Art Nouveau Style Transfer with Face Alignment

[Elena Tuzhilina] Stanford University, Department of Statistics

Motivation

Flourished throughout Europe and the United States at the turn of 19th and 20th centuries, Art Nouveau still remains one of the most beautiful decorative art movements. Promulgating the idea of art and design as part of everyday life and inspired by natural forms and patterns of plants and flowers, it has influenced different aspects of art and architecture, such as interior, furnishings and glass design, as well as graphic work, posters, and illustration. This project inspired by Henri de Toulouse-Lautrec and Alphonse Mucha works of art is aimed to develop a deep learning tool transforming already boring photos into a bright and bold Art Nouveau fine art posters.

Data

An example of *style image*. Alphonse Mucha paintings obtained from "Painter by numbers" Kaggle competition, 200 in total.

An example of *content image*. Images downloaded from Flickr using "women, vintage dress" tag, 2000 in total.

Baseline method

Input: content image C, style image S

Output: generated image G

Features: via VGG-16

- Content $a[\ell](C)$ - output of ℓ-th activation layer
- Style $GM[\ell](S)$ - gram matrix of layer ℓ, measures the correlation across the channels

Loss:

- Content $L_{content} = \frac{1}{2}\|a[L](C) - a[L](G)\|_2^2$
- Style $L_{style} = \sum_{l=1}^{L} \frac{\|GM[\ell](S) - GM[\ell](G)\|_F^2}{\#elements\ in\ GM[l](\cdot)}$
- Regularization $TV(G)$

$$L(G) = \alpha L_{content}(C, G) + \beta L_{style}(S, G) + \gamma TV(G)$$

Face Detection + alignment

MTCNN output format:

```
[{'box': [192, 188, 93, 121],
'confidence': 0.99922275,
'keypoints': {'left_eye': (218,
234), 'right_eye': (264, 239),
'nose': (237, 265), 'mouth_left':
(217, 277), 'mouth_right': (256,
281)}}]
```

IOU alignment: use `'box'`

1. Detect the faces on the content and style images.
2. Pick the face box with the highest confidence value.
3. Find the linear transformation that it maximizes the IoU of the transformed bounding boxes.
4. Crop only the "necessary" parts of the aligned pictures.

