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

Sanghyuck Na
May, 29, 2020
Dongguk University
Artificial Intelligence Laboratory
shna@Dongguk.edu

0

Contents

1. Style Transfer
2. Super-Resolution
3. Perceptual Loss for Style Transfer and Super Resolution
4. Result
5. Reference

1 Style Transfer

The goal of style transfer is to generate an image $\hat{y}$ that combines the content of a target content image y_c with the style of a target style image y_s.

1. **Style transfer using Pre-trained networks**
   - It is possible with only two images(content image & style image).
   - When generating a new image, it needs to be newly optimized.

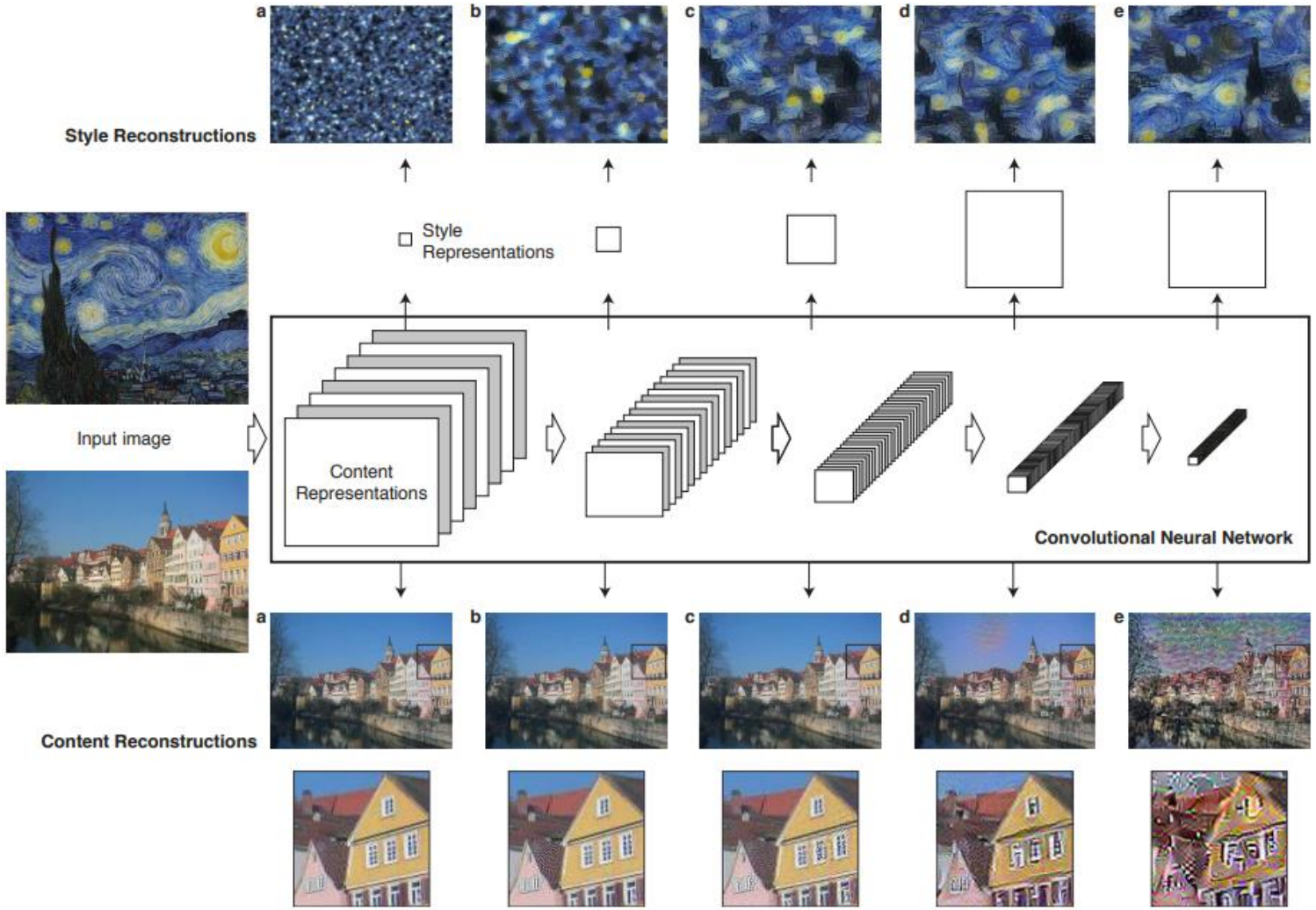
2. **Train a style transfer network**
   - When training a network and applying it to a new image, you only need to feed forward.
   - It needs to train the network, so I need a lot of images and it takes a long time to train.

1 Style Transfer

Red bounding box : Style Image

Content image

1 Style Transfer



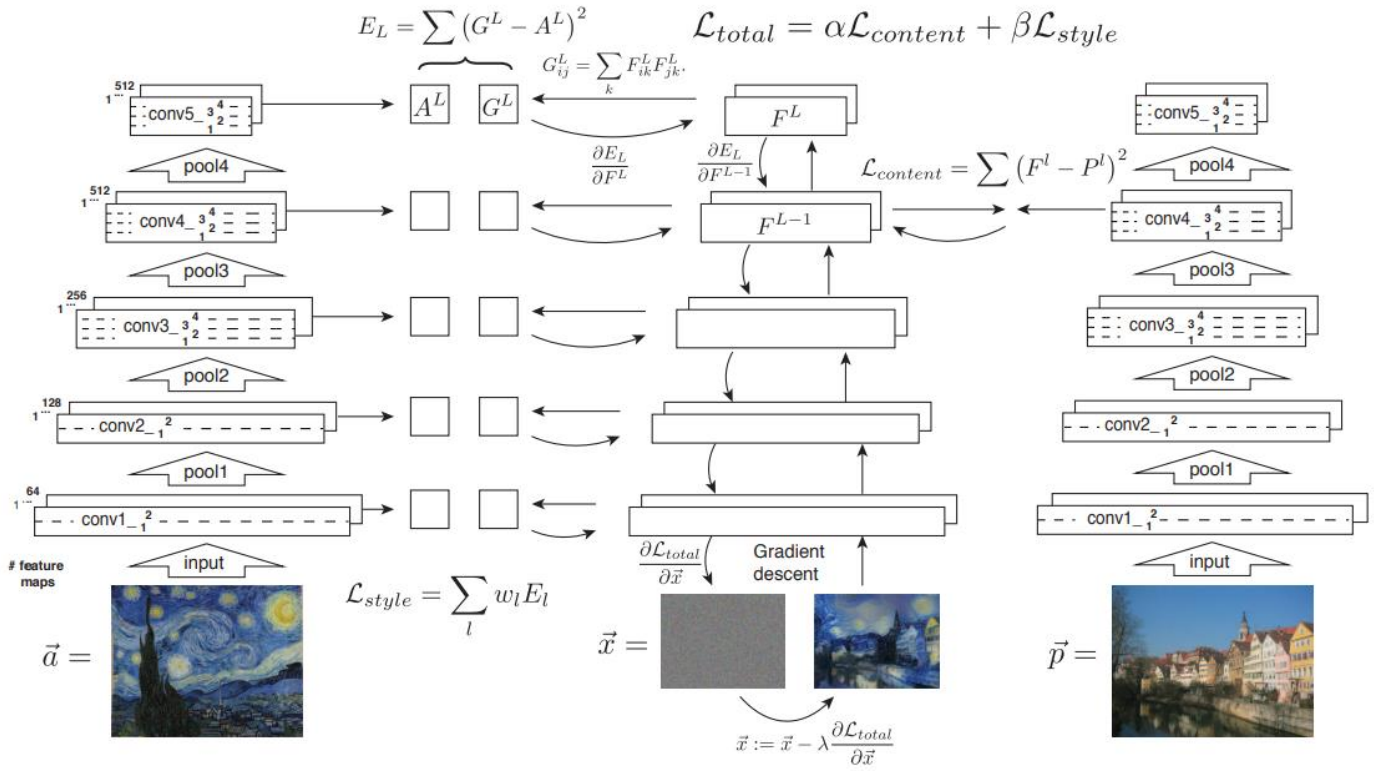
Style Reconstructions

'conv1_1(a)',
'conv1_1, conv2_1(b)',
'conv1_1, conv2_1, conv3_1(c)',
'conv1_1, conv2_1, conv3_1, conv4_1(d)',
'conv1_1, conv2_1, conv3_1, conv4_1, conv5_1(e)

