SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

Keshigeyan Chandrasegaran, Ngoc-Trung Tran, Ngai-Man Cheung

{ keshigeyan, ngoctrung_tran, ngaiman_cheung } @sutd.edu.sg

CVPR 2021 Oral

CNN-generated images have become indistinguishable from real images.

Left to right : Real, StyleGAN, StyleGAN2, PGGAN, VQ-VAE2, and ALAE generated images. [Dzanic et al. NeurIPS 2020]

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

A Fun Exercise : Which of these images are CNN-generated?

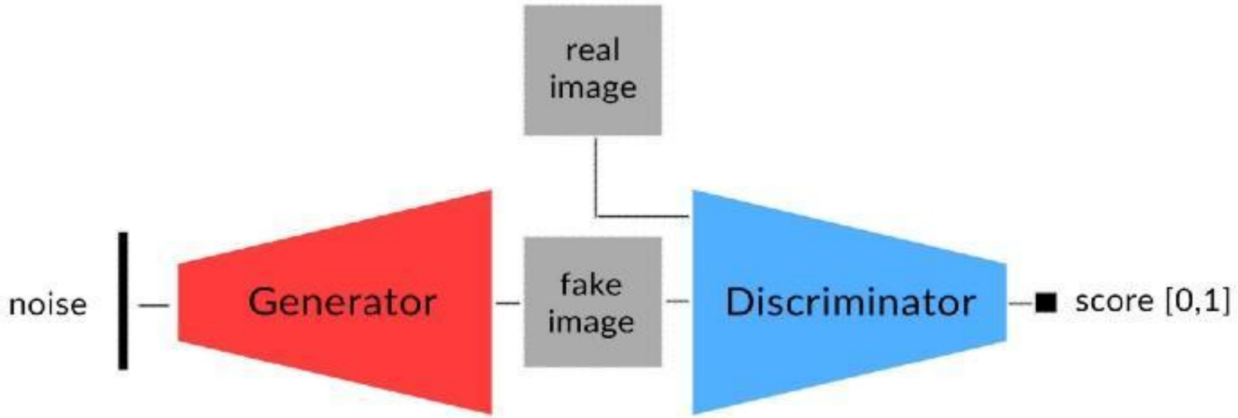
* Synthetic images generated by architectural variant of ProGAN (Zero-insertion based upsampling for all steps)

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

A Fun Exercise : Which of these images are CNN-generated? **Only 1 real, rest are synthetic**

* Synthetic images generated by architectural variant of ProGAN (Zero-insertion based upsampling for all steps)

Background : Generative Adversarial Networks (GAN)



GANs simultaneously train two models :

- A generative model G that captures the data distribution, and
- A discriminative model D that estimates the probability that a sample came from the training data rather than G.

The training procedure for G is to maximize the probability of D making a mistake. [Goodfellow, NeurIPS 2014]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

With serious concerns over Deepfakes being widely used for malicious purposes,
detection of deepfake multimedia content
has become an important research field.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

How to build detectors to identify CNN-generated images?

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

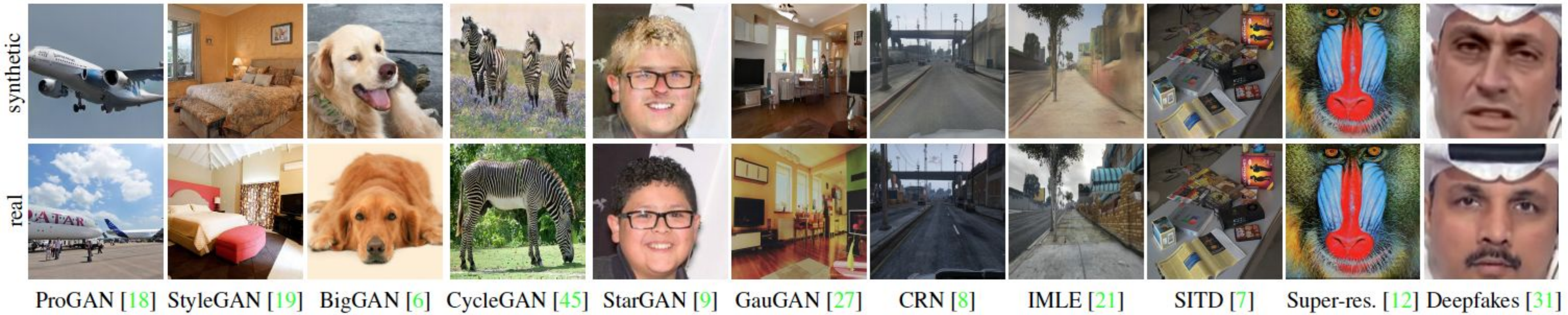
How to build detectors to identify CNN-generated images?

Deep Neural Network
based detectors

Hand-crafted feature
based detectors

Deep Neural Network (ResNet-50) based detector

[Wang et al. CVPR 2020]



A classifier trained to detect images generated by only one CNN (ProGAN, far left) can detect those generated by many other models (remaining columns).

This suggests the possibility that today's CNN-generated images share some common systematic flaws.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

How to build detectors to identify CNN-generated images?

Deep Neural Network based detectors

- Good detection accuracies
- Complex DNN architectures
- Computationally intensive
- Black box : learnt features used for detection are not explicit

Hand-crafted feature based detectors

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

How to build detectors to identify CNN-generated images?

Deep Neural Network based detectors

- Good detection accuracies
- Complex DNN architectures
- Computationally intensive
- Black box : learnt features used for detection are not explicit

Hand-crafted feature based detectors

- Good detection accuracies (Dzanic et al. NeurIPS 2020, Durall et al. CVPR 2020)
- Simple
- Lightweight
- Transparent

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

How to build detectors to identify CNN-generated images?

Deep Neural Network based detectors

- Good detection accuracies
- Complex DNN architectures
- Computationally intensive
- Black box : learnt features used for detection are not explicit

Hand-crafted feature based detectors

- Good detection accuracies (Dzanic et al. NeurIPS 2020, Durall et al. CVPR 2020)
- Simple
- Lightweight
- Transparent

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Recently some works have observed that CNN-generated images share a systematic shortcoming in replicating high frequency Fourier spectrum decay attributes.

Background : Fourier Transform

Fourier analysis is the study of how general functions can be decomposed into trigonometric or exponential functions with definite frequencies. There are two types of Fourier expansions:

- Fourier series: If a (reasonably well-behaved) function is periodic, then it can be written as a discrete sum of trigonometric or exponential functions with specific frequencies.
- Fourier transform: A general function that isn't necessarily periodic (but that is still reasonably well-behaved) can be written as a continuous integral of trigonometric or exponential functions with a continuum of possible frequencies.

https://scholar.harvard.edu/files/david-morin/files/waves_fourier.pdf

Background : Fourier Transform

Fourier analysis is the study of how general functions can be decomposed into trigonometric or exponential functions with definite frequencies. There are two types of Fourier expansions:

- Fourier series: If a (reasonably well-behaved) function is periodic, then it can be written as a discrete sum of trigonometric or exponential functions with specific frequencies.

- Fourier transform: A general function that isn't necessarily periodic (but that is still reasonably well-behaved) can be written as a continuous integral of trigonometric or exponential functions with a continuum of possible frequencies.

- An Image is a 2D discrete signal.

https://scholar.harvard.edu/files/david-morin/files/waves_fourier.pdf

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

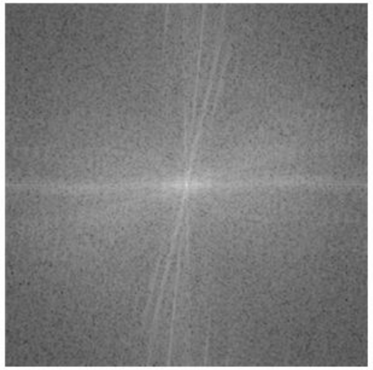
Background : 2D Fourier Transform and Image Frequency

Definition

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}\,dx\,dy,$$
$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j2\pi(ux+vy)}\,du\,dv$$

f(x,y)

|F(u,v)|



https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf

Background : 2D Fourier Transform and Image Frequency

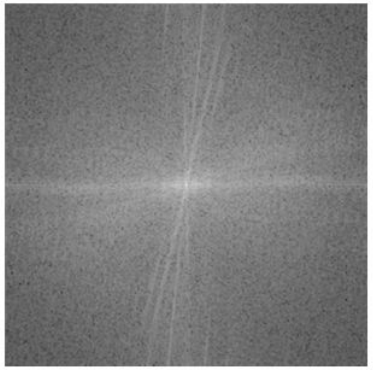
Definition

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}\,dx\,dy,$$
$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j2\pi(ux+vy)}\,du\,dv$$

