

Can We Find Strong Lottery Tickets in Generative Models?

Sangyeop Yeo¹, Yoojin Jang¹, Jy-yong Sohn², Dongyoon Han³, Jaejun Yoo^{1*}

¹LAIT, Ulsan National Institute of Science and Technology (UNIST)

²University of Wisconsin-Madison

³Naver AI Lab

sangyeop377@gmail.com, softjin@unist.ac.kr, jysohn1108@gmail.com, dongyoon.han@navercorp.com, jaejun.yoo@unist.ac.kr (*: corresponding author)

Abstract

Yes. In this paper, we investigate *strong lottery tickets* in generative models, the subnetworks that achieve good generative performance without any weight update. Neural network pruning is considered the main cornerstone of model compression for reducing the costs of computation and memory. Unfortunately, pruning a generative model has not been extensively explored, and all existing pruning algorithms suffer from excessive weight-training costs, performance degradation, limited generalizability, or complicated training. To address these problems, we propose to find a strong lottery ticket via moment-matching scores. Our experimental results show that the discovered subnetwork can perform similarly or better than the trained dense model even when only 10% of the weights remain. To the best of our knowledge, we are the first to show the existence of strong lottery tickets in generative models and provide an algorithm to find it stably. Our code and supplementary materials are publicly available at <https://lait-cvlab.github.io/SLT-in-Generative-Models/>.

Introduction

State-of-the-art generative models tend to use extremely large and complex structures for better performance (Brock, Donahue, and Simonyan 2018; Karras, Laine, and Aila 2019; Karras et al. 2020; Ramesh et al. 2021; Radford et al. 2021). One downside of large models is the high computational costs for training, which limits their application to edge devices such as mobile environments. This naturally calls for the design of a new lightweight architecture or a new compression method in generative modeling.

In this work, we focus on the model pruning techniques that fall into the latter category. Unlike discriminative models where various pruning techniques (LeCun, Denker, and Solla 1989; Hassibi and Stork 1992; Han et al. 2015; Frankle and Carbin 2018; Ramanujan et al. 2020; Sreenivasan et al. 2022) have been actively studied, pruning generative models have not been extensively explored. Moreover, it has been found that naïve application of existing pruning methods (developed for discriminative models) to generative models leads to performance degradation and/or unstable training (Wang et al. 2020; Li et al. 2021).

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



(a) Dense (11 M params) (b) Ours (1.1 M params)

Figure 1: Visualization of dense models and our strong lottery tickets. For various datasets including CelebA, FFHQ, LSUN Bedroom, and CIFAR-10, we compare the images generated by (a) a trained dense generative feature matching network (GFMN); (b) a subnetwork discovered by our method, which uses only 10% of parameters within a randomly-initialized GFMN. Intriguingly, our method finds the strong lottery tickets which achieve comparable or even better generative performances over the baseline dense model qualitatively and quantitatively (Table 1).

Recently, several methods for pruning generative models have been proposed and showed that it is possible to obtain a lightweight model by following the “train, prune, re-train” paradigm when tuned to generative models with special care (Li et al. 2020; Wang et al. 2020; Liu et al. 2021; Li et al. 2021; Tuli et al. 2021; Hou et al. 2021). For example, to overcome training instability, Hou *et al.* (Hou et al. 2021) introduced multiple shared discriminators to train a slimmable generator that can flexibly change its capacity at runtime. Li *et al.* (Li et al. 2021) proposed a cooperative scheme between the generator and the discriminator to stabilize the compression during the adversarial training.

However, because their basic strategy inevitably involves subtle balancing between training and pruning procedures, all existing methods suffer from excessive computational costs (Liu et al. 2021; Tuli et al. 2021; Hou et al. 2021), performance degradation (Wang et al. 2020; Li et al. 2021;

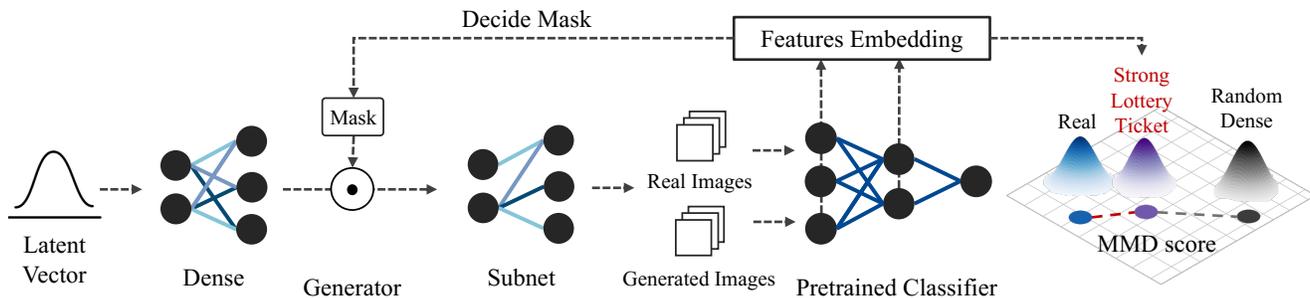


Figure 2: A schematic overview of our method. Our method finds a strong lottery ticket (SLT) via moment-matching scores. By exploiting the power of the pretrained classifier, our method assigns scores to the randomly initialized weights and finds a sparse mask so that the discovered subnetwork performs similarly or better than the trained dense generator.