Content Reconstructions

'conv1_2(a)',
'conv2_2(b)',
'conv3_2(c)',
'conv4_2(d)',
'conv5_2(e)

1 Style Transfer



Content representation

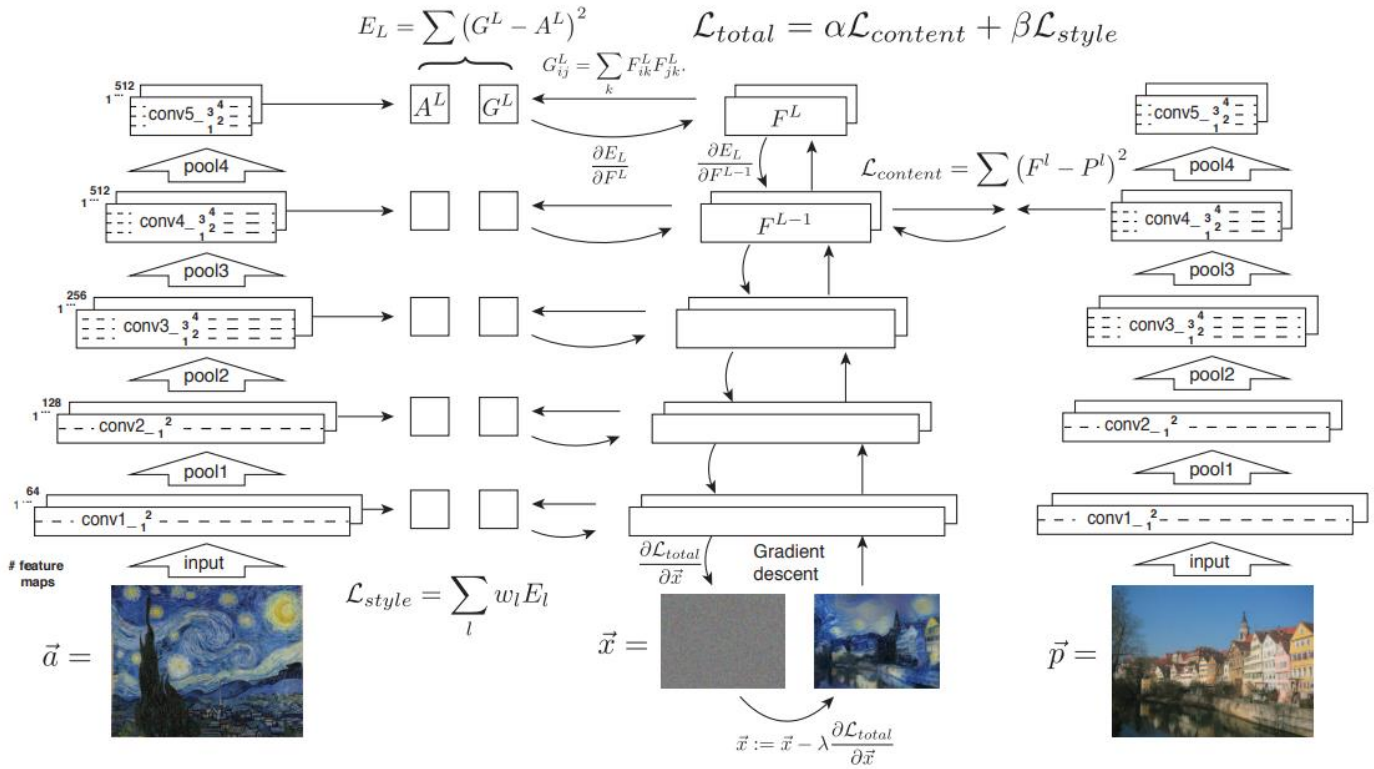
$$L_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F^l_{ij} - P^l_{ij})^2$$

$$\frac{\partial L_{content}}{\partial F^l_{ij}} = \begin{cases} (F^l - P^l)_{ij} & \text{if } F^l_{ij} > 0 \\ 0 & \text{if } F^l_{ij} < 0 \end{cases}$$

F^l_{ij} : *activation of the* i^{th} *filter at position j in layer l using content image*

P^l_{ij} : *activation of the* i^{th} *filter at position j in layer l using noise image*

1 Style Transfer



Style representation

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk}$$

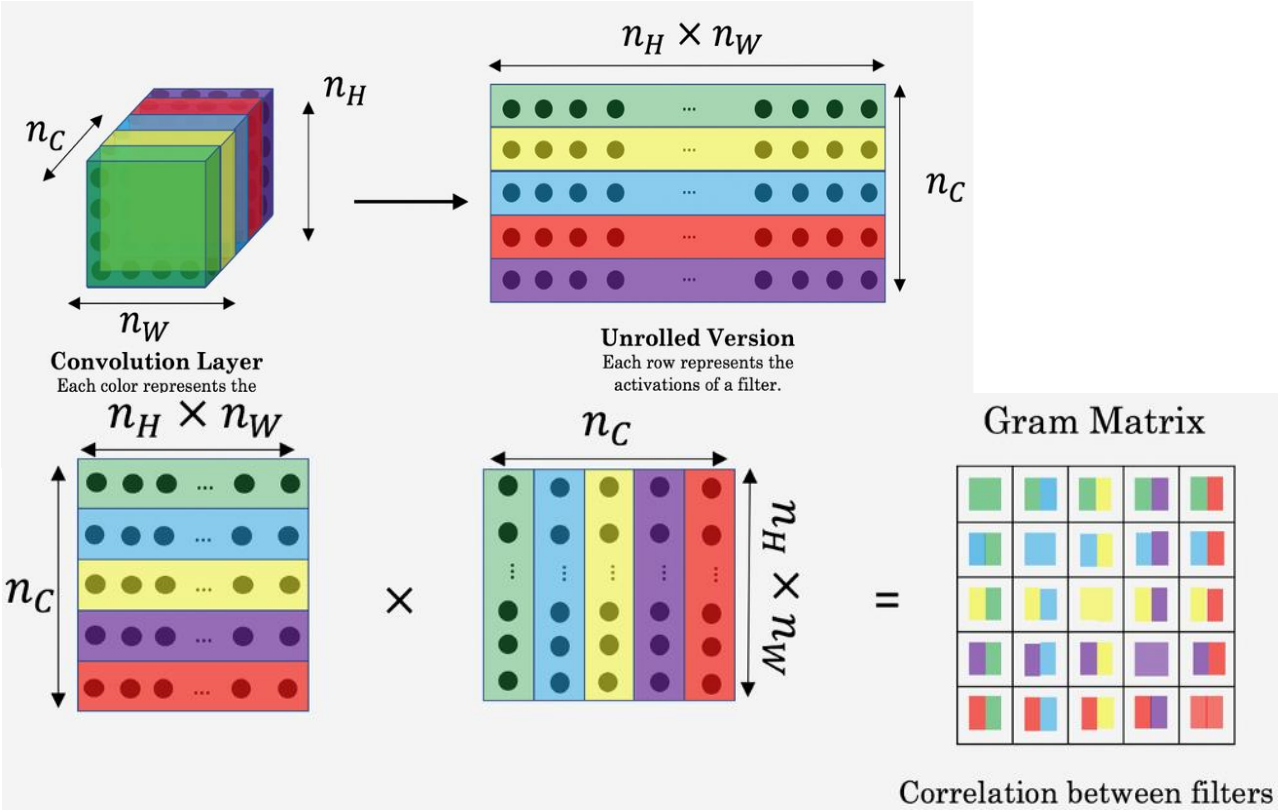
$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G^l_{ij} - A^l_{ij})^2$$

$$L_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l$$

$$\frac{\partial E_l}{\partial F^l_{ij}} = \begin{cases} \frac{1}{N_l^2 M_l^2} \left((F^l)^T (G^l - A^l) \right)_{ij} & \text{if } F^l_{ij} > 0 \\ 0 & \text{if } F^l_{ij} < 0 \end{cases}$$

