Generative Adversarial Nets: Learning and Applications

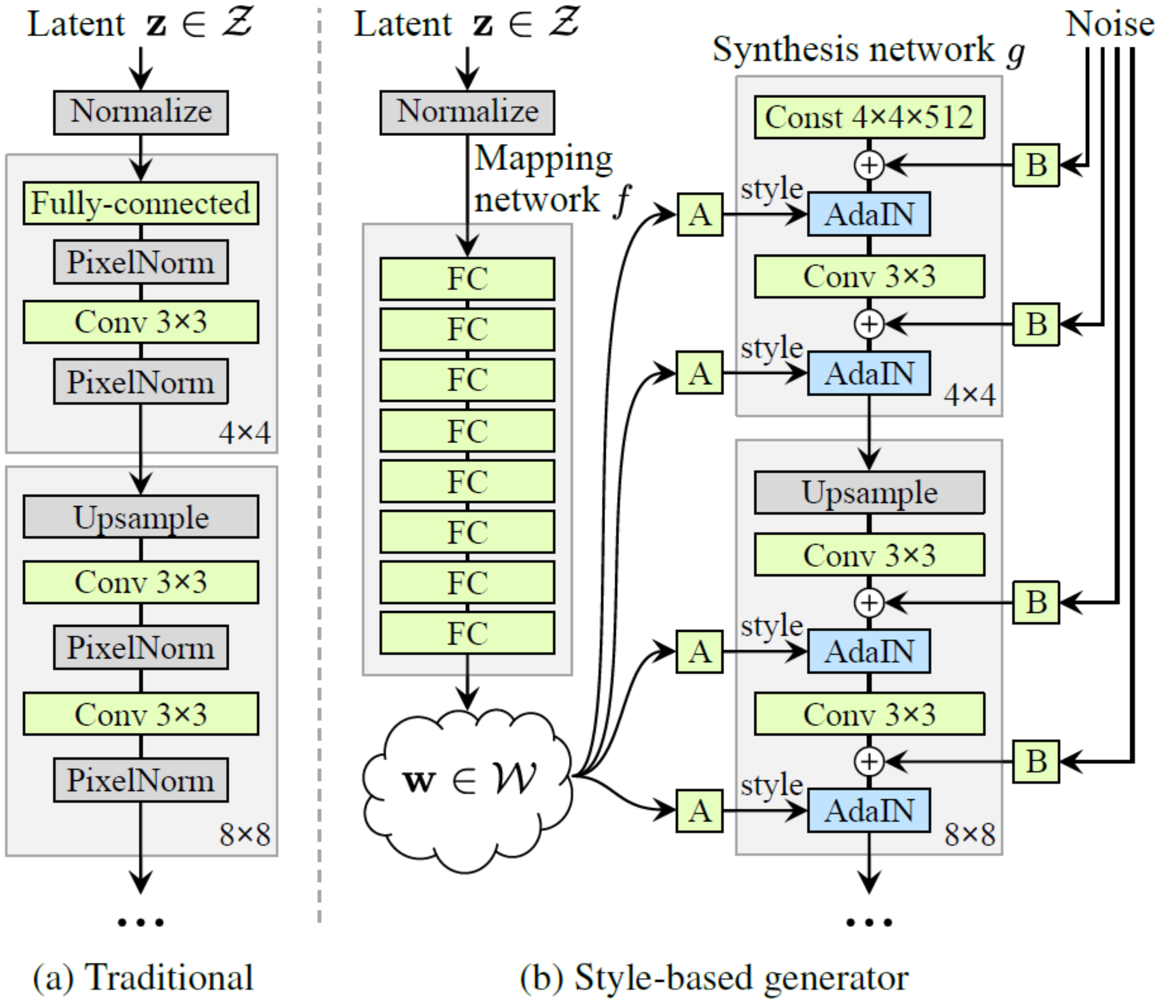
Wangmeng Zuo

Centre on Machine Learning Research
Harbin Institute of Technology

GAN Applications

StyleGAN (Karras, CVPR 2019)

- Progressive GAN
- Multi-layer manipulation on style and noise
- Truncation trick



Tero Karras, Samuli Laine, Timo Aila, A style-based generator architecture for generative adversarial networks, CVPR 2019.

Image2StyleGAN

- Robustness of latent space (Affine / Mask / Class)
- Analyze StyleGAN space and propose $\mathcal{W}^+$ space
- $\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathcal{L}_{percept}(G(\mathbf{w}), \mathbf{x}) + \lambda_{mse} \|G(\mathbf{w}) - \mathbf{x}\|_2^2$

Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?

Rameen Abdal
KAUST
rameen.abdal@kaust.edu.sa

Yipeng Qin
KAUST
yipeng.qin@kaust.edu.sa

Peter Wonka
KAUST
pwonka@gmail.com

Image2StyleGAN

- Applications *Style Mixing (last 9 layers)*

Image2StyleGAN

- Applications *Expression Transfer*

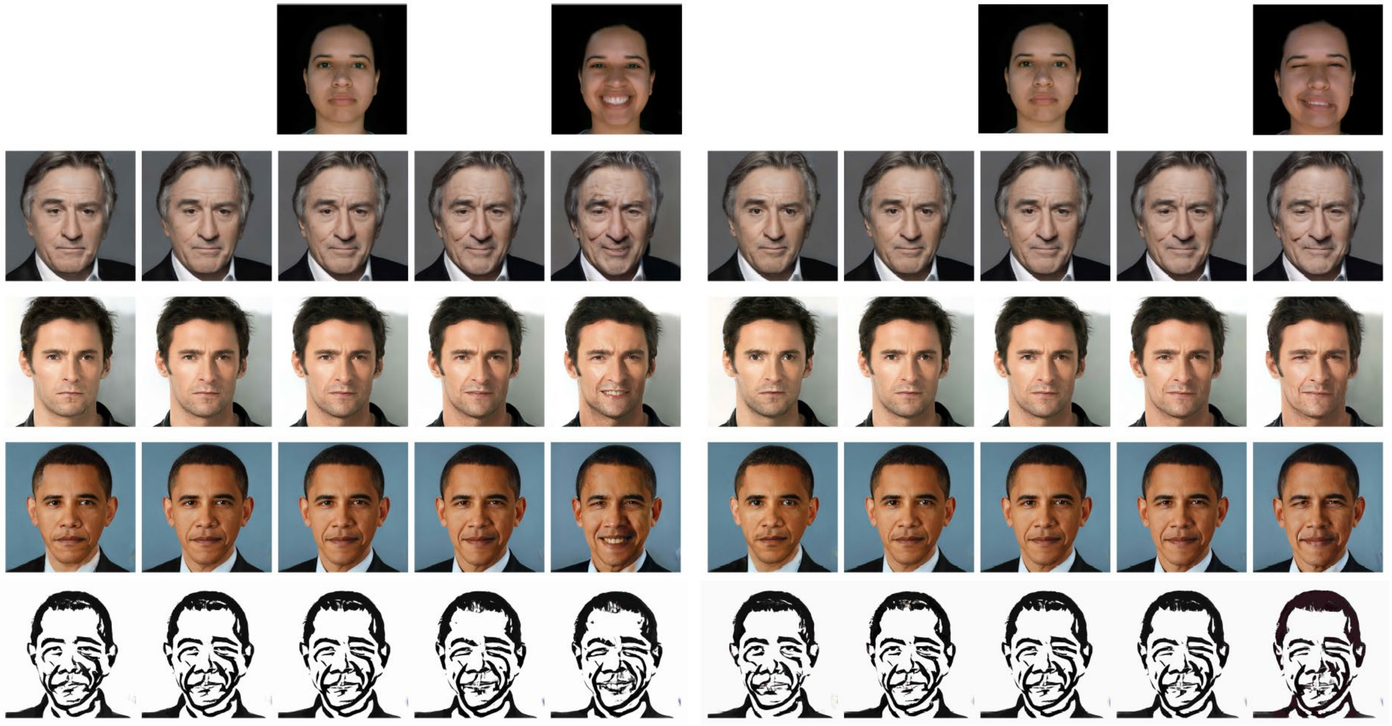


Image2StyleGAN++

- Introducing $\mathcal{N}$ space
- Optimization objective for editing

Image2StyleGAN++: How to Edit the Embedded Images?

