Focal Frequency Loss for Image Reconstruction and Synthesis

Liming Jiang[1] Bo Dai[1] Wayne Wu[2] Chen Change Loy[1]

[1]S-Lab, Nanyang Technological University [2]SenseTime Research

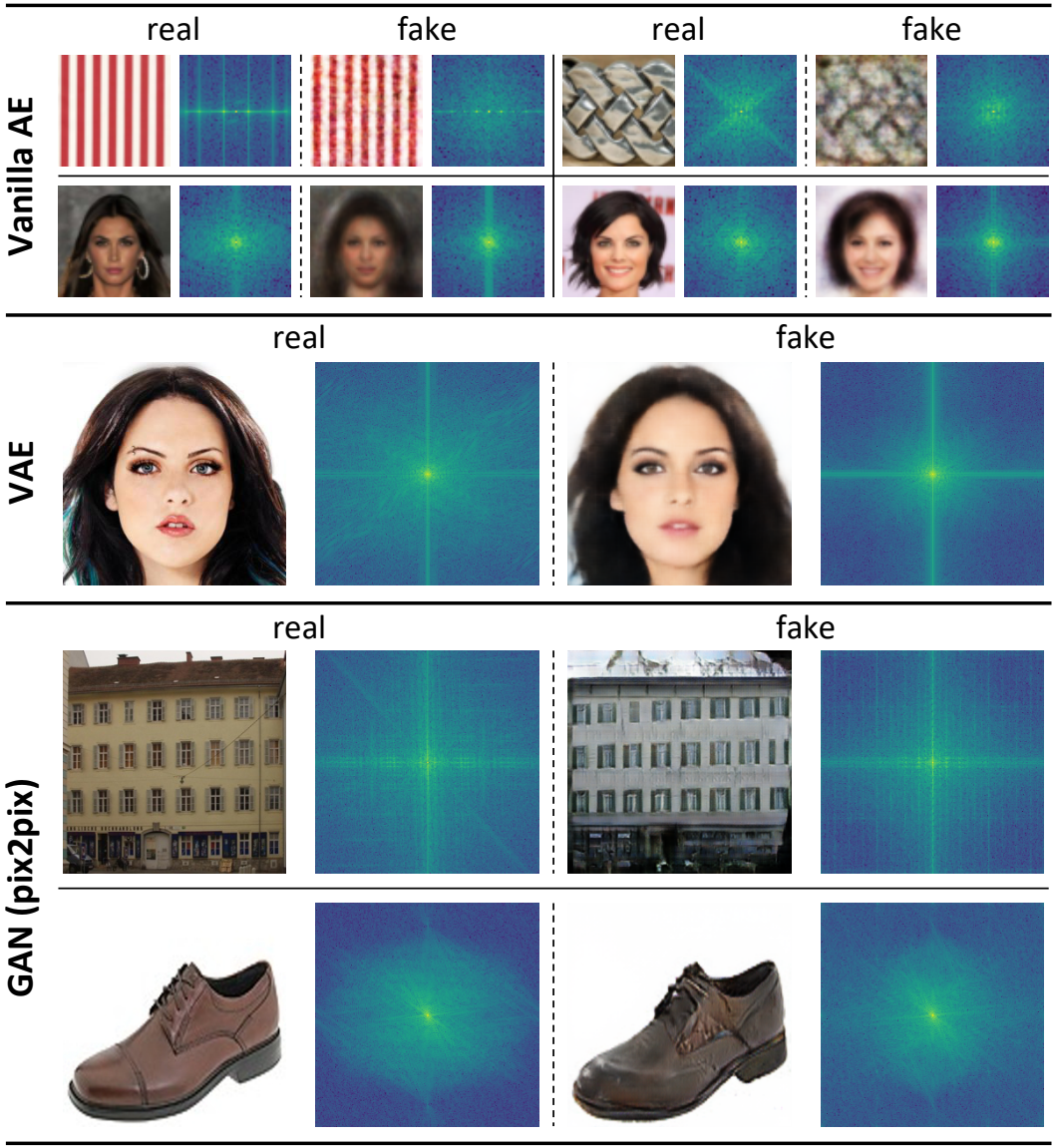
{liming002, bo.dai, ccloy}@ntu.edu.sg wuwenyan@sensetime.com

