DSA5204 Project Report

SSD-GAN: Measuring the Realness in the Spatial and Spectral Domains

Group 19: Cai Anqi, Chen Xinyi, Guan Youyou, Li Yixuan, Yao Yuhan, Zhang Youyang, Zhao Tian Qi

1. Background

Generative Adversarial Networks (GANs) have gained much attention in recent years due to their ability to reproduce realistic images. There have been numerous refinements and improvements to the original GAN structure since its invention in 2014 [6]. GAN consists of two primary components: a generator and a discriminator. The generator is responsible for synthesising images while the discriminator aims to evaluate generated images' quality and distinguish them from real images.

It is observed that the discriminator in standard GAN lacks the ability to capture high frequencies in images which is caused by the downsampling layers in the architecture, thus leading to the lack of incentive for generators to learn the high frequency content [1]. The high frequency content of input images is removed during the downsampling process, due to low-pass filtering or the content itself being aliased [14]. This therefore causes a discrepancy in the spectrum which makes generated images less realistic. Liu et al. proposed an approach to ensure a neural network learns all the information of an image, by using discrete wavelet transformation (DWT) [12]. With a similar purpose, an improved GAN architecture is proposed to incorporate both the spatial and spectral domains of training images, by embedding a frequency-aware classifier in the discriminator. The new structure is called SSD-GAN as the discriminator captures both the spatial and spectral information.

2. Technical Approach

Our technical approach to reproduce SSD-GAN includes gathering appropriate datasets for training and preprocessing the data, which includes resizing the images and normalisation process. In the modelling part, we use three variants of GAN: SNGAN, SNGAN+REG and StyleGAN to embed the SSD structure and compare with the corresponding baseline models. Hinge loss is used as the loss function. The models are trained on a Tesla P100 GPU. SNGAN and SNGAN+REG are trained for 100 thousand steps with batch size 64, while StyleGAN is trained from 4*4*3 images to 32*32*3 images for 40 epochs with decreasing batch sizes from 64 to 16. Finally, for evaluation of model performances, we use Fréchet Inception Distance (FID), fast Fourier Transform (FFT) and power spectrum inspection as our metrics. Example code for StyleGAN, SNGAN, SNGAN+REG and their SSD versions using CIFAR10 can be found in this Github repository.

3. Dataset & Preprocessing

We have 2 types of datasets. One is the built-in datasets of Torchvision. We focus on CIFAR10 and STL10 in our project. Both datasets consist of more than 50,000 of coloured images and are the datasets used in the original paper. We wish to explore them for reproduction purposes.

We also source another dataset known as the Abstract Art Gallery dataset as an extension. It contains about 3000 images painted in abstract shapes and colours, which differ greatly from real-world pictures. The reason why we choose this dataset as an extension is that we consider it perfectly fit with SSD-GAN. It enhances evaluation of high-frequency learning because abstract images with richer high-frequency information can provide a more challenging and comprehensive test for the algorithm. In addition, applying SSD-GAN to a largely different dataset can showcase its full potential, as it will provide evidence of how well the algorithm can generate fine details and complex textures. Furthermore, we can identify limitations and areas for improvement by working with a more demanding dataset, which can lead to the development of new techniques or improvements to the existing algorithm.

For data preprocessing, we mainly focus on resizing and normalisation. Shapes need to be standardised across all images in order to feed into the GAN models. Here we choose to resize all the images to shape 32x32x3 for the balance between information loss due to downsampling and computational cost. After that, we normalise the pixel values of the images. By centering the pixel values around zero and scaling them into the range [-1, 1], we ensure that all the images contribute equally to the model and better maintain stability. For the Art Gallery extension data, an additional processing step is required. We convert it to the same form as CIFAR10, that is, a matrix representing image data followed by an integer representing categories, to ensure the universality of the model, even though the categories of images play almost no role in GANs.

