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

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

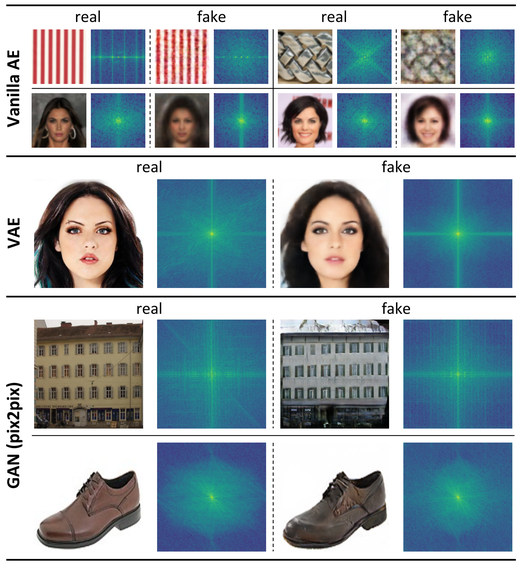
商汤 sensetime

2021 ICCV OCTOBER 11-17 VIRTUAL

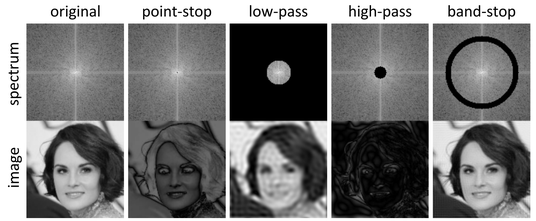
GitHub (Code & Model)

Project Page

Motivation



Frequency Domain Gaps: Common Issue



Standard Bandlimiting: "Missing Frequencies"

- Despite remarkable performance, gaps could still exist between the real and generated images.
- Some gaps are visible, while others may only be revealed through the frequency spectrum analysis.
- Inherent bias of neural networks: "spectral bias", "F-Principle", *etc.*

Methodology

- **Step 1: Frequency Representation of Images** - facilitates optimization in the frequency dimension.
  - Discrete Fourier transform (DFT):

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

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

- **Step 2: Frequency Distance** - quantifies the differences between images in the frequency domain.

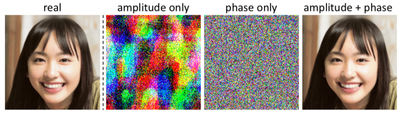
$$F(u,v) = R(u,v) + I(u,v)\,i = a + bi$$

- Amplitude:

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

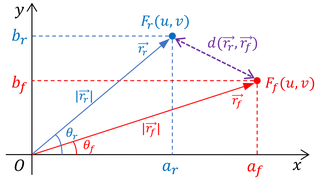
- Phase:

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



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,

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

- **Step 3: Dynamic Spectrum Weighting** - dynamically focuses a model on "hard frequencies" during training.
  - Spectrum weight matrix ($\alpha = 1$ by default):

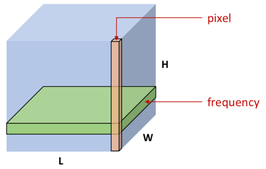
$$w(u,v) = |F_r(u,v) - F_f(u,v)|^\alpha$$

  - The *full* form of Focal Frequency Loss (FFL):

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

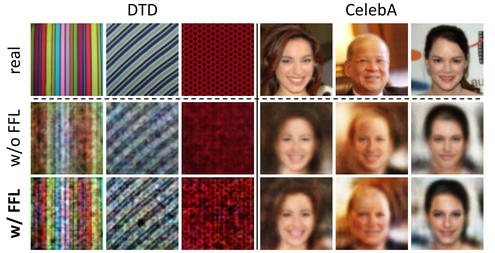
  * Other variants of FFL for flexibility: adjusting α, patch-based FFL, ...

- More intuitive illustration:

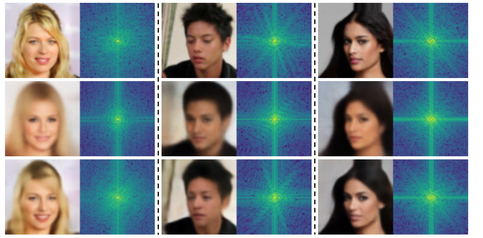


Results

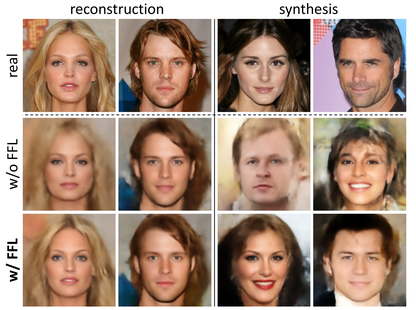
- **Image Reconstruction (Vanilla AE)**



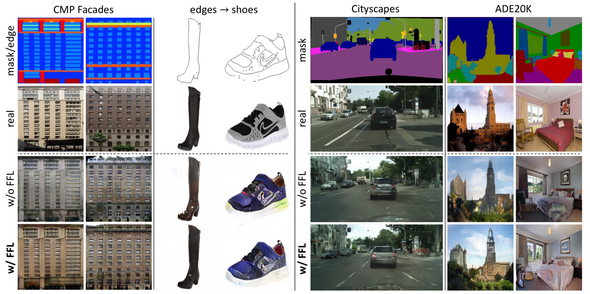
- **Analysis on Frequency Domain Gaps (VAE)**



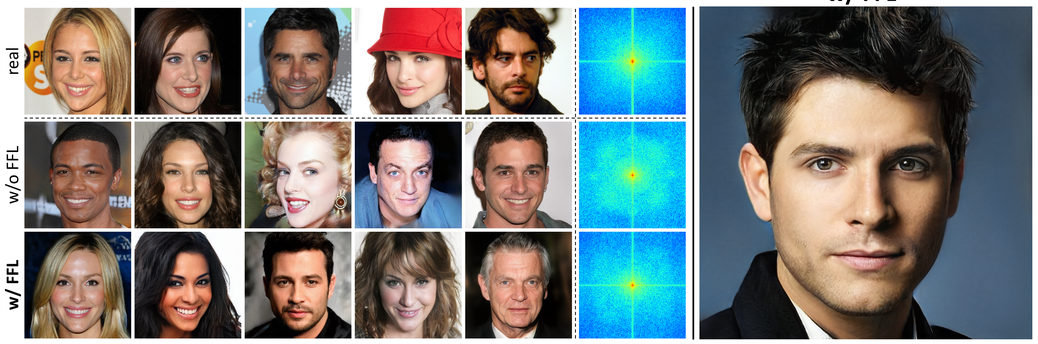
- **Reconstruction and Synthesis (VAE)**



- **Image-to-Image Translation (pix2pix | SPADE)**



- **Unconditional Image Synthesis (StyleGAN2, State of the Art)**



Code: https://github.com/EndlessSora/focal-frequency-loss `pip install focal-frequency-loss` is all you need!