2021 ICCV OCTOBER 11-17 VIRTUAL

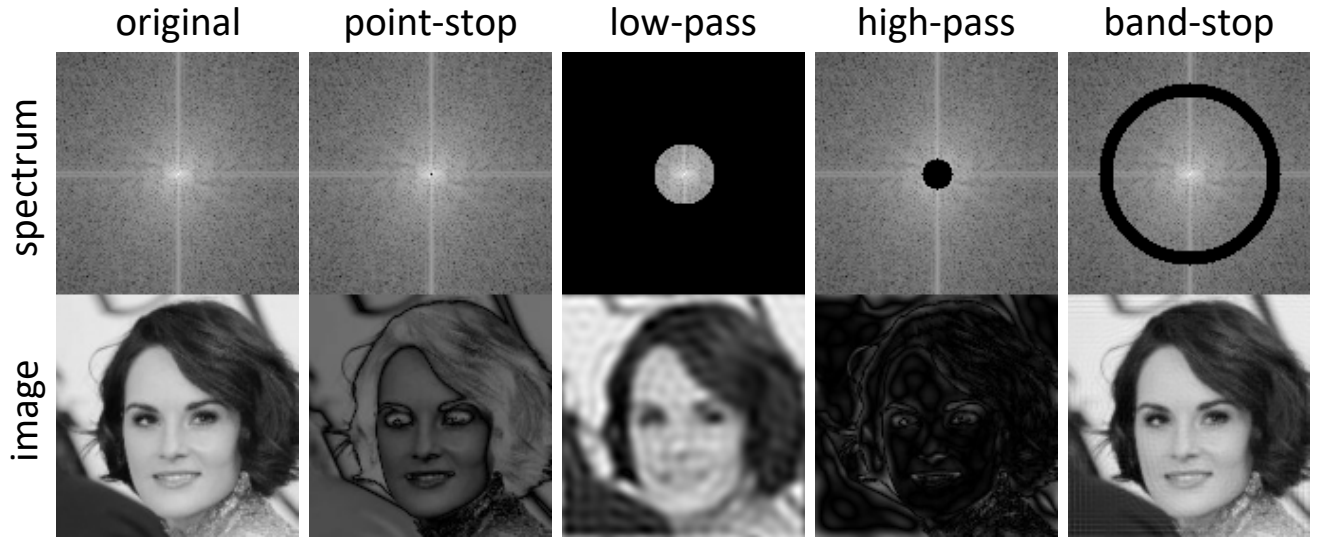
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

商汤 sensetime

Motivation



- Frequency Domain Gaps

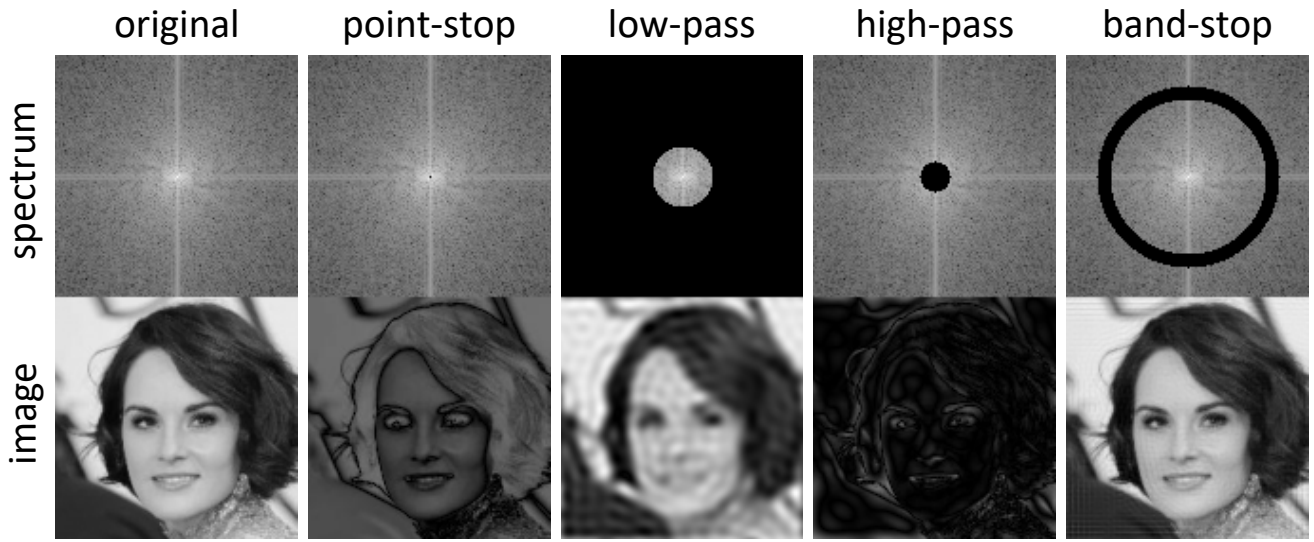


- Standard Bandlimiting: “Missing Frequencies”

