PPCD-GAN: Progressive Pruning and Class-Aware Distillation for Large-Scale Conditional GANs Compression
Supplementary Material

Procedure 1 Training PPCD-GAN

```
Input: dataset $data$, PPCD-GAN with generator $G^{\mathrm{S}}$ and
    discriminator $D^{\mathrm{S}}$, pre-trained Big-GAN $G^{\mathrm{T}}$, compres-
    sion ratio threshold $\alpha$, epoch to train $T$
Output: optimized $G^{*\mathrm{S}}$
  for $t \leftarrow 1$ to $T$ do
    Get a batch $\{I_{\text{real}}, cls\}$ from $data$
    $D^{\mathrm{S}} \leftarrow$ UPDATEDISCRIMINATOR($I_{\text{real}}$, $cls$)
    $z \leftarrow$ RANDOMNOISE()
    $I_{\text{fake}}, F^{\mathrm{S}} \leftarrow G^{\mathrm{S}}(z, cls)$
    $F^{\mathrm{T}} \leftarrow G^{\mathrm{T}}(z, cls)$
    $\mathcal{L}_{\mathrm{PP}} \leftarrow$ COMPUTELOSSPP($G^{\mathrm{S}}$) (Eqs.2, 5)
    $\mathcal{L}_{\mathrm{CD}} \leftarrow$ COMPUTELOSSCD($F^{\mathrm{S}}$, $F^{\mathrm{T}}$) (Eqs.4, 6)
    $\mathcal{L}_{\mathrm{ADV}} \leftarrow$ COMPUTELOSSADV($D^{\mathrm{S}}$, $I_{\text{fake}}$, $cls$) (Eq.7)

    $G^{\mathrm{S}} \leftarrow$ UPDATEGENERATOR($\mathcal{L}_{\mathrm{PP}}$, $\mathcal{L}_{\mathrm{CD}}$, $\mathcal{L}_{\mathrm{ADV}}$)
    (Eq.8)
    for all mask $\mathbf{m}$ in $G^{\mathrm{S}}$ do
      BINARIZEMASK($\mathbf{m}$, $\alpha$) (Eq.3)
    end for
  end for
```

I. Training procedure

Procedure 1 illustrates the training procedure which we jointly train our model in the end-to-end manner to minimize Eq.(8).

II. Additional experiment

Although the ablation study presented in the main manuscript already illustrates the contribution of the class-aware distillation, we further set up an additional experiment to investigate its effectiveness. To this end, we trained our PPCD-GAN with class-irrelevant distillation, which is a modification of our class-aware distillation. While we employ the same class condition (as used in the generator) for both PPCD-GAN and the teacher model (i.e., Big-GAN) in executing class-aware distillation, we feed different class conditions to PPCD-GAN and Big-GAN to obtain class-irrelevant distillation. More precisely, let cls be a class condition and cls^- be a different class condition (an irrelevant class to cls), we use cls and cls^- when computing attention maps of our PPCD-GAN and Big-GAN, respectively. As a result, Eqs. (4) and (5) in the main manuscript are easily modified to class-irrelevant distillation. Accordingly, in **Procedure 1**, we replace $F^{\mathrm{T}} \leftarrow G^{\mathrm{T}}(z, cls)$ with $F^{\mathrm{T}} \leftarrow G^{\mathrm{T}}(z, cls^-)$.

PPCD-GAN$^-$ denotes the model trained with class-irrelevant distillation (we used $\alpha = 0.5$), PCCD-GAN$^-$ w/o $\mathcal{L}_{\mathrm{PP}}$ denotes the model dropping PP-Res while using class-irrelevant distillation (i.e., we trained SA-GAN [2] with class-irrelevant distillation). We remind that PPCD-GAN (complete) indicates our complete model and PPCD-GAN w/o $\mathcal{L}_{\mathrm{PP}}$ is the model dropping PP-Res (refer to the main manuscript for more details). Table A shows the quantitative comparison of all the models. We see that the models using class-irrelevant distillation perform far worse than those using class-aware distillation. We also visually observe that generated images by the models using class-irrelevant distillation are not clear in details and suffer from class leakage (Fig. A). This is because the class-irrelevant distillation guides our model with inaccurate knowledge, leading to the models' inability to clearly distinguish classes. These observations strongly confirm the importance of the usage of class-aware distillation.

