2021 ICCV OCTOBER 11-17 VIRTUAL

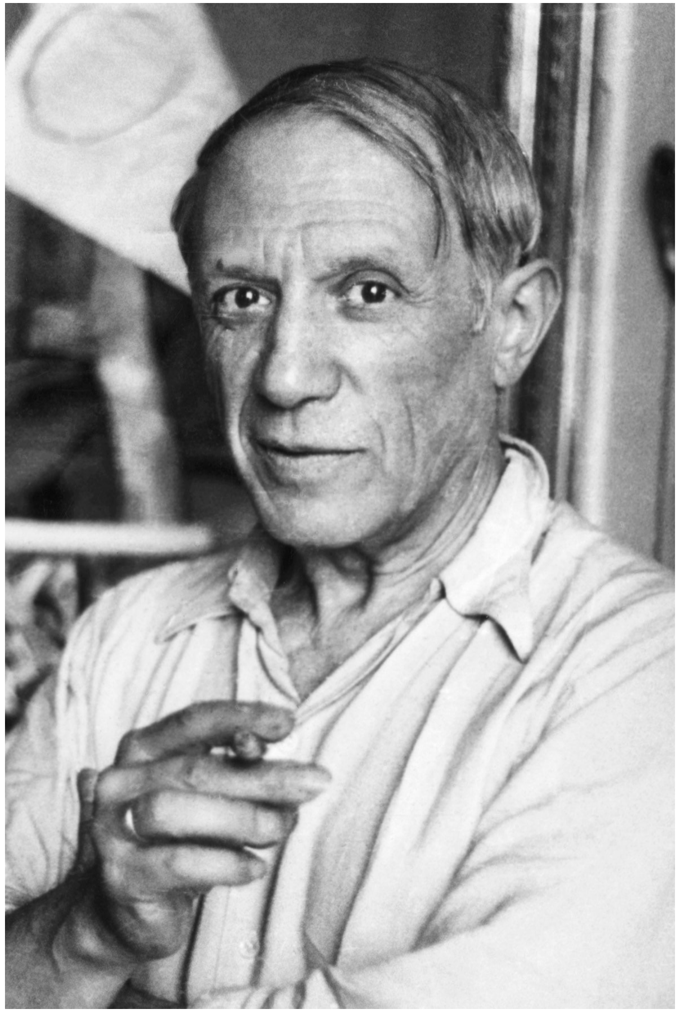
DRB-GAN: A Dynamic ResBlock Generative Adversarial Network for Artistic Style Transfer

Wenju Xu[1], Chengjiang Long[2], Ruisheng Wang[3], Guanghui Wang[4],

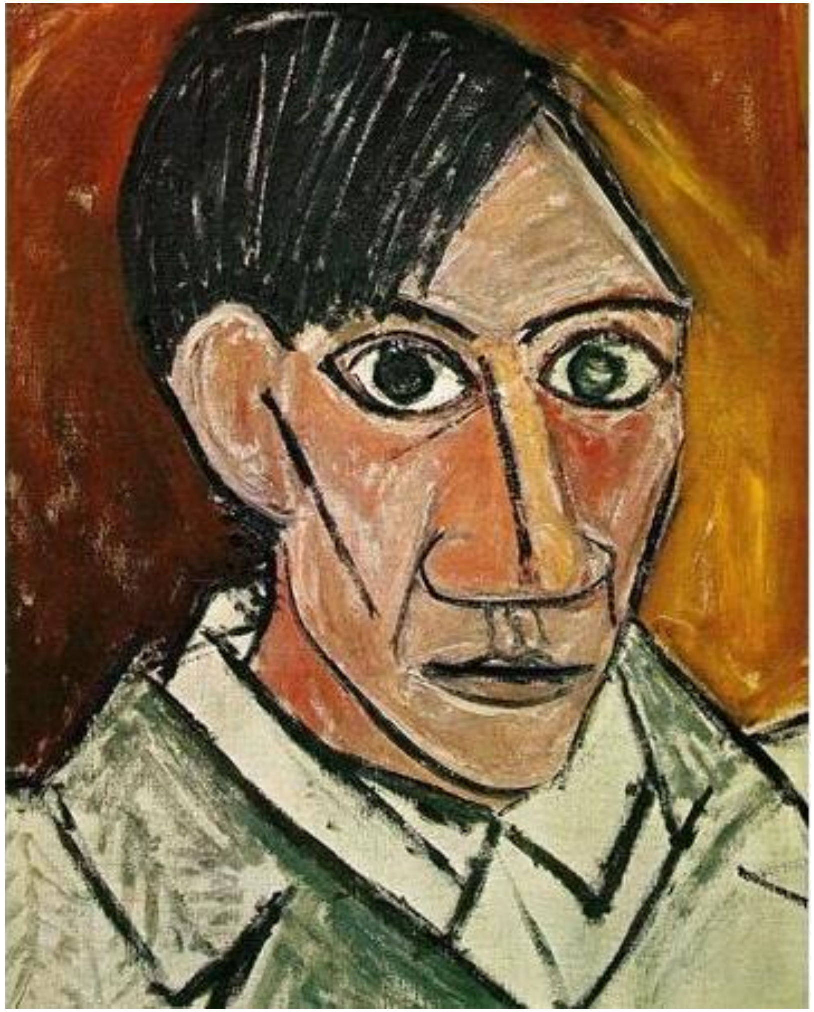
[1]OPPO US Research Center, InnoPeak Technology Inc
[2]JD Finance America Corporation
[3]University of Calgary
[4]Ryerson University

