

Spectral Regularization for Combating Mode Collapse in GANs

Appendix

1. Proof of Corollary 1

Corollary 1. Let P_r and P_g be two distributions in \mathcal{X} , a compact metric space. A linear and 1-Lipschitz constrained function $f^* = Wx$, is the optimal solution of $\max_{\|f\|_{Lip} \leq 1} E_{x \sim P_r}[f(x)] - E_{x \sim P_g}[f(x)]$. Then all the singular values of the weight matrix W are 1.

Proof: Proposition 1 in [Gulrajani et al., Improved Training of Wasserstein GANs, Advances in Neural Information Processing Systems, 5769-5779, 2017.](#) has proven that the optimal solution to 1-Lipschitz discriminator function f^* has gradient norm 1 almost everywhere. In other words, f^* is obtained at the upper bound of 1-Lipschitz constraint: $\|f(x_1) - f(x_2)\| \leq \|x_1 - x_2\|$.

Because f is a linear function: $f(x) = Wx$. The 1-Lipschitz constraint for f can be expressed as:

$$\|Wx\| \leq \|x\| \quad (1)$$

Equation 1 is equivalent to:

$$\|Wx\|^2 \leq \|x\|^2 \quad (2)$$

and,

$$\|Wx\|^2 = x^T W^T W x = x^T V \Sigma V^T x \quad (3)$$

where columns of V , $[v_1, \dots, v_n]$ are eigenvectors of $W^T W$, and diagonal entries of diagonal matrix Σ are eigenvalues of $W^T W$.

Taking $y = V^T x$, then

$$\|Wx\|^2 = y^T \Sigma y = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \quad (4)$$

where λ_i is the i -th eigenvalue, and y_i is the i -th element of y .

Because $W^T W$ is symmetric, $V^T = V^{-1}$, then

$$\|y\|^2 = y^T y = x^T V V^T x = x^T x = \|x\|^2 \quad (5)$$

Finally, $\|Wx\|^2 \leq \|x\|^2$ is equivalent to $\lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \leq y_1^2 + \dots + y_n^2$. We can see that the upper bound of 1-Lipschitz constraint can be obtained only when all eigenvalues of $W^T W$ are 1. In other words, all the singular values of W are 1.

2. Architecture and Optimization Settings

In this paper, we employ SN-GAN architecture, which is illustrated in Figure 1. The weight in the convolutional layer is in the format $[out, in, h, w]$, where out is the output channel, in represents the input channel, h and w are kernel sizes. Particularly, there are 10 convolutional layers ($layer_0 \sim layer_9$) in discriminator network, and CH in Figure 1(b) corresponds to channel size of discriminator function in main text, where extensive experiments are conducted with different settings of CH. All the experiments are conducted based on the following architecture. Image generation on STL-10 shares the same architecture with that on CIFAR-10. Thus, images in STL-10 are compressed to 32×32 pixels, identical to the resolution of images in CIFAR-10. The purpose is to evaluate how different data affect mode collapse and spectral distribution, regardless of the effect of architecture.

The optimization settings follow SN-GANs. To be specific, the learning rate is taken as 0.0002, the number of updates of the discriminator per one update of the generator n_{critic} is 5, the batch size is taken as 64, and Adam optimizer is used as the optimization with the first and second order momentum parameters as 0 and 0.9, respectively.

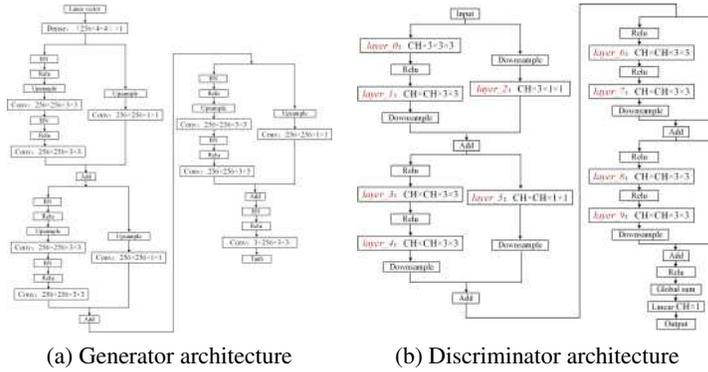


Figure 1. Architecture of GANs.

3. Spectral Distributions

In Figure 2 ~ Figure 6, we show the spectral distributions for different settings in $layer_9$. We can see that spectral collapse and mode collapse always go side by side. In Figure 7 ~ Figure 11, we show the spectral distributions of each layer except $layer_2$ and $layer_5$. Because $layer_2$ and $layer_5$ act as the role of skip connection, and the intense correlation between mode collapse and spectral distortion is not observed in $layer_2$ and $layer_5$.

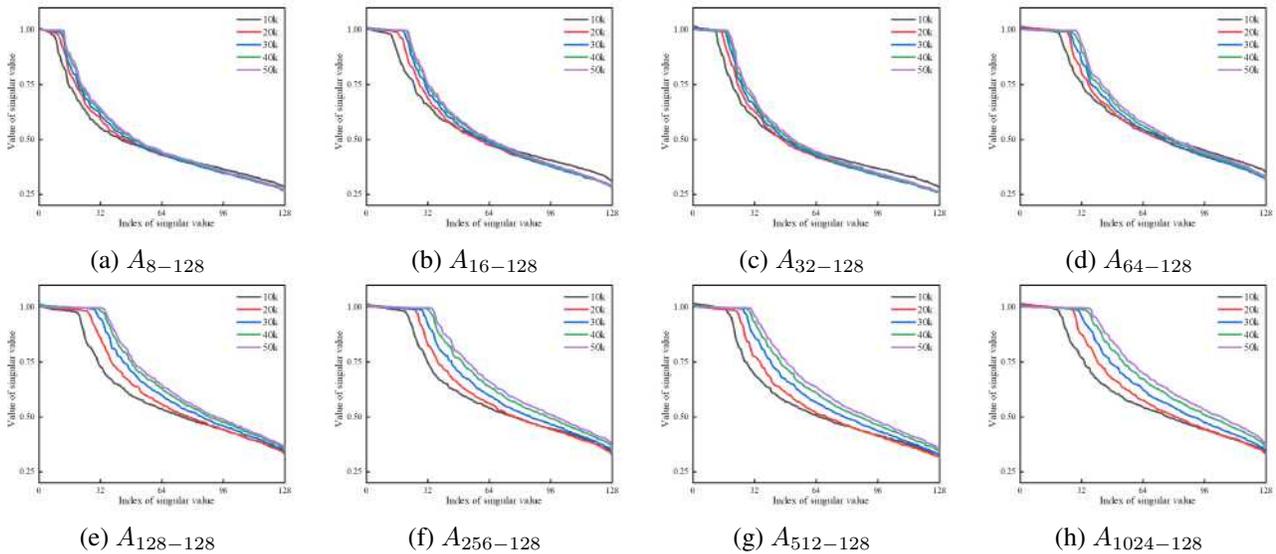


Figure 2. Spectral distributions in $layer_9$ for settings in group A. No mode collapse and spectral collapse is observed in group A.

