Raphael Painting Analysis
Transfer learning and Visualization

HU Wei, ZHAO Yuqi, YE Rougang, HAN Ruijian

Hong Kong University of Science and Technology

March 13, 2018
Outline

1. Data Description
2. Methodology
3. Visualization
Data Description

Raphael Paintings: 12 authentic, 9 fake and 7 disputed paintings.

Goal: Investigate the secret of Raphael!
Data Description

Preprocessing: crop (224, 224) patches from original paintings, remove almost blank parts (simply thresholding at variance of patches).
Data Description

Sequentially cropping v.s. random cropping

crop_size = (224, 224)
picture_dir = 'Raphael_Project'
disputed_id = [1, 7, 10, 20, 23, 25, 26]
authentic_id = [2, 3, 4, 5, 6, 8, 9, 21, 22, 24, 27, 28]
fake_id = [11, 12, 13, 14, 15, 16, 17, 18, 19]
labels = dict(**{str(x):'fake' for x in fake_id}, **{str(x):'disputed' for x in disputed_id}, **{str(x):'authentic' for x in authentic_id})
low_var_filter = True
low_var_threshold = 200

from Crop_Images import crop_Images
crop = crop_Images(picture_dir, labels, low_var_filter, low_var_threshold)

crop.random_crop(crop_size, n_multiple = 2)

folder: data/train/disputed does not exists, creating
folder: data/train/authentic does not exists, creating
folder: data/train/fake does not exists, creating
totally 18072 pictures created

crop.sequential_crop(crop_size, offset=(180, 180))

Folder: data/train\disputed does not exist, creating
Folder: data/train\authentic does not exist, creating
Folder: data/train\fake does not exist, creating
totally 12034 pictures created, we ignore 1055 low variance pictures
Both validation and test sets consist of one authentic and one fake paintings.

crop.shuffle()

take 24 and 12 as validation pictures
take 3 and 16 as test pictures
Outline

1. Data Description
2. Methodology
3. Visualization
We borrow pretrained ResNet18 from PyTorch, reset FC layer.
Resnet18 has 4 such Layers. Next, we shall tune the number of freeze Layers.
Results

<table>
<thead>
<tr>
<th>Typical models</th>
<th>Good model</th>
<th>Bad model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layers Trained</td>
<td>train</td>
<td>val</td>
</tr>
<tr>
<td>FC layer</td>
<td>86.98</td>
<td>97.98</td>
</tr>
<tr>
<td>Layer 4, FC layer</td>
<td>93.87</td>
<td>99.36</td>
</tr>
<tr>
<td>Layers 3 &amp; 4, FC layer</td>
<td>99.90</td>
<td>99.79</td>
</tr>
</tbody>
</table>

Good model: Val: 21,18 Test:9,12
Bad model: Val:24, 12 Test:3,16
We compare 8 popular methods in Manifold Learning on the test sets. The result of the Good model (Layers 3, 4 and FC layer) is as follows:
Manifold Learning

The result of the bad model (Layers 3 & 4, FC layer) is as follows:
Outline

1. Data Description
2. Methodology
3. Visualization
Visualization directly on painting
Visualization

Motivation

- The performance of model highly depends on the choice of data segmentation.
- Lack of data - prior knowledge - visualization.
- Visualization bridge the gap between art master and data scientist.
Bad Model: Validation

(a) #12 Fake

(b) #24 Authentic
Visualization

Bad Model: Test

(c) #16 Fake

(d) #3 Authentic
Possible Reasons

- These are the only 2 landscape paintings in datasets.
- Model did not learn any features for landscape painting.
Our model gives 48% of patches to be real.
Model mis-recognize contaminated patches.