Chen et al. 2021), limited generalizability (Li et al. 2020; Hou et al. 2021), or complicated training (Liu et al. 2021). This, combined with the notorious instability of generative adversarial networks (GANs), which most methods target, makes it more challenging to develop a pruning method for generative models.

To address these problems, we propose to find *strong lottery tickets* in generative models. A strong lottery ticket is a subnetwork at initialization (*i.e.*, no weight update) that performs similarly or even better than its dense counterpart whose weights are trained. Here, we employ the edge-popup (EP) algorithm (Ramanujan et al. 2020), which is the earliest method to find a strong lottery ticket in discriminative models. The EP algorithm selects a subnetwork mask based on the idea that one can “score” the importance of each weight. Once such a score is assigned, one simply keeps the weights of high scores according to the desired target sparsity.

Because the performance of the EP algorithm largely depends on the updated scores that serve as pruning criteria, it is essential to use a proper score function that gives a representative feature for pruning generative models. One may easily think of the adversarial loss, a commonly used criterion for training high-quality generators, but it is extremely unstable and hinders the search for appropriate scores. Instead, we propose to utilize a technique from statistical hypothesis testing known as maximum mean discrepancy (MMD) (Gretton et al. 2006, 2012), which leads to a simple moment-matching score using features extracted from a fixed, pretrained ConvNets (Li, Swersky, and Zemel 2015; Li et al. 2017; Bińkowski et al. 2018; Wang, Sun, and Halgamuge 2018; Santos et al. 2019; Ren, Luo, and Zhu 2021).

By combining the EP algorithm with the moment-matching score, we propose a stable algorithm that finds a subnetwork with good generative performance in a very sparse regime. Note that our method can avoid the challenging problem of balancing between training and pruning procedures because it does not involve any weight update. In addition, thanks to the stable characteristic of the moment-matching score, our method can find a Strong Lottery Ticket (SLT) in generative models without bells and whistles. To the best of our knowledge, we are the first to show the existence of strong lottery tickets in generative models and pro-

vide an algorithm to find it stably. Our extensive experiments show that one can find a subnetwork of 10% sparsity while maintaining the generative performance of its dense version (see Figure 1). More surprisingly, we find that our method can also be used to find a well-performing subnetwork in the pretrained generative models. This implies that one can scale down off-the-shelf generative models to have less memory consumption with comparable or even better performance.

Main Contributions. Our contributions can be summarized as follows:

- We show that there exist strong lottery tickets in generative models. By searching for strong lottery tickets via moment-matching scores, we avoid the joint optimization of pruning and training, which is complicated.
- We provide an algorithm that can stably find a good subnetwork in generative models; *i.e.*, one can prune a randomly initialized generative model (without any weight updates) and find a sparse subnetwork that achieves comparable or better performance than the dense, fully trained counterpart.
- We further find that our method can even improve pretrained generative models. Starting from a densely pretrained model, our method can produce its lighter and stronger counterpart in various experimental settings.

Method

This section proposes a simple method to find a strong lottery ticket (SLT) in generative models. The schematic overview of our method is shown in Figure 2. Here, we consider a neural network $G(z; \theta)$ with randomly initialized weights $\theta \in \mathbb{R}^d$. We then aim at finding a strong lottery ticket: a mask $m \in \{0, 1\}^d$ which satisfies that the pruned network $G(z; \theta \odot m)$ performs well on the generative task.

A Simple Algorithm for Finding Strong Lottery Tickets in Generative Models

Edge-popup (EP) algorithm (Ramanujan et al. 2020) is the earliest method to find strong lottery tickets in randomly initialized discriminative networks. With a proper score function, we show that the EP algorithm can be successfully

applied to generative models as well. In the EP algorithm, we first assign a random score s_i for each weight θ_i where $\theta = [\theta_1, \dots, \theta_d]$. Suppose we want to remain $k\%$ of the weights. Then, at each forward path, we sort the score s_i at each layer and assign $m_i = 1$ if $|s_i|$ is in the top $k\%$ within the corresponding layer, and assign $m_i = 0$ otherwise. In each backward path, we compute the loss of the network and update the score s_i by using back-propagation. Here we use straight-through estimator (Bengio, Léonard, and Courville 2013) to handle the indicator function that maps s_i to m_i .

Modeling a Stable Score via Moment-Matching

Now we are pruning generative models, we need to devise a proper score-updating function instead of the cross-entropy loss used for discriminative models. To this end, we utilize a kernel maximum mean discrepancy (MMD) (Gretton et al. 2006, 2012), which is known to give a stable optimization for learning generative models (Li, Swersky, and Zemel 2015; Li et al. 2017; Bińkowski et al. 2018; Wang, Sun, and Halgamuge 2018; Santos et al. 2019; Ren, Luo, and Zhu 2021).

