FreeU--Free Lunch in Diffusion U-Net

FreeU: Free Lunch in Diffusion U-Net

Chenyang Si Ziqi Huang Yuming Jiang Ziwei Liu✉
S-Lab, Nanyang Technological University
{chenyang.si, ziqi002, yuming002, ziwei.liu}@ntu.edu.sg

Abstract

*In this paper, we uncover the untapped potential of diffusion U-Net, which serves as a "free lunch" that substantially improves the generation quality on the fly. We initially investigate the key contributions of the U-Net architecture to the denoising process and identify that its main backbone primarily contributes to denoising, whereas its skip connections mainly introduce high-frequency features into the decoder module, causing the network to overlook the backbone semantics. Capitalizing on this discovery, we propose a simple yet effective method—termed "**FreeU**" — that enhances generation quality without additional training or finetuning. Our key insight is to strategically re-weight the contributions sourced from the U-Net's skip connections and backbone feature maps, to leverage the strengths of both components of the U-Net architecture. Promising results on image and video generation tasks demonstrate that our FreeU can be readily integrated to existing diffusion models, e.g., Stable Diffusion, DreamBooth, ModelScope, Rerender and ReVersion, to improve the generation quality with only a few lines of code.* ***All you need is to adjust two scaling factors during inference.*** *Project page: https://chenyangsi.top/FreeU/.*

在本文中，我们发现了diffusion U-Net尚未开发的潜力，它是一种"free lunch"，在实质上提高了动态生成质量。我们初步研究了U-Net架构对去噪过程的关键贡献，并确定**其main backbone有助于去噪**，而其skip connect主要将high-frequency feature引入到去decoder module中，导致网络忽略了骨干语义。

利用这一发现，我们提出了一种简单而有效的方法，称为“FreeU”——在没有额外训练或微调的情况下提高生成质量。我们的key insight是战略性地re-weight来自U-Net的skip connect和backbone feature map的贡献，以利用U-Net架构的两个组件的优势。图像和视频生成任务的有希望的结果表明，我们的FreeU可以很容易地集成到现有的扩散模型中，例如Stable Diffusion、DreamBooth、ModelScopeRender和ReVersion，只需几行代码就可以提高生成质量。

Observations regarding frequency domain

Issue: finer details are markedly sensitive to noise.

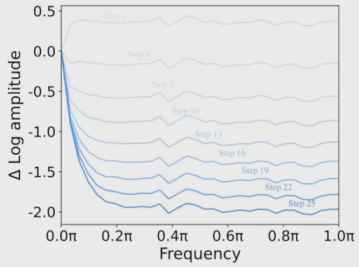


Figure 3. **Relative log amplitudes of Fourier for diffusion intermediate steps.** At each denoising step t, we visualize the relative log amplitudes of Fourier of recovered date x_t. We observe that the high-frequency components of x_t drops drastically during the denoising process.

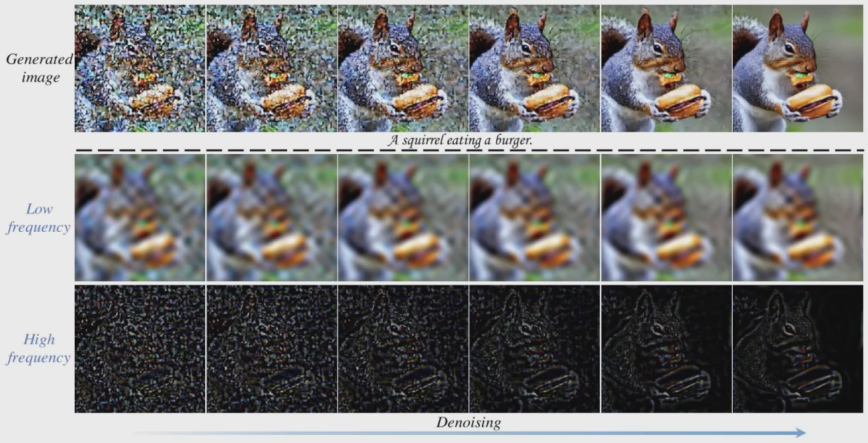


Figure 2. **The denoising process.** The top row illustrates the image's progressive denoising process across iterations, while the subsequent two rows display low-frequency and high-frequency components after the inverse Fourier Transform, matching each step. It's evident that low-frequency components change slowly, whereas high-frequency components exhibit more significant variations during the denoising process.

