Microsoft

Neural Style Transfer for Images and Videos

Jing Liao(Visual Computing Group, MSRA)

worked with Lu Yuan, Gang Hua, Sing Bing Kang，Dongdong Chen*，Yuan Yao*，(*: interns)

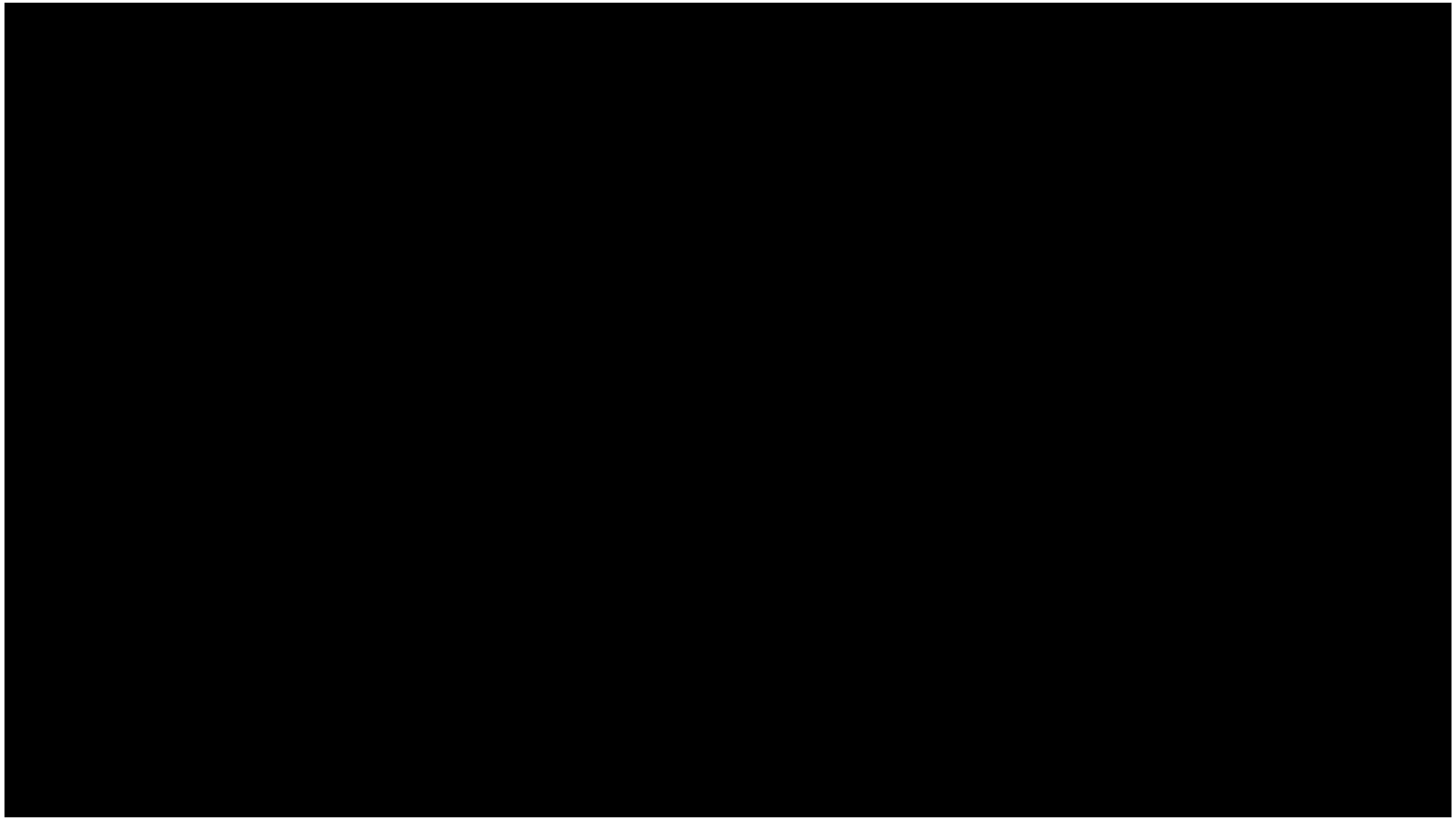
Style Transfer

Transfer the artistic style of a painting to another image?

e.g. combining an image with Vincent van Gogh's *The Starry Night*.

Style Transfer

Manual simulation: “Loving Vincent” has **65,000** individual frames painted by **125** artists, took **6** years.



Style Transfer

Automatic simulation: Traditional NPR methods.

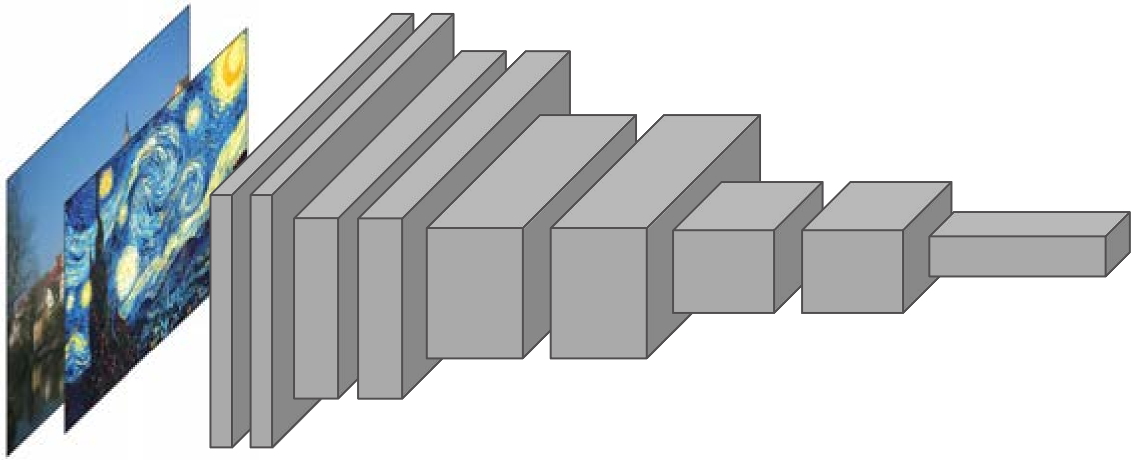
Cartoon
[Winnemoller et al. 2006] SIGGRAPH

Oil Painting
[Zeng et al. 2009] ACM TOG

Pencil drawing
[Lu et al. 2012] NPAR

Style Transfer

Automatic simulation: Deep neural network methods.



Success in market:

Prisma, Pikazo, Lucid, Painnt, Artisto, Icon8, DeepArt, Malevich, Ostagram

[Gatys et al. 2015], [Li & Wang 2016], [Ulyanov et al. 2016], [Johnson et al. 2016], [Dumoulin et al. 2016]

Style Transfer by Convolutional Neural Networks

[Gatys et al. 2015]

$$\underset{I}{\operatorname{argmin}} L(I, content, style)$$

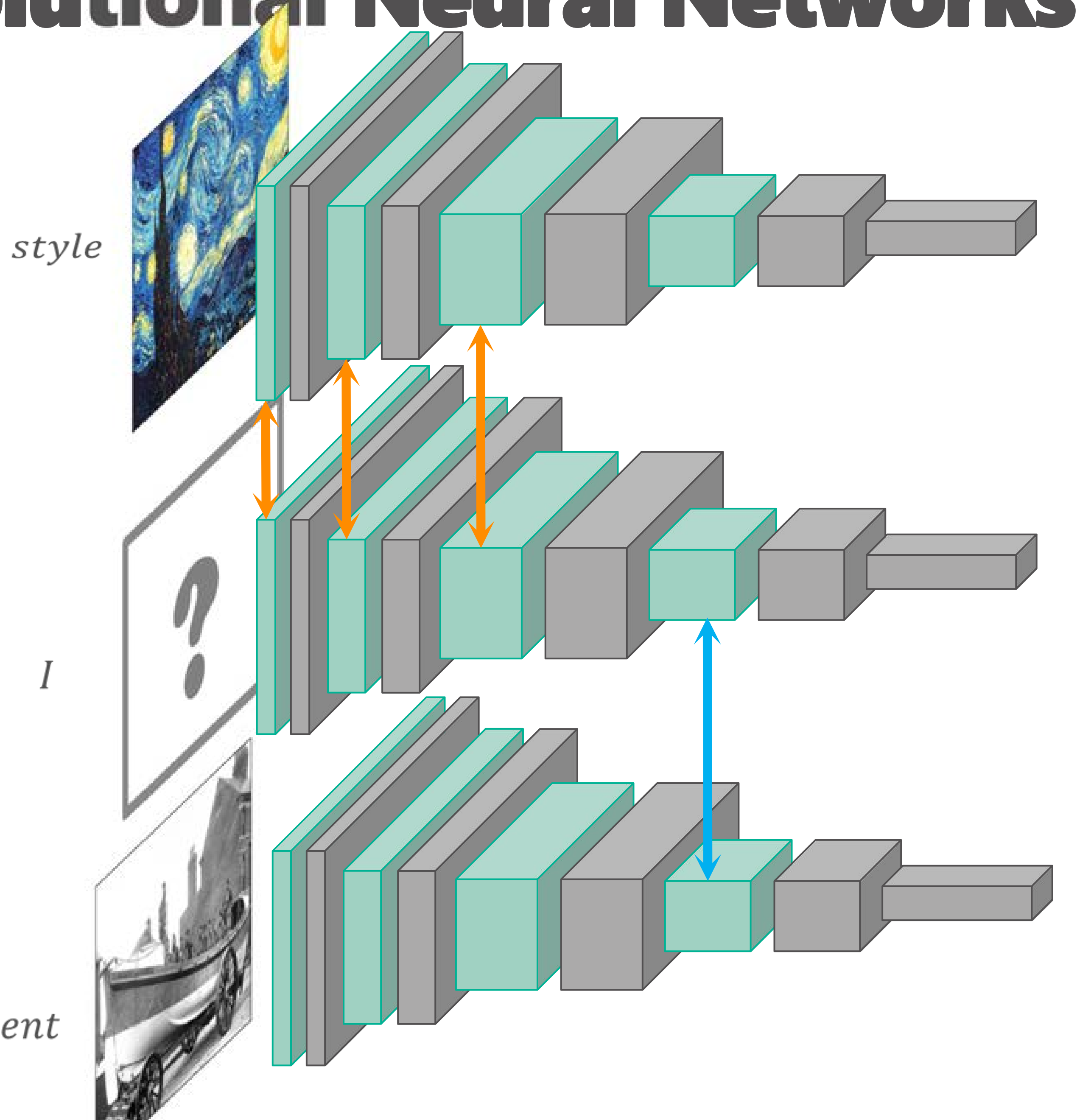
$$= \underset{I}{\operatorname{argmin}}(\alpha L_{content} + \beta L_{style})$$

High-level

L2 dis between features

Low-level

L2 dis between Features Gram matrix



Style Transfer by Convolutional Neural Networks

Limitations of [Gatys et al. 2015]

1. Slow： requires hundreds of forward and backward passes through the CNN

   StyleBank [CVPR 2017]

2. Temporal incoherent： flickering artifacts

   Coherent Video Style Transfer [ICCV 2017]

3. Local incorrectness：

   Deep Image Analogy [Siggraph 2017]