Given two sets of real and fake samples $\{r_i\}_{i=1}^N$ and $\{f_j\}_{j=1}^M$, minimizing the MMD loss $\mathcal{L}_{\mathcal{MMD}}$ can be interpreted as matching all moments of the model distribution to the empirical data distribution:

$$\mathcal{L}_{\mathcal{MMD}} = \left\| \frac{1}{N} \sum_{i=1}^N \phi(r_i) - \frac{1}{M} \sum_{j=1}^M \phi(f_j) \right\|^2, \quad (1)$$

where $\phi(\cdot)$ denotes a function that leads to matching high order moments. Ideally, $\phi(\cdot)$ must be calculated with infinite orders. To compute MMD efficiently, we rephrase the expression (1) via kernel trick:

$$\begin{aligned} \mathcal{L}_{\mathcal{MMD}} &= \frac{1}{N^2} \sum_{i=1}^N \sum_{i'=1}^N \psi(r_i, r_{i'}) - \frac{2}{NM} \sum_{i=1}^N \sum_{j=1}^M \psi(r_i, f_j) \\ &\quad + \frac{1}{M^2} \sum_{j=1}^M \sum_{j'=1}^M \psi(f_j, f_{j'}), \end{aligned} \quad (2)$$

where we use the pretrained VGG network as a fixed kernel ψ and match the mean μ and covariance σ of real and fake sample features in the VGG embedding space:

$$\mathcal{L}_{\mathcal{MMD}} = \sum_{j=1}^L \left(\|\mu_r^j - \mu_f^j\|^2 + \|\sigma_r^j - \sigma_f^j\|^2 \right). \quad (3)$$

We define I_v , w_{uv} , σ , and α as the input of node v , network parameter for node u and node v , activation function, and learning rate, respectively. At time step t , the amount of changes in the score can be expressed as

$$\mathbf{s}_{t+1, \mathbf{uv}} = \mathbf{s}_{t, \mathbf{uv}} - \alpha \frac{\partial \mathcal{L}_{\mathcal{MMD}}}{\partial I_v} w_{uv} \sigma(I_u). \quad (4)$$

It is worth noting that our method uses the MMD loss for finding nodes of low importance, not for learning weights.

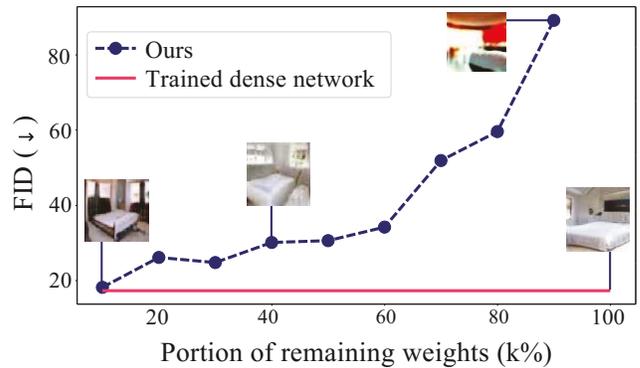


Figure 3: Comparison of FID scores of the subnetworks and the trained dense network (GFMN; LSUN-Bedroom). Recall that our method prunes a randomly initialized neural network without any weight update. Here, we visualize the FIDs for various k , which is the portion (%) of the remaining weights in the pruned subnetwork.

Experimental Evaluation

In this section, we provide empirical results on the proposed pruning method. First, we show that our method finds strong lottery tickets in randomly initialized generative models. Second, we show that our pruning method can be used to lighten pretrained generative models. Finally, we demonstrate that the strong lottery tickets in generative models found by our method are not fine-tunable to reach better performance, similar to the observation made for discriminative models (Ramanujan et al. 2020).

Datasets. We use LSUN Bedroom (Yu et al. 2015), FFHQ (Karras, Laine, and Aila 2019), CIFAR-10 (Krizhevsky, Hinton et al. 2009), CelebA (Liu et al. 2015), and BabyImageNet (Kang, Shin, and Park 2022) datasets. Image resolution is set to 64×64 for every dataset.

Baselines. Following the setup of the generative feature matching network (GFMN) (Santos et al. 2019), we adopt the ResNet-based architecture as our default generator that serves as a dense model. We also use other off-the-shelf pretrained generative models (BigGAN, SNGAN) trained on BabyImageNet, whose weights are provided in the official StudioGAN (Kang, Shin, and Park 2022) code¹. The model setup is configured by the codebase implemented for reproducible GANs.

Evaluation metrics. We evaluate the visual quality and the diversity of generated images with Fréchet Inception Distance (FID) (Heusel et al. 2017), Precision & Recall (Kynkäänniemi et al. 2019), and Density & Coverage (Naeem et al. 2020), where we use InceptionV3 as the evaluation backbone model (Szegedy et al. 2016). Here, we use 10,000 samples of real and generated images, respectively. The details on evaluation metrics and protocols are further described in Supplementary Materials.

¹<https://github.com/postech-cvlab/pytorch-studiogan>

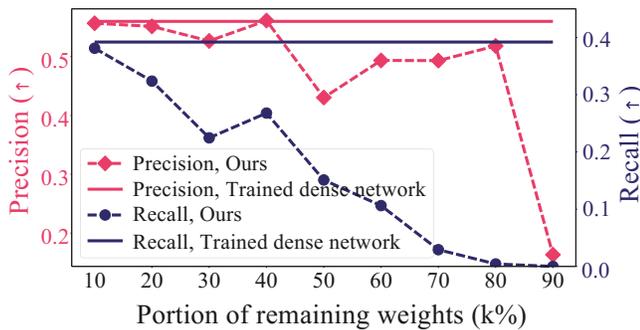


Figure 4: Comparison of Precision & Recall values of the subnetworks and the trained dense network (GFMN; LSUN-Bedroom). The discovered SLT performs well in both Precision & Recall. That is, SLT can generate “various” images of “good quality” without weight training.

