Anisotropic Stroke Control for Multiple Artists Style Transfer

MM '20: Proceedings of the 28th ACM International Conference on Multimedia

October 2020 Pages 3246–3255

Keywords: Style Transfer, Deep Learning

https://doi.org/10.1145/3394171.3413770

IPHD YuanFu Yang

Outline

01 Background

02 Art Statement

03 Method

04 Connection/Demo

Background - Author

Xuanhong Chen

Shanghai Jiao Tong University

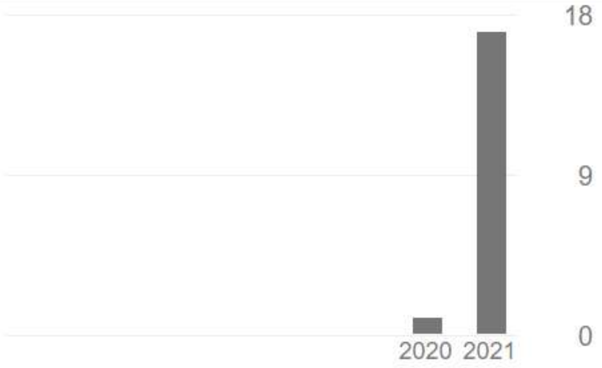
Most frequent co-Author	Most cited colleague	Top subject	Top keyword	Most frequent Affiliation
Bingbing Ni Shanghai Jiao ...	Bingbing Ni Shanghai Jiao ...	**Computational photography**	**image translation**	Shanghai Jiao Tong University 2 Papers
View author →	View author →	View research →	View research →	View affiliation →

Bibliometrics

Average Citation per Article	Citation count	Publication counts	Publication Years	Available for Download	Average Downloads per Article	Downloads (6 weeks)	Downloads (12 months)	Downloads (cumulative)
2	3	2	2020 - 2020	2	260	78	361	519

Background - Author

TITLE	CITED BY	YEAR
SimSwap: An Efficient Framework For High Fidelity Face Swapping R Chen, X Chen, B Ni, Y Ge Proceedings of the 28th ACM International Conference on Multimedia, 2003-2011	7	2020
CooGAN: A Memory-Efficient Framework for High-Resolution Facial Attribute Editing X Chen, B Ni, N Liu, Z Liu, Y Jiang, L Truong, Q Tian European Conference on Computer Vision, 670-686	4	2020
Fast Optimal Transport Artistic Style Transfer T Qiu, B Ni, Z Liu, X Chen International Conference on Multimedia Modeling, 37-49	2	2021
Incipient fault detection and variable isolation based on subspace decomposition and distribution dissimilarity analysis C Zhao, X Chen, L Lu, S Zhang, Y Sun 2017 6th Data Driven Control and Learning Systems (DDCLS), 48-53	2	2017
X-volution: On the unification of convolution and self-attention X Chen, H Wang, B Ni arXiv preprint arXiv:2106.02253	1	2021
Sketch Generation with Drawing Process Guided by Vector Flow and Grayscale Z Tong, X Chen, B Ni, X Wang Proceedings of the AAAI Conference on Artificial Intelligence 2021	1	2020
A phase division strategy for multiphase batch process monitoring based on particle swarm optimizer (PSO) X Chen, C Zhao, Y Sun 2017 29th Chinese Control And Decision Conference (CCDC), 4515-4520	1	2017
Image Translation via Fine-grained Knowledge Transfer X Chen, Z Liu, T Qiu, B Ni, N Liu, X Hu, Y Li		2020



