Style Image Retrieval Using CNN-based Style Vector

Shin Matsuo, and Keiji Yanai
The University of Electro-Communications, Tokyo
JAPAN

Background

- Style recognition will help analyzing of images/videos.

- However, style recognition is more difficult than content recognition.
    - Some images have common contents with different styles for misleading elements.

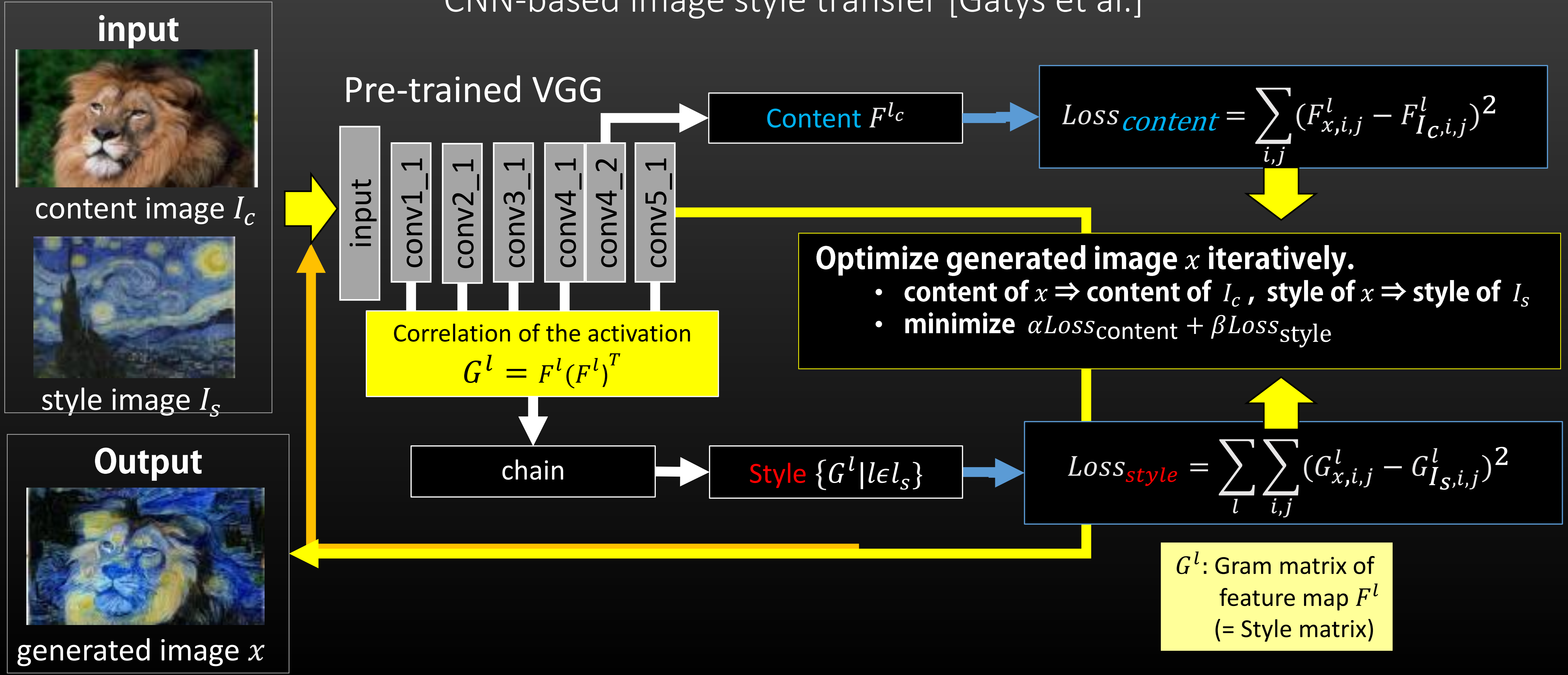
Objective

Recognition of images with their style, using novel image representation.

- We propose “style vector” based on activation of CNN for this goal.
- Outperform CNN features for style image retrieval.

Related Work

CNN-based image style transfer [Gatys et al.]