	FFHQ	LSUN	CIFAR10	CelebA
Dense Model	11.52	17.32	18.86	9.52
Ours (10%)	13.32	20.21	15.06	10.93

Table 1: Comparison of FID values of trained dense models and strong lottery tickets (SLT). SLT is found by our pruning method on a ResNet-based generator for various datasets. Smaller FID numbers indicate better performance. Here percentage denotes the portion of remaining weights. This shows that we can obtain a decent generative model by pruning 90% of weights in a randomly initialized neural network without any weight update.

Experiment 1: Can We Find Strong Lottery Tickets in Generative Models?

To investigate this question, we need a dense model that serves as the reference. On the one hand, following the setup of GFMN (Santos et al. 2019), we train a ResNet-based generator using the MMD loss. Here, the loss is used for training model weights. On the other hand, we apply our method to find a subnetwork from the generator of the same architecture but with randomly initialized weights. Note that the same MMD loss is used here, but it just serves as a score function to find a subnetwork mask—the loss does not affect the weights. Figure 3 and Figure 4 show how the generative performance of our pruned network changes as a function of k , the portion (%) of remaining weights in the subnetwork. When most of the random weights remain, *e.g.*, $k = 90%$, the pruned network is almost identical to the untrained dense network and thus shows poor generative performance. As k decreases, one can see that the pruned network starts to generate realistic images and achieve its best FID values at around $k = 10%$. The similar trend is observed in Precision & Recall (see Figure 4).

In Table 1, we compare the FID values of the trained dense models and the strong lottery tickets (SLT) that are obtained by our method (using $k = 10%$ of the weights). This indicates that in a randomly initialized dense network, there is a subnetwork with similar or better performance than

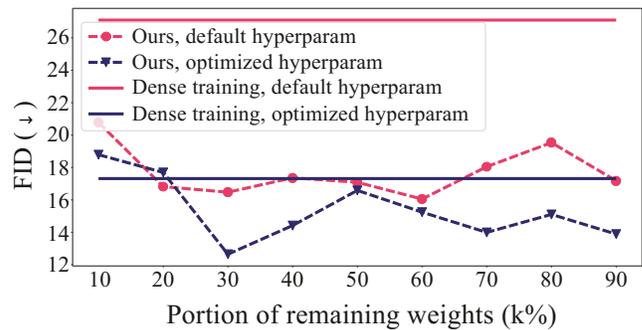


Figure 5: The impact of applying our method on pretrained models (GFMN; LSUN-Bedroom). Here, we test two pre-trained models: (1) one using the default hyperparameter (solid red line) (Santos et al. 2019); (2) the other using a more optimized hyperparameter found from our experiments (solid blue line). Each dashed line shows the performance of the subnetwork obtained by our pruning method for each pretrained model with different target k . Our pruning method finds a subnetwork that performs better than the corresponding pretrained dense model in various sparsity.

the trained dense model while having only 10% of the total number of parameters in the dense network. Note that in the previous literature (Ramanujan et al. 2020), Ramanujan et al. (2020) found the best performing SLT for discriminative networks in the $k = 50%$ region but failed to obtain SLT in a sparser region like $k = 10%$. In contrast, SLTs found by our method for generative models achieve their best performance when $k = 10%$. This difference of discriminate/generative models in the optimal sparsity regime for finding SLTs is an interesting observation, which can be further analyzed in future work.

Experiment 2: Can We Use Our Method to Lighten the Pretrained Models?

Recall that one of the biggest questions in the literature on model compression is whether we can find a good subnetwork within a fully trained model without losing performance (LeCun, Denker, and Solla 1989; Han, Mao, and Dally 2016; Wiedemann et al. 2020; Isik, No, and Weissman 2021). Focusing on this fundamental question that has been discussed over decades, we investigate whether our pruning method can provide a positive answer for this question in generative models.

Figure 5 shows the performance of our pruning method when applied to dense GFMN models trained on the LSUN dataset. Two pretrained models are considered in this experiment: (1) the model trained with the default hyperparameter suggested by (Santos et al. 2019), and (2) the model trained with the optimized hyperparameter found in our experiments. In both dense pretrained models, applying our pruning method maintains or even improves the performance when the portion of remaining weights is chosen within 10% – 90% regime. This shows the practical importance of our method in that an off-the-shelf generative model (that already has reasonable performance) can be lightened

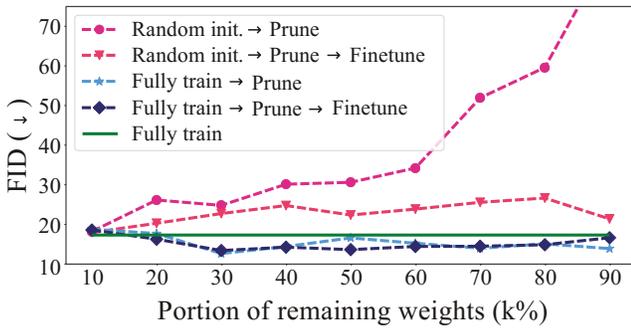


Figure 6: The impact of fine-tuning on subnetworks (GFMN; LSUN-Bedroom). We show the FID performances for two scenarios: (1) when the dense model is randomly initialized; (2) when the dense model is fully trained.

to achieve 10x efficiency while having similar or better performance compared with the dense pretrained model.