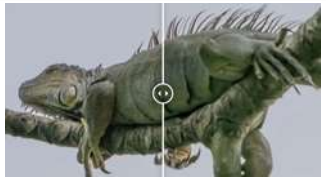
Public access VIEW ALL

2 articles — not available

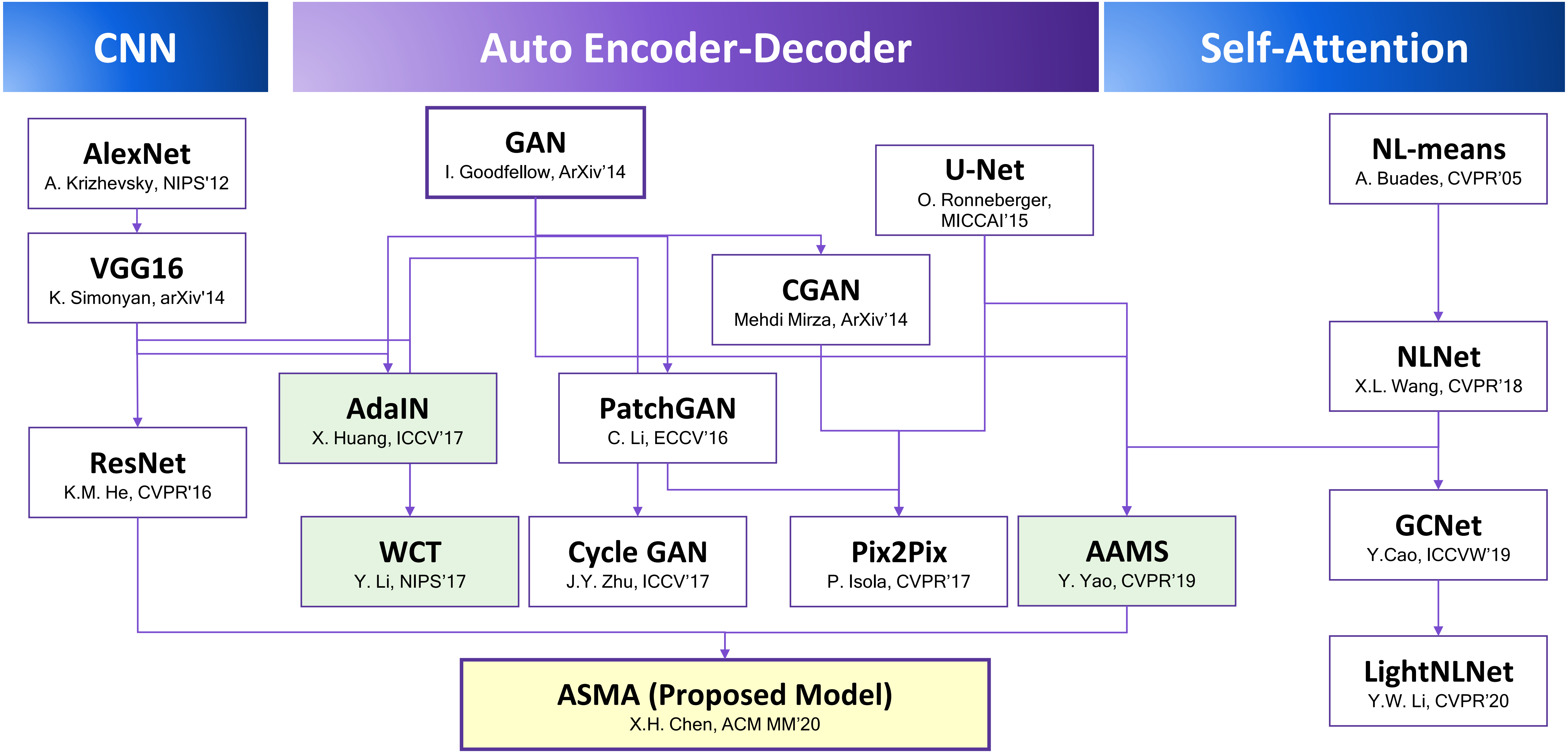
4 articles — available

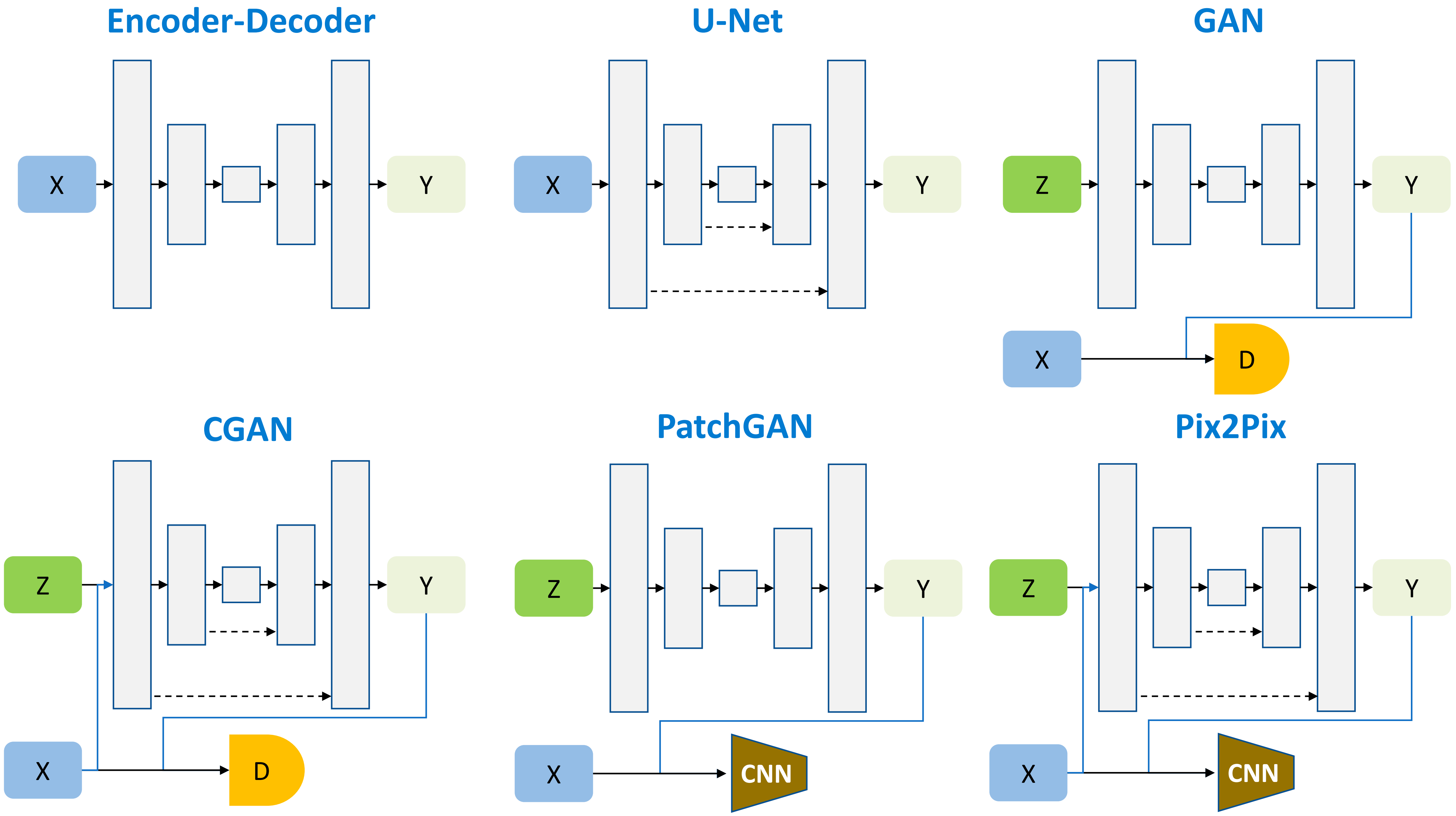
Based on funding mandates

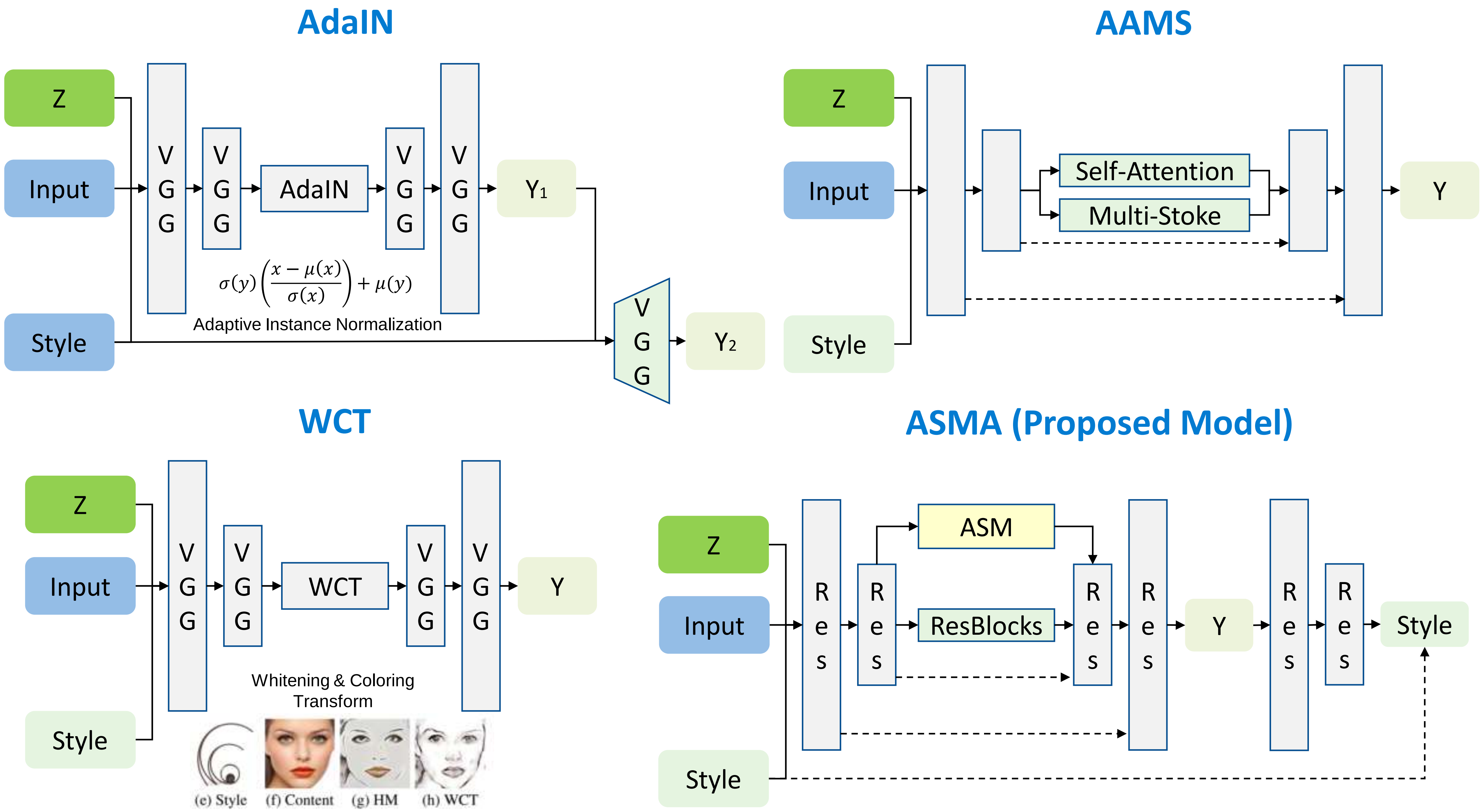
Background – Image Process

Cate	Paper Title	Authors	Conference/Journal	Method
Super Resolution 	Automatic photo adjustment using deep neural networks	Zhicheng Yan et al.	ACM Transactions on Graphics	DNN, CNN
	Residual Dense Network for Image Super-Resolution	Yulun Zhang, et al.	CVPR'18	ResNet
Sketch Simplification 	Learning to simplify: fully convolutional networks for rough sketch cleanup	Simo-Serra et al.	ACM Transactions on Graphics	CNN
	Sketch simplification by classifying strokes	Toru Ogawa et al.	ICPR'16	CNN, Auto Encoder
Style Transfer 	Image style transfer using convolutional neural networks.	Gatys et al.	CVPR'16	CNN, VGG
	Painting style transfer for head portraits using convolutional neural networks	Selim et al.	ACM Transactions on Graphics (TOG)	CNN, Auto Encoder
Inpainting 	Context Encoders: Feature Learning by Inpainting	Pathak et al.	CVPR'16	CNN, Auto Encoder
	Foreground-Aware Image Inpainting	Wei Xiong et al.	CVPR'19	CNN
Image blending 	Learning a Discriminative Model for the Perception of Realism in Composite Images	Zhu et al.	ICCV'15	GAN
	GP-GAN: Towards Realistic High-Resolution Image Blending	Huikai Wu et al.	ACMMM'19	GAN
Denoising 	Deep joint demosaicking and denoising	Gharbi et al.	ACM Transactions on Graphics	Auto Encoder
	Real Image Denoising with Feature Attention	Saeed Anwar and Nick Barnes	ICCV'19	CNN, Self-Attention
Colorization 	Real-Time User-Guided Image Colorization with Learned Deep Priors	Richard Zhang et al.	SIGGRAPH'17	U-Net

Background – Related Works




Encoder-Decoder
X
Y
U-Net
X
Y
GAN
Z
Y
X
D
CGAN
Z
Y
X
D
PatchGAN
Z
Y
X
CNN
Pix2Pix
Z
Y
X
CNN