1. Despite remarkable performance, gaps between the real and fake still exist.
2. Some gaps are visible, while others may only be revealed through the frequency spectrum analysis.
3. Inherent bias of neural networks: “spectral bias”, “F-Principle”, *etc.*

Methodology: *Step 1*

- Frequency Representation of Images



- Discrete Fourier transform (DFT):

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cdot e^{-i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)},$$

$$e^{-i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} = \cos 2\pi \left(\frac{ux}{M}+\frac{vy}{N}\right) - i \sin 2\pi \left(\frac{ux}{M}+\frac{vy}{N}\right).$$

Methodology: *Step 2*

- Frequency Distance

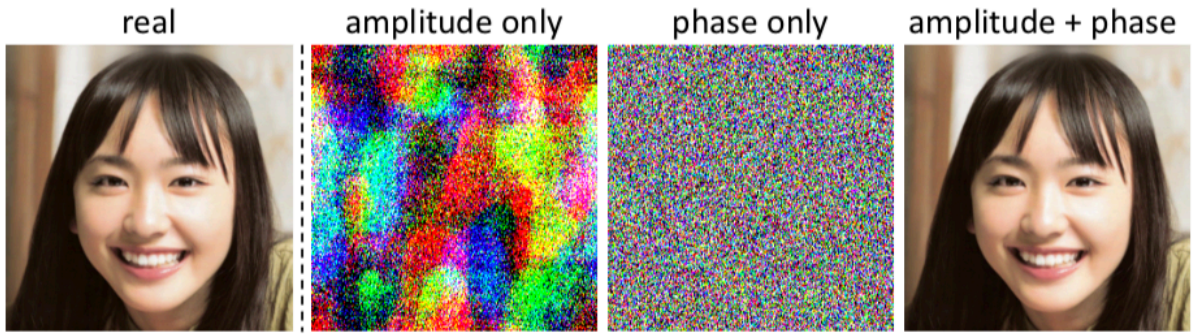
$$F(u,v) = R(u,v) + I(u,v)\,i = a + bi$$

- Amplitude:

$$|F(u,v)| = \sqrt{R(u,v)^2 + I(u,v)^2} = \sqrt{a^2 + b^2}$$

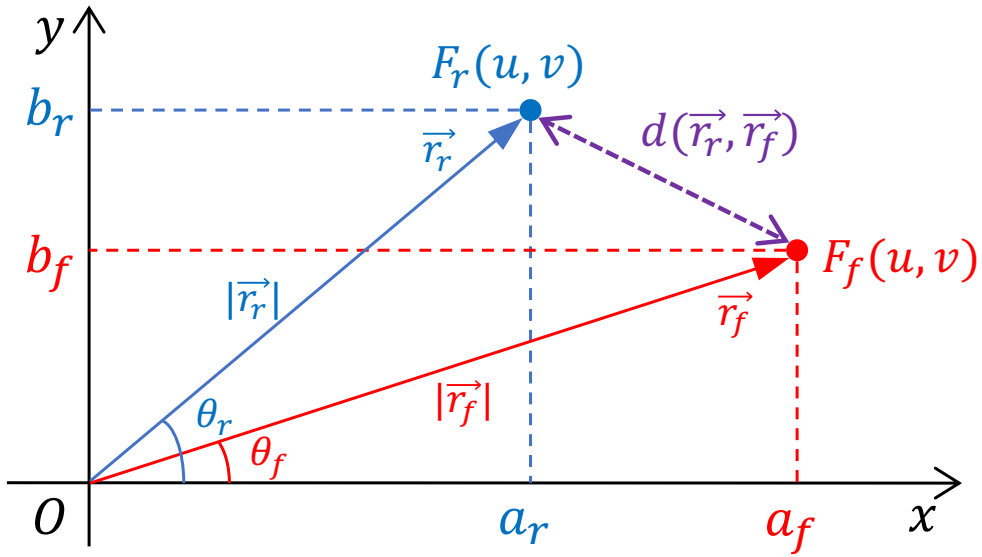
- Phase:

$$\angle F(u,v) = \arctan\left(\frac{I(u,v)}{R(u,v)}\right) = \arctan\frac{b}{a}$$



