Generic and Privacy-free Synthetic Data Generation for Pretraining GANs

Anonymous Author(s)
Affiliation
Address
email

Abstract

Transfer learning for GANs successfully improves low-shot generation performance. However, existing studies show that the pretrained model using a single benchmark dataset is not generalized to various datasets. More importantly, the pretrained model can be vulnerable to copyright or privacy risks. To resolve both issues, we propose an effective and unbiased data synthesizer, namely `Primitives-PS`, inspired by the generic characteristics of natural images. Since `Primitives-PS` only considers the generic properties of natural images, the images are free from copyright and privacy issues. In addition, the single model pretrained on our dataset can be transferred to various target datasets. Extensive analysis demonstrates that each component of our data synthesizer is effective, and provides insights on the desirable nature of the pretrained model for the transferability of GANs.

1 Introduction

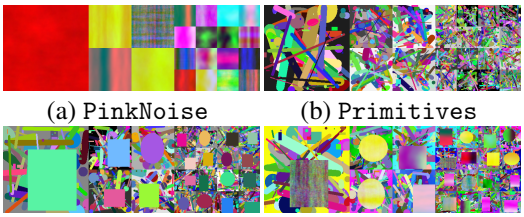
Generative adversarial networks (GANs) [7] are a powerful generative model that can synthesize complex data by learning the implicit density distribution with adversarial training. Despite the remarkable quality, GANs require at least several thousand, mostly several hundred thousand images for training. This requirement for data collection is often infeasible in practical applications (e.g., many pictures of a treasure, endangered species, or the medical images of rare disease).

The idea of transfer learning has been recently introduced to GAN training [35, 20] for low-shot generation. Following the common practice, GAN transfer learning 1) pretrains GANs on a large-scale source dataset (e.g., FFHQ) and then 2) finetunes GANs with a relatively small target dataset. As a result, developing GANs with transfer learning clearly improves the generation quality and diversity over the models trained from scratch only with the target dataset.

However, the effectiveness highly depends on how similar the source dataset is to the target dataset. According to TransferGAN [35], transfer learning can achieve the best performance when the source shares characteristics with the target. Other than the performance issue, we argue that the current pretrained models can be vulnerable to copyright and privacy issues [40]. Even for public benchmark datasets, employing them for commercial purposes is not always permitted. For this reason, one might compose her own dataset via web crawling, but filtering out the copyrighted samples is practically difficult. Besides, unresolved copyright and privacy issues might cause legal issues [28].

Recent studies [9, 12, 5] also show that the deep generative models are vulnerable to membership inference attacks, implying that privacy issues still remains beyond the copyright issues. An adversary can reconstruct a face even without additional prior information [37]. That is, we can reveal individual training samples by attacking the trained model. As the network capacity of GANs increases rapidly to improve performance, the risk of memorization also grows quickly. Memorization effects make GANs more vulnerable to membership inference attacks [4]. Since we consider transfer learning,

Submitted to NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research. Do not distribute.



(a) `PinkNoise` (b) `Primitives`

(c) `Primitives-S` (d) `Primitives-PS`

Figure 1: Visualization of our synthetic datasets. We visualize four variants of synthetic datasets and `Primitives-PS` is finally chosen for the best performance. Images are resized in three different scales.

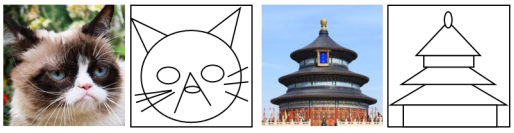


Figure 2: Potentials of primitive shapes for representing things. We only use a line, ellipse, and rectangle to express a cat and a temple. These examples motivate us to develop `Primitives`, which generates the data by a simple composition of the shapes.

someone might argue that the membership inference on the source (e.g., pretraining) dataset should not be a critical issue. However, Zou et al. [40] reported that the membership inference of the source dataset could be conducted even after the transfer learning.

In this work, we dive into tackling the two undiscovered but critical issues of transfer learning for GANs: 1) the lack of generalization for the pretrained model and 2) the copyright or privacy issue of the pretraining dataset. Towards this goal, we adopt the generic property of the natural images in the frequency spectrum and structure. We develop our data generation strategy, namely `Primitives-PS`, inspired by the analysis and observations on natural images from previous studies [23, 30, 19]. We pretrain GANs using the synthetic dataset generated by our data synthesizer. Then, the effectiveness of the proposed method is evaluated by repurposing the pretrained model to various low-shot datasets. The experimental results verify that our synthetic dataset pretrained model achieves comparable to FFHQ pretrained model or outperforms in various low-shot datasets. We also provide the analysis on the diversity of the filters to support the result in Appendix.