$f(x,y)$

Spatial Domain

$|F(u,v)|$



Frequency Domain

https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf

Background : 2D Fourier Transform and Image Frequency

Definition

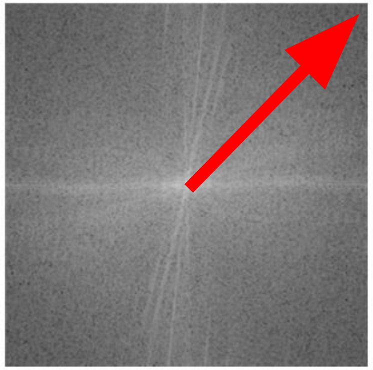
$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}\,dx\,dy,$$

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j2\pi(ux+vy)}\,du\,dv$$

$f(x,y)$

$|F(u,v)|$

Magnitude spectrum
(No phase information)

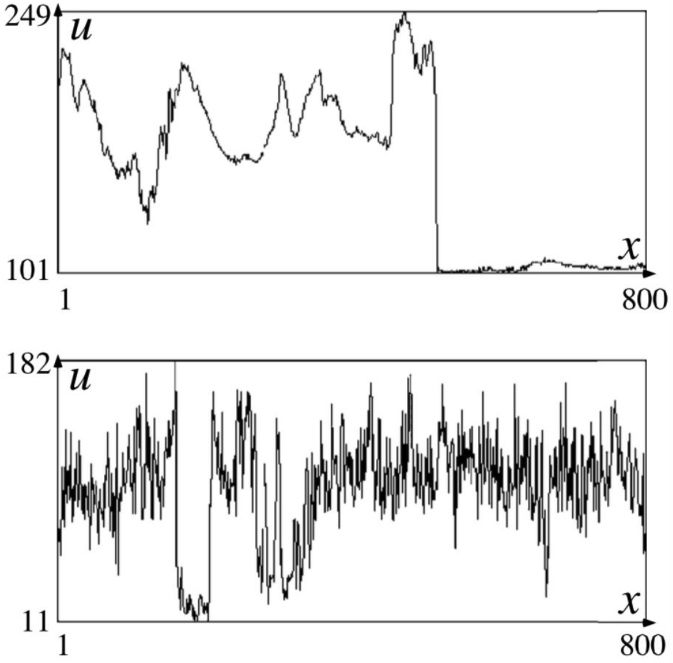


Frequency increases (radially)

https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf

Background : Image Frequency Example (Horizontal frequency)

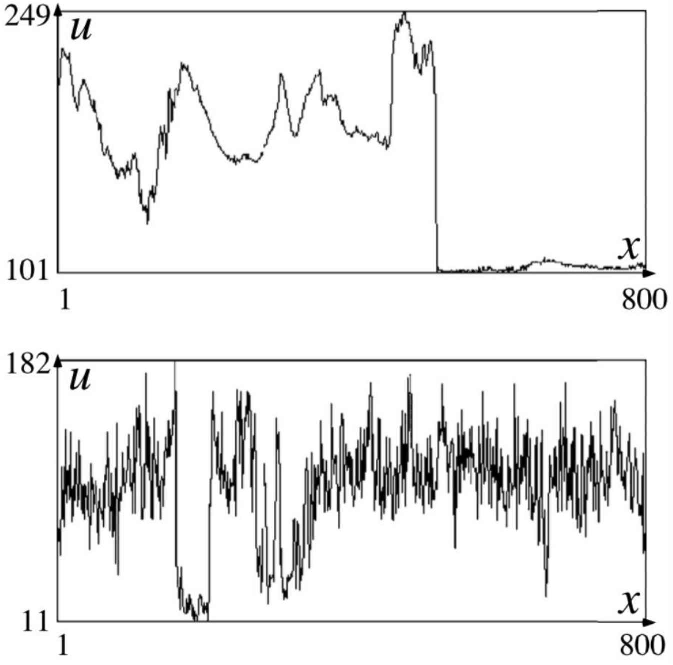
Which row has higher spatial frequency?



https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf

Background : Image Frequency Example (Horizontal frequency)

Low spatial frequency

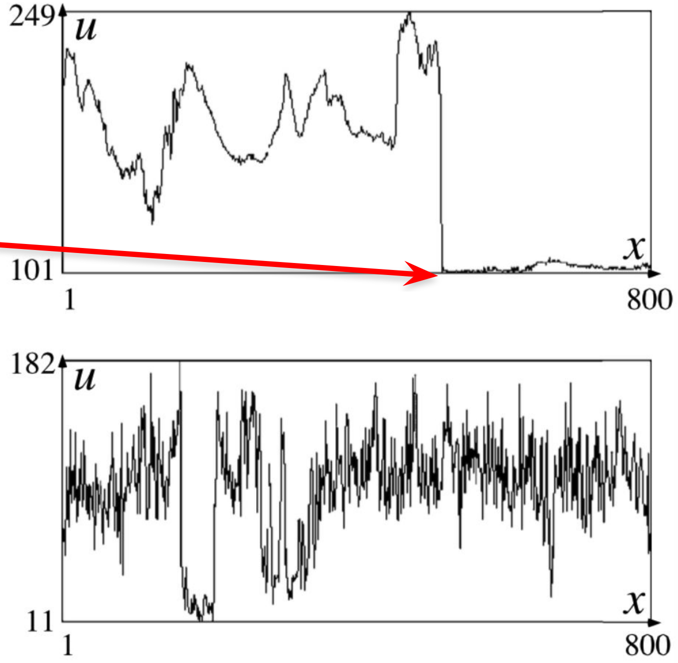


High spatial frequency

https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf

Background : Image Frequency Example (Horizontal frequency)

Low spatial frequency

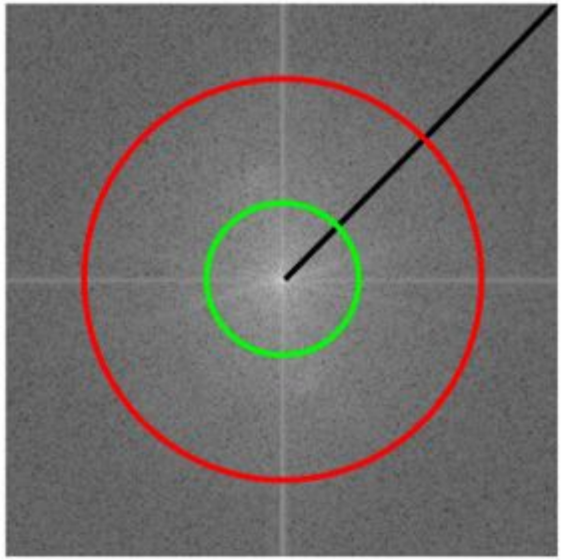


High spatial frequency

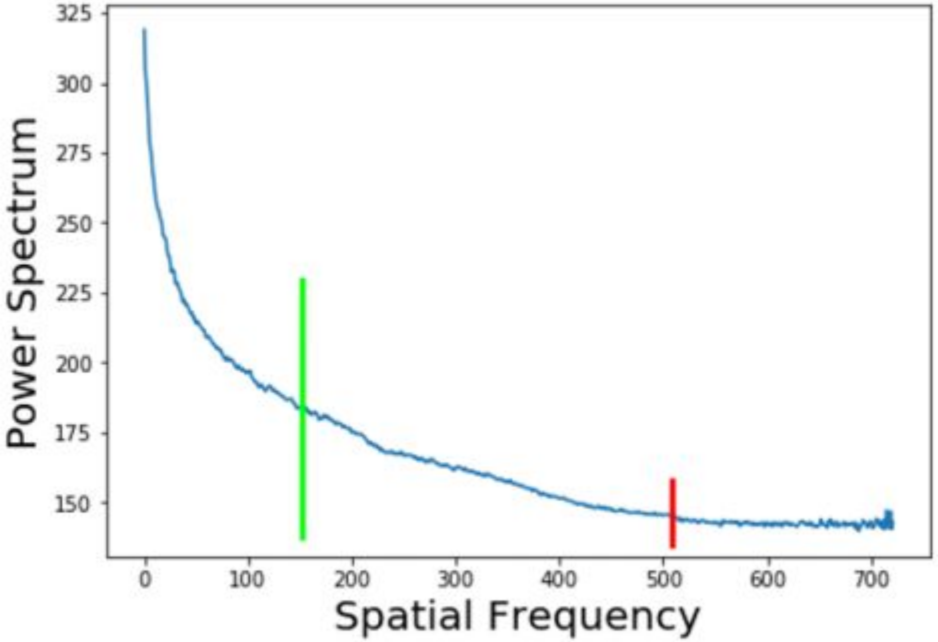
https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf

Background : 2D Fourier Transform and Azimuthal Averaging

Original Image



Magnitude spectrum

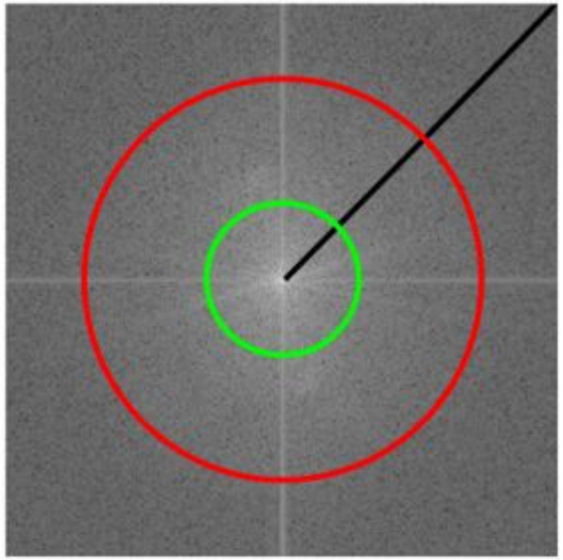


Azimuthal average of Magnitude spectrum over radial frequencies

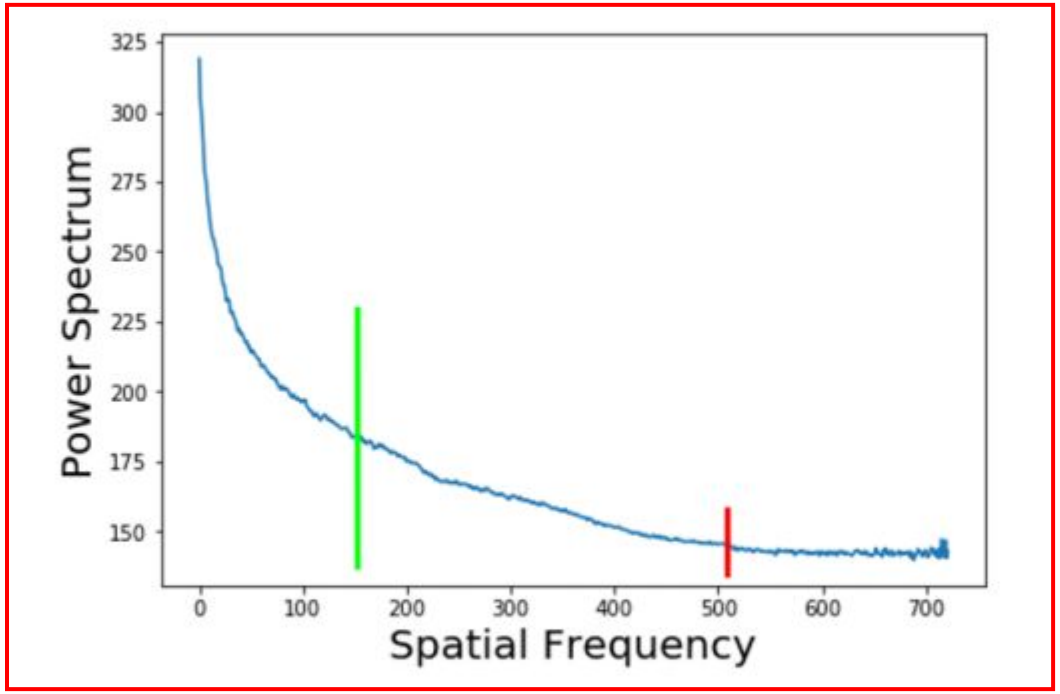
[Durall et al. CVPR 2020]

Background : 2D Fourier Transform and Azimuthal Averaging

Original Image



Magnitude spectrum



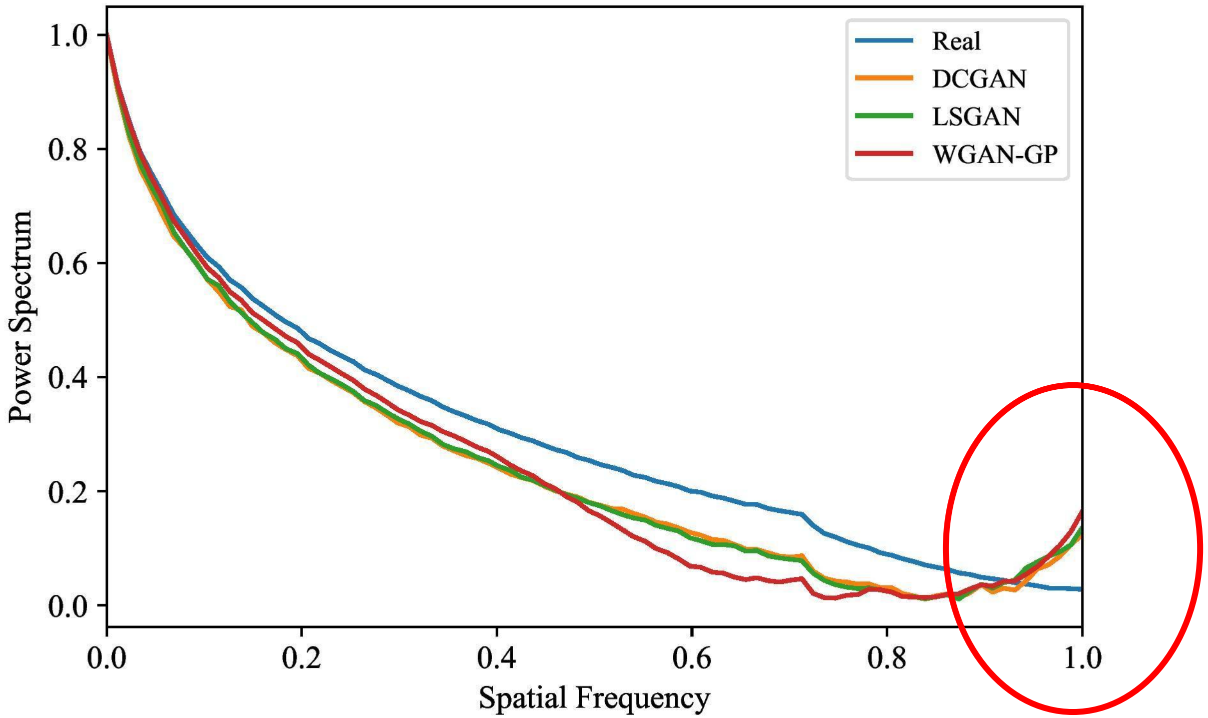
Azimuthal average of Magnitude spectrum over radial frequencies

[Durall et al. CVPR 2020]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

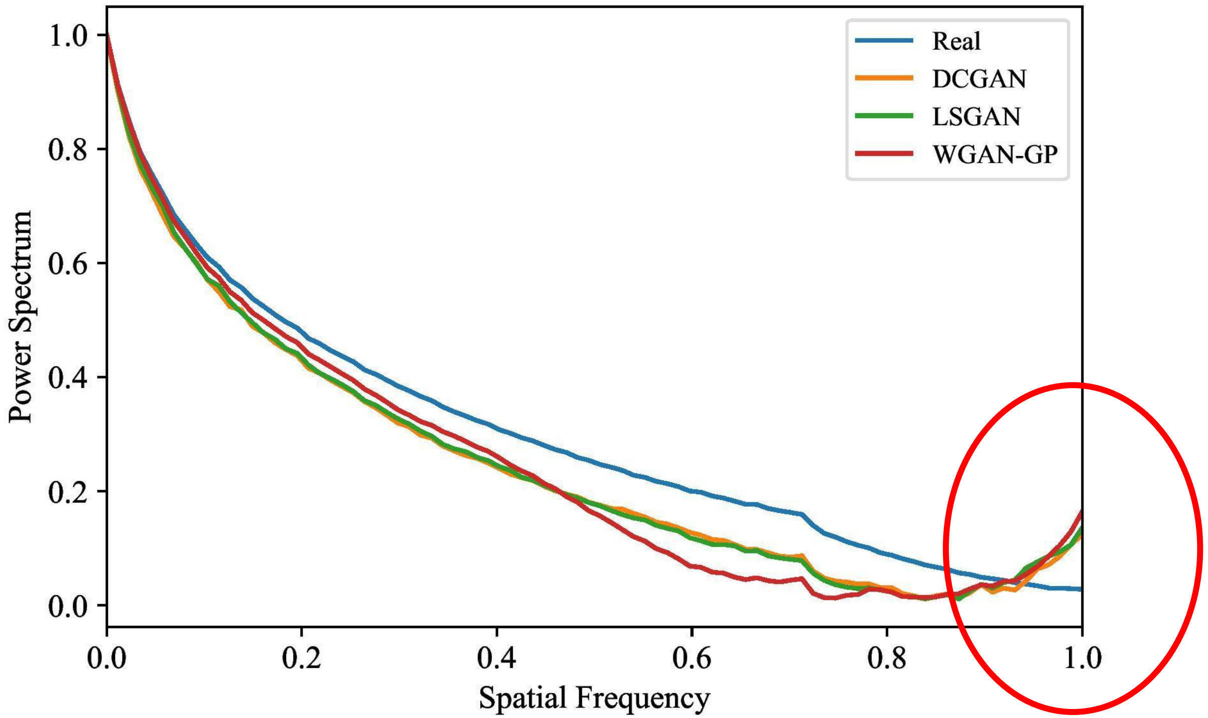
Recently some works have observed that CNN-generated images share a systematic shortcoming in replicating high frequency Fourier spectrum decay attributes.