Related Work

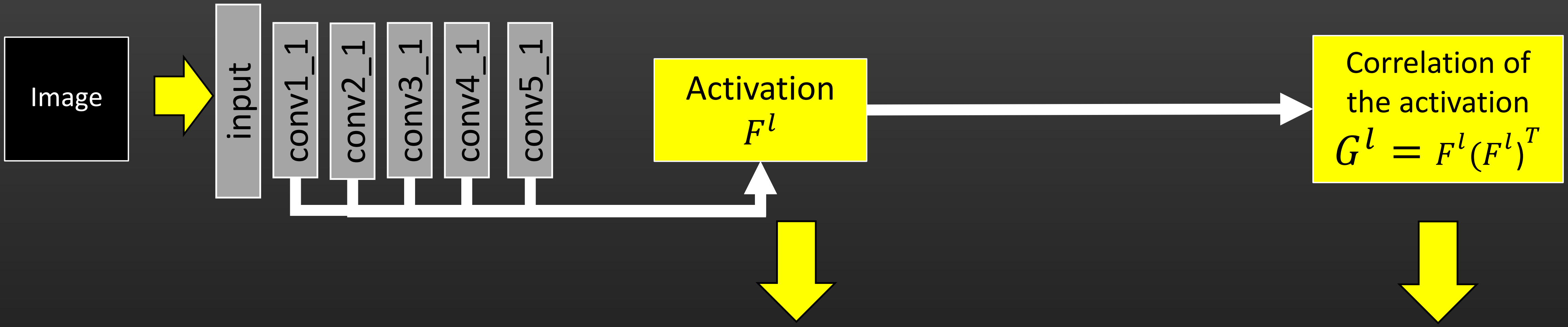
Recognizing image style [S. Karayev et al, British Machine Vision Conference 2013]

- Classifying style images using CNN activation features.
- We use same dataset and compare the performance.

Visualizing and Understanding Deep Texture Representations[Tsung-Yu Lin et al, CVPR 2016]

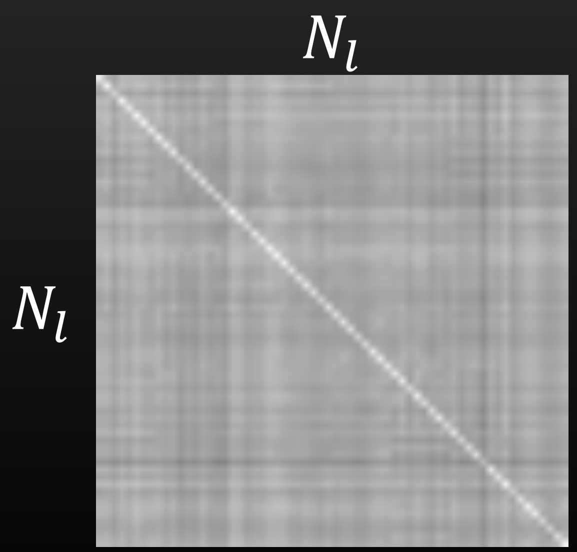
- Texture recognition with Bilinear-CNN feature.

Style Vector



Filter activation F^l

N_l: the number of filters $\quad$ M_l : the number of pixels



Gram matrix G^l

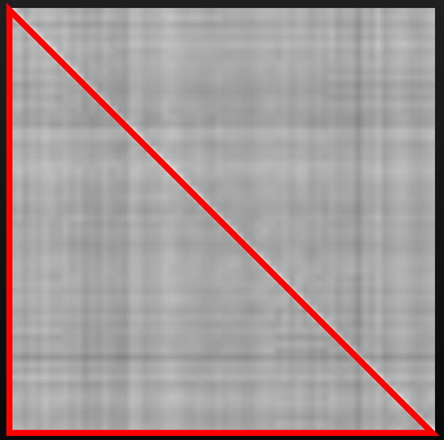
Style Vector

Convert G^l into vector V^l excluding symmetrical elements.