2 Related work

Utilizing synthetic datasets. The samples and labels of synthetic datasets can be generated automatically and unlimitedly by a pre-defined process. Since generating synthetic data can bypass the cumbersome data crawling and pruning for data collection, previous works have utilized synthetic datasets for training the model and then achieved performance improvement on real datasets [14, 13, 27, 25, 26, 36, 31].

Although the previous methods improved the performance of the model on the real dataset, generating such synthetic datasets requires expertise in domain knowledge or a specific software (e.g., GTA-5 game engine [24]). To handle the issue, Kataoka et al. [17] utilized the iterated function system to generate fractals and used the fractals as a pretraining dataset for classification. As a concurrent work, Baradad et al. [2] observe that the unsupervised representation learning [10] trains the model using patches, and these patches are visually similar to the noise patches (from the noise generation model) or the patches drawn from GANs. However, none of the existing studies have investigated synthetic data generation for training GANs.

Transfer learning in GANs. GANs involve a unique architecture and a training strategy; consisting of a discriminator and generator trained via adversarial competition. Therefore, the GAN transfer learning method should be developed by considering the unique characteristics of GANs [35, 21, 34, 20, 38, 22]. TransferGAN [35] trains GANs with a small number of samples by transferring the weights trained on a relatively large dataset. FreezeD [20] fixes several layers of the discriminator and then finetunes the remaining layers. FreezeD improved the generation performance of transferring from the FFHQ pretrained model to various animals. Despite the improvement in GAN transfer learning, the model still requires a large-scale pretraining dataset. Consequently, they commonly suffer from copyright issues, and their performance is sensitive to the relationship between the source and target dataset. In contrast, our goal is to tackle both issues simultaneously by introducing an effective data synthesizer.

Low-shot learning in GANs. For high-quality image generation, GANs require a large-scale dataset, and such a requirement can limit the practical use of GANs. To reduce the number of samples for training, several recent studies have introduced data augmentation for training the

(a) (b)

Figure 3: Comparison between (a) `Primitives` and (b) `Primitives-PS` on Obama dataset. The model pretrained with `Primitives` generates multiple faces in a single image.

S \ T	Obama	Grumpy cat	Bridge	Panda
Scratch + DiffAug	48.98	27.51	57.72	15.82
`PinkNoise`	50.32	29.47	73.82	15.65
`Primitives`	<u>43.20</u>	27.97	59.89	12.78
`Primitives-S`	43.29	<u>26.57</u>	<u>57.24</u>	**11.95**
`Primitives-PS`	**41.62**	**26.01**	**54.02**	<u>12.23</u>

Table 1: The FID score of transferring to low-shot datasets from the proposed pretraining datasets. The lower is the better. Bold and underlined text indicates the best and second best performance among the pretraining datasets. It will be the same convention throughout the paper. S and T represent the source and target dataset, respectively.

discriminator[33, 39, 15]. Then, the generator can produce images with a small number of samples without reflecting an unwanted transformation in the results (i.e., augmentation leakage [15]). In this work, we tackle low-shot generation using GANs via transfer learning; GANs are trained with a small number of samples by transferring a pretrained network into a low-shot dataset.

3 Towards an effective data synthesizer

In this work, our primary goal is to develop an unbiased and effective data synthesizer. The synthetic dataset secured by our synthesizer is then used to pretrain GANs, which facilitates low-shot data generation. To accomplish unbiased data generation, we only consider the generic properties of natural images because the inductive bias in a pretraining dataset is harmful to transfer learning of GANs. In the following, we introduce three design choices of our synthesizer inspired by the common characteristics of natural images: 1) modeling the power spectrum of images, 2) exploiting the shape primitives from natural images, and 3) adopting the existence of saliency in images. Figure 1 shows several example images by adding each aspect.