StyleBank: An Explicit Representation for Neural Image Style Transfer

CVPR 2017

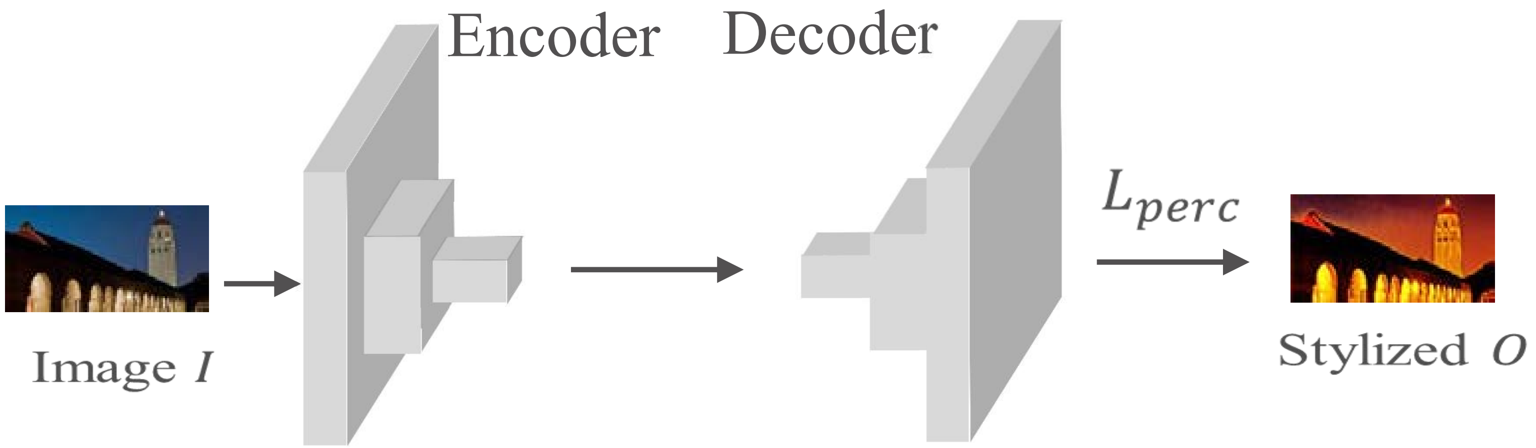
Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, Gang Hua

+

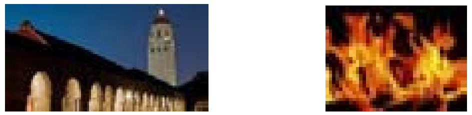
=

Feed-forward Baseline vs. StyleBank

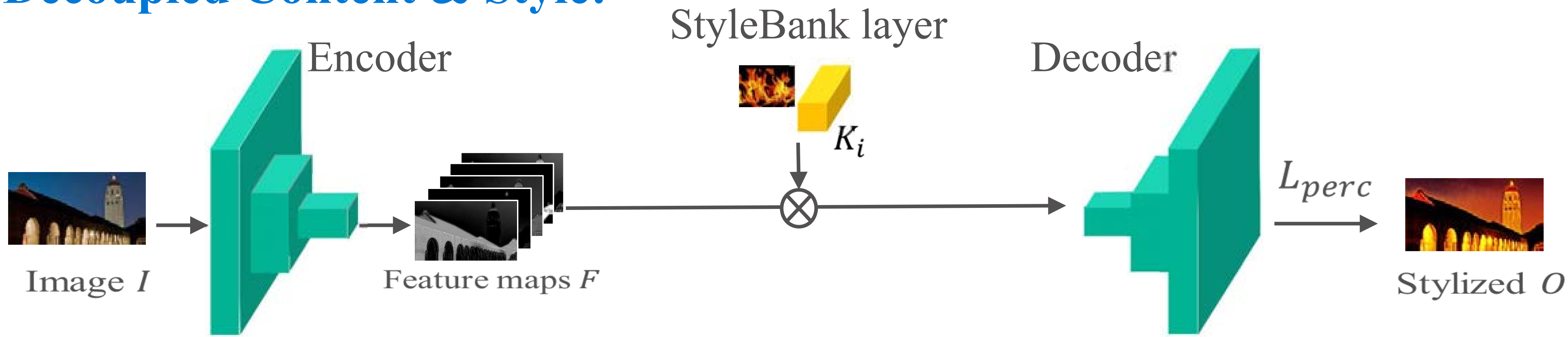
- **Coupled Content & Style:**



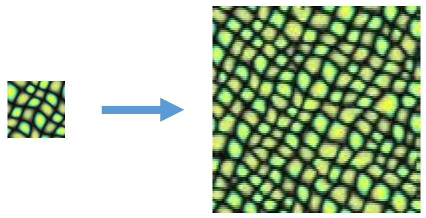
$$L_{perc} = L_{content} + \lambda L_{style}$$



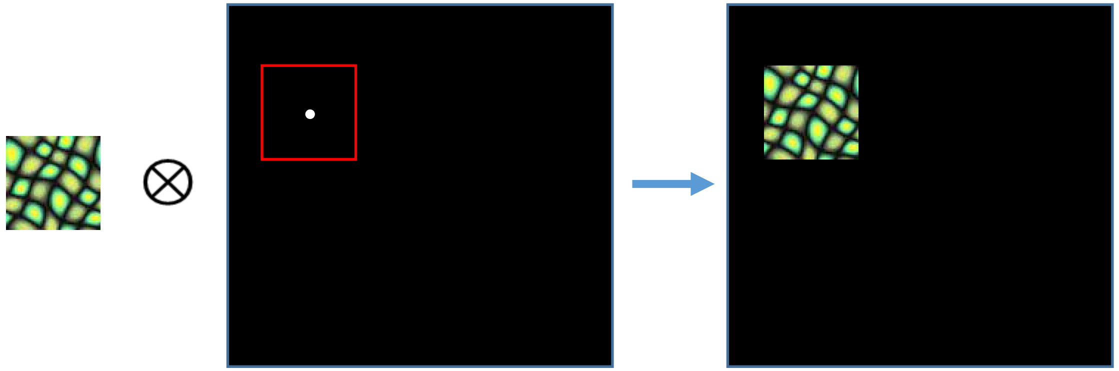
- **Decoupled Content & Style:**



Motivation

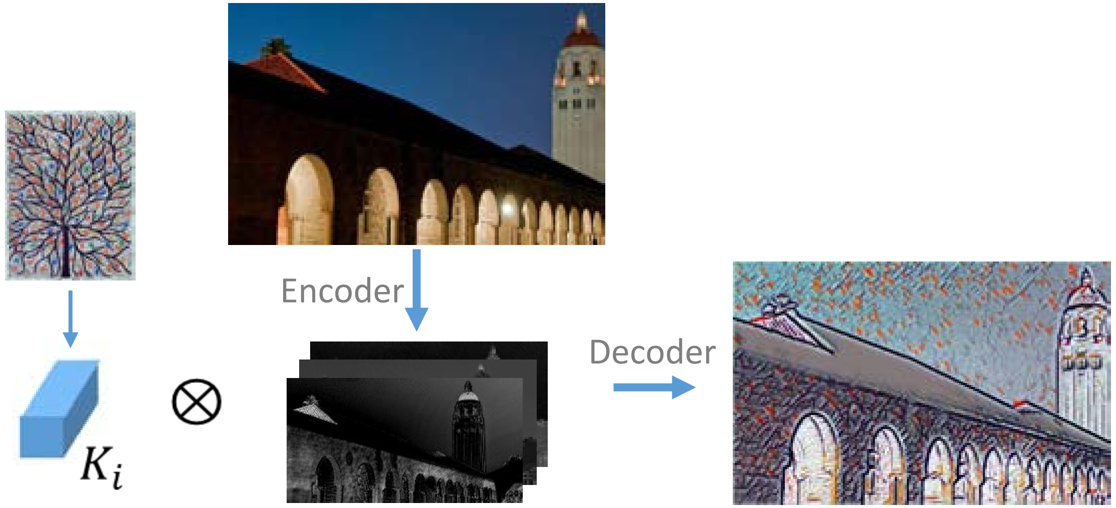


Texture synthesis can be considered as a convolution between *Texton* and sampling function.



Texture Synthesis in Image Space

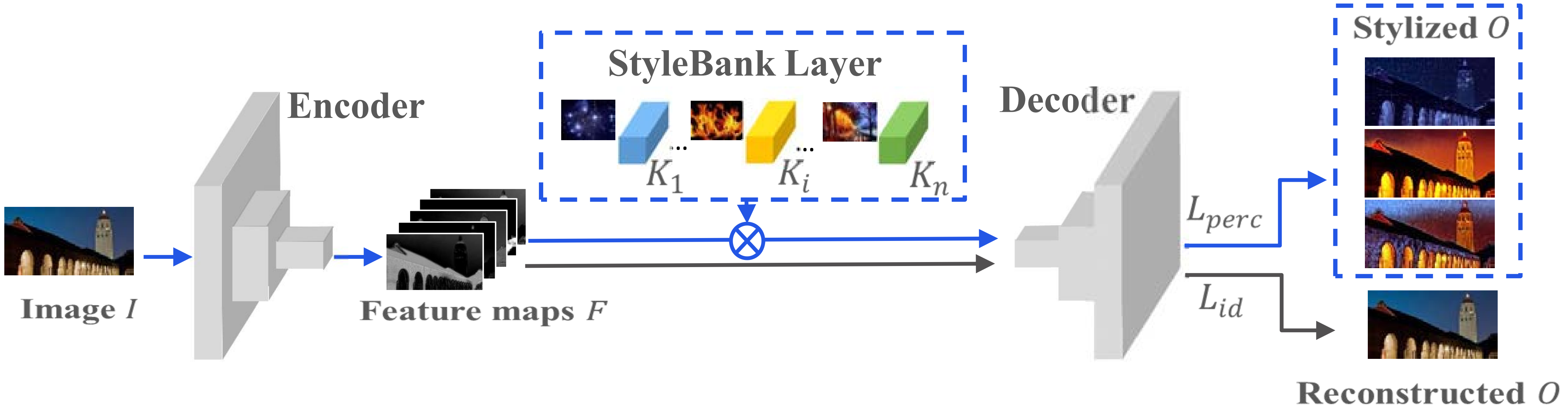
Can this idea can be applied in deep feature space for texture/style transfer?