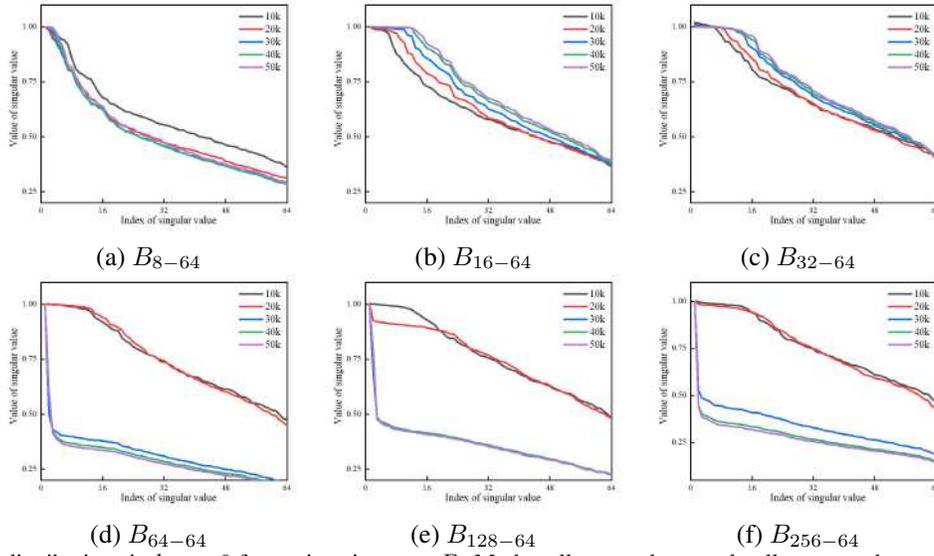


Figure 3. Spectral distributions in *layer_9* for settings in group *B*. Mode collapse and spectral collapse are observed in setting B_{64-64} , B_{128-64} and B_{256-64} .

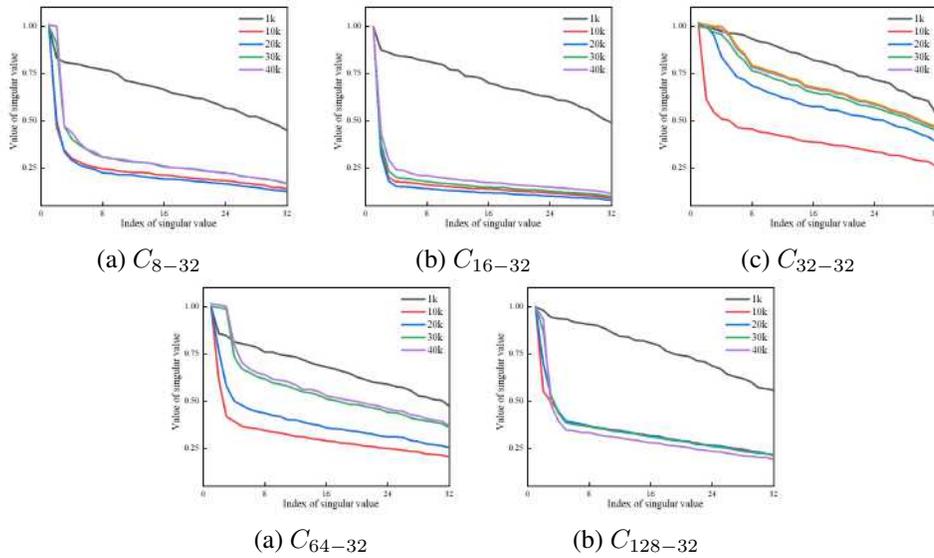


Figure 4. Spectral distributions in *layer_9* for settings in group *C*. Mode collapse and spectral collapse are observed in all settings of group *C*.

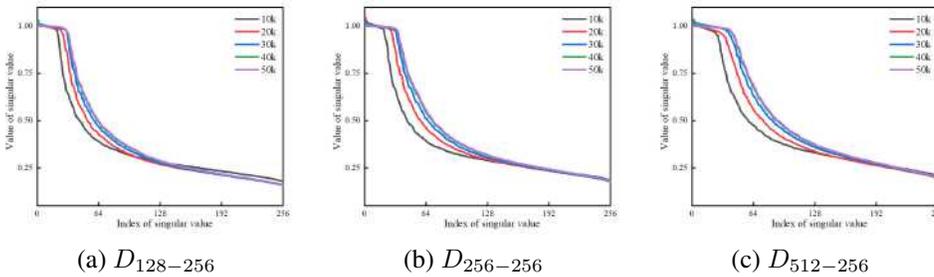


Figure 5. Spectral distributions in *layer_9* for settings in group *D*. No mode collapse and spectral collapse is observed in group *D*.

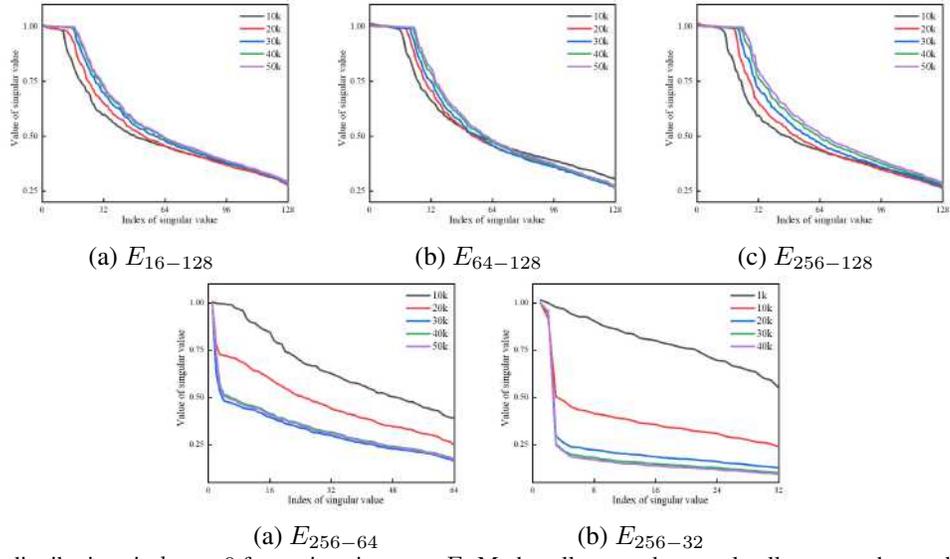


Figure 6. Spectral distributions in *layer_9* for settings in group *E*. Mode collapse and spectral collapse are observed in setting E_{256-64} and E_{256-32} .