oppo JDT 京东科技 UNIVERSITY OF CALGARY Ryerson University

Artistic style transfer

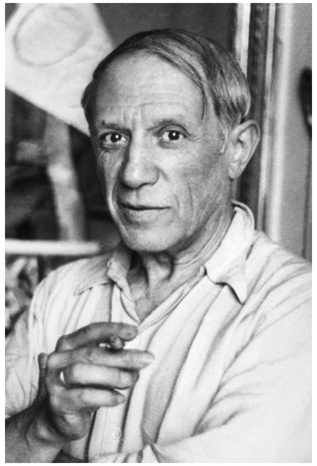


Picasso

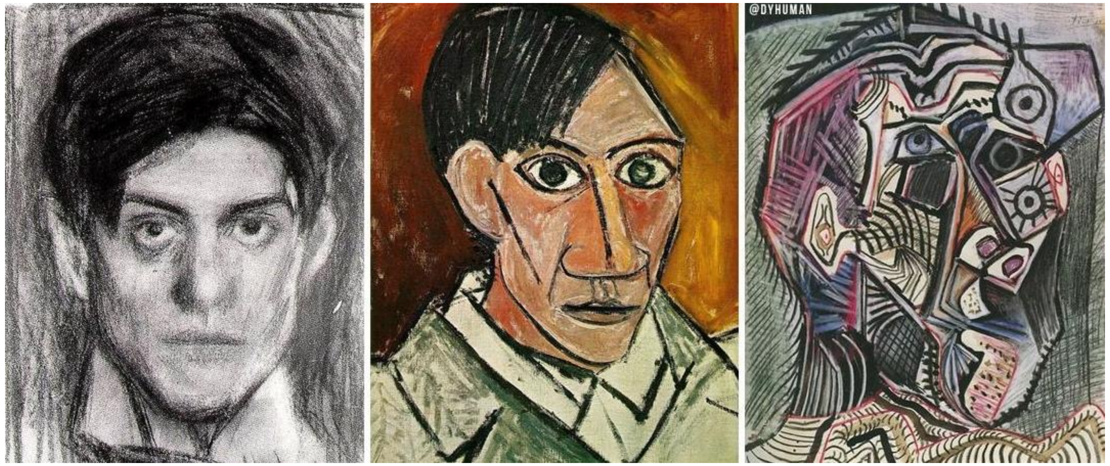


Picasso’s self-portrait

Artistic style transfer



The Evolution of Picasso's self portrait

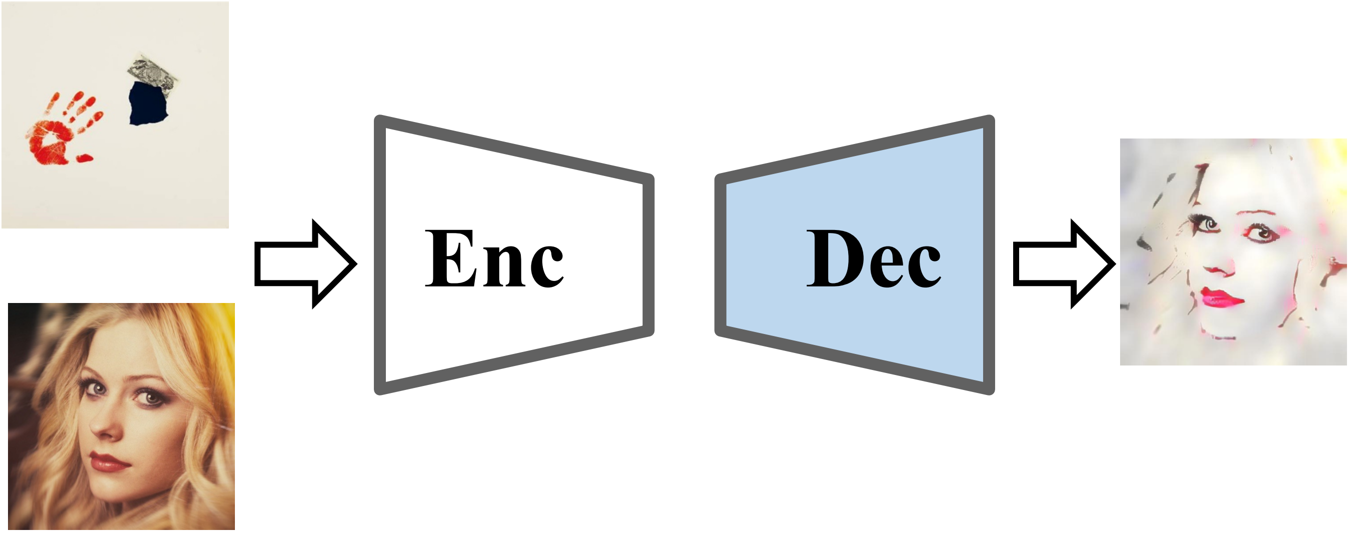


Age 18 Age 25 Age 90

Style: Surrealism

Art collection

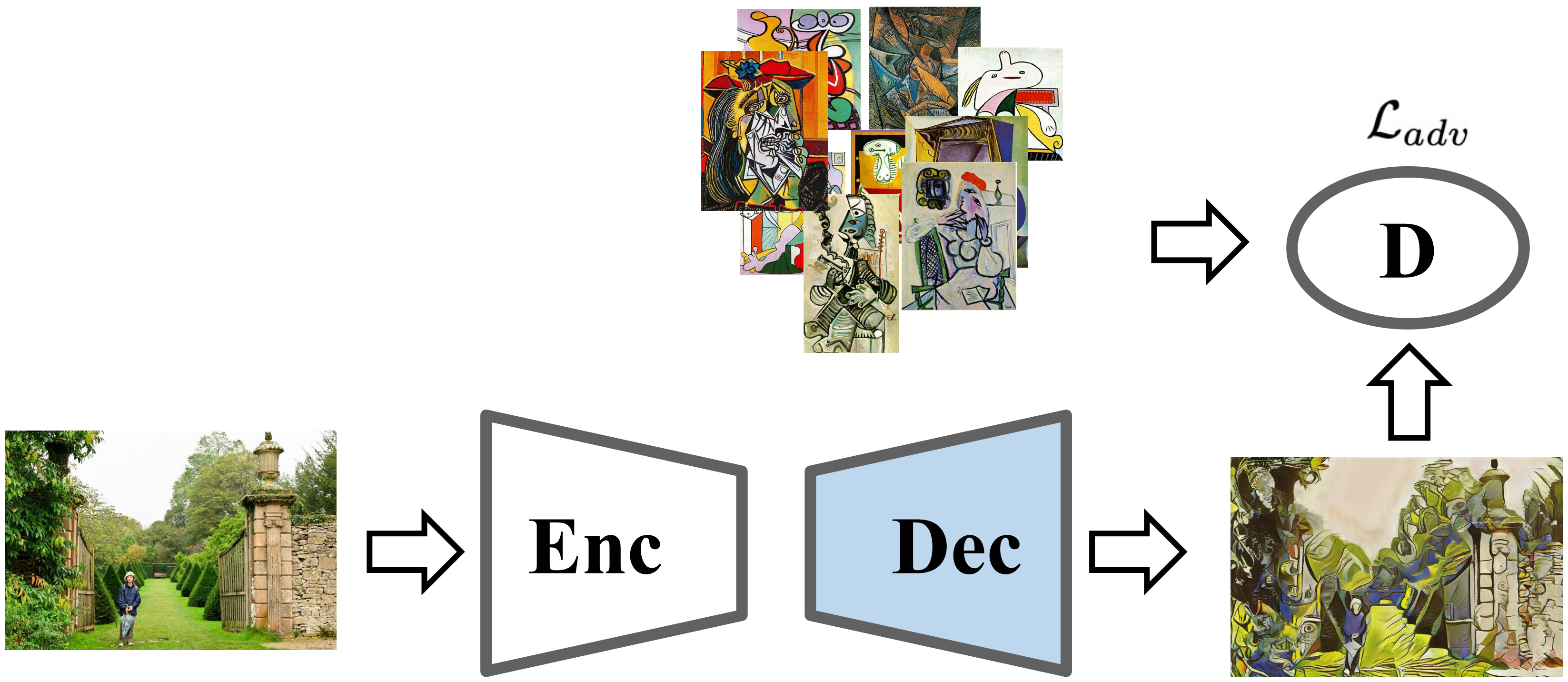
Arbitrary style transfer



- Cannot benefit from other style images sharing similar style.
- Cannot well obtain style consistency and maintain content structure similarity.