A natural follow-up question is whether having a better performance at the dense model implies having a better performance after applying our pruning method. The examples in Figure 5 show this is true in our experimental setting. However, exploring the answer to this question in various settings is out of the scope of this paper. We leave this as future work.

Experiment 3: What Happens When We Further Train a Strong Lottery Ticket?

The pioneering work (Ramanujan et al. 2020) on finding strong lottery tickets (SLTs) on discriminative networks had an interesting observation: SLTs found in their work are not fine-tunable, *i.e.*, the performance of SLT does not improve even after weight training. We test whether this observation is true in SLTs found in generative models by fine-tuning the subnetwork found by our algorithm.

Figure 6 compares performances of the subnetwork (obtained by our pruning method) before and after the fine-tuning. We test on two subnetworks: one obtained from a randomly initialized network, and the other one obtained from the network having fully trained weights. For the subnetwork obtained from a randomly initialized network, fine-tuning the network improves the performance in all sparsity regimes except when the portion of remaining weights is 10%. The performance of the subnetworks obtained from a fully trained network does not improve even after fine-tuning in all sparsity regimes. From these experiments, one can confirm that subnetworks that already achieve the performance of the fully trained dense model cannot be improved by fine-tuning the survived weights. This coincides with the observation made in (Ramanujan et al. 2020).

Discussions

Here, we provide some interesting discussion topics on finding strong lottery tickets in generative models.

Factor analysis The generative performance of the subnetwork found by our method depends on multiple factors.

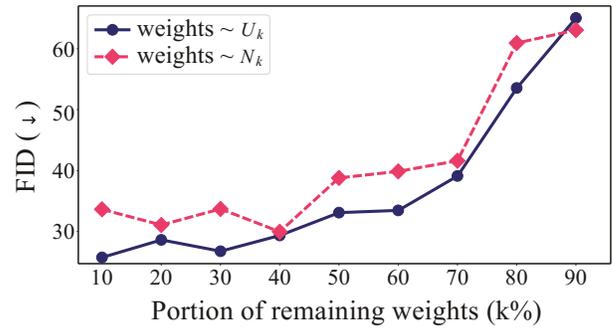


Figure 7: The impact of various initializations (GFMN; LSUN-Bedroom). We visualize the performance of our method according to different weight initializations: Kaiming normal N_k and signed Kaiming constant U_k . The latter option shows better performance overall.

Ch. Multiplier (n)	0.4	1.0	1.4	1.6	1.8	2.0
FID (↓)	26.96	10.93	10.12	8.83	8.97	8.31

Table 2: The impact of channel width on the performance of subnetworks (GFMN; CelebA). The performance improves as the channel width increases.

The first factor is how we initialize weights of a neural network. Figure 7 compares two random weight initialization methods: “Kaiming normal (N_k)” and “signed Kaiming constant (U_k)”. One can confirm that in most sparsity regimes, the signed Kaiming constant has better performance (lower FID value). This observation is consistent with the results in (Ramanujan et al. 2020) for discriminative models. Intuitively, weight initialization can be considered important because subnetwork search space varies depending on weight initialization. However, understanding which weight initialization can perform well still requires deeper research on the structure of neural networks.

The second factor is the channel width n of the network. The default network is denoted by $n = 1$, and we test on various networks having n times larger channel width at each layer. Table 2 shows the performance of a strong lottery ticket found by our method for various n . One can confirm that as channel width increases, the subnetwork’s performance improves. This observation makes sense: as the width increases, the number of subnetworks in a randomly initialized network increases exponentially; the probability of finding a better subnetwork therefore increases.

Does it have to be the EP algorithm and the moment-matching score? Because our method is the first algorithm to address prune-at-initialization in generative models, there is no base method for comparing. A possible naive alternative to our method would be random pruning; some might wonder if it has to be the EP algorithm and if it is also possible with random pruning. To address this point, we compare random pruning and our method and show that our results are not something that can be achieved by mere

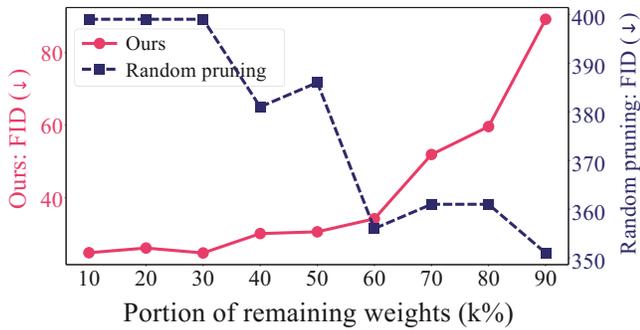


Figure 8: Performance comparison of random pruning versus our pruning method (GFMN; LSUN-Bedroom). The performance gap in FID between ours and random pruning is huge. Unlike our method, random pruning fails to find SLT and loses performance as the portion of remaining weights gets decreased.

chance (see Figure 8). While our method finds a sufficiently meaningful subnetwork that produces realistic images, random pruning does not find such a subnetwork.

Another natural follow-up question would be whether we can substitute the MMD score function with the adversarial loss, which is commonly used for training modern generative models. In our experiments, we find that when using the adversarial loss, network pruning simply fails due to its notorious instability as the score keeps changing (see Figure 9). Here, the subnetwork that we obtained via adversarial loss suffers from unrealistic image generation and mode-collapse phenomena.