A^l: *where* A^l_{ij} *is the inner product between* F^l_i F^l_j *in layer* l *using style image*
G^l: *where* G^l_{ij} *is the inner product between* F^l_i F^l_j *in layer* l *using noise image*
N_l: *where # of feature maps at layer* l
M_l: *height* $*$ *width of feature maps at layer* l

1 Style Transfer



Style representation

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$$

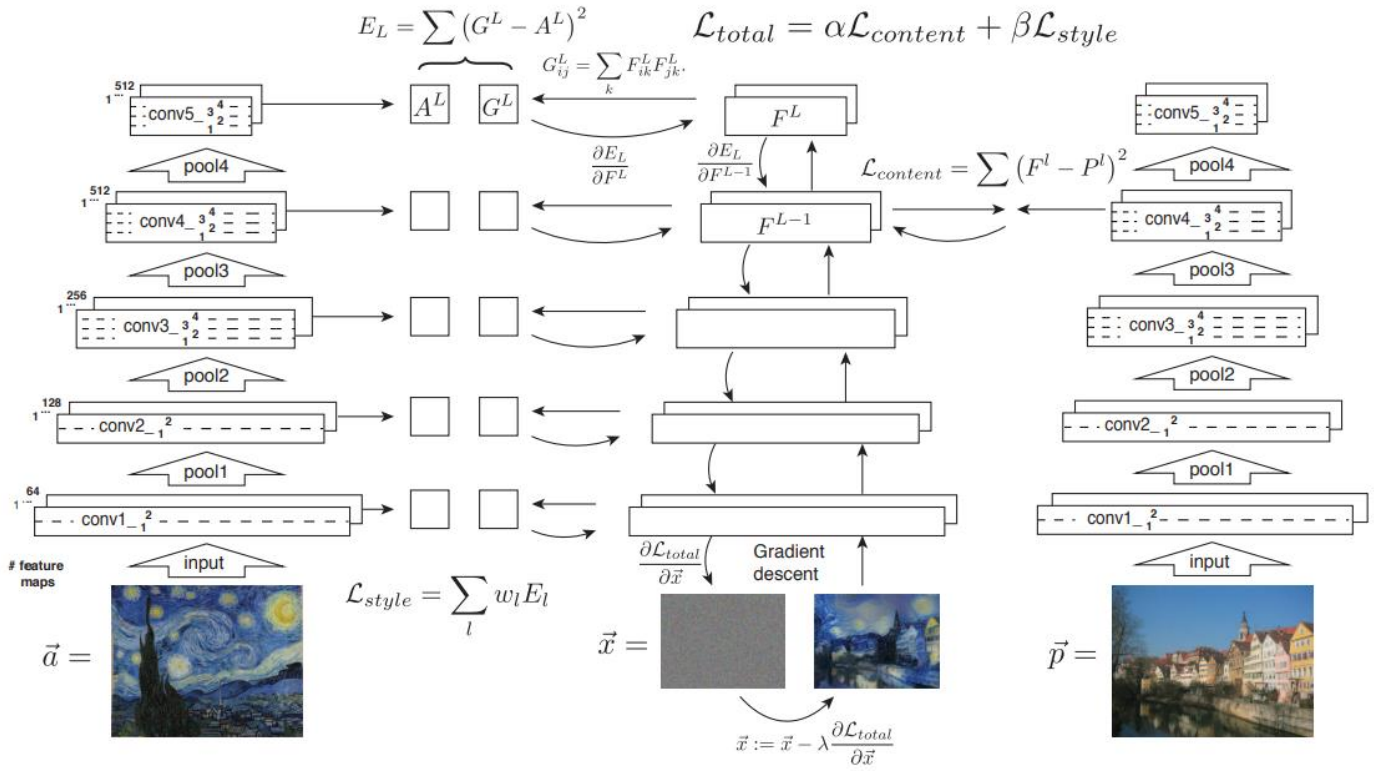
$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

$$L_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l$$

$$\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2} \left((F^l)^T (G^l - A^l) \right)_{ij} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases}$$

A^l: where A_{ij}^l is the inner product between F_i^l F_j^l in layer l using style image
G^l: where G_{ij}^l is the inner product between F_i^l F_j^l in layer l using noise image
N_l: where # of feature maps at layer l
M_l: height $*$ width of feature maps at layer l

1 Style Transfer



Style transfer

$$L_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha L_{content}(\vec{p}, \vec{x}) + \beta L_{style}(\vec{a}, \vec{x})$$