[1] Arbitrary style transfer (Huang et al., 2017)
[2] Neural style transfer (Gatys et al., 2016)

Collection style transfer

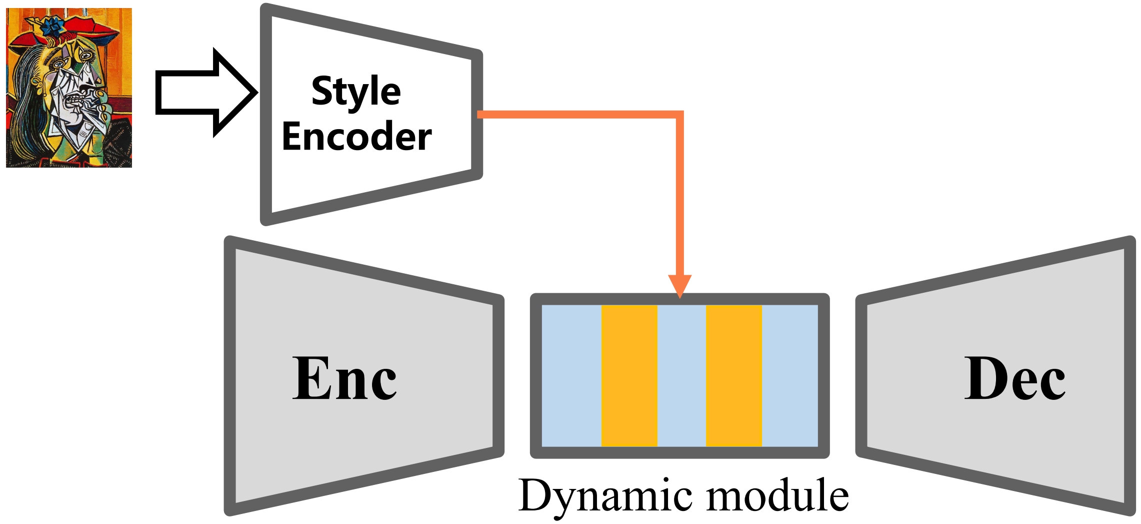


- Recognize and transfer the dominant style clues;
- Lack the flexibility of exploring style manifold.

[1] Adaptive Style Transfer (Sanakoyeu et al., 2018)
[2] CycleGAN (Zhu et al., 2017)

Insights

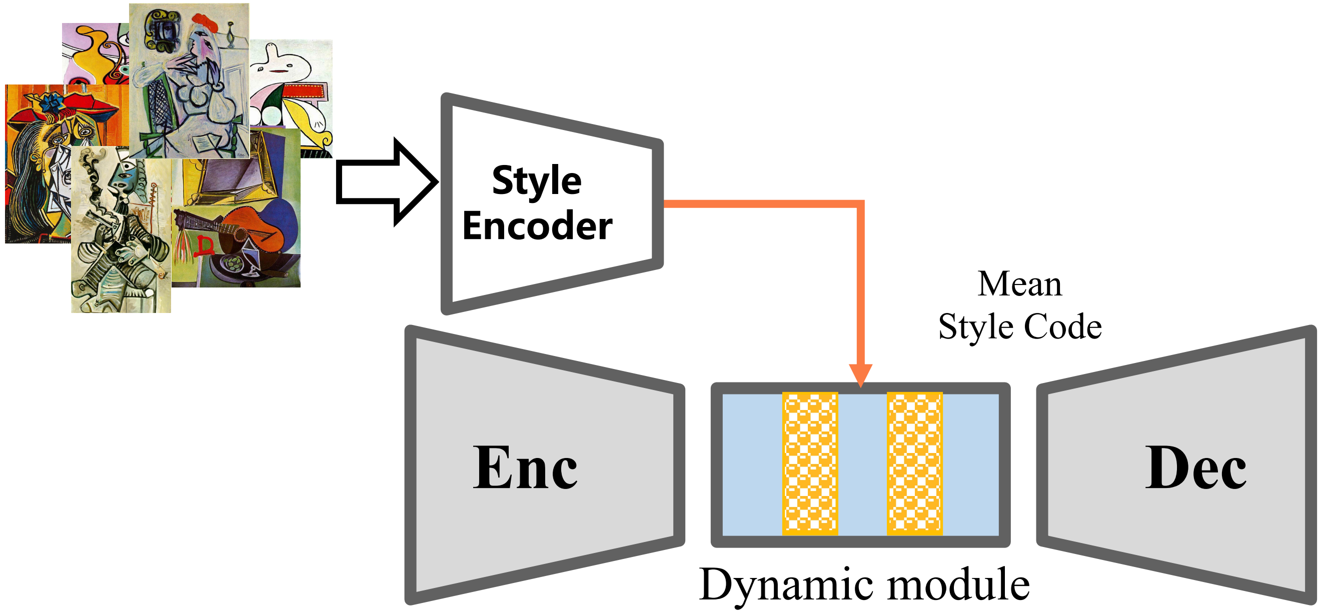
- Handle arbitrary style transfer and collection style transfer in a unified model.
- Ensure style consistency and content structural similarity.



- “style codes” is modeled as the dynamic parameters within dynamic modules.

Insights

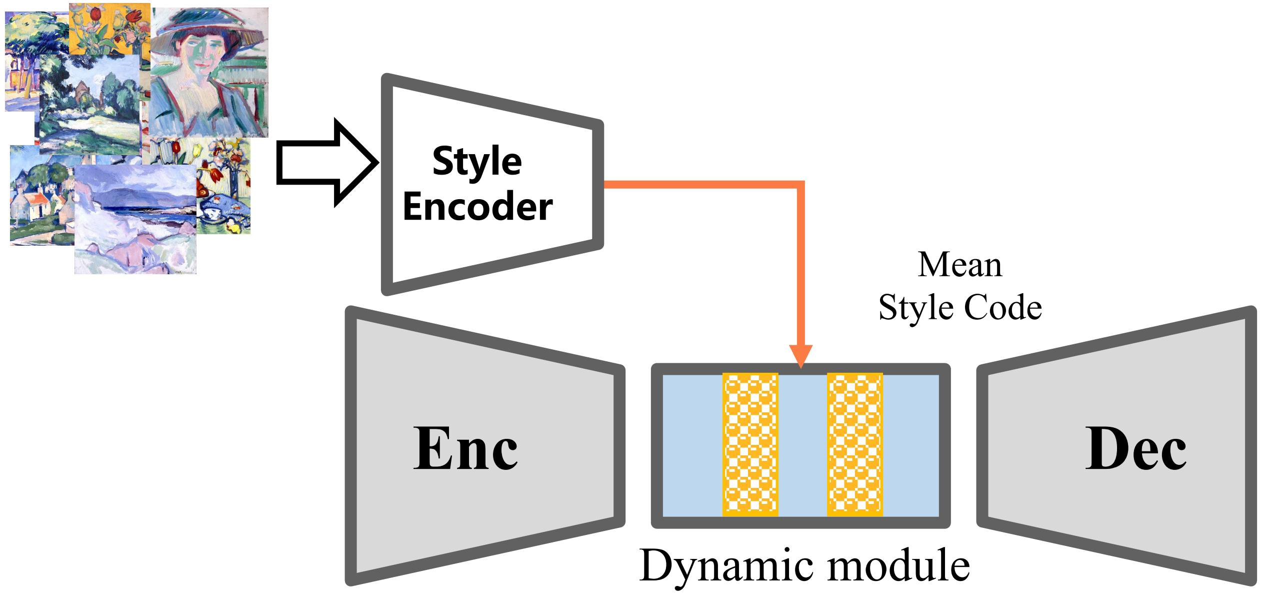
- Handle arbitrary style transfer and collection style transfer in a unified model.
- Ensure style consistency and content structural similarity.



- “style codes” is modeled as the dynamic parameters within dynamic modules.