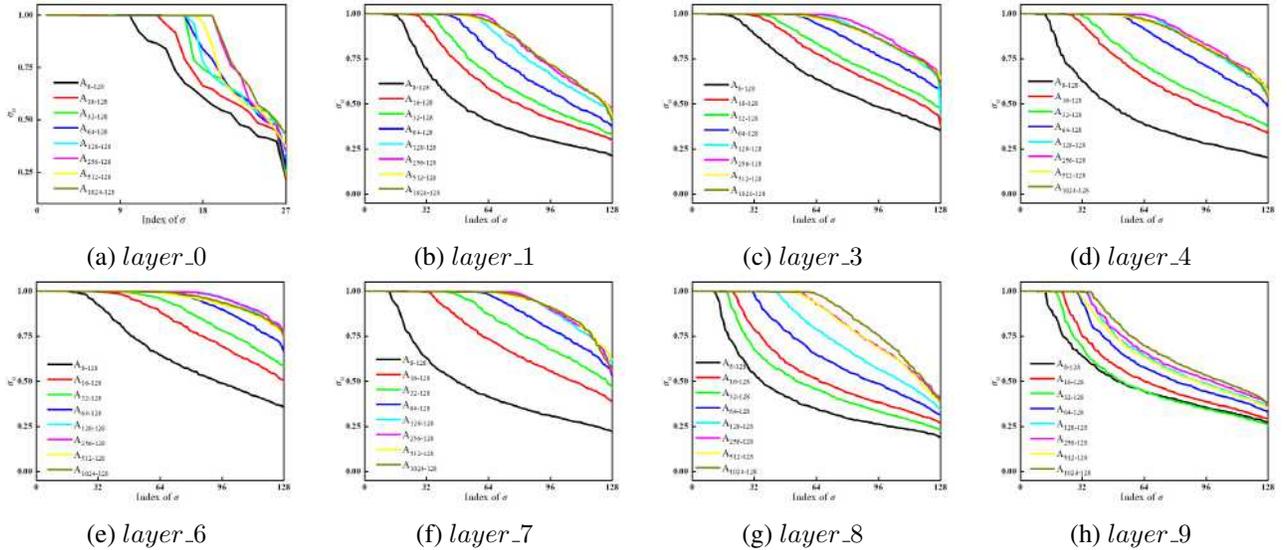


Figure 7. Spectral distributions (after 50k iterations) in each layer for settings in group *A*. No mode collapse and spectral collapse is observed in group *A*.

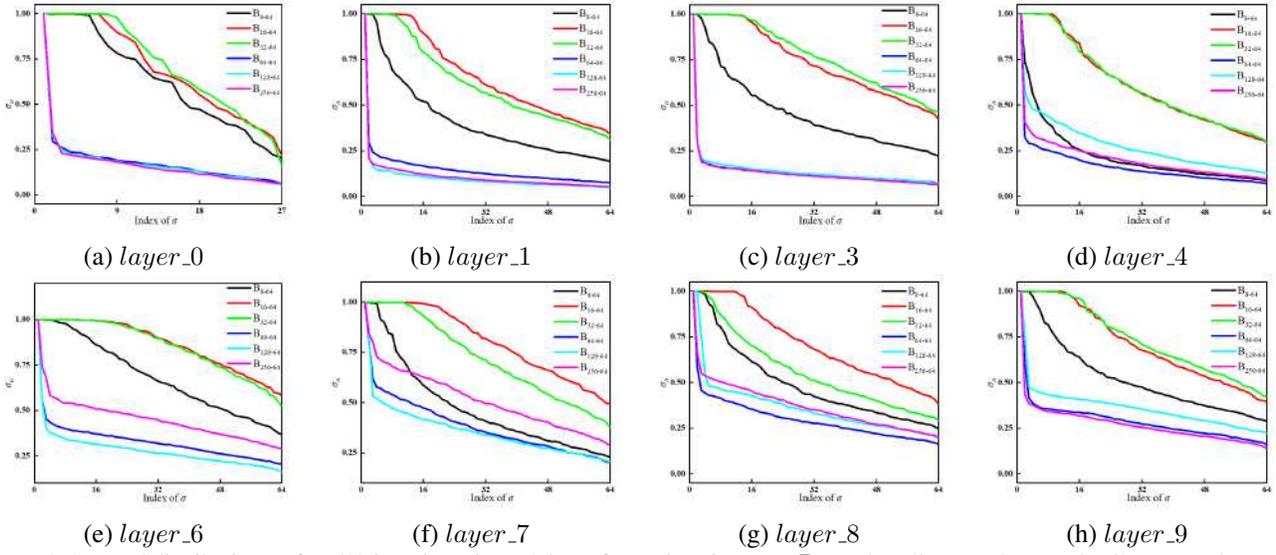


Figure 8. Spectral distributions (after 50k iterations) in each layer for settings in group B . Mode collapse and spectral collapse are observed in setting B_{64-64} , B_{128-64} and B_{256-64} .