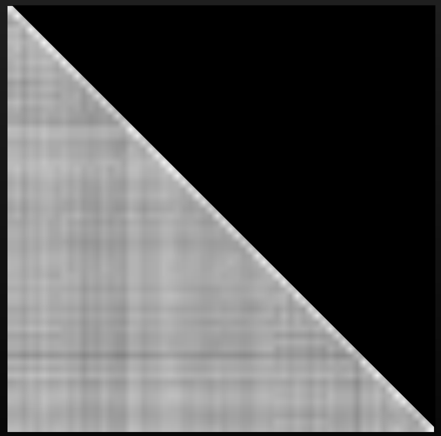
$$V^l = [G^l_{1,1}, G^l_{2,1}, G^l_{2,2}, \dots, G^l_{N_l,1}, G^l_{N_l,2}, \dots G^l_{N_l,N_l}]$$

$$|V^l| = (hurf\ elements) + (diagnal\ elements) = N_l * (N_l + 1)/2$$

N_l



N_l



Style vector V^l

PCA
(to 4096, 2048, 1024)

Ex, at conv5_1
$N_{conv5\_1} = 512$,
$|V^{conv5\_1}|$=131,328

Style Vector

Normalize V^l with several ways

L2-norm

$$S^{l_{L2}} = \frac{V^l}{\|V^l\|}$$

signed square root + L2-norm.

$$S^{l_{sgnsqrt}} = \frac{sgn(V^l)\sqrt{V^l}}{\|sgn(V^l)\sqrt{V^l}\|}$$

$$\Rightarrow V^l(\text{raw}),\ S^l_{L2}\ (\text{L2norm}),\ S^l_{sgnsqrt}(\text{sgnsqrt})$$

Compare three normalizations in experiment

Experiment

1. Style retrieval with direct style vector

2. Style retrieval with PCA

3. Comparison with other work[2]

Dataset

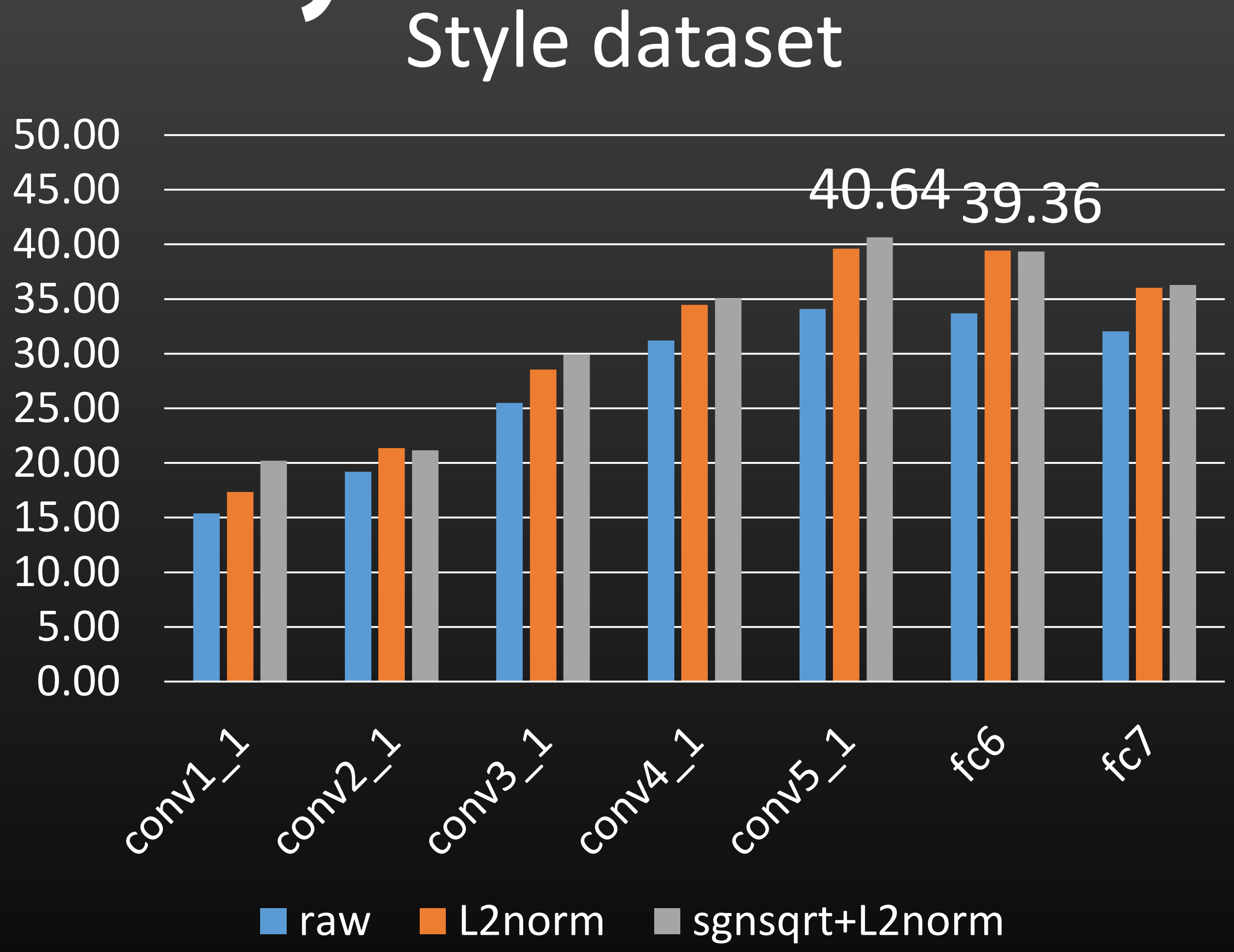
	Style/Karayev* Dataset	Artist Dataset
Classes	Style	Artist
The number of class	25	10
The number of Images	2500 / 82437	1000
examples	 Abstract Art  Baroque  Ukiyo-e	 Camille Pissarro  Pablo Picasso  Salvador Dali