4. Models

4.1. The SSD-GAN Structure

In order to make the model aware of the spectral feature, the algorithm suggests modifying the discriminator by adding spectral loss. Therefore, the model will introduce incentive for the generator to create images that preserve the spectral feature. In normal GAN algorithms, the adversarial loss is computed by

$$L_D = -\mathbb{E}_{x\sim p_{data}(x)}[log(D(x)] --\mathbb{E}_{x\sim p_{data}(x)}[log(1 - D(x)]$$

where D represents the discriminator, and $D(x)$ measures the realness of data x. Here the discriminator only measures spatial realness. Spectral realness is ignored in the model.

In order to address the gap, the SSD method introduce a new discriminator C for spectral feature, and hence develop an overall discriminator and loss function:

$$D_{ss}(x) = \lambda D(x) + (1 - \lambda)C(\phi(x)),\ \lambda\in(0,\ 1)$$

$$L_{adv} = -\mathbb{E}_{x\sim p_{data}(x)}[log(D_{ss}(x)] -- \mathbb{E}_{x\sim p_{data}(x)}[log(1 - D_{ss}(x)]$$

$\phi(x)$ in the equation above is a vector that represents the spectral feature of x. It is computed from the Fourier Transformation which is a technique that captures the image spectrum. The spectral vector is computed in 3 steps:

1. Compute discrete Fourier Transform of the image f of shape $M \times N$:

   $$F(k, l) = \sum_{m=0}^{M}\sum_{n=0}^{N} f(m, n)e^{-2\pi i\frac{km}{M}}e^{-2\pi i\frac{ln}{N}}$$

2. Convert the Fourier Transform to polar representation:

   $$F(r, \theta) = F(k, l), r = \sqrt{k^2 + l^2}, \theta = atan2(k, l)$$

   Where $atan2(\cdot, \cdot)$ compute the angle between its input

3. Take azimuthal average of the polar Fourier Transform over θ:

   $$v(r) = \frac{1}{2\pi}\int_{0}^{2\pi} |F(r, \theta)|\, d\theta$$

$v(r)$ is effective to highlight the difference between the spectral characteristics of real and deep network generated images [4][5]. Hence, we use the vector v from grayscale image x to denote $\phi(x)$, since grayscale images preserve spectral features.

4.2. Modelling

We have applied the SSD technique on 3 different GAN algorithms to examine its effect, namely SNGAN[11], SNGAN with regularisation, and StyleGAN. According to the original paper on SSD-GAN, there is another discriminator introduced for spectral features of the images, and the overall loss is computed by combining the 2 discriminators. In order to examine whether the same spectral discriminator has a consistent effect on different GAN models, the spectral norm is applied to represent the loss of spectral features. The parameter $\boldsymbol{\lambda}$ is set to 0.5 so that the spectral loss and spatial loss has the same weightage.

4.3. Extension

In addition to the original paper, we have established extensions by modifying the SSD algorithm. In the original algorithm, the spectral feature is represented by taking azimuthal average of processed Fourier Transformations of the input images, which returns a vector representation of spectrum. This intrigues another idea: represent spectrum using the Fourier Transform directly, instead of using the vector. Our experiment shows that the generated images have slightly better performance in preserving spectral features, but with much higher time cost. Hence, using the processed spectral vector $\phi(x)$ is more efficient. Details of the experiment result can be found in Section 5 and Appendix A.

