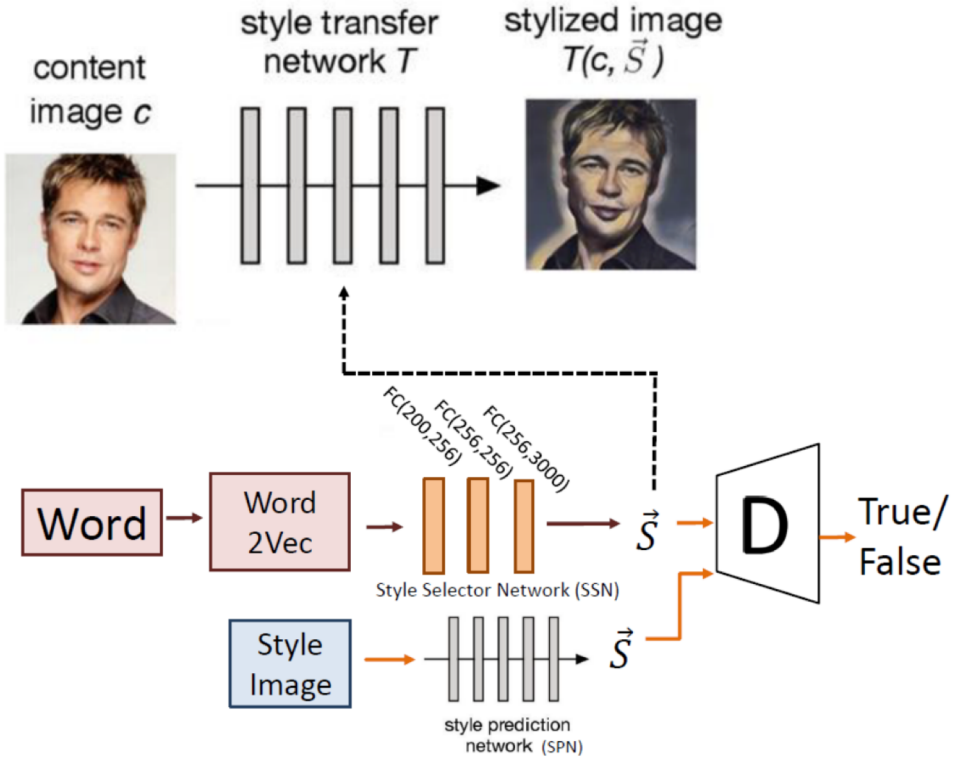
Style Selector Network

Train Dataset

- Stands by Yahoo100M wild images data
- Random select 1000 images for 500 categories
- Adverbs tag annotation
- Contains 500,000 image-word pairs