Rameen Abdal
KAUST
rameen.abdal@kaust.edu.sa

Yipeng Qin
Cardiff University
qiny16@cardiff.ac.uk

Peter Wonka
KAUST
pwonka@gmail.com

Image2StyleGAN++

- Optimization objective for editing

$$\mathcal{L} = \lambda_{mse_1} \left\| M_m \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{x}) \right\|_2^2 + \lambda_p \mathcal{L}_{percept}(M_p, G(\mathbf{w},\mathbf{n}), \mathbf{x})$$

$$+\lambda_{mse_2} \left\| (1 - M_m) \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{y}) \right\|_2^2 + \lambda_s \mathcal{L}_{style}(M_s, G(\mathbf{w},\mathbf{n}), \mathbf{y})$$

where M_s, M_m, M_p are spatial masks, $\mathcal{L}_{style}$ means $conv3\_3$ of VGG-16, $\mathcal{L}_{percept}$ uses $conv1\_1,1\_2,2\_2,3\_3$ of VGG-16

  - $\mathcal{W}^+$ optimization: $\lambda_s = \lambda_{mse_2} = 0$, $\lambda_{mse_1} = \lambda_p = 10^{-5}$
  - $\mathcal{N}$ optimization: $\lambda_s = \lambda_p = 0$, $\lambda_{mse_1} = \lambda_{mse_2} = 10^{-5}$
  - Style Transfer: $\lambda_s = 5 \times 10^{-7}$, $\lambda_{mse_1} = \lambda_{mse_2} = \lambda_p = 0$

Image2StyleGAN++

$$\mathcal{L} = \lambda_{mse_1} \|M_m \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{x})\|_2^2 + \lambda_p \mathcal{L}_{percept}(M_p, G(\mathbf{w},\mathbf{n}), \mathbf{x})$$
$$+ \lambda_{mse_2} \|(1 - M_m) \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{y})\|_2^2 + \lambda_s \mathcal{L}_{style}(M_s, G(\mathbf{w},\mathbf{n}), \mathbf{y})$$

- Reconstruction

$\mathcal{W}^+$ for 5k iters, and $\mathcal{N}$ for 3k iters. PSNR 44~45 dB

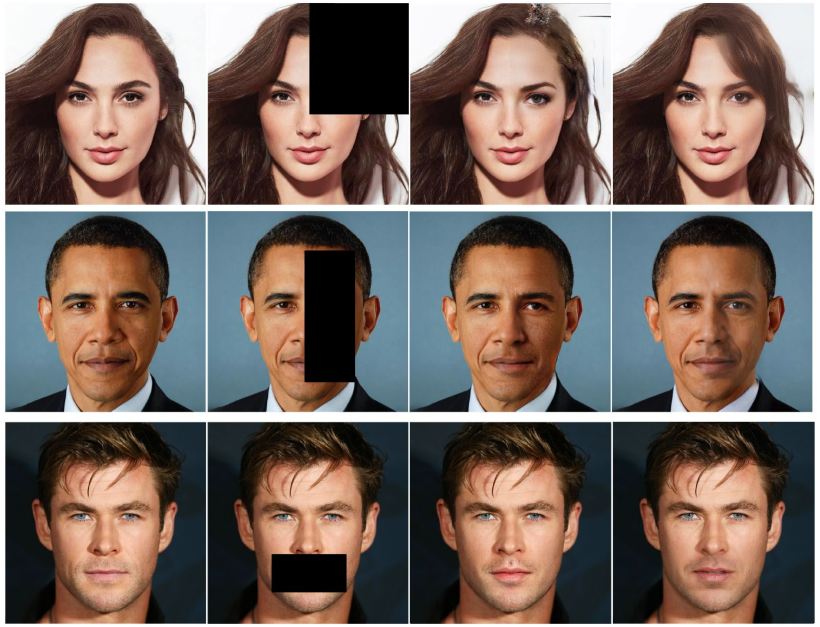
Input I2S I2S++

Image2StyleGAN++

$$\mathcal{L} = \lambda_{mse_1} \left\| M_m \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{x}) \right\|_2^2 + \lambda_p \mathcal{L}_{percept}(M_p, G(\mathbf{w},\mathbf{n}), \mathbf{x})$$
$$+ \lambda_{mse_2} \left\| (1 - M_m) \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{y}) \right\|_2^2 + \lambda_s \mathcal{L}_{style}(M_s, G(\mathbf{w},\mathbf{n}), \mathbf{y})$$

- Image Inpainting

Only optimize 1-9 and 17/18 th *w*



GT Input PConv I2S++

Algorithm 4: Image Inpainting

Input: image $I_{def} \in \mathbb{R}^{n \times m \times 3}$; masks M, M_{blur+}
Output: the embedded code (w_{out}, n_{out})

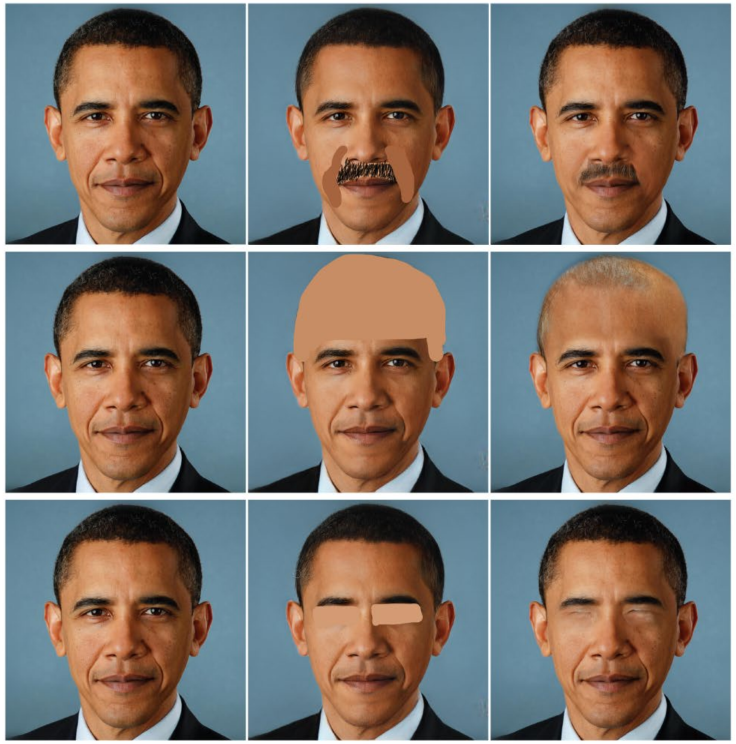
$(w_{ini}, n_{ini}) \leftarrow$ initialize();
$w_{out} = W_l(1 - M, 1 - M, w_m, w_{ini}, n_{ini}, I_{def})$;
$n_{out} =$
$Mk_n(1 - M_{blur+}, w_{out}, n_{ini}, I_{def}, G(w_{out}))$;

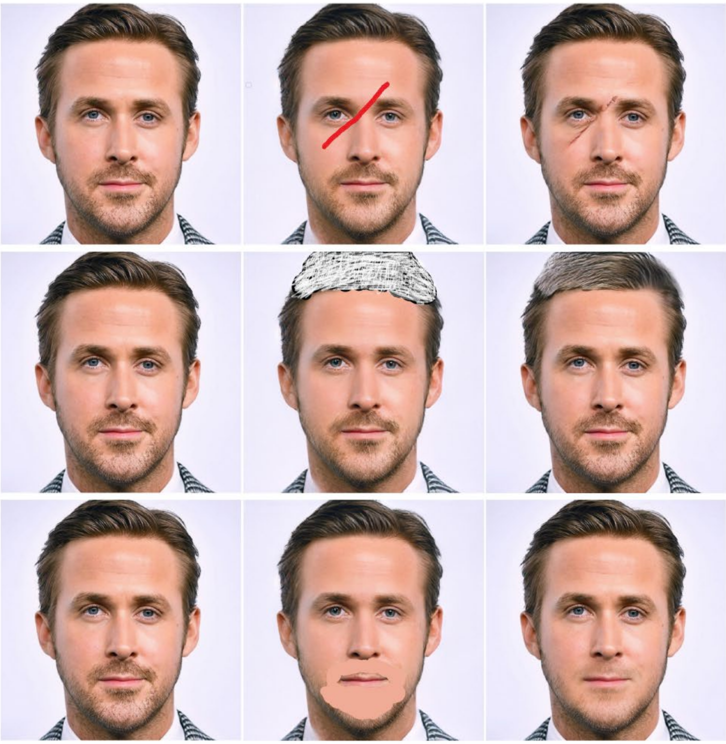
Image2StyleGAN++

$$\mathcal{L} = \lambda_{mse_1} \left\| M_m \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{x}) \right\|_2^2 + \lambda_p \mathcal{L}_{percept}(M_p, G(\mathbf{w},\mathbf{n}), \mathbf{x})$$
$$+ \lambda_{mse_2} \left\| (1 - M_m) \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{y}) \right\|_2^2 + \lambda_s \mathcal{L}_{style}(M_s, G(\mathbf{w},\mathbf{n}), \mathbf{y})$$

- Local editing

Only optimize 4-6 th w





Algorithm 5: Local Edits using Scribble

Input: image $I_{scr} \in \mathbb{R}^{n \times m \times 3}$; masks M_{blur}

Output: the embedded code (w_{out}, n_{out})

$(w^*, n_{ini}) \leftarrow$ initialize();

$w_{out} = W_l(1, 1, w_m, w^*, n_{ini}, I_{scr}) + \lambda \|w^* - w_{out}\|_2$;

$n_{out} = Mk_n(M_{blur}, w_{out}, n_{ini}, I_{scr}, G(w_{out}))$;