5. Results & Evaluation

5.1. Result Comparison

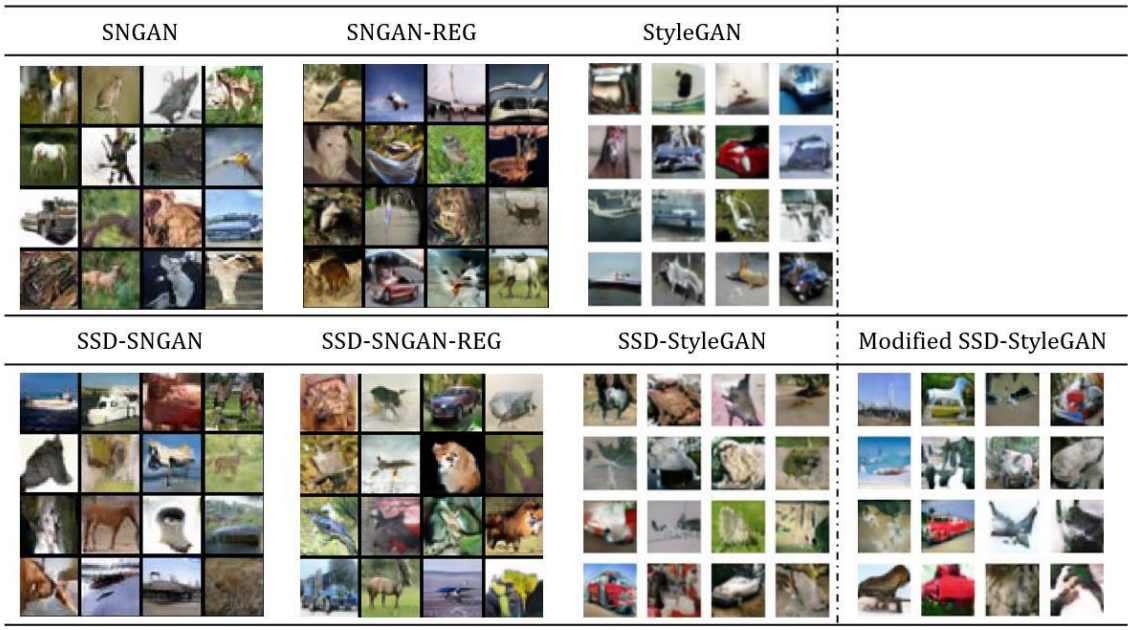

Fig. 1: Snapshots of output images by various GAN models trained on CIFAR10 dataset. Typically a high-frequency image contains many sharp edges, fine details and rapid changes in colour or brightness, while a low-frequency image is smoother with less details.

By inspection, images on the second row change more in colour and brightness and have more details as compared to images on the first row. This suggests that SSD encourages the generator to learn high-frequency components of the images. Moreover, when we compare the output images between SSD-StyleGAN and modified SSD-StyleGAN, we could also observe a little improvement in terms of the high-frequency component.

5.2. Frechet Inception Distance

Frechet Inception Distance(FID)[7] is a common metric used to evaluate the quality of images generated by GANs. It calculates the Fréchet distance between feature representations of real and generated images using an inception network, with lower FID score indicating better image quality.

Although FID score does not directly measure the high-frequency components of images, it can indirectly evaluate them by comparing the similarity between the generated and real images in terms of their high-frequency components as the ultimate goal of SSD-GAN is to generate more realistic images. Consequently, if a GAN model generates images that have high-frequency components similar to real images, the FID score will show a lower value to reflect this.

Table 1 contains the FID scores for CIFAR-10 and Gallery dataset using different models, from which we can see the FID score of SSD-GAN models has decreased as compared to the baseline GAN models. However, there exist cases where SSD-GAN shows a higher FID score, which could

be due to different discriminators applied to the spectral feature, and the training dataset as well.

Table 2 compares the FID scores between SSD-GAN models and modified SSD-GAN models. A slight improvement on FID indicates that using the Fourier Transform directly is effective.

Model	CIFAR-10	Abstract Gallery
StyleGAN	174.55	177.50
SSD-StyleGAN	166.52	185.42
SNGAN	73.04	725.24
SSD-SNGAN	69.56	689.39
SNGAN+REG	62.23	683.77
SSD-SNGAN+REG	61.35	654.22

Table1: FID scores on CIFAR-10 & Gallery dataset

Model	SSD	Modified SSD
StyleGAN	166.52	153.47
SNGAN	69.56	66.89
SNGAN_reg	61.35	59.41

Table2: FID scores on Modified SSD

5.3. Fast Fourier Transform

The frequency magnitude of the Fast Fourier Transform(FFT) for the output images allows us to directly examine the strongness of high-frequency components of the images. The larger the magnitude, the stronger the component. We randomly select 50 output images of each GAN model, convert them into grey scales and compute the magnitude of 2D FFT. An average of the magnitudes is taken across all the images for different models correspondingly.

