Leveraging Frequency Analysis for Deep Fake Image Recognition

Joel Frank, Thorsten Eisenhofer, Lea Schönherr,
Asja Fischer, Dorothea Kolossa, Thorsten Holz

RUHR
UNIVERSITÄT
BOCHUM

RUB

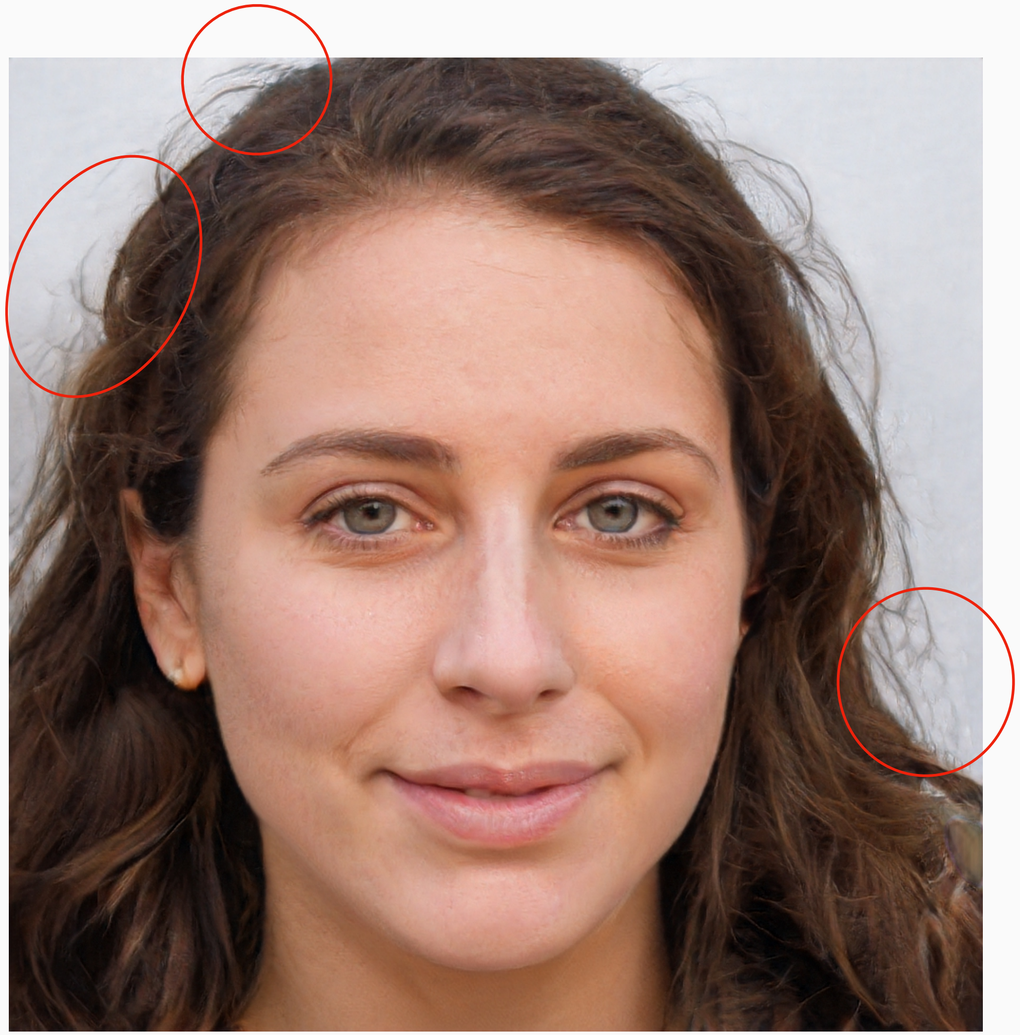
CASA

Which Face is Real?

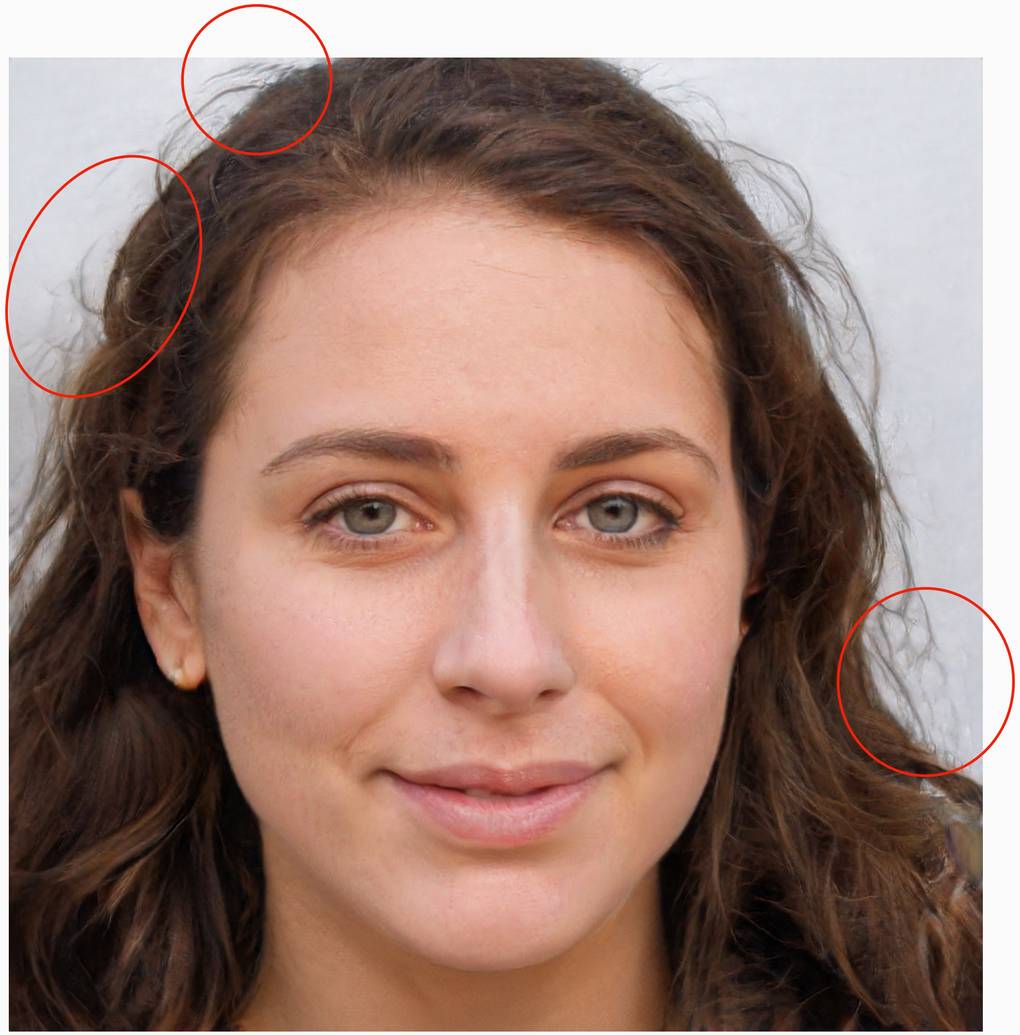
PLAY ABOUT METHODS LEARN PRESS CONTACT BOOK CALLING BS

Click on the person who is real.

Which Face is Real?

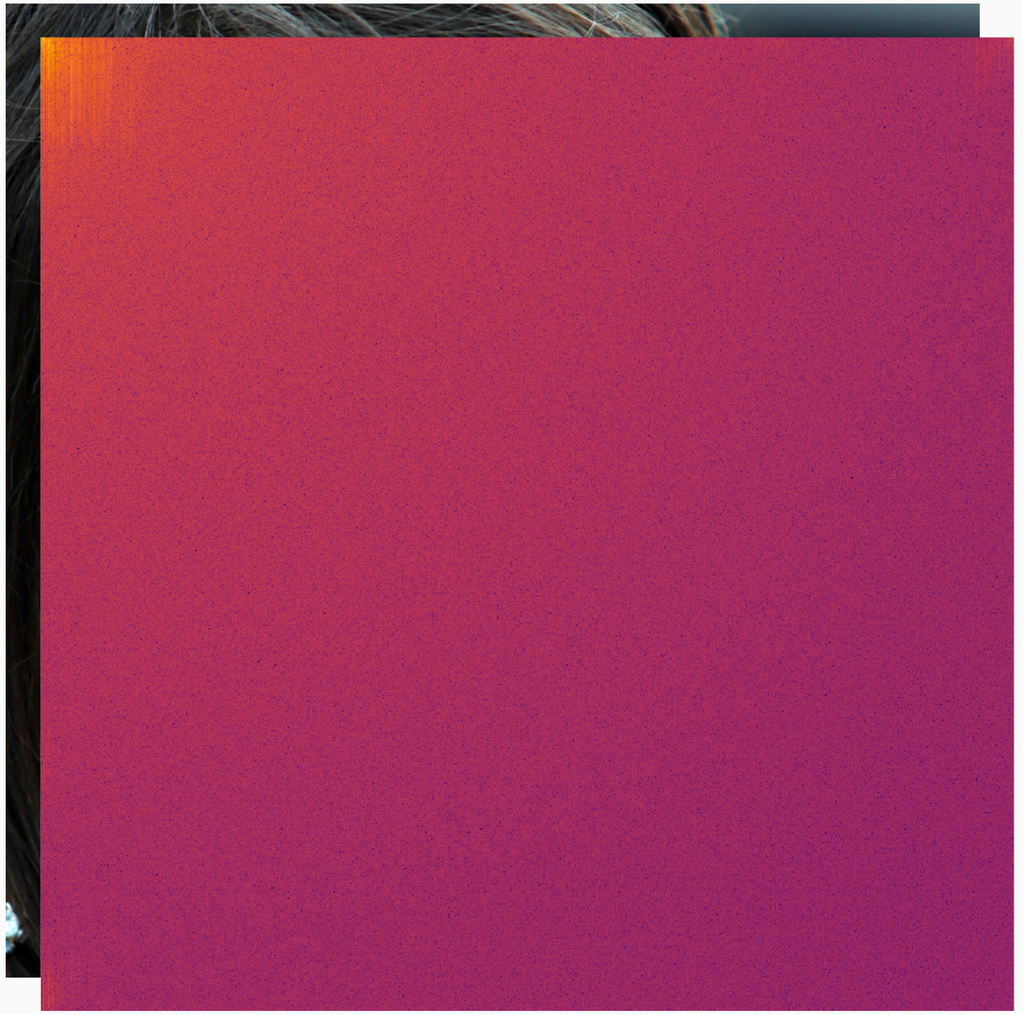


Which Face is Real?