Image2StyleGAN++

$$\mathcal{L} = \lambda_{mse_1} \left\| M_m \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{x}) \right\|_2^2 + \lambda_p \mathcal{L}_{percept}(M_p, G(\mathbf{w},\mathbf{n}), \mathbf{x})$$
$$+\lambda_{mse_2} \left\| (1 - M_m) \odot (G(\mathbf{w},\mathbf{n}) - \mathbf{y}) \right\|_2^2 + \lambda_s \mathcal{L}_{style}(M_s, G(\mathbf{w},\mathbf{n}), \mathbf{y})$$

- Local style transfer

Algorithm 6: Local Style Transfer

Input: images $I_1, I_2 \in \mathbb{R}^{n \times m \times 3}$; masks M_{blur}

Output: the embedded code (w_{out}, n_{out})

1 $(w^*, n_{ini}) \leftarrow$ initialize();

2 $w_{out} = W_l(M_{blur}, M_{blur}, 1, w^*, n_{ini}, I_1)$
$+M_{st}(1 - M_{blur}, w^*, n_{ini}, I_2)$;

3 $n_{out} = Mk_n(M_{blur}, w_{out}, n_{ini}, I_1, G(w_{out}))$;

PULSE

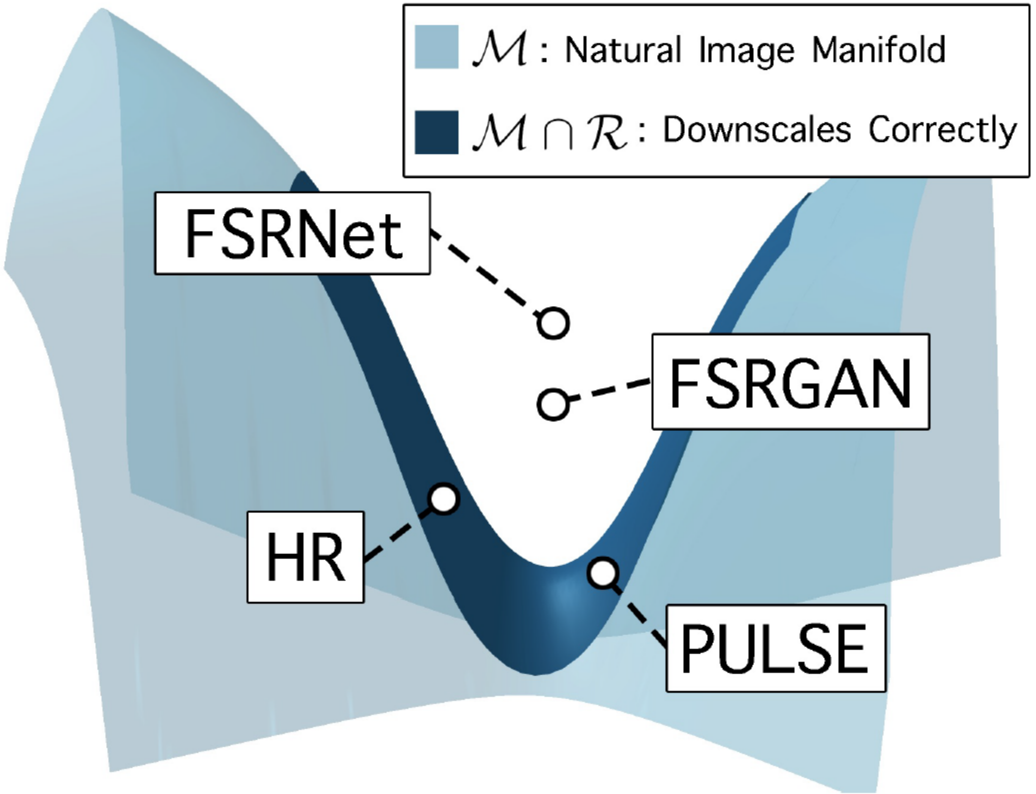
- Typical work using StyleGAN for SR (Obama SR)
- Down-scale Loss

PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models

Sachit Menon*, Alexandru Damian*, Shijia Hu, Nikhil Ravi, Cynthia Rudin
Duke University
Durham, NC
{sachit.menon,alexandru.damian,shijia.hu,nikhil.ravi,cynthia.rudin}@duke.edu

PULSE

- Natural Image Manifold by *pre-trained GAN*



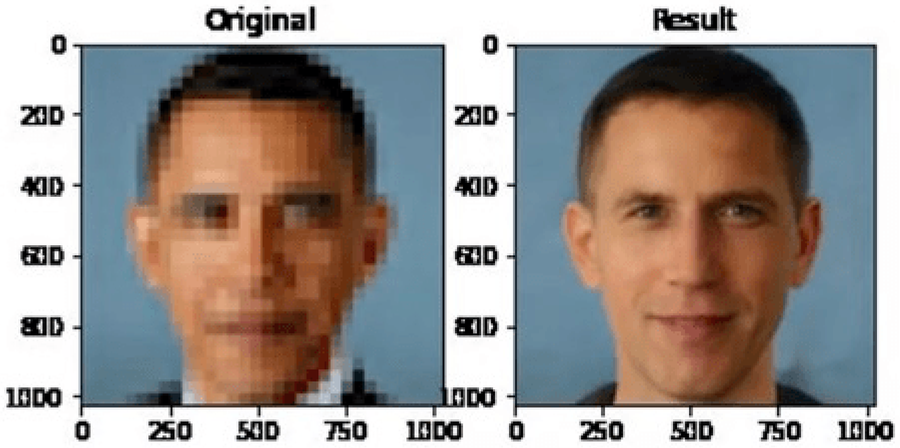
PULSE

- Optimization objective

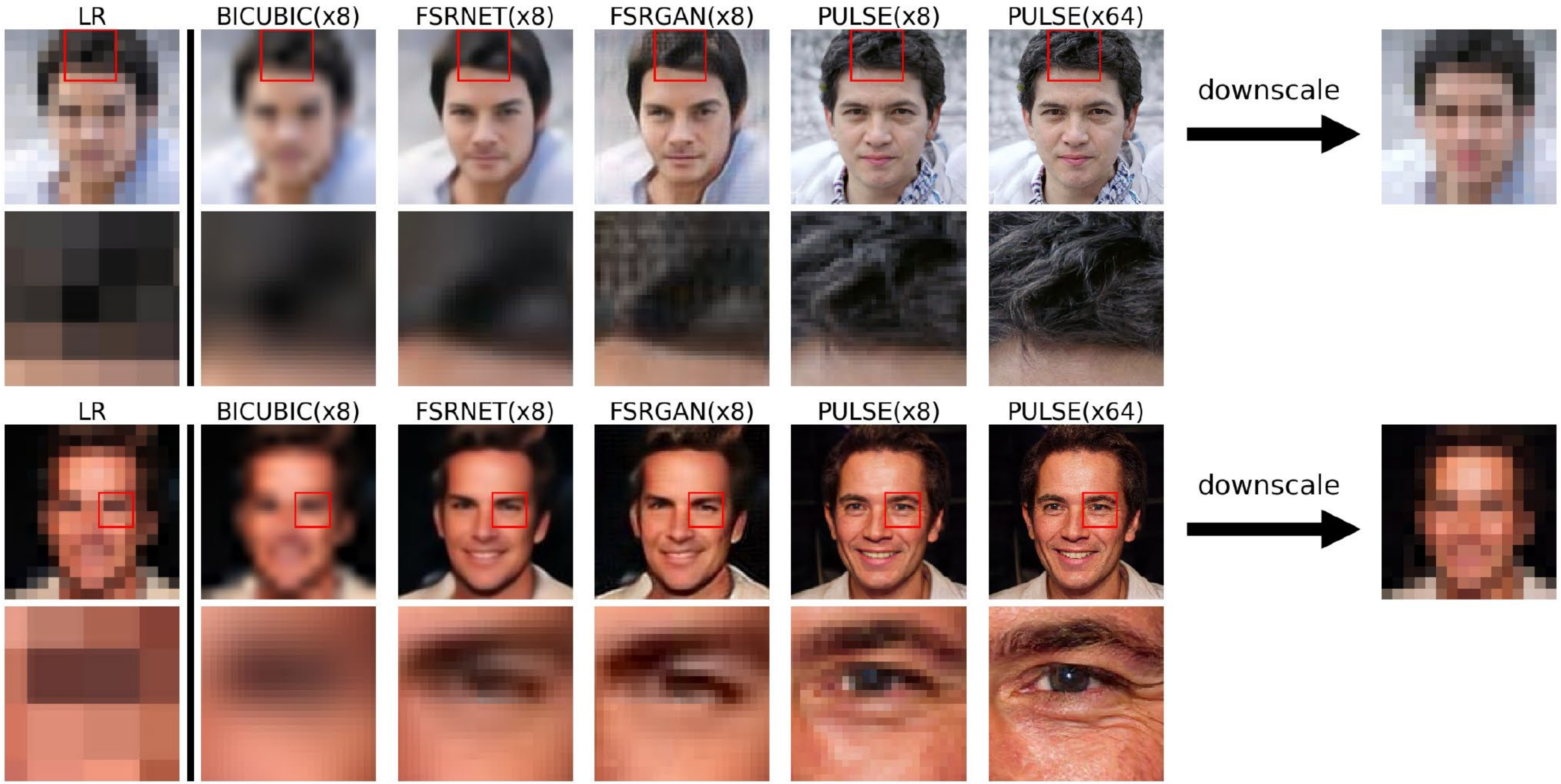
$$\mathcal{L}_{DS}(I_{SR}, I_{LR}) = \left\| DS(I_{SR}) - I_{LR} \right\|_p^p$$