3.1 Learning the power spectrum of natural images

Several previous works reported the magnitude of natural images in the frequency domain [6, 3, 32] roughly obeys $w_m = \frac{1}{|f_x|^a + |f_y|^a}$ where a is a constant, well approximated to one. Inspired by this finding, we generate synthetic images by randomly drawing a from the uniform distribution of $\mathcal{U}(0.5, 3.5)$, as also suggested in [2]. Specifically, random white noise is sampled, and then its magnitude signal after applying the Fast Fourier Transform (FFT) is weighted by w_m. By applying the inverse FFT to the weighted signal, we can easily compute the synthetic image. We repeat this for RGB color channels and finally produce synthetic images. Originally, the image with $a = 1$ was named a pink noise. We call this method of generating images with $a \sim \mathcal{U}(0.5, 3.5)$ as `PinkNoise`. Since we only utilize the generic properties of natural images, no inductive bias toward any specific dataset influences `PinkNoise`. As shown in Figure 1(a), `PinkNoise` produces interesting patterns with vertical, horizontal orientation, or color blobs.

3.2 Shape primitives inspired by natural images

> *"Everything in nature is formed upon the sphere, the cone, and the cylinder. One must learn to paint these simple figures, and then one can do all that he may wish."*
>
> Paul Cézanne

Considering the importance of phase in images (e.g., determining the unique appearance of the image [23]), `PinkNoise` alone is insufficient to represent the rich characteristics of natural images; `PinkNoise` is random noise on a phase spectrum. To have a meaningful signal even in its phase, we can consider 1) modeling the phase of natural images independently or 2) developing the different generation strategies to model the magnitude and phase simultaneously. Unlike the magnitude spectrum, we seldom find regularity in the phase of images. Besides, separately modeling the phase

(a) From scratch (b) TransferGAN (c) FreezeD (d) `Primitives-PS`

Figure 4: Qualitative evaluation on Obama, Grumpy cat, Temple, and Wuzhen.

and magnitude may not produce meaningful images, preserving the proper structures [30]. For this reason, we focus on finding structural regularity in natural images because it can affect both magnitude and phase. Specifically, we are inspired by the observation that natural images can be represented by the composition of the elementary shapes [19]. The common practice in artistic drawings also utilizes elementary shapes as the basis for representing things (inspired by Paul Cézanne).

Figure 2 demonstrates the abstraction examples of various images using elementary shapes, such as ellipses, lines, and rectangles. We find the potential of abstraction via elementary shapes to encode the structural information of natural images and to remove the bias to a specific dataset. We then devise the data synthesizer to produce images consisting of various elementary shapes. The outputs of this synthesis procedure are akin to those of the dead leaves model [8, 18]. The dead leaves model is an early generative model, which closely mimics natural images by conducting tessellation, where their sizes and positions are determined by sampling from the Poisson process. Unlike the dead leaves model, we do not fill all the regions and use different distributions for sampling because the resultant images are quite sensitive to the hyperparameter of the Poisson process. For position, we use the uniform distribution. To prevent the large shapes in the later stage from completely overwriting those in the early stage, we gradually decrease the maximum shape size over multiple stages; drawing the small objects toward the end. In addition, it is conversely proportional to the number of currently injected shapes. We name this generation strategy `Primitives`, and Figure 1(b) visualizes the representative examples.

3.3 Combining saliency as prior

In addition to the natural images, we investigate the benchmark datasets and find that they commonly have saliency, target objects of interest to determine the class. These salient objects are usually located nearly in the middle of the image. For example, the animal face on the cat and panda dataset can be the saliency. To reflect the nature of benchmark datasets, we insert a large shape after applying `Primitives` and name it as `Primitives-S` (`Primitives` with Saliency).

By utilizing the three design factors, we develop four variants of our data synthesizer. They are 1) `PinkNoise` adopting the nature of magnitude spectrum of natural images only as shown in Figure 1(a), 2) `Primitives` generating various elementary (monotone) shapes randomly as illustrated in Figure 1(b), and 3) `Primitives-S` adding a salient object into `Primitives` in Figure 1(c).

Finally, we apply a `PinkNoise` pattern onto the salient object and the background of `Primitives-S`, which is called (4) `Primitives-PS` (`Primitives` with Patterned Saliency) as shown in Figure 1(d). Since the size of the salient object is considerable, having a salient monotone object may induce an unwanted texture bias. Focusing on the visual effects, inserting the monotone object can be similar to the regional dropout [29, 1] in the weakly-supervised object localization (WSOL) task. When training a network with the regional dropout, previous WSOL methods suggest filling the dropped region with mean statistics or with other regions from the same image to prevent distribution bias. Motivated by the practice in WSOL, we apply `PinkNoise` to the salient object.