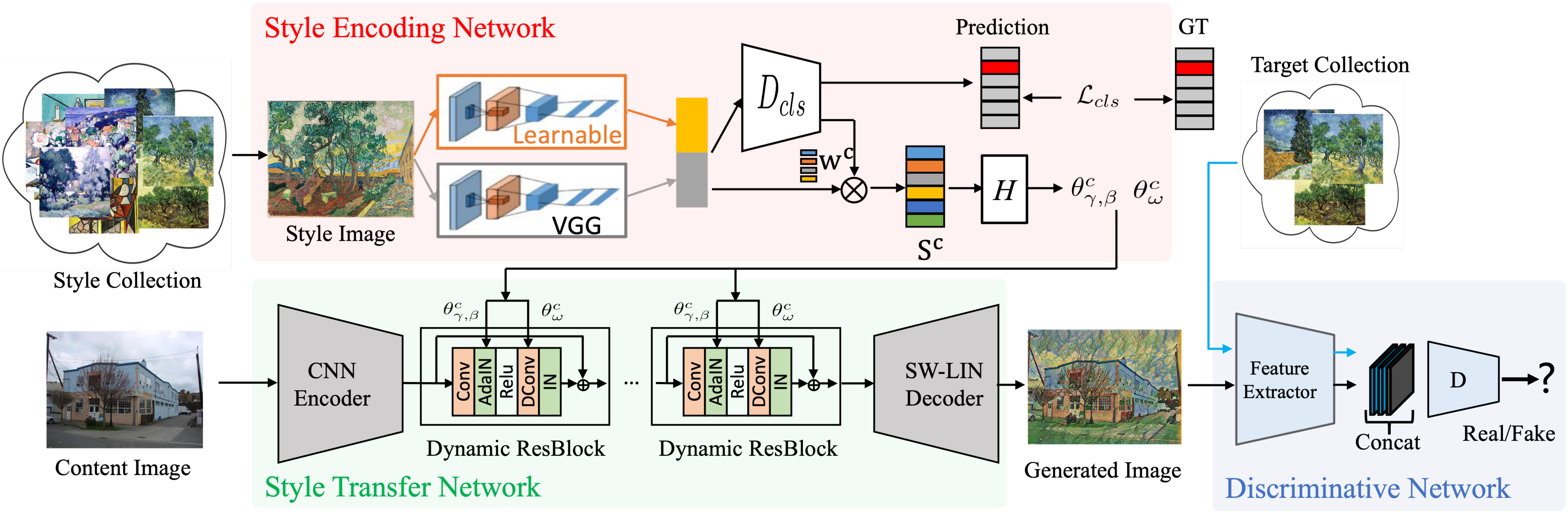
Insights

- Handle arbitrary style transfer and collection style transfer in a unified model.
- Ensure style consistency and content structural similarity.



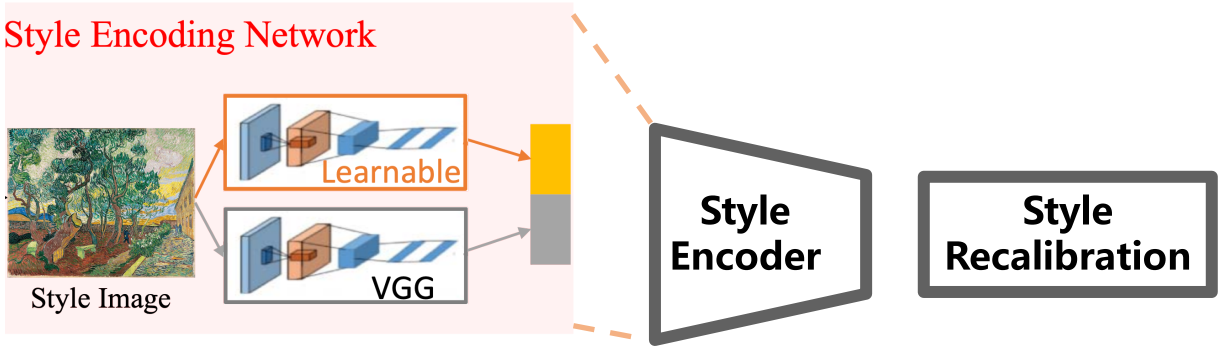
- “style codes” is modeled as the dynamic parameters within dynamic modules.

DRB-GAN



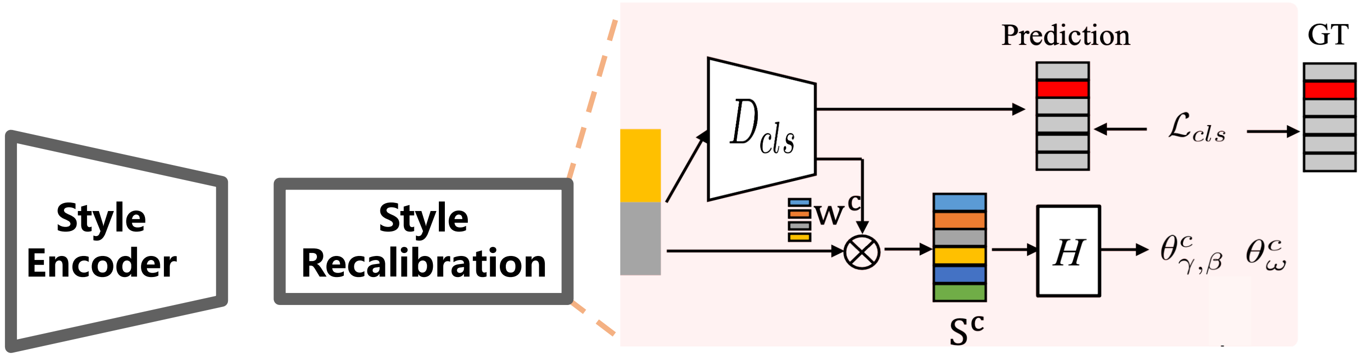
- Three Components: style encoding network, style transfer network and discriminative network.

Style Encoding Network



- Style encoder: learnable CNN & pretrained VGG.

