Diversity-aware Channel Pruning for StyleGAN Compression

Jiwoo Chung, Sangeek Hyun, Sang-Heon Shim, Jae-Pil Heo*
Sungkyunkwan University
{wldn0202, hsi1032, ekzmwww, jaepilheo}@g.skku.edu

Abstract

StyleGAN has shown remarkable performance in unconditional image generation. However, its high computational cost poses a significant challenge for practical applications. Although recent efforts have been made to compress StyleGAN while preserving its performance, existing compressed models still lag behind the original model, particularly in terms of sample diversity. To overcome this, we propose a novel channel pruning method that leverages varying sensitivities of channels to latent vectors, which is a key factor in sample diversity. Specifically, by assessing channel importance based on their sensitivities to latent vector perturbations, our method enhances the diversity of samples in the compressed model. Since our method solely focuses on the channel pruning stage, it has complementary benefits with prior training schemes without additional training cost. Extensive experiments demonstrate that our method significantly enhances sample diversity across various datasets. Moreover, in terms of FID scores, our method not only surpasses state-of-the-art by a large margin but also achieves comparable scores with only half training iterations. Codes are available at github.com/jiwoogit/DCP-GAN.

1. Introduction

Thanks to recent progress in generative artificial intelligence, there have been numerous techniques in the research field of image generation [1, 3, 10, 15, 18, 26, 30, 31, 33, 36]. Especially for Generative Adversarial Networks (GANs) [5], StyleGAN [12–14] is one of the most successful approaches and leads to the development of diverse applications of GANs including image editing [7, 27, 39], super resolution [29], and even 3D generation [2, 6, 32]. However, due to their high computation costs, the deployment of these applications on the edge devices such as mobile phones and embedded systems is still challenging and remains a significant research problem. To handle this challenge, the research area of StyleGAN compression [24, 37] is recently introduced.

Most of prior StyleGAN compression techniques including CAGAN [24] and StyleKD [37] have two stages: (1) channel-pruning and (2) distillation stages. In detail, the channel-pruning stage initializes the compressed model (student) from the pre-trained network (teacher) by selectively removing channels. Subsequently, in the distillation stage, channel-pruned student model undergoes further training by adversarial objectives and knowledge distillation from the teacher model if needed.

The primary goal of the GAN compression is to maintain the diversity and fidelity of generated images from the teacher model. However, we observe that the compressed generator often struggles to preserve the diversity compared to teacher generator (i.e. low recall as reported in Table. 1). Specifically, in the case of StyleKD, although it retains the latent space and mapping network of the teacher model during the pruning stage, the synthesis network of student model hardly preserves the diversity (recall) in the latent-to-image rendering process. We hypothesize that such degradation in diversity is caused by improper initialization of the synthesis network (e.g. random initialization or inadequate pruning method). Hence, in this paper we focus on development of an appropriate channel pruning scheme for synthesis network toward preserving the diversity of teacher.

The diversity of the synthesis network stems from the variations observed in images generated from different latent signals. Our hypothesis is that the channel-wise behavior differs for different latent vectors, implying that each channel contributes to the diversity of generated images differently. For instance, channels associated with common characteristics in the dataset (e.g., the number of eyes in a facial dataset) may not be sensitive to the latent vector and its impact on diversity, despite their significant contribution to the generated images. On the other hand, channels related to semantic attributes exhibit distinct behavior with respect to the latent vector. Furthermore, the common characteristics within a dataset can be highly recoverable during further training even if their relevant channels are pruned. Therefore, to achieve a compressed student model