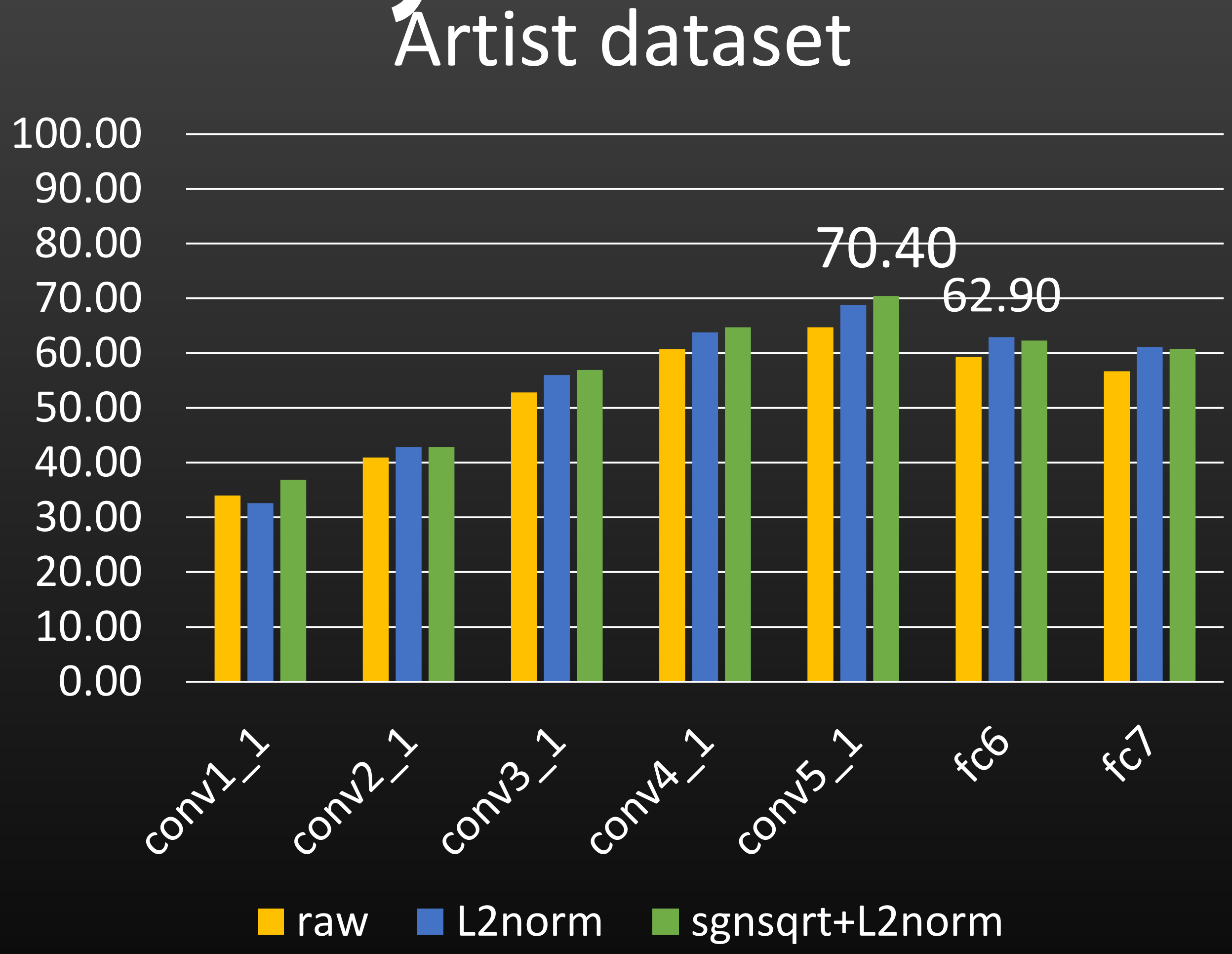
- **The Image Dataset is collected in wikiart.org.**
- *** the same dataset as Karayev [2]**