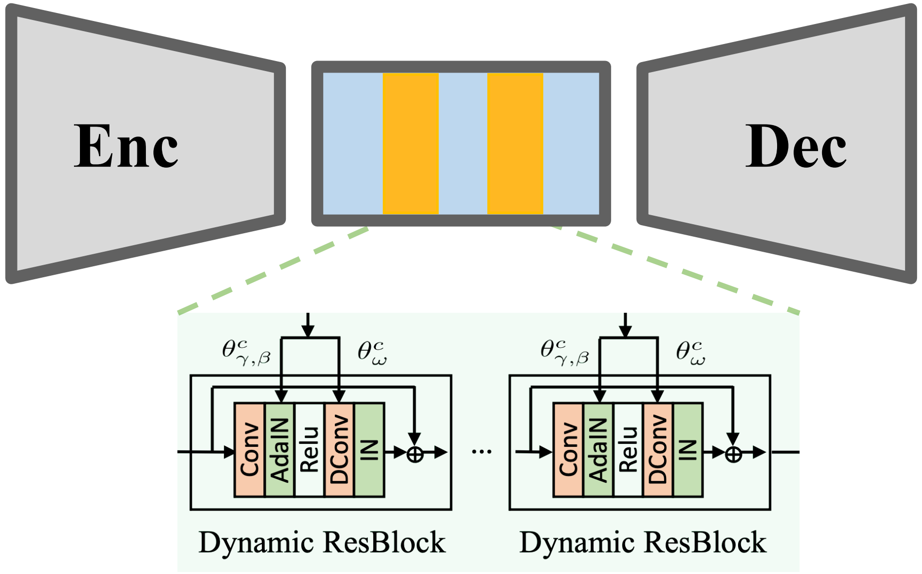
Style Encoding Network



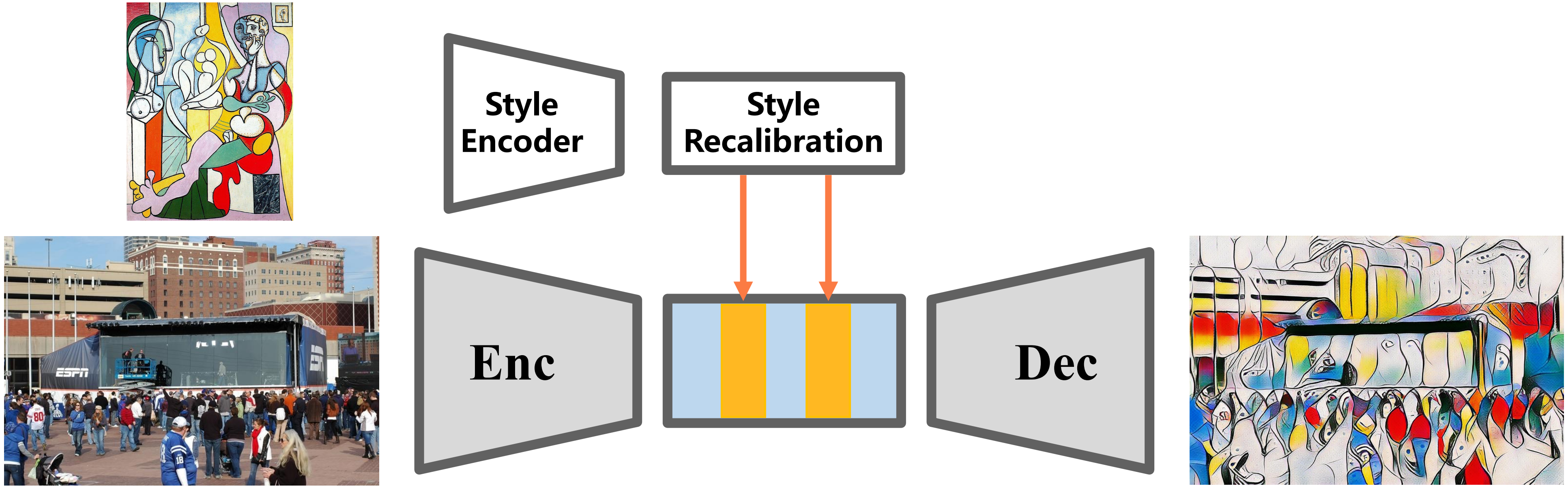
- Style recalibration: refine the style code with the class attention.

Style transfer network

- Dynamic ResBlock: dynamic convolutional layer and AdaIN.



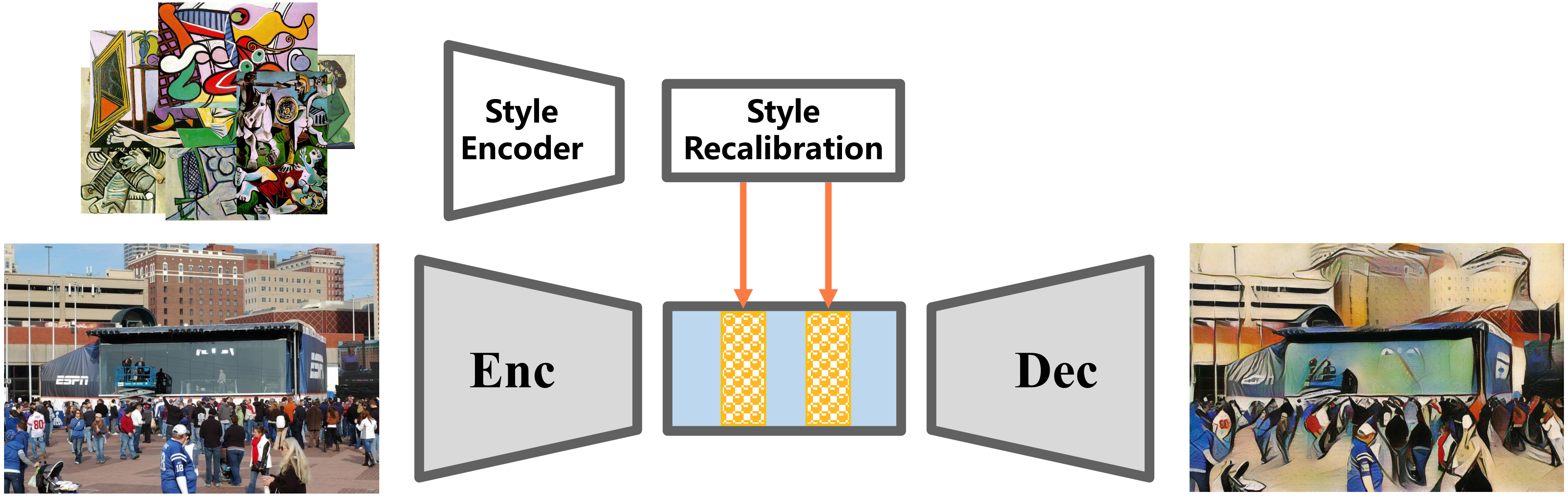
Style code



- “style code” in dynamic ResBlocks:

$$\{\theta_{\omega}^{c}, \theta_{\gamma,\beta}^{c}\} = \{H_{\omega}(s^{c}), H_{\gamma,\beta}(s^{c})\}$$

Collection style code



- “collection style code” as a weighted mean of the “style codes”:

$$\{\bar{\theta}^c_{\omega},\ \bar{\theta}^c_{\gamma,\beta}\} = \{\frac{1}{K}\sum_{k=0}^{K}\pi_k\theta^c_{\omega_k},\ \frac{1}{K}\sum_{k=0}^{K}\pi_k\theta^c_{\gamma_k,\beta_k}|c \sim N\}$$

Style transfer network

- SW-LIN Decoder: spatial window layer-instance normalization layer.
- Preserve local feature and remove artifacts in generated images.