Texture/Style Transfer in Feature Space

Method

- Stylizing branch: StyleBank (only style)
- Auto-encoder branch (only content)



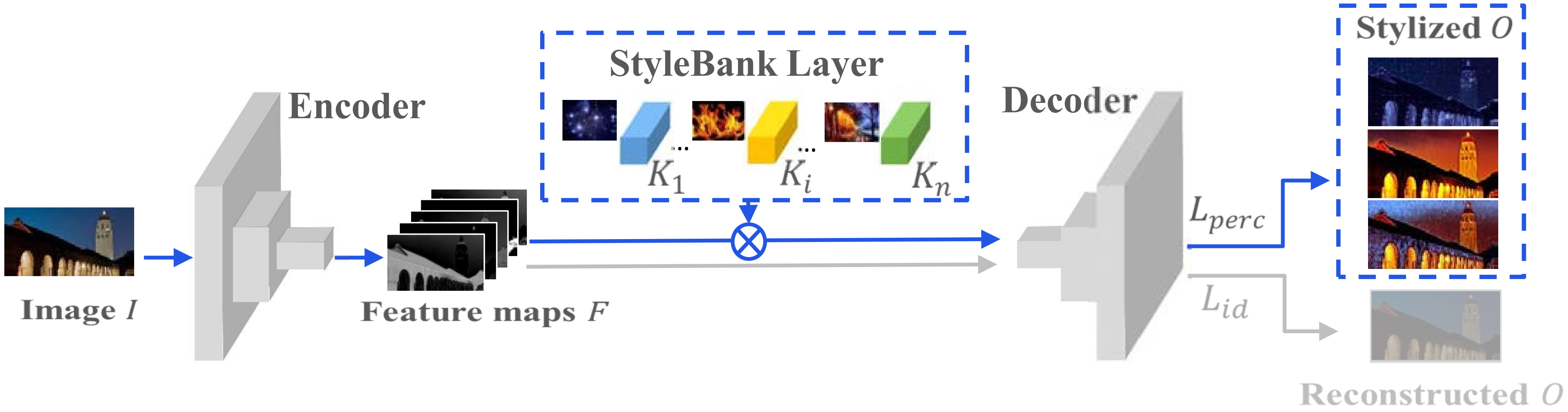
$$L_{id} = \|O - I\|^2$$

$$L_{perc} = \alpha L_{content} + \beta L_{style}$$

two branches share the same encoder & decoder

Method

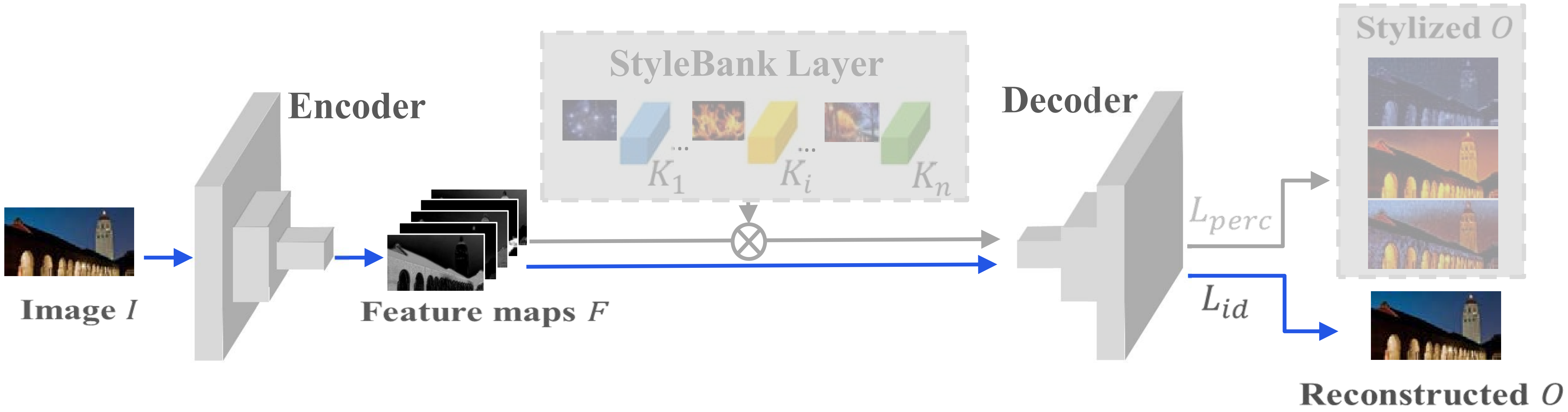
- “T+1” Training Strategy



T iterations for stylizing branch

Method

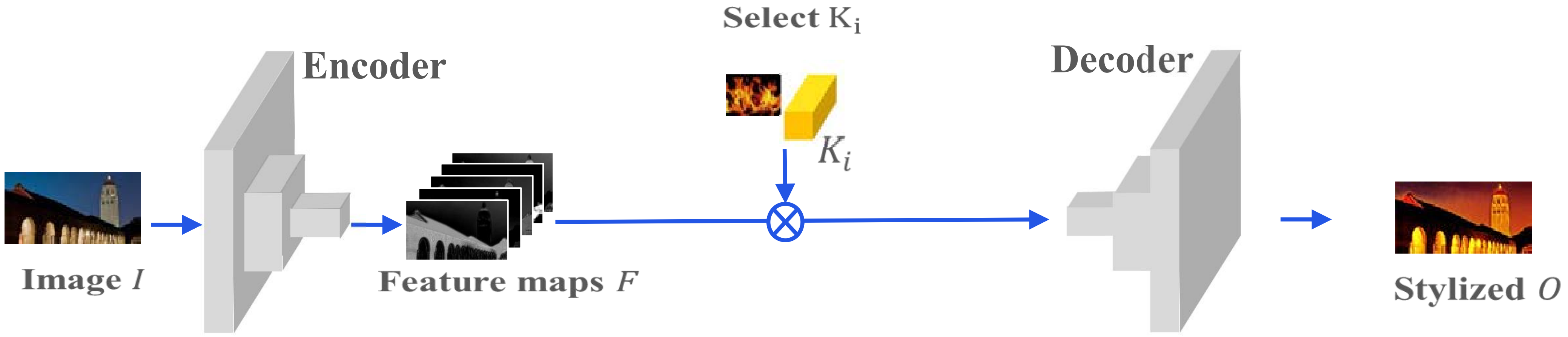
- "T+1" Training Strategy



1 iterations for auto-encoder branch

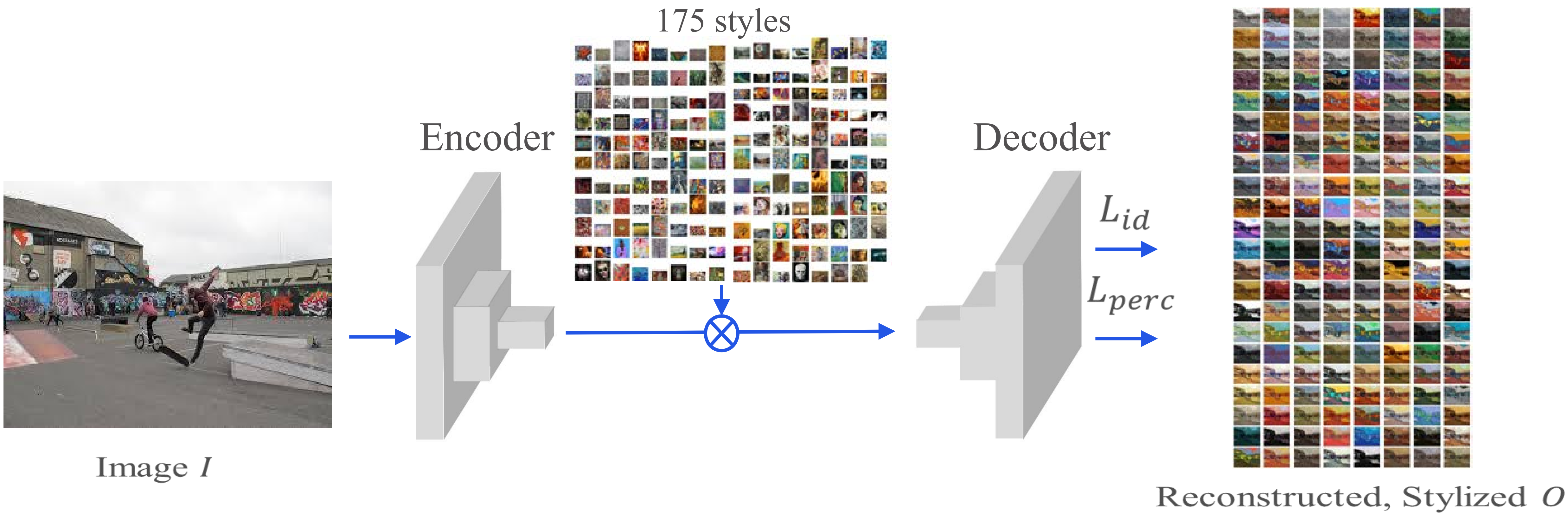
Method

- Test Strategy



Advantages:

1. Simultaneously learn multiple styles in one network

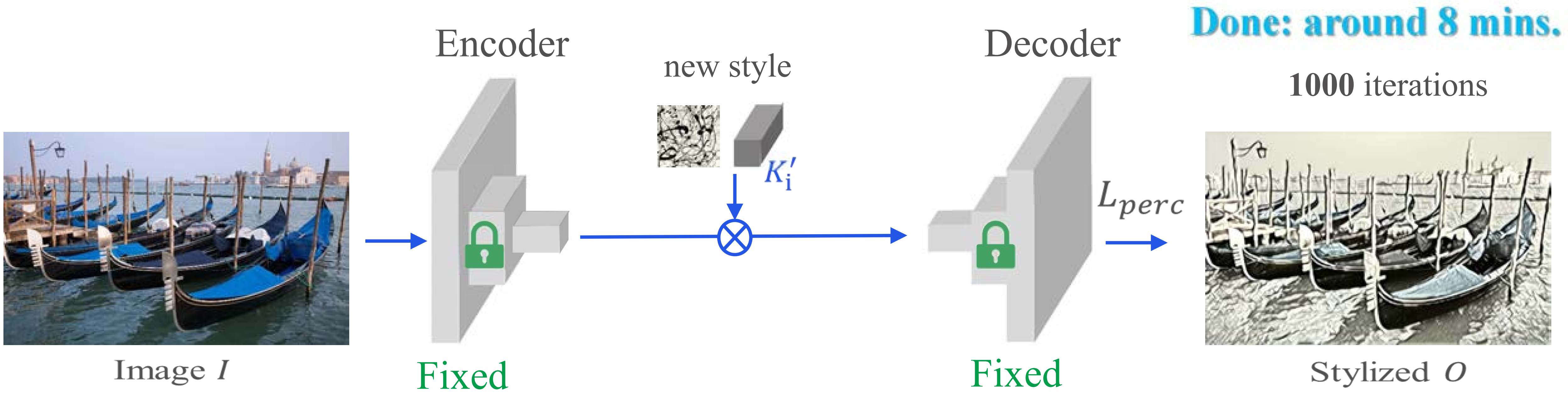


[Johnson et al. 2016]: 800 hours, 1,120 Mbytes
Ours: 36 hours, 120 Mbytes

Results

Advantages:

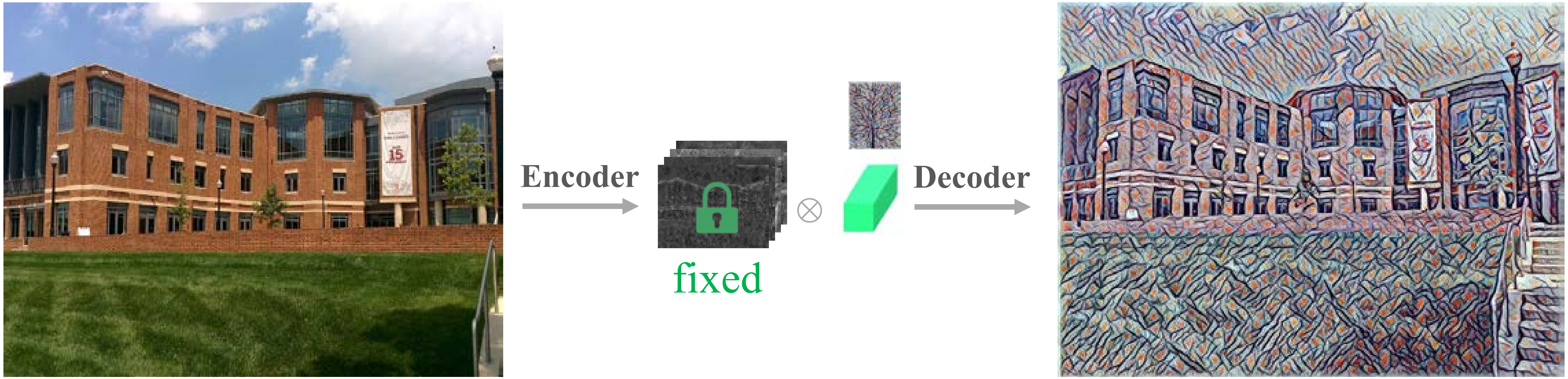
2. Faster training for new styles: only learn StyleBank layer