Fig.2 shows the average FFT magnitude plots of 50 output images from SNGAN+REG and SSD-SNGAN+REG. The middle of the frequency spectrum corresponds to low-frequency components, and the high-frequency components are represented by the corners, which is our interest in this research. The colour intensity represents the amplitude of each frequency component. A higher amplitude component will result in a brighter pixel. Based on the observation of 4 corners of Fig 2(b), Fig 2(d) and Fig 2(e), we observe an increase in the intensity of the pixel colours across them. This demonstrates that SSD and modified SSD does produce images with stronger high-frequency content and is effective to address the high-frequency missing problem.

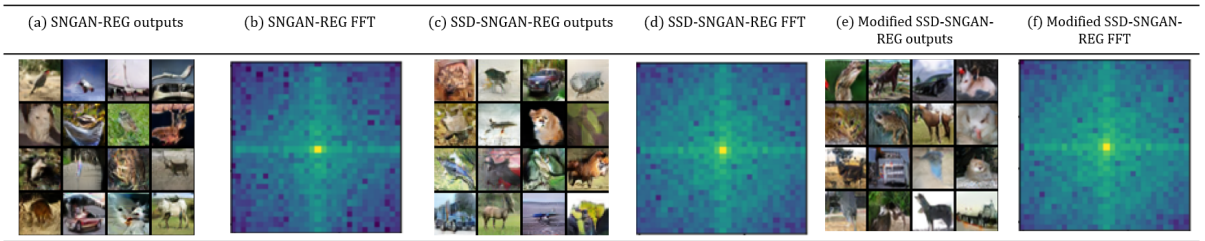

Fig. 2: FFT of images generated by SNGAN+REG, SSD-SNGAN+REG & Modified SSD-SNGAN+REG

5.4. Power Spectrum

Computed by taking the Fourier transform of a signal and then squaring the magnitude of the resulting complex values, the power spectrum reflects how much energy is contained in different frequency components of a signal. The resulting power spectrum can be visualised as a plot of power versus frequency.

Following prior works[2][3], we also employ the azimuthal integration over the Fourier power spectrum to analyse the spectral properties of generated images. We plot for models without and with the spectral classifier embedded. The left plot in Fig. 3 shows the statistic (mean) after azimuthal integration over the power-spectrum of real and GAN generated images over CIFAR-10 dataset. It obviously indicates that the current GANs fail to reproduce the spectral distributions with higher frequency and will introduce heavy spectral distortions into generated images. However, after we add a spectral classifier into our model and do the same experiment, as the plot on the right shows, SSD-GANs display good spectral properties indicated by a response more similar to real data even under high frequency. This verifies the effectiveness of our improved model.