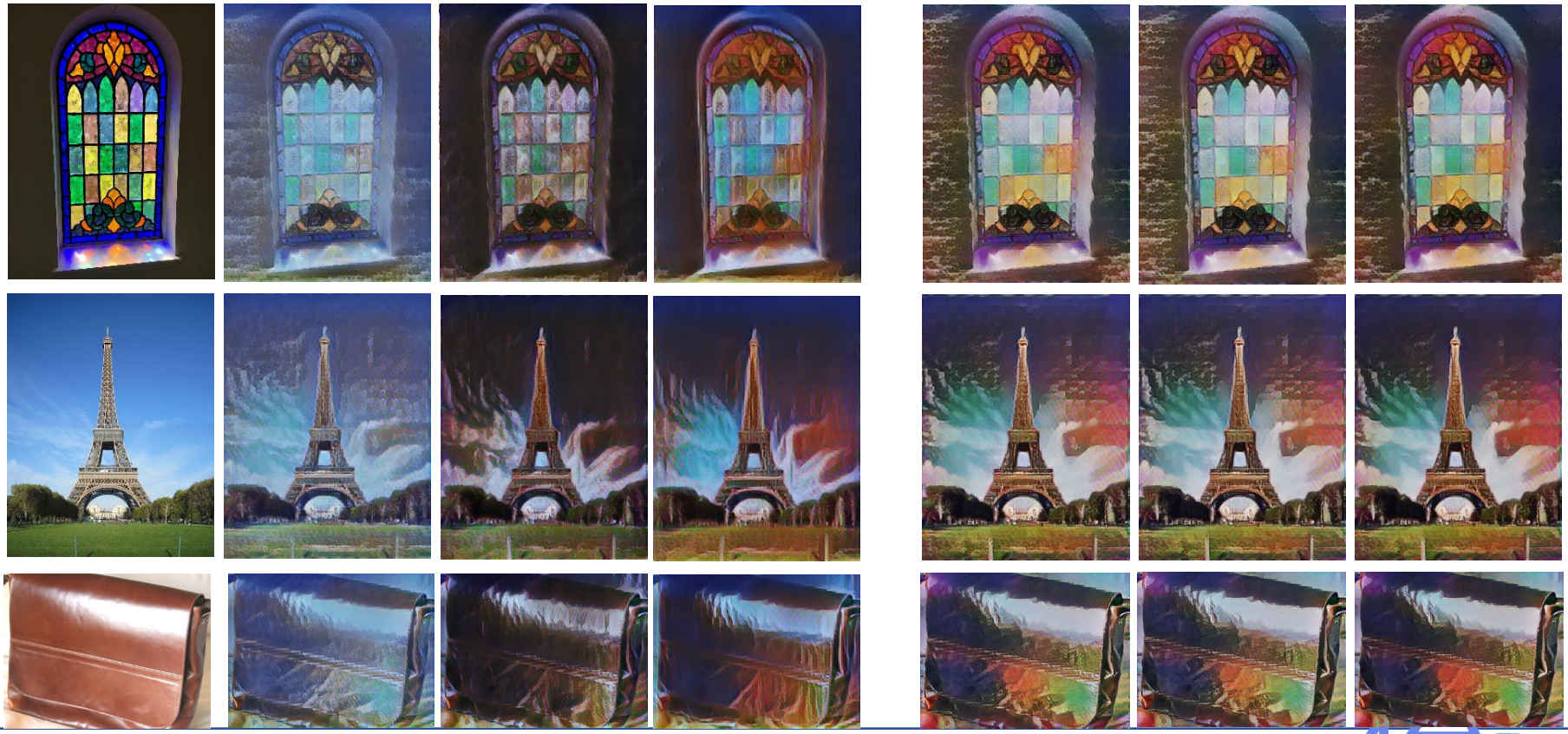
Experiments - Detail

- Optimize with Adam
- Style transfer network and style prediction network are pre-trained
- Train Only 3 FC layers
- Training takes 10 min



Experiments

- L2 minimalize instead of adversarial loss



Experiments

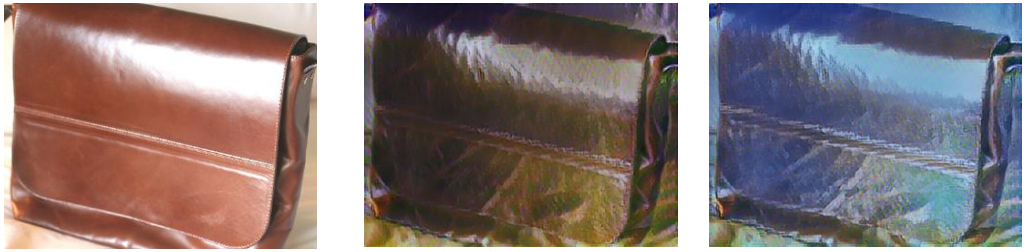
- There are some mismatches between content images and style words