Feedforward nets [Johnson et al. 2016]: 4 ~ 5 hours
Ours: 8 mins

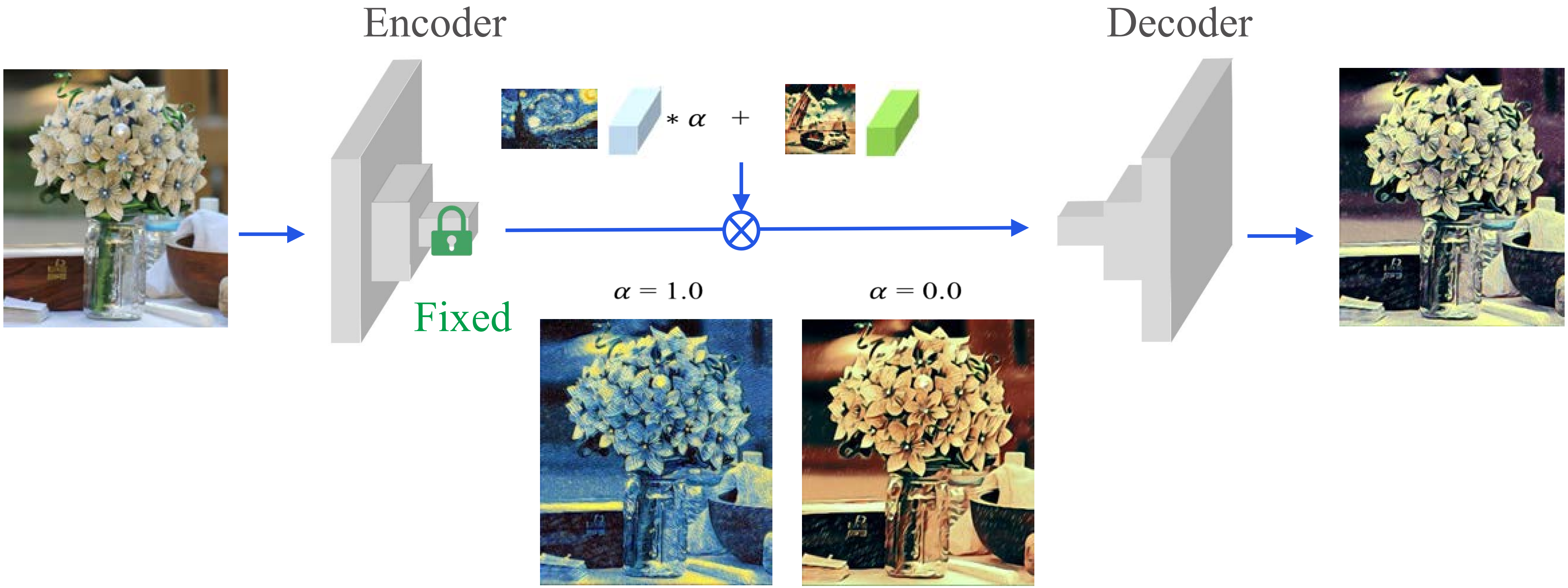
Advantages:

3. Faster synthesis in switching various styles



Advantages:

4. Style fusion: linear fusion of style filter banks



Advantages:

4. Style fusion: linear fusion of style filter banks

Coherent Online Video Style Transfer

ICCV 2017

Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, Gang Hua

+

=

Per-frame Method *vs.* Our Method

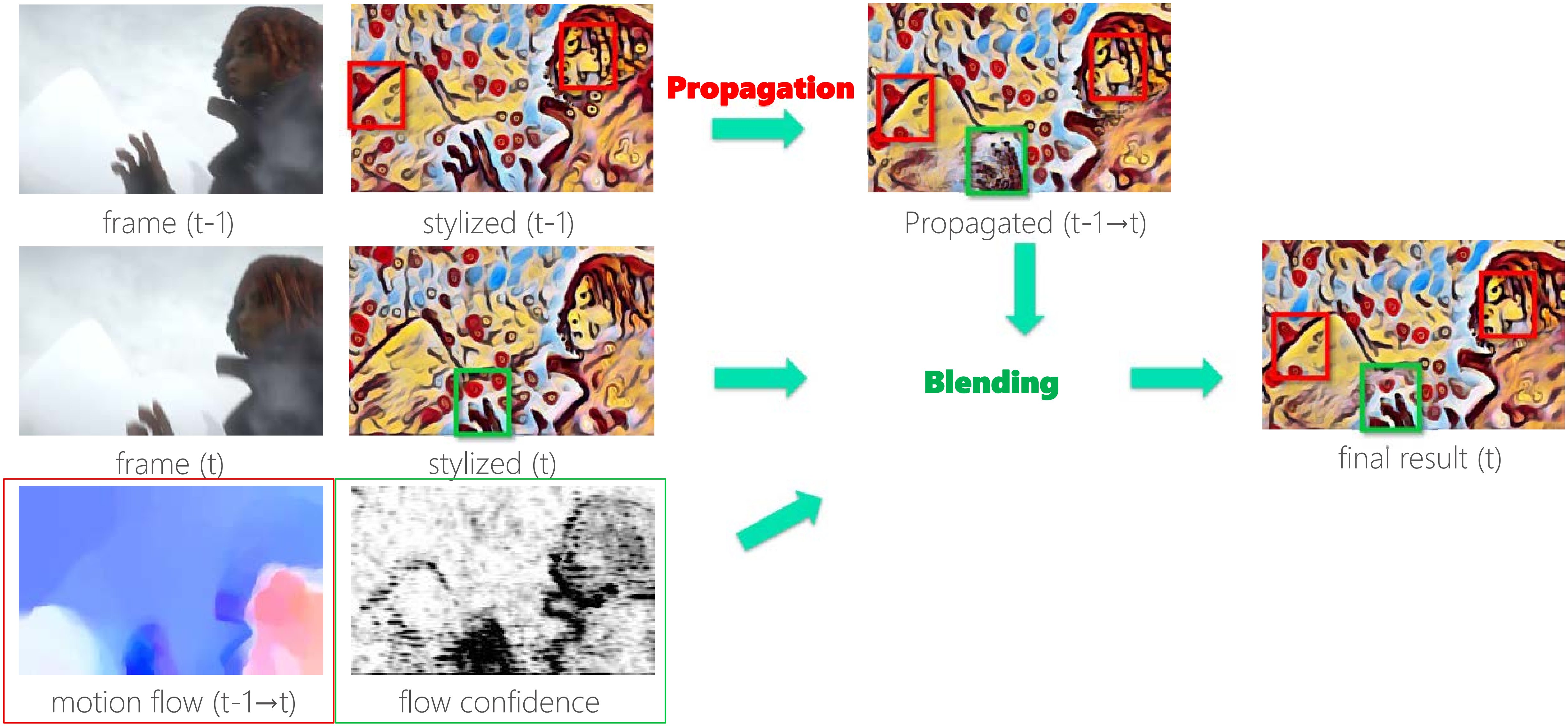
input:

+

per-frame [Johnson et al. 2016]

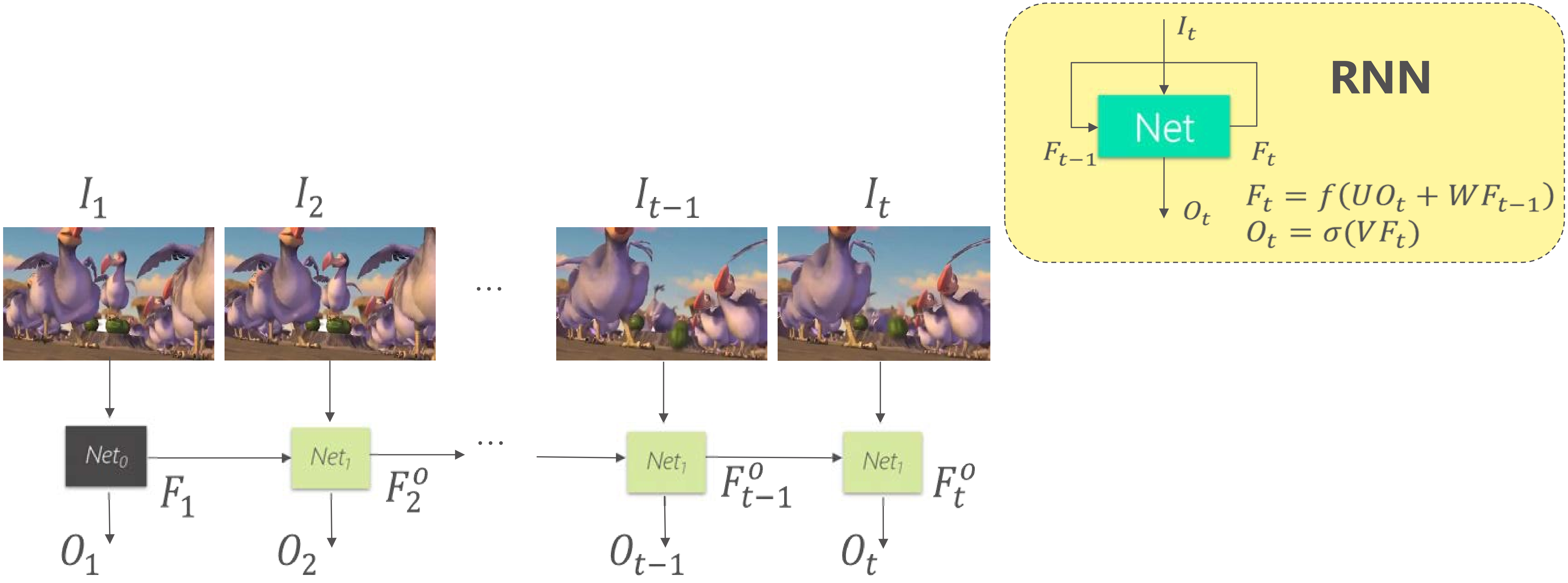
our method (online processing)

Our Idea: Propagation + Blending



Our Method:

Short-term consistency approximates long-term consistency by propagation



Comparisons

input:

+

per-frame [Johnson et al. 2016]

our method

Comparisons

input:

+

per-frame (StyleBank) [CVPR 2017]

our method (StyleBank)

Comparisons

input:

+

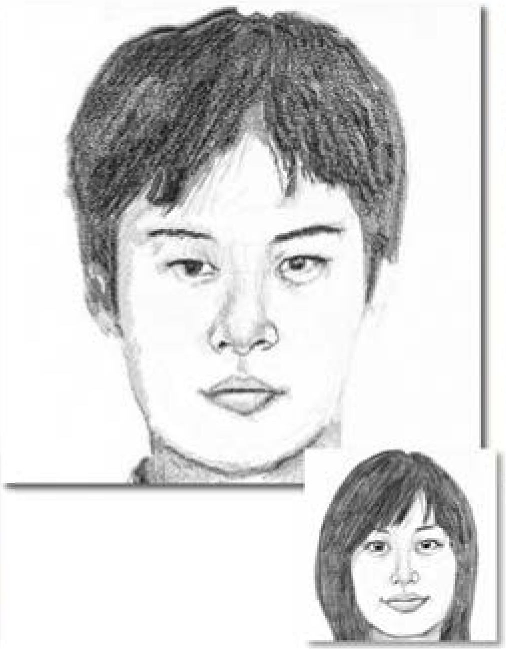
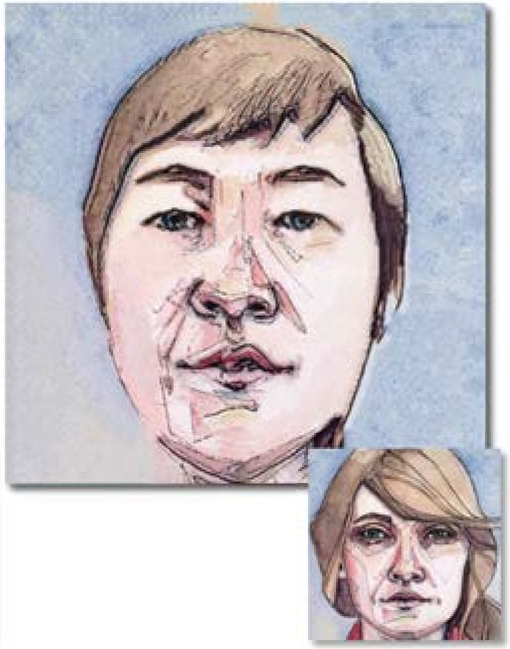
global optimization [Ruder et al. 2015]

our method

Transfer Visual Attribute Transfer through Deep Image Analogy

Siggraph 2017

Jing Liao　Yuan Yao　Lu Yuan　Gang Hua　Sing Bing Kang

Three Generations of Neural Style Transfer

Source

Reference

Global Statistics

Gatys et al. [2015]

App: Ostagram
Deep Style

Fast Approximation

Johnson et al. [2016]

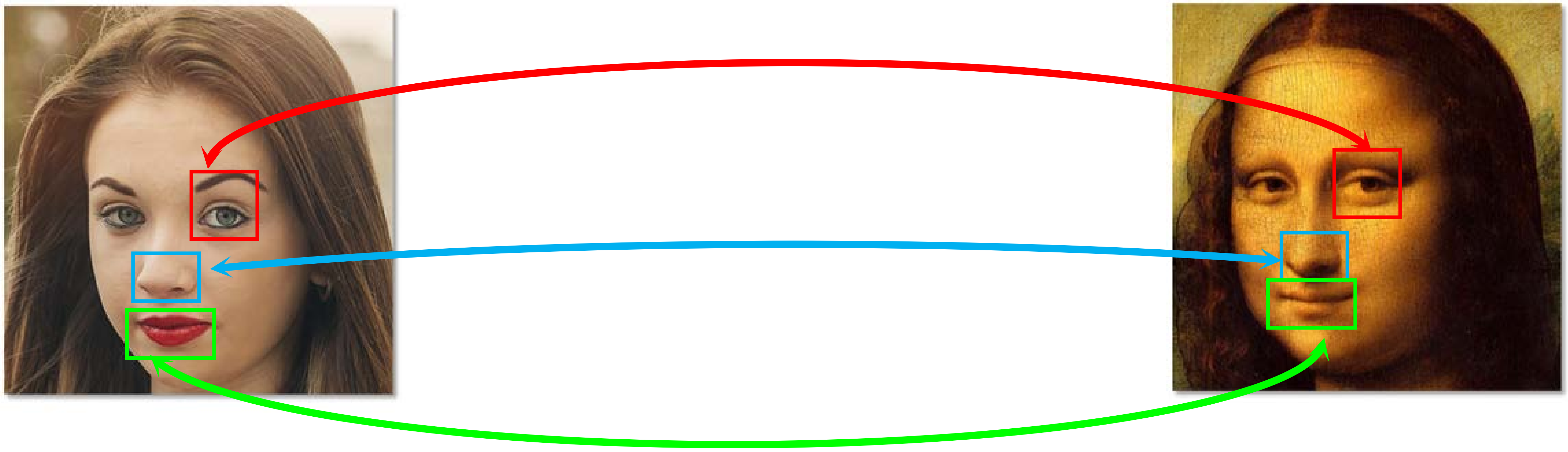
App: Prisma

Local Semantics

Ours

Core Problem in Style Transfer

Cross-domain matching: **difficult!**



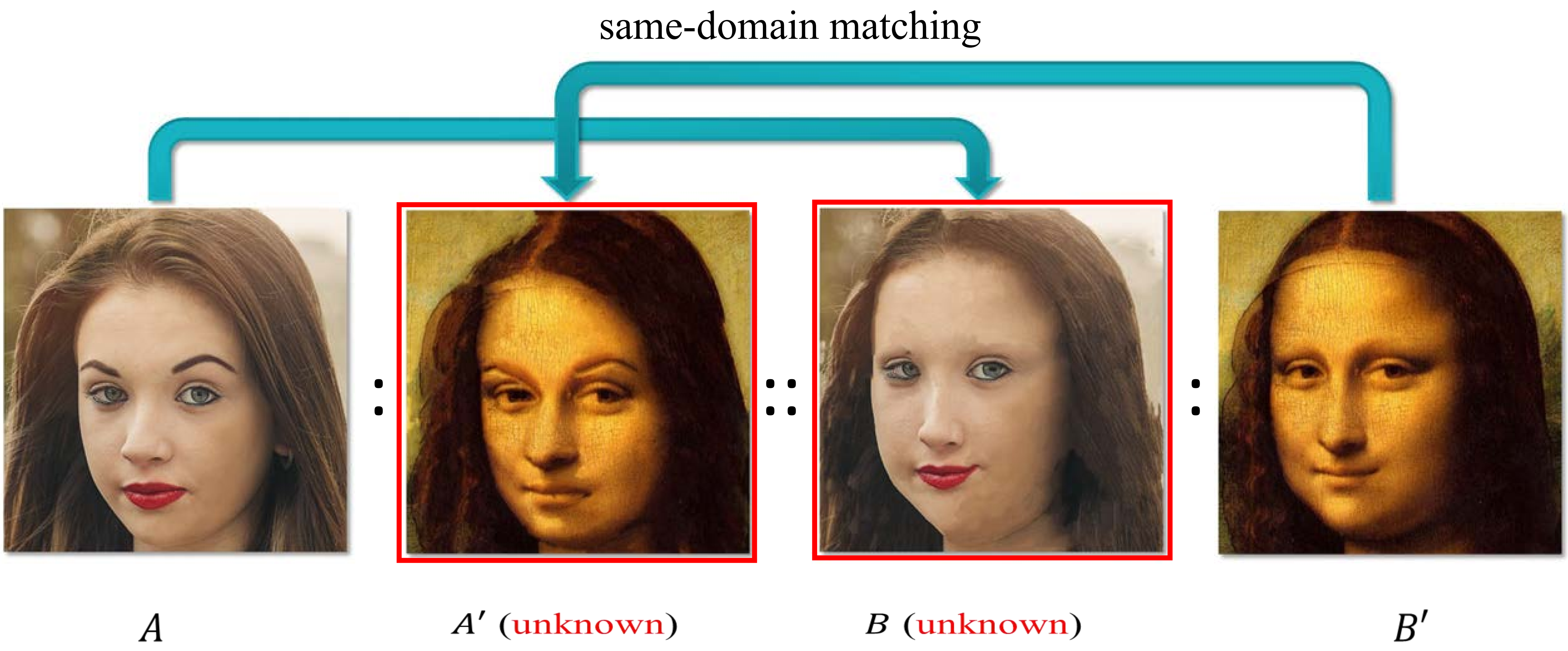
Traditional Image Analogy : Hertzmann et al.[2001]

same-domain matching

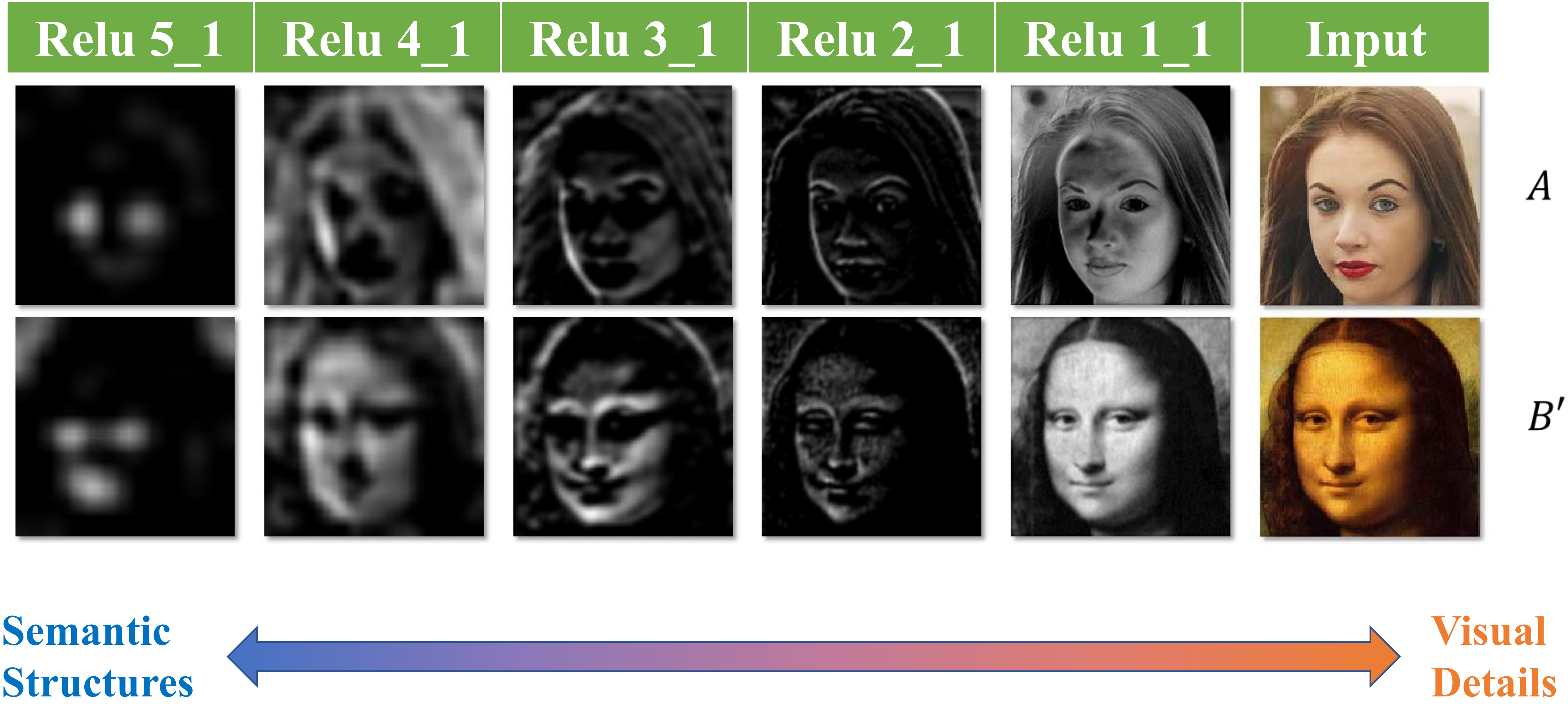
cross-domain matching

A A' (unknown) B B'