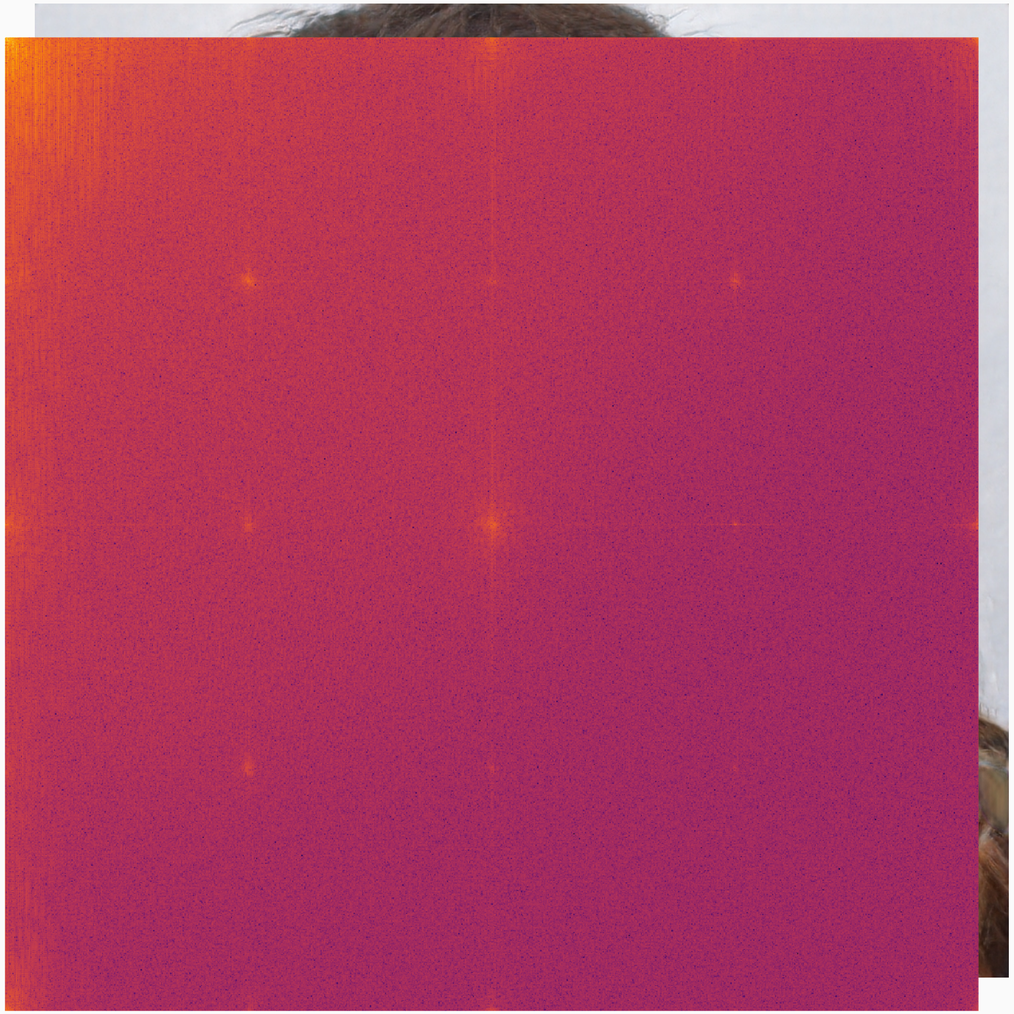
$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

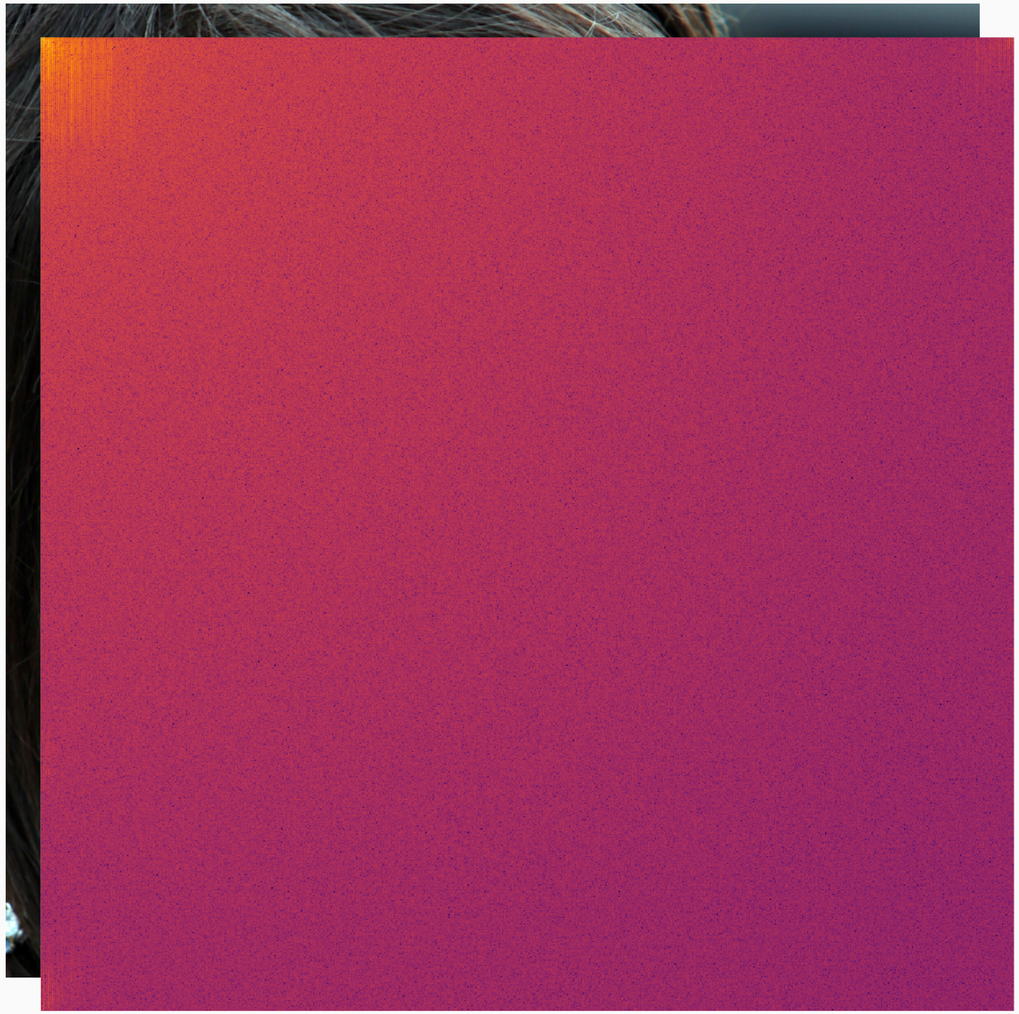
Which Face is Real?



$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

Which Face is Real?

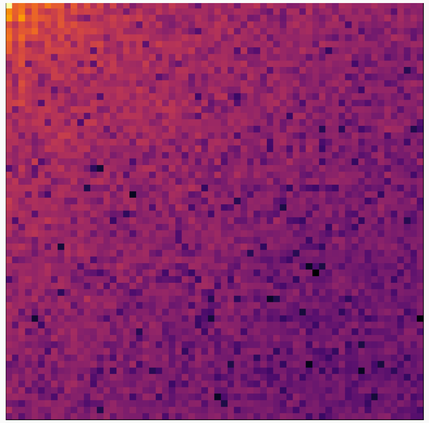




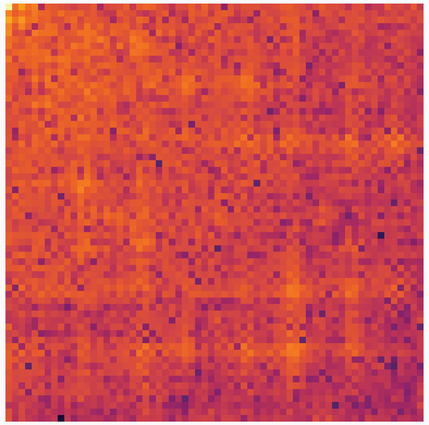
$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

Specific to StyleGAN?

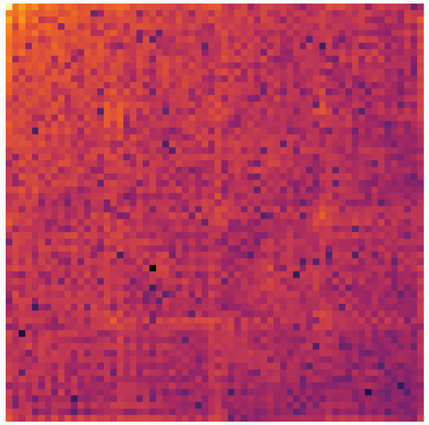
Specific to StyleGAN?