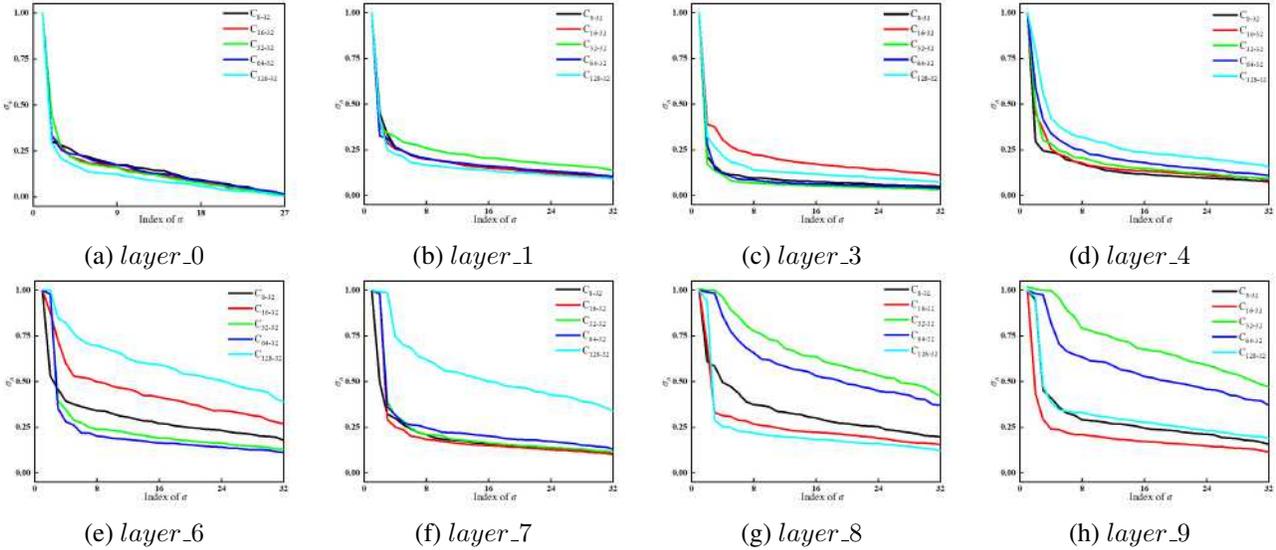


Figure 9. Spectral distributions (after 50k iterations) in each layer for settings in group C . Mode collapse and spectral collapse are observed in all settings of group C .

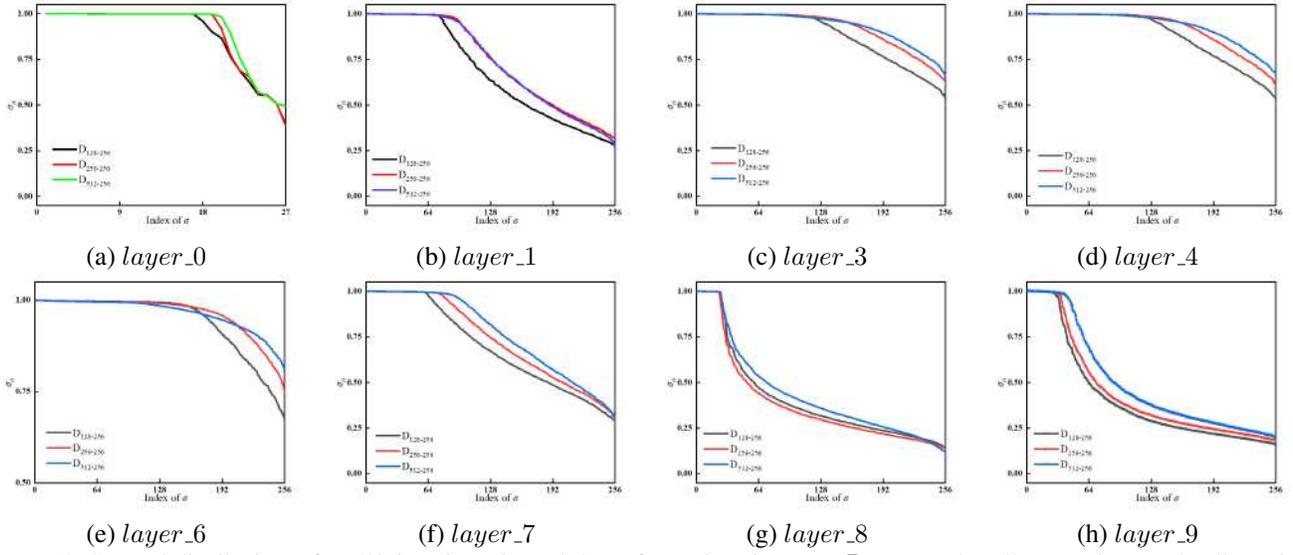


Figure 10. Spectral distribution (after 50k iterations) in each layer for settings in group D . No mode collapse and spectral collapse is observed in group D .

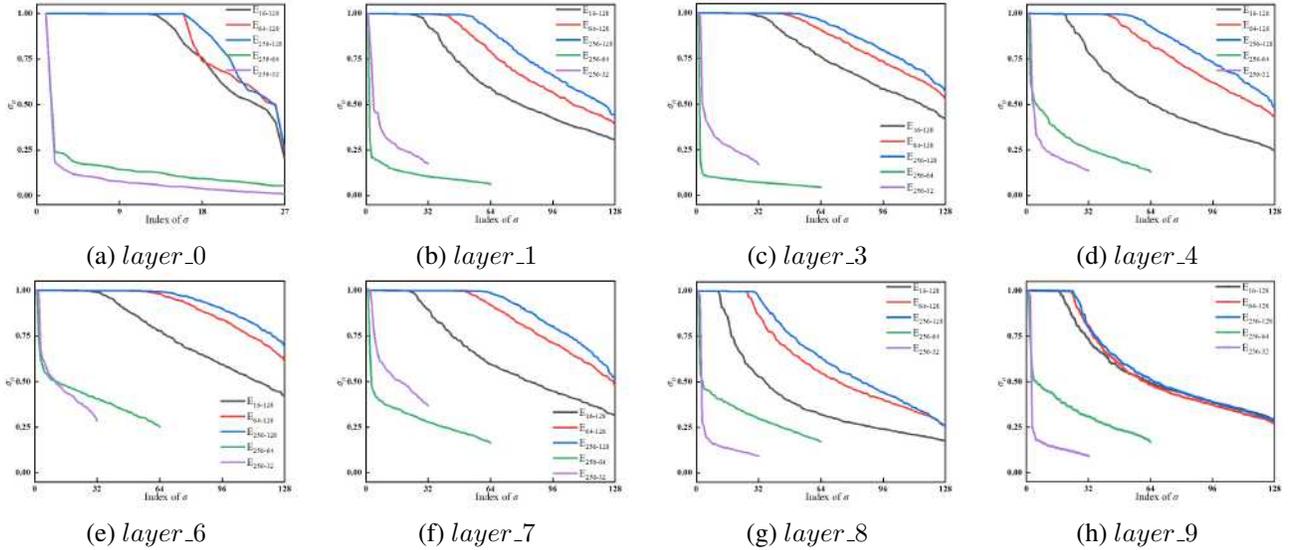


Figure 11. Spectral distribution (after 50k iterations) in each layer for settings in group E . Mode collapse and spectral collapse are observed in setting E_{256-64} and E_{256-32} .

4. Statistics of $D(x)$ and Discriminator Objective L_D

In the main text, we primarily show the statistics of $D(x)$ and discriminator objective L_D for setting $A_{128-128}$ and B_{128-64} . In Figure 12 ~ Figure 13, we show the mean and variance of $D(x)$ and L_D . To be specific, we feed the discriminator function with generated data and data in the training set, and obtain the output of the discriminator objective, then we calculate its mean and variance, finally we show their variation with i in Figure 12. To investigate the performance of discriminator on test set, we monitor $D(x)|_{x \sim q_{train}}$ and $D(x)|_{x \sim q_{test}}$, where q_{train} , q_{test} represent the training and test set, respectively. Then, we calculate the mean and variance, and show the variation with i in Figure 13.