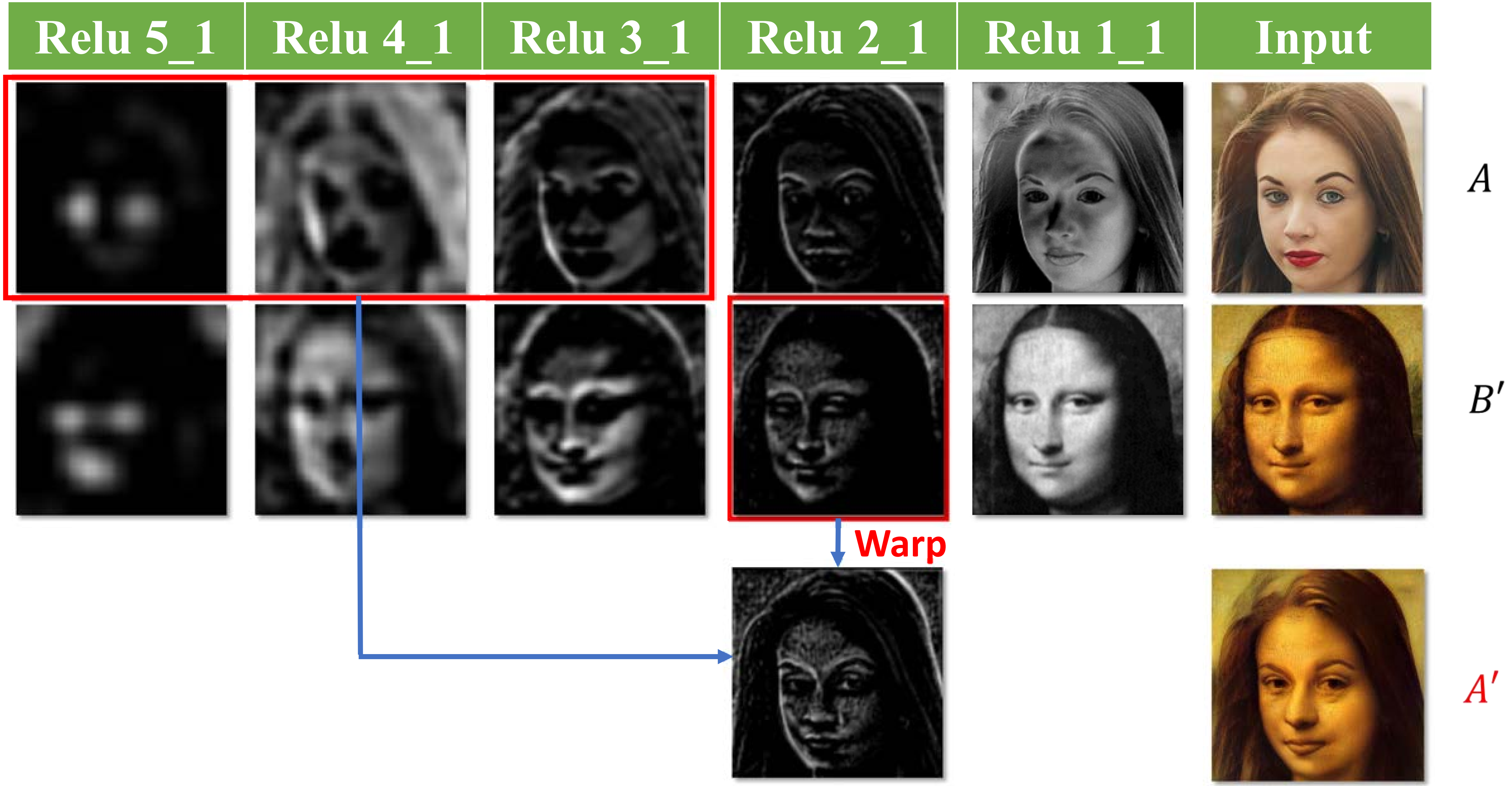
Deep Image Analogy



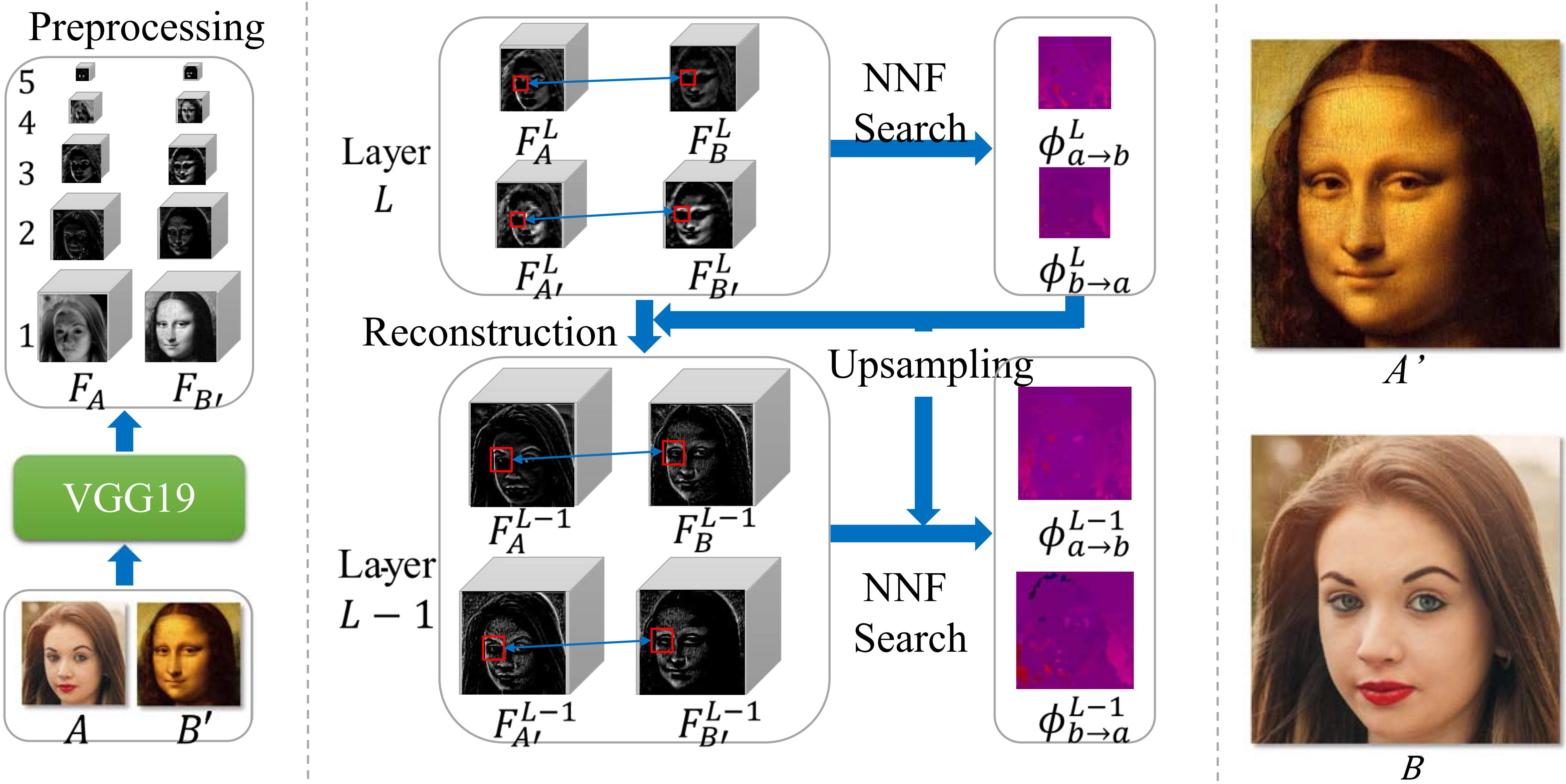
Decouple structures and details with neural networks



Decouple structures and details with neural networks

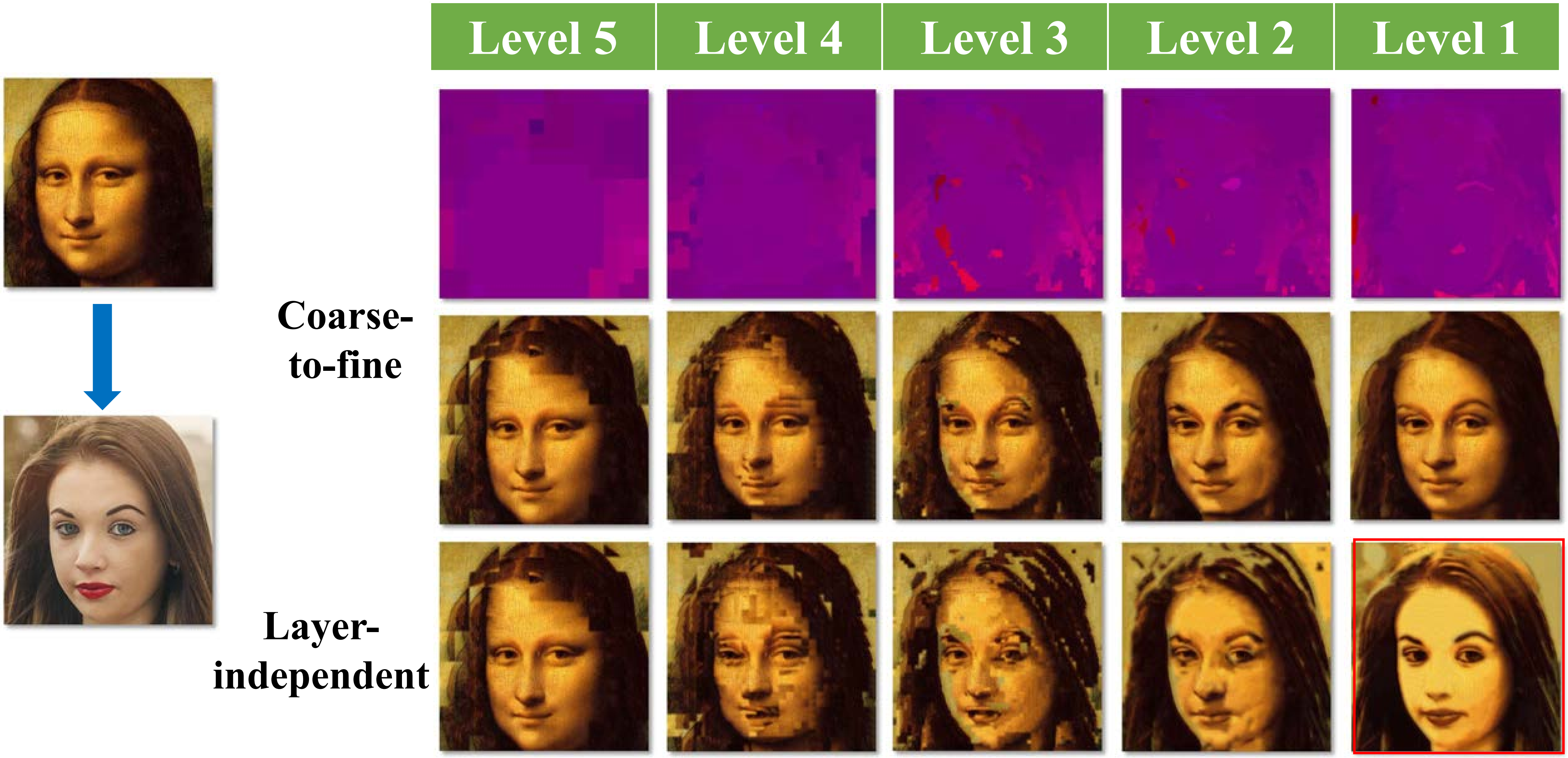


Method



Preprocessing
5
4
3
2
1
F_A
$F_{B'}$
VGG19
A
B'
Layer L
F_A^L
F_B^L
$F_{A'}^L$
$F_{B'}^L$
NNF Search
$\phi_{a\to b}^L$
$\phi_{b\to a}^L$
Reconstruction
Upsampling
Layer $L-1$
F_A^{L-1}
F_B^{L-1}
$F_{A'}^{L-1}$
$F_{B'}^{L-1}$
NNF Search
$\phi_{a\to b}^{L-1}$
$\phi_{b\to a}^{L-1}$
A'
B