III. More examples

We show some more examples in this supplementary material. Figures B, C, and D show results obtained by our PPCD-GAN, Big-GAN [1], SA-GAN [2], SN-GAN [3], Tiny-GAN [5], Distill + NAS [6], and Distill only [34]. These figures consistently illustrate that our method generates high-quality images comparable to Big-GAN [1] and much better than the other methods (SA-GAN [2], SN-GAN [3], Tiny-GAN [5], Distill + NAS [6], and Distill only [34]).

Next, Fig. E illustrates the generated images obtained by the ablation models. We observe that all the ablation models generate reasonable images.

We show the outputs obtained by interpolating between

Table A: Comparison of the models using class-aware distillation and the models using class-irrelevant distillation. We used $\alpha = 0.5$ for all the models except for PPCD-GAN w/o $\mathcal{L}_{\mathrm{PP}}$ and PPCD-GAN^{-} w/o $\mathcal{L}_{\mathrm{PP}}$.

Models	Class-aware	Class-irrelevant	IS ↑	FID ↓	#Par. ↓
PPCD-GAN (complete)	✓		81.47	14.26	25M
PPCD-GAN w/o $\mathcal{L}_{\mathrm{PP}}$	✓		82.29	13.02	42M
PPCD-GAN^{-}		✓	23.40	37.06	25M
PPCD-GAN^{-} w/o $\mathcal{L}_{\mathrm{PP}}$		✓	34.50	24.01	42M

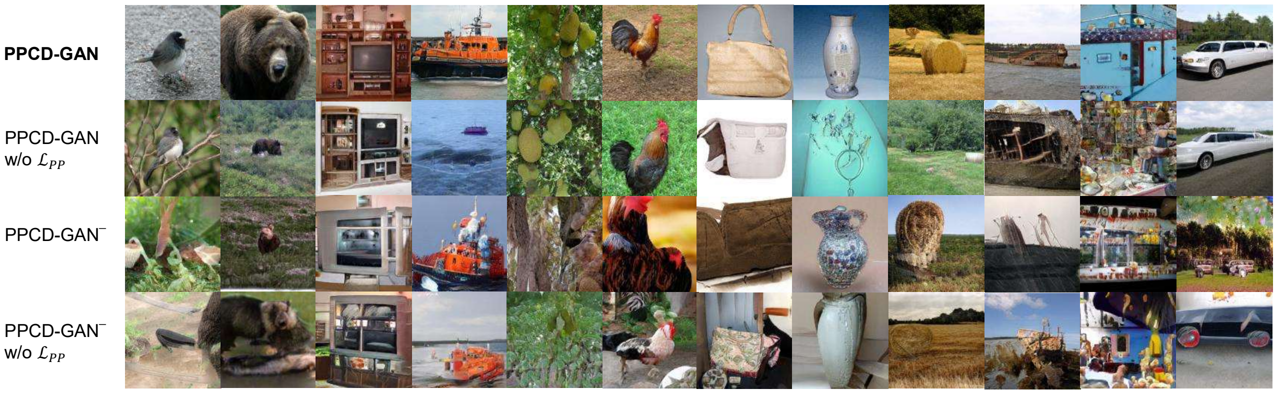
PPCD-GAN
PPCD-GAN w/o $\mathcal{L}_{PP}$
PPCD-GAN^{-}
PPCD-GAN^{-} w/o $\mathcal{L}_{PP}$

Figure A: Example results by the models using class-aware distillation and the models using class-irrelevant distillation. From left to right, conditional classes are junco, brown bear, entertainment center, lifeboat, jackfruit, cock, mailbag, vase, hay, wreck, toyshop, and limousine.

the noise vector and the class condition vector in Figs. F and G. These figures further illustrate the generalization ability of our PPCD-GAN.

Finally, we show the generated images obtained by changing compression ratio threshold α by 0.1 from 0.5 to 0.8, see Fig. H. We see that our model is capable of generating high-quality images up to $\alpha = 0.7$. As we mentioned in the main manuscript, our model fails in the case of $\alpha = 0.8$ because the size of the model becomes too small to precisely handle all the classes. Different from our PPCD-GAN, the quality of the images generated by the other methods (i.e., Big-GAN [1], SA-GAN [2] and PPCD-GAN*) quickly and drastically degrades as α increases. These observations further confirm the sustainable capability of PPCD-GAN even under different compression ratio thresholds. We show more examples obtained by our method with different compression ratio thresholds in Fig. I.