In Figure 12, we can see that discriminator objective has a decreasing tendency with the increase of i . As we can see in Figure 13, $D(x)|_{x \sim q_{train}}$ and $D(x)|_{x \sim q_{test}}$ diverge when i is excessively large. Thus, excessively increasing i potentially leads to over-fitting, especially when i is taken as N . As we can see in Figure 13 (d) ~ Figure 13 (f), $D(x)|_{x \sim q_{train}}$ and $D(x)|_{x \sim q_{test}}$ agree well, indicating that no over-fitting is observed in group B .

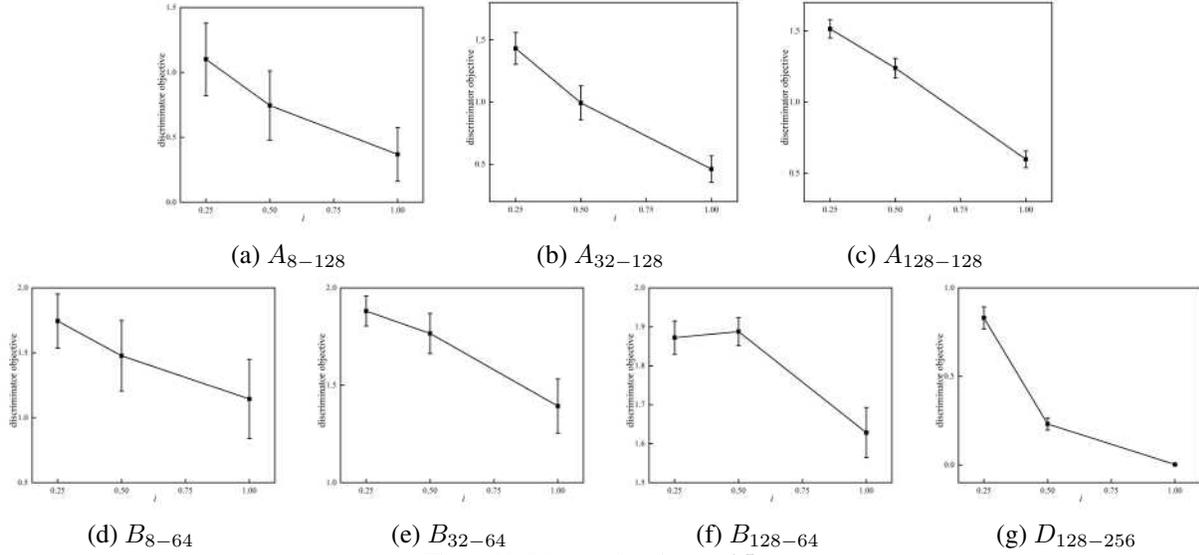


Figure 12. Mean and variance of L_D .

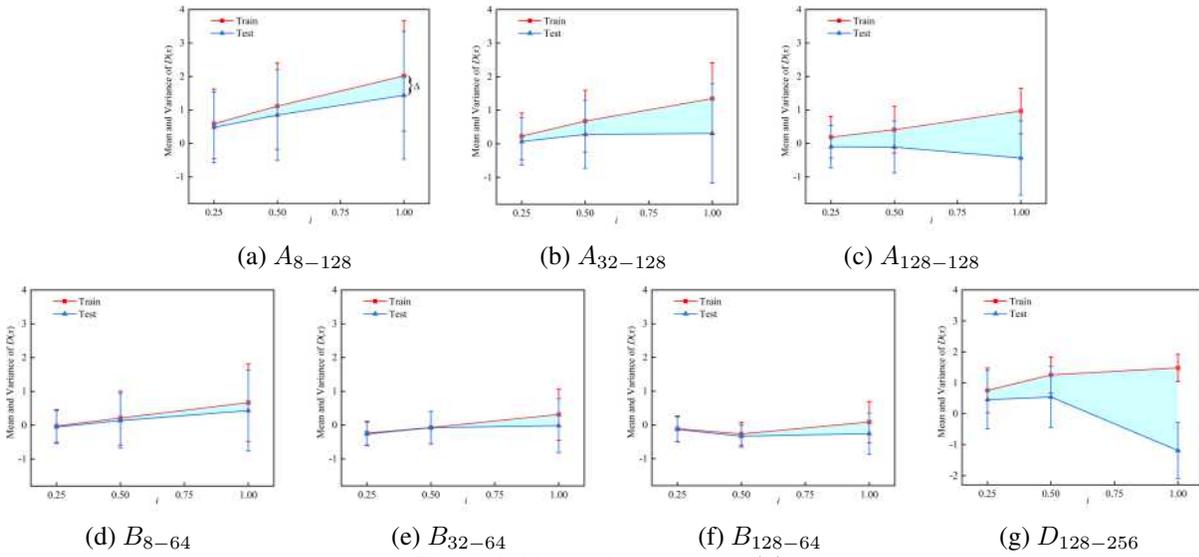


Figure 13. Mean and variance of $D(x)$.