* Corresponding author

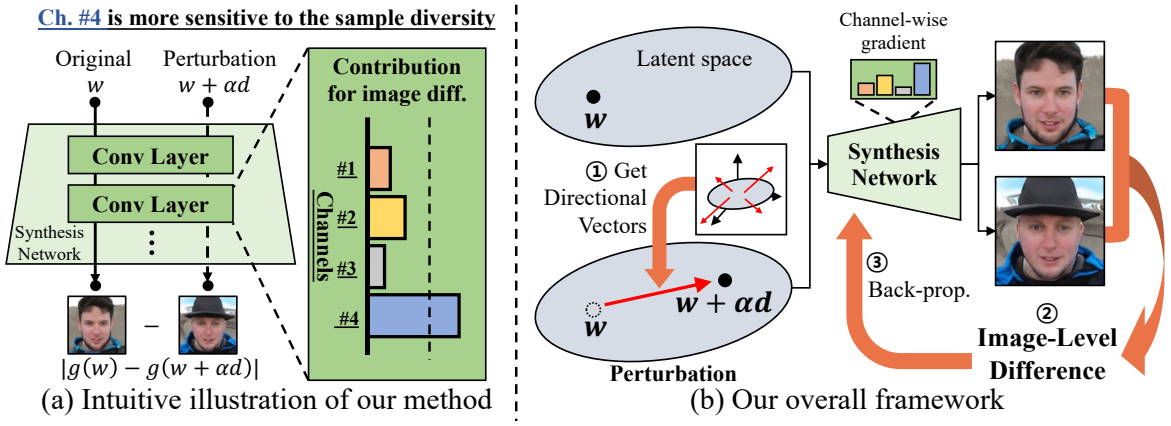


Figure 1. **(a) Intuitive illustration of our method.** We compare four channels (Ch. #1, 2, 3, 4) by evaluating their responses when we pass the same latent vector w and its perturbed counterpart $(w + \alpha d)$. By investigating the contribution of each channel to resulting image difference, we determine the sensitivity of channels to the latent perturbation. In this example, Ch. #4 is highly sensitive to the perturbation, while Ch. #1, 2, 3 exhibit low sensitivity. Consequently, in terms of preserving sample diversity, Ch. #4 is suitable for retaining. **(b) Our overall framework.** We aim to assess the contribution of each channel to the sample diversity by measuring its sensitivity to latent vector perturbation. In detail, 1) we sample a directional vector for the perturbation, 2) we compute the image-level difference caused by the latent vector perturbation, and 3) we calculate channel-wise gradient magnitudes induced by the difference image. The channel-wise sensitivity to the sample diversity is determined by its gradient magnitudes. As a result, we can estimate the channel-wise sensitivity against diversity.

that maintains a highly similar generative distribution to the teacher model with fewer channels, it is reasonable to prioritize the preservation of channels that are sensitive to the latent vector.

In this paper, we present a channel-pruning method that considers the sensitivity of each channel to latent vector perturbations. Specifically, we investigate gradients induced by the image-level difference between two generated samples: one from the original latent vector and the other from its perturbated counterpart, as shown in Fig. 1. The intuition behind this is that a larger gradient magnitude for a generator's parameter indicates a higher sensitivity to the latent vector perturbation. Building on this primitive idea, we define a channel-wise importance score by aggregating the gradient magnitudes of parameters within the channel for diverse latent vectors and their perturbations. This score serves as a measure of each channel's contribution to the sample diversity. Based on these importance scores, we prune channels accordingly.

We conduct extensive evaluations in various datasets including FFHQ, LSUN-Church, and LSUN-Horse. The experimental results demonstrate that our compressed model not only achieves state-of-the-art performance but also preserves significantly higher sample diversity. Moreover, our method reaches the previous state-of-the-art FID score with only half the number of training iterations.

2. Related Work

One of the network compression techniques, channel pruning [19, 22, 23, 25, 34, 38, 40] proves highly effective in significantly reducing both memory usage and computational expenses. For example, MeanGrad [22] aims to preserve model performance despite pruning in network parameters and evaluates the importance of convolution weights using the mean gradient criterion in classification tasks.