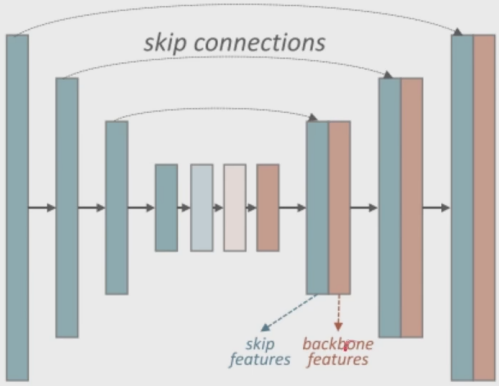
问题和假设：

观察到生成图像时，从random noise 到真正的图像，在频域中，变化的现象。在刚开始做diffusion的时候，高频的信息量是很高的，随时间step T的不断增加，高频的信息量逐渐减少，大部分的**能量**集中在了低频的区域。这是符合直觉的，因为图像和random noise相比是比较平滑的。随着del-noise的过程中，不断的把高频的信息去掉，这些高频的信息可能就被当作noise和random noise一起被去掉了。

How does diffusion U-Net work?

Backbone features => Low freq.

Skip features => High freq.



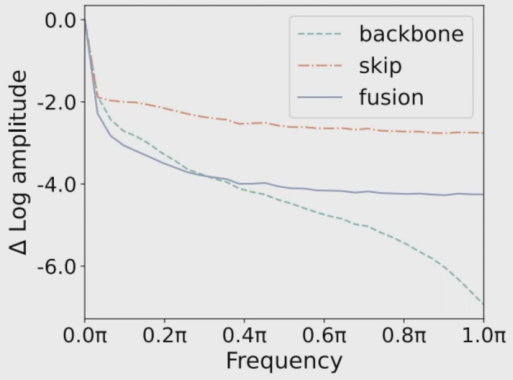
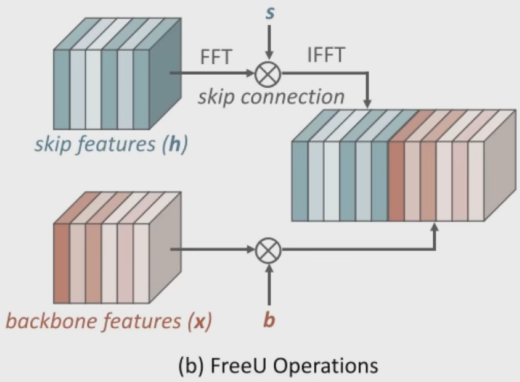


Figure 7. **Fourie relative log amplitudes of backbone, skip, and their fused feature maps.** The features, forwarded by skip connections directly from earlier layers of the encoder block to the decoder contain a large amount of high-frequency information.

FreeU

Adjust the ratio of two features in decoding stage



(b) FreeU Operations

Backbone features:

$$x'_{l,i} = \begin{cases} b_l \cdot x_{l,i}, & \text{if } i < C/2 \\ x_{l,i}, & \text{otherwise} \end{cases}$$

Skip features (suppress low-freq part):

$$\mathcal{F}(h_{l,i}) = \text{FFT}(h_{l,i})$$
$$\mathcal{F}'(h_{l,i}) = \mathcal{F}(h_{l,i}) \odot \alpha_{l,i} \quad \alpha_{l,i}(r) = \begin{cases} s_l & \text{if } r < r_{\text{thresh}}, \\ 1 & \text{otherwise.} \end{cases}$$
$$h'_{l,i} = \text{IFFT}(\mathcal{F}'(h_{l,i}))$$

对于backbone feature，C在这里表示channel，对前一半的channel采取直接乘以一个b_l在上面，这个l是decoder的层数。对于skip feature，抑制它的低频部分，具体做法是做FFT到频域里面，然后乘上了一个滤波器（这个滤波器是一个圆形的mask，不知道为什么这么设计，也没有说清楚半径r选择多少），滤波之后再逆变换回来。

可以通过调整这两个参数去得到一个相对满意的效果，实验说明了调高backbone的权重会让图片生成的更加平滑好看，而增大skip connect的权重对于图像的生成没有太明显的改善。