5. Synthetic Images

We show some examples generated by SN-GANs and SR-GANs in Figure 14 ~ Figure 22

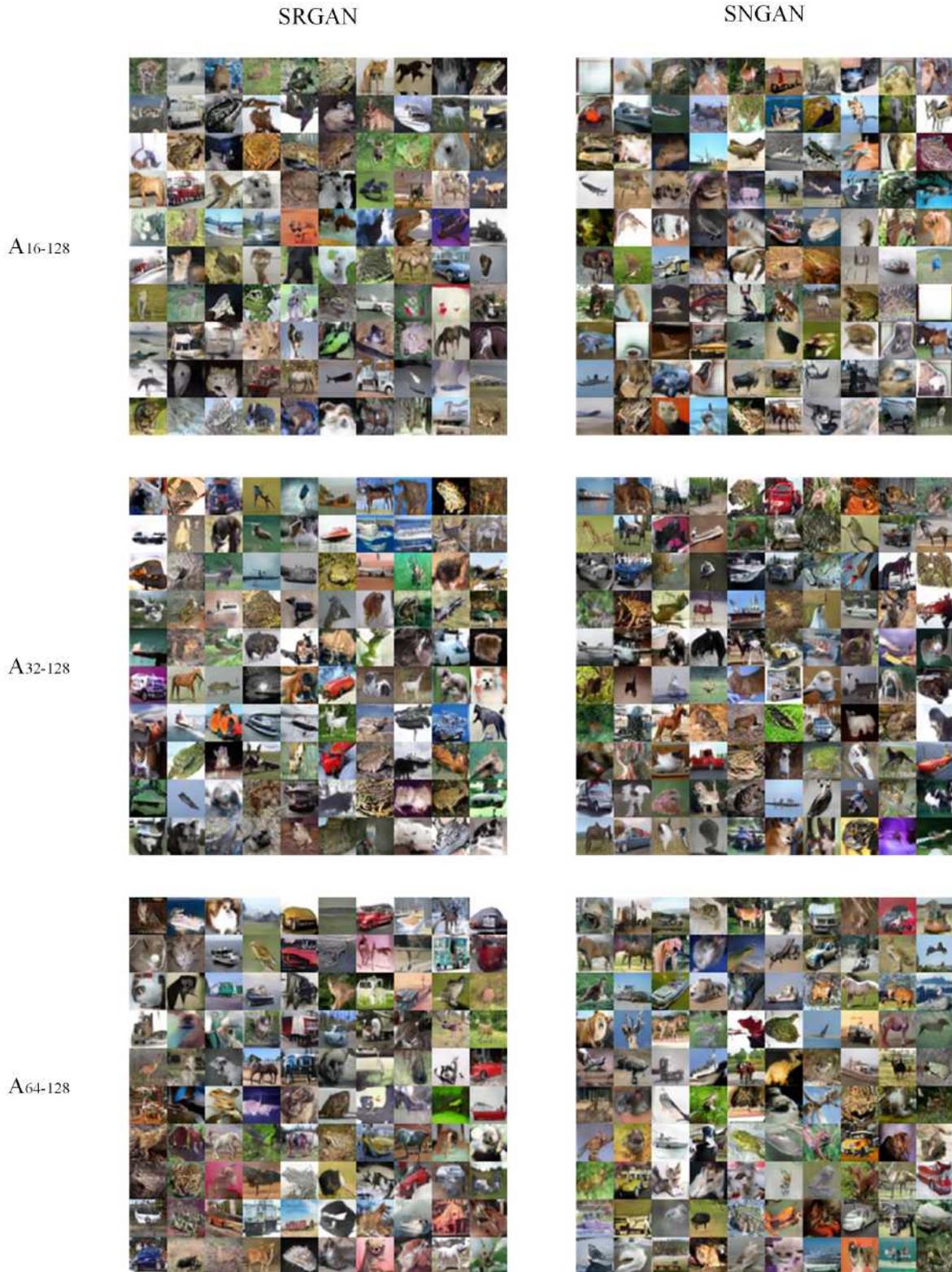


Figure 14. Synthetic images.

SRGAN

SNGAN

A₁₂₈₋₁₂₈



A₂₅₆₋₁₂₈



A₅₁₂₋₁₂₈

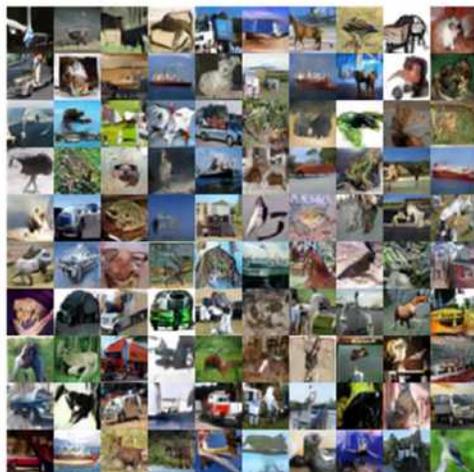


Figure 15. Synthetic images.

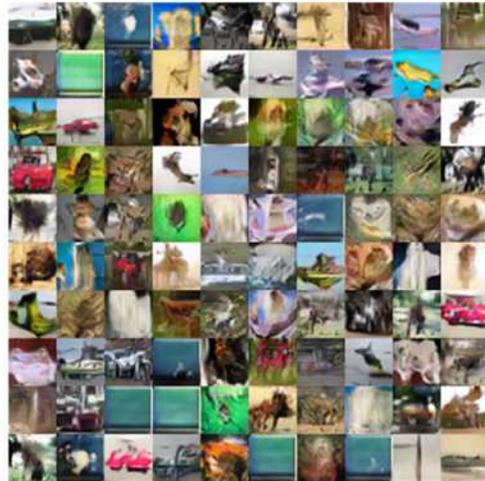
SRGAN

SNGAN

A1024-128



B8-64



B16-64



Figure 16. Synthetic images.