Fourier Spectrum Decay Discrepancies in CNN-generated Images



CNN generated images at the highest frequencies do not decay as usually observed in real images.

Fourier Spectrum Decay Discrepancies in CNN-generated Images

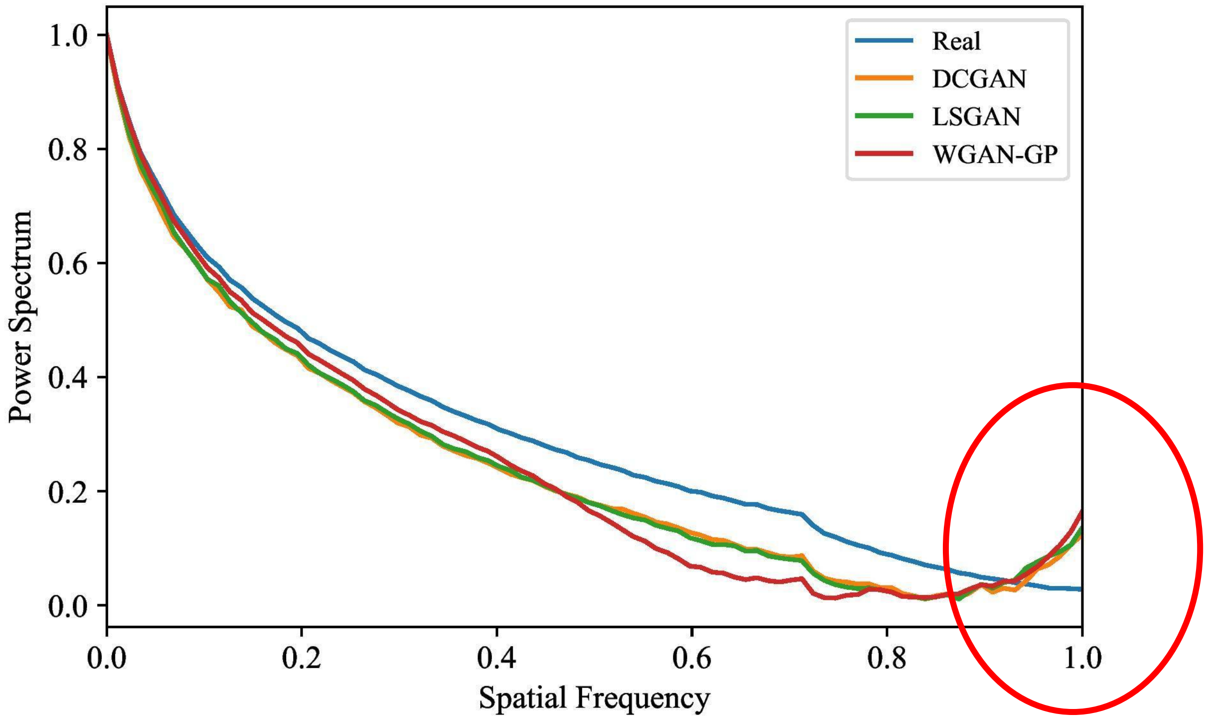


“ CNN-based generative deep neural networks are failing to reproduce spectral distributions. ”

“ This effect is independent of the underlying architecture ”

[Durall et al. CVPR 2020]

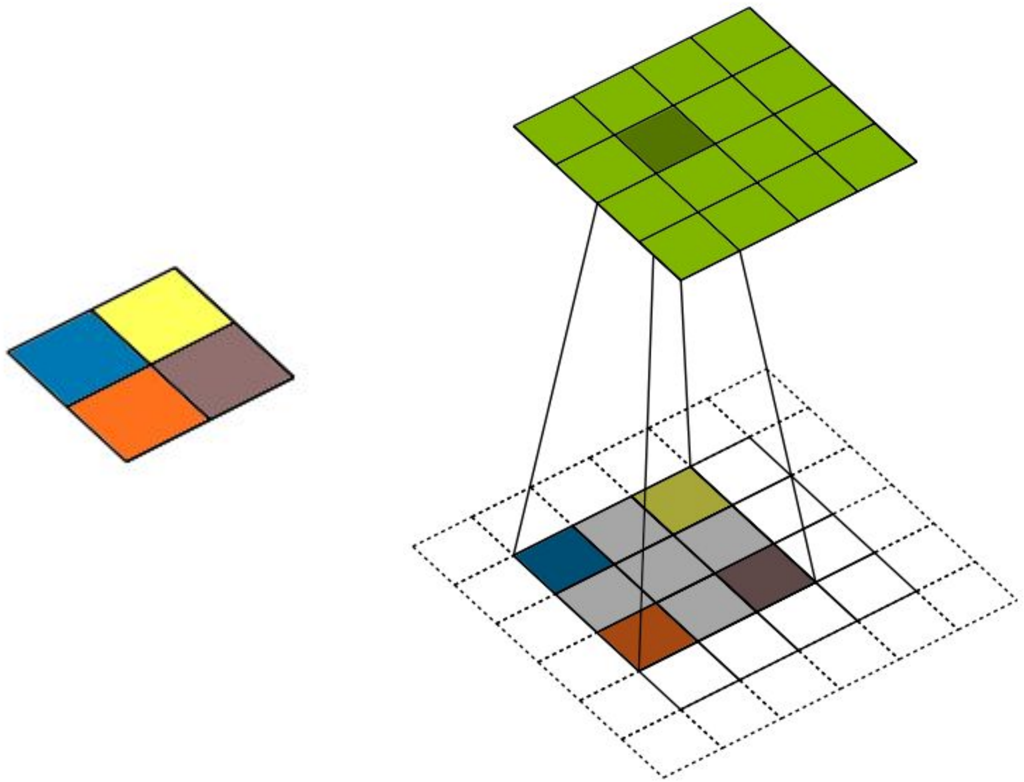
Fourier Spectrum Decay Discrepancies in CNN-generated Images



"deep network generated images share an observable, systematic shortcoming in replicating the attributes of these high-frequency modes. "

[Dzanic et al. NeurIPS 2020]

Fourier Spectrum Decay Discrepancies in CNN-generated Images



Transposed convolution: up-sampling with zero insertion + convolution with a filter kernel creates high-frequency artifacts

[Durall et al. CVPR 2020]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

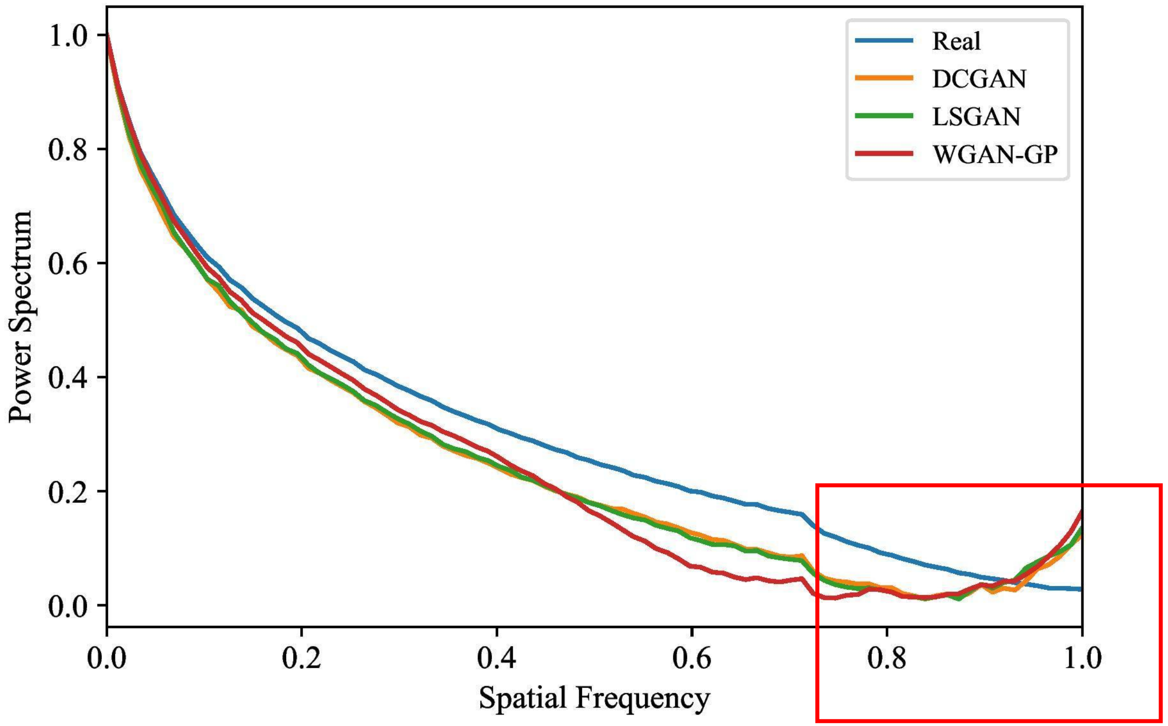
Detector : High frequency decay attributes + kNN classifier