Procrustes alignment: use `'keypoints'`

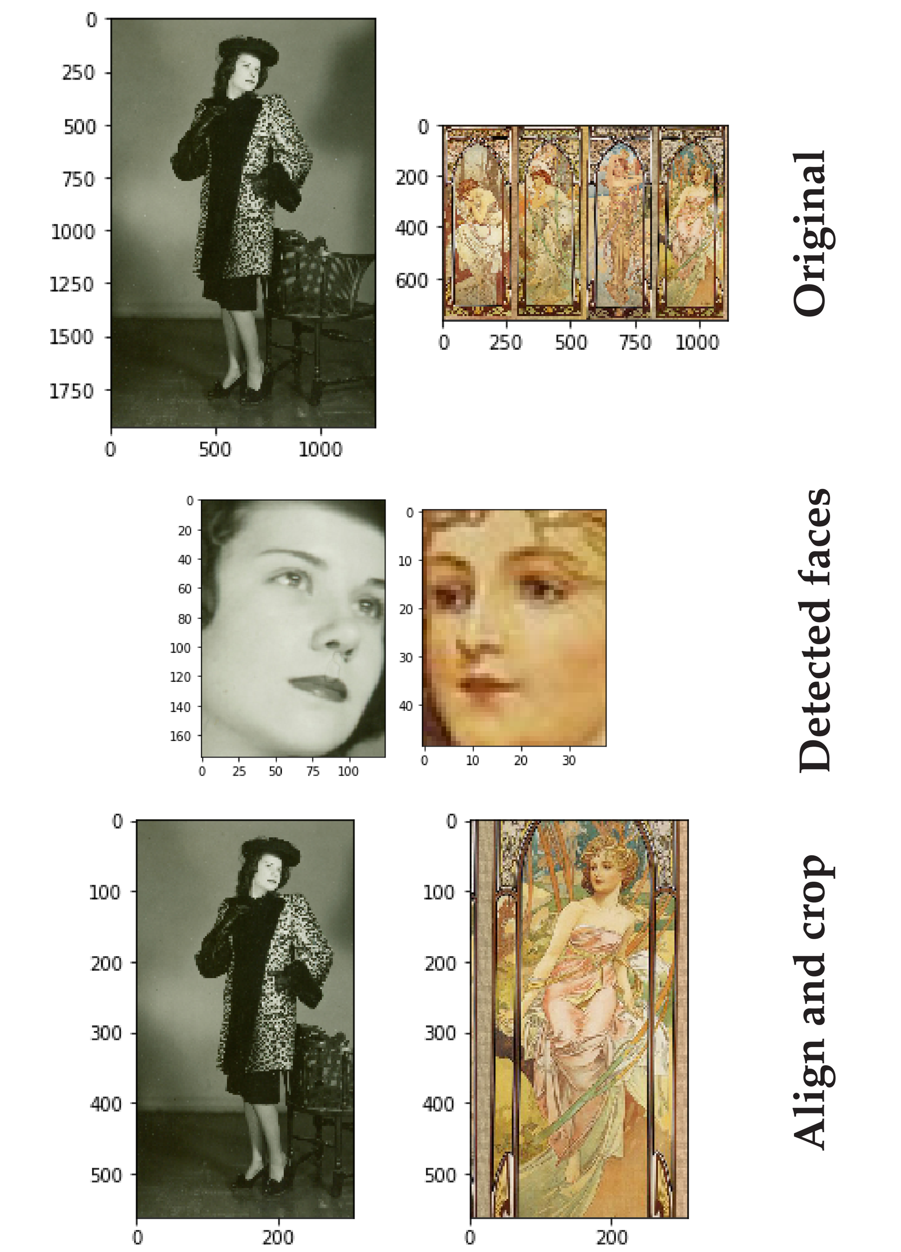
1. Detect the faces on the content and style images.
2. Pick the face box with the highest confidence value.
3. Create the matrices $X_{content}, X_{style} \in \mathbb{R}^{5\times 2}$ containing `'keypoints'` coordinates.
4. Solve the Procrustes optimization problem:

   minimize $\|X_{content} - s \cdot X_{style}R - b\|_F$ w.r.t. b, s and R.
5. Scale, shift and rotate the content and style images.
6. Crop the images to the same size.