1 Style Transfer

$\alpha/\beta = 0.0001$

$\alpha/\beta = 0.001$

$\alpha/\beta = 0.01$

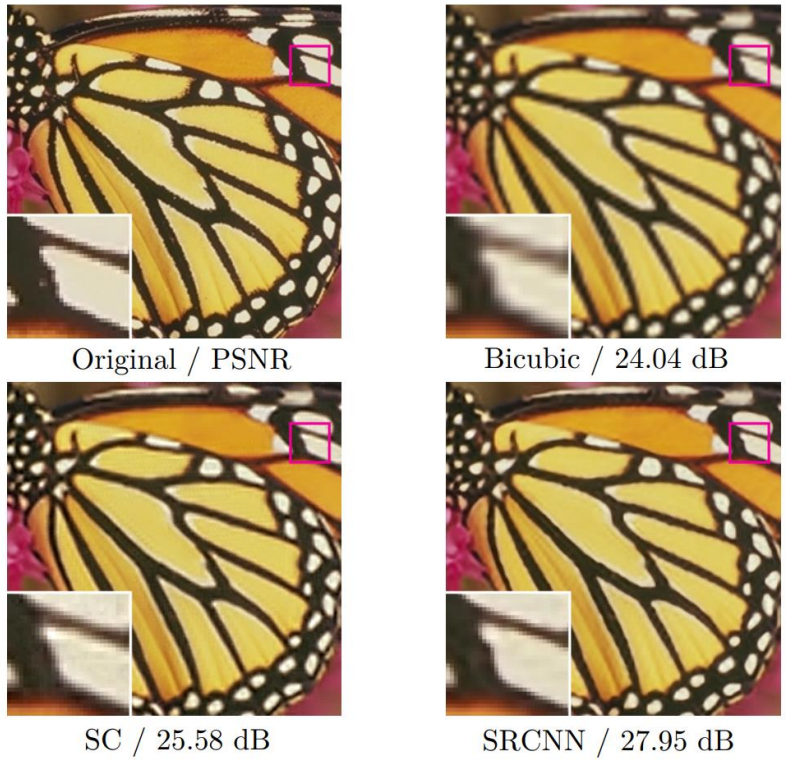
$\alpha/\beta = 0.1$

Style Image

Content Image

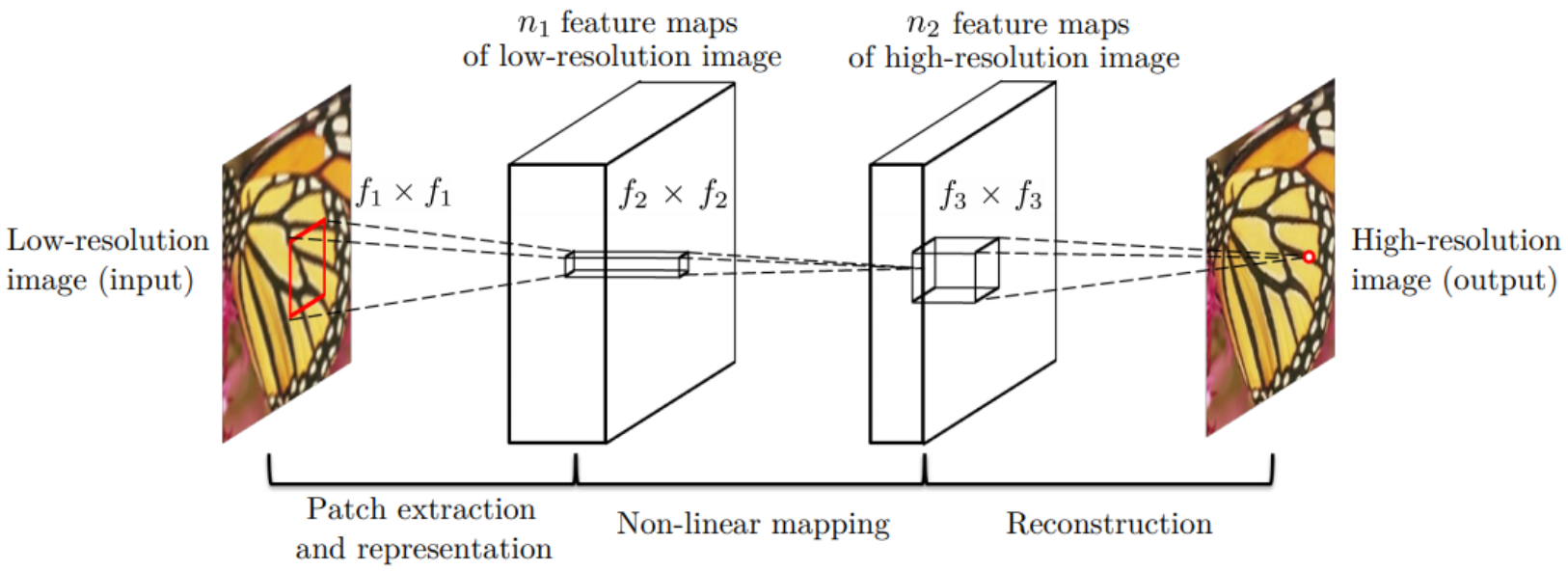
$\alpha/\beta = 0.01$

2 Super-Resolution



The task of estimating high-resolution (HR) images from low-resolution (LR) counterpart is referred to as super-resolution (SR).

2 Super-Resolution

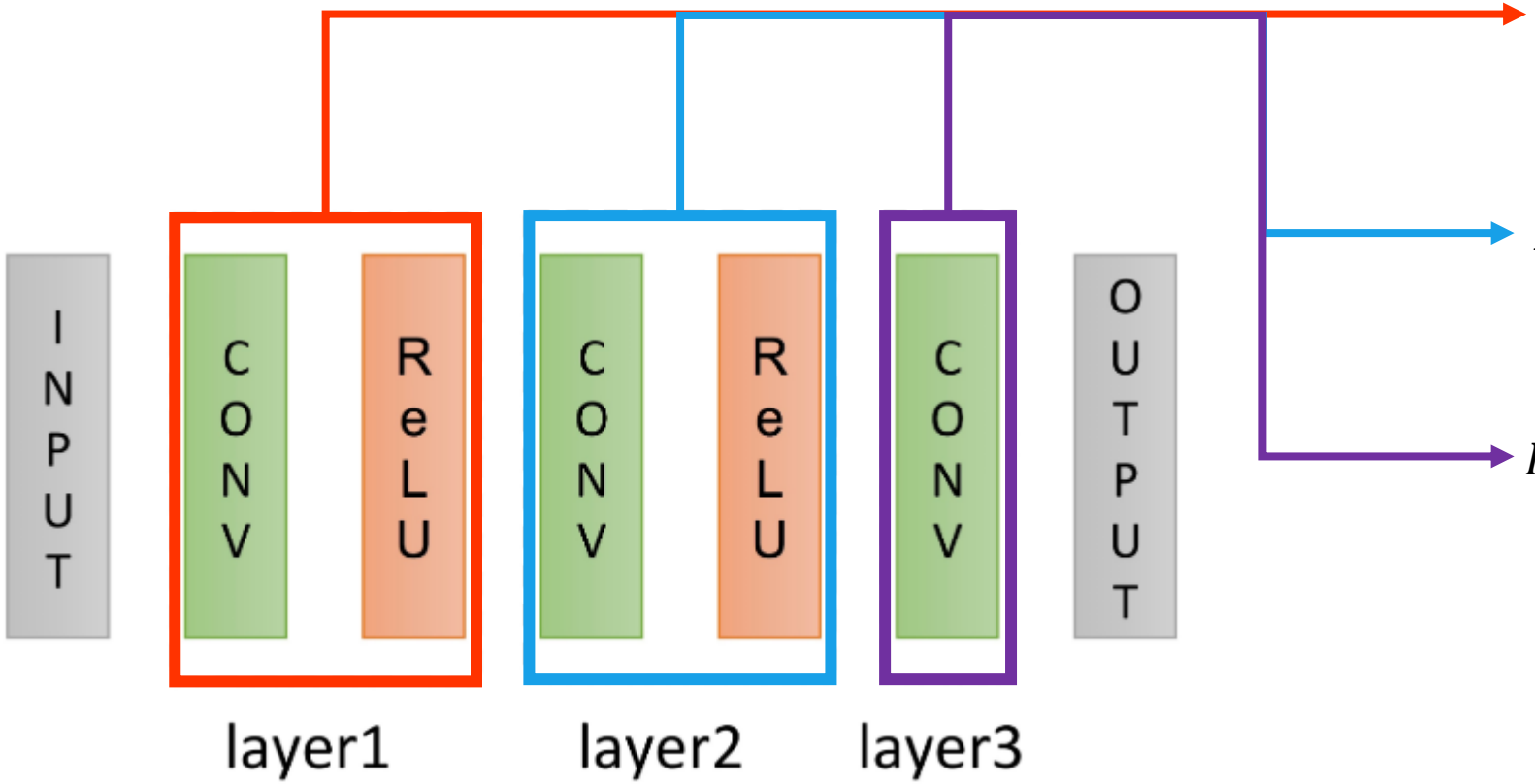


Patch extraction and representation : Patch extraction rom low-resolution image Y

Non-linear mapping : mapping to high dimension patch vector to other high dimension

Reconstruction : generate the super-resolution image from high dimension patch vector

2 Super-Resolution



Patch extraction and representation

$$F_1(Y) = \max(0, W_1 * Y + B_1)$$

Non-linear mapping

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2)$$

Reconstruction

$$F(Y) = W_3 * F_2(Y) + B_3$$

Mapping function

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \|F(Y_i;\Theta) - X_i\|^2$$

Y : low resolution image
F(Y) : Super-resolution image
X :Ground truth high resolution image

3 Perceptual Loss for Style Transfer and Super Resolution

Overview

Propose the used of perceptual loss for training feed-forward networks for image transformation.

Per-pixel loss functions compare two images base on their pixel values.

Perceptual loss function compare two images base on high-level representation from pretrained neural net.