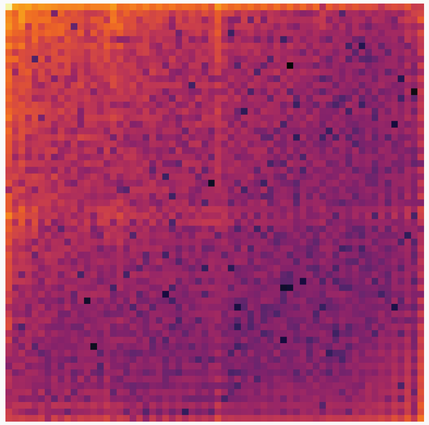
Stanford Dogs



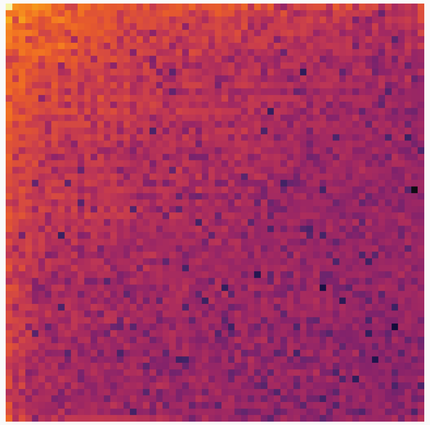
BigGAN



ProGAN

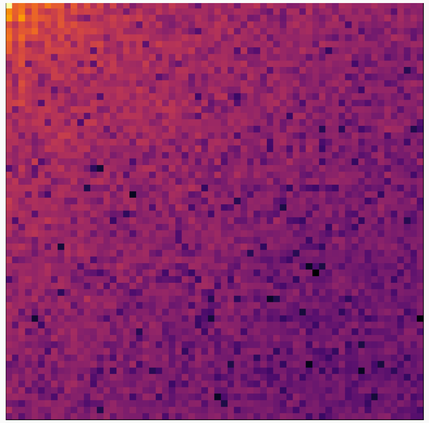


SN-DCGAN

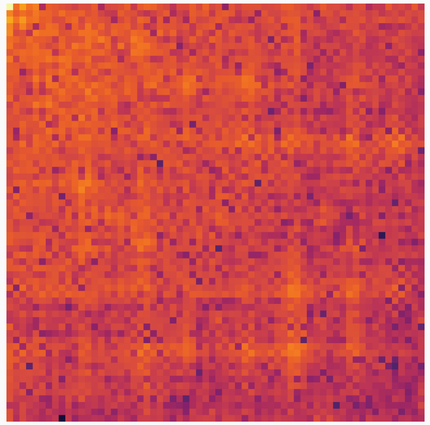


StyleGAN

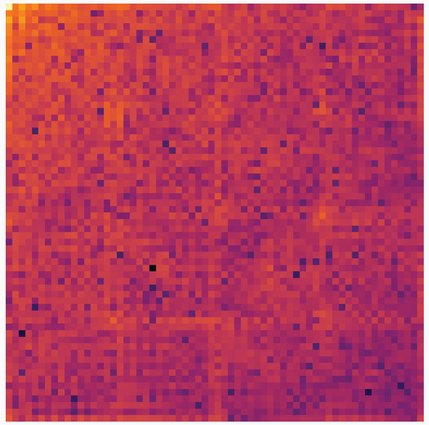
Specific to StyleGAN?



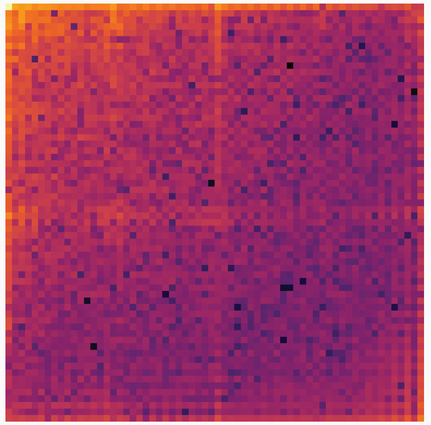
Stanford Dogs



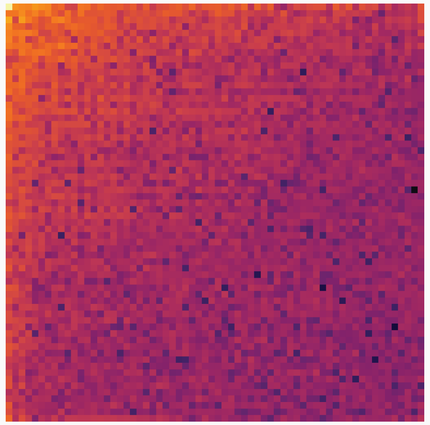
BigGAN



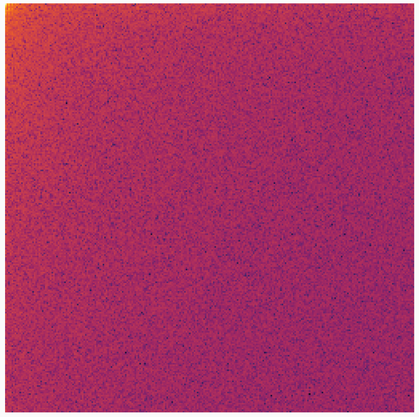
ProGAN



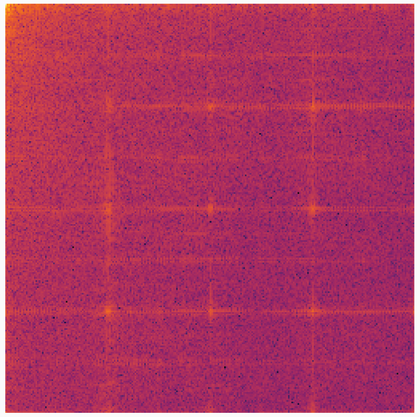
SN-DCGAN



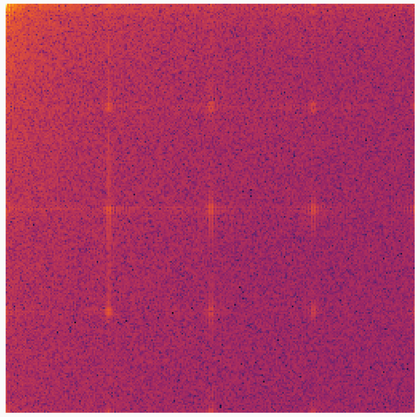
StyleGAN



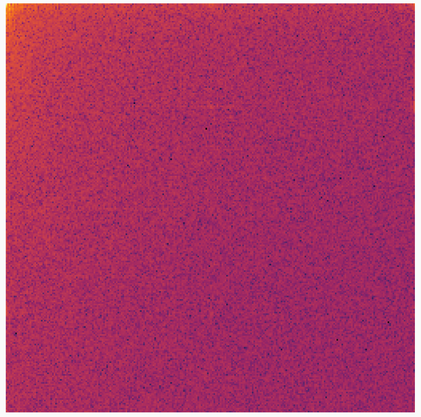
LSUN Bedrooms



Nearest Neighbour



Bilinear



Binomial

Advantages of the Frequency Domain

Advantages of the Frequency Domain

Domain	Accuracy
Image	75.78%
Frequency	100.00%

Advantages of the Frequency Domain

Domain	Accuracy
Image	75.78%
Frequency	100.00%

- Experiments on corrupted data

Advantages of the Frequency Domain

Domain	Accuracy
Image	75.78%
Frequency	100.00%

- Experiments on corrupted data
  - Blurring, cropping, jpeg compression, noise, combination

Advantages of the Frequency Domain

Domain	Accuracy
Image	75.78%
Frequency	100.00%

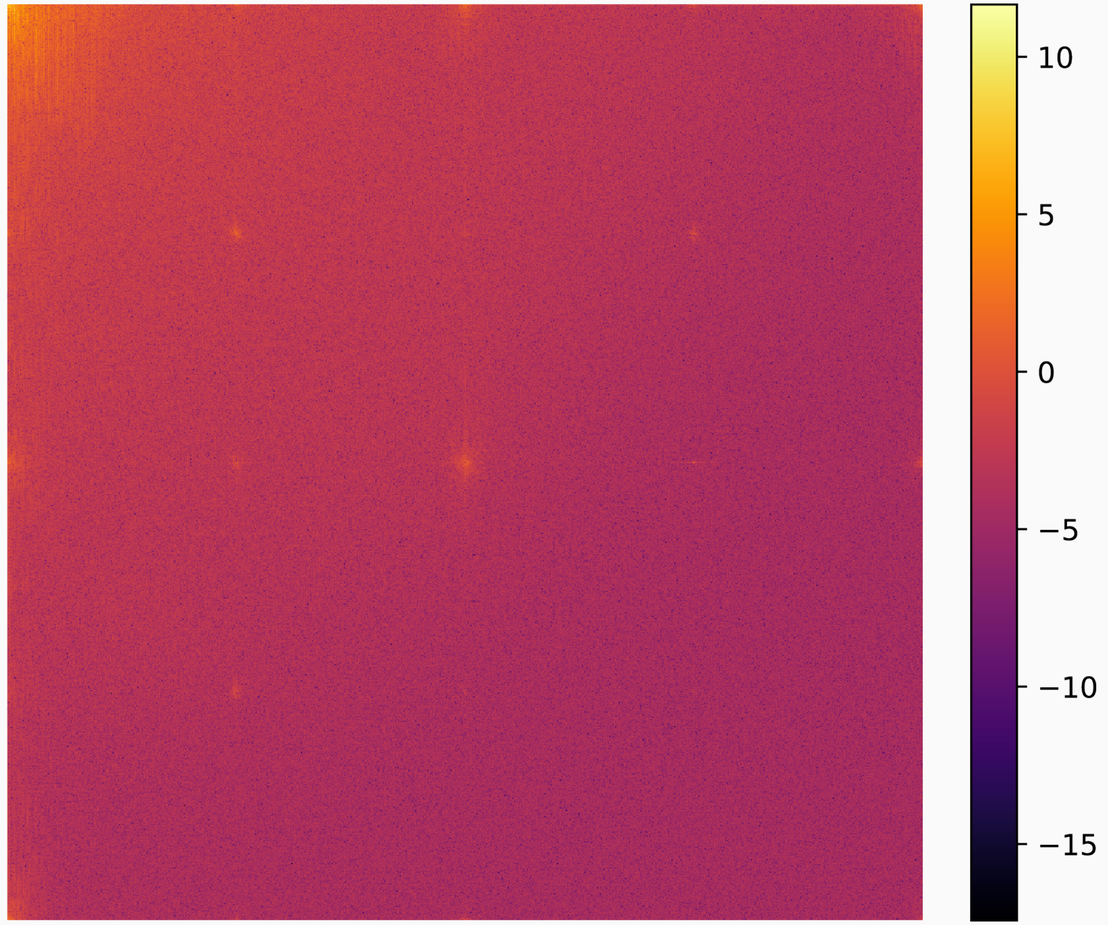
- Experiments on corrupted data
  - Blurring, cropping, jpeg compression, noise, combination
  - Frequency representation performs better (bar one exception)