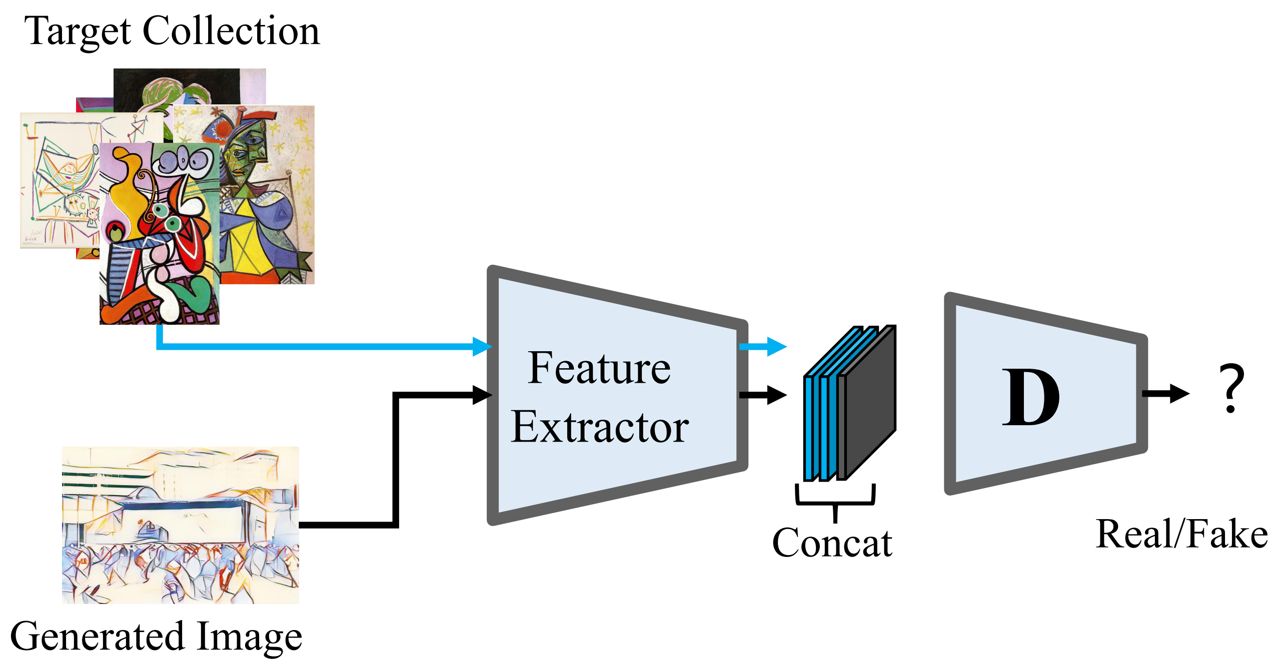
Style transfer network

$$\text{SW-LIN}(\gamma, \beta, \rho) = \gamma(\rho \phi_{sw}^{c} + (1-\rho)\phi_{sw}^{l}) + \beta$$

$$\phi_{sw} = \frac{\mathbf{h} - E_{x_i \in sw}[\mathbf{h}(x_i)]}{\sqrt{Var_{x_i \in sw}[\mathbf{h}(x_i)]}}$$

- SW-LIN Decoder: spatial window layer-instance normalization layer.
- Preserve local feature and remove artifacts in generated images.

Discriminative network :



$$\mathcal{L}_{adv} = E_{y^c, y_i^c \sim Y, c \sim N}[-\log D(y^c, \{y_i^c\}_{i=0}^{M})] + E_{\tilde{x}^c \sim G(x), y_j^c \sim Y, c \sim N}[-\log(1 - D(\tilde{x}^c, \{y_j^c\}_{j=0}^{M}))]$$

- Objective function

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_{per}\mathcal{L}_{per} + \lambda_{cls}\mathcal{L}_{cls}$$

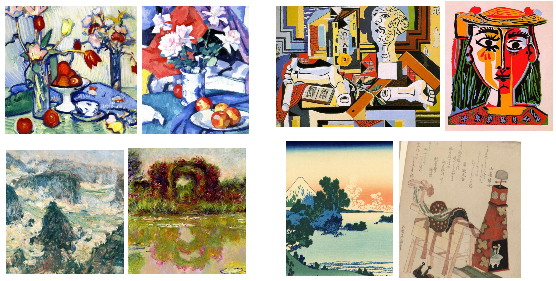
Comparison with other approaches

- Dataset
  - Content image: Place365 dataset
  - Style image: Wikiart dataset
- Metrics: Deception rate, inference time and human study.
- Model is trained on 768x768 and inferred on arbitrary resolution.

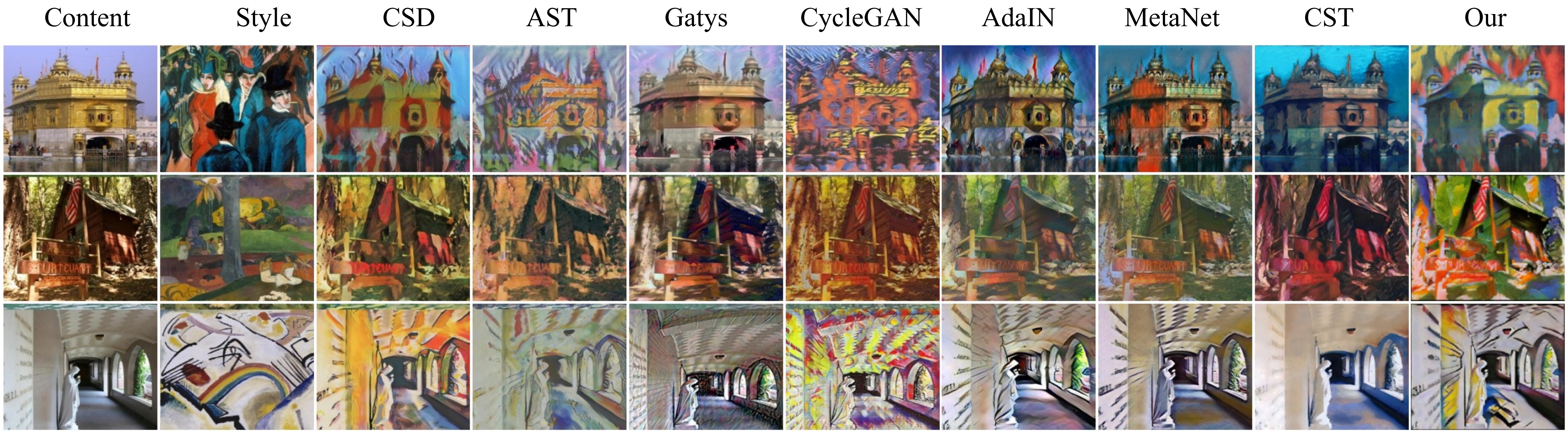
Method	Time (sec)	GPU memory (MiB)	Model	Deception rate	Human studies Content score	Style score
Wikiart test				0.626	-	-
Gatys *et al.*	200	3887	PSPM	0.251	67.1%	0.127
AdaIN	0.16	8872	ASPM	0.061	43.6%	0.019
WCT	5.22	10720	ASPM	0.023	39.2%	0.013
PatchBased	8.70	4159	ASPM	0.063	53.4%	0.043
Johnson	0.06	671	ASPM	0.080	38.5%	0.021
CycleGAN	0.07	1391	PDPM	0.130	43.2%	0.012
AST	0.07	1043	PDPM	0.450	63.9%	0.312
DRB-GAN	0.08	1324	MDPM	**0.573**	**72.2%**	**0.453**

Place365 dataset

Wikiart dataset



Comparison with other approaches



- Our method: no artifacts in the regions and preserve the structural similarity.

Arbitrary style transfer

- Style consistency & Content structural similarity.

Collection style transfer

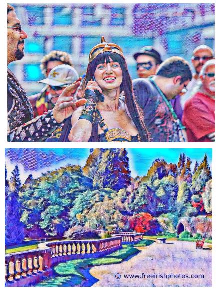
Table 2. Quantitative comparison of different methods. SD stands for style distance metric; DS represents deception score.

Setting	Arbitrary Style (SD↓)	Collection style (DS↑)			
		K=2	5	10	20
AdaIN	263.4	0.066	0.045	0.013	0.011
MetaNet	271.8	0.032	0.026	0.023	0.020
DRB-GAN	**241.2**	**0.576**	**0.580**	**0.581**	**0.583**

Collection

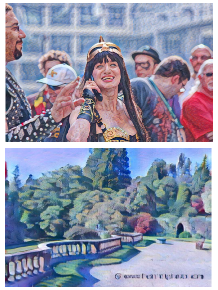
K=2

Content



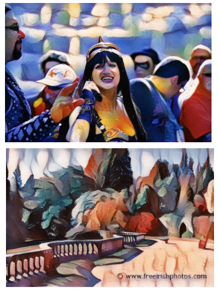
MetaNet

K=2



AdaIN

K=2



Our

- The number of style images used to calculate the mean style code.