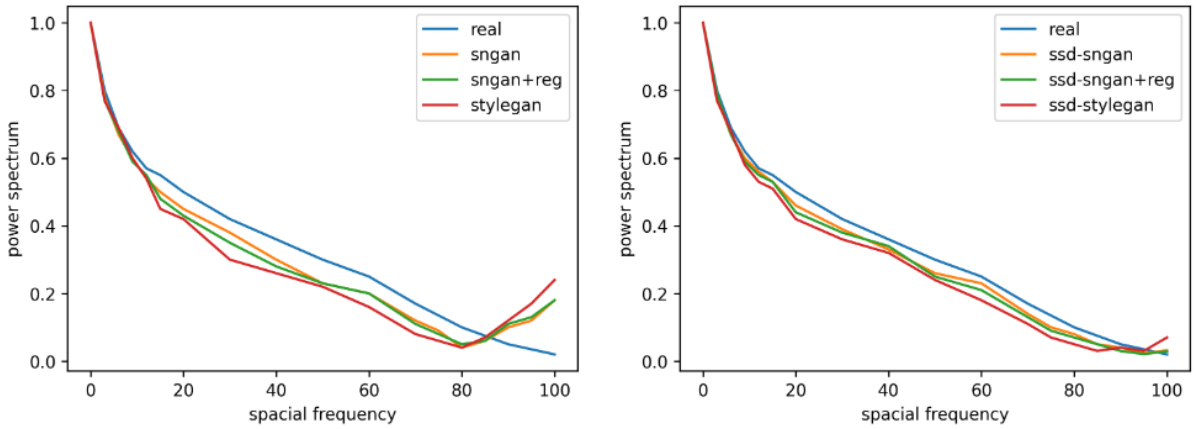

Fig. 3: Power spectrum results of real and generated images on CIFAR-10

6. Discussion

6.1. Conclusion of reproduction

By reproducing the SSD-GAN structure, we generate images for comparison between different models. By comparing FID scores, FFT and power spectrum of generated images, SSD-GAN is proven to be able to encourage the generator to learn high-frequency content of images and this finding can be generalised to different datasets. Our extension idea could further address the high-frequency missing problem.

6.2. Future Work

There remain things that can be done in the future to improve the reproduction and extension work. For example, applying SSD-GAN to other GAN models such as CycleGAN, and training SSD-GAN with various datasets.

References

[1] Chen, Y., Li, G., Jin, C., Liu, S., & Li, T. (2020). SSD-GAN: Measuring the Realness in the Spatial and Spectral Domains. AAAI. arXiv. 2012.05535

[2] Durall, R., Keuper, M., & Keuper, J. (2020). Watch your Up-Convolution: CNN Based Generative Deep Neural Networks are Failing to Reproduce Spectral Distributions. *Computer Vision and Pattern Recognition*. arXiv. 2003.01826

[3] Durall, R., Keuper, M., Pfreundt, F.-J., & Keuper, J. (2019). Unmasking DeepFakes with simple Features. *Machine Learning*. arXiv. 1911.00686

[4] Dzanic, T., Shah, K., & Witherden, F. (2020). Fourier Spectrum Discrepancies in Deep Network Generated Images. *Neural Information Processing Systems*, *33*.

[5] Frank, J., Eisenhofer, T., Sch¨onherr, L., Fischer, A., Kolossa, D., & Holz, T. (2020). Leveraging frequency analysis for deep fake image recognition. International Conference on Machine Learning. arXiv. 2003.08685

[6] Goodfellow, I. J., Jean Pouget-Abadie, J., Mizra, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *NIPS*, *27*. arXiv. 1406.2661

[7] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Advances in Neural Information Processing Systems, 17(NIPS), 6629-6640. arXiv. 1706.08500

[8] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive Growing of GANs for Improved Quality, Stability, and Variation. International Conference on Learning Representations. arXiv. 1710.10196

[9] Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. The Conference on Computer Vision and Pattern Recognition. arXiv. 1812.04948

[10] Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. International Conference on Learning Representations. arXiv. 1412.6980

[11] Lee, k. S., & Town, C. (2020). Mimicry: Towards the Reproducibility of GAN Research. CVPR. arXiv. 2005.02494

[12] Liu, P., Zhang, H., Zhang, K., Lin, L., & Zuo, W. (2018). Multi-level Wavelet-CNN for Image Restoration. *CVPR*. arXiv. 1805.07071

[13] Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral Normalization for Generative Adversarial Networks. International Conference on Learning Representations. arXiv. 1802.05957

[14] Zhang, R. (2019). Making Convolutional Networks Shift-Invariant Again. *ICML*, *97*. arXiv. 1904.11486

Appendix A

Following is the comparison of GAN algorithms using normal SSD method and modified SSD method on the CIFAR10 dataset, in terms of result performance and time cost:

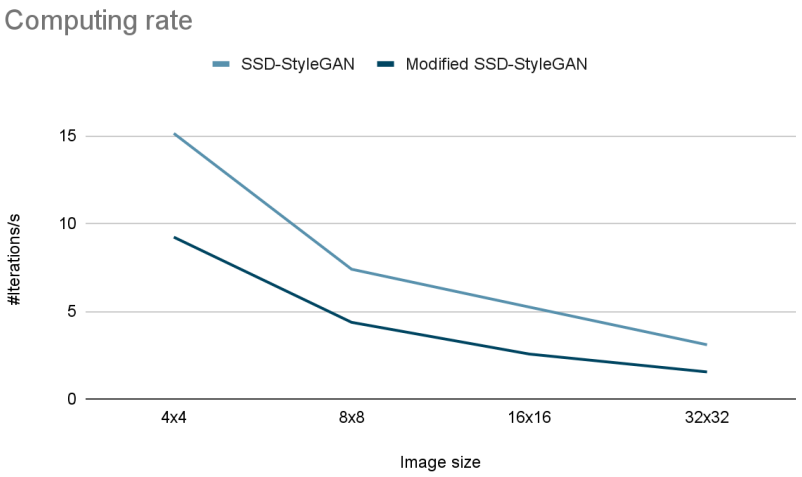

Fig. 4: Computing rate of SSD-StyleGAN and modified SSD-StyleGAN

Appendix B

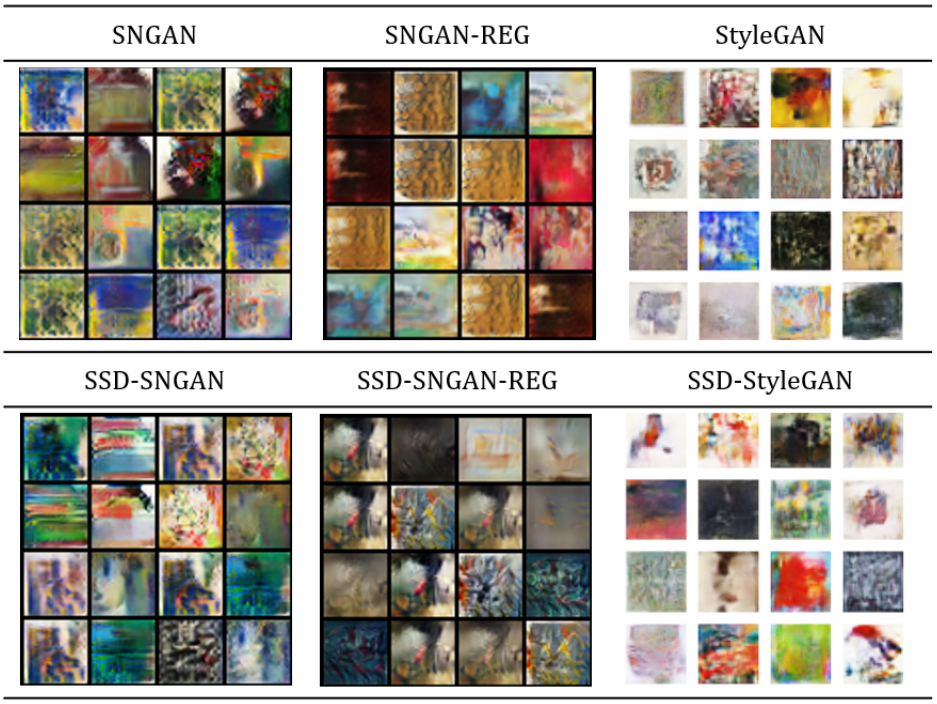

Fig. 5: Snapshots of output images by various GAN models trained on Gallery dataset

Appendix C

Roles of group members

Data Collection and Preprocessing	Li Yixuan, Guan Youyou
Implementation of Algorithm	Zhang Youyang, Zhao Tian Qi, Chen Xinyi
Statistics Collection and Evaluation	Cai Anqi, Yao Yuhan