References

[1] A. Brock, J. Donahue, and K. Simonyan, “Large scale GAN training for high fidelity natural image synthesis,” in ICLR, 2019.

[2] H. Zhang, I. J. Goodfellow, D. N. Metaxas, and A. Odena, “Self-attention generative adversarial networks,” in ICML, 2019.

[3] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” in ICLR, 2018.

[5] T.-Y. Chang and C.-J. Lu, “Tinygan: Distilling biggan for conditional image generation,” in ACCV, 2020.

[6] M. Li, J. Lin, Y. Ding, Z. Liu, J.-Y. Zhu, and S. Han, “Gan compression: Efficient architectures for interactive conditional gans,” in CVPR, 2020

[34] A. Aguinaldo, P.-Y. Chiang, A. Gain, A. D. Patil, K. Pearson, and S. Feizi, “Compressing gans using knowledge distillation,” ArXiv, vol. abs/1902.00159, 2019.

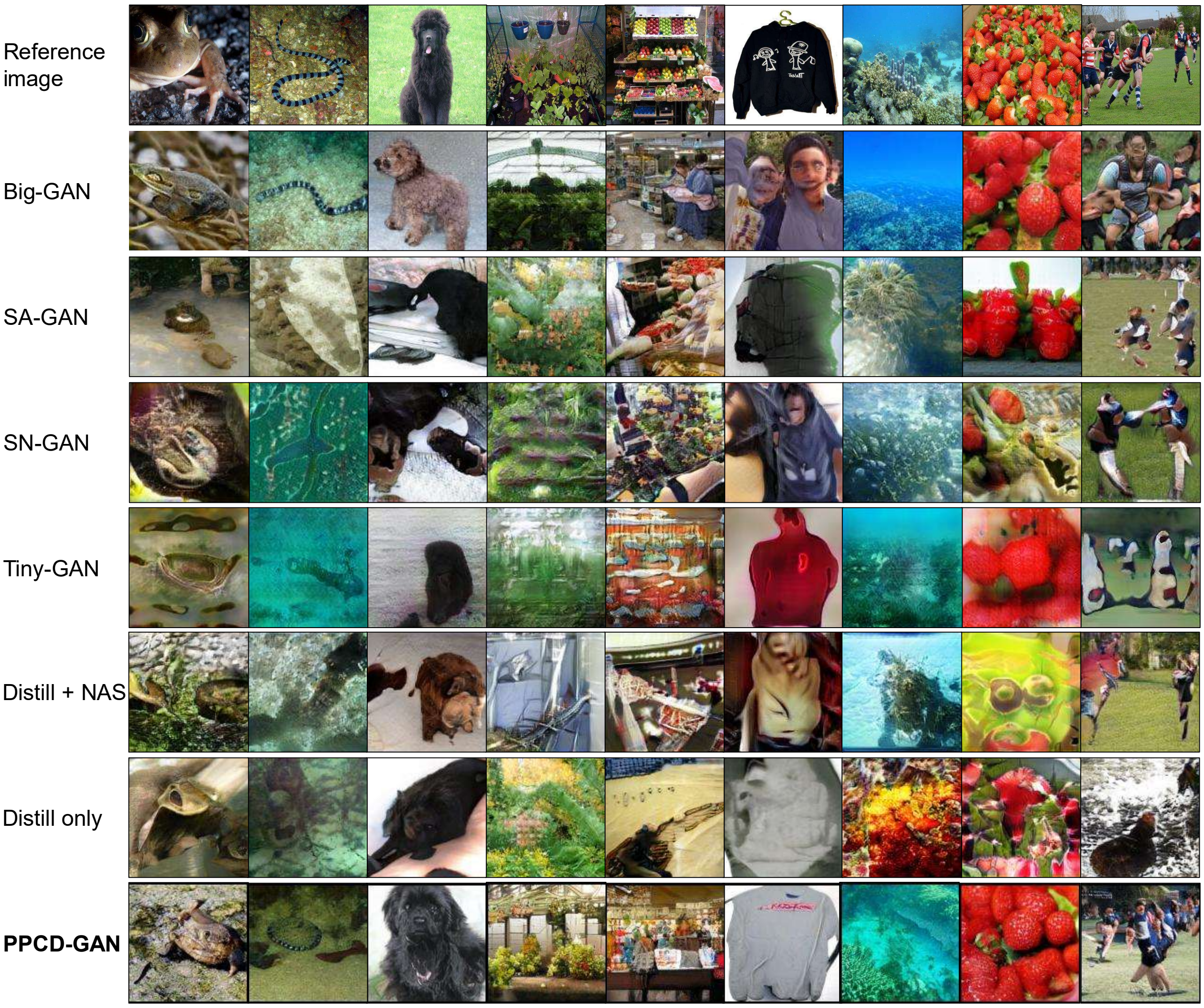

Figure B: Visual comparison of PPCD-GAN against Big-GAN [1], SA-GAN [2], SN-GAN [3], Tiny-GAN [5], Distill + NAS [6], and Distill only [34]. From left to right, conditional classes are bullfrog, sea snake, Newfoundland dog, greenhouse, grocery store, sweatshirt, coral reef, strawberry, and rugby ball.