Efficient multi-domain generation One nice property of our method is that it enables efficient multi-domain generation. Although several studies have proposed a model for multi-domain generation, all require a specific architecture to do so (Liu et al. 2019; Choi et al. 2020; Baek et al. 2021). On the other hand, since our method only finds a subnetwork from the randomly initialized generator, one can use the same architecture and perform multi-domain generation simply by changing the mask found a priori. Unlike the other methods, our framework does not require any modification in the architecture to add a new domain to generate; one can simply find another mask, which is more efficient than developing a new model or fine-tuning the model architecture to a new domain. All results in Figure 10 on various domains (FFHQ, LSUN, CIFAR-10, and CelebA) are generated by simply changing the mask to the generator with the same weights. In Figure 10 (c), we show that the GFMN generators pruned via our method show decent performance in various datasets with a very small number of parameters.

Can we improve the performance of our method? In this paper, we shed light on the potential of finding strong lottery tickets in generative models by using the edge-popup (EP) algorithm and the MMD loss. Here, we discuss possible ways of improving the performance of our method. Note that a recent work (Sreenivasan et al. 2022) on pruning discriminative networks found that there are two methods to



Figure 9: Visualization of generated images with the adversarial loss (BigGAN, BabyImageNet). Here, we leverage the adversarial loss to obtain scores for selecting a subnetwork mask. The resultant generator fails to generate diverse images and gets mode-collapsed.

improve the performance of EP: (1) using global EP (pruning weights by sorting the scores globally) instead of vanilla EP (pruning weights by sorting the scores at each layer), and (2) using gradual pruning (moving from dense regime to sparse regime gradually during pruning) instead of vanilla EP which moves to the sparse regime from the beginning. Inspired by this observation, we expect applying EP with these two variants (global pruning and gradual pruning) in our method has the potential to improve the performance of the SLT in generative models.

Are there strong lottery tickets in high resolution? To explore SLT at high resolution, we apply our method for 128 resolution and observe that the trend remains the same. That is, we can also find SLT in high-resolution generative models. We include this result in Appendix E. Although we can find SLT at high resolution, we observe that our method converges slowly due to high-dimensional feature maps, and the performance of the dense network is not satisfactory. We can improve performance by investigating stronger kernels (*i.e.*, feature extractors), but this is beyond the scope of this paper and is an interesting future research direction.

Why is pruning generator challenging? While pruning discriminative models has shown remarkable results, pruning generators still has various challenges: 1) There are no obvious criteria. Unlike supervised learning which has labels, it is difficult to provide clear criteria for what to prune in generative models. 2) Training is unstable. Many studies have shown that pruning generators severely reduces performance due to training instability. 3) Generative models are mostly decoder structures. Pruning the weights in the decoder can have a more significant impact on the final output than the encoder due to the expanded output space. In this paper, our framework provides obvious criteria and obtains pruned generators stably. Appendix D contains the results showing the stability of our method.

Related Work

Neural Network Pruning. Conventional iterative train-prune-retrain framework incurs massive *training costs*, even