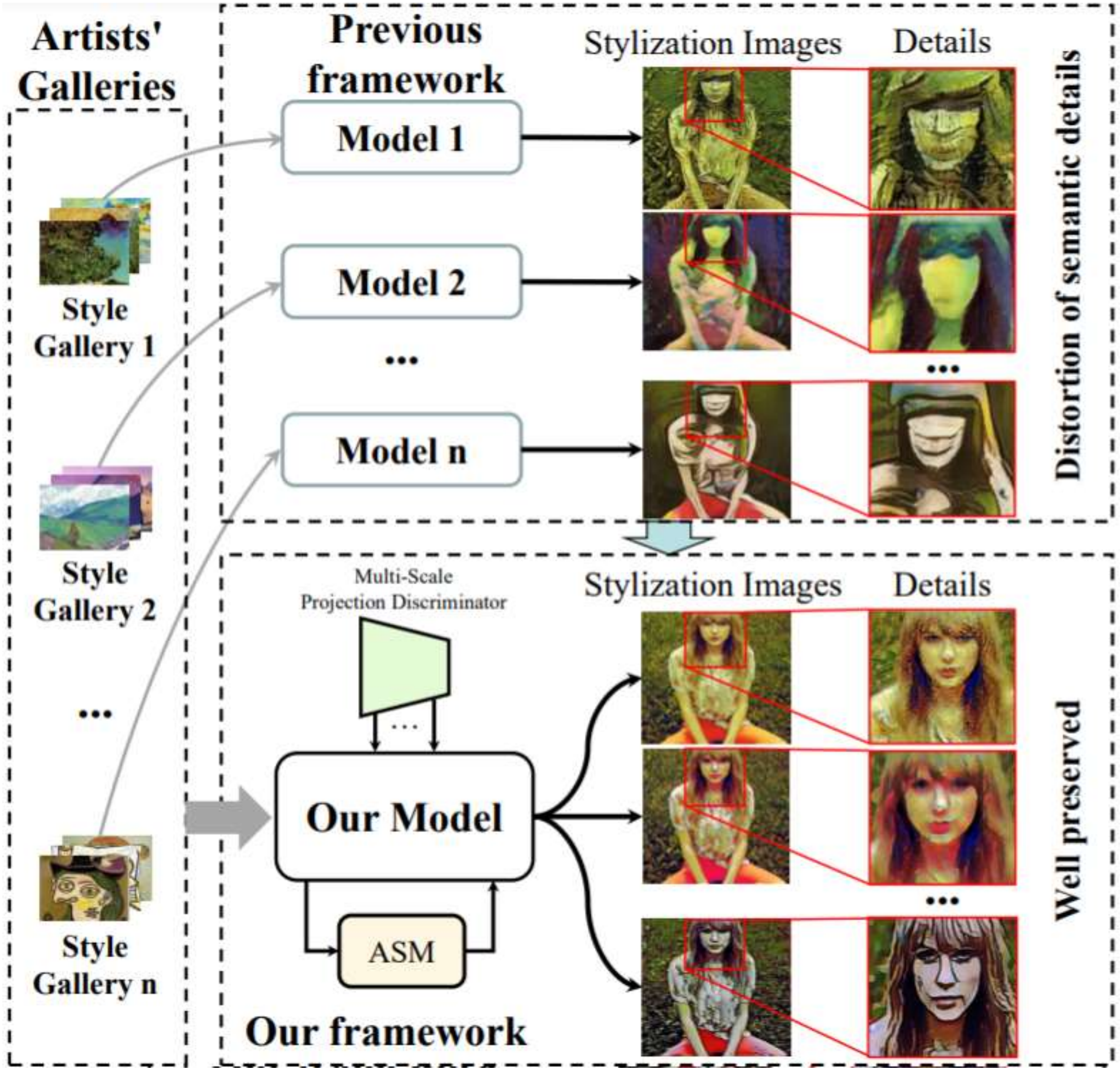

AdaIN
Z
Input
Style
VGG
VGG
AdaIN
VGG
VGG
Y1
$\sigma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right)+\mu(y)$
Adaptive Instance Normalization
VGG
Y2
AAMS
Z
Input
Style
Self-Attention
Multi-Stoke
Y
WCT
Z
Input
Style
VGG
VGG
WCT
VGG
VGG
Y
Whitening & Coloring Transform
(e) Style
(f) Content
(g) HM
(h) WCT
ASMA (Proposed Model)
Z
Input
Style
Res
Res
ASM
ResBlocks
Res
Res
Y
Res
Res
Style

Background – Related Works

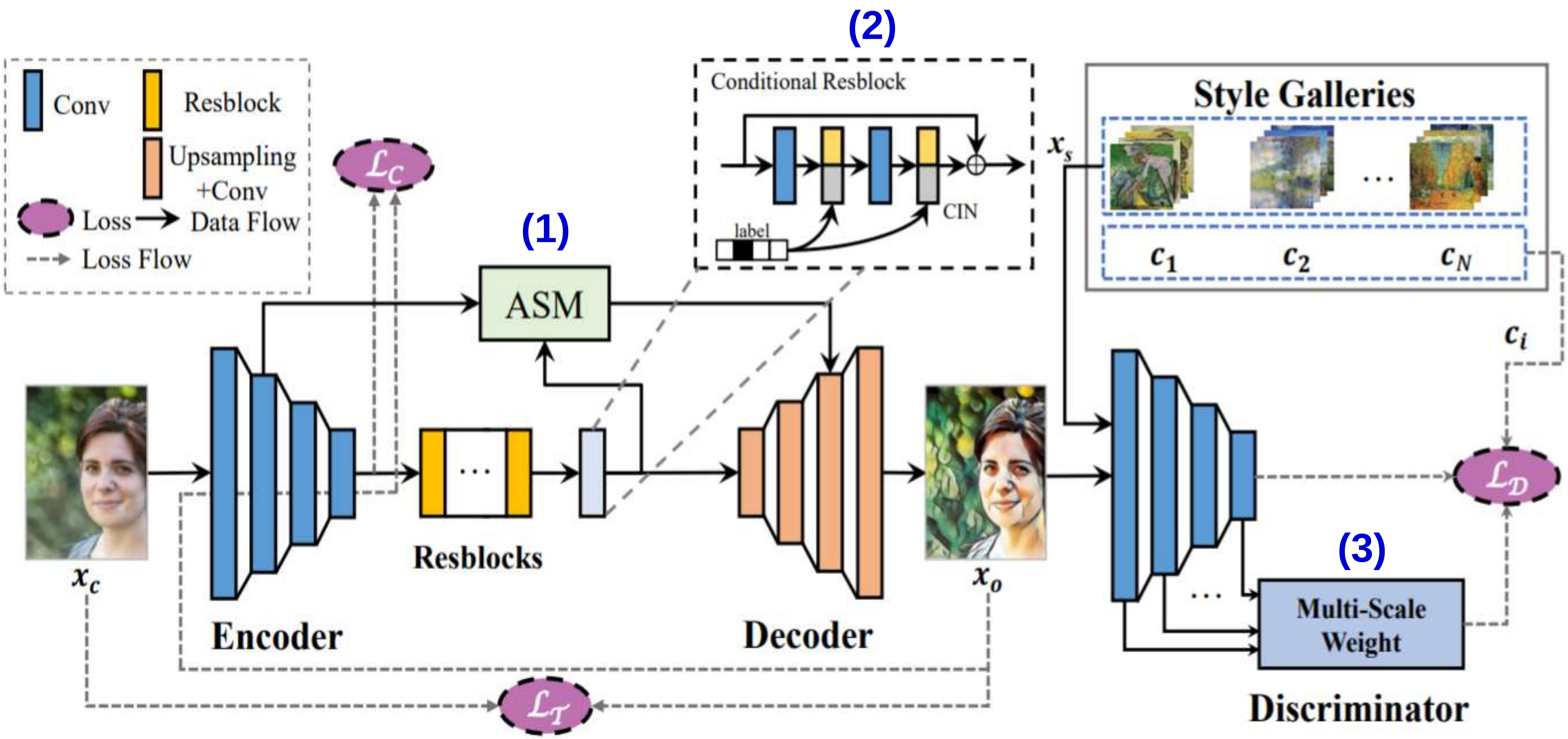
Auto-Encoder	Conference	Loss Function	
U-Net	MICCAI'15	Binary Cross Entropy Loss	$\sum_{x\in\Omega} w(x)\log(P_{\ell(x)}(X)))$
GAN	ArXiv'14	GAN Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$
CGAN	ArXiv'14	Condition Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x\|y)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z\|y)\right)\right)\right]$
PatchGAN	ECCV'16	GAN Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$
Pix2Pix	CVPR'17	Condition Loss+L1 Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x\|y)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z\|y)\right)\right)+\lambda\mathcal{L}_{L1}(G)\right]$
Cycle GAN	ICCV'17	Cycle Loss	$\min_{\theta_g}\max_{\theta_d}[\mathcal{L}_{GAN}(G,D_Y,X,Y)+\mathcal{L}_{GAN}(F,D_X,Y,X)+\lambda\mathcal{L}_{cyc}(G,F)]$
AdaIN	ICCV’17	GAN Loss+ Context Loss + Style Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)+\lambda_1\mathcal{L}_c(G)+\lambda_2\mathcal{L}_s(G)\right]$
WCT	NIPS’17	GAN Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$
AAMS	CVPR’19	GAN Loss + Context Loss + + Attention Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)+\lambda_1\mathcal{L}_{con}(G)+\lambda_2\mathcal{L}_{att}(G)\right]$
ASMA (proposed)	ACM MM’20	GAN Loss + Transform Loss + Style-aware Context Loss	$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)+\lambda_1\mathcal{L}_T(G)+\lambda_2\mathcal{L}_C(G)\right]$