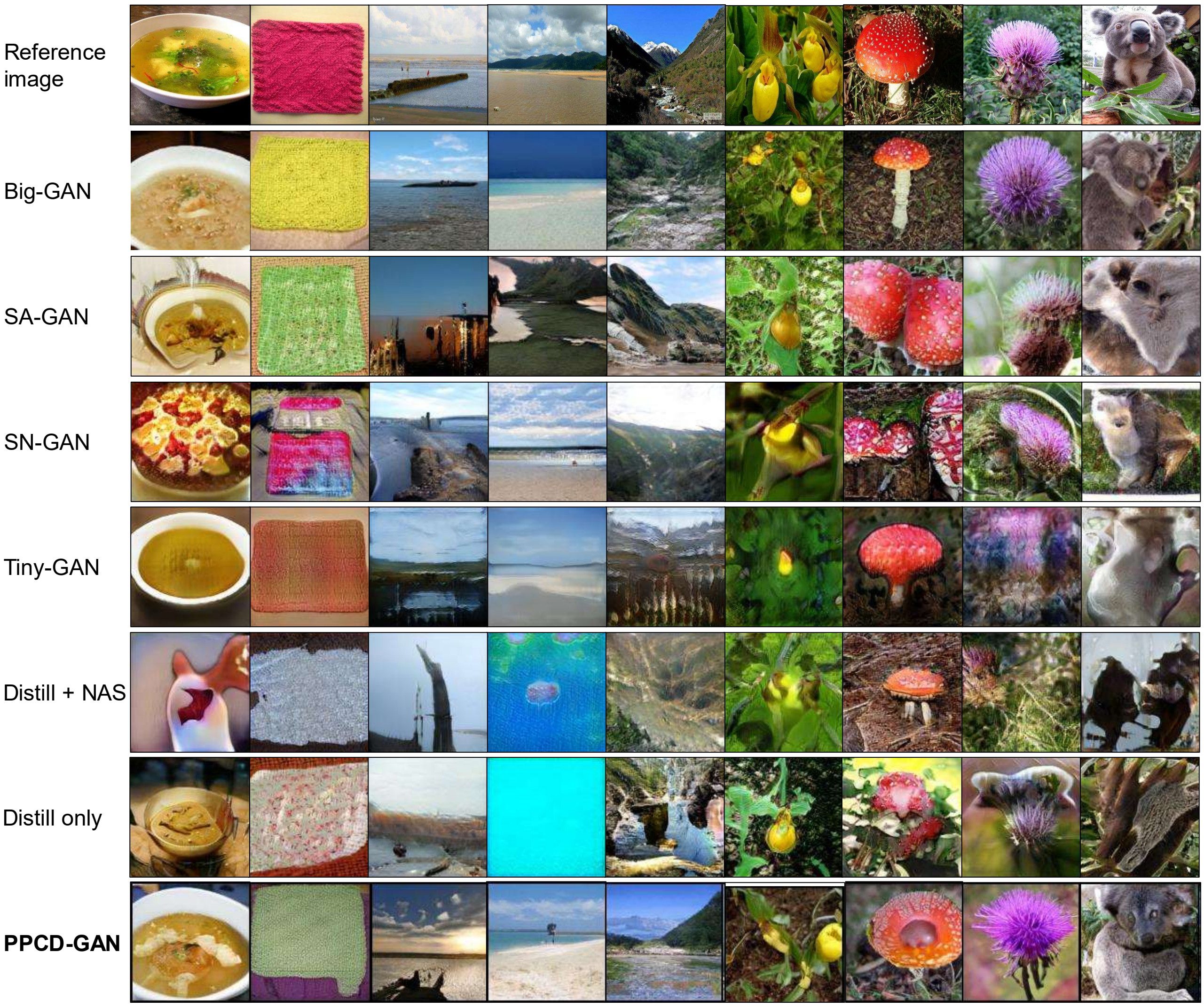

Figure C: Visual comparison of PPCD-GAN against Big-GAN [1], SA-GAN [2], SN-GAN [3], Tiny-GAN [5], Distill + NAS [6], and Distill only [34]. From left to right, conditional classes are soup bowl, dishrag, breakwater, sandbar, valley, yellow lady's slipper, agaric, cardoon, and koala.