Style retrieval with direct style vector

data	**style dataset and artist dataset**
layer	conv1_1, …, conv5_1
classification method	k-nearest neighbor
normalization	raw, L2norm and sgnsqrt + L2norm
baseline	CNN features (fc6, fc7 of VGG-16)

Style retrieval with direct style vector





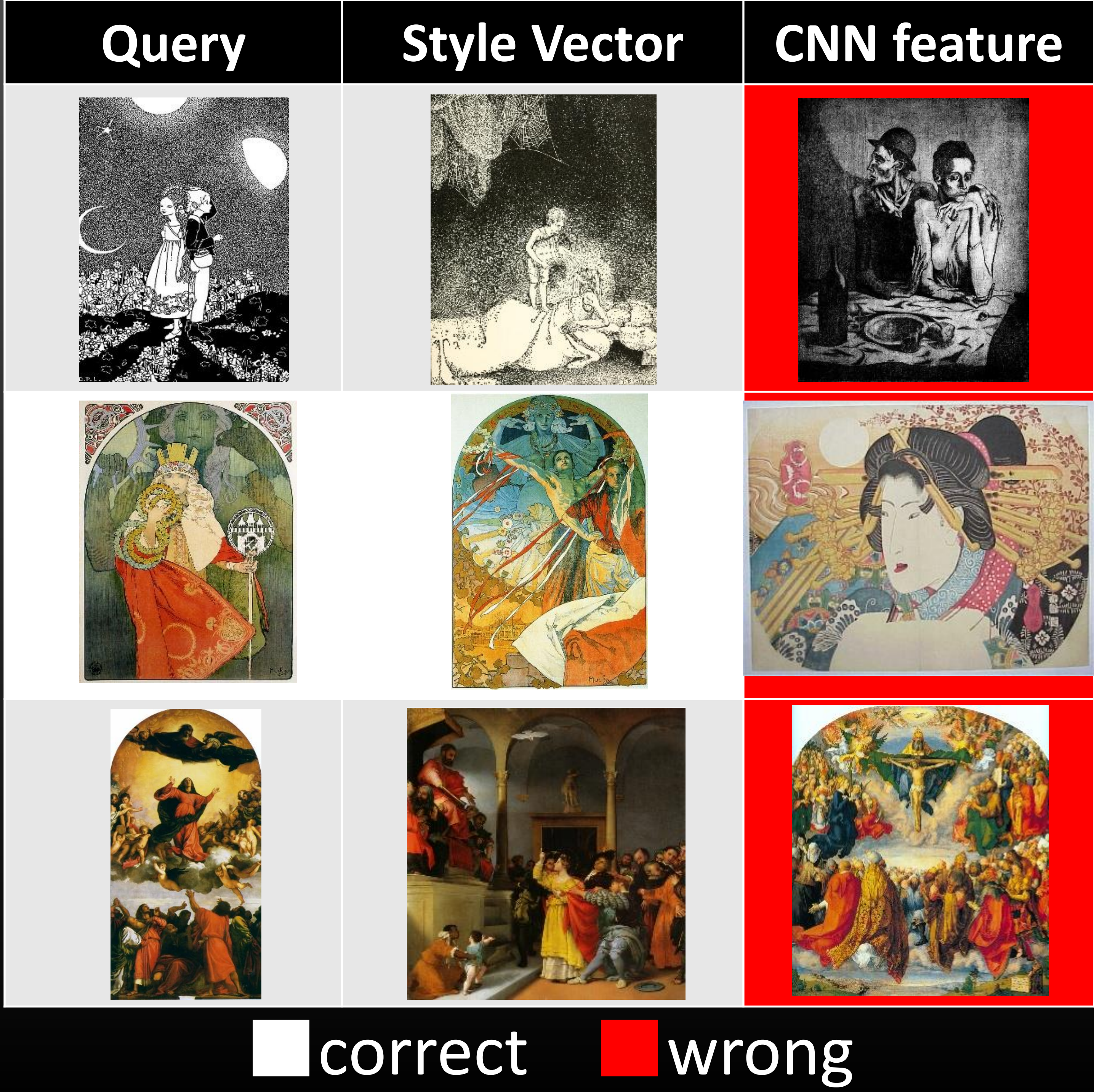
- **Conv5_1 layer and sgnsqrt+L2norm normalization was the best.**