Art Statement

This work present an novel Multi-Scale Projection Discriminator to realize the texture-level conditional generation. In contrast to the single-scale conditional discriminator, this discriminator is able to capture multi-scale texture clue to effectively distinguish a wide range of artistic styles. Their framework can transform a photograph into different artistic style oil painting via only ONE single model. Furthermore, the results are with distinctive artistic style and retain the anisotropic semantic information.



Method



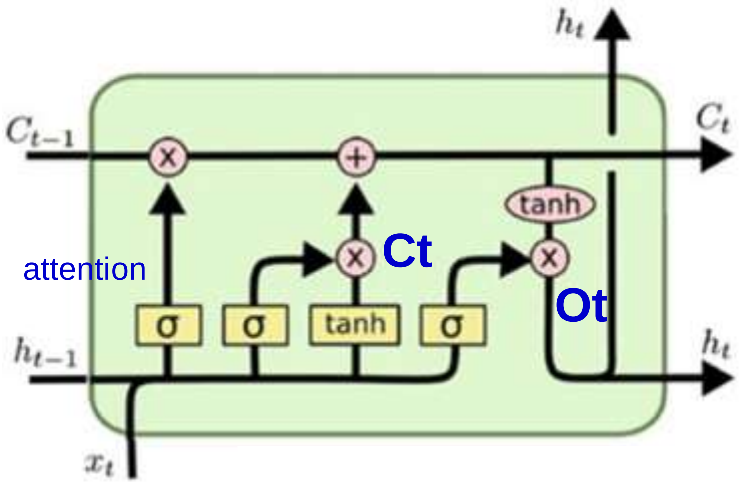
(2)
Conditional Resblock
label
CIN
x_s
Style Galleries
c_1
c_2
c_N
Conv
Resblock
Upsampling +Conv
Loss
Data Flow
Loss Flow
$\mathcal{L}_C$
(1)
ASM
c_i
x_c
Encoder
Resblocks
Decoder
x_o
$\mathcal{L}_D$
(3)
Multi-Scale Weight
Discriminator
$\mathcal{L}_T$

Method

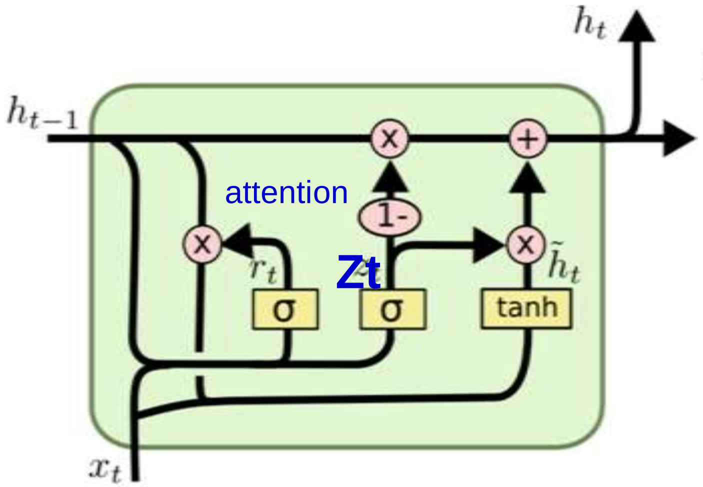
- **ASM:**

  ASM is a variant of GRU with much lighterweight. It can fuse features of two different scales to achieve dynamic adjustment of style-stroke.