3 Perceptual Loss for Style Transfer and Super Resolution

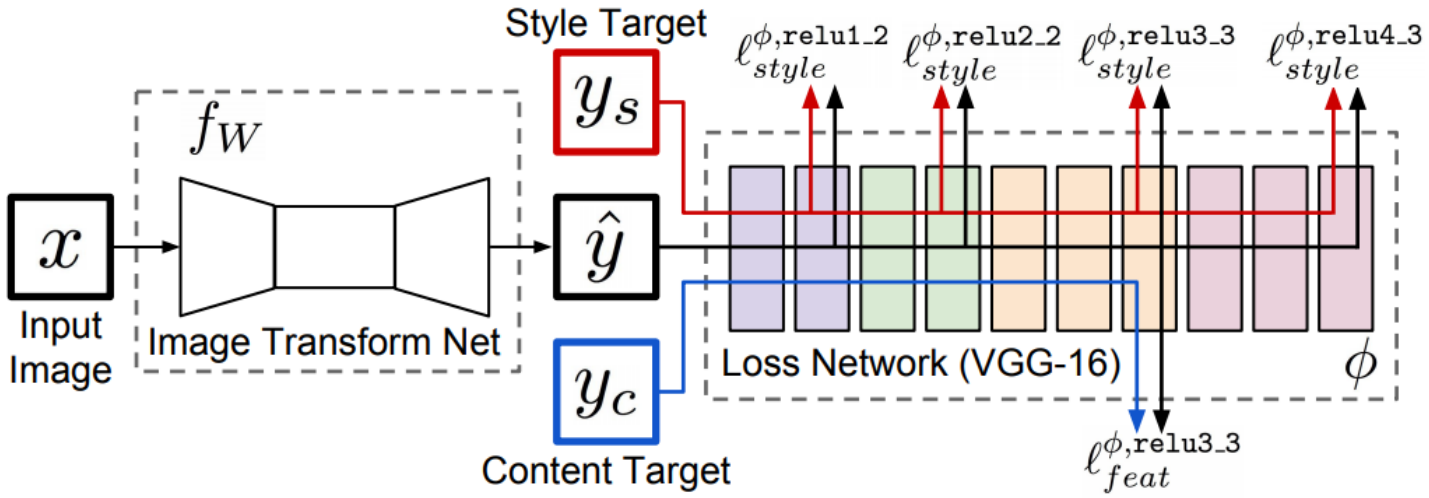


Image transform network

$$W^* = \arg\min_W E_{x,\{y_i\}}\left[\sum_{i=1} \lambda_i \ell_i(f_W(x), y_i)\right]$$

Feature Reconstruction Loss

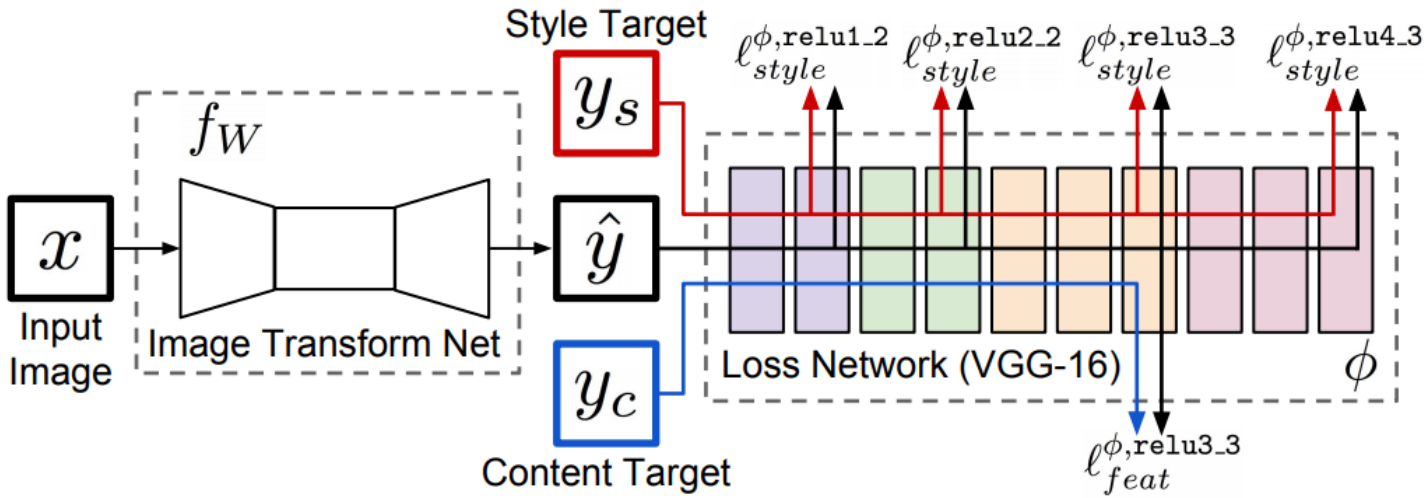
$$\ell_{feat}^{\emptyset,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j}\left\|\emptyset_j(\hat{y}) - \emptyset_j(y)\right\|_2^2$$

Style Reconstruction Loss

$$G_j^{\emptyset}(x)_{c,c'} = \frac{1}{C_j H_j W_j}\sum_{h=1}^{H_j}\sum_{w=1}^{W_j} \emptyset_j(x)_{h,w,c}\emptyset_j(x)_{h,w,c'}$$

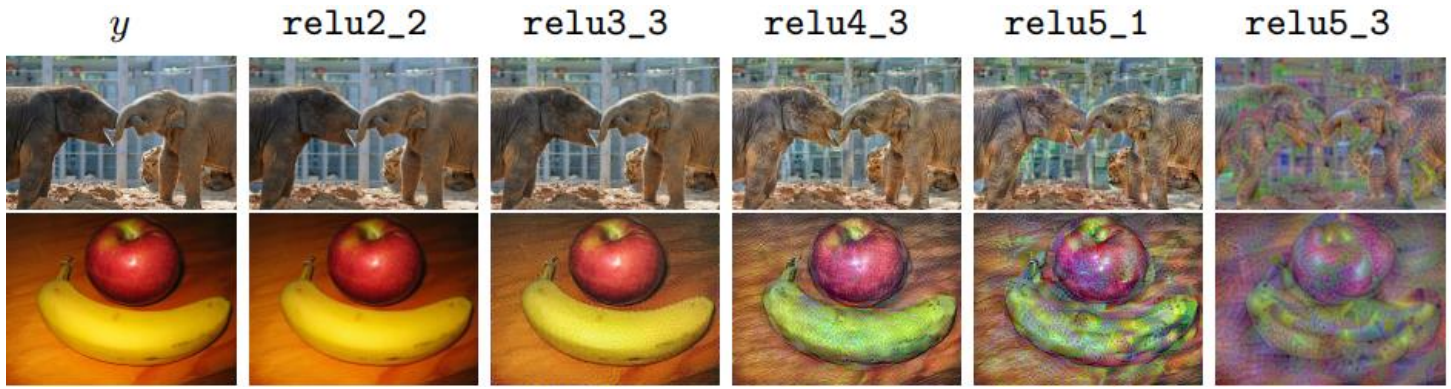
$$\ell_{feat}^{\emptyset,j}(\hat{y}, y) = \left\|G_j^{\emptyset}(\hat{y}) - G_j^{\emptyset}(y)\right\|_F^2$$