Retrieval examples

Success example

(style vector worked better)

Common object like humen in many arts, with unique style



Retrieval examples

Failure example

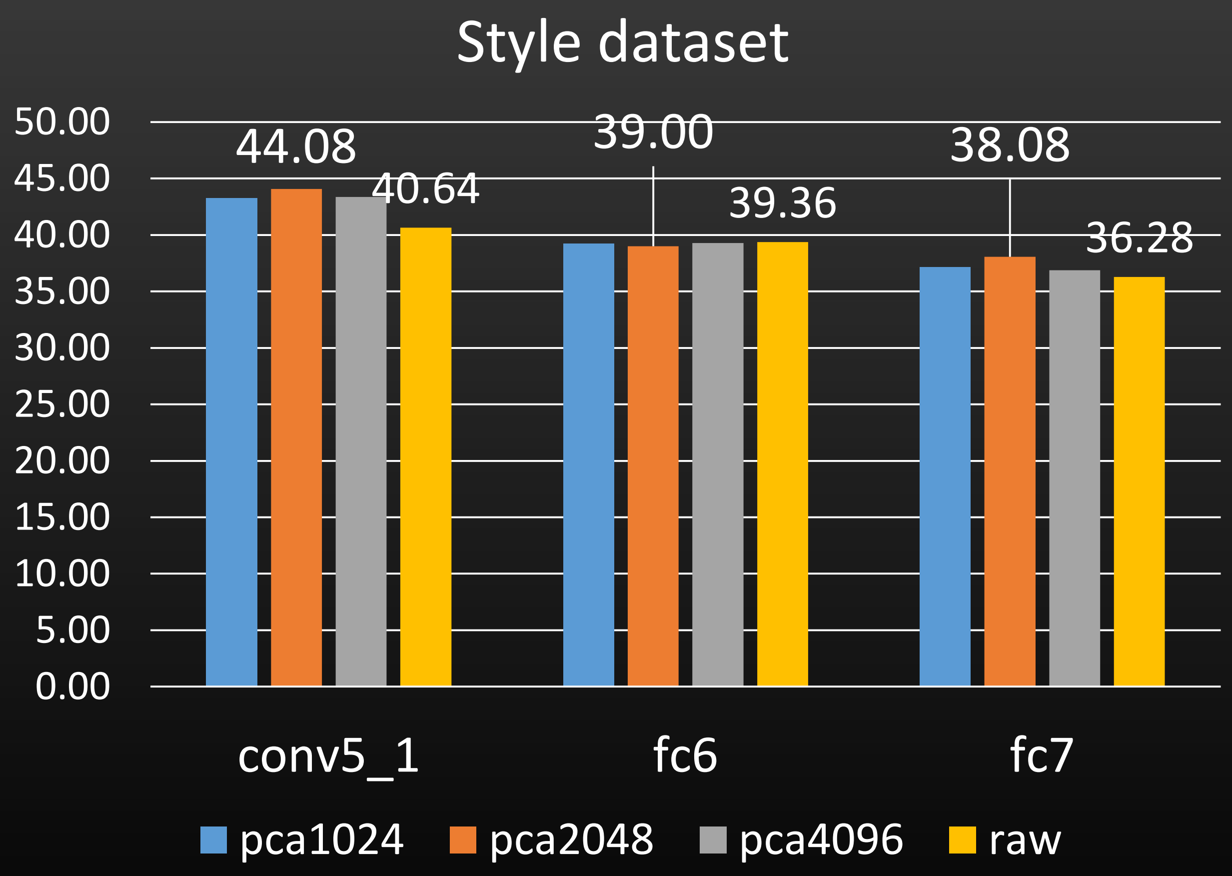
(CNN feature worked better)