Single-image reconstruction

- Definition:



For a single frequency,

$$d(\vec{r_r},\vec{r_f}) = \|\vec{r_r} - \vec{r_f}\|_2^2 = |F_r(u,v) - F_f(u,v)|^2$$

For the real and fake images,

$$d(F_r,F_f) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}|F_r(u,v) - F_f(u,v)|^2$$

Methodology: *Step 3*

- Dynamic Spectrum Weighting

  - Spectrum weight matrix ($\alpha = 1$ by default):

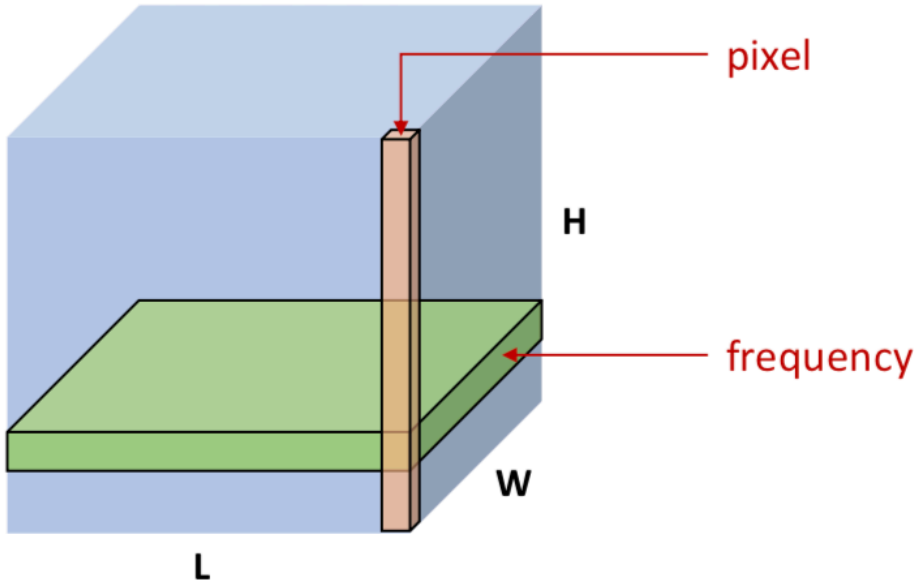
$$w(u,v) = |F_r(u,v) - F_f(u,v)|^{\alpha}$$

  - The *full* form of the focal frequency loss (FFL):

$$\mathrm{FFL} = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} w(u,v) |F_r(u,v) - F_f(u,v)|^2.$$

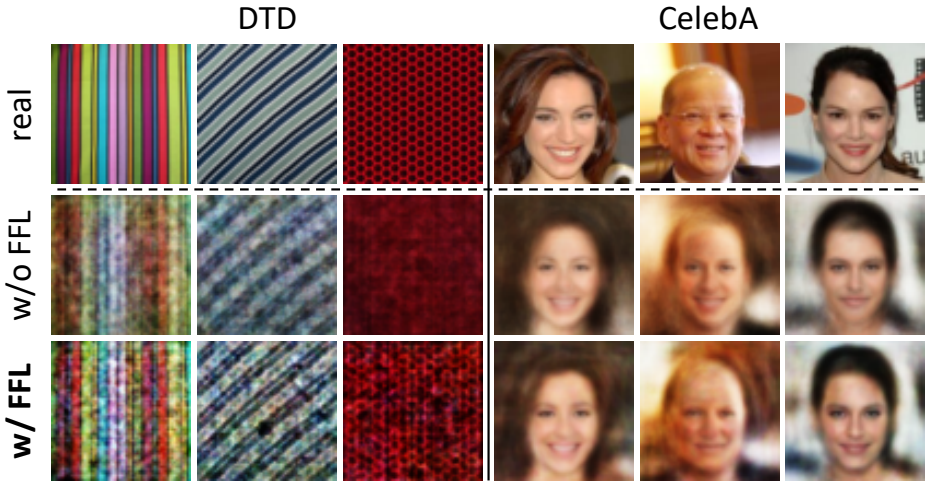
    * Other variants of FFL for the flexibility: adjusting α, patch-based FFL, …

  - More intuitive illustration:

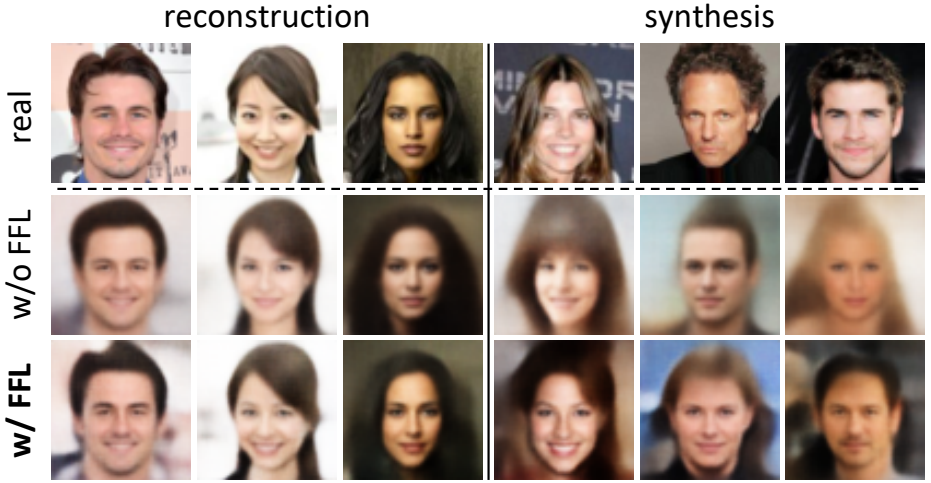


Results and Analysis

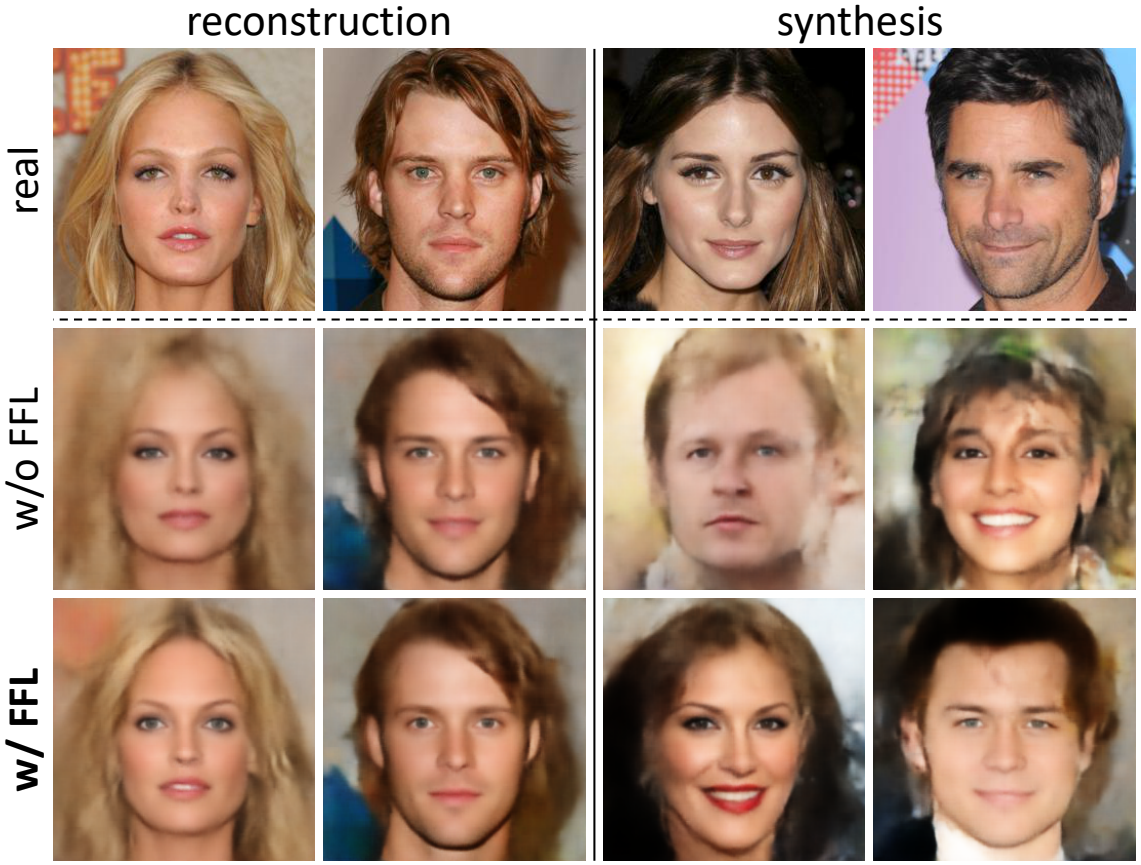
- Image Reconstruction and Unconditional Synthesis (Autoencoders)