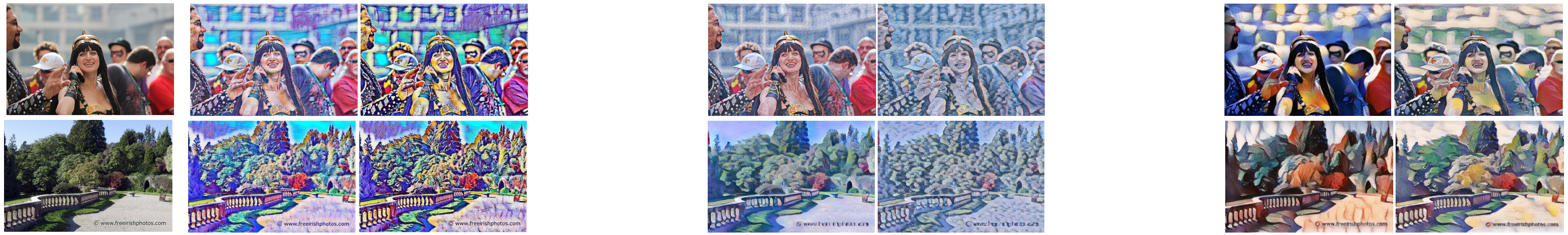
Collection style transfer

Table 2. Quantitative comparison of different methods. SD stands for style distance metric; DS represents deception score.

Setting	Arbitrary Style (SD↓)	Collection style (DS↑)			
		K=2	5	10	20
AdaIN	263.4	0.066	0.045	0.013	0.011
MetaNet	271.8	0.032	0.026	0.023	0.020
DRB-GAN	**241.2**	**0.576**	**0.580**	**0.581**	**0.583**

Collection

K=10 K=10 K=10



Content MetaNet AdaIN Our

- The number of style images used to calculate the mean style code.

Collection style transfer

Table 2. Quantitative comparison of different methods. SD stands for style distance metric; DS represents deception score.

Setting	Arbitrary Style (SD↓)	Collection style (DS↑)			
		K=2	5	10	20
AdaIN	263.4	0.066	0.045	0.013	0.011
MetaNet	271.8	0.032	0.026	0.023	0.020
DRB-GAN	**241.2**	**0.576**	**0.580**	**0.581**	**0.583**

Collection

Content MetaNet AdaIN Our

- The number of style images used to calculate the mean style code.

Collection style transfer

Picasso

Content

AST

Ours

Collection style transfer

Ablation study

- SW-LIN Decoder: preserve local feature and remove artifacts.
- w/o $\mathcal{L}_{adv}$: improve the style consistency.
- w/o vgg : capture the dominant style clues without subtle details.
- w/o $\mathcal{L}_{cls}$: causes slight degradation on stroke size variations.

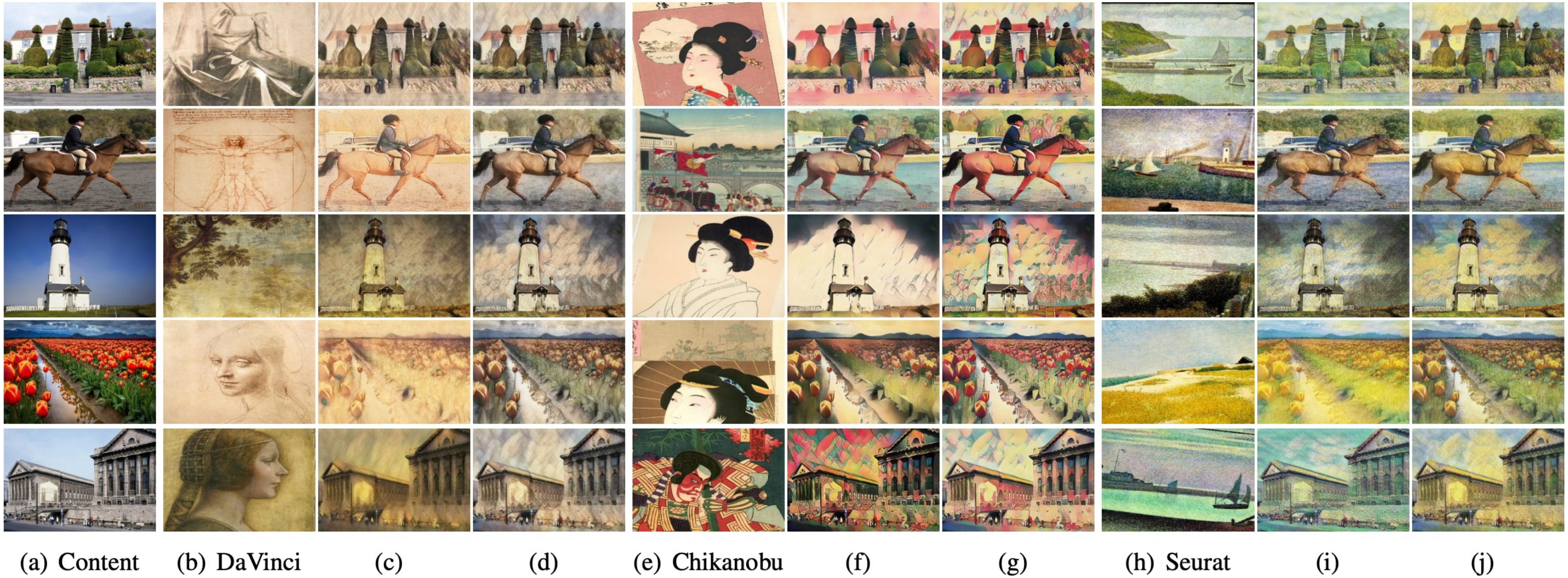
Discriminative network

- Collection discriminator: improve style consistency.

[1] CST (Jan Svoboda, 2020)

Evaluation with unseen styles

2021 ICCV OCTOBER 11-17 VIRTUAL



- (c) (f) (i): arbitrary style transfer.
- (d) (g) (j): collection style transfer.

Evaluation with different resolutions

- Style consistency.
- Structural similarity.