- *Gaussian Prior Regularization*

$$\mathcal{L}' = \sqrt{d}\, S^{d-1}$$



PULSE


LR
BICUBIC(x8)
FSRNET(x8)
FSRGAN(x8)
PULSE(x8)
PULSE(x64)
downscale
LR
BICUBIC(x8)
FSRNET(x8)
FSRGAN(x8)
PULSE(x8)
PULSE(x64)
downscale

Pre-trained GANs for Restoration

pSp

- Learning-based Encoder $pSp(\mathbf{x}) \coloneqq G(E(\mathbf{x}) + \overline{\mathrm{w}})$
- A U-Net shaped Encoder
- LPIPS Loss + $\overline{w}$ regularization + ArcFace-based ID Loss
- *"Invert first, edit later"* → *One stop operation*

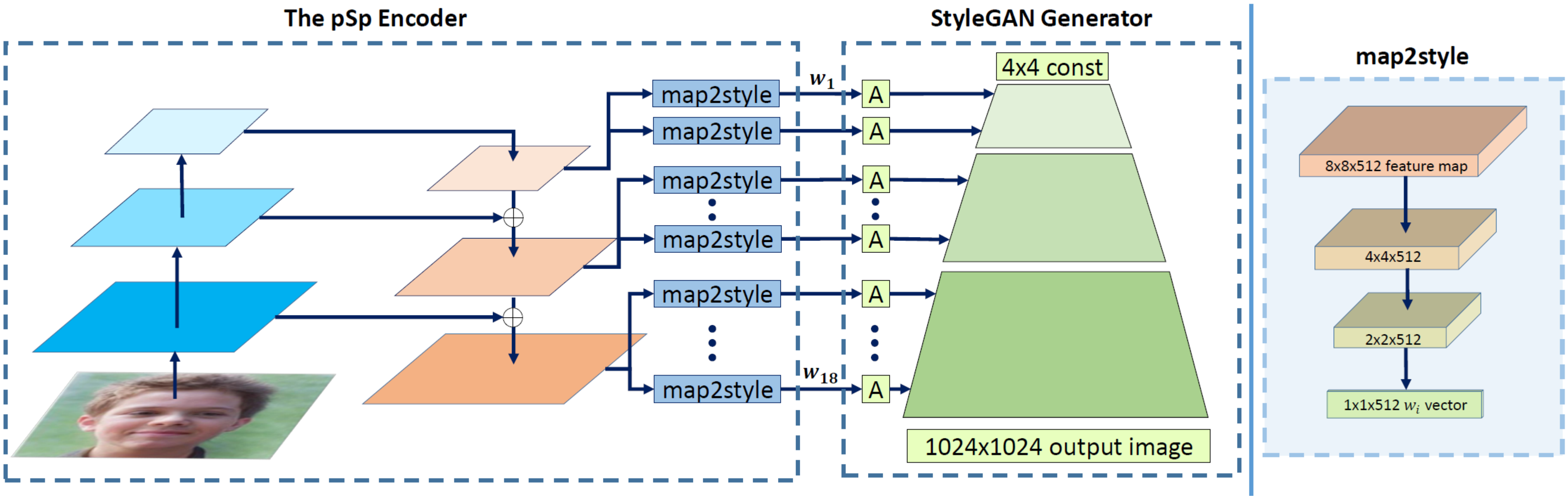
Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation

Elad Richardson[1] Yuval Alaluf[1,2] Or Patashnik[1,2] Yotam Nitzan[2]
Yaniv Azar[1] Stav Shapiro[1] Daniel Cohen-Or[2]

[1]Penta-AI [2]Tel-Aviv University

pSp

- Left: ResNet-50 coarse/medium/fine: 0-2/3-6/7-18



pSp

- Loss function
  - L2

$$\mathcal{L}_2(\mathbf{x}) = \left\| \mathbf{x} - pSp(\mathbf{x}) \right\|_2, \text{ where } pSp(\mathbf{x}) = G(E(\mathbf{x}) + \overline{\mathbf{w}})$$

  - LPIPS

$$\mathcal{L}_{LPIPS}(\mathbf{x}) = \left\| F(\mathbf{x}) - F(pSp(\mathbf{x})) \right\|_2$$

  - $\overline{\mathbf{w}}$ Regularization

$$\mathcal{L}_{reg}(\mathbf{x}) = \left\| E(\mathbf{x}) - \overline{\mathbf{w}} \right\|_2$$

  - ID (R is pre-trained ArcFace Model)

$$\mathcal{L}_{id} = 1 - \left\langle R(\mathbf{x}), R(pSp(\mathbf{x})) \right\rangle$$

pSp

- Reconstruction

Method	↑ Similarity	↓ LPIPS	↓ MSE	↓ Runtime
Karras *et al.* [21]	0.77	0.11	0.02	182.1
ALAE [32]	0.06	0.32	0.15	0.207
IDInvert [42]	0.18	0.22	0.06	0.032
pSp	0.56	0.17	0.03	0.105

Table 1. Quantitative results for image reconstruction.

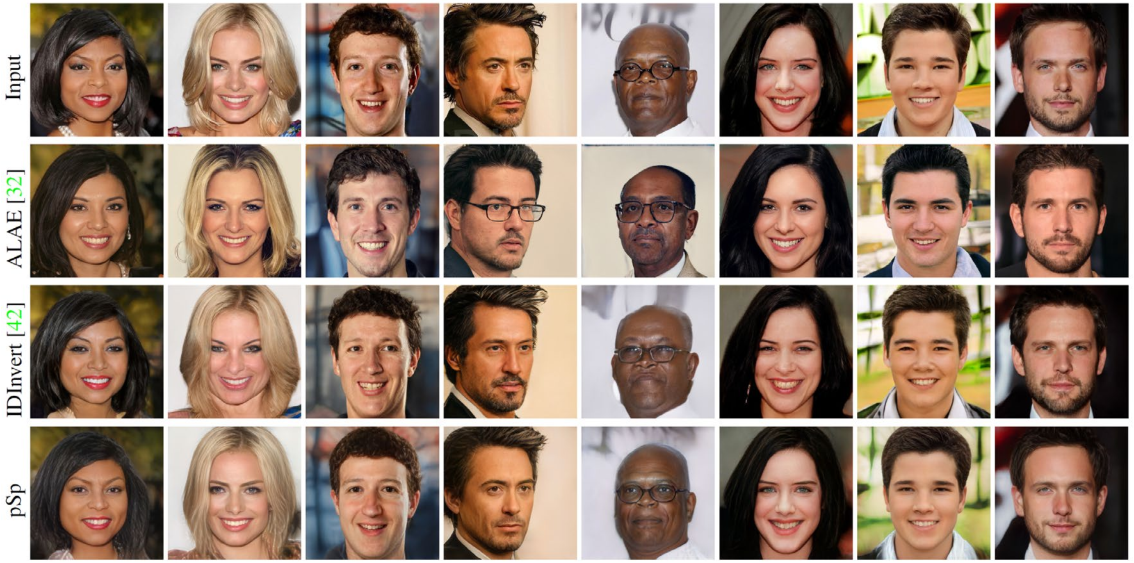
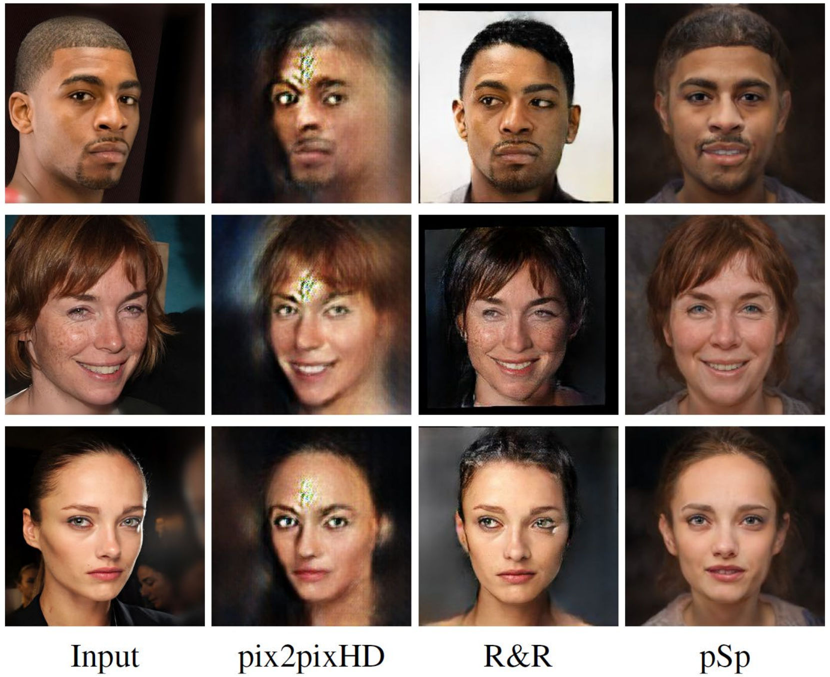


Figure 4. Results of pSp for StyleGAN inversion compared to other encoders on CelebA-HQ.