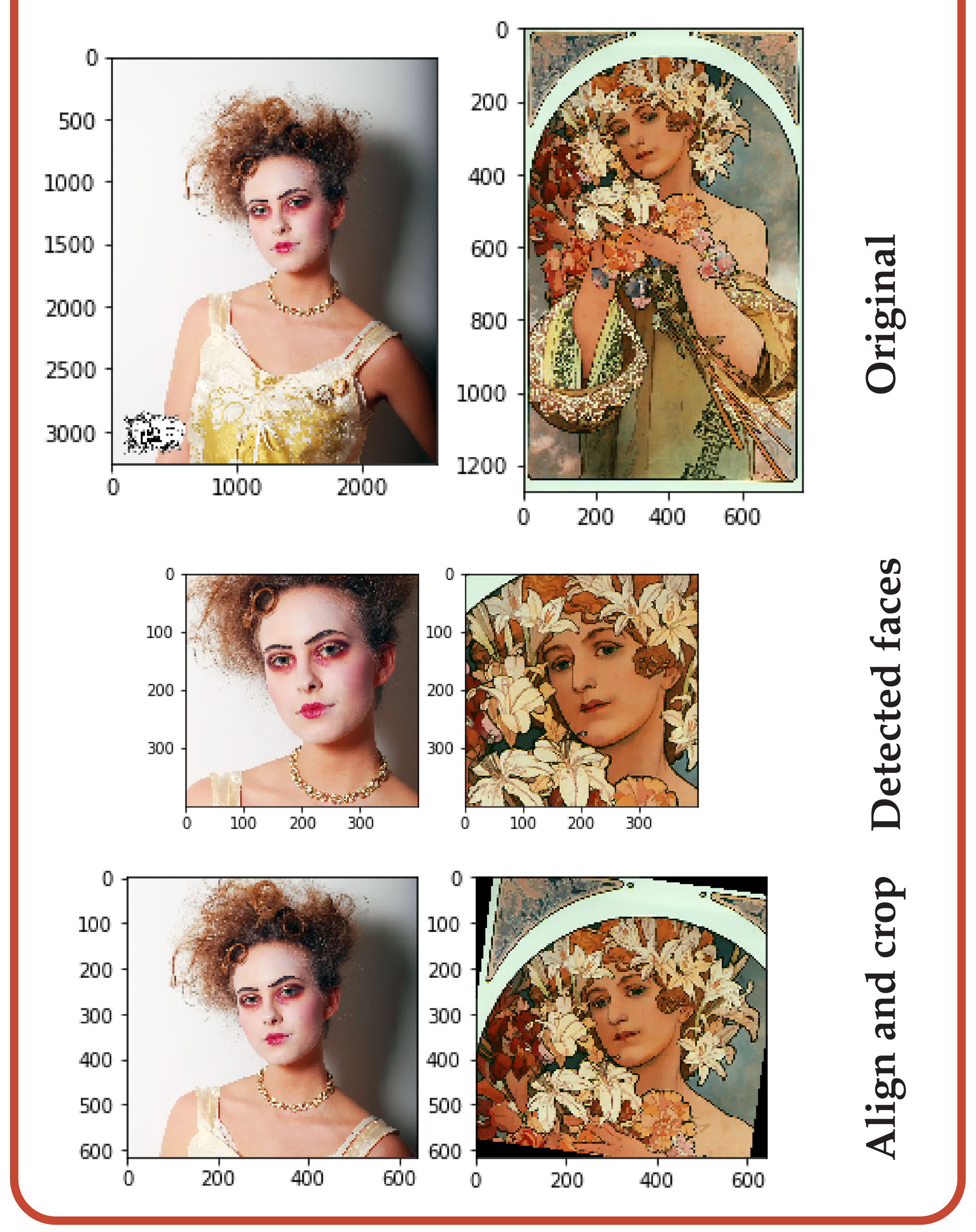
Use face `'box'` to build new loss:

$$L(G) = \alpha L_{content}(C_{aligned}, G) + \beta L_{style}(S_{aligned}, G) + \gamma TV(G|_{face\ box})$$

IOU example



Procrustes example



Hyperparameter tuning

- Learning rate $lr = 0.002, 0.05, 0.1$
- VGG-16 Layers $L_c = 1, \dots, 5$
  $L_s = 1, \dots, 5$
- Loss weights $\alpha = 10^3, 10^4, 10^5, 10^6$
  $\beta = 1$
  $\gamma = 0, 30, 300, 3000$

Results for 20 epochs, each 100 steps

(a) $\alpha = 10^3, \gamma = 300$ (b) $\alpha = 10^4, \gamma = 0$

(a) $\alpha = 10^4, \gamma = 300$ (b) $\alpha = 10^5, \gamma = 0$

Results (epoch = 50)

(a) Content (b) Style (c) NST (d) FST (e) RST (f) IOU (g) Procrustes

Discussion

- experiment with facial penalty: add pixel-to-pixel penalty measuring the deviation of G from C
- add normalization, e.g.adapt Fast Neural Style Transfer
- try Markov Random Fields approach to encode stylistic features

References

[1] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge, *A Neural Algorithm of Artistic Style* (2015), arXiv, https://arxiv.org/abs/1508.06576

[2] Justin Johnson, Alexandre Alahi, Li Fei-Fei *Perceptual Losses for Real-Time Style Transfer and Super-Resolution* (2016), arXiv, https://arxiv.org/abs/1603.08155

[3] Chuan Li, Michael Wand, *Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis* (2016), arXiv, https://arxiv.org/abs/1601.04589

[4] *MTCNN* https://github.com/ipazc/mtcnn