Intermediate results



Qualitative Evaluations

- Category 1: same scene with varied views or motions

Input (src)

SIFT flow

DeepFlow2

Input (ref)

PatchMatch

Ours

Qualitative Evaluations

- Category 2: same scene with different colors or tones


Input (src)


SIFT flow


DeepFlow2


PatchMatch


Input (ref)


NRDC


RFM


Ours

Qualitative Evaluations

- Category 3: semantically related but visually different scenes

Input (src)

SIFT flow

DeepFlow2

PatchMatch

Input (ref)

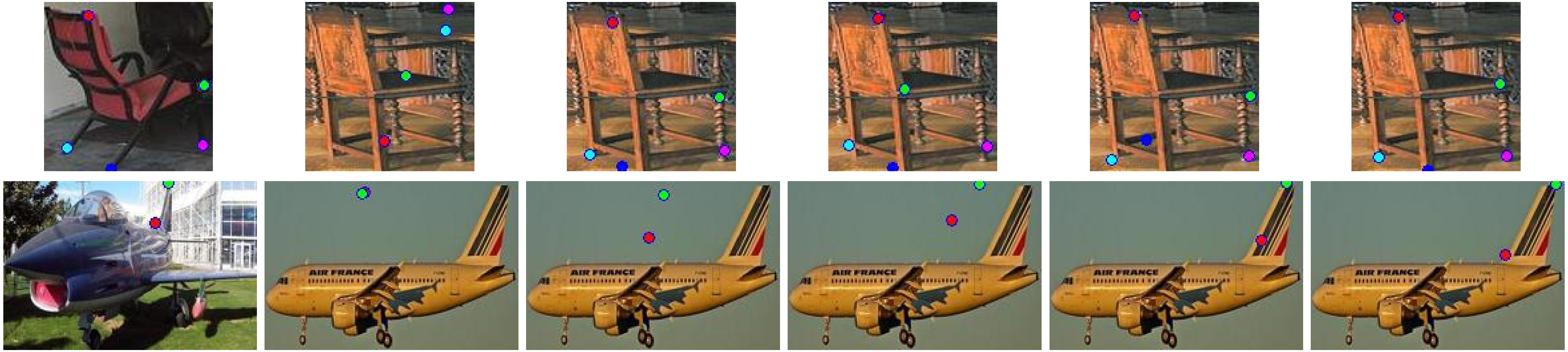
Daisy flow

Halfway

Ours

Quantitative Evaluations

➢ Pascal 3D+ dataset (20 color image pairs for each category, 12 categories):



Source PatchMatch SIFT flow cycle consistent ours reference

	aero	bike	boat	bottle	bus	car	chair	table	mbike	sofa	train	tv	**mean**
PatchMatch (Barnes et al. 2009)	6.5	6.3	2.6	2.9	2.3	4.7	3.3	12.5	2.0	0.0	4.2	2.6	4.2
SIFT Flow (Liu et al. 2011)	8.1	**14.3**	5.1	26.1	25	20.9	13.3	6.3	14.3	15.4	4.2	44.7	16.5
Cycle consistency (Zhou et al. 2016)	12.9	6.3	10.3	**39.1**	27.3	**23.3**	13.3	**12.5**	6.1	**19.2**	**12.5**	36.8	18.3
Ours	**19.4**	7.9	**15.4**	27.5	**47.7**	11.6	**20.0**	6.3	**18.4**	15.4	**12.5**	**50.0**	**21.0**

Table 2. Correspondence accuracy measured in PCK ($\alpha = 0.1$). The test is conducted on randomly selected 20 pairs of each category of PASCAL3D+ dataset.

Results: Photo to Style


Source


Reference


Source


Reference


Output


Output

Results: Photo to Style

Source

Reference

Source

Reference

Source

Reference

Output

Output

Output

Results: Photo to Style

Source

Reference

Source

Reference

Source

Reference

Output

Output

Output

Comparisons

Source

Neural style

Deep style

Ours

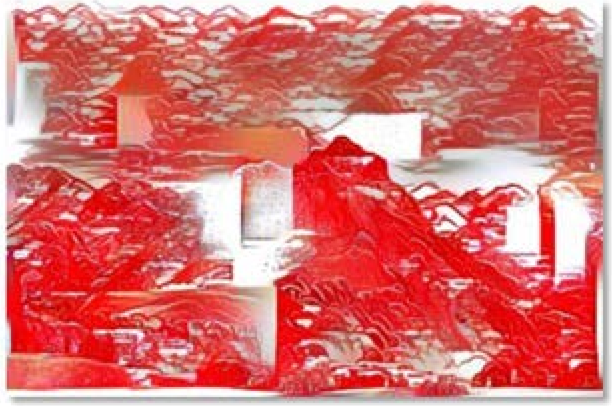
Reference

MRF

Ostagram

Comparisons


source


Neural style


Perceptual loss


Ours

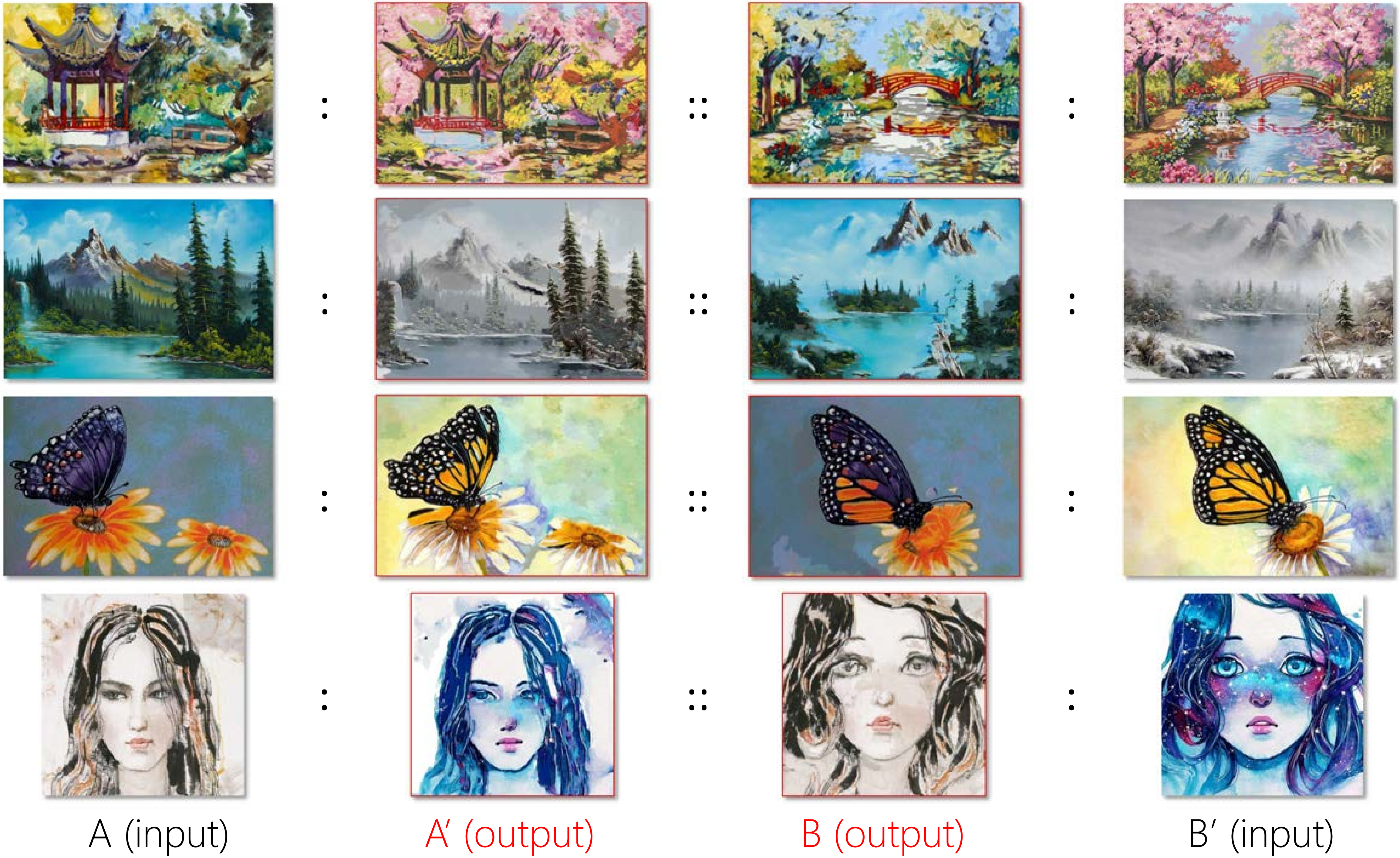

reference


MRF

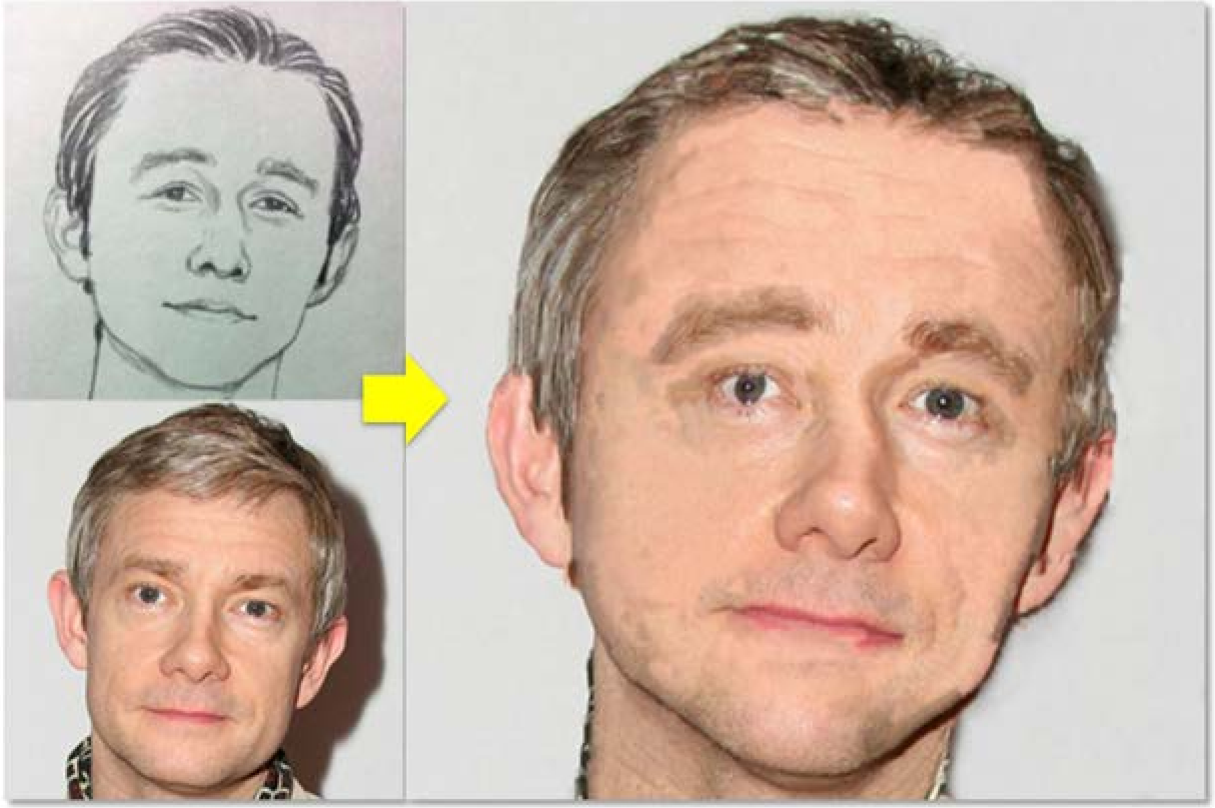

Ostagram

Results: Style to Style

A (input) : A′ (output) :: B (output) : B′ (input)