HD Stylization

Content

Style

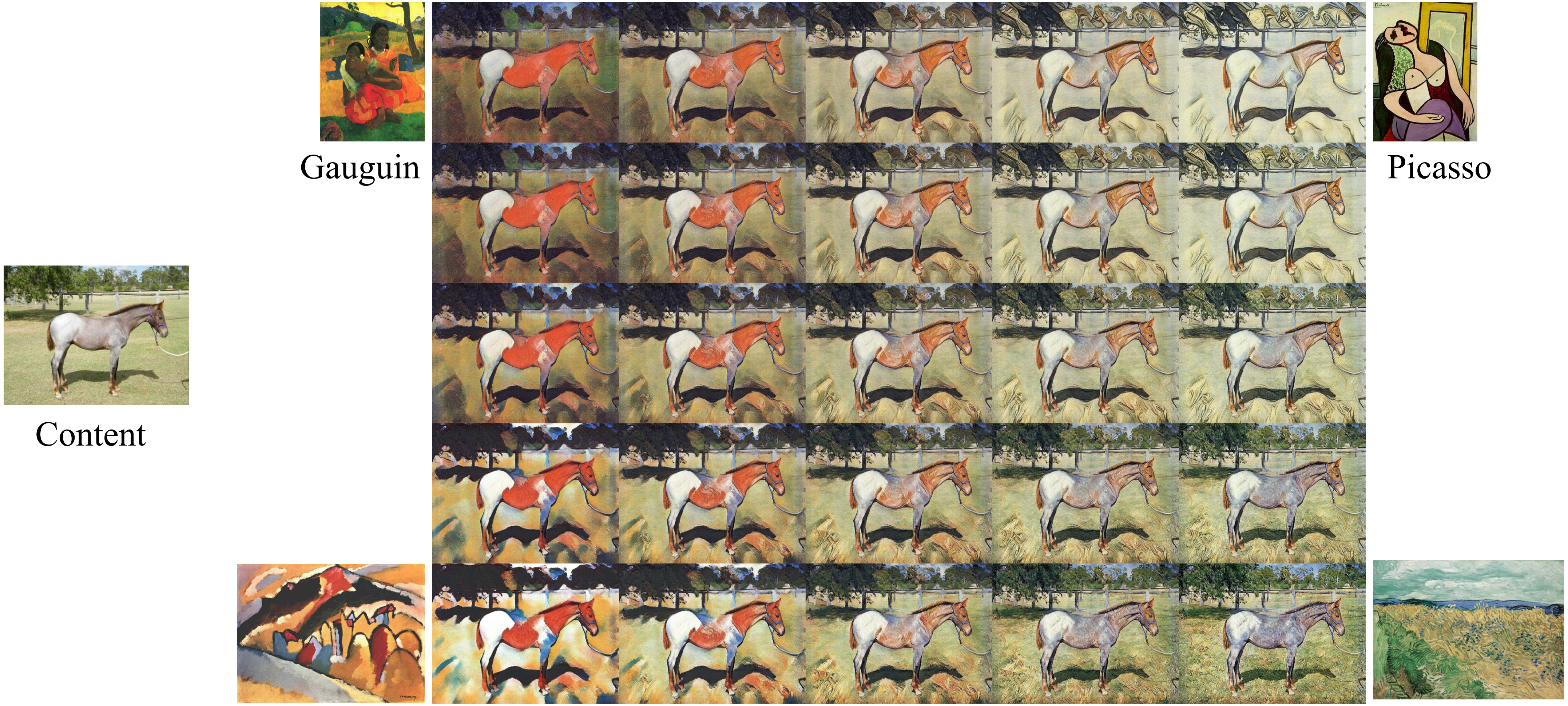
1024x2560

3072x7680

768x1920

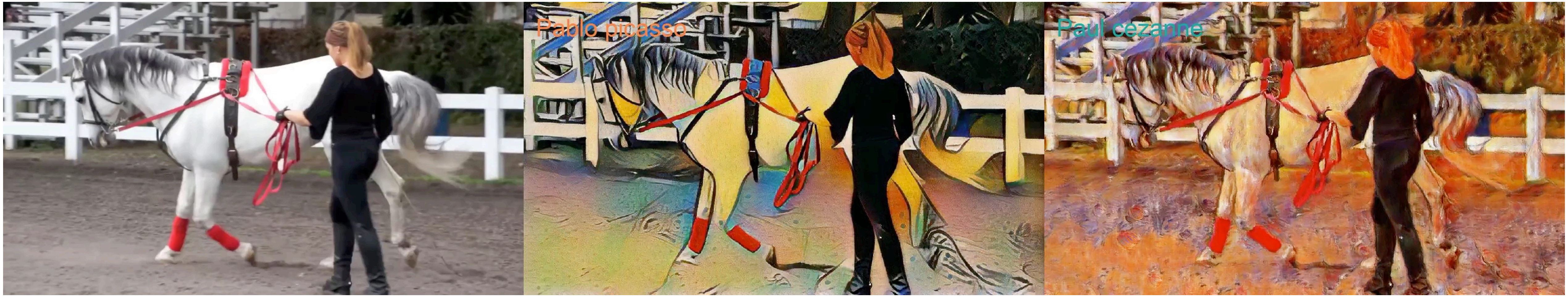
2048x5120

Four-Way Style Interpolation



- Our model creates a smooth manifold structure.

Video Style Transfer @1920x1080



Input video

Collection style transfer

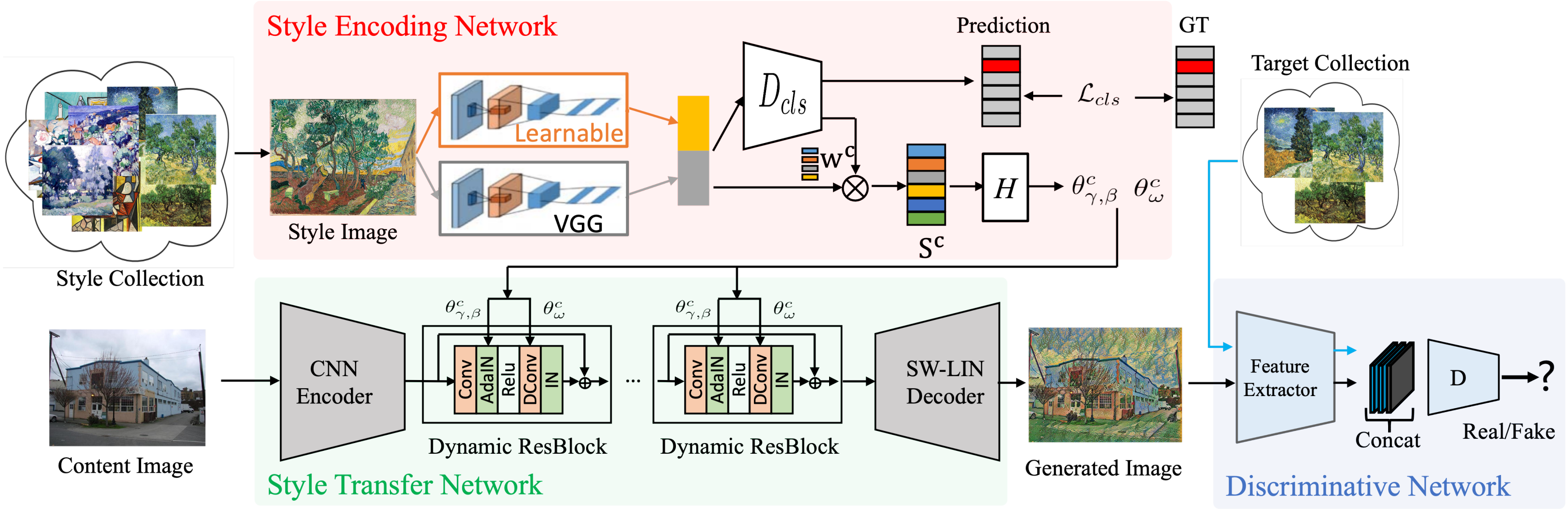
Style interpolation in arbitrary style transfer

- All stylizations come from one trained model.

Conclusions

2021 ICCV OCTOBER 11-17 VIRTUAL

- A unified Model that handle arbitrary style transfer and collection style transfer.
- "style codes" is modeled as the dynamic parameters within Dynamic ResBlocks.
- Style consistency & Content structural similarity.



QR Code for our project:

https://github.com/xuwenju123/DRB-GAN

Thank you!