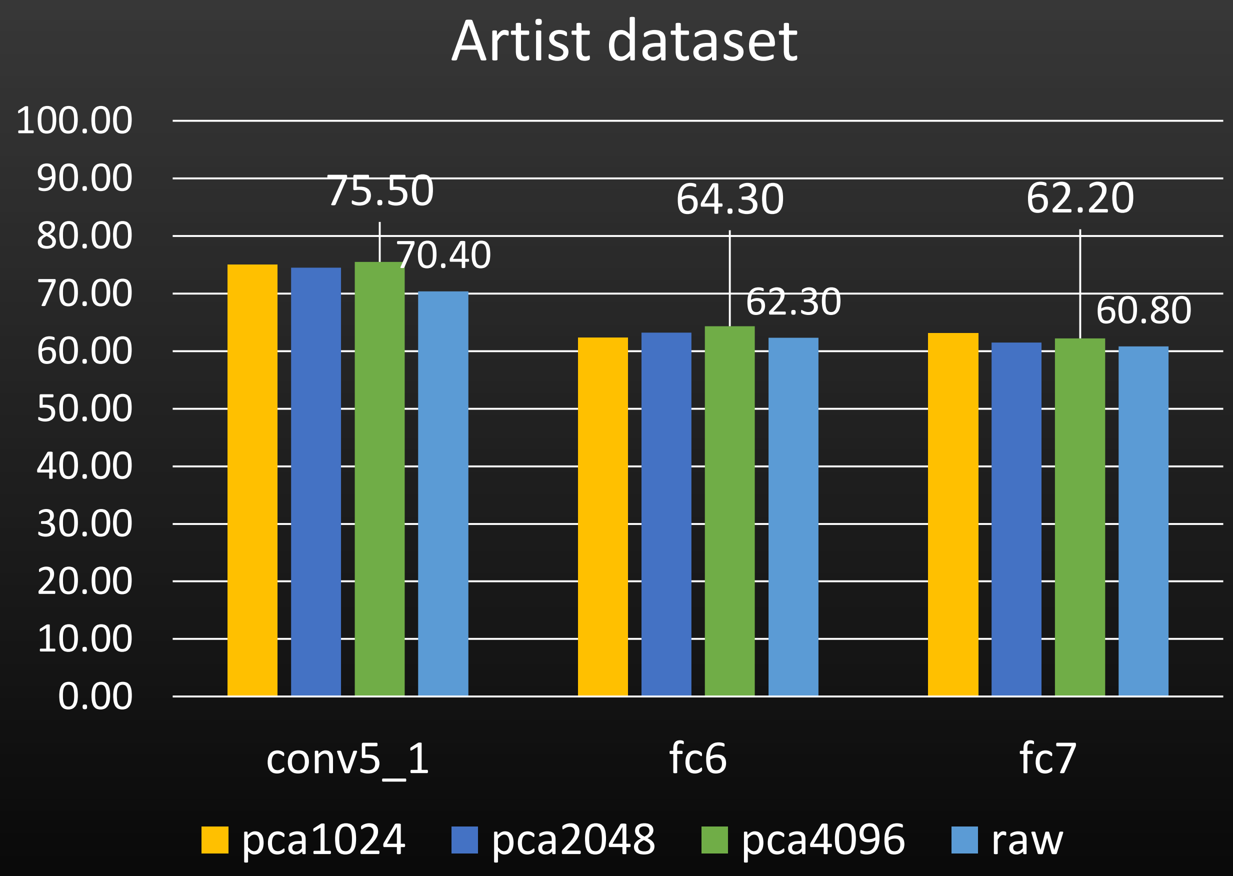
Whole shapes are more important than style.