3 Perceptual Loss for Style Transfer and Super Resolution



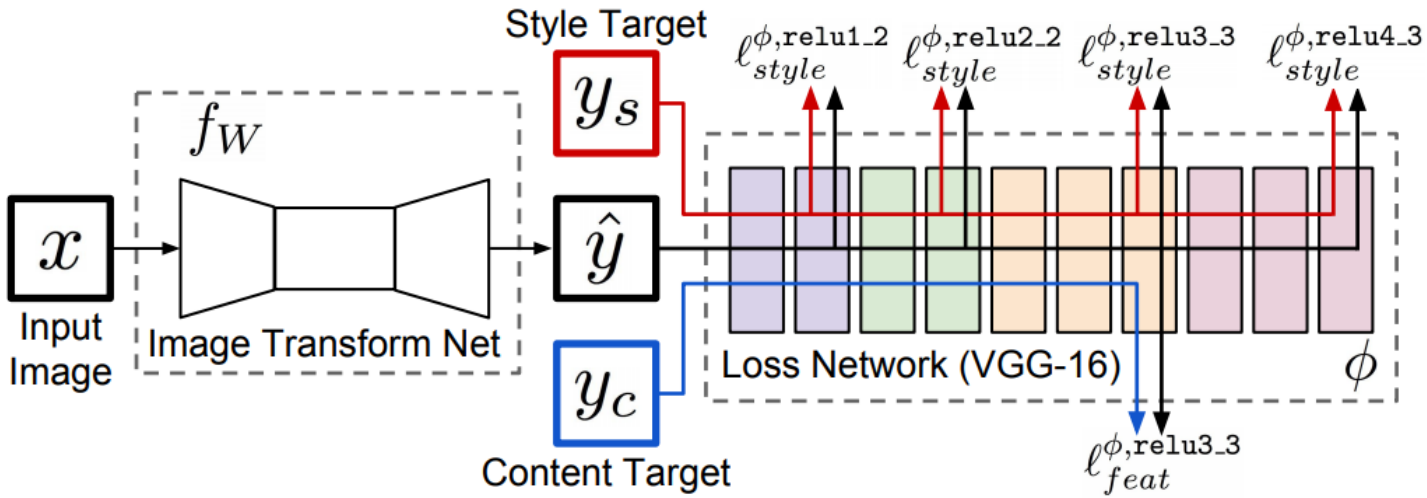
Feature Reconstruction Loss

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \left\| \phi_j(\hat{y}) - \phi_j(y) \right\|_2^2$$



C : channel
H : Height
W : Width

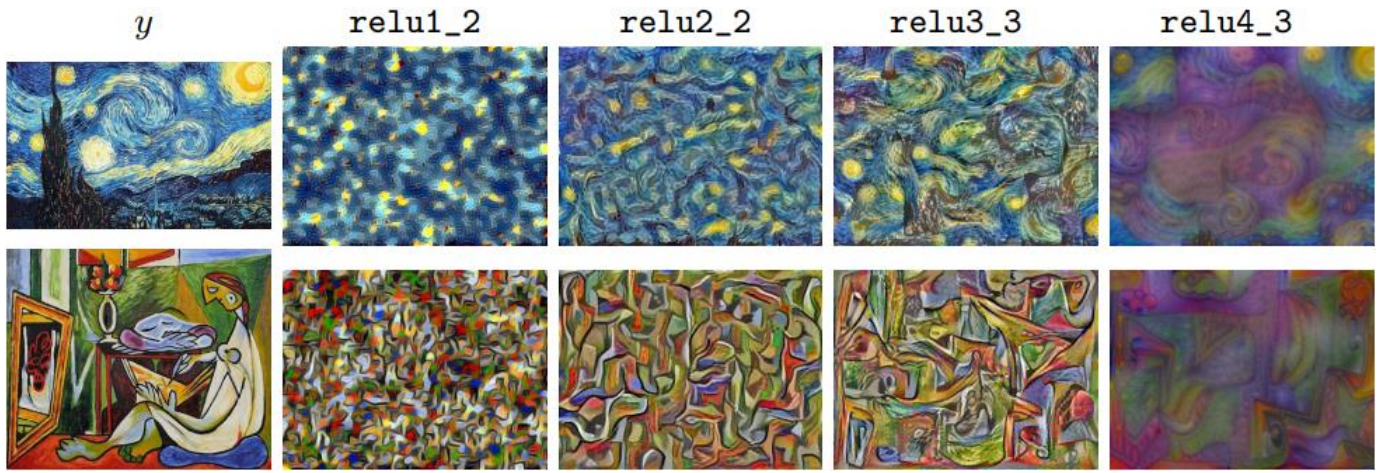
3 Perceptual Loss for Style Transfer and Super Resolution



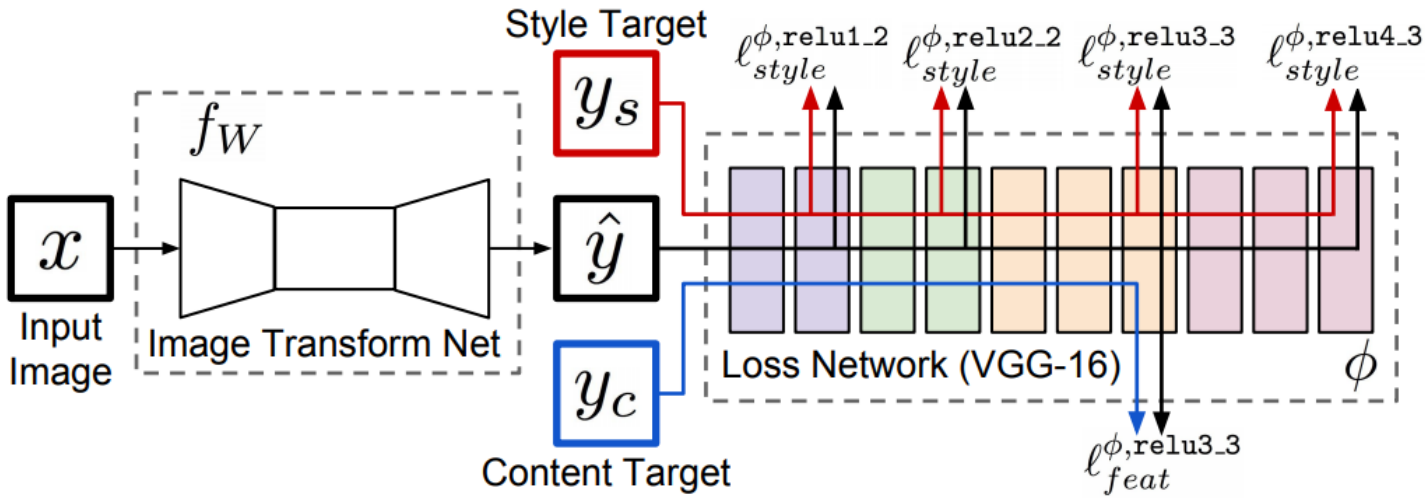
Style Reconstruction Loss

$$G_j^{\emptyset}(x)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \emptyset_j(x)_{h,w,c} \emptyset_j(x)_{h,w,c'}$$

$$\ell_{feat}^{\emptyset,j}(\hat{y}, y) = \left\| G_j^{\emptyset}(\hat{y}) - G_j^{\emptyset}(y) \right\|_F^2$$



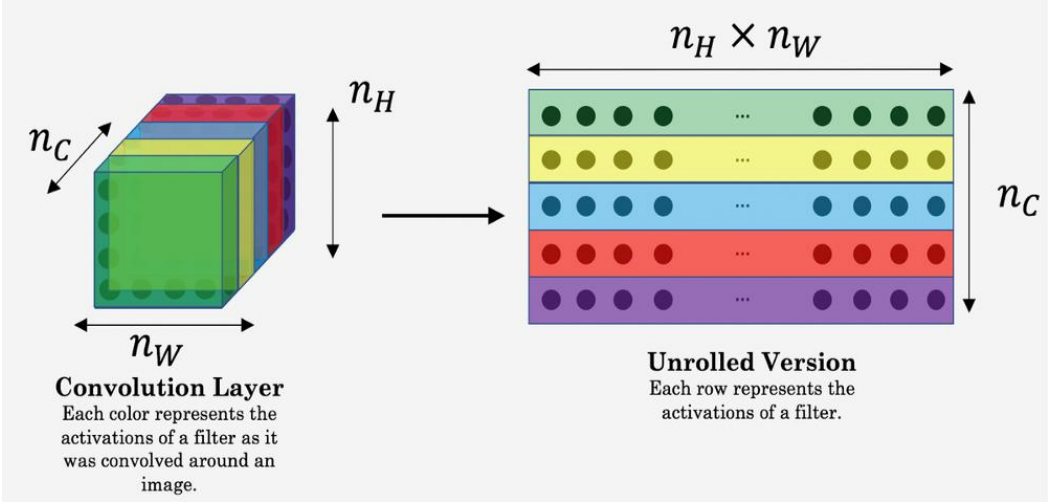
3 Perceptual Loss for Style Transfer and Super Resolution