Vanilla AE reconstruction on DTD and CelebA (64 × 64)



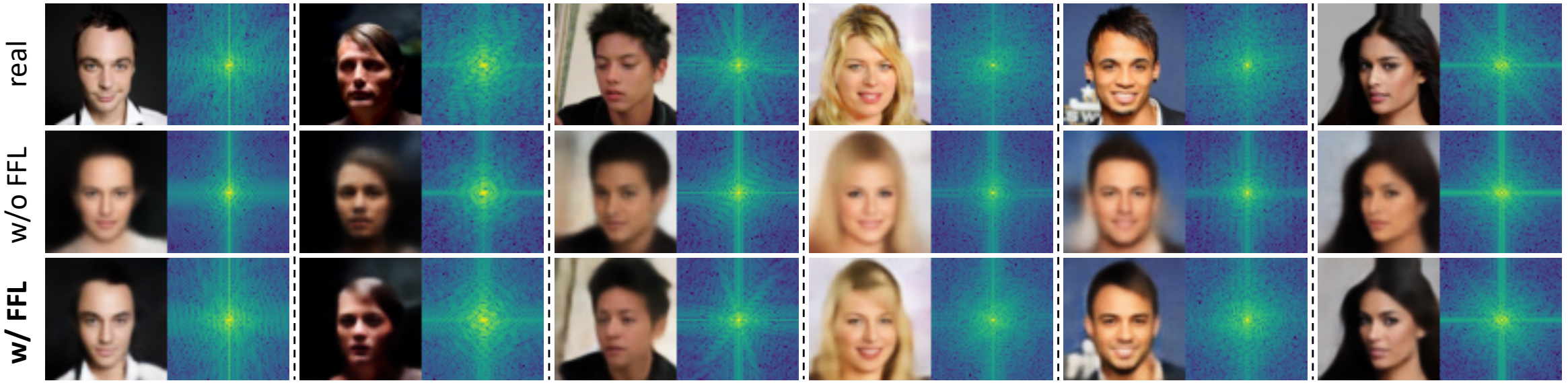
VAE reconstruction and **synthesis** on CelebA (64 × 64)



VAE reconstruction and **synthesis** on CelebA-HQ (256 × 256)