Mismatch

input old new

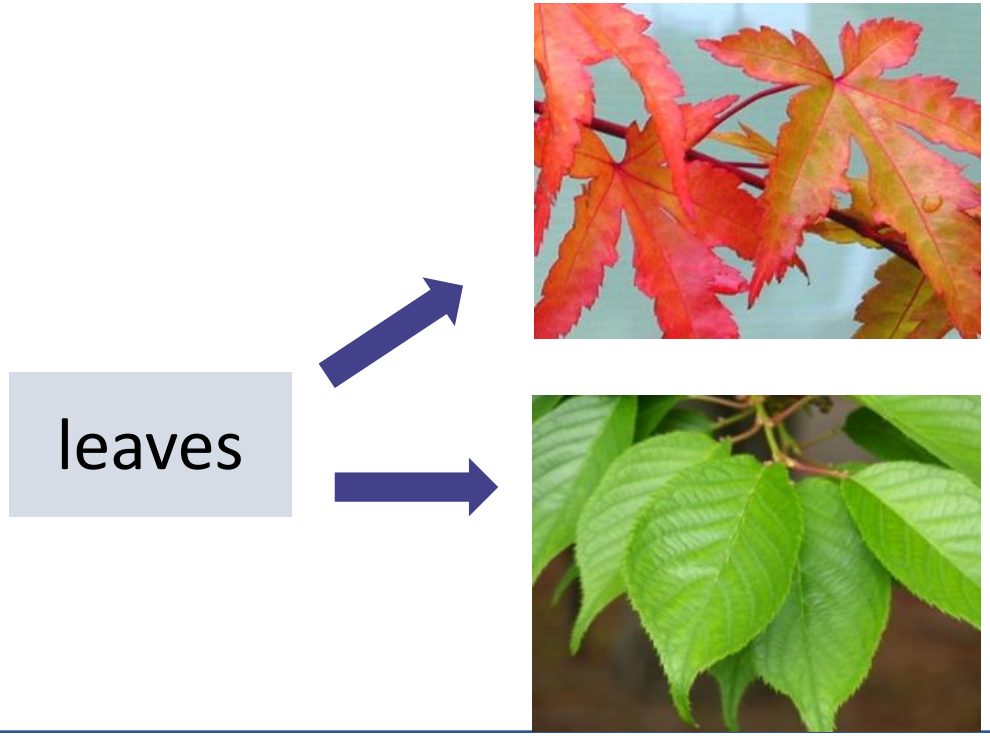
match

input old new



Experiments - problem

- Word2Vec cannot distinct “summer leaves” and “fall leaves” from only “leaves”
- One word is not enough explain visual feature



Experiments

- Leather images dataset
- Crawled by Google image search
- Search 84 keywords
- About 500 images each keywords

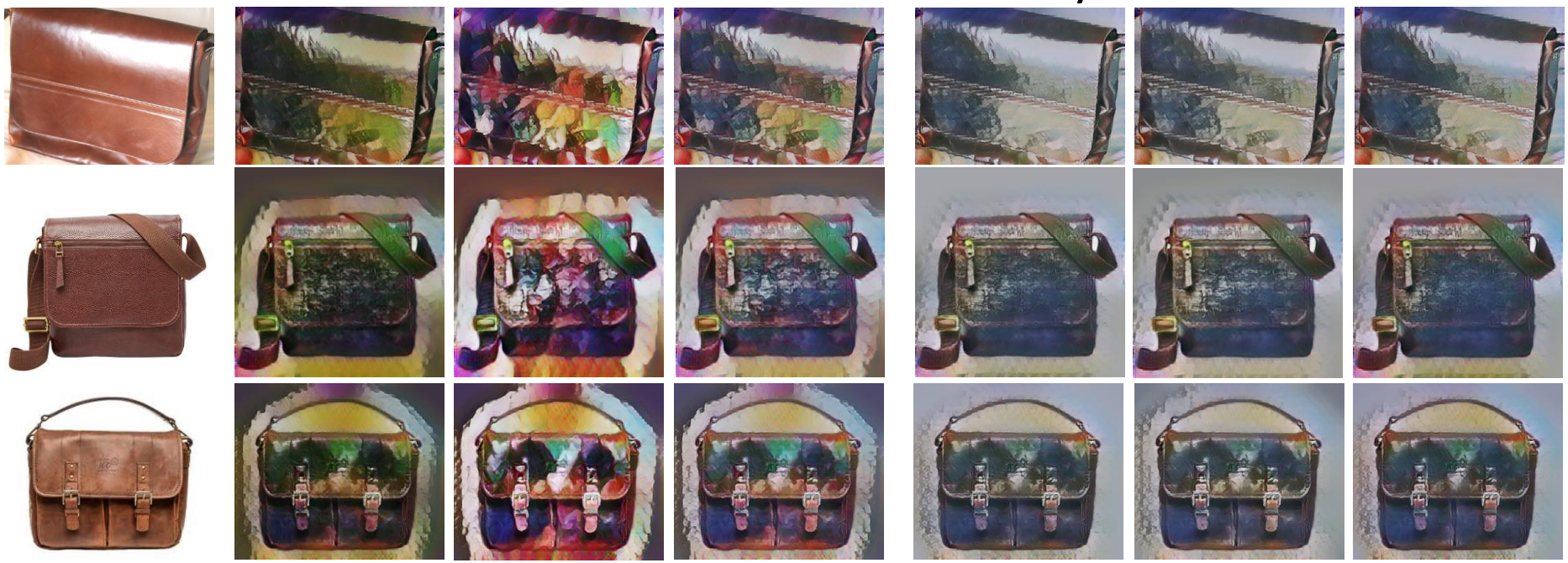
Leather advanced bag

Leather ancient bag

Leather	*	Advanced Ancient Recent elderly etc.	*	Bags Shoes wallet

Experiments

- Trained SSN only leather images model



Results and Discussions

- Adversarial loss generate superior feature map than L2 loss
  - L2 loss model generates mean feature of trained images
- No confidence to match the word meaning and visual feature
  - leather experiments solved several problems
- Cannot preserve background and object domain
  - Style transfer architectures are not match perfect.

Conclusion and Future Work

Conclusion

- Image Stylization with words without conditional approach
- There are question the transformation is it right for our feelings.

Future Work

- LSTM units for use sentences not only words
- Other domain transfer techniques for image synthesis network

Typical Bad results

- Input “red” --> image styled darker without color change
- Input “black” --> sometime image styled blight
- I think dataset not only that words area
  - “leather red wallet” image --> black wallet with red emblem
- Attention model can solve this problem
  - Segment “red” area of images

Experiments