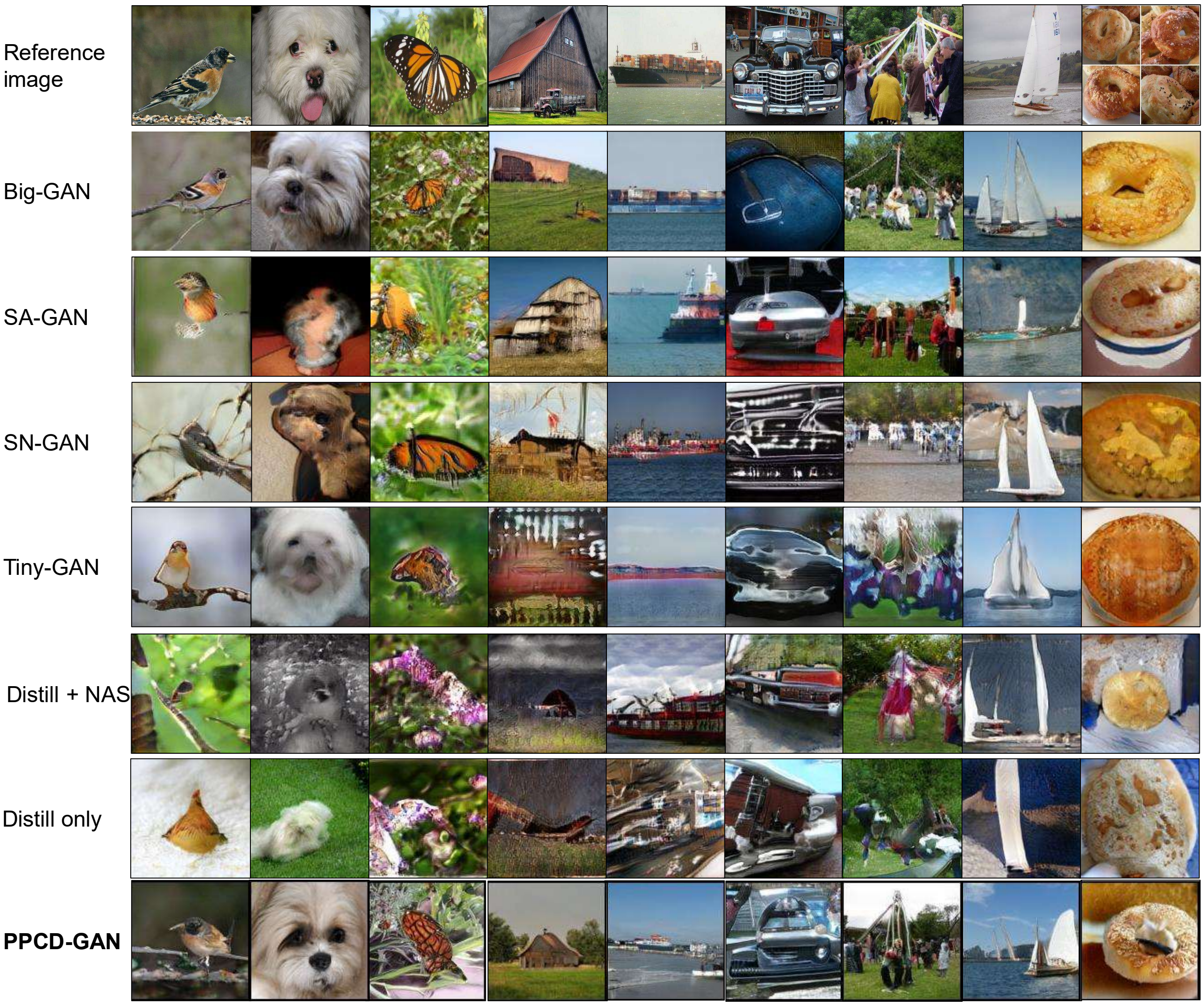

Figure D: Visual comparison of PPCD-GAN against Big-GAN [1], SA-GAN [2], SN-GAN [3], Tiny-GAN [5], Distill + NAS [6], and Distill only [34]. From left to right, conditional classes are brambling, Lhasa dog, monarch, barn, container ship, grille, maypole, yawl, and bagel.