[Dzanic et al. NeurIPS 2020]

Experiment	Resolution	Compression Quality	Overall Class. Acc.	StyleGAN Class. Acc.	StyleGAN2 Class. Acc.	PGGAN Class. Acc.	VQ-VAE2 Class. Acc.	ALAE Class. Acc.
A	1024^2	100	**99.2%**	99.9%	99.5%	97.4%	99.8%	99.8%
B	1024^2	95	**94.4%**	99.2%	88.5%	88.5%	100 %	99.7%
C	1024^2	85	**83.9%**	78.9%	65.9%	78.7%	99.6%	87.4%
D	768^2	100	**98.5%**	100 %	99.1%	95.9%	99.9%	99.9%
E	768^2	95	**93.0%**	97.9%	85.4%	87.3%	100 %	99.5%
F	768^2	85	**84.6%**	77.1%	68.6%	79.3%	99.6%	85.7%
G	256^2	100	**88.8%**	85.0%	87.4%	69.0%	92.0%	90.7%
H	256^2	95	**88.1%**	81.7%	83.4%	68.2%	92.2%	87.7%
I	256^2	85	**87.4%**	67.8%	79.3%	64.8%	87.7%	80.6%

Attributes from 0.75 to 1.0 Spatial frequency are used.

Fourier Spectrum Decay Discrepancies in CNN-generated Images



Attributes from 0.75 to 1.0 Spatial frequency are used.

[Dzanic et al. NeurIPS 2020]

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

"High frequency spectral decay discrepancies are not intrinsic for CNN-generated images. Therefore, we urge re-thinking in using such features for CNN generated image detection."

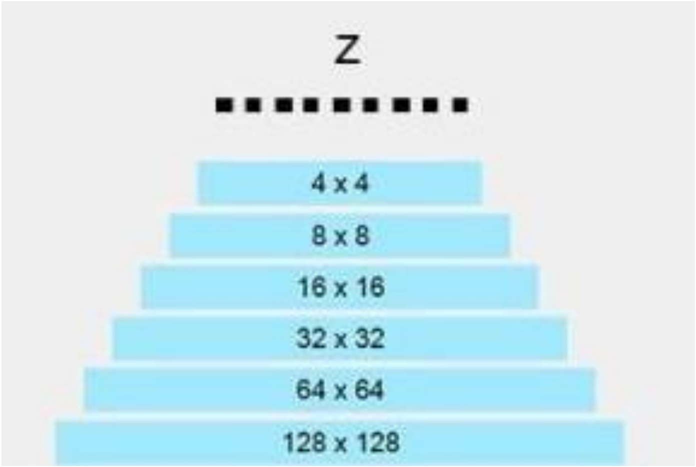
A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

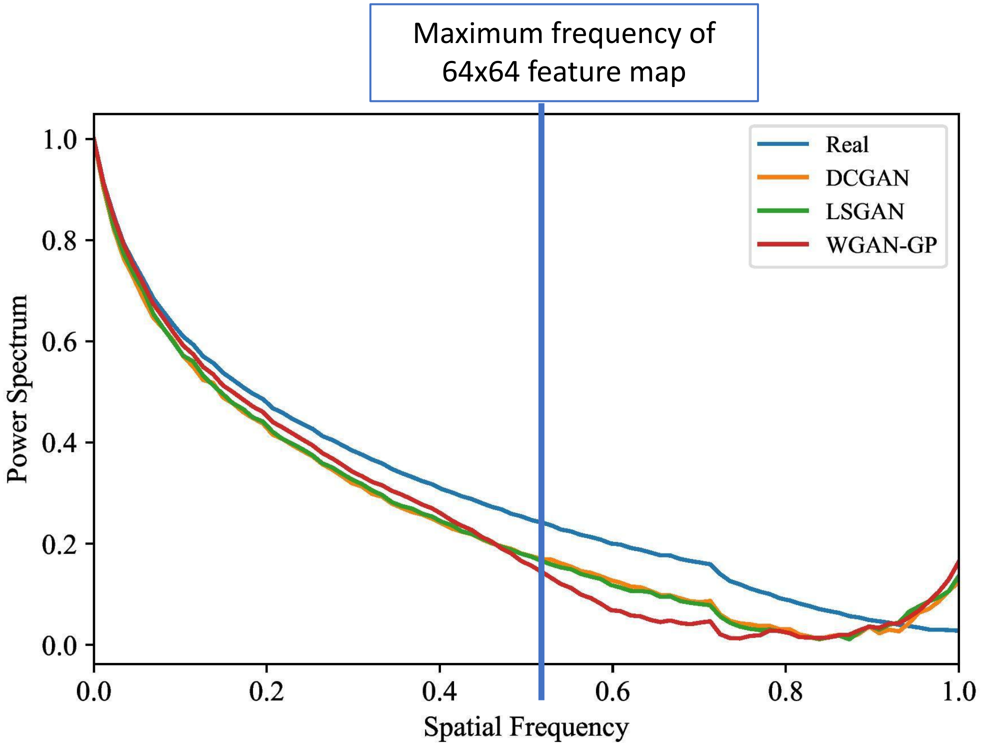
[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

- Given discrepancies reported in the highest frequencies, we hypothesize that inner generator layers that produce lower resolution outputs may not be directly responsible for the high frequency discrepancies. (Based on Sampling theorem)

- Therefore, we focus on the last upsampling step.





A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

In particular, we split the last upsampling step into 2 operations,

1) Feature map scaling
2) Convolution

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Feature map scaling

- Zero-Insert scaling
- Nearest interpolation.
- Bilinear interpolation.

In terms of amount of high frequency content injection into feature maps
Zero-insertion > Bilinear interpolation > Nearest interpolation

Convolution

The subsequent convolution operation learns kernels in order to satisfy the optimization objective.
Convolutional kernels are capable of suppressing/ amplifying high frequencies.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Test Bed to study the effect of feature map scaling and convolution in the last upsampling step of the generator.

Code	**Details**
Baseline	Transpose convolution (4x4 kernel)
N.1.5	Nearest Upsampling + 1 conv block of 5x5 kernel
Z.1.5	Zero insert Upsampling + 1 conv block of 5x5 kernel
B.1.5	Bilinear Upsampling + 1 conv block of 5x5 kernel
N.1.3	Nearest Upsampling + 1 conv block of 3x3 kernel
N.1.7	Nearest Upsampling + 1 conv block of 7x7 kernel
Z.1.3	Zero insert Upsampling + 1 conv block of 3x3 kernel
Z.1.7	Zero insert Upsampling + 1 conv block of 7x7 kernel
B.1.3	Bilinear Upsampling + 1 conv block of 3x3 kernel
B.1.7	Bilinear Upsampling + 1 conv block of 7x7 kernel
N.3.5	Nearest Upsampling + 3 conv blocks of 5x5 kernel
Z.3.5	Zero insert upsampling + 3 conv blocks of 5x5 kernel
B.3.5	Bilinear upsampling + 3 conv blocks of 5x5 kernel

Table 1. Test Bed to study the effect of feature map scaling and convolution in the *last* upsampling step of the generator, for different GANs: DCGAN, LSGAN, WGAN-GP, StarGAN. "Baseline" refers to public released code of the GAN model, which uses transpose convolution of 4×4 kernel. For other models, we replace the last transpose convolution in Baseline with the corresponding configurations shown. We emphasize that we only modify the last step specified as above; the algorithms, learning objectives and architectures (except the last step) are kept identical as the public released code for different GAN models.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Results : Effect of Feature Map Scaling Methods