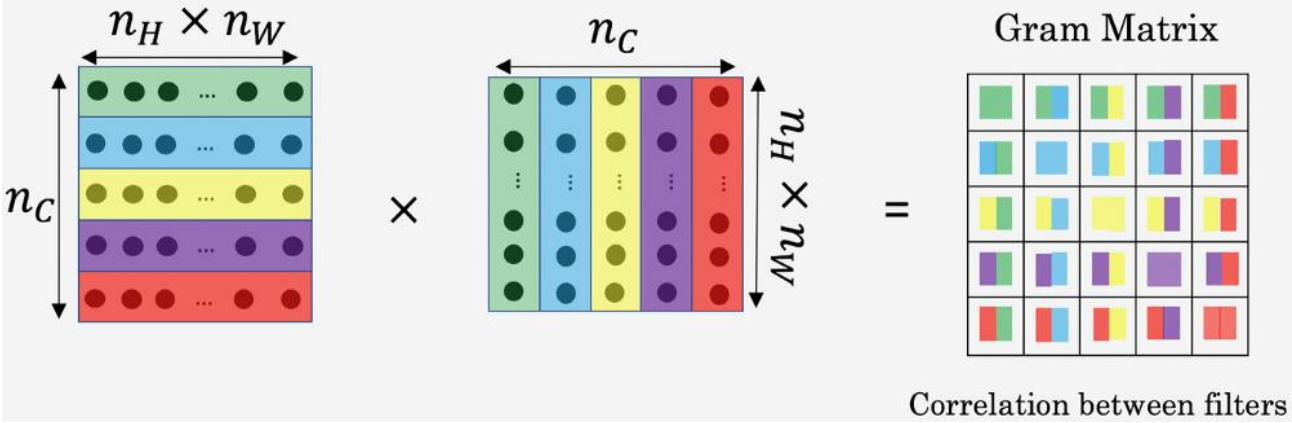


Style Reconstruction Loss

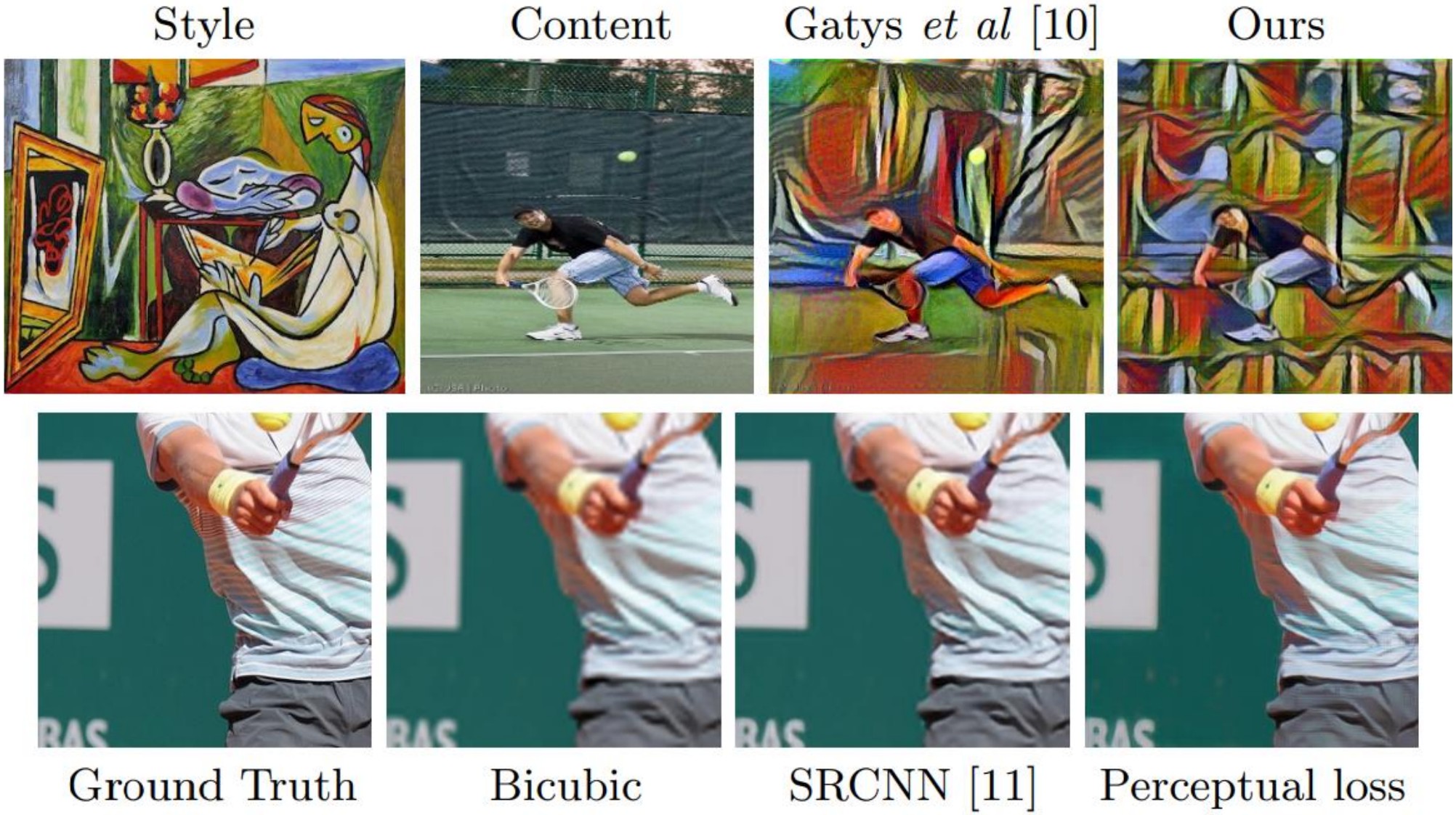
$$G_j^{\emptyset}(x)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \emptyset_j(x)_{h,w,c} \emptyset_j(x)_{h,w,c'}$$

$$\ell_{feat}^{\emptyset,j}(\hat{y}, y) = \left\| G_j^{\emptyset}(\hat{y}) - G_j^{\emptyset}(y) \right\|_F^2$$





3 Perceptual Loss for Style Transfer and Super Resolution



4 Result

Image Size	Gatys *et al* [10] 100	300	500	Ours	Speedup 100	300	500
256×256	3.17	9.52s	15.86s	**0.015s**	212x	636x	**1060x**
512×512	10.97	32.91s	54.85s	**0.05s**	205x	615x	**1026x**
1024×1024	42.89	128.66s	214.44s	**0.21s**	208x	625x	**1042x**