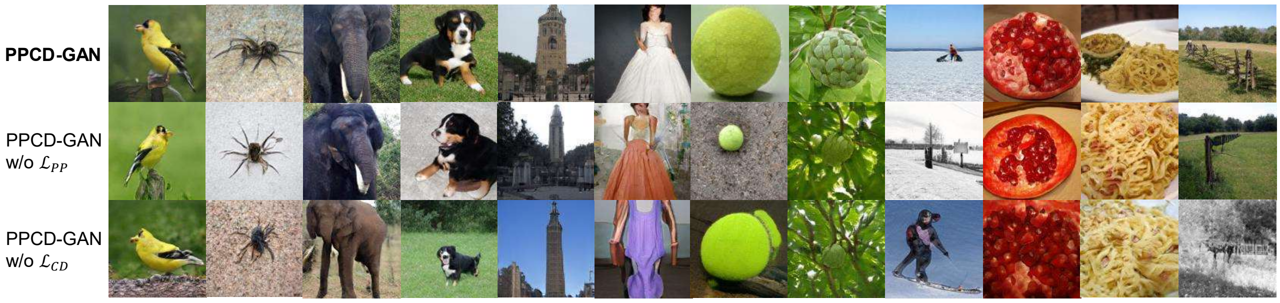
PPCD-GAN
PPCD-GAN w/o $\mathcal{L}_{PP}$
PPCD-GAN w/o $\mathcal{L}_{CD}$

Figure E: Example results by the ablation models. From left to right, conditional classes are goldfinch, wolf spider, tusker, bell cote, hoopskirt, tennis ball, custard apple, ski, pomegranate, carbonara, and worm fence.