In generation tasks, various methods [9, 11, 20, 21, 28] have been also proposed to compress the generator network in both conditional and unconditional GANs to leverage its generative capabilities on devices with limited computational resources. One of the recent works, CAGAN [24] proposed a two-stage compression method for unconditional GANs. They first pruned the network parameters and then fine-tuned the pruned student model using knowledge distillation from the teacher model. Additionally, they proposed a content-aware pruning technique to preserve channels that are highly activated in salient regions of images. More recently, StyleKD [37] addressed the issue of output discrepancy between teacher and student models by maintaining the mapping network while randomly initializing the synthesis network with a reduced number of parameters.

Our approach revisits the pruning stage in compressing the generator of StyleGAN. Specifically, we introduce a novel algorithm for selecting channels to be pruned based on their contributions to the sample diversity measured by the sensitivity to latent vector perturbations. Additionally, our method requires no extra training costs during the pruning stage, as it operates without the need for additional supervision, such as a pre-trained semantic segmentation model, while CAGAN requires it. In StyleKD, despite discarding all parameters in the synthesis network and utilizing latent directions to mimic the teacher model in the distillation stage, we observe that they still face challenges

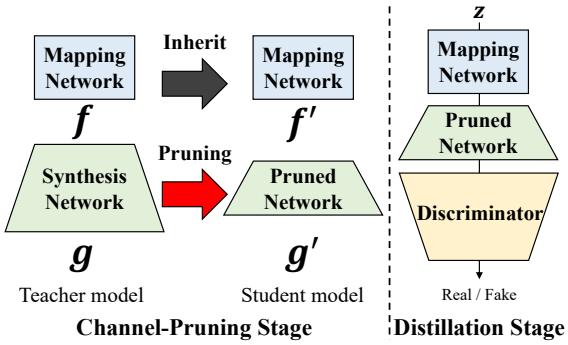


Figure 2. **Overall compression framework.** Overall compression framework consists of two stages: channel pruning and distillation. Channel-pruning stage initializes a compact student model by pruning channels of a larger teacher model. Specifically for the StyleGAN2 architecture, pruning process usually focuses on reducing channels in the synthesis network, while retraining the mapping network. Distillation stage further trains student model with several training objectives such as adversarial and distillation losses. In this paper, we focus on the channel-pruning stage.

in maintaining sample diversity. In contrast, our proposed method directly retains diversity-aware parameters in the synthesis network while discarding diversity-unrelated parameters in the pruning stage. Thus, our approach effectively inherits the generative capabilities of the teacher network and is twice as fast as the baselines thanks to our enhanced initialization.

3. Method

In this paper, we present a novel channel pruning strategy for compressing StyleGAN. we focus on preserving sample diversity by considering the relationship between the latent and image spaces of the GAN, which has been overlooked by previous methods in the pruning process. Before introducing details of our proposed method, we provide an overview of the typical pipeline of GAN compression. Subsequently, we describe our channel selection process.

3.1. Preliminaries

Unconditional GAN Compression. As in Fig 2, GAN compression methods typically involve two networks: a teacher model (large and pretrained) and a student model (small). The goal is to train the student model with fewer parameters by distilling the knowledge from the pretrained teacher network, while maintaining performance as much as possible. This approach generally consists of two stages: (1) channel pruning and (2) distillation. In the channel pruning stage, the less important channels of the teacher model are pruned to initialize the student model. In the distillation stage, the student model is further trained using both adversarial learning and distillation losses from the teacher.