Style retrieval with PCA

data	**style dataset and artist dataset**
layer	conv5_1
classification method	k-nearest neighbor
normalization	sgnsqrt + L2norm
dimension	4096, 2048, 1024
baseline	CNN features (fc6, fc7 of VGG-16)

Style retrieval with PCA



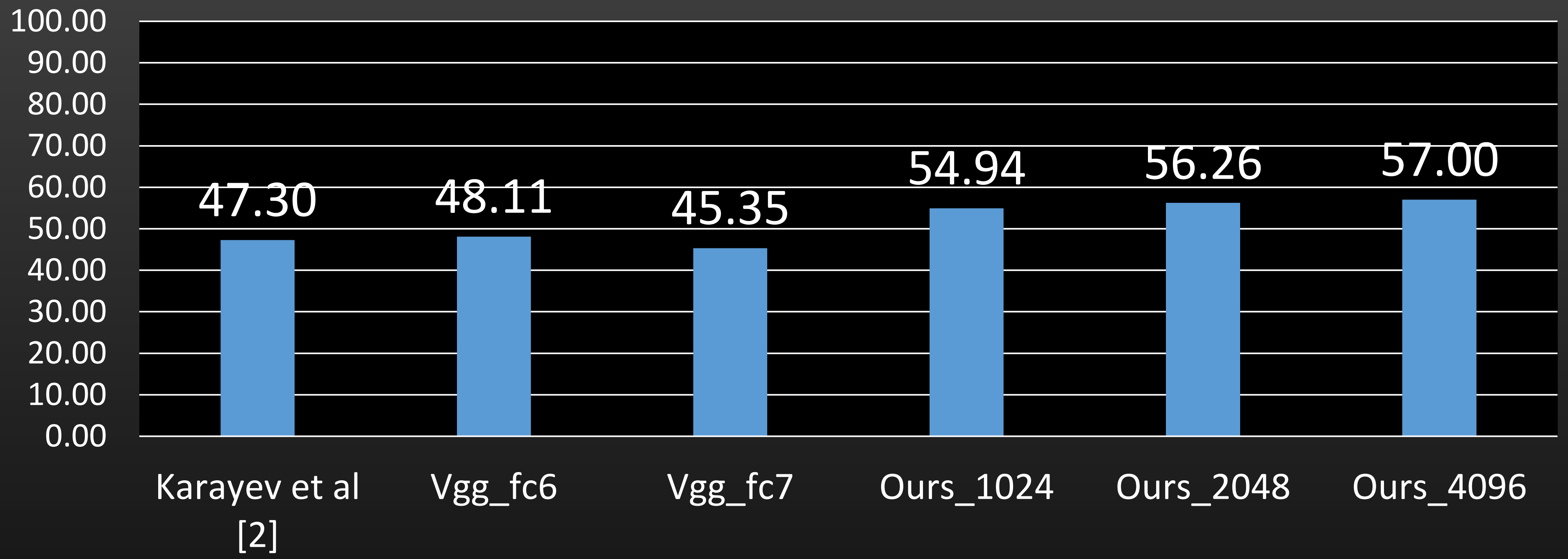


- **The performance was boosted by PCA dimension reduction.**

Comparison with other work

data	**Karayev dataset**
layer	conv5_1
classification method	SVM
normalization	sgnsqrt + L2norm
dimension	4096, 2048, 1024
baseline	CNN features (fc6, fc7 of VGG-16)

Comparison with other work



Baselines			**Ours**		
Karayev et al [2]	VGG16_fc6	VGG16_fc7	pca1024	pca2048	pca4096
47.30	48.11	45.35	54.94	56.26	57.00

Style vector outperformed the previous work.

Conclusions

- **Style vector outperformed CNN features in style retrieval of art images.**
- **The performance was boosted by introducing PCA dimension reduction**
- **Style vector worked well for the images with common objects and unique styles.**