Deep Learning and Practice 2023 Summer
Final project

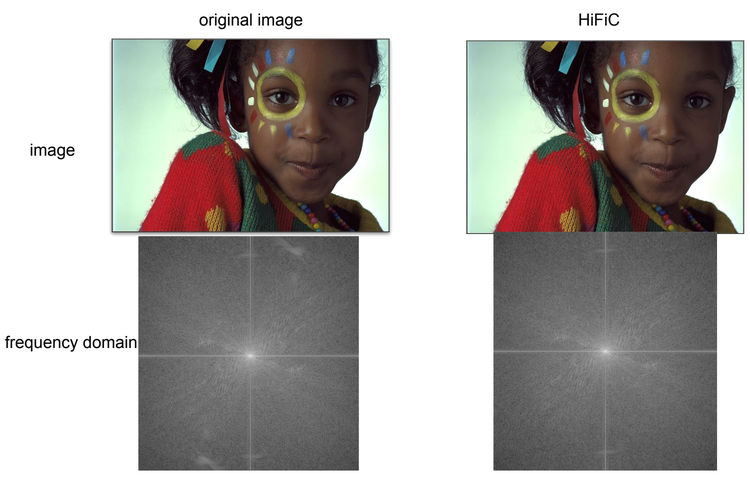
FA-HiFiC: Frequency Augmented High-Fidelity Generative Image Compression

何承原 - 許瑋哲(Wesley) - N.Q.Sang

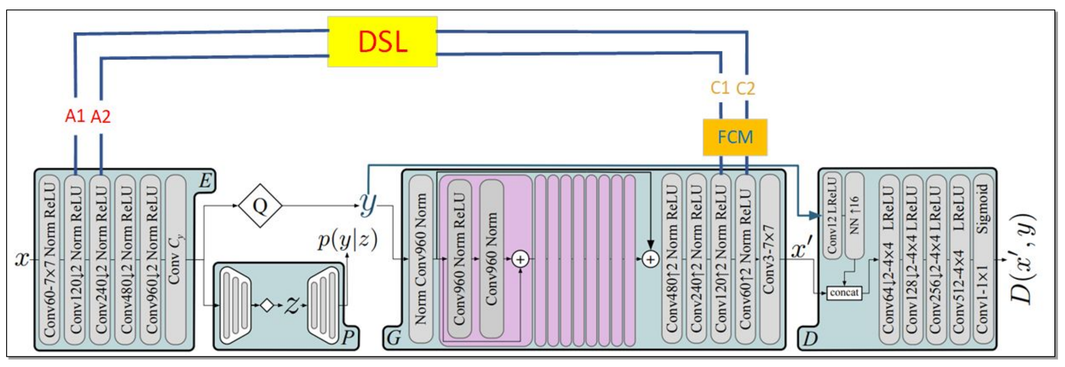
Abstract

The High-Fidelity Generative Image Compression Network combines Generative Adversarial Networks with learned compression to achieve a state-of-the-art generative lossy compression system as of 2020. In their paper, the authors investigated normalization layers, generator and discriminator architectures, training strategies, and perceptual losses. It is also recognized for delivering strong visual performance.

However, upon examining the frequency domain closely, it becomes evident that HiFiC does not learn high-frequency domain information effectively. Therefore, our objective is to minimize the disparity between the frequency domain of the original and compressed images.



Objective and Proposal



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

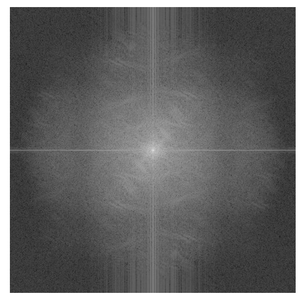
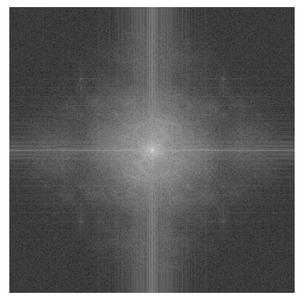
$$\mathrm{SL}(\mathcal{A}_i, \mathcal{C}_i) = \mathrm{FFL}(\hat{\mathcal{A}}_i, \hat{\mathcal{C}}_i)$$

$$(\hat{\mathcal{A}}_i, \hat{\mathcal{C}}_i) = (K_i(\mu, \sigma_i) \star \mathcal{A}_i, K_i(\mu, \sigma_i) \star \mathcal{C}_i)$$

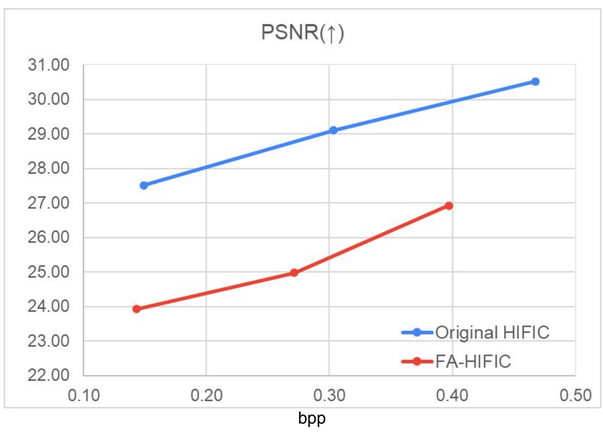
The paper, Catch Missing Details: Image Reconstruction with Frequency Augmented Variational Autoencoder(CVPR 2023), enhance the compression performance in high frequency by adding the focal frequency loss in pixel domain and adding the dynamic spectrum loss in feature domain. These two loss function are used to evaluate the gap of frequency between real and generated image.

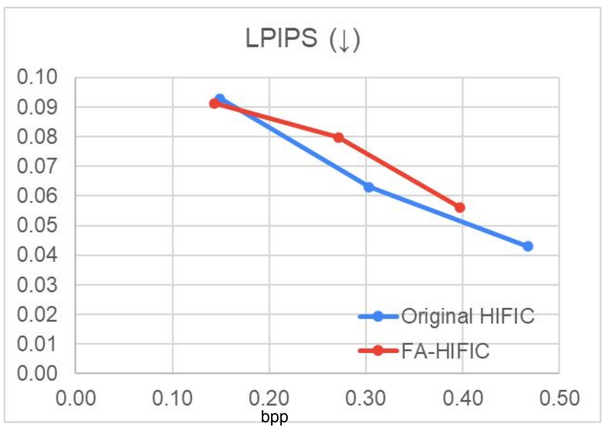
Inspired by the paper, we tried to apply focal frequency loss and dynamic spectrum loss in HiFiC model to reduce the gap of frequency and try to compress the image with more details. And FCM is needed to evaluate dynamic spectrum loss.

Reconstructed Image

	Original	FA-HIFIC
Image		
Frequency Domain		

Experiment Result





Conclusion

Our result doesn't perform as well as the original HiFiC in terms of PSNR. However, when considering the alternative evaluation metric, LPIPS, which aims to capture human perception of image similarity, it performs quite well and even manages to outperform HiFiC at low bit rates.

In the frequency domain, it's evident that our model can capture a significantly greater amount of high-frequency information, aligning well with the purpose of utilizing the frequency loss. Nevertheless, we struggle to strike a balance between the frequency loss and the mean squared error loss. This imbalance is likely the primary reason we are unable to achieve performance on par with HiFiC in terms of PSNR.