Channel-Pruning Stage. The channel-pruning stage involves the estimation of channel importance within the generator network, followed by pruning of less important channels based on the determined importance scores. In the context of StyleGAN compression, since most of methods are built on the StyleGAN2 framework [13], we provide a detailed description of the pruning procedure. The generator architecture of StyleGAN2 consists of a mapping network, denoted as $f(\cdot) : \mathcal{Z} \mapsto \mathcal{W}$, and a synthesis network, denoted as $g(\cdot) : \mathcal{W} \mapsto \mathcal{I}$. Here, $\mathcal{Z} \in \mathbb{R}^{512}$ represents the input noise space following a standard normal distribution $\mathcal{N}(0, 1)$. $\mathcal{W} \in \mathbb{R}^{512}$ corresponds to the intermediate latent space, and $\mathcal{I} \in \mathbb{R}^{h \times w \times 3}$ represents the image space. In the case of StyleGAN2, the focus of pruning is primarily on reducing the channels within the synthesis network while maintaining the mapping network. This strategy is guided by two reasons. Firstly, the mapping network accounts for only a small portion of the parameters and floating-point operations (FLOPs) compared to the synthesis network, making the benefits of pruning the mapping network negligible. Secondly, previous work [37] has observed that pruning the mapping network results in a significant drop in performance as it disrupts the consistency of output images between teacher and student models. Therefore, most of the existing research in this field has concentrated on developing techniques to prune the synthesis network. In alignment with this approach, our paper also focuses on pruning the synthesis network.

3.2. Our Approach

As discussed earlier, we observed a notable degradation in sample diversity in the compressed student model compared to the teacher network, even though the mapping network remains uncompressed. We hypothesize that a critical factor in this decline is the lack of consideration for sample diversity in previous network initialization (or channel pruning). Consequently, our primary goal is to address this issue by introducing a novel channel-pruning method aimed at enhancing the diversity of generated images.

To accomplish this, we introduce the concept of latent perturbation-induced gradients. These gradients reflect the sensitivity of each channel to perturbations in the latent vector. Since latent perturbations produce the image-level differences of generated samples, the gradients induced by these perturbations enable us to measure the contribution of each channel to the sample diversity. Based on these gradients, we define diversity-sensitive importance scores for each channel using various latent vectors and their perturbations. Finally, we select channels to be pruned according to their importance scores.

Note that, while we explain the method using a single convolution weight W for simplicity, in practice, our pruning method is applied to all weights in convolutional layers.

3.2.1 Latent Perturbation-induced Gradients

In our context, sample diversity refers to the variations observed among images generated from different latent vectors. To assess the contribution of each channel to sample diversity, we examine the gradients induced by the image-level difference between two generated samples from the latent vector and its perturbation. Specifically, we consider the latent representations in the intermediate latent space $\mathcal{W}$ inherited from the teacher model.

To produce the perturbated latent vectors, we first establish a set of independent directional vectors D in $\mathcal{W}$. The perturbation is accomplished by shifting the latent vector w along a directional vector d sampled from D as $w + \alpha d$, where α is a scalar constant. Among a variety of feasible strategies to define D, we specifically employ principal components inspired by GANSpace [7]. We also use a probabilistic sampling for the selection of directional vectors, with the selection probabilities proportional to the variance ratios corresponding principal components, since directions represent different degrees of semantic variation in this scheme. Note that, we experimentally found that our algorithm also works effectively even with a set of randomly drawn directional vectors (i.e. ($d_{\text{rand}} \sim \mathcal{N}(0, 1)$), and the performance gap between principal components and random vectors is not significant.

Given a latent vector w and a sampled directional vector d, we compute the difference between two images $g(w)$ and $g(w + \alpha d)$ generated by the teacher generator $g(\cdot)$ based on L_1 distance as follows:

$$\mathcal{L}_{\text{diff}} = |g(w) - g(w + \alpha d)|, \quad g(\cdot) \in \mathbb{R}^{h \times w \times 3}, \tag{1}$$

where $|\cdot|$ denotes the absolute value, and h and w are the height and width of the images, respectively.

To investigate the sensitivity of learnable parameters W within a convolution layer of g to the latent vector perturbation, we calculate gradients of weights by back-propagating $\mathcal{L}_{\text{diff}}$. These gradients are referred to as latent perturbation-induced gradients $\mathbf{G}_{\text{perturb}}$, and expressed as follows:

$$\mathbf{G}_{\text{perturb}} = |\frac{\partial \mathcal{L}_{\text{diff}}}{\partial W}|, \quad \mathbf{G}_{\text{perturb}} \in \mathbb{R}^{c^{\text{in}} \times c^{\text{out}} \times 3 \times 3}, \tag{2}$$

where c^{in} and c^{out} represent the number of input channels and output channels of W, respectively.