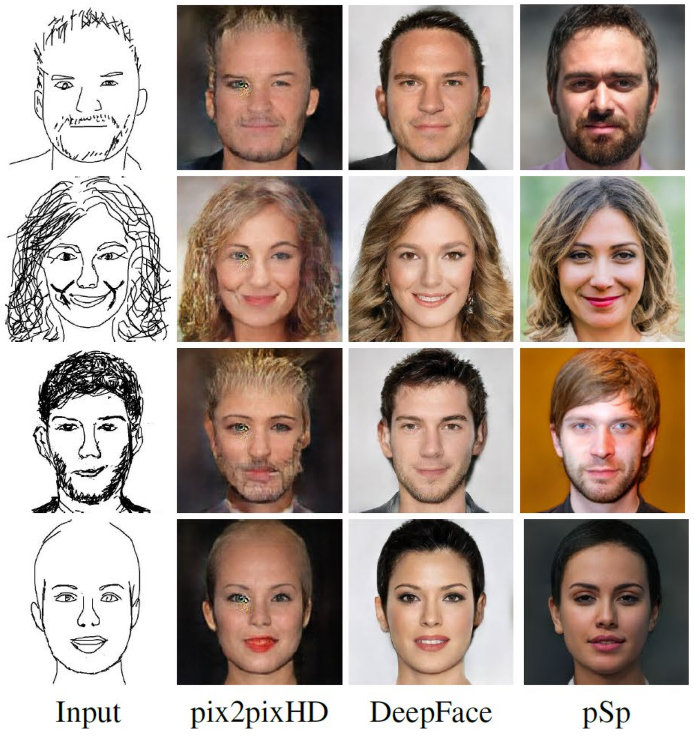
Face Frontalization (*random flip*)

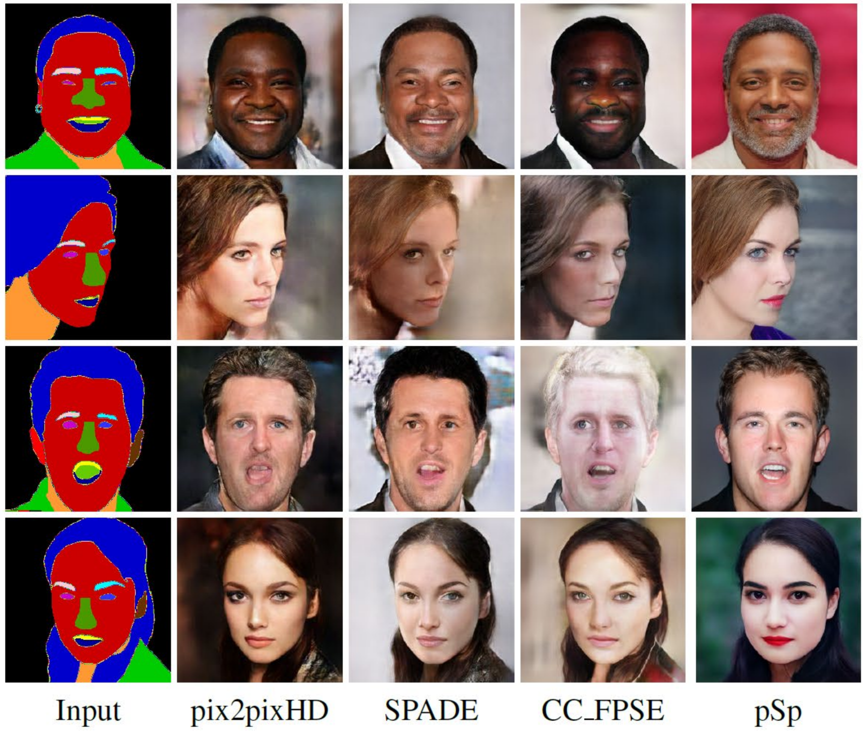


pSp

- sketch & label to image

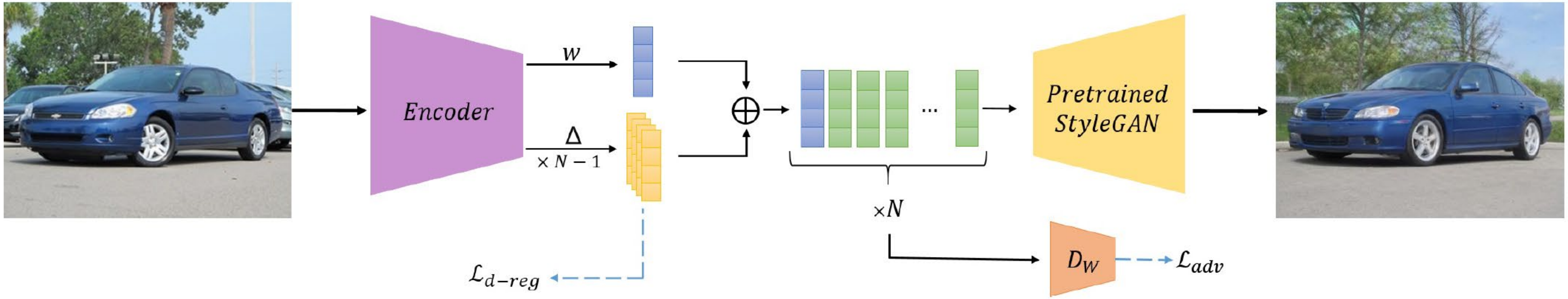
$$w = [w_{input}^{1-7}, w_{rand}^{8-18}]$$





E4E: encoder4editing

- Minimize Variation: Outputs a single style code together with a set of offsets



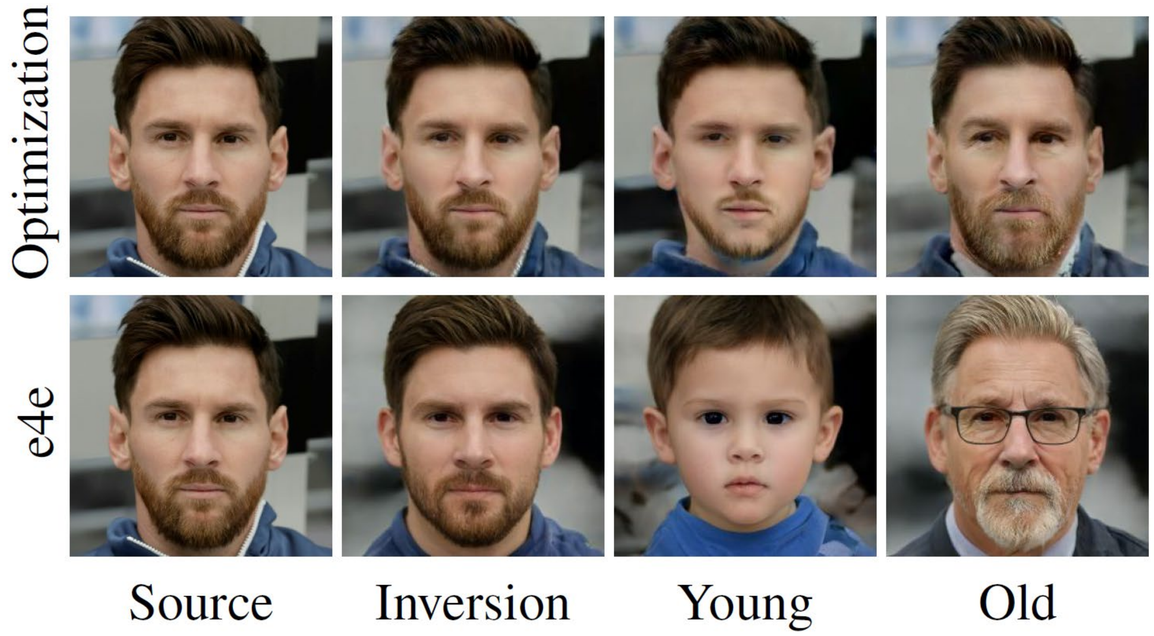
$$\mathcal{L}_{\text{adv}}^{D} = -\underset{w\sim\mathcal{W}}{\mathbb{E}}[\log D_{\mathcal{W}}(w)] - \underset{x\sim p_X}{\mathbb{E}}[\log(1 - D_{\mathcal{W}}(E(x)_i)] + \frac{\gamma}{2}\underset{w\sim\mathcal{W}}{\mathbb{E}}\left[\|\nabla_w D_{\mathcal{W}}(w)\|_2^2\right], \quad (2)$$