Advantages of the Frequency Domain

Domain	Accuracy
Image	75.78%
Frequency	100.00%

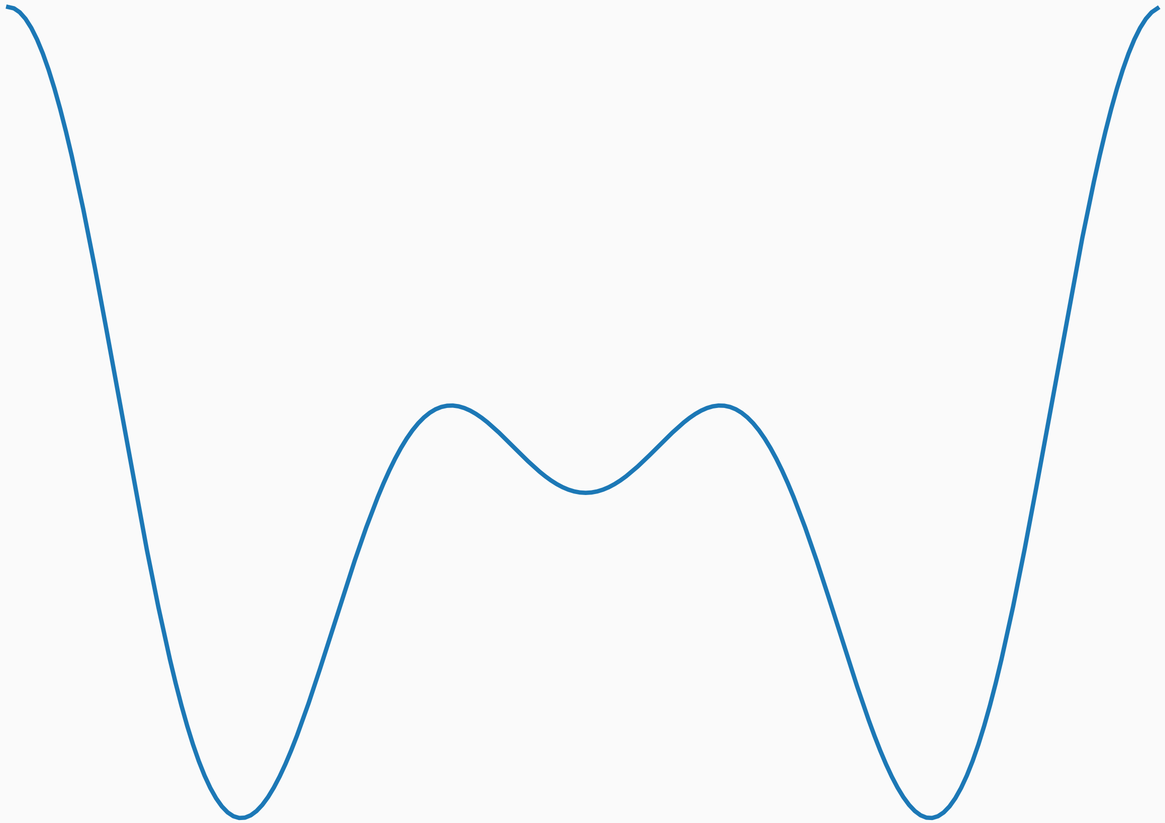
- Experiments on corrupted data
  - Blurring, cropping, jpeg compression, noise, combination
  - Frequency representation performs better (bar one exception)
  - When trained on corrupted data, frequency representation recovers higher accuracy

Frequency Domain

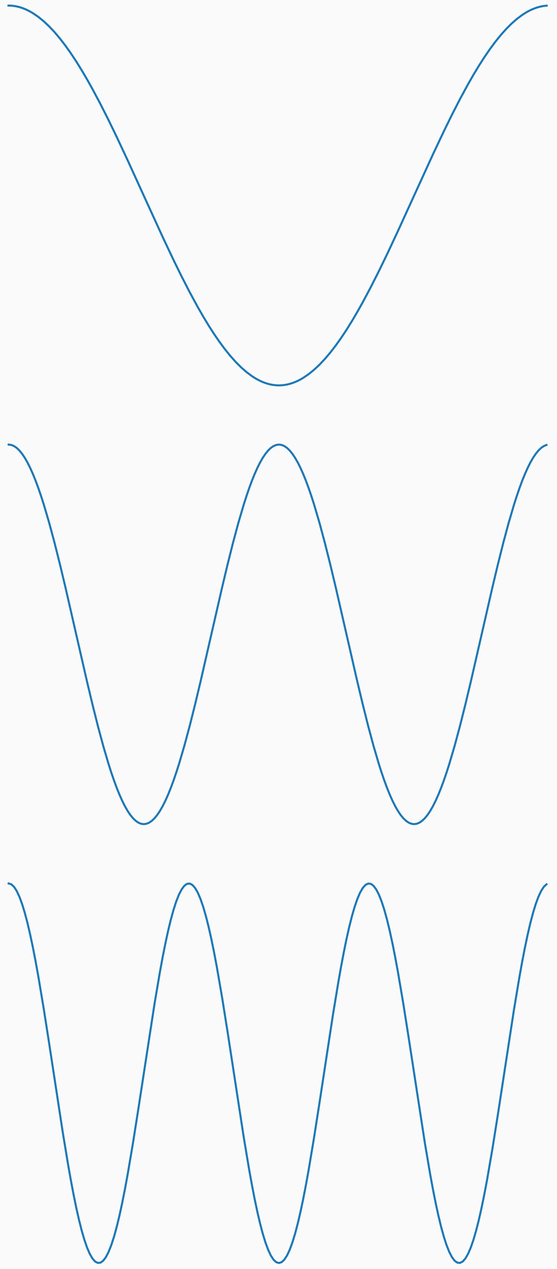


$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

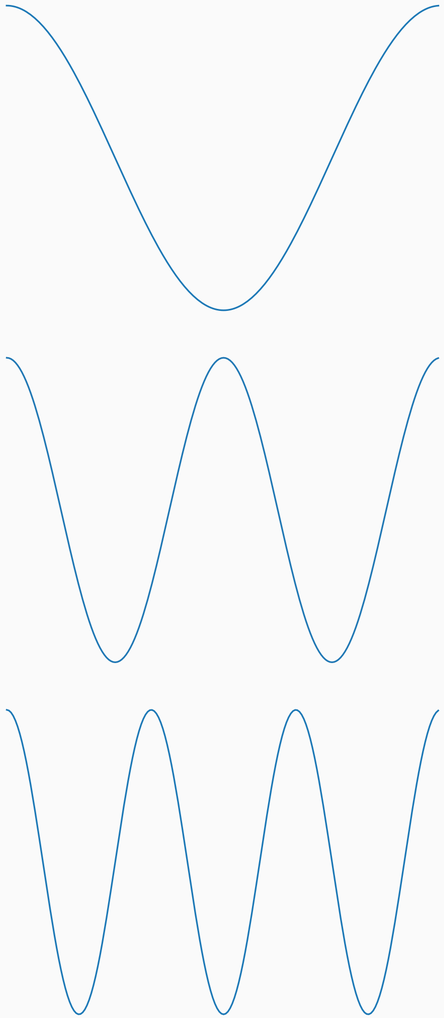
Frequency Domain

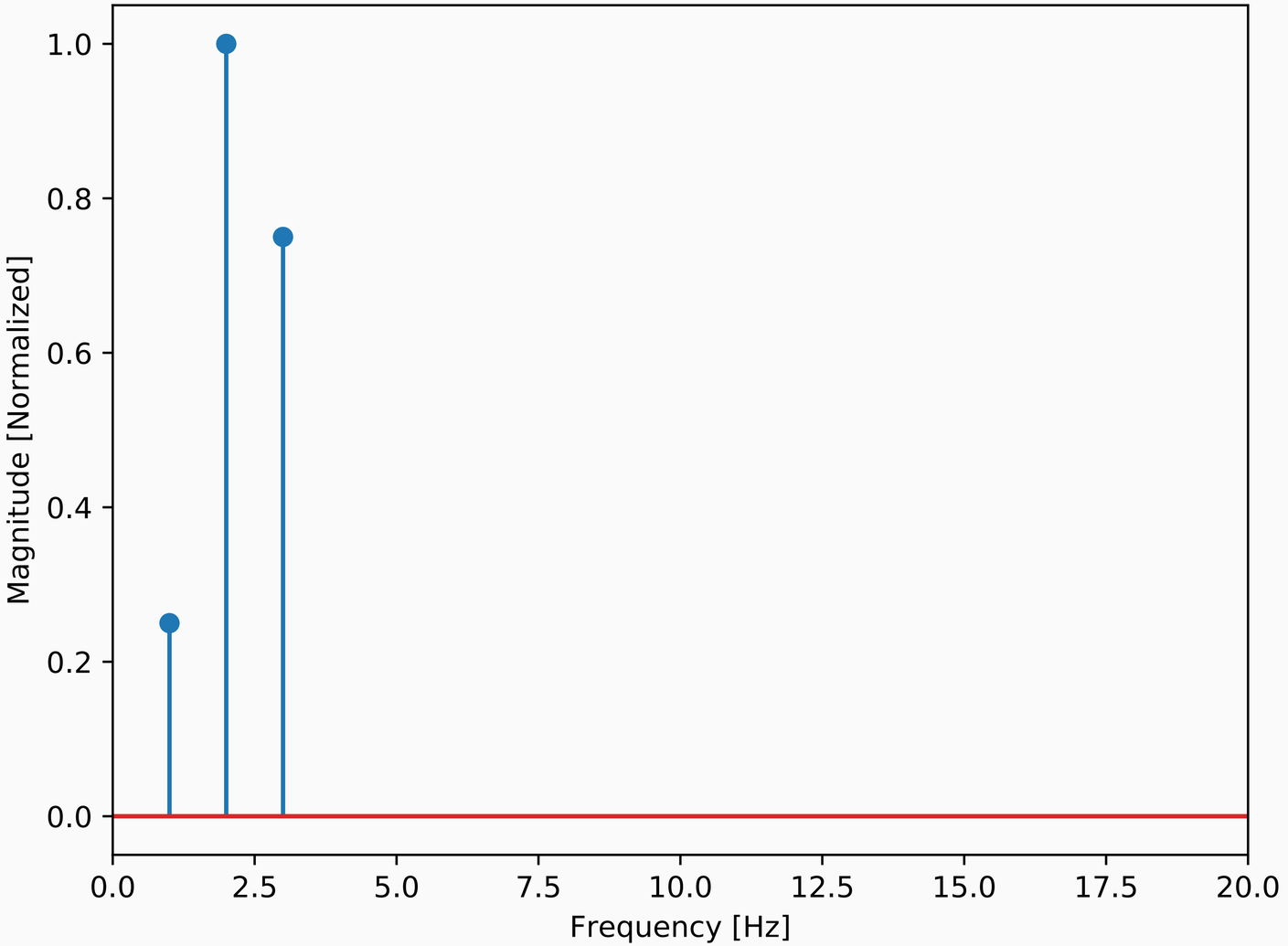


Discrete Cosine Transform



Frequency Domain

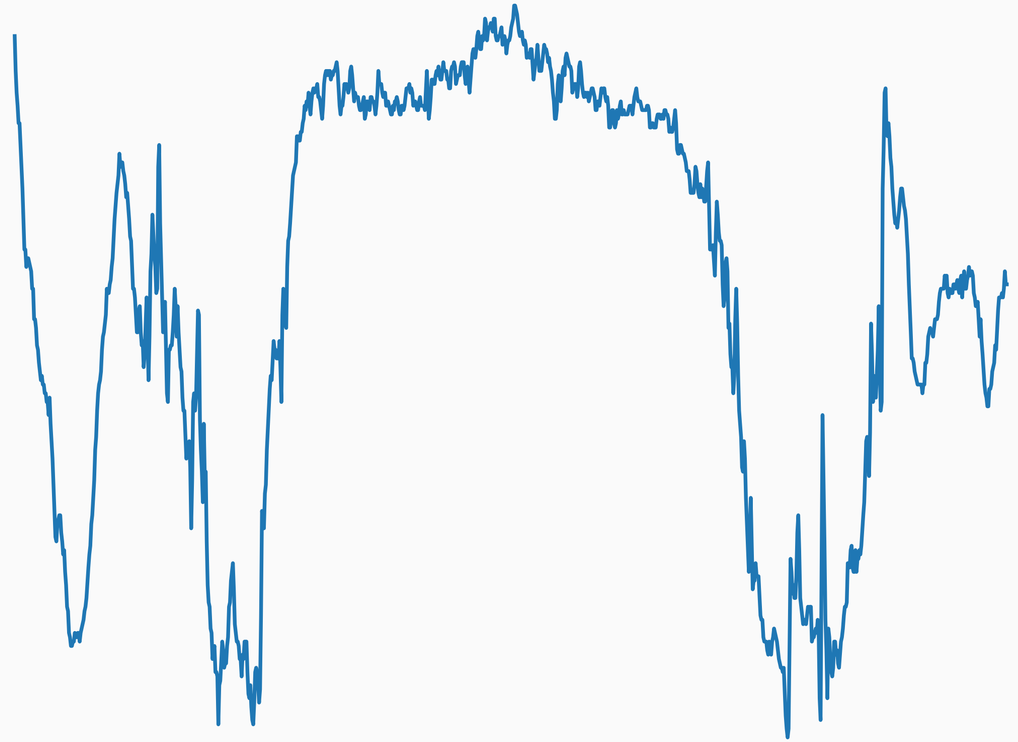




Frequency Domain

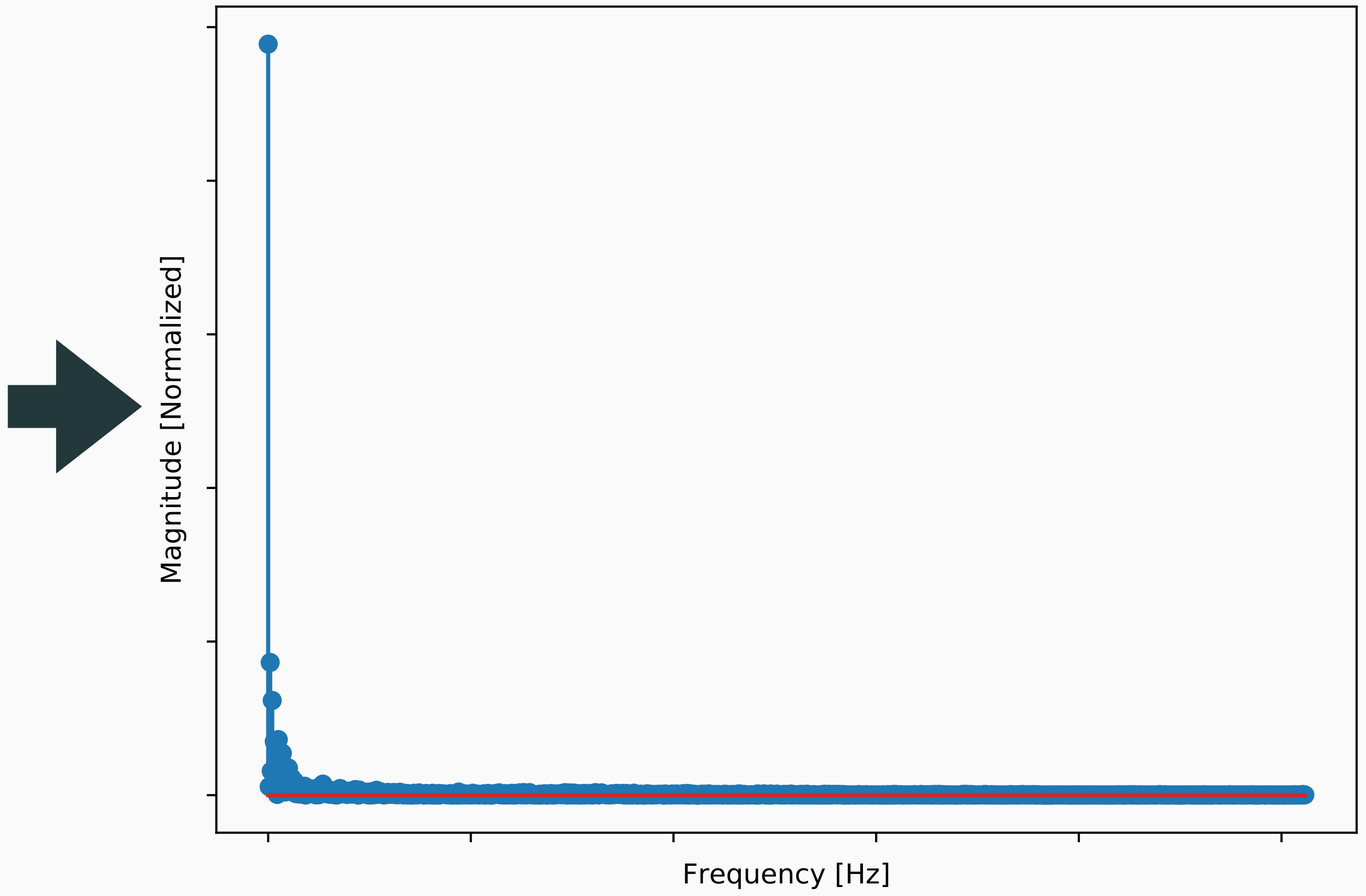
$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

Frequency Domain



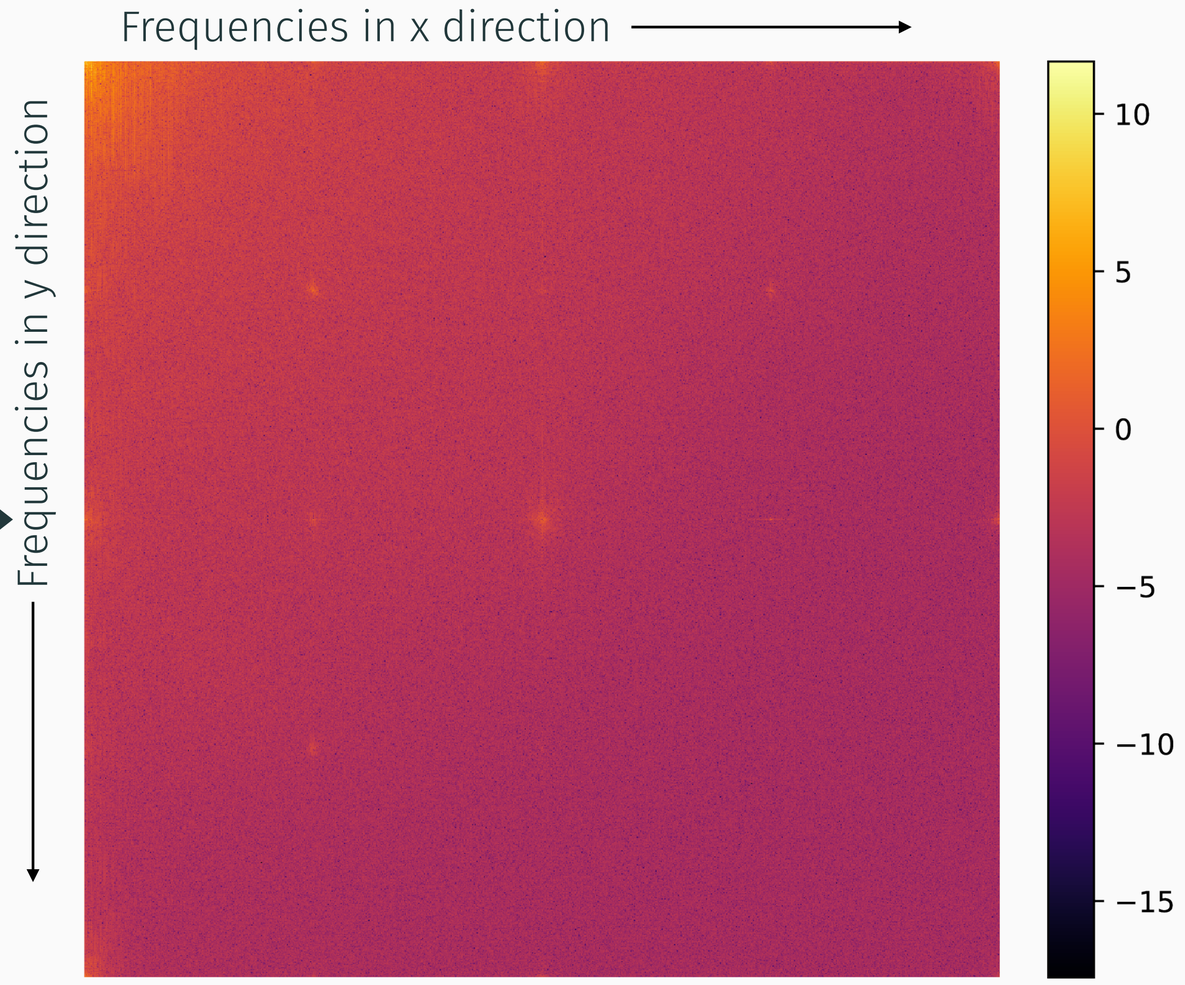
$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

Frequency Domain



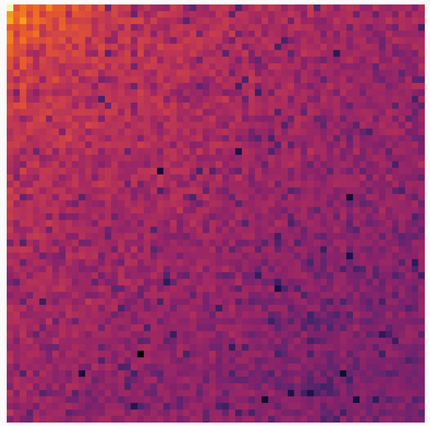
$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

Frequency Domain

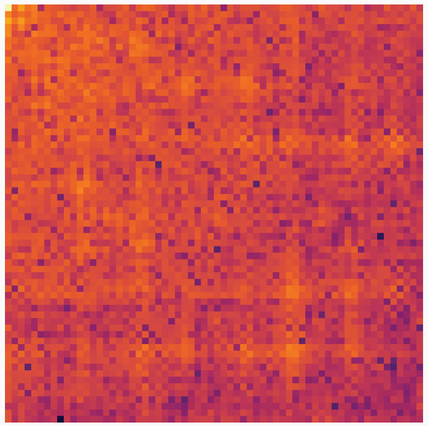


$$D_{k_x,k_y} = \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} I_{x,y} \cos\left[\frac{\pi}{N_1}\left(x+\frac{1}{2}\right)k_x\right] \cos\left[\frac{\pi}{N_2}\left(y+\frac{1}{2}\right)k_y\right].$$

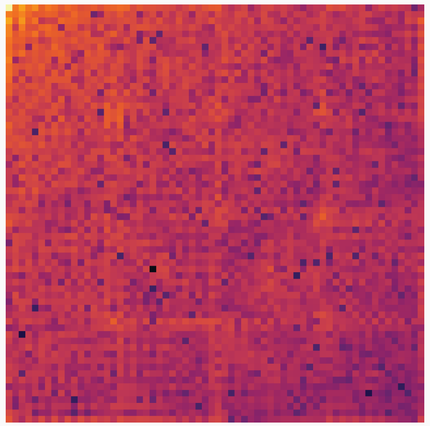
Specific to StyleGAN?



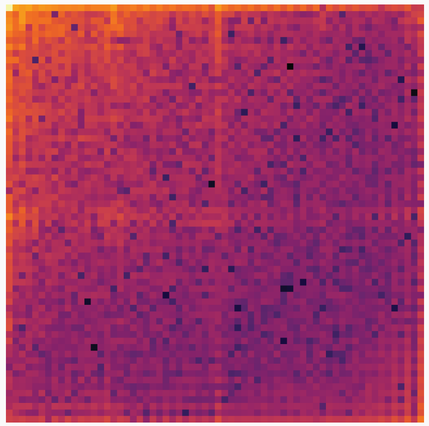
Stanford dogs



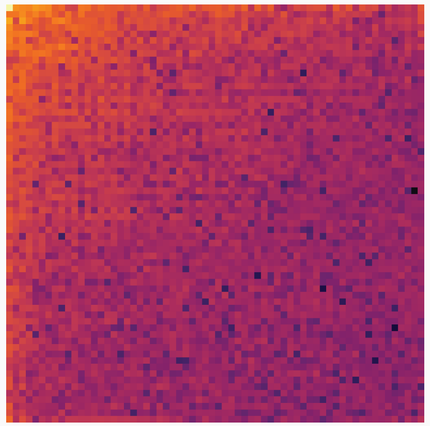
BigGAN



ProGAN



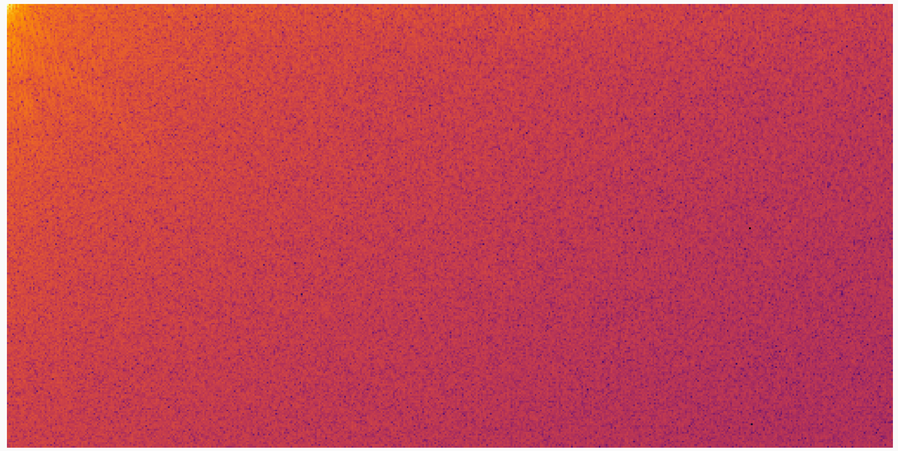
SN-DCGAN

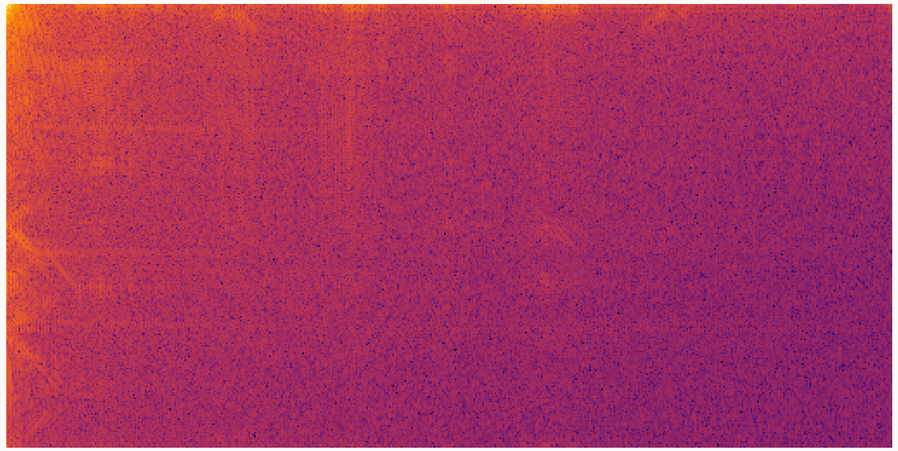


StyleGAN

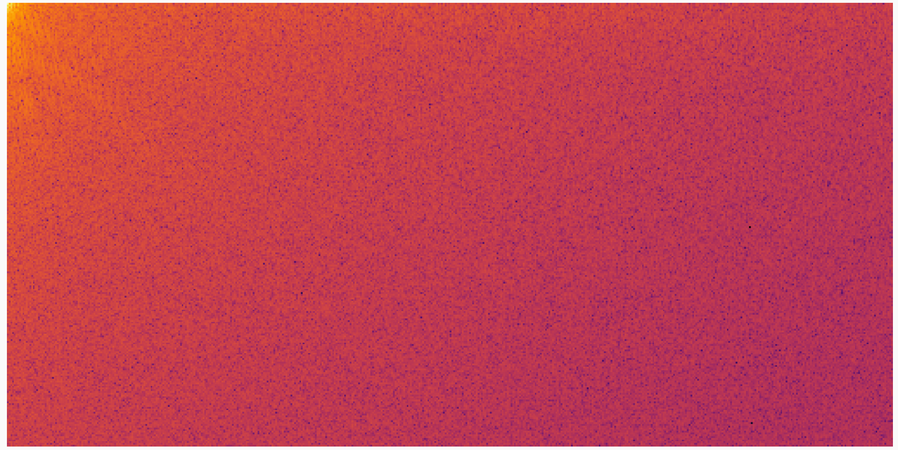
Specific to GANs?

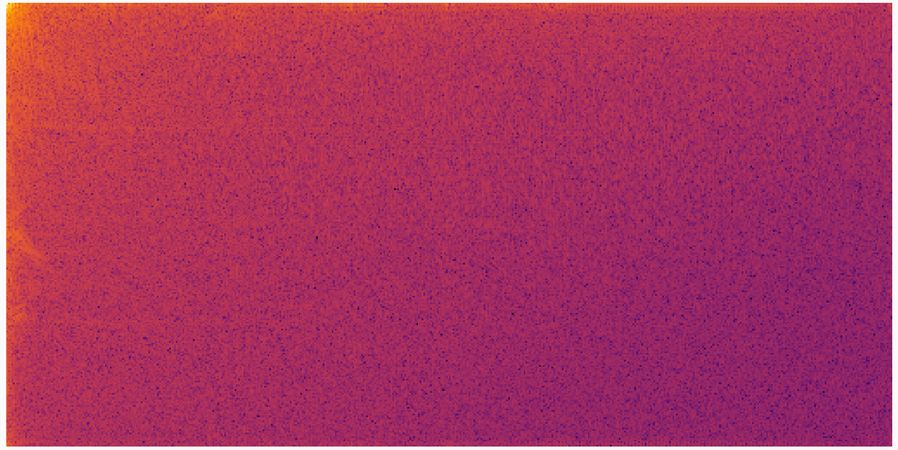
Cascaded Refinement Networks





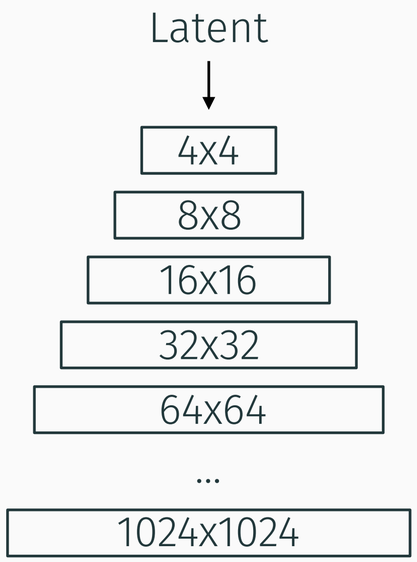
Implicit Maximum Likelihood Estimation





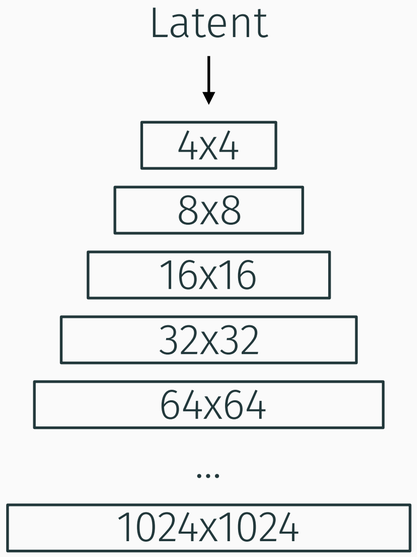
Wang, et al., "CNN-generated images are surprisingly easy to spot... for now", CVPR 2020

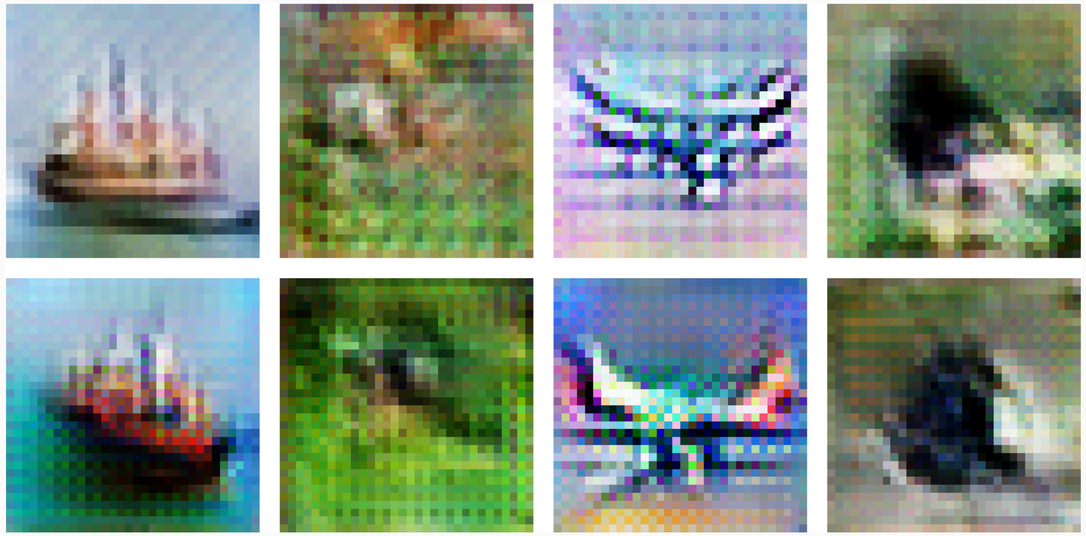
Upsampling?



Upsampling?

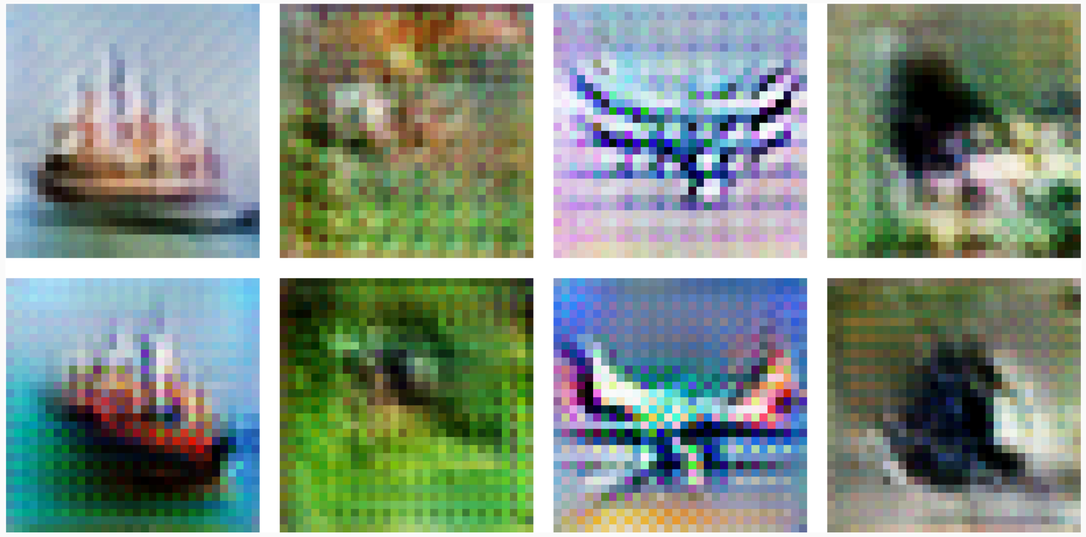
Odena, et al., "Deconvolution and Checkerboard Artifacts", Distill 2016



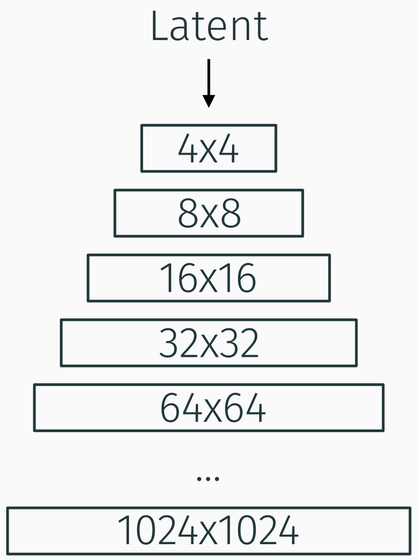


Upsampling?

Odena, et al., "Deconvolution and Checkerboard Artifacts", Distill 2016

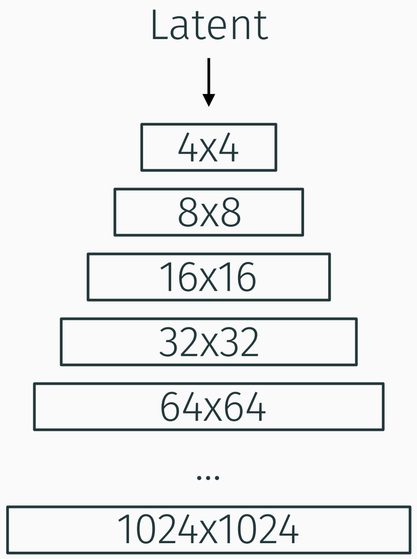


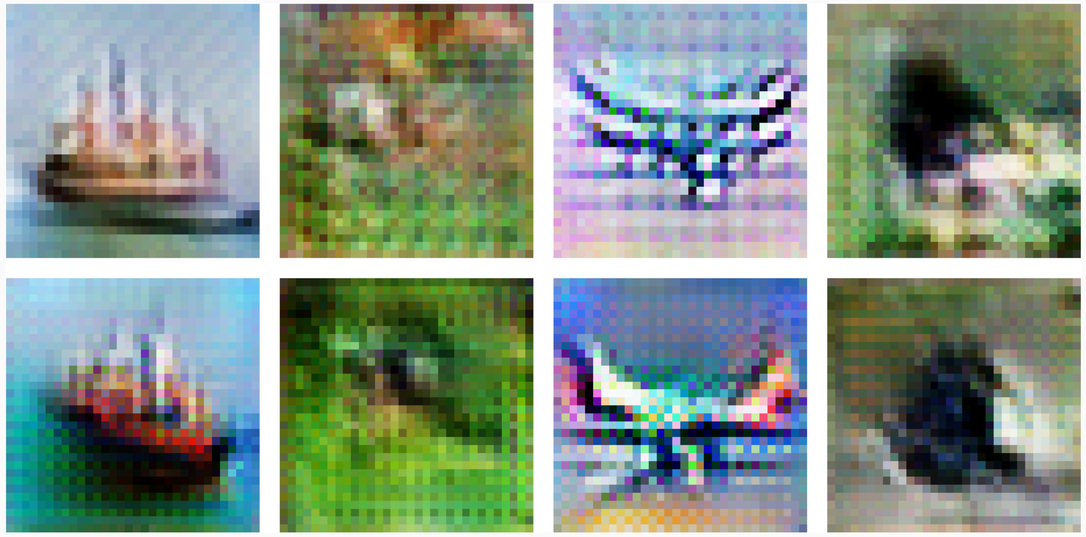
Strided Transposed Convolution → Upsampling + Convolution



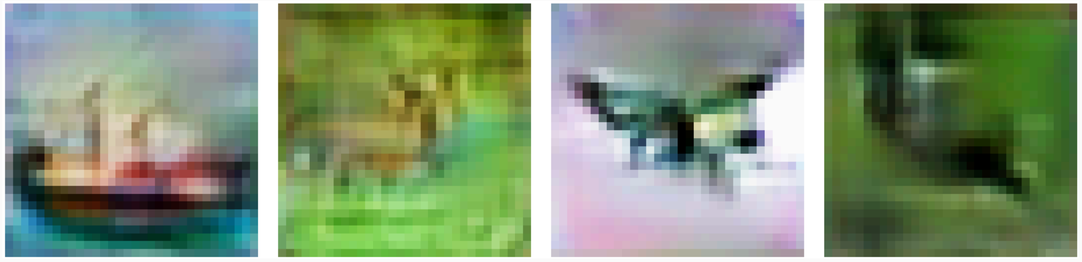
Upsampling?

Odena, et al., "Deconvolution and Checkerboard Artifacts", Distill 2016





Strided Transposed Convolution → Upsampling + Convolution



Durall, et al., "Watch your Up-Convolution: CNN Based Generative Deep Neural Networks are Failing to Reproduce Spectral Distributions", CVPR 2020

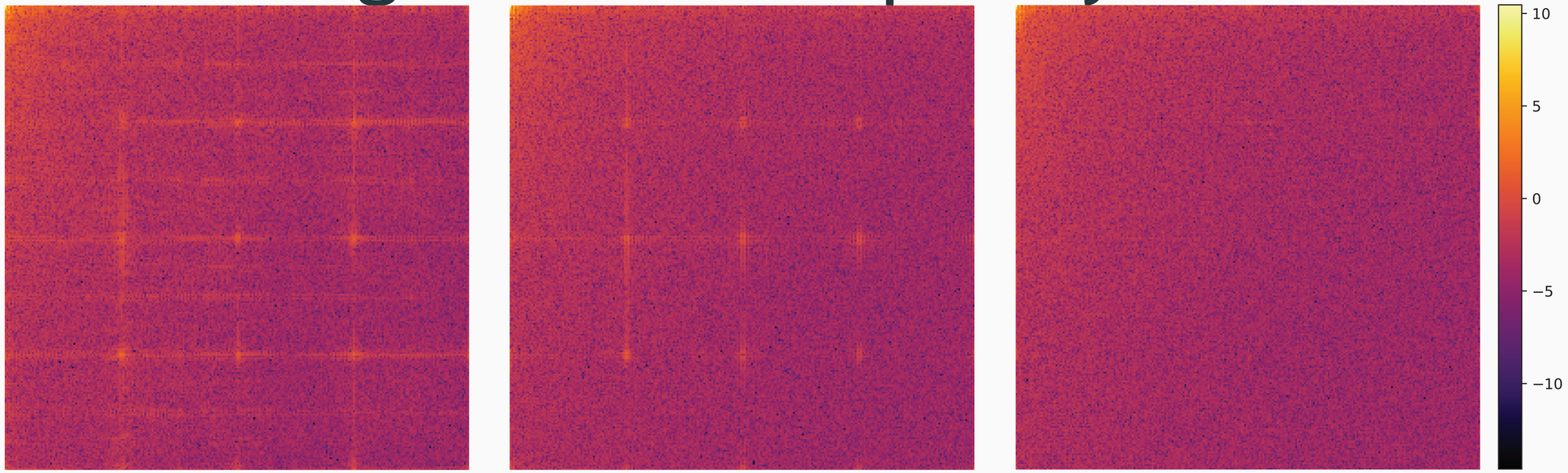
Advantages of the Frequency Domain

Domain	Accuracy
Image	75.78%
Frequency	100.00%

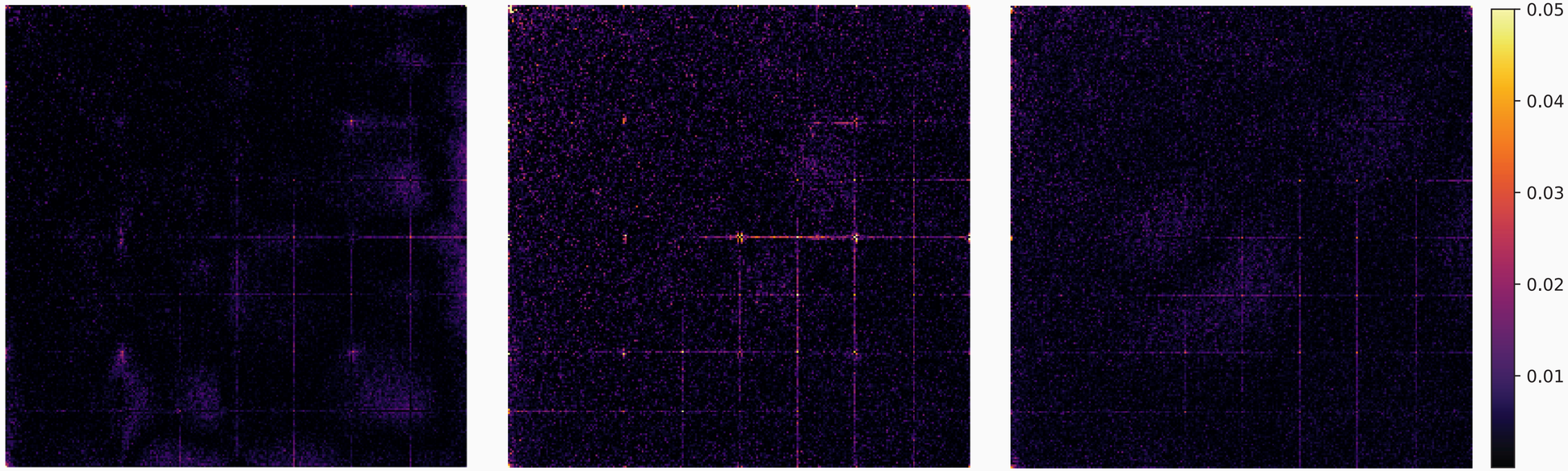
Advantages of the Frequency Domain

- Frequency domain enables linear separability

Advantages of the Frequency Domain



Nearest Neighbour | Bilinear | Binomial



Advantages of the Frequency Domain

- Frequency domain enables linear separability
- Still artifacts for more elaborate upsampling techniques

Advantages of the Frequency Domain

- Frequency domain enables linear separability
- Still artifacts for more elaborate upsampling techniques
- For existing source attribution tasks, we can reduce the error rate by up to 75%

Advantages of the Frequency Domain

- Frequency domain enables linear separability
- Still artifacts for more elaborate upsampling techniques
- For existing source attribution tasks, we can reduce the error rate by up to 75%
- Neural network training is easier and needs less training data

Advantages of the Frequency Domain

- Frequency domain enables linear separability
- Still artifacts for more elaborate upsampling techniques
- For existing source attribution tasks, we can reduce the error rate by up to 75%
- Neural network training is easier and needs less training data
- Experiments on corrupted data