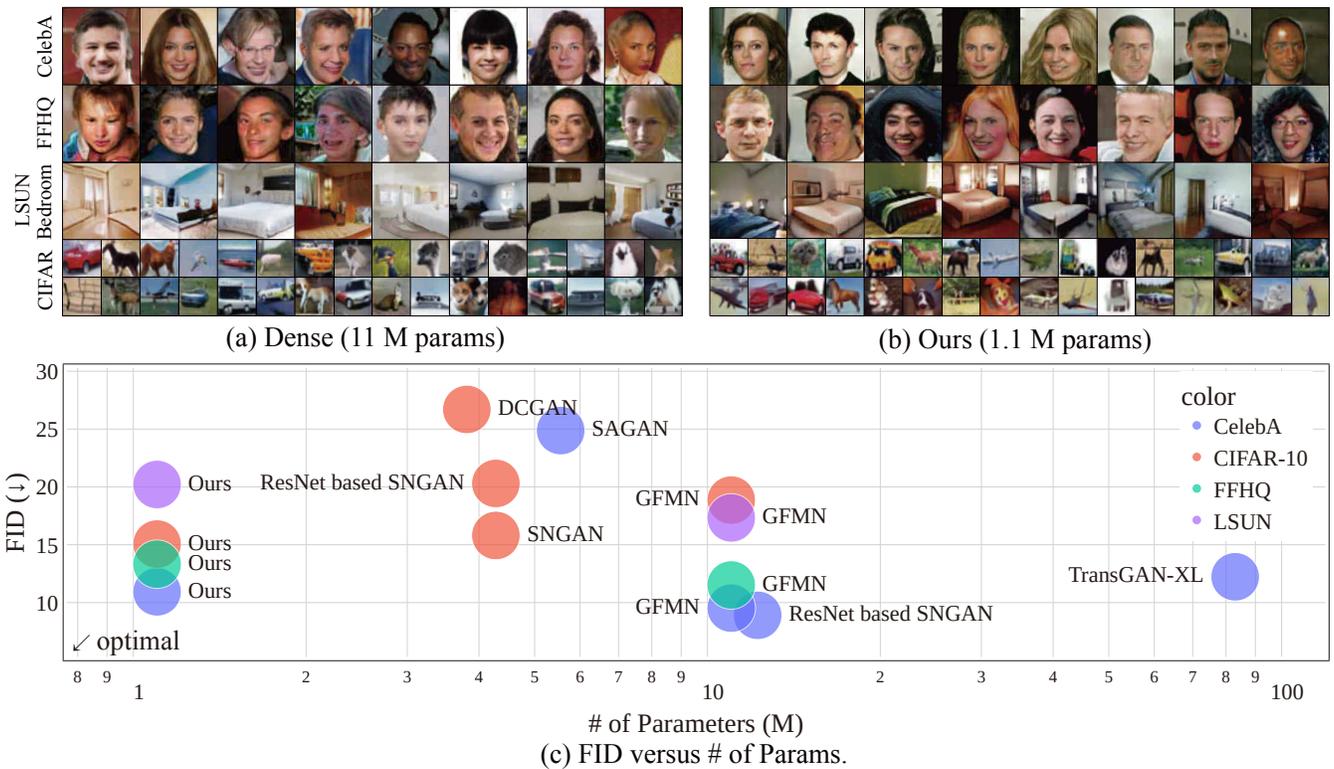


Figure 10: Qualitative and quantitative comparison (GFMN; CelebA, FFHQ, and CIFAR-10). Qualitative results for trained dense model (top left) and ours (top right). Ours can find SLT of good image quality and diversity. We also show that the GFMN generators pruned via ours show decent performance with a very small number of parameters (bottom). Note that the left bottom is the optimal point, where both the FID value and the number of parameter are small.

though it significantly reduces computational costs at test time. To address this issue, two categories of network pruning methods have been suggested in recent years: (1) pruning a random network in a way that the subnetwork is trainable to have a good performance, and (2) pruning a random network in a way that the subnetwork itself is having good performance without any weight updates. A subnetwork obtained from the former category is called *weak* lottery ticket (Frankle and Carbin 2018) and a subnetwork in the latter category is called *strong* lottery ticket (Ramanujan et al. 2020). There have been extensive works on developing theories and algorithms on weak/strong lottery tickets in discriminative networks (Frankle and Carbin 2018; Frankle et al. 2020; Lee, Ajanthan, and Torr 2018; Chen et al. 2020), but results on generative models were limited so far.

Compressing generative models. Some recent works focus on finding weak lottery tickets for obtaining lightweight generative models. For GANs and VAEs, the authors of (Kalibhat, Balaji, and Feizi 2021; Chen et al. 2021) used iterative magnitude pruning (Frankle and Carbin 2018). However, unlike *strong* lottery tickets, *weak* lottery tickets require additional weight updates for achieving reasonable performance. *Weak* lottery tickets show slow convergence speed compared to *strong* lottery tickets due to additional

weight updates (see Appendix D). Our work differs from these works in two perspectives: (1) we focus on finding strong lottery tickets that perform well without any weight update, and (2) we do not rely on GAN loss, thus finding a good subnetwork stably. To the best of our knowledge, the present paper is the first work that shows the existence of a strong lottery ticket in generative models.

Conclusion

In this paper, we investigated strong lottery tickets (SLT) in generative models. While all existing works on building lightweight generative models suffer from huge weight update costs, performance degradation, or complicated training, we circumvented these problems by searching for SLT. To the best of our knowledge, we are the first to show the existence of SLT in generative models; SLT was previously observed only in discriminative models. By exploiting the moment-matching approach for scoring important weights in a randomly initialized generator, our framework finds SLT stably without bells and whistles. Our experimental results showed that our method could successfully find a sparse subnetwork in various datasets, and the discovered subnetwork achieved similar or even better performance than the trained dense model even when only 10% of the weights remained.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2.220574.01), Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-01336, Artificial Intelligence Graduate School Program (UNIST), No.2021-0-02068, Artificial Intelligence Innovation Hub, No.2022-0-00959, (Part 2) Few-Shot Learning of Causal Inference in Vision and Language for Decision Making), No.2022-0-00264, Comprehensive Video Understanding and Generation with Knowledge-based Deep Logic Neural Network).