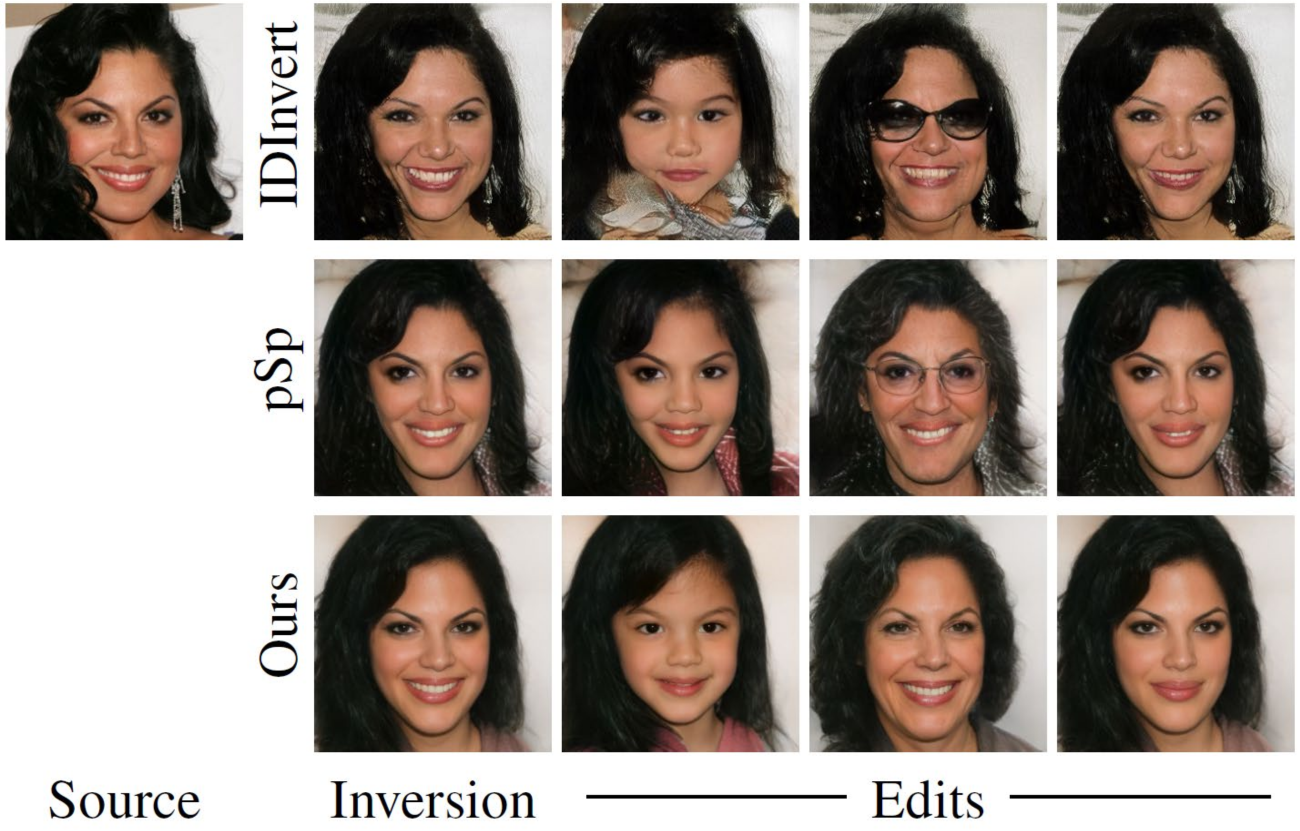
- Minimize Deviation

$$\mathcal{L}_{\text{adv}}^{E} = -\underset{x\sim p_X}{\mathbb{E}}[\log D_{\mathcal{W}}(E(x)_i)]. \quad (3)$$

Designing an Encoder for StyleGAN Image Manipulation, ACM TOG 2021

E4E: Results


Optimization
e4e
Source
Inversion
Young
Old


IDInvert
pSp
Ours
Source
Inversion
Edits

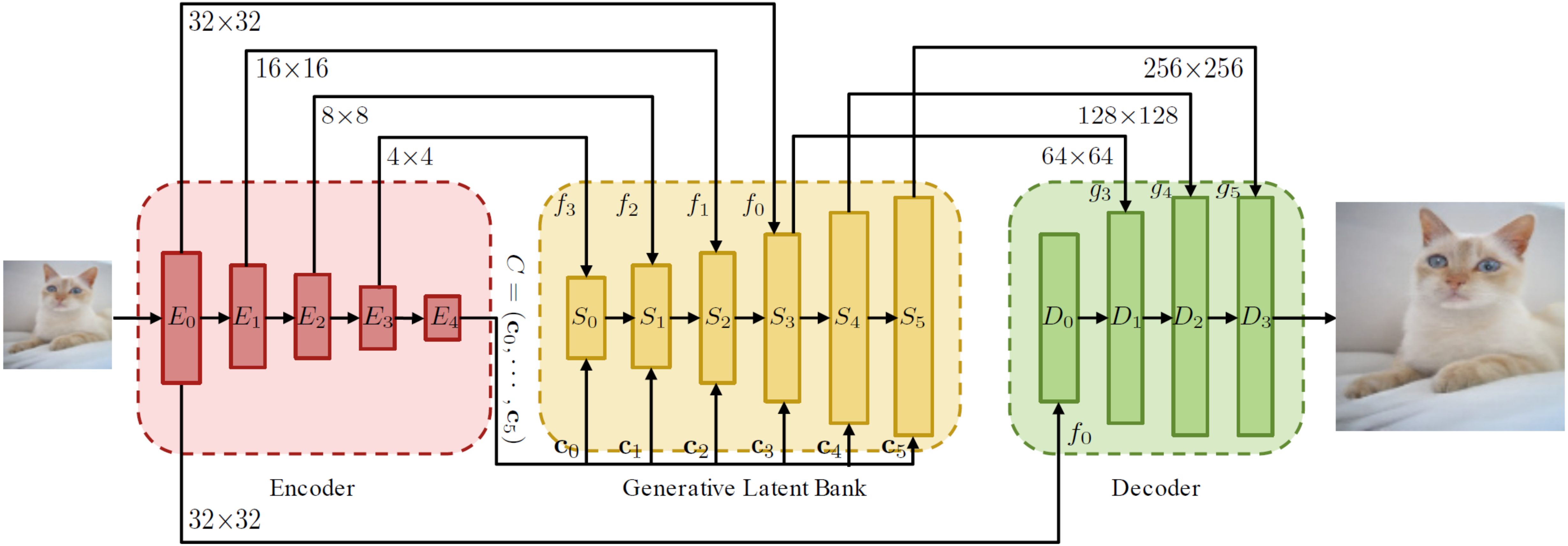
GLEAN

- Learning-based encoder for IR (so many in CVPR 2021)
- Extra refinement network behind StyleGAN

GLEAN: Generative Latent Bank for Large-Factor Image Super-Resolution

Kelvin C.K. Chan[1] Xintao Wang[2] Xiangyu Xu[1] Jinwei Gu[3,4] Chen Change Loy[1*]
[1]S-Lab, Nanyang Technological University
[2]Applied Research Center, Tencent PCG [3]Tetras.AI. [4]Shanghai AI Laboratory
{chan0899, xiangyu.xu, ccloy}@ntu.edu.sg xintao.wang@outlook.com gujinwei@tetras.ai

GLEAN (l2+perceptual+adversarial loss)