A (input)
A′ (output)
B (output)
B′ (input)

Results: Style to Photo

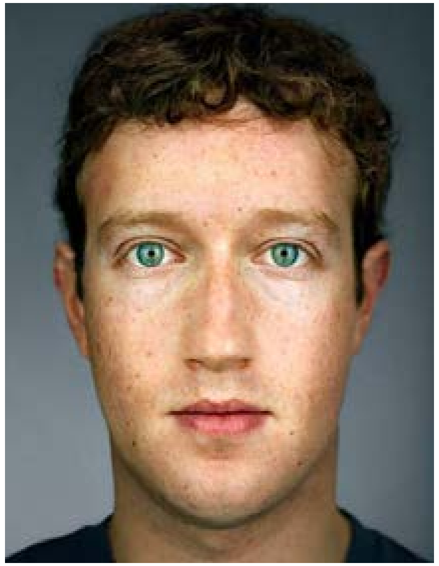


Results: Style to Photo

Results: Photo to Photo

Source

Reference



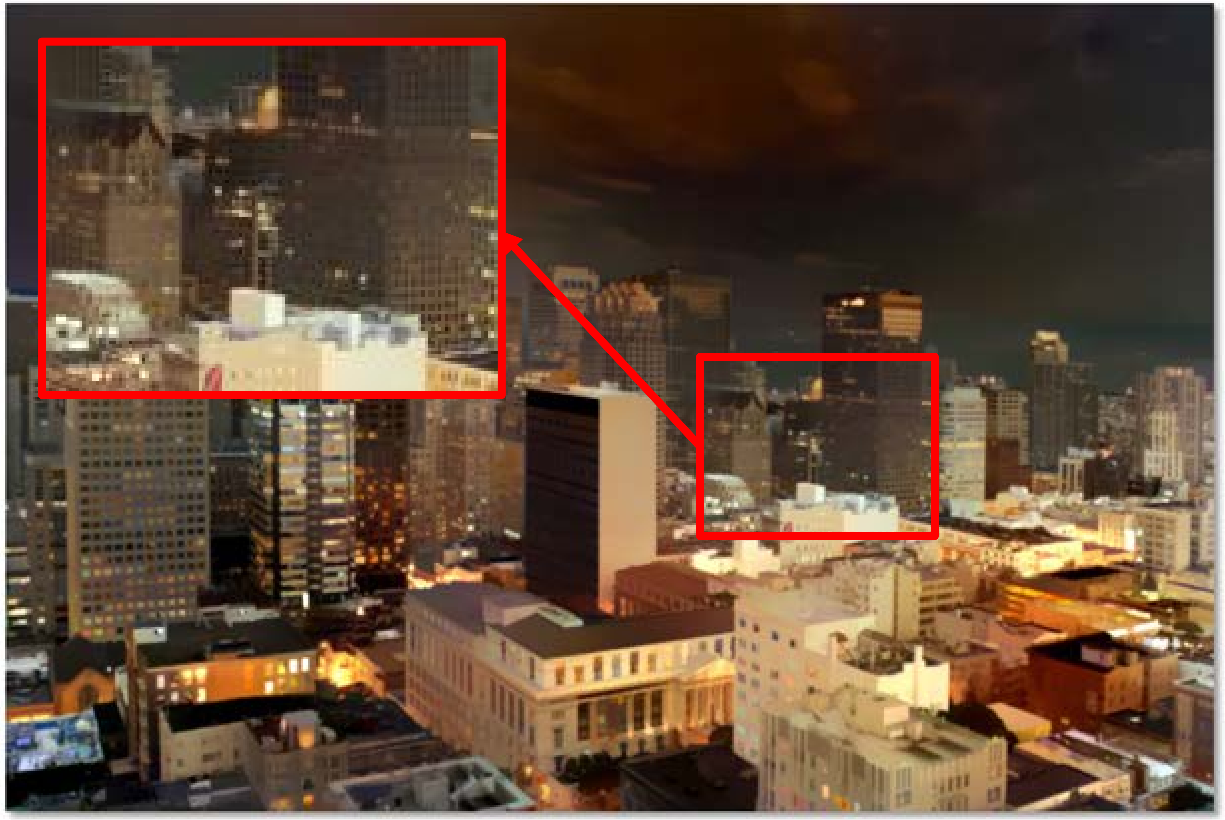
Portrait style transfer [Shih et al. 2014]

Our result

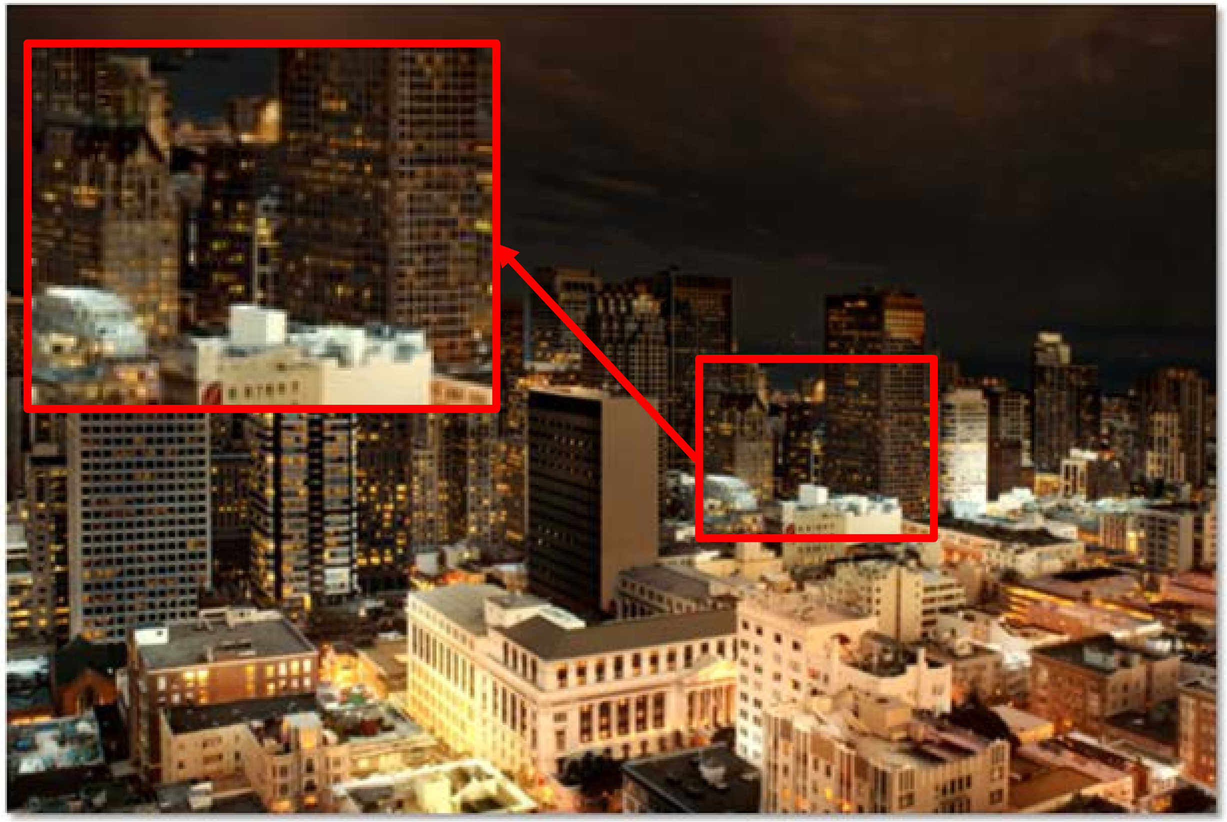
Results: Photo to Photo

Source

Reference



Deep photo style [Luan et al. 2017]



Our result

Results: Photo to Photo

Source 

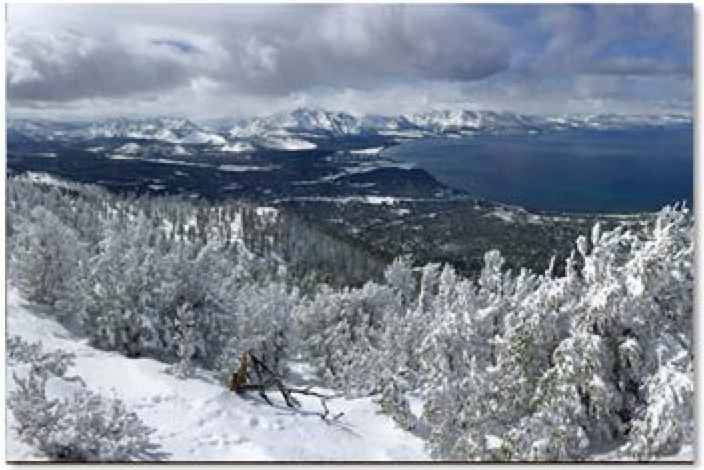
Reference 

Deep photo style [Luan et al. 2017]

Our result

Results: Time Lapse

Input (src)



Input (ref 4)

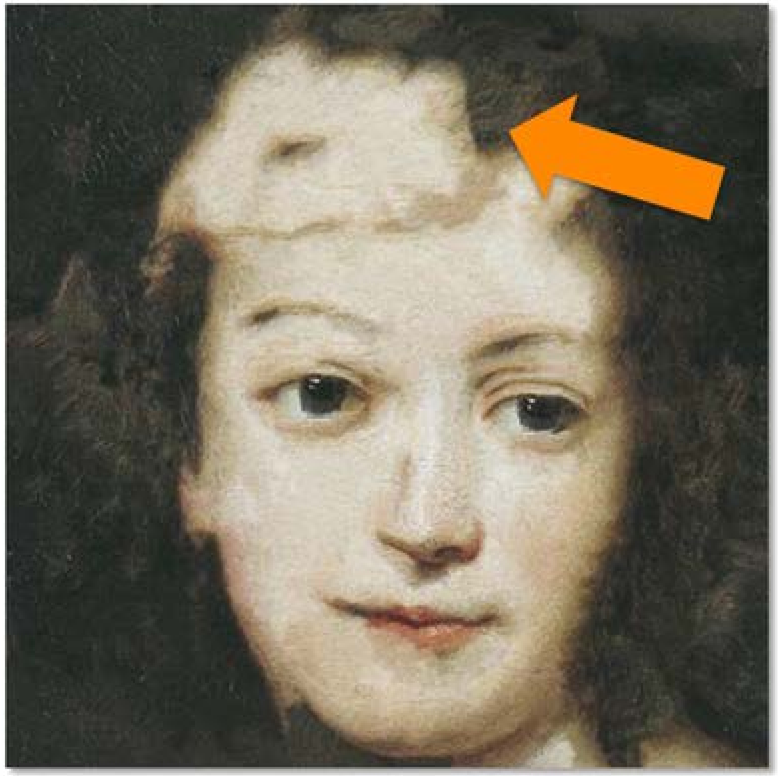
Output 4

Limitation:

Fails to find correct matches for the object which is missing in the reference

source

reference



our result

Limitation:

Fails to build correspondences between scenes varying a lot in scales

source

reference

our result

Limitation:

No geometry style transfer

source

reference

our result

Thanks!

https://github.com/msracver/Deep-Image-Analogy

Q & A