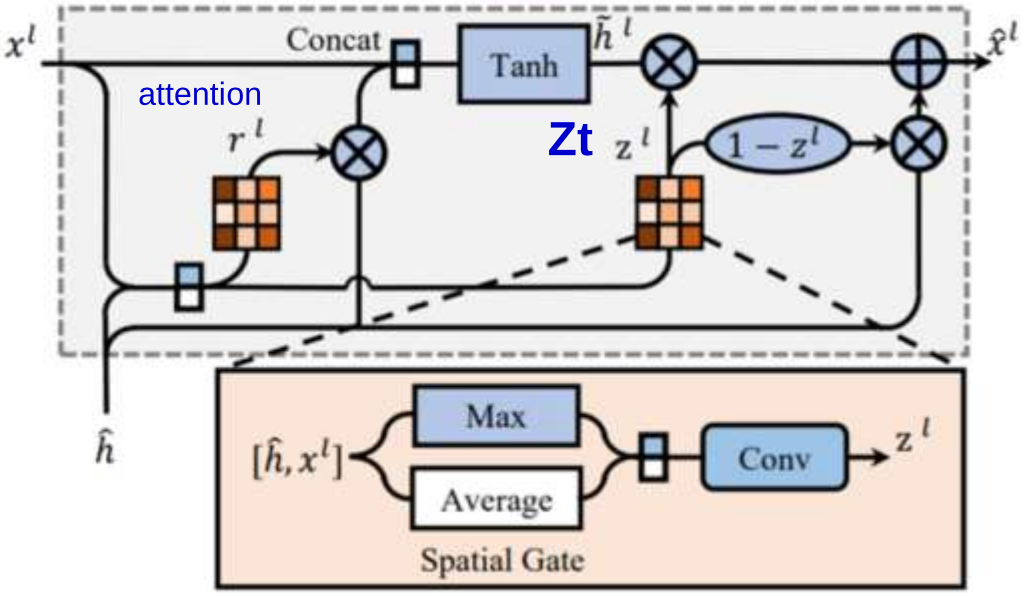
LSTM



GRU



ASM



CT: Current Memory State Gate

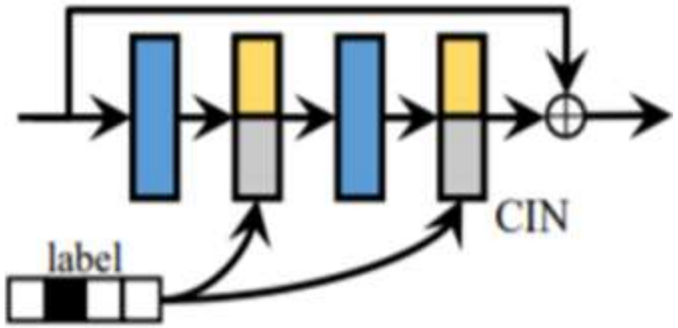
Ot: Forget Gate

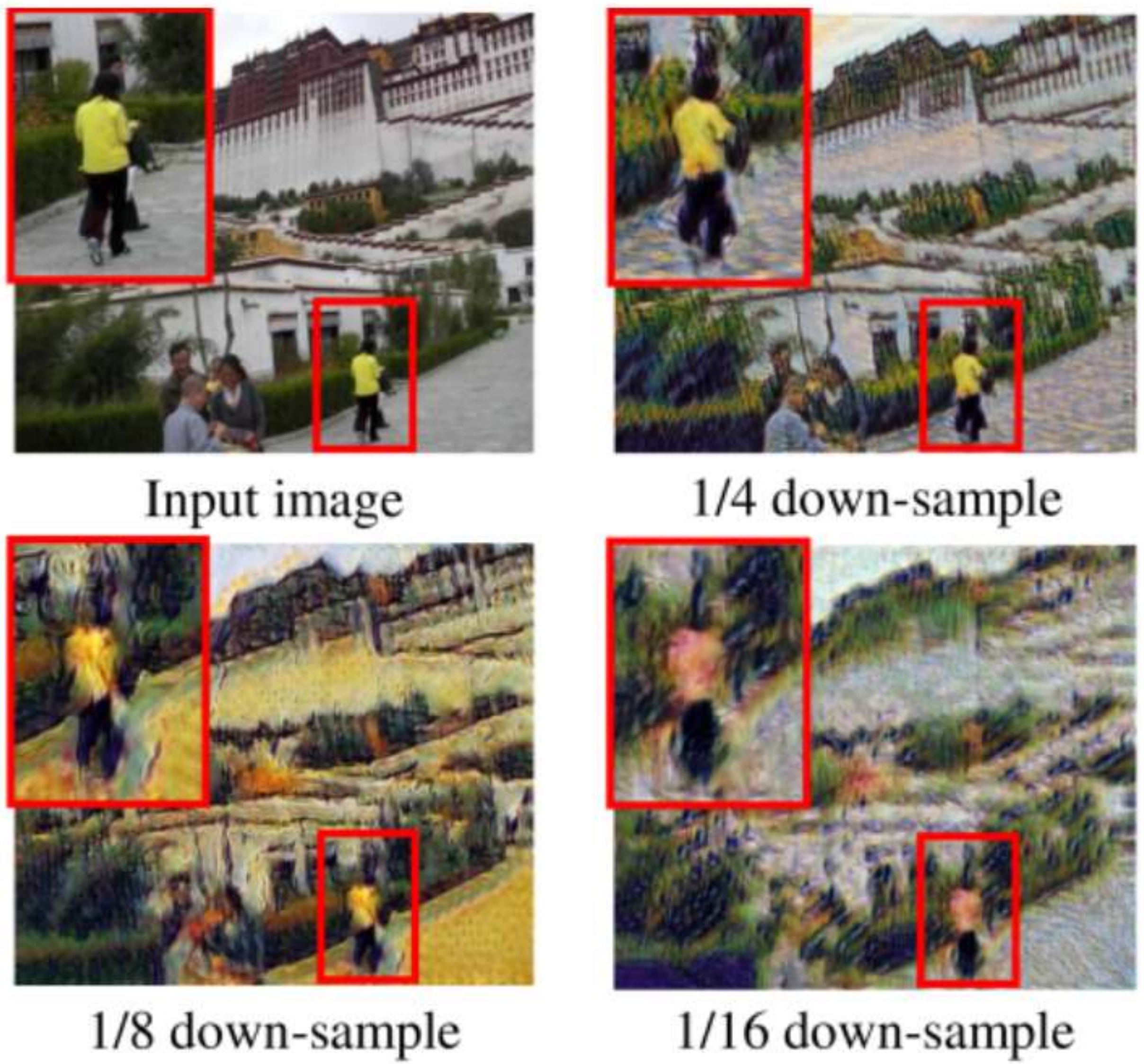
Zt: Update Gate

Method

- **Conditional ResBlock:**

  Different sizes of style-stroke regard to different down-sample rates. They put ResBlocks in places of different down-sample rates between the Encode and the Decoder to show the relationship between down-sample rate and stylization extent. As shown above, granularity of stylization increases as down-sample rate grows.

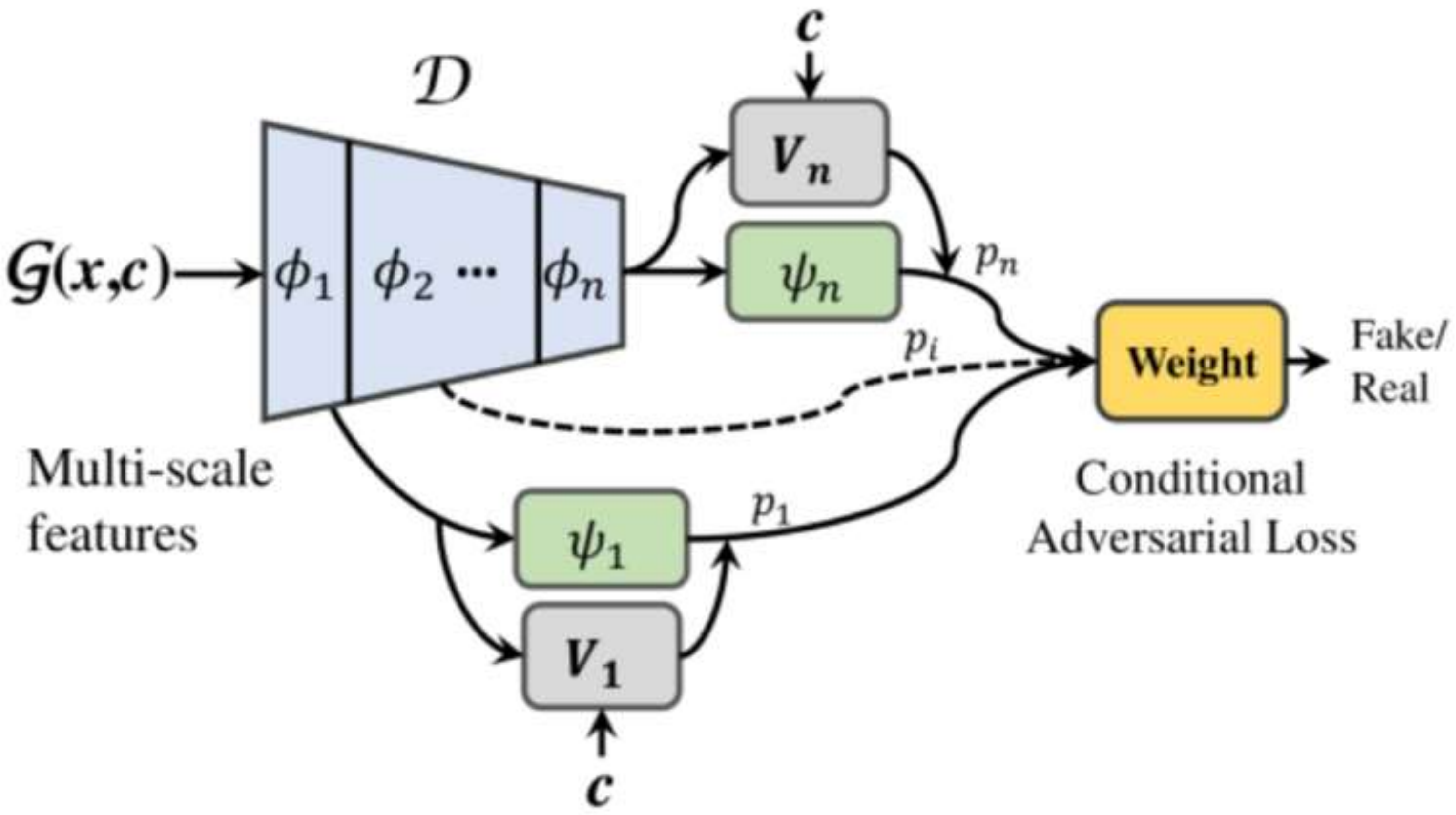




Method

- **Multi-Scale Projection Discriminator :**

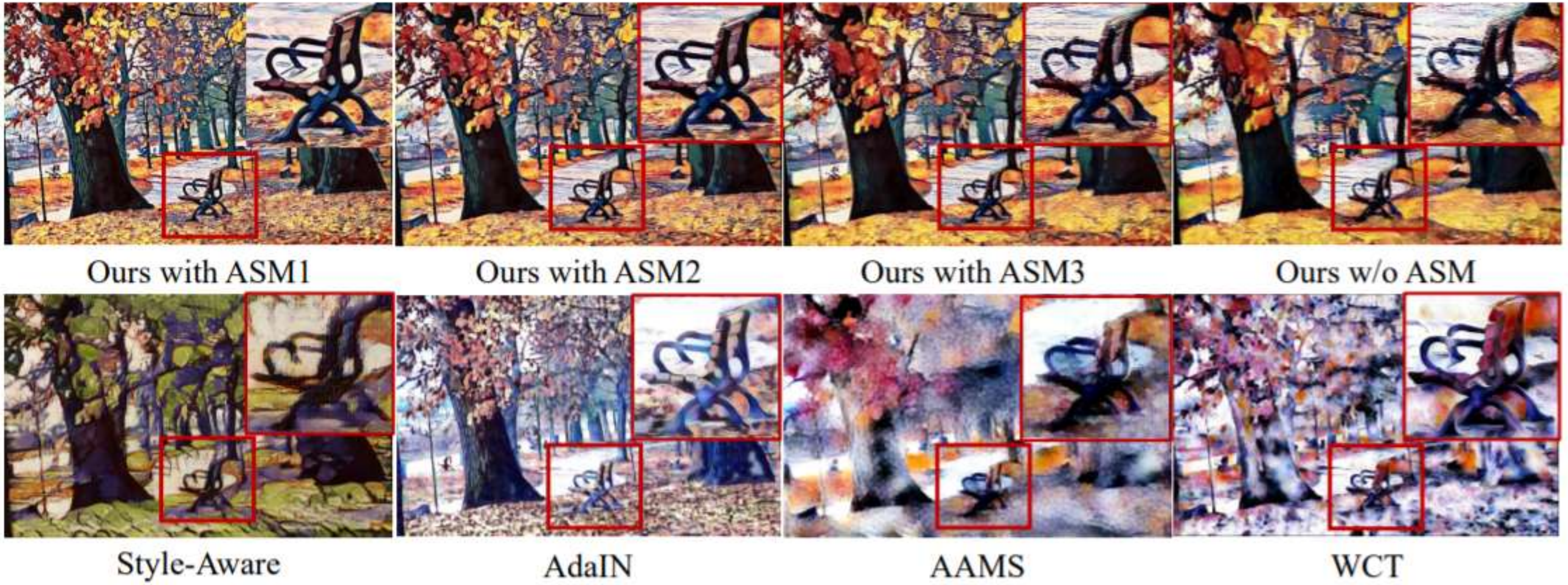
  Projection Discriminator comprehensively uses the features of different scales, which greatly strengthens the discriminator to recognize the stroke textons of different scales in the painting.