GLEAN

- SR

(a) Low-Resolution (b) ESRGAN (c) PULSE (d) GLEAN (ours) (e) Ground-truth

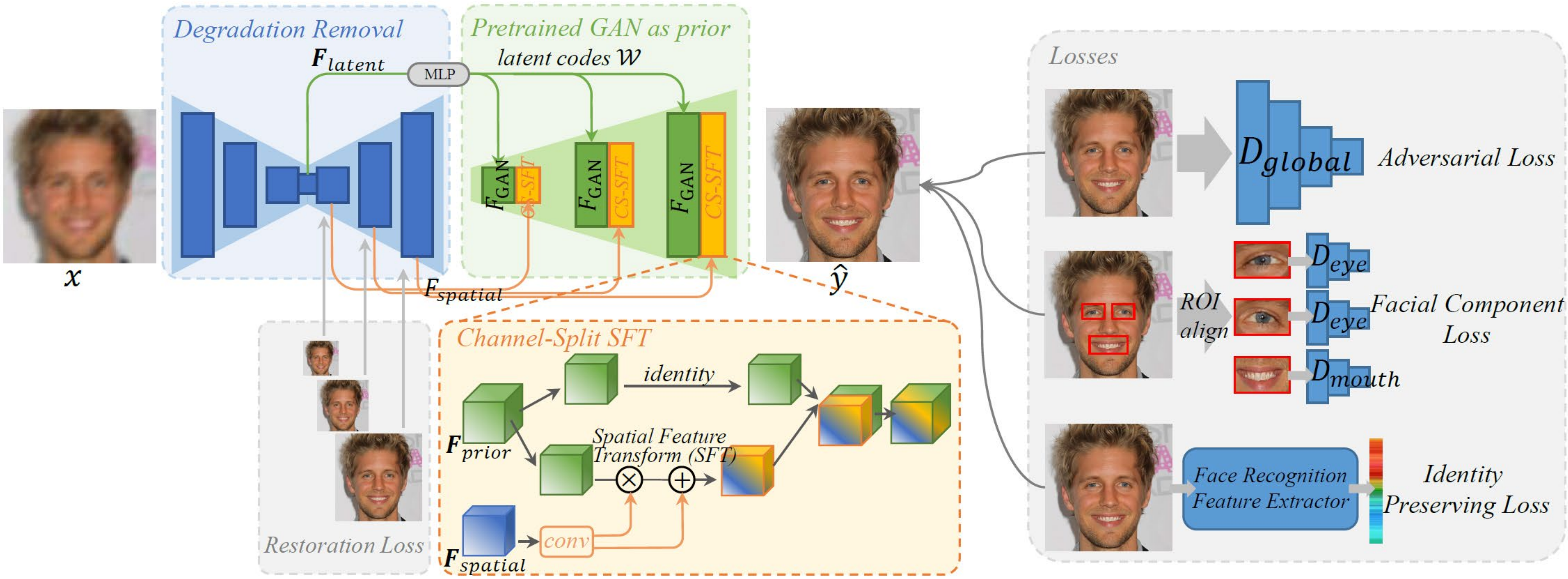
GFPGAN

- A U-Net based Encoder for *degradation removal*
- Applying Spatial Feature Transform (STF) in Pre-trained GAN

Towards Real-World Blind Face Restoration with Generative Facial Prior

Xintao Wang Yu Li Honglun Zhang Ying Shan
Applied Research Center (ARC), Tencent PCG
{xintaowang, ianyli, honlanzhang, yingsshan}@tencent.com

GFPGAN


Degradation Removal
F_{latent}
MLP
Pretrained GAN as prior
latent codes $\mathcal{W}$
F_{GAN}
CS-SFT
x
$\hat{y}$
$F_{spatial}$
Restoration Loss
Channel-Split SFT
identity
F_{prior}
Spatial Feature
Transform (SFT)
conv
$F_{spatial}$
Losses
D_{global}
Adversarial Loss
ROI
align
D_{eye}
D_{eye}
D_{mouth}
Facial Component
Loss
Face Recognition
Feature Extractor
Identity
Preserving Loss

GFPGAN

- Learning Objective

$$\mathcal{L}_{rec} = \lambda_{l1} \left\| \hat{y} - y \right\|_1 + \lambda_{per} \left\| \phi(\hat{y}) - \phi(y) \right\|_1$$

$$\mathcal{L}_{adv} = -\lambda_{adv} \mathbb{E}_{\hat{y}} \mathrm{softplus}(D(\hat{y}))$$

$$\mathcal{L}_{comp} = \sum_{\mathrm{ROI}} \lambda_{local} \mathbb{E}_{\hat{y}_{\mathrm{ROI}}} [\log(1 - D_{\mathrm{ROI}}(\hat{\mathbf{y}}_{\mathrm{ROI}})] + \lambda_{fs} \left\| Gram(\psi(\hat{y}_{ROI})) - Gram(\psi(y_{ROI})) \right\|_1$$

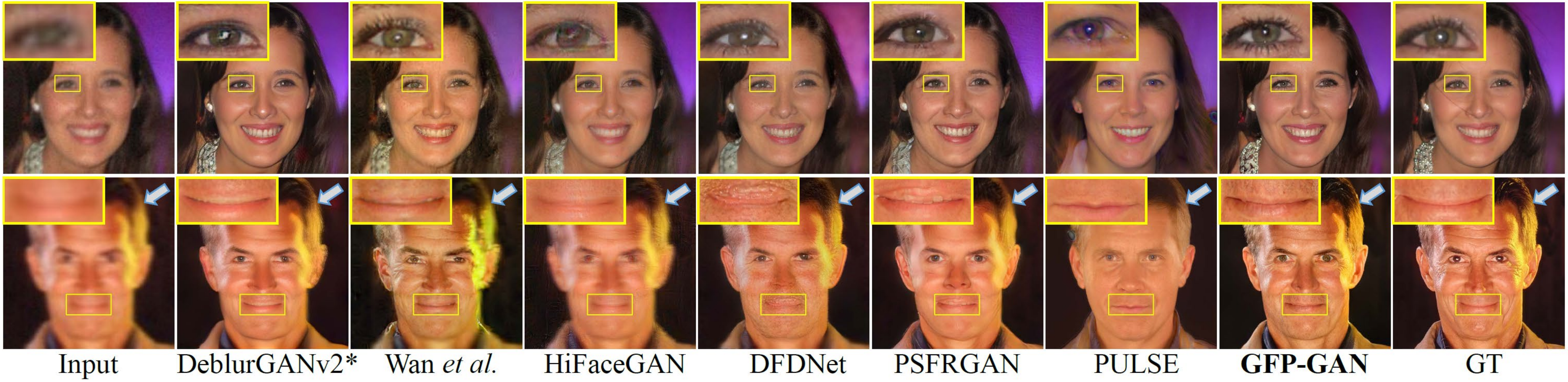
- Degradation Model

$$\mathbf{x} = [(\mathbf{y} * \mathbf{k}_{\sigma}) \downarrow_r + \mathbf{n}_{\delta}]_{JPEG_q}$$

$$\sigma = \{0.2 : 10\}, r = \{1 : 8\}, \delta = \{0 : 15\}, q = \{60 : 100\}$$

GFPGAN

- SR results



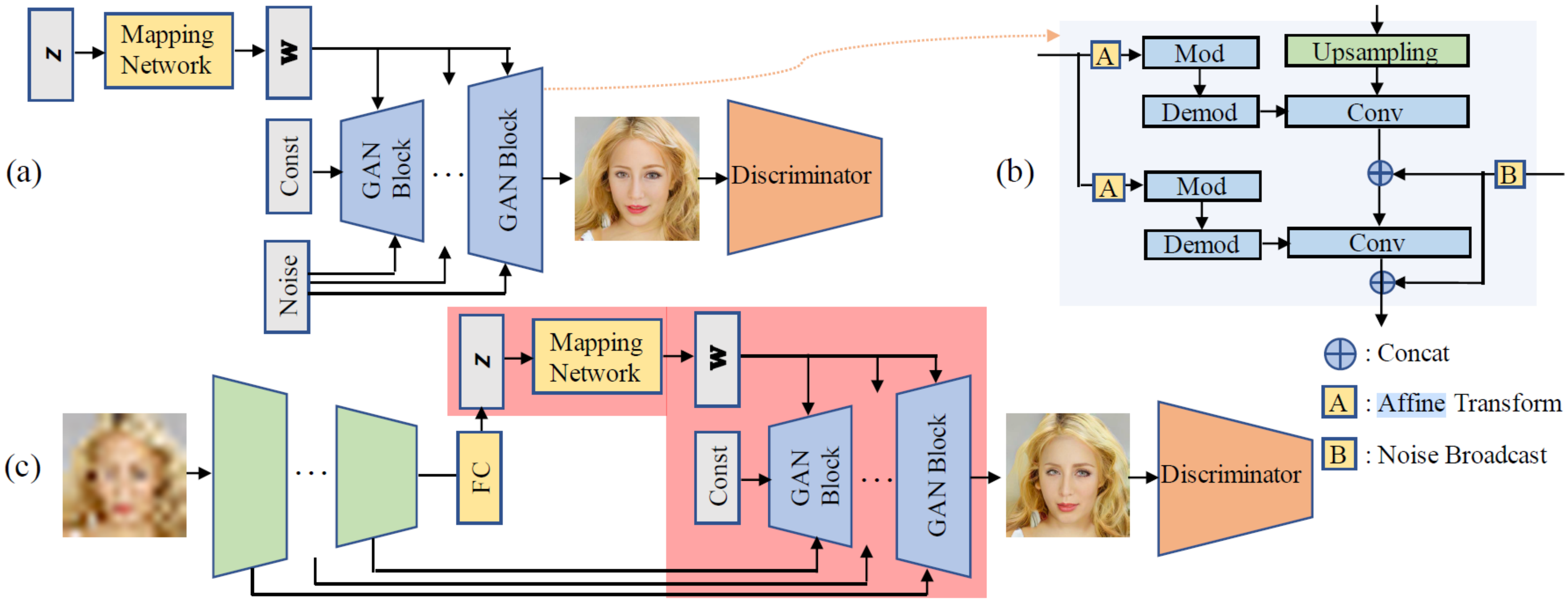
GPEN

- Both $\mathcal{W}^+$ and $\mathcal{N}$ space
- Jointly fine-tune the Generator

GAN Prior Embedded Network for Blind Face Restoration in the Wild

Tao Yang[1], Peiran Ren[1], Xuansong Xie[1], and Lei Zhang[1,2]*
[1]DAMO Academy, Alibaba Group
[2]Department of Computing, The Hong Kong Polytechnic University
yangtao9009@gmail.com, peiran_r@sohu.com, xingtong.xxs@taobao.com, cslzhang@comp.polyu.edu.hk