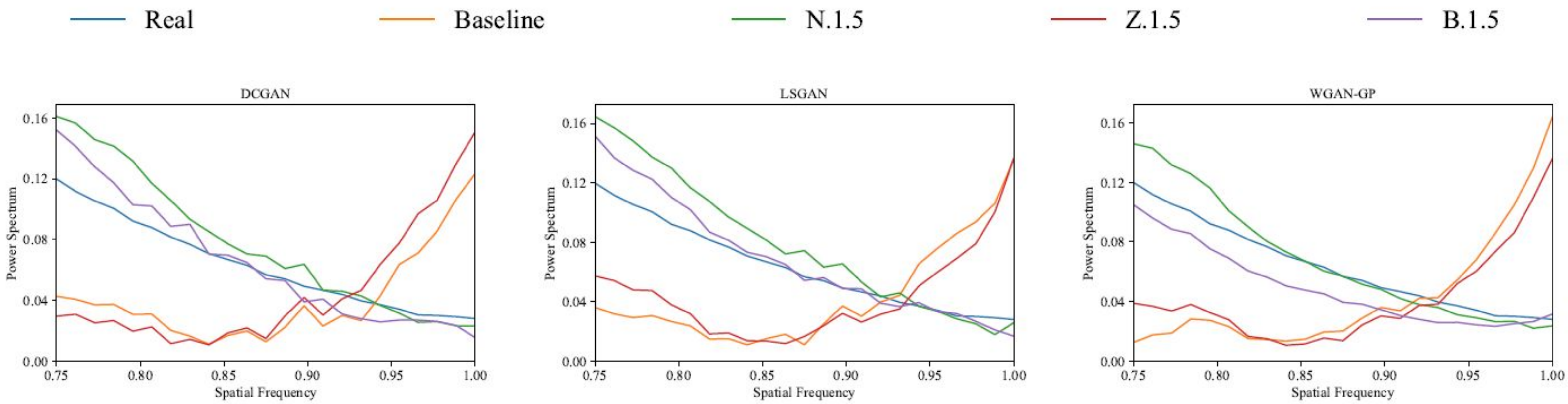


Figure 3. Feature Map Scaling Results. We observe that experiments using nearest and bilinear interpolation methods in the last step are able to produce spectral consistent GANs. Refer to table 1 for experiment details.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Results : Effect of Kernel Size (3x3)

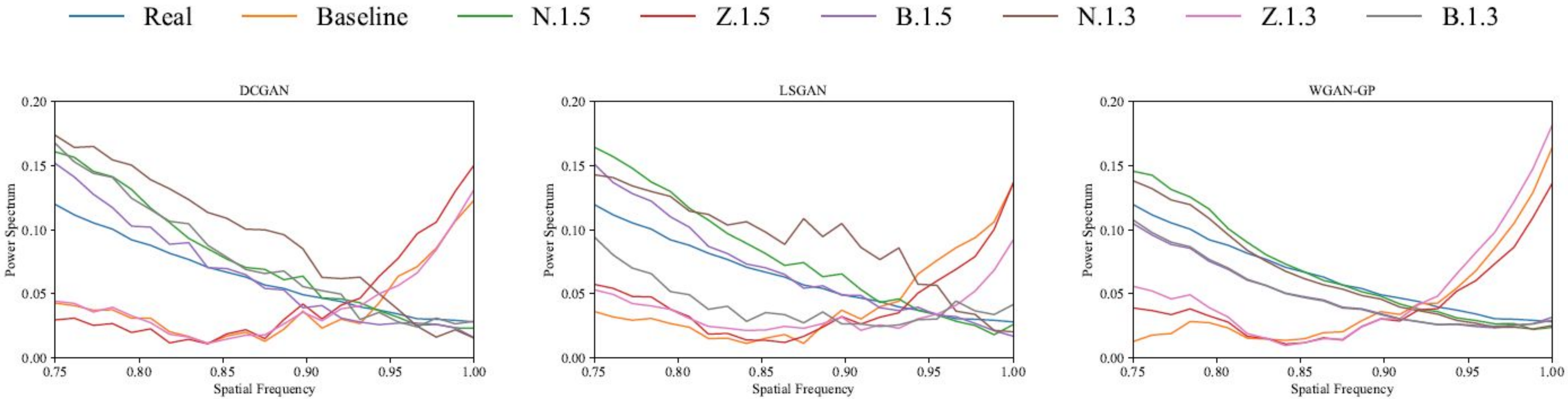


Figure 4. Smaller Kernel (3x3) Results. We observe that smaller kernels do not substantially deteriorate spectral consistent GANs except for some turbulent behaviour observed in LSGAN for N.1.3 and B.1.3 experiments. Refer to table 1 for experiment details.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Results : Effect of Kernel Size (7x7)

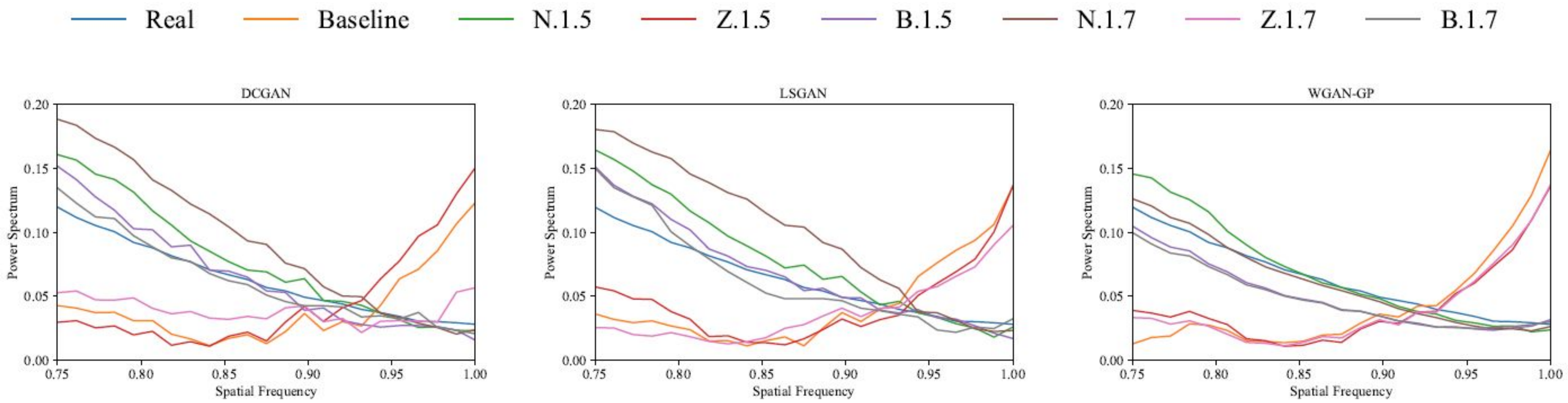


Figure 5. Larger Kernel (7x7) Results. We observe that larger kernels do not substantially manipulate the discrepancies in Z.1.7 experiments. Refer to table 1 for experiment details.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Results : Effect of Number of Kernels