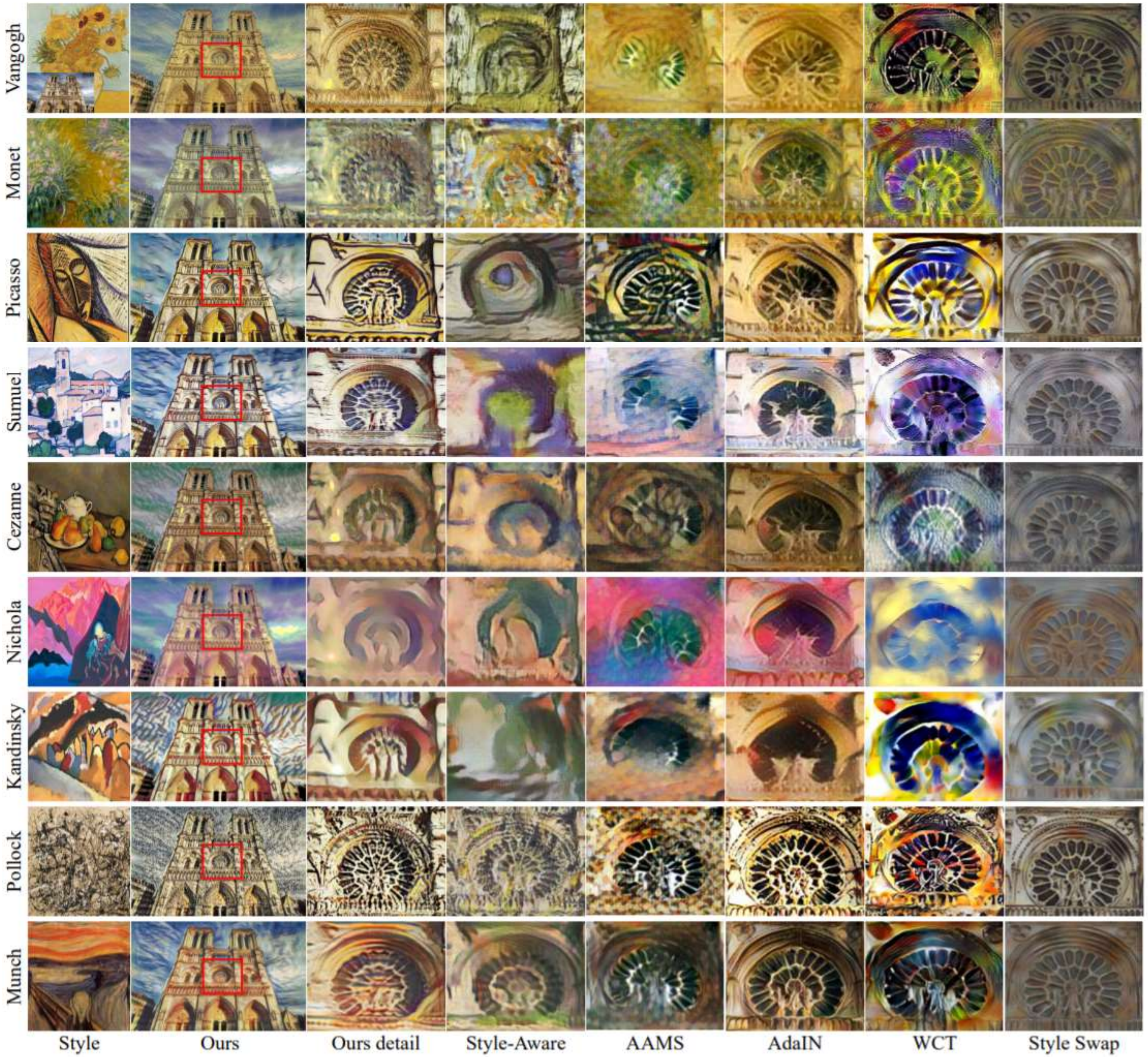
Result

- **Comparison with State-of-Art:**

  Comparison of anisotropic semantic preserving effect from Style-Aware, AAMS, AdaIN, WCT and ours. ASM1, ASM2, ASM3 indicate that ASM is placed in different layers of Generator.



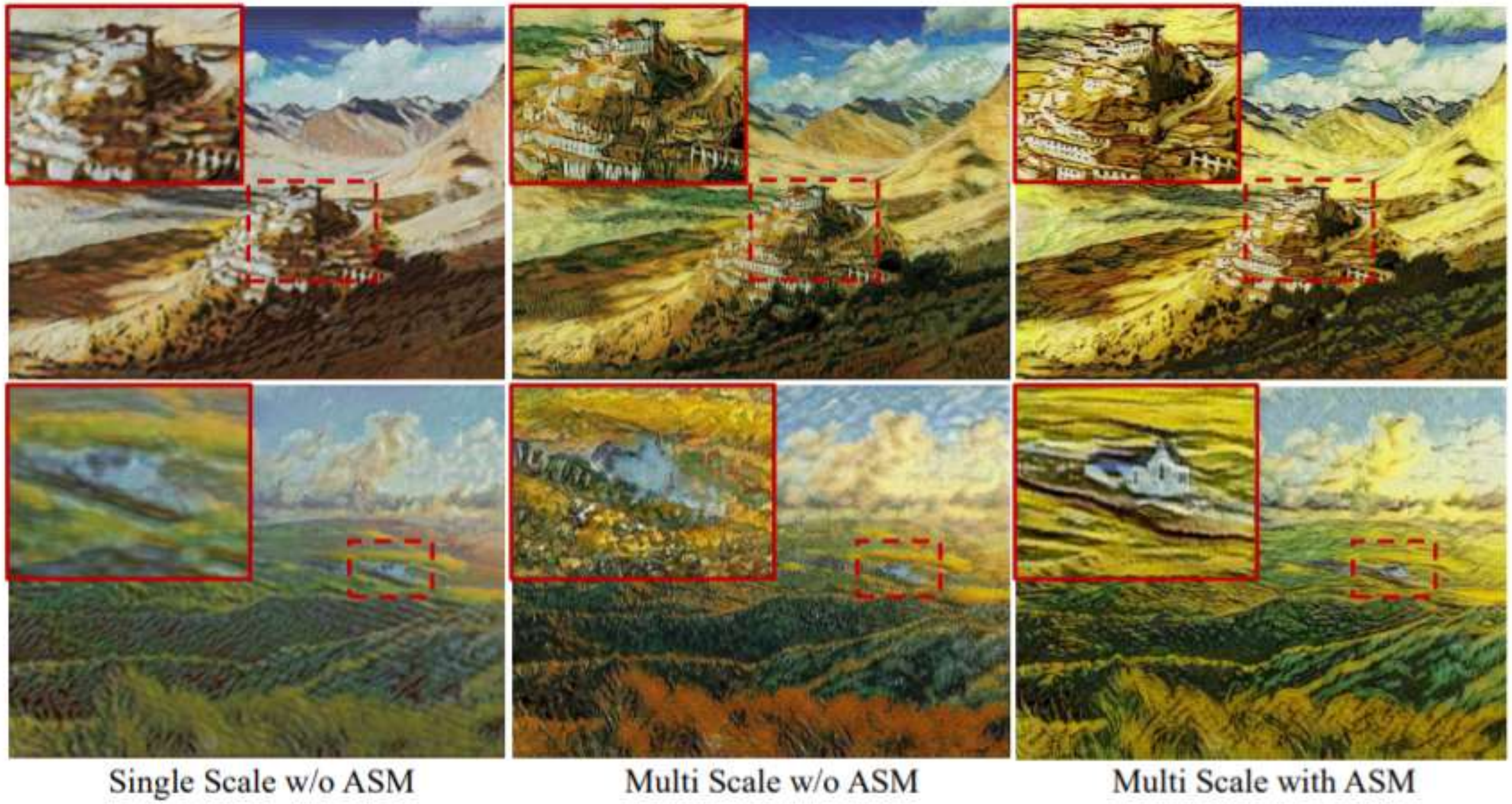
Result


Vangogh
Monet
Picasso
Sumuel
Cezanne
Nichola
Kandinsky
Pollock
Munch
Style
Ours
Ours detail
Style-Aware
AAMS
AdaIN
WCT
Style Swap