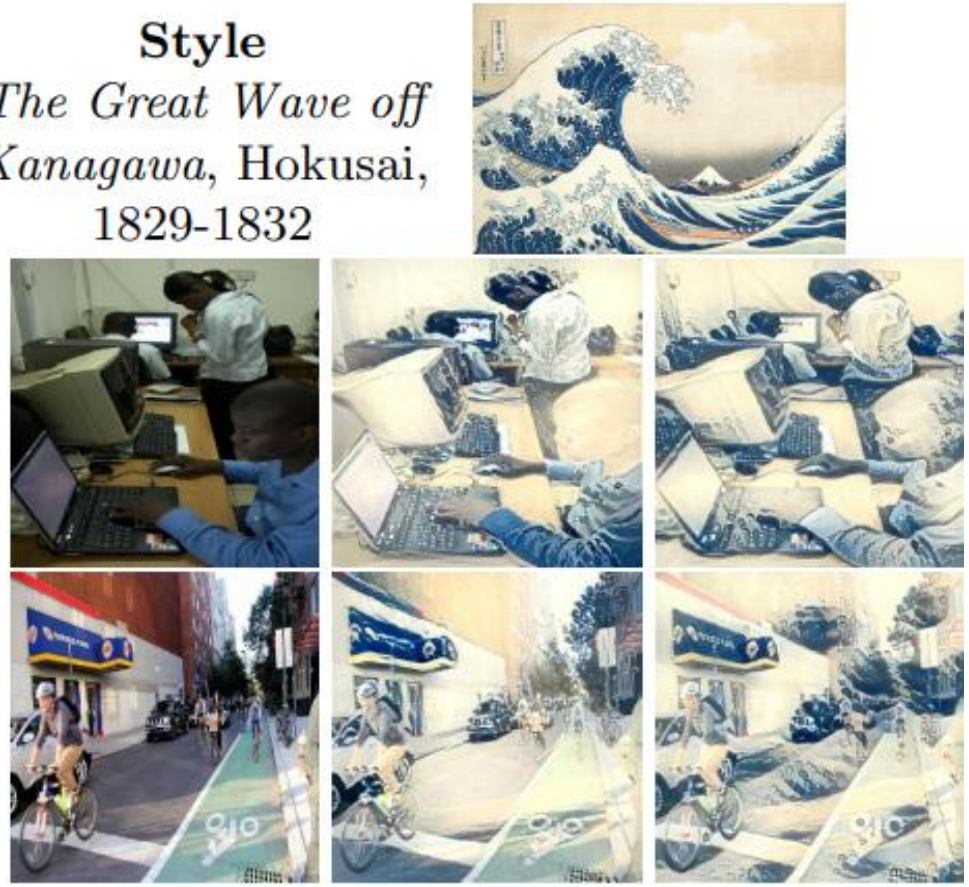
4 Result

Style
The Starry Night,
Vincent van Gogh,
1889

Style
The Muse,
Pablo Picasso,
1935

4 Result


Style
Composition VII,
Wassily
Kandinsky, 1913

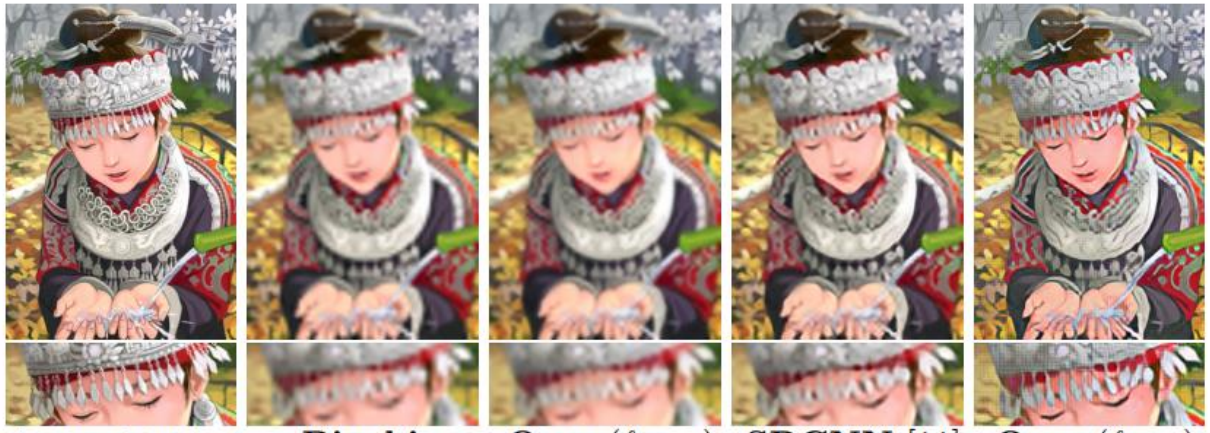

Style
The Great Wave off
Kanagawa, Hokusai,
1829-1832

4 Result

4 Result

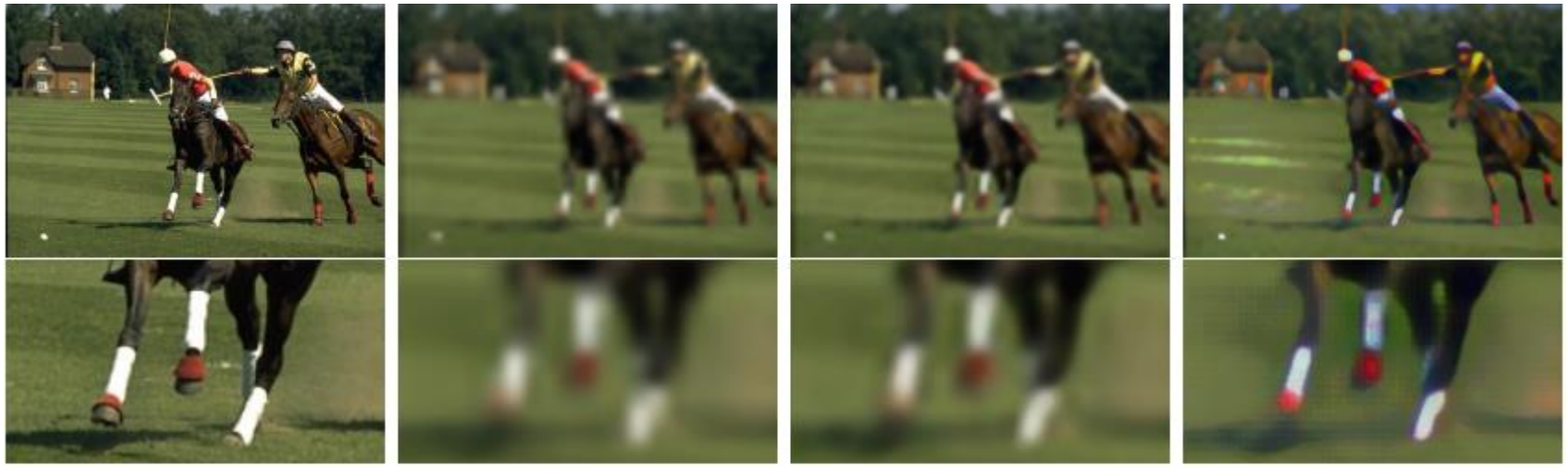
4 Result

Ground Truth	Bicubic	Ours (ℓ_{pixel})	SRCNN [11]	Ours (ℓ_{feat})
This image	31.78 / 0.8577	31.47 / 0.8573	32.99 / 0.8784	29.24 / 0.7841
Set5 mean	28.43 / 0.8114	28.40 / 0.8205	30.48 / 0.8628	27.09 / 0.7680



Ground Truth	Bicubic	Ours (ℓ_{pixel})	SRCNN [11]	Ours (ℓ_{feat})
This Image	21.69 / 0.5840	21.66 / 0.5881	22.53 / 0.6524	21.04 / 0.6116
Set14 mean	25.99 / 0.7301	25.75 / 0.6994	27.49 / 0.7503	24.99 / 0.6731
BSD100 mean	25.96 / 0.682	25.91 / 0.6680	26.90 / 0.7101	24.95 / 63.17

4 Result



Ground Truth	Bicubic	Ours (ℓ_{pixel})	Ours (ℓ_{feat})
This image	22.75 / 0.5946	23.42 / 0.6168	21.90 / 0.6083
Set5 mean	23.80 / 0.6455	24.77 / 0.6864	23.26 / 0.7058
Set14 mean	22.37 / 0.5518	23.02 / 0.5787	21.64 / 0.5837
BSD100 mean	22.11 / 0.5322	22.54 / 0.5526	21.35 / 0.5474

Reference

- https://arxiv.org/pdf/1603.08155.pdf
- https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf
- https://arxiv.org/pdf/1508.06576.pdf
- https://blog.lunit.io/2017/04/27/style-transfer/
- https://www.popit.kr/neural-style-transfer-%EB%94%B0%EB%9D%BC%ED%95%98%EA%B8%B0/
- https://hoya012.github.io/blog/Fast-Style-Transfer-Tutorial/