Intuitively, if a parameter has a larger gradient magnitude, the parameter is considered more sensitive to the difference between samples produced by latent perturbation, thereby contributing more to the sample diversity.

3.2.2 Diversity-Sensitive Importance Score

Based on the latent perturbation-induced gradients $\mathbf{G}_{\text{perturb}}$, we introduce channel-wise diversity-sensitive importance score for pruning.

One straightforward approach is to average the gradient magnitudes over the parameters associated with each channel for multiple latent vectors $w \sim f(z)$ and directions d. We refer this scheme as the average-based importance score, expressed as follows:

$$S^{\mu}(c) = \|\mathbb{E}_z \mathbb{E}_d[\mathbf{G}_{\text{perturb}}]_c\|_1, S^{\mu}(c) \in \mathbb{R}^1, \tag{3}$$

where $c \in [0, c^{\text{in}})$ denotes the channel index, $[\cdot]_c$ represents c^{th} channel of $[\cdot]$, and $\mathbf{G}_{\text{perturb}}$ denotes the gradient magnitude. In practice, we compute $\mathbb{E}_d[\cdot]$ using N directional vectors sampled from D.

If a specific latent vector significantly influences $[\mathbf{G}_{\text{perturb}}]_c$ regardless of its perturbations, the channel would receive a high importance score even if it is not actually sensitive to the perturbations. Consequently, the influence of such channels can greatly disturb desired channel selection.

To address this problem, we revise the average-based score by considering the individual effect of each latent vector. Specifically, for each latent vector, we compute the average gradient magnitudes over N perturbations and utilize them to penalize the induced gradients for the learnable parameters. These average gradient magnitudes for a latent vector are referred to as the gradient offset. As a result, our diversity-sensitive importance score for c^{th} channel, $S^{\sigma}(c)$, is defined as follows:

$$S^{\sigma}(c) = \|\mathbb{E}_z \mathbb{E}_d[[\mathbf{G}_{\text{perturb}} - \mathbb{E}_d[\mathbf{G}_{\text{perturb}}]]^2]_c\|_1, S^{\sigma}(c) \in \mathbb{R}^1, \tag{4}$$

where $\mathbb{E}_d[\mathbf{G}_{\text{perturb}}]$ represents the gradient offset of the original latent vector w, and $[\cdot]^2$ denotes element-wise squaring. Once the diversity-sensitive importance scores are computed, we prune channels with low scores according to the pruning ratio p_r.

Aside from the calibration by the gradient offset, $S^{\sigma}(c)$ is equivalent to the variance of $\mathbf{G}_{\text{perturb}}$ over perturbations. This approach further reduces the impact of small noisy values and grants greater importance to channels that are highly sensitive to a particular direction in the latent space. Such directions can be highly related to specific semantic attributes. Intuitively, it is crucial to retain channels that are highly sensitive to directional vectors in the latent space relevant to certain semantic attributes, such as glasses or age.

3.3. Objective Functions for Distillation Stage

After establishing the initial student model through a channel-pruned teacher generator, we fine-tune the student model using adversarial and knowledge distillation objectives. As our primary focus in this paper lies on the pruning algorithm, we simply follow the training scheme of StyleKD [37] incorporating four objectives: $\mathcal{L}_{\text{GAN}}$, $\mathcal{L}_{\text{rgb}}$, $\mathcal{L}_{\text{lpips}}$, and $\mathcal{L}_{\text{LD}}$.

$\mathcal{L}_{\text{GAN}}$ denotes the minimax objective for adversarial training. Meanwhile, $\mathcal{L}_{\text{rgb}}$ and $\mathcal{L}_{\text{lpips}}$ strive to replicate

teacher-generated images in image and feature spaces, respectively. $\mathcal{L}_{