Result

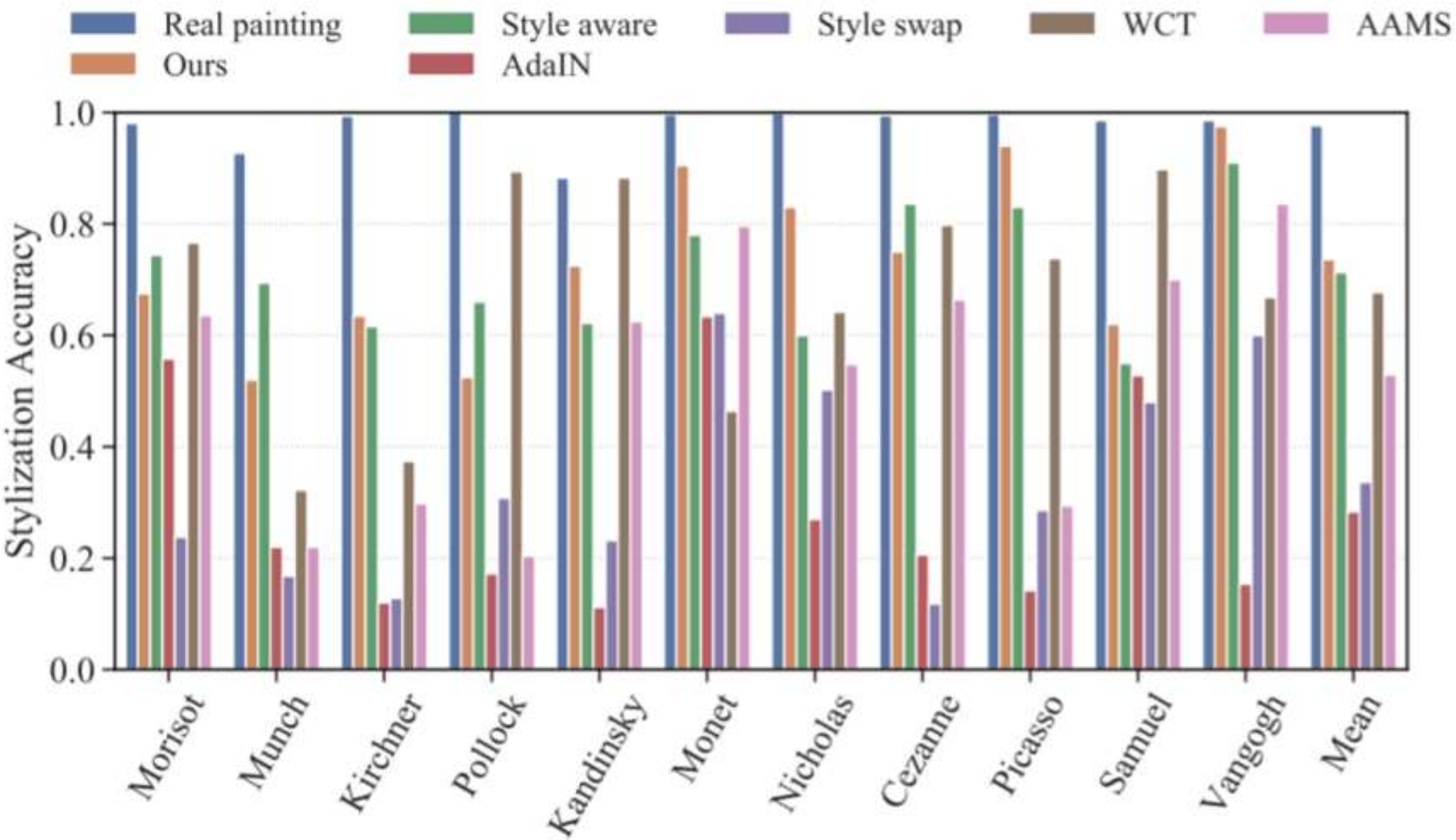
- **Comparison with Multi-Scale and Single-Scale Projection Discriminator :**

  Comparison of the Multi-Scale Projection Discriminator and the Single-Scale Projection Discriminator.



Result

- Author generate 200 result images for each artist's style, and measure the Style Accuracy of the stylization by sending these result images to the style classifier. The higher the accuracy of the classification result, the closer the class is to the corresponding painting style.



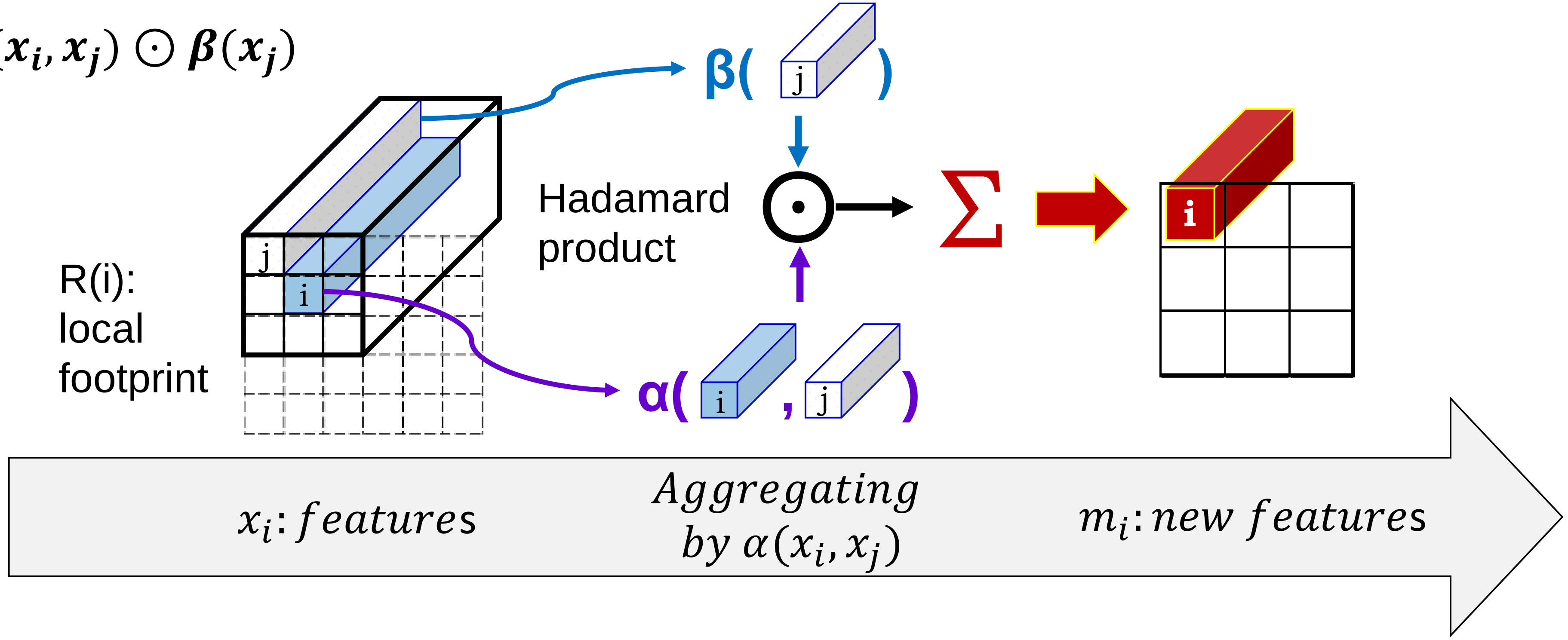
Connection

- Anisotropic Stroke Control
- One Model
- Not only Channel-wise Attention, but Spatial-wise Attention
- It might be better to add smooth L1-Loss

Connection

- **Spatial-wise Attention** generalizes standard dot-product attention and is fundamentally a set operator, which has the following form.
  - ✓ ⊙ is the Hadamard product, i is the spatial index of feature vector xi (i.e., its location in the feature map).
  - ✓ R(i) is the local footprint of the aggregation. The footprint R(i) is a set of indices that specifies which feature vectors are aggregated to construct the new feature mi.
  - ✓ The function β produces the feature vectors β(xj) that are aggregated by the adaptive weight vectors α(xi,xj).