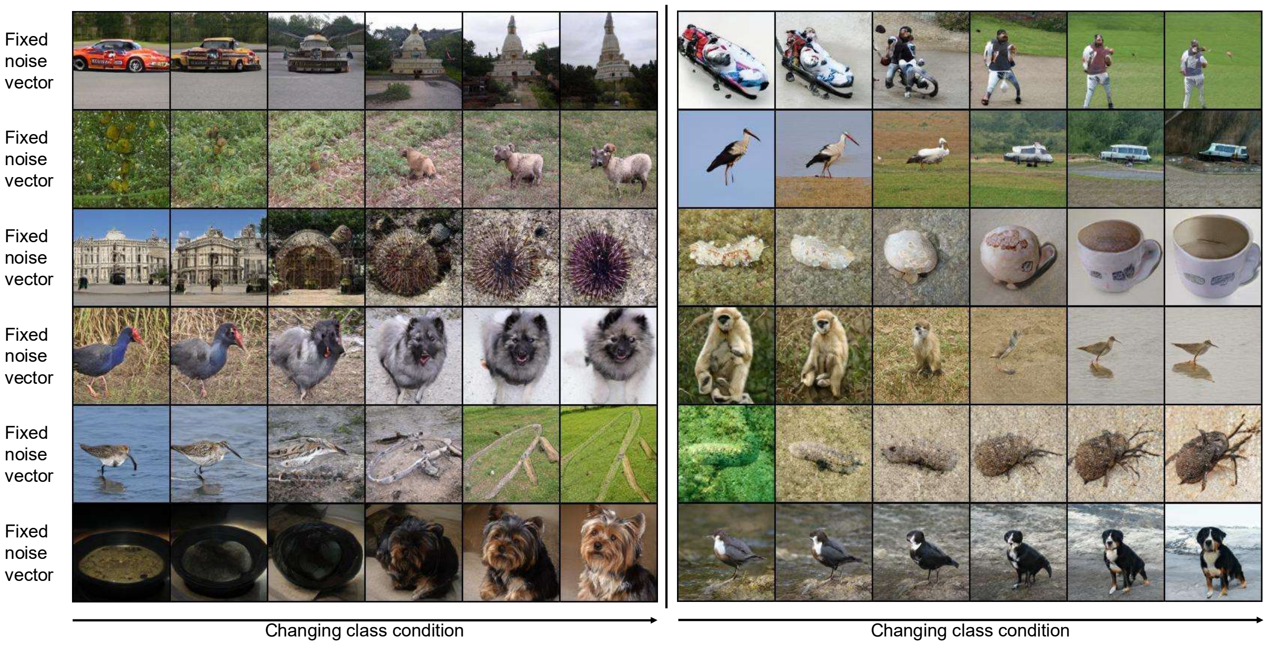

Figure F: Interpolated results by changing class condition while fixing noise vector.

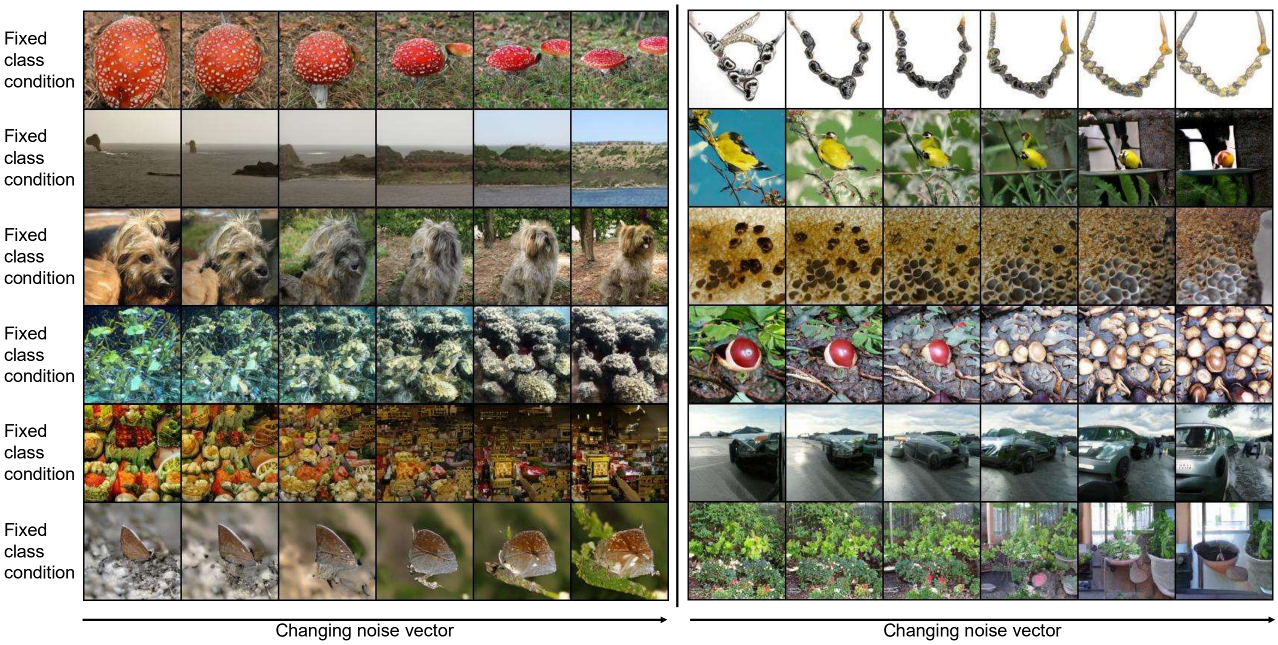

Figure G: Interpolated results by fixing class condition while changing noise vector.