GPEN


(a)
z
Mapping Network
w
Const
GAN Block
Noise
GAN Block
Discriminator
(b)
A
Mod
Upsampling
Demod
Conv
A
Mod
B
Demod
Conv
: Concat
A : Affine Transform
B : Noise Broadcast
(c)
z
Mapping Network
w
FC
Const
GAN Block
GAN Block
Discriminator

GPEN

- Learning Objective ($LR_{enc}: LR_{dec}: LR_{dis} = 100: 10: 1$)

$$L_C = \min_G \left\| X - \tilde{X} \right\|_1$$

$$L_A = \min_G \max_D E_X \log(1 + \exp(-D(G(\tilde{X}))))$$

$$L_F = \min_G E_X \left(\sum_{i=0}^{T} \left\| D^i(X) - D^i(G(\tilde{X})) \right\|_2\right)$$

- Degradation Model

$$I^d = ((I \otimes \mathbf{k}) \downarrow_s + \mathbf{n}_\sigma)_{JPEG_q}$$

$$s = \{10:200\}, \sigma = \{0:25\}, q = \{5:50\}$$

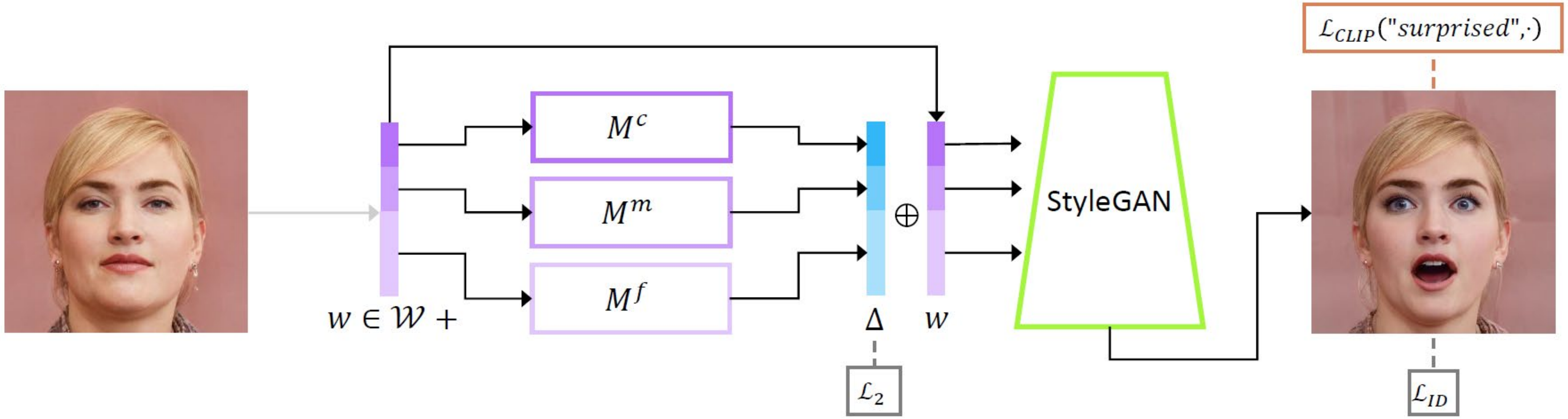
GPEN

StyelCLIP

$$\underset{w\in W^{+}}{\arg\min}\, D_{CLIP}(G(w),t)+\lambda_{L2}\left\|w-w_s\right\|_2+\lambda_{ID}\mathcal{L}_{ID}(w)$$

$$\mathcal{L}_{ID}(w)=1-\left\langle R(G(w_s)),R(G(w))\right\rangle$$

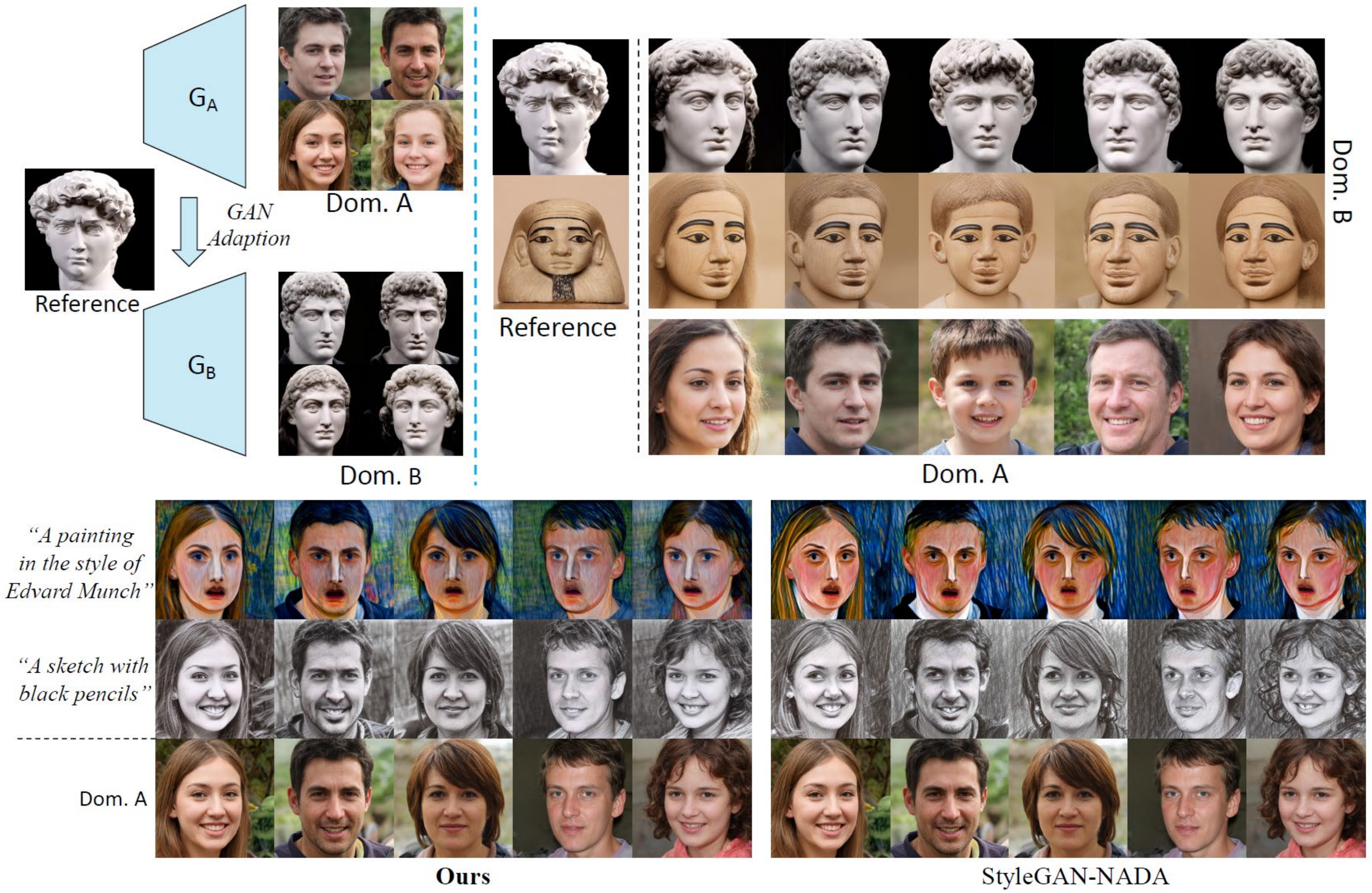
- Optimization



Input | "Beyonce" (0.004, 0) | "A woman without makeup" (0.008, 0.005) | "Elsa from Frozen" (0.004, 0)

Input | "A man with a beard" (0.008, 0.005) | "A blonde man" (0.008, 0.005) | "Donald Trump" (0.0025, 0)

DFGDA (Ours)



G_A
GAN
Adaption
G_B
Reference
Dom. A
Dom. B
Reference
Dom. B
Dom. A
"A painting
in the style of
Edvard Munch"
"A sketch with
black pencils"
Dom. A
Ours
StyleGAN-NADA

总结

- 理论研究
  - 如果更好更稳定地训练GAN
  - 网络结构、损失函数与正则化

- 应用研究
  - 预训练网络
  - 理解和应用：编辑/生成、算法->网络学习、更好地理解设计GAN Inversion网络