The effectiveness of the proposed synthetic datasets is evaluated by transferring GANs in Section 4. We first pretrain GANs using the randomly generated images via our `Primitives-PS`, and then finetune the pretrained model on low-shot datasets. While finetuning, all competitors and our pretrained model utilize DiffAug (translation, cutout, and color jittering).

Source \ Target	Obama	Grumpy cat	Bridge	Panda	Temple	Wuzhen	Fountain	Buildings
Scratch + DiffAug [39]	48.98	27.51	57.72	15.82	46.69	146.81	44.46	93.71
TransferGAN [35]	36.50	30.60	60.29	14.53	40.58	95.83	46.61	81.63
FreezeD [20]	**35.90**	29.41	59.47	13.39	42.09	93.54	45.70	80.48
`Primitives-PS`	41.62	**26.01**	**54.02**	**12.23**	**40.42**	**88.14**	**43.06**	**78.74**

Table 2: The FID score of transferred models to low-shot datasets. We use FFHQ pretrained weight for TransferGAN and FreezeD. For all models, we apply DiffAug.

4 Experiments

We first demonstrate the effectiveness of four variants of our data synthesizer. Then, we choose the best strategy among the four variants and use it for pretraining GANs. Our model is compared with other models pretrained on a natural benchmark dataset in the transfer learning scenario.

Datasets. For the comparison between our synthesizers, we adopt four datasets, including Obama, Grumpy cat, Panda, and Bridge of sighs (Bridge) [39]. To compare with transfer learning methods, we also use Wuzhen, Temple of heaven (Temple), and Medici fountain (Fountain). Each dataset has 100 images. In addition, we create a dataset, namely Buildings, by merging a subset of four datasets; Bridge of sighs, Wuzhen, Temple of heaven, and Medici fountain. Buildings is used to evaluate the performance under highly diverse conditions.

Evaluation protocols. StyleGAN2 architecture [16] with DiffAug [39] is applied when evaluating all models in the low-shot generation task. The baseline is the model trained from scratch with DiffAug. The strong competitors are TransferGAN [35] and FreezeD [20], where both methods suggest finetuning strategies. To reproduce the competitors, we used StyleGAN2 model pretrained on FFHQ and then finetune the model using TransferGAN and FreezeD, respectively. We stress that all competitors, baseline and `Primitives-PS` use DiffAug. We follow the configuration of DiffAug for `Primitives-PS` and the baseline (from scratch with DiffAug). Otherwise, we use the configuration of TransferGAN and FreezeD as described in [39] for the best performance. As an evaluation metric, we use Fréchet inception distance (FID) [11] and report the FID score of the best model during training as suggested by DiffAug [39].

4.1 Effects of different data synthesizers

We developed four variants of data synthesizer: `PinkNoise`, `Primitives`, `Primitives-S`, and `Primitives-PS`. We evaluate their effectiveness in the low-shot generation scenario– pretraining with the synthetic dataset and then finetuning on target datasets with DiffAug. Table 1 summarizes the FID scores of four data synthesizers and the baseline under four different low-shot datasets.

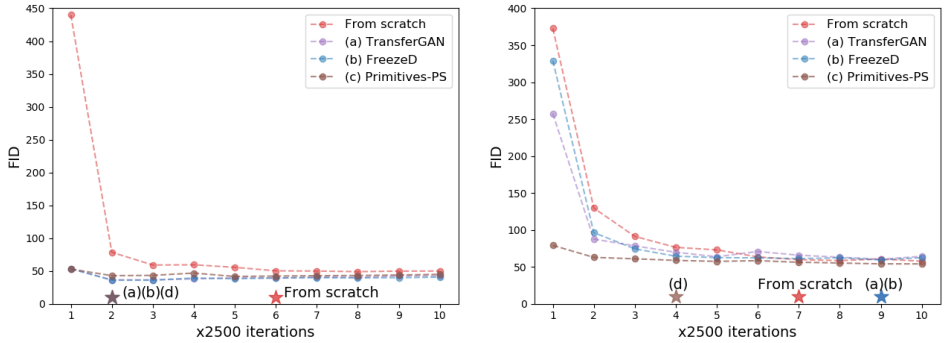
In general, `PinkNoise` fails to improve the FID score (worse than the baseline). Unlike `PinkNoise`, `Primitives` clearly improves the generation performance in Obama and Panda. However, it is not effective on Grumpy cat and Bridge. Compared to `Primitives`, `Primitives-S` further improves the FID scores, demonstrating the effectiveness of saliency prior. Finally, `Primitives-PS` clearly improves the low-shot generation performance on all datasets by about 15% on average over the baseline. From these results, we observe that 1) a naïve synthesizer (`PinkNoise`) is even worse than simply using the low-shot dataset, and 2) the combination of our three design factors (`Primitives-PS`) remarkably improves the baseline, supporting the effectiveness and importance of each factor.