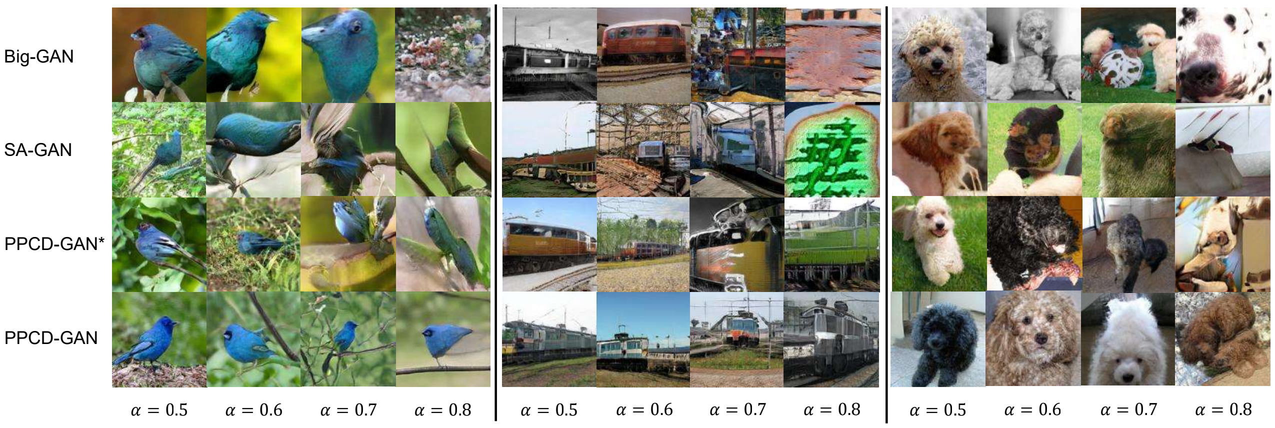

Figure H: Examples results obtained by Big-GAN [1], SA-GAN [2], PPCD-GAN* and PPCD-GAN by changing compression ratio α by 0.1 from 0.5 to 0.8.

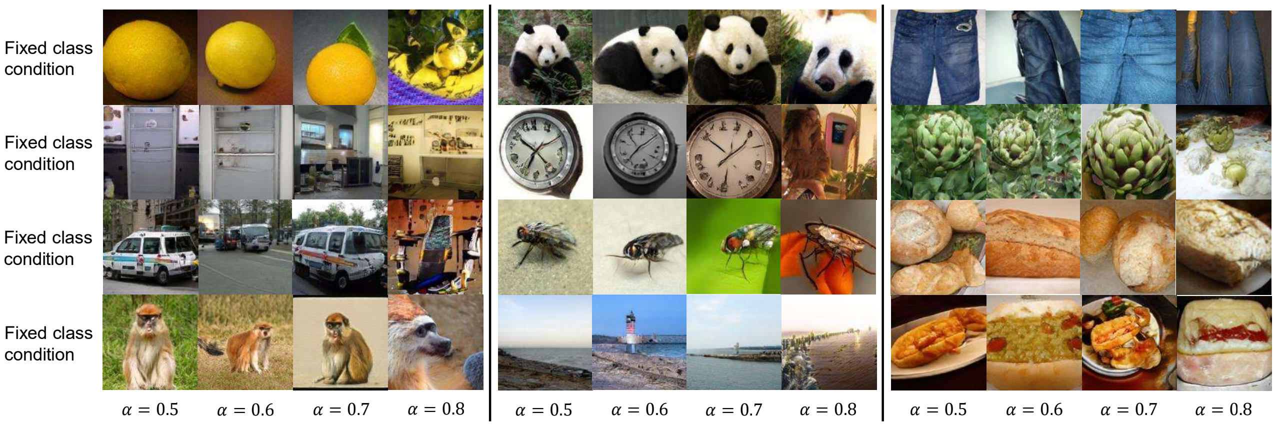

Figure I: Examples results obtained by our PPCD-GAN by changing compression ratio α by 0.1 from 0.5 to 0.8.