Results and Analysis

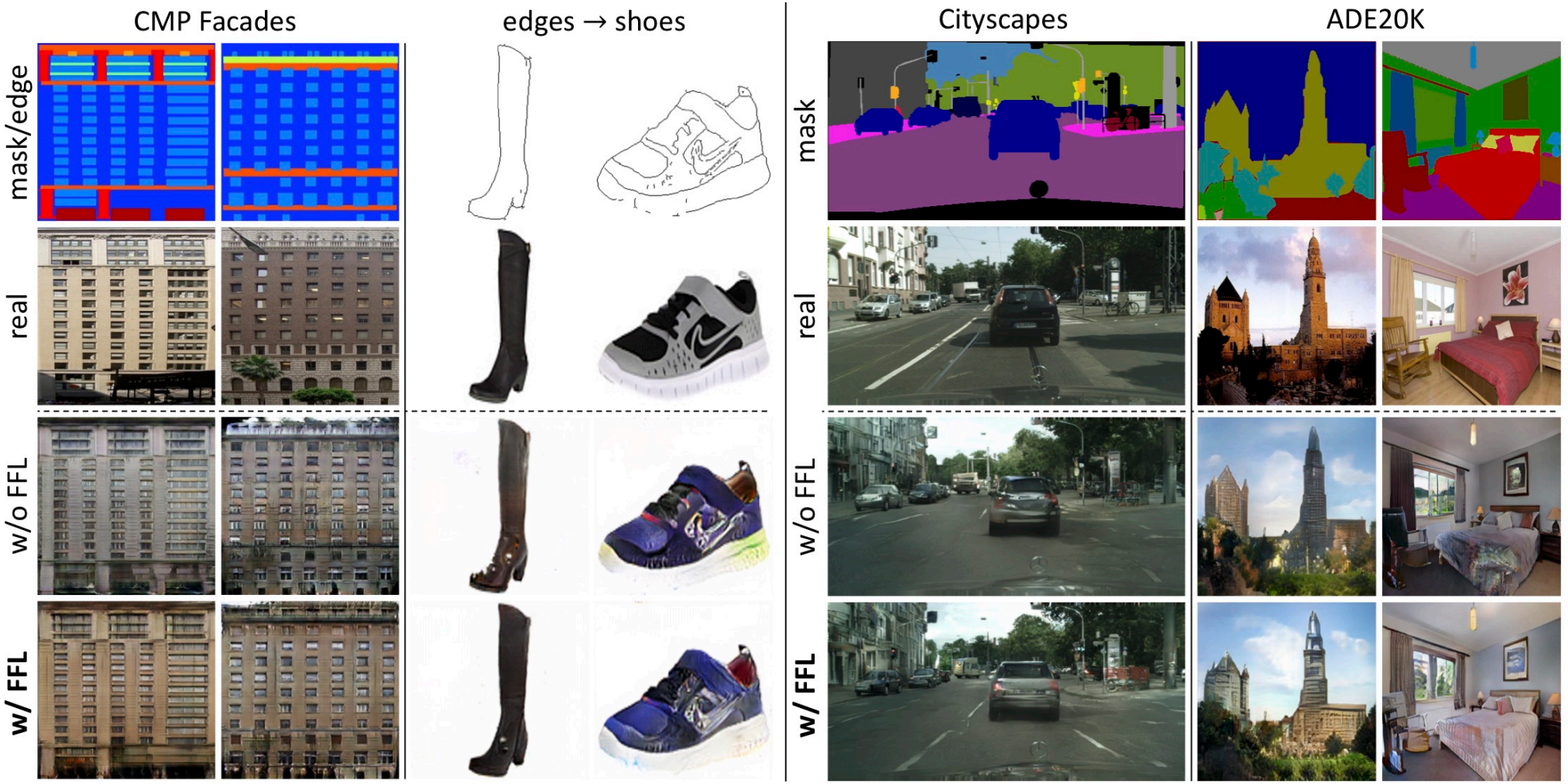
- Analysis on Frequency Domain Gaps (VAE)



As an example, frequency domain gaps are narrowed by FFL for VAE image reconstruction on CelebA

Results and Analysis

- Conditional Image Synthesis (pix2pix | SPADE)



GAN-based image-to-image translation on various datasets (256 pix in short edge)