We also verify the effect of the saliency prior by comparing `Primitives` and `Primitives-PS`. Because `Primitives` includes multiple shapes, meaning all can be candidates for the main object, the outputs often contain multiple faces in the middle of training (e.g., the top-left and the middle in Figure 3(a)). On the other hand, `Primitives-PS` focuses on generating a single face and eventually exhibits improved quality.

Considering all, we confirm that `Primitives-PS` is the best data synthesizer, and thus it is chosen as our final model for comparative evaluations with competitors.

4.2 Comparisons with the state-of-the-arts

We pretrain a model using `Primitives-PS` and compare it with state-of-the-art models pretrained with FFHQ in a transfer learning task to low-shot datasets.



(a) Obama (b) Bridge

Figure 5: FID per training iterations. The star marker (★) indicates the point where the model reaches 95% of the best FID score of the from scratch model with DiffAug (baseline).

Table 2 reports the quantitative results and Figure 4 shows the qualitative comparison. As expected, TransferGAN [35] and FreezeD [20] show outstanding performance on the Obama dataset because they are pretrained with FFHQ that is a superset of the target. Except for the Obama dataset, `Primitives-PS` pretrained model outperforms all competitors. Unless the inductive bias in the source dataset is advantageous to the target (e.g., Obama), FreezeD does not consistently outperform the baseline (from scratch with DiffAug). In fact, the performances of existing methods highly vary upon target datasets. Contrarily, our pretrained model with `Primitives-PS` consistently outperforms the competitors in each dataset, except Obama. This implies that our pretrained model has strong transferability. Since `Primitives-PS` does not use any inductive bias for modeling human faces, the performance drawback on Obama can be acceptable.

We emphasize that our achievement is impressive and meaningful in two aspects: 1) `Primitives-PS` uses no real but all synthetic images, which possesses all the attractive nature in application scenarios and 2) our results show the great potential of a single pretrained model for GAN transfer learning.

Training convergence speed. We investigate the convergence speed of transfer learning by examining FID upon training iterations. Figure 5 describes the evolution of the FID scores during the training. To save space, we provide two different datasets; Obama and Bridge. For Obama, all pretrained models converge faster than the baseline (from scratch with DiffAug). Meanwhile, only our model converges faster than the baseline for Bridge. Compared to the baseline, the model pretrained with `Primitives-PS` reaches 95% of the best baseline performance within the first 30% of iterations. Interestingly, other pretrained models cannot reach 95% of the best baseline performance earlier than the baseline. This shows that our model effectively reduces the required iterations for convergence, and the overhead for pretraining can be sufficiently deducted.

5 Limitation and conclusion

Limitation. Our `Primitives-PS` is devised based on the observations from natural images. Hence, it is possible that more effective observations can further improve the data generation quality. Developing a metric to quantify the transferability of the model and deriving a data generation process by optimizing the metric might be a good way to overcome this limitation.

Conclusion. Existing studies for GAN transfer learning utilize a model trained with natural images and thereby suffer from 1) biased pretrained model that can be harmful to the resultant performance and 2) copyright or privacy issues with both the model and dataset. To overcome these limitations, we introduce a new image synthesizer, namely `Primitives-PS`, inspired by the three generic properties of natural images: 1) following the power spectrum of natural images, 2) abstracting the image via the composition of primitive shapes (e.g., line, circle, and rectangle), and 3) having saliency in the image. Experimental comparisons and analysis show that our strategy effectively improves both the generation quality and the convergence speed.

References

[1] Kyungjune Baek, Minhyun Lee, and Hyunjung Shim. Psynet: Self-supervised approach to object localization using point symmetric transformation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10451–10459, 2020.

[2] Manel Baradad, Jonas Wulff, Tongzhou Wang, Phillip Isola, and Antonio Torralba. Learning to see by looking at noise. In *Advances in Neural Information Processing Systems*, 2021.