$$\boldsymbol{m_i} = \sum_{j \in \boldsymbol{R}(\boldsymbol{i})} \boldsymbol{\alpha}(\boldsymbol{x_i}, \boldsymbol{x_j}) \odot \boldsymbol{\beta}(\boldsymbol{x_j})$$



Connection

- **Smooth L1 Loss**

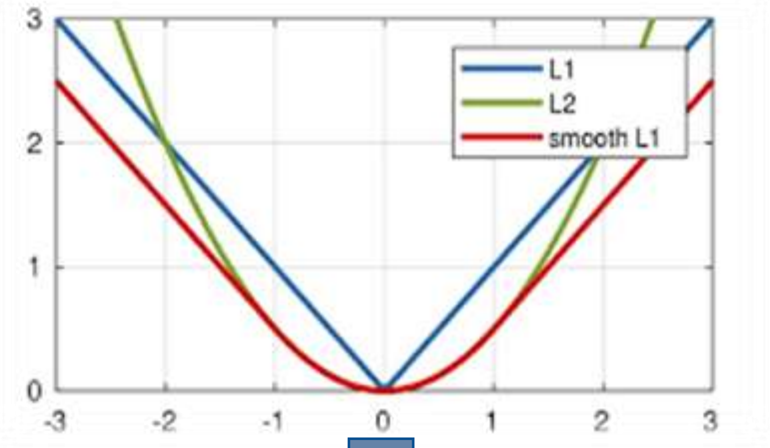
$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p(x)}\log\left(D_{\theta_d}(x|y)\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z|y)\right)\right)+\lambda Smooth\mathcal{L}_{L1}(G)\right]$$

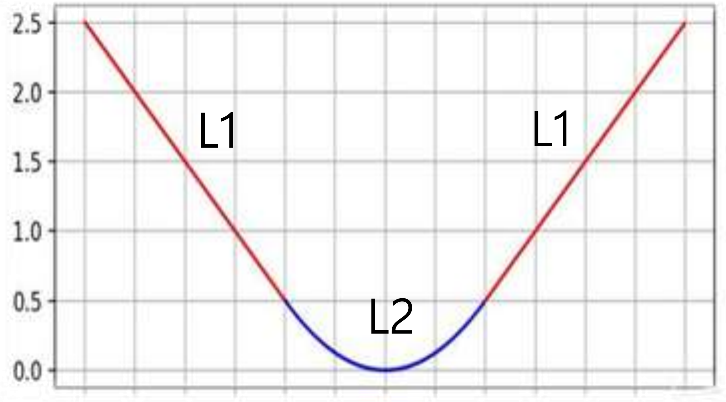
$$\mathcal{L}(\mathcal{F}(X,U;\theta),Y)=\sum_{h,w}\sum_{q}\ell_\delta(\mathcal{F}(X,U;\theta)_{h,w,q},Y_{h,w,q})$$

$$\ell_\delta(x,y)=\frac{1}{2}(x-y)^2\mathbb{1}_{\{|x-y|<\delta\}}+\delta(|x-y|-\frac{1}{2}\delta)\mathbb{1}_{\{|x-y|\geq\delta\}})$$

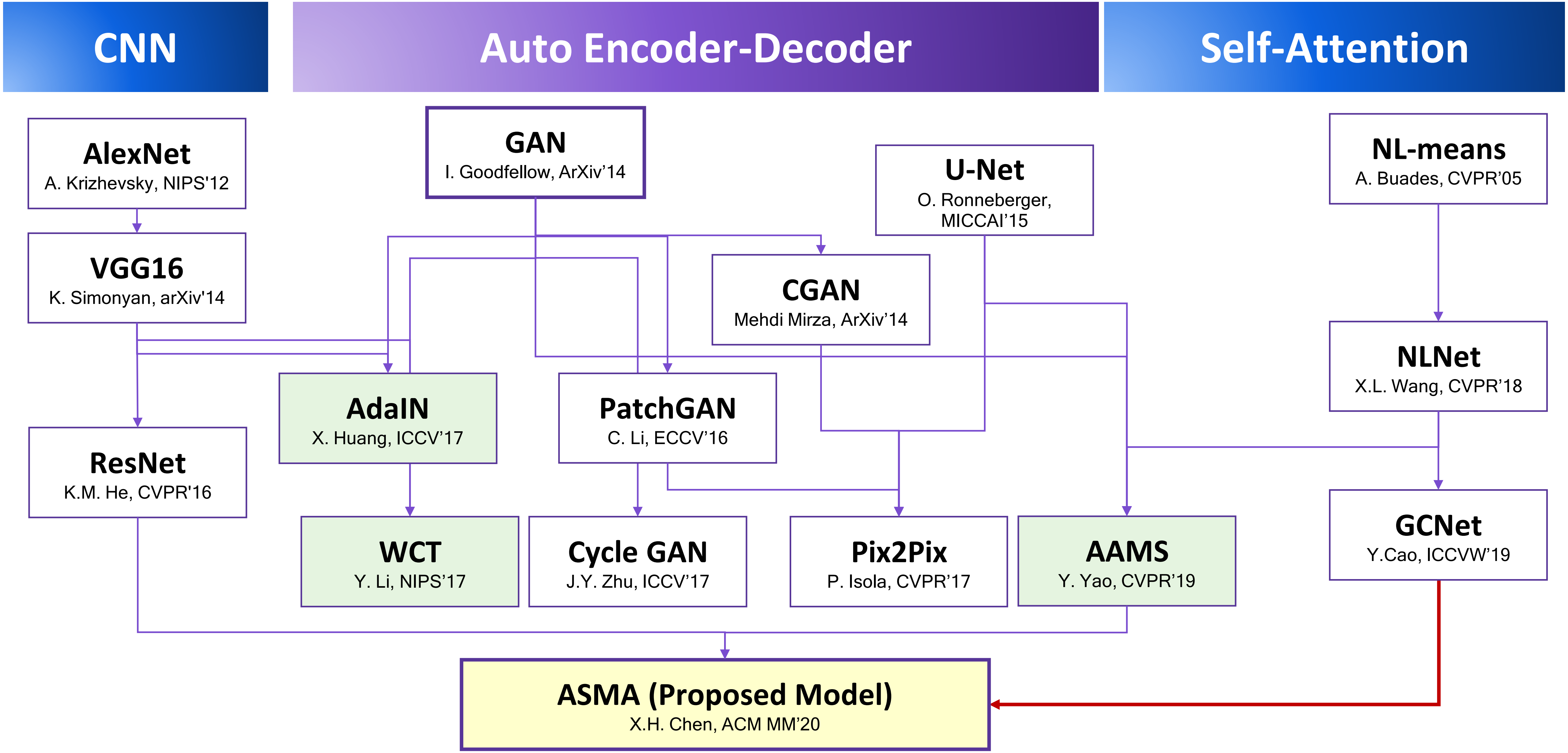
δ=1

Loss Function	Derivative of Loss Function
$L_2(x)=x^2$	$\frac{dL_2(x)}{dx}=2x$
$L_1(x)=\lvert x\rvert$	$\frac{dL_1(x)}{dx}=\begin{cases}1, if\ x\geq 0\\ -1, otherwise\end{cases}$
$smoothL_1(x)=\begin{cases}0.5x^2, if\lvert x\rvert<1\\ \lvert x\rvert-0.5, otherwise\end{cases}$	$\frac{dsmooth_{L1}(x)}{dx}=\begin{cases}\text{x}, if\ \lvert x\rvert<0\\ \pm 1, otherwise\end{cases}$





Demo



CNN
Auto Encoder-Decoder
Self-Attention
AlexNet
A. Krizhevsky, NIPS'12
VGG16
K. Simonyan, arXiv'14
ResNet
K.M. He, CVPR'16
GAN
I. Goodfellow, ArXiv'14
U-Net
O. Ronneberger, MICCAI'15
CGAN
Mehdi Mirza, ArXiv'14
AdaIN
X. Huang, ICCV'17
PatchGAN
C. Li, ECCV'16
WCT
Y. Li, NIPS'17
Cycle GAN
J.Y. Zhu, ICCV'17
Pix2Pix
P. Isola, CVPR'17
AAMS
Y. Yao, CVPR'19
NL-means
A. Buades, CVPR'05
NLNet
X.L. Wang, CVPR'18
GCNet
Y.Cao, ICCVW'19
ASMA (Proposed Model)
X.H. Chen, ACM MM'20

Berthe Morisot | Edvard Munch | Jackson Pollock | Ludwig Kirchner | Vangogh

Picasso Paul Cezanne Samuel

Kandinsky

Monet