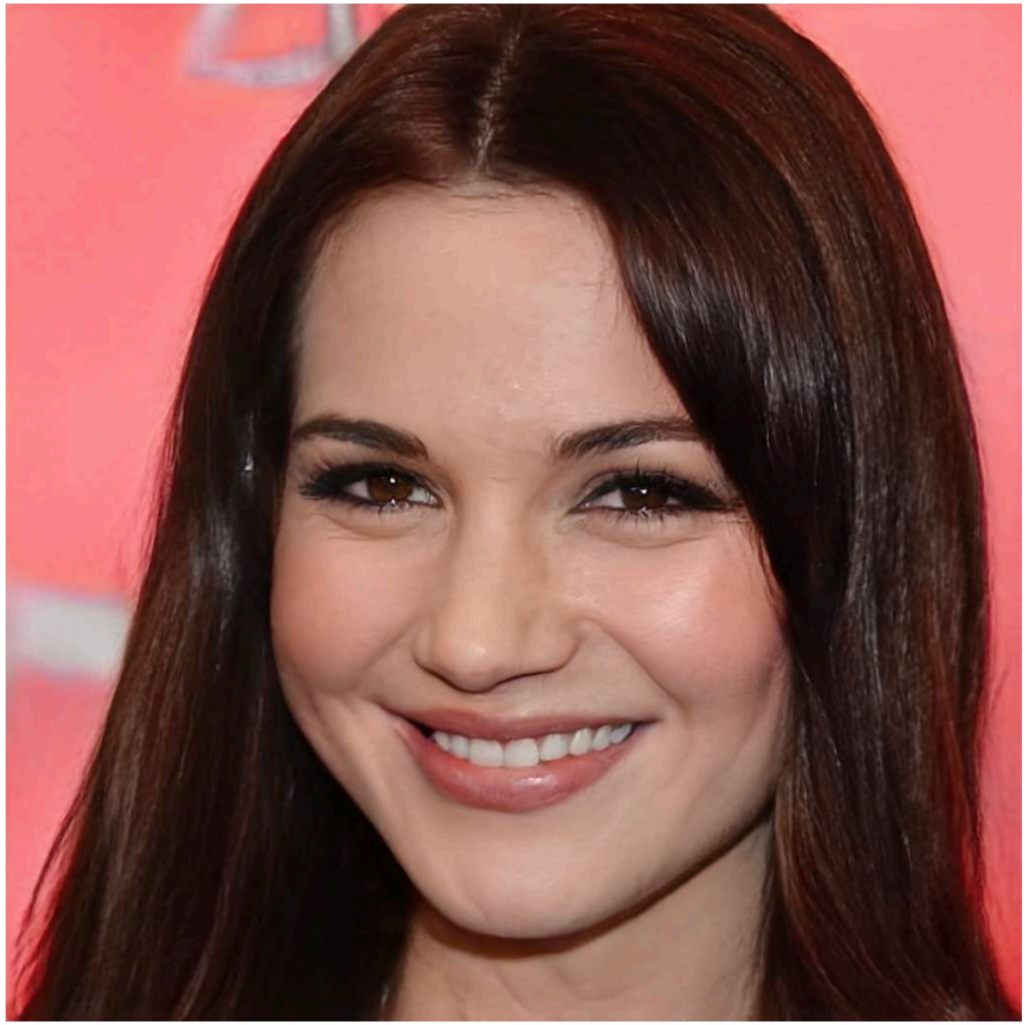
Results and Analysis

- Potential on the State of the Art (StyleGAN2)

StyleGAN2 unconditional image synthesis on CelebA-HQ (256 × 256)

Results and Analysis

- High-Resolution Examples (StyleGAN2)



Synthesized images by **StyleGAN2 trained with FFL** on CelebA-HQ (1024 × 1024)

Results and Analysis

- Quantitative Evaluations

Vanilla AE image reconstruction

Dataset	FFL	PSNR↑	SSIM↑	LPIPS↓	FID↓	LFD↓
DTD	w/o	20.133	**0.407**	0.414	246.870	14.764
	w/	**20.151**	0.400	**0.404**	**240.373**	**14.760**
CelebA	w/o	20.044	0.568	0.237	97.035	14.785
	w/	**21.703**	**0.642**	**0.199**	**83.801**	**14.403**

VAE image reconstruction

Dataset	FFL	PSNR↑	SSIM↑	LPIPS↓	FID↓	LFD↓
CelebA	w/o	19.961	0.606	0.217	69.900	14.804
	w/	**22.954**	**0.723**	**0.143**	**49.689**	**14.115**
CelebA-HQ	w/o	21.310	0.616	0.367	71.081	17.266
	w/	**22.253**	**0.637**	**0.344**	**59.470**	**17.049**

VAE unconditional image synthesis

Dataset	FFL	FID↓	IS↑
CelebA	w/o	80.116	1.873
	w/	**71.050**	**2.010**
CelebA-HQ	w/o	93.778	2.057
	w/	**84.472**	**2.060**

pix2pix image-to-image translation

Dataset	FFL	FID↓	IS↑
CMP Facades	w/o	128.492	1.571
	w/	**123.773**	**1.738**
edges → shoes	w/o	80.279	2.674
	w/	**74.359**	**2.804**

SPADE semantic image synthesis

Method	Cityscapes			ADE20K		
	mIoU↑	accu↑	FID↓	mIoU↑	accu↑	FID↓
CRN [5]	52.4	77.1	104.7	22.4	68.8	73.3
SIMS [49]	47.2	75.5	**49.7**	N/A	N/A	N/A
pix2pixHD [66]	58.3	81.4	95.0	20.3	69.2	81.8
SPADE [47]	62.3	81.9	71.8	38.5	79.9	33.9
SPADE + FFL	**64.2**	**82.5**	59.5	**42.9**	**82.4**	**33.7**

StyleGAN2 unconditional image synthesis

Dataset	FFL	FID↓	IS↑
CelebA-HQ (256 × 256)	w/o	5.696	3.383
	w/	**4.972**	**3.432**

2021 ICCV OCTOBER 11-17 VIRTUAL

Focal Frequency Loss for Image Reconstruction and Synthesis

Thanks!

GitHub (Code & Model)

https://github.com/EndlessSora/focal-frequency-loss

Project Page

https://www.mmlab-ntu.com/project/ffl/index.html

P.S. `pip install focal-frequency-loss` is all you need!