[3] Geoffrey J Burton and Ian R Moorhead. Color and spatial structure in natural scenes. *Applied optics*, 26(1):157–170, 1987.

[4] Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song. The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th {USENIX} Security Symposium ({USENIX} Security 19)*, pages 267–284, 2019.

[5] Dingfan Chen, Ning Yu, Yang Zhang, and Mario Fritz. Gan-leaks: A taxonomy of membership inference attacks against generative models. In *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, pages 343–362, 2020.

[6] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, 4(12):2379–2394, 1987.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] Robert M Haralick, Stanley R Sternberg, and Xinhua Zhuang. Image analysis using mathematical morphology. *IEEE transactions on pattern analysis and machine intelligence*, (4):532–550, 1987.

[9] Jamie Hayes, Luca Melis, George Danezis, and Emiliano De Cristofaro. Logan: Membership inference attacks against generative models. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*, volume 2019, pages 133–152. De Gruyter, 2019.

[10] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

[11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[12] Benjamin Hilprecht, Martin Härterich, and Daniel Bernau. Monte carlo and reconstruction membership inference attacks against generative models. *Proc. Priv. Enhancing Technol.*, 2019(4):232–249, 2019.

[13] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. PMLR, 2018.

[14] Yunzhong Hou and Liang Zheng. Visualizing adapted knowledge in domain transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13824–13833, 2021.

[15] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12104–12114. Curran Associates, Inc., 2020.

[16] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.

[17] Hirokatsu Kataoka, Kazushige Okayasu, Asato Matsumoto, Eisuke Yamagata, Ryosuke Yamada, Nakamasa Inoue, Akio Nakamura, and Yutaka Satoh. Pre-training without natural images. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.

[18] Ann B Lee, David Mumford, and Jinggang Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1):35–59, 2001.

[19] Ravish Mehra, Qingnan Zhou, Jeremy Long, Alla Sheffer, Amy Gooch, and Niloy J Mitra. Abstraction of man-made shapes. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–10. 2009.

[20] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Freeze the discriminator: a simple baseline for fine-tuning gans. In *CVPR AI for Content Creation Workshop*, 2020.

[21] Atsuhiro Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2750–2758, 2019.

[22] Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot image generation via cross-domain correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10743–10752, 2021.

[23] Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

[24] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *European Conference*

on Computer Vision (ECCV), volume 9906 of *LNCS*, pages 102–118. Springer International Publishing, 2016.

[25] Subhankar Roy, Evgeny Krivosheev, Zhun Zhong, Nicu Sebe, and Elisa Ricci. Curriculum graph co-teaching for multi-target domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5351–5360, 2021.

[26] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018.

[27] Rui Shu, Hung Bui, Hirokazu Narui, and Stefano Ermon. A DIRT-t approach to unsupervised domain adaptation. In *International Conference on Learning Representations*, 2018.

[28] Natasha Singer and Mike Isaac. Facebook to pay $550 million to settle facial recognition suit. *The New York Times*, 2019.

[29] Krishna Kumar Singh and Yong Jae Lee. Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In *International Conference on Computer Vision (ICCV)*, 2017.

[30] Y Tadmor and DJ Tolhurst. Both the phase and the amplitude spectrum may determine the appearance of natural images. *Vision research*, 33(1):141–145, 1993.

[31] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 23–30. IEEE, 2017.

[32] DJ Tolhurst, Y_ Tadmor, and Tang Chao. Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, 12(2):229–232, 1992.

[33] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. On data augmentation for gan training. *IEEE Transactions on Image Processing*, 30:1882–1897, 2021.

[34] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. Minegan: effective knowledge transfer from gans to target domains with few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9332–9341, 2020.

[35] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring gans: generating images from limited data. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 218–234, 2018.

[36] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020.

[37] Yuheng Zhang, Ruoxi Jia, Hengzhi Pei, Wenxiao Wang, Bo Li, and Dawn Song. The secret revealer: Generative model-inversion attacks against deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 253–261, 2020.

[38] Miaoyun Zhao, Yulai Cong, and Lawrence Carin. On leveraging pretrained gans for generation with limited data. In *International Conference on Machine Learning*, pages 11340–11351. PMLR, 2020.

[39] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient gan training. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[40] Yang Zou, Zhikun Zhang, Michael Backes, and Yang Zhang. Privacy analysis of deep learning in the wild: Membership inference attacks against transfer learning. *arXiv preprint arXiv:2009.04872*, 2020.