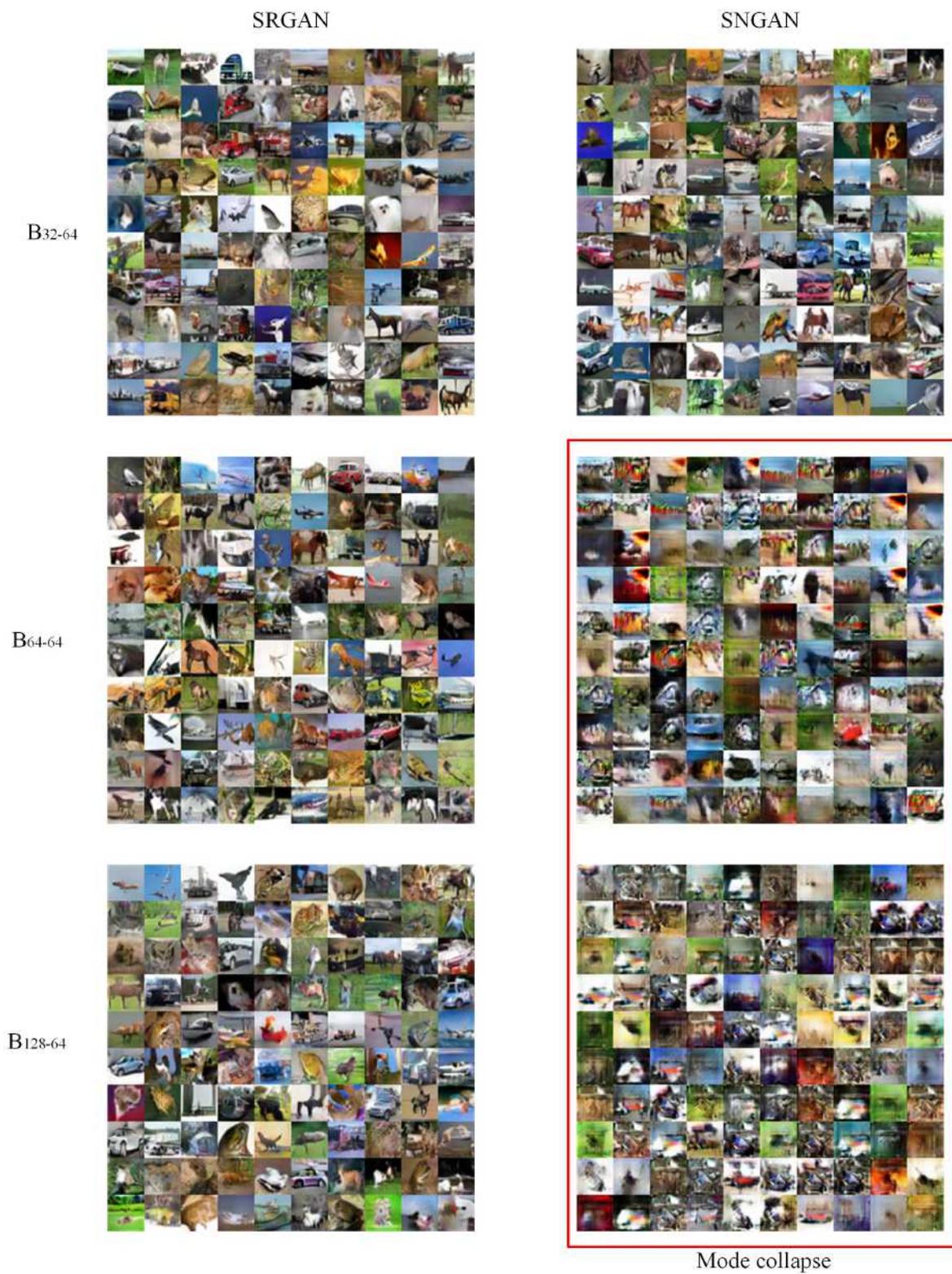


Figure 17. Synthetic images.

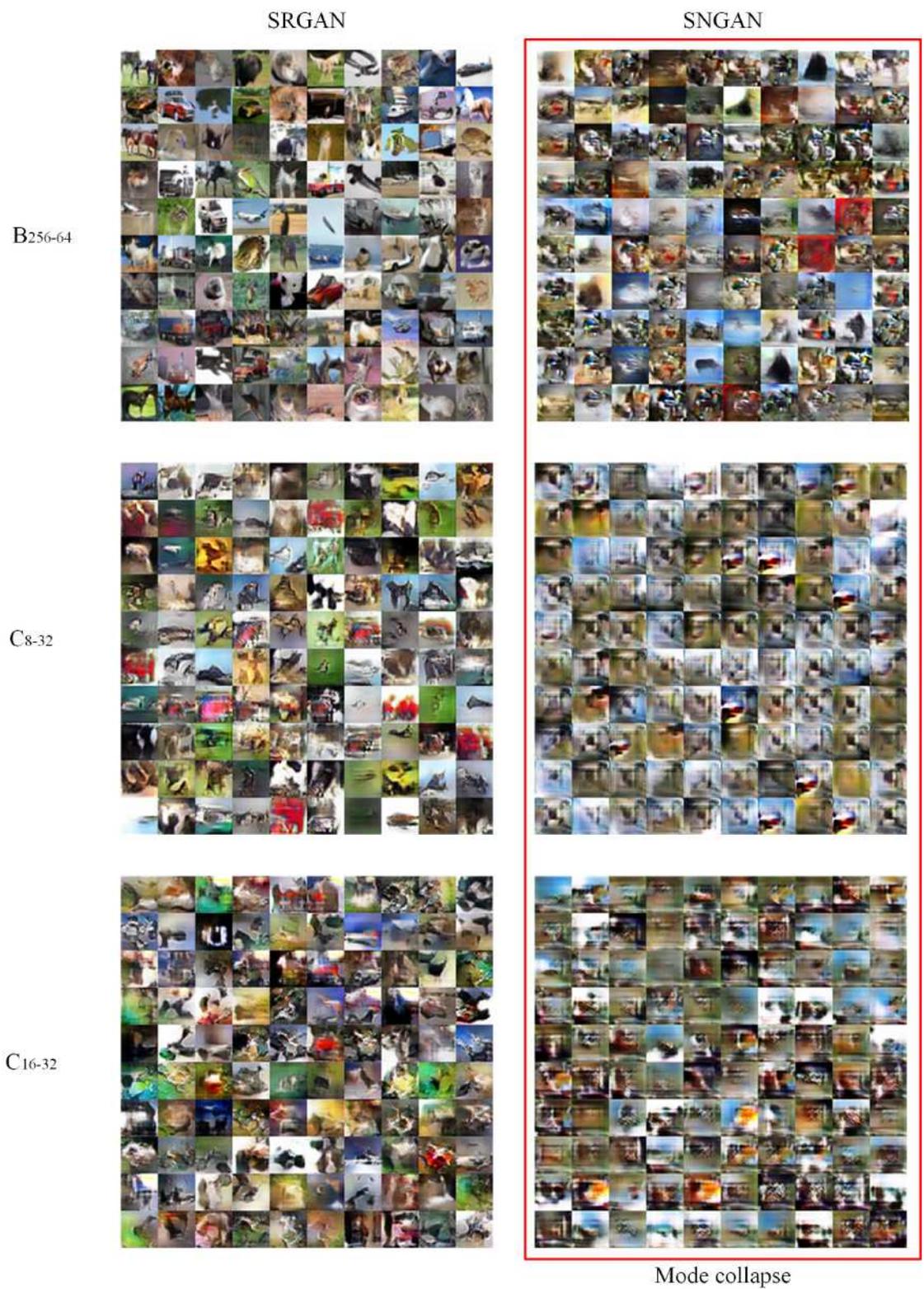


Figure 18. Synthetic images.