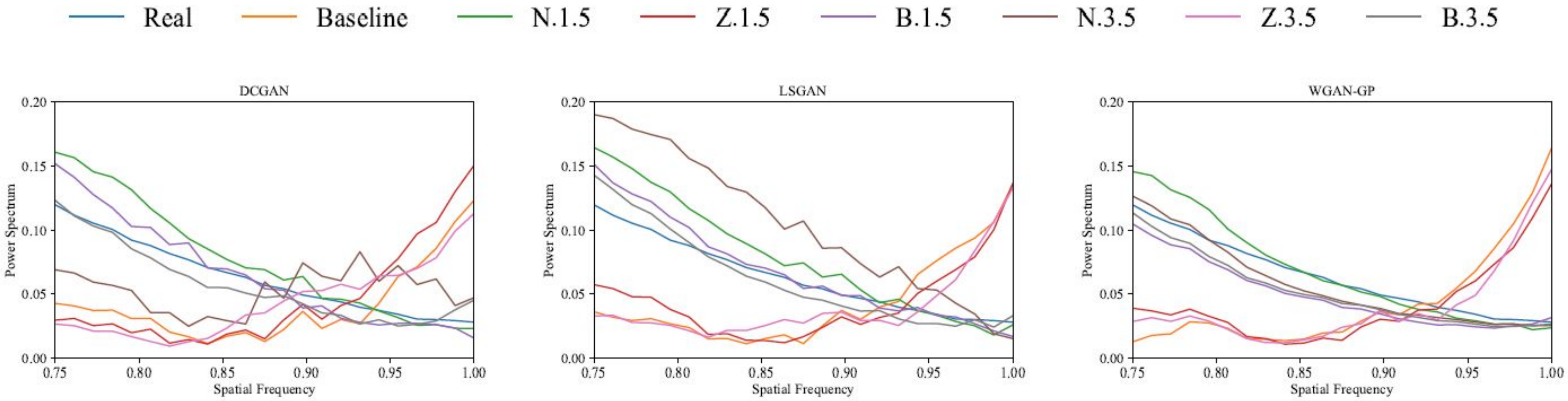
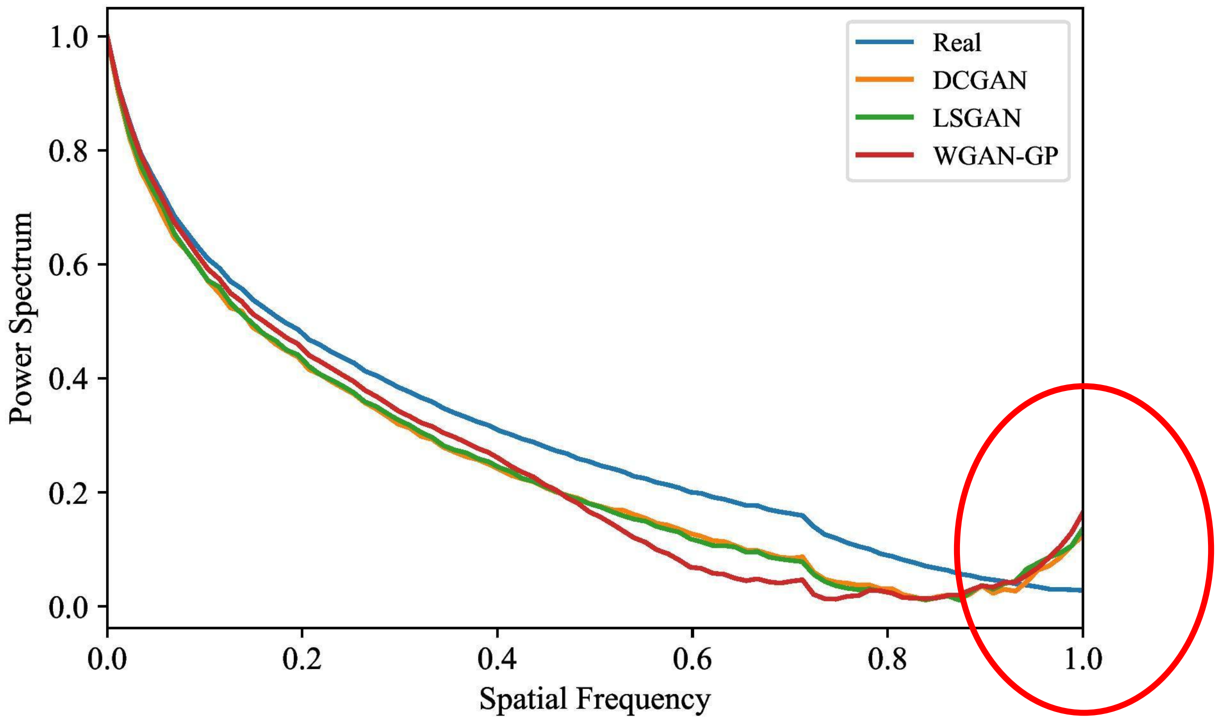


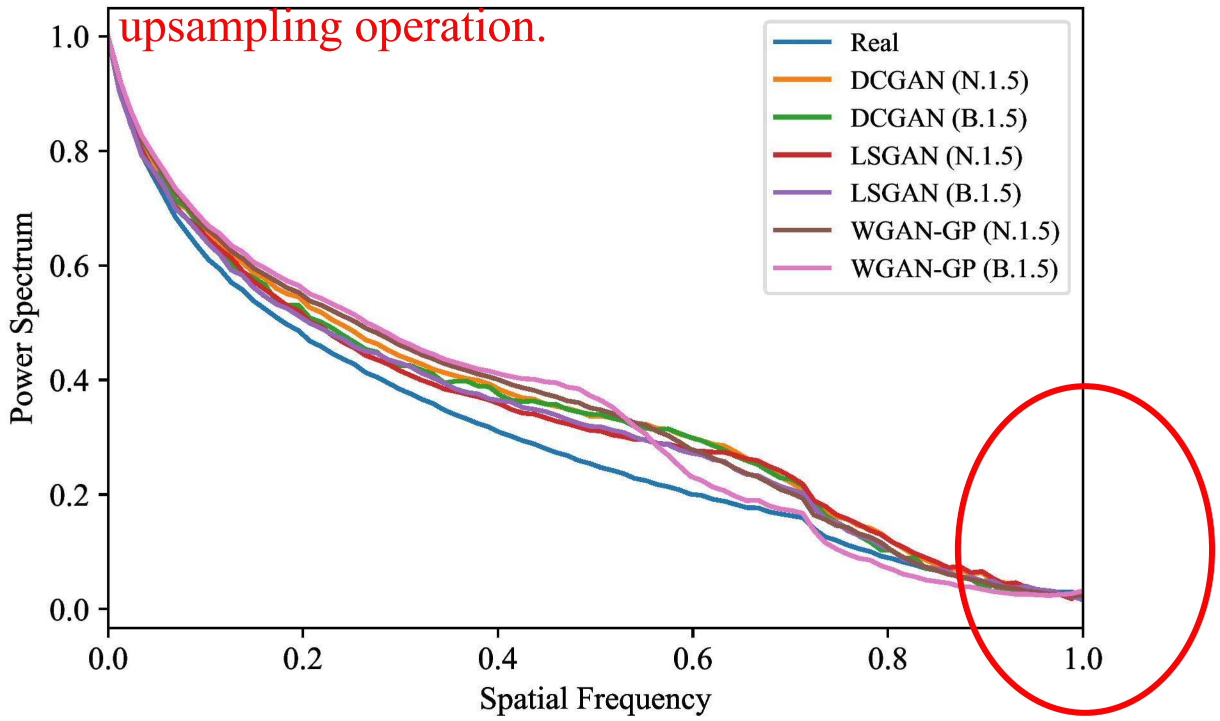
Figure 6. Increased number of kernels (3 conv blocks) Results. We see that even with more number of kernels in the last upsampling step, Z.3.5 experiment is not able to produce spectral consistent GANs. Refer to table 1 for experiment details.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]



High frequency spectral decay discrepancy avoided by a simple architecture change in the last upsampling operation.



A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Setup	DCGAN	LSGAN	WGAN-GP
N.1.5	**0.09 ± 0.03%**	**0.34 ± 0.08%**	**0.14 ± 0.05%**
Z.1.5	84.82 ± 3.72%	88.16 ± 3.98%	99.75 ± 0.14%
B.1.5	**0 ± 0%**	**0.1 ± 0%**	**0.2 ± 0.12%**
N.1.3	**0 ± 0%**	**0.06 ± 0.05%**	**0.24 ± 0.13%**
N.1.7	**0 ± 0%**	**0 ± 0%**	**0.06 ± 0.05%**
Z.1.3	98.73 ± 0.56%	73.09 ± 3.5%	97.94 ± 0.87%
Z.1.7	97.23 ± 1.1%	95.66 ± 1.93%	99.94 ± 0.07%
B.1.3	**0 ± 0%**	**0.19 ± 0.1%**	**0.07 ± 0.05%**
B.1.7	**0 ± 0%**	**0.1 ± 0%**	**0.17 ± 0.13%**
N.3.5	**0.16 ± 0.05%**	**0 ± 0%**	**0 ± 0%**
Z.3.5	77.67 ± 6%	67.66 ± 11.9%	99.9 ± 0.19%
B.3.5	**0.03 ± 0.05%**	**0.48 ± 0.04%**	**0.13 ± 0.05%**

Table 2 : Detection results for the detector proposed by Dzanic et al.

- We successfully by-pass the detector proposed by Dzanic et al. (NeurIPS 2020) that uses high frequency decay attributes as features.
- We emphasize that we obtain these results using identical training algorithms, objective functions and network architectures (except using a different upscaling in the last upsampling operation).

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

We also successfully by-pass stronger classifiers that use high frequency decay attributes as features

Setup	DCGAN	LSGAN	WGAN-GP
N.1.5	**$0.1 \pm 0\%$**	**$0.31 \pm 0.06\%$**	**$0.23 \pm 0.16\%$**
Z.1.5	$82.22 \pm 1.98\%$	$87.33 \pm 2.77\%$	$99.45 \pm 0.21\%$
B.1.5	**$0 \pm 0\%$**	**$0.11 \pm 0.09\%$**	**$0.25 \pm 0.17\%$**
N.1.3	**$0.01 \pm 0.03\%$**	**$0.07 \pm 0.05\%$**	**$0.35 \pm 0.22\%$**
N.1.7	**$0 \pm 0\%$**	**$0 \pm 0\%$**	**$0.05 \pm 0.05\%$**
Z.1.3	$98.3 \pm 0.45\%$	$72.13 \pm 2.21\%$	$96.81 \pm 1.63\%$
Z.1.7	$95.81 \pm 0.93\%$	$95.55 \pm 1.23\%$	$99.24 \pm 0.43\%$
B.1.3	**$0 \pm 0\%$**	**$0.25 \pm 0.12\%$**	**$0.15 \pm 0.15\%$**
B.1.7	**$0 \pm 0\%$**	**$0.11 \pm 0.03\%$**	**$0.3 \pm 0.27\%$**
N.3.5	**$0.1 \pm 0\%$**	**$0 \pm 0\%$**	**$0 \pm 0\%$**
Z.3.5	$74.27 \pm 3.32\%$	$65.37 \pm 6.5\%$	$93.82 \pm 0.6\%$
B.3.5	**$0.04 \pm 0.07\%$**	**$0.5 \pm 0.05\%$**	**$0.21 \pm 0.14\%$**

Table 10 : Detection rates using SVM (RBF kernel) using same features as Dzanic et al.