References

- Baek, K.; Choi, Y.; Uh, Y.; Yoo, J.; and Shim, H. 2021. Rethinking the truly unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14154–14163.
- Bengio, Y.; Léonard, N.; and Courville, A. 2013. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*.
- Bińkowski, M.; Sutherland, D. J.; Arbel, M.; and Gretton, A. 2018. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*.
- Brock, A.; Donahue, J.; and Simonyan, K. 2018. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.
- Chen, T.; Frankle, J.; Chang, S.; Liu, S.; Zhang, Y.; Wang, Z.; and Carbin, M. 2020. The lottery ticket hypothesis for pre-trained BERT networks. *arXiv preprint arXiv:2007.12223*.
- Chen, X.; Zhang, Z.; Sui, Y.; and Chen, T. 2021. GANs can play lottery tickets too. *arXiv preprint arXiv:2106.00134*.
- Choi, Y.; Uh, Y.; Yoo, J.; and Ha, J.-W. 2020. StarGAN v2: Diverse Image Synthesis for Multiple Domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Frankle, J.; and Carbin, M. 2018. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*.
- Frankle, J.; Dziugaite, G. K.; Roy, D.; and Carbin, M. 2020. Linear mode connectivity and the lottery ticket hypothesis. In *International Conference on Machine Learning*, 3259–3269. PMLR.
- Gretton, A.; Borgwardt, K.; Rasch, M.; Schölkopf, B.; and Smola, A. 2006. A kernel method for the two-sample problem. *Advances in neural information processing systems*, 19.
- Gretton, A.; Borgwardt, K. M.; Rasch, M. J.; Schölkopf, B.; and Smola, A. 2012. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1): 723–773.
- Han, S.; Mao, H.; and Dally, W. J. 2016. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. *arXiv:1510.00149 [cs]*.
- Han, S.; Pool, J.; Tran, J.; and Dally, W. 2015. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28.
- Hassibi, B.; and Stork, D. 1992. Second order derivatives for network pruning: Optimal brain surgeon. *Advances in neural information processing systems*, 5.
- Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.
- Hou, L.; Yuan, Z.; Huang, L.; Shen, H.; Cheng, X.; and Wang, C. 2021. Slimmable generative adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7746–7753.
- Isik, B.; No, A.; and Weissman, T. 2021. Rate-Distortion Theoretic Model Compression: Successive Refinement for Pruning. *arXiv preprint arXiv:2102.08329*.
- Kalibhat, N. M.; Balaji, Y.; and Feizi, S. 2021. Winning lottery tickets in deep generative models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8038–8046.
- Kang, M.; Shin, J.; and Park, J. 2022. StudioGAN: A Taxonomy and Benchmark of GANs for Image Synthesis. *arXiv preprint arXiv:2206.09479*.
- Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401–4410.
- Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8110–8119.
- Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images. In *Tech Report*. Citeseer.
- Kynkäänniemi, T.; Karras, T.; Laine, S.; Lehtinen, J.; and Aila, T. 2019. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems*, 32.
- LeCun, Y.; Denker, J.; and Solla, S. 1989. Optimal brain damage. *Advances in neural information processing systems*, 2.
- Lee, N.; Ajanthan, T.; and Torr, P. H. 2018. Snip: Single-shot network pruning based on connection sensitivity. *arXiv preprint arXiv:1810.02340*.
- Li, C.-L.; Chang, W.-C.; Cheng, Y.; Yang, Y.; and Póczos, B. 2017. Mmd gan: Towards deeper understanding of moment matching network. *Advances in neural information processing systems*, 30.
- Li, M.; Lin, J.; Ding, Y.; Liu, Z.; Zhu, J.-Y.; and Han, S. 2020. GAN compression: Efficient architectures for interactive conditional gans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5284–5294.
- Li, S.; Wu, J.; Xiao, X.; Chao, F.; Mao, X.; and Ji, R. 2021. Revisiting discriminator in GAN compression: A generator-discriminator cooperative compression scheme. *Advances in Neural Information Processing Systems*, 34: 28560–28572.

- Li, Y.; Swersky, K.; and Zemel, R. 2015. Generative moment matching networks. In *International conference on machine learning*, 1718–1727. PMLR.
- Liu, M.-Y.; Huang, X.; Mallya, A.; Karras, T.; Aila, T.; Lehtinen, J.; and Kautz, J. 2019. Few-shot unsupervised image-to-image translation. In *Proceedings of the IEEE International Conference on Computer Vision*, 10551–10560.
- Liu, Y.; Shu, Z.; Li, Y.; Lin, Z.; Perazzi, F.; and Kung, S.-Y. 2021. Content-aware gan compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12156–12166.
- Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, 3730–3738.
- Naeem, M. F.; Oh, S. J.; Uh, Y.; Choi, Y.; and Yoo, J. 2020. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning*, 7176–7185. PMLR.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 8748–8763. PMLR.
- Ramanujan, V.; Wortsman, M.; Kembhavi, A.; Farhadi, A.; and Rastegari, M. 2020. What’s hidden in a randomly weighted neural network? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11893–11902.
- Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; and Sutskever, I. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, 8821–8831. PMLR.
- Ren, Y.; Luo, Y.; and Zhu, J. 2021. Improving Generative Moment Matching Networks with Distribution Partition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 9403–9410.
- Santos, C. N. d.; Mroueh, Y.; Padhi, I.; and Dognin, P. 2019. Learning implicit generative models by matching perceptual features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4461–4470.
- Sreenivasan, K.; Sohn, J.-y.; Yang, L.; Grinde, M.; Nagle, A.; Wang, H.; Lee, K.; and Papailiopoulos, D. 2022. Rare Gems: Finding Lottery Tickets at Initialization. *arXiv preprint arXiv:2202.12002*.
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.
- Tuli, S.; Tuli, S.; Casale, G.; and Jennings, N. R. 2021. Generative optimization networks for memory efficient data generation. *arXiv preprint arXiv:2110.02912*.
- Wang, H.; Gui, S.; Yang, H.; Liu, J.; and Wang, Z. 2020. GAN slimming: All-in-one gan compression by a unified optimization framework. In *European Conference on Computer Vision*, 54–73. Springer.
- Wang, W.; Sun, Y.; and Halgamuge, S. 2018. Improving MMD-GAN training with repulsive loss function. *arXiv preprint arXiv:1812.09916*.
- Wiedemann, S.; Kirchhoffer, H.; Matlage, S.; Haase, P.; Marban, A.; Marinč, T.; Neumann, D.; Nguyen, T.; Schwarz, H.; Wiegand, T.; et al. 2020. Deepcabac: A universal compression algorithm for deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 14(4): 700–714.
- Yu, F.; Zhang, Y.; Song, S.; Seff, A.; and Xiao, J. 2015. LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop. *arXiv preprint arXiv:1506.03365*.