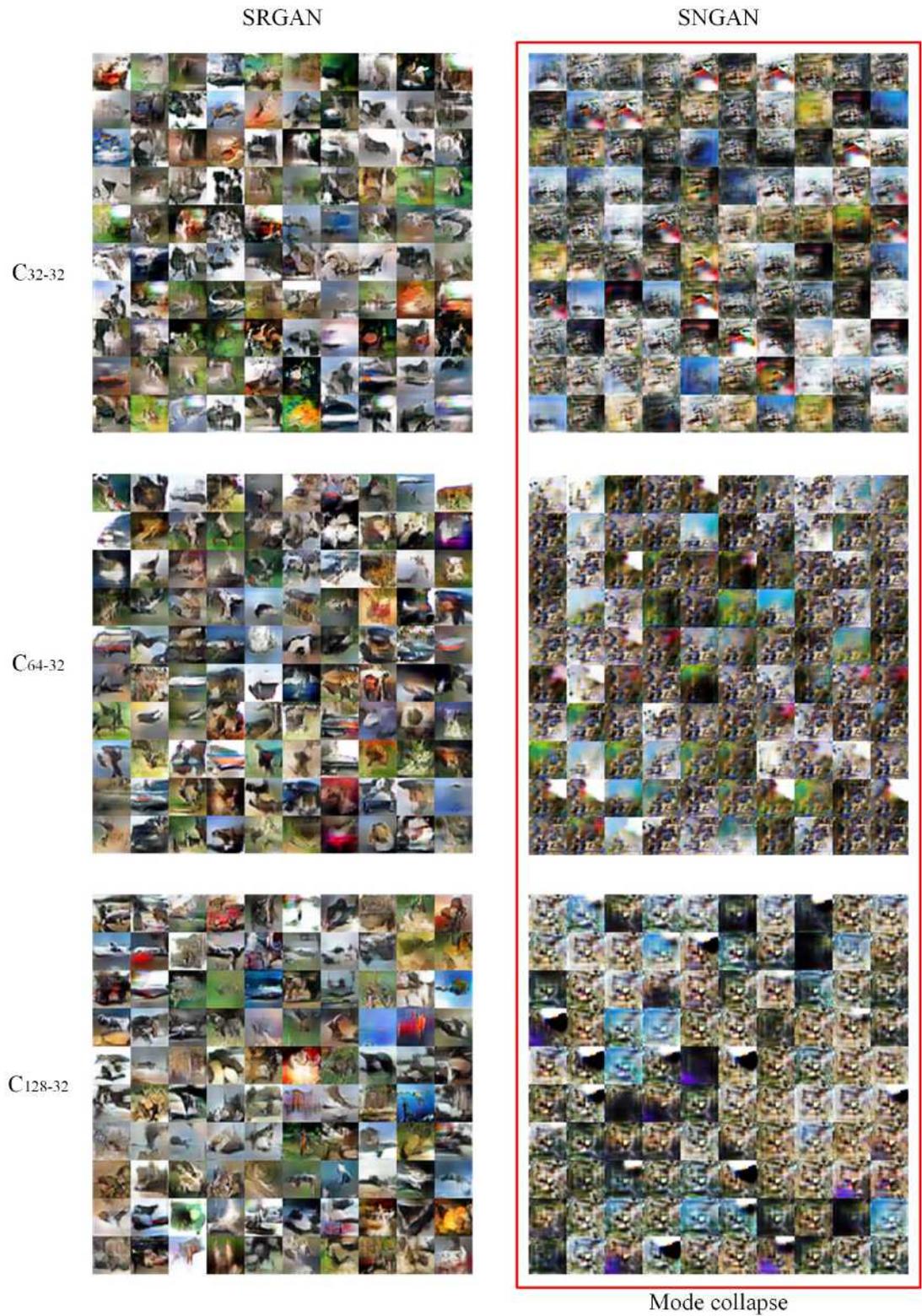


Figure 19. Synthetic images.

SRGAN

SNGAN

D128-128



D256-128



D512-128

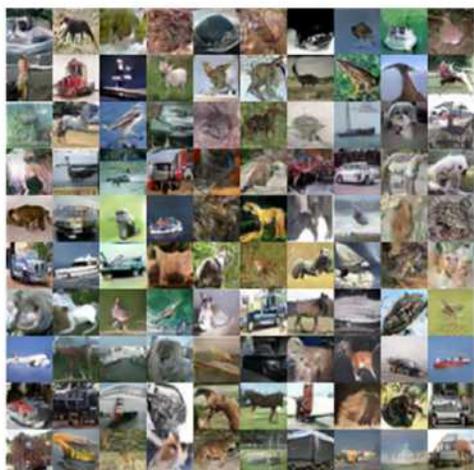
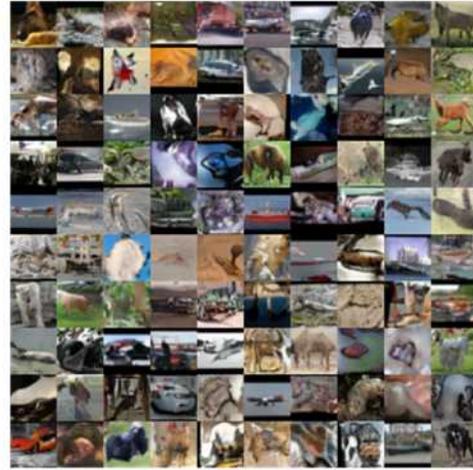


Figure 20. Synthetic images.

SRGAN

SNGAN

E16-128



E64-128



E256-128

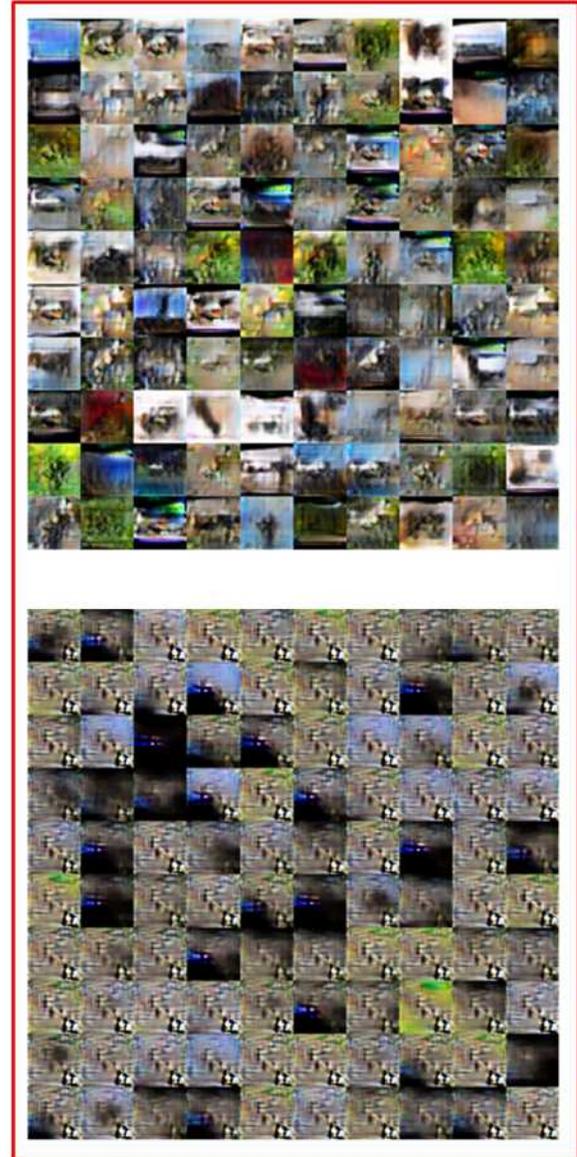


Figure 21. Synthetic images.

SRGAN

SNGAN

E256-64



E256-32



Mode collapse

Figure 22. Synthetic images.