Setup	DCGAN	LSGAN	WGAN-GP
N.1.5	**$0.1 \pm 0\%$**	**$0.77 \pm 0.15\%$**	**$1.53 \pm 0.32\%$**
Z.1.5	$81.14 \pm 2.9\%$	$83.88 \pm 2.59\%$	$99.77 \pm 0.09\%$
B.1.5	**$0.04 \pm 0.1\%$**	**$0.87 \pm 0.46\%$**	**$3.03 \pm 0.82\%$**
N.1.3	**$0.18 \pm 0.04\%$**	**$0.05 \pm 0.13\%$**	**$1.4 \pm 0.2\%$**
N.1.7	**$0 \pm 0\%$**	**$0.04 \pm 0.05\%$**	**$0.67 \pm 0.18\%$**
Z.1.3	$97.54 \pm 0.41\%$	$72.65 \pm 2.64\%$	$98.11 \pm 0.44\%$
Z.1.7	$94.53 \pm 0.97\%$	$93.07 \pm 1.6\%$	$99.97 \pm 0.05\%$
B.1.3	**$0.03 \pm 0.09\%$**	**$1.6 \pm 0.54\%$**	**$2.79 \pm 0.5\%$**
B.1.7	**$0.01 \pm 0.03\%$**	**$0.42 \pm 0.29\%$**	**$4.63 \pm 1.01\%$**
N.3.5	**$0.17 \pm 0.05\%$**	**$0 \pm 0\%$**	**$0.37 \pm 0.27\%$**
Z.3.5	$74.88 \pm 2.79\%$	$71.22 \pm 4.46\%$	$99.8 \pm 0\%$
B.3.5	**$0.28 \pm 0.14\%$**	**$1.89 \pm 0.45\%$**	**$3.66 \pm 1.19\%$**

Table 11 : Detection rates using MLP (2 hidden layers of size 10 with sigmoid activation) using same features as Dzanic et al.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Quality of GAN images with nearest and bilinear interpolation methods are comparable or better than the Baseline method.

Figure 4: WGAN-GP samples for Baseline (Left), N.1.5 (Middle) and B.1.5 (Right) for CelebA. We observe that the visual quality is comparable when replacing the last transpose convolutions with nearest and bilinear methods. FID scores for Baseline (Left), N.1.5 (Middle) and B.1.5 (Right) setups are 60.6, 48.69 and 52.18 respectively. (Measured using 50k real and generated samples.)

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Setup Code	DCGAN	LSGAN	WGAN-GP
Baseline	**88.6**	**73.26**	**60.6**
N.1.5	87.52	70.69	48.69
Z.1.5	69.14	60.29	47.73
B.1.5	84.65	78.66	52.18
N.1.7	90.8	73.09	60.11
Z.1.7	71.45	59.55	43.1
B.1.7	79.92	76.33	55.28
N.1.3	93.54	74.06	58.35
Z.1.3	65.46	61.45	56.91
B.1.3	76.04	81.97	58.55
N.3.5	73.63	78.31	55.47
Z.3.5	68.41	66.27	57.59
B.3.5	80.89	72.29	54.84
SR	99.2	86.16	60.81

Table 5. FID scores of GAN images trained on CelebA [29] dataset. We include the FID scores of Spectral Regularized GANs (indicated as SR) for comparison.

- FID scores of nearest and bilinear interpolation methods are comparable or better than the Baseline FID.
- Lower FID is better.

Extended Experiments

- We extend our experiments to LSUN dataset and confirm our findings.
- We also extend our analysis to Image-to-Image translation using StarGAN and confirm our findings.

All these additional results can be found in the paper.

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

“High frequency spectral decay discrepancies are not intrinsic for CNN-generated images. Therefore, we urge re-thinking in using such features for CNN generated image detection.”

Code and pre-trained models available

https://keshik6.github.io/Fourier-Discrepancies-CNN-Detection/

A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection

[Chandrasegaran, Tran, Cheung ; CVPR 2021 Oral]

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

CVPR Paper Session 6 : 23 June 6 pm – 8 pm SGT (Paper ID: 7067)

Q & A

References

- Tarik Dzanic, Karan Shah, & Freddie Witherden (2020). Fourier Spectrum Discrepancies in Deep Network Generated Images. *(NeurIPS)*.
- Durall, R., Keuper, M., & Keuper, J. (2020). Watch Your Up-Convolution: CNN Based Generative Deep Neural Networks Are Failing to Reproduce Spectral Distributions. *(CVPR)*.
- Wang, S.Y., Wang, O., Zhang, R., Owens, A., & Efros, A. (2020). CNN-Generated Images Are Surprisingly Easy to Spot... for Now. *(CVPR)*.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc..
- Tero Karras, Timo Aila, Samuli Laine, & Jaakko Lehtinen (2018). Progressive Growing of GANs for Improved Quality, Stability, and Variation. In *International Conference on Learning Representations*.

- https://towardsdatascience.com/a-new-way-to-look-at-gans-7c6b6e6e9737
- https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf
- https://scholar.harvard.edu/files/david-morin/files/waves_fourier.pdf

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Thank you

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

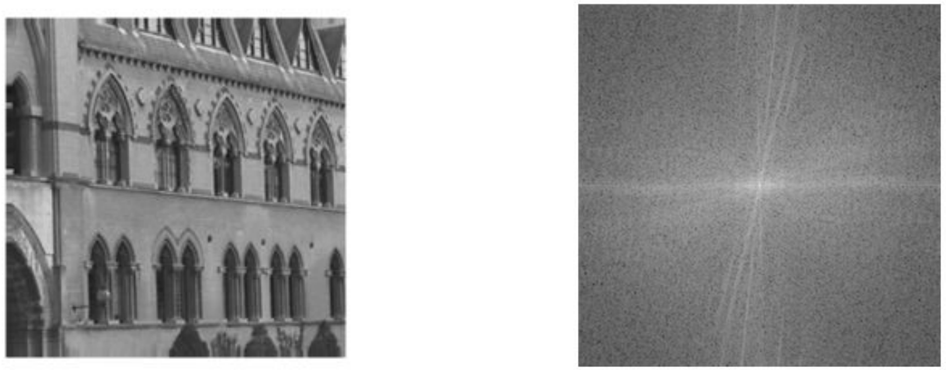
Appendix

Background : 2D Fourier Transform and Image Frequency

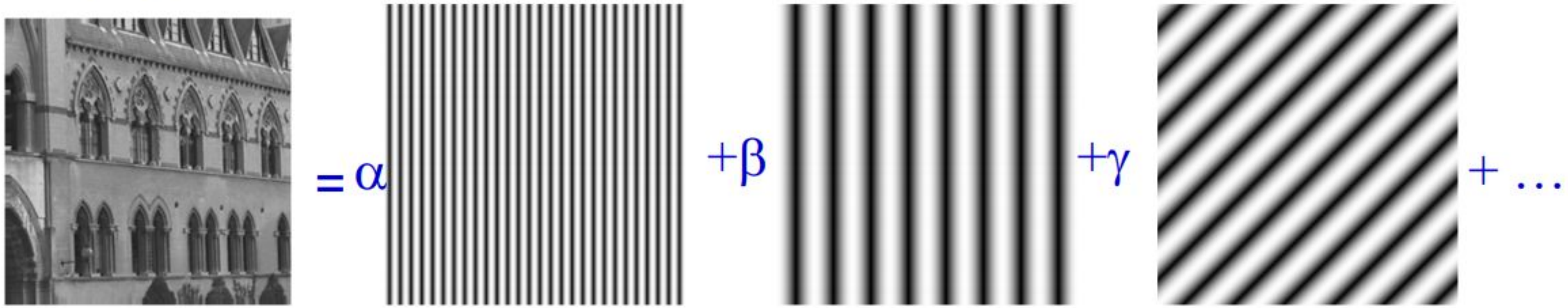
Definition

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}\,dx\,dy,$$
$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j2\pi(ux+vy)}\,du\,dv$$

$f(x,y)$ $|F(u,v)|$



$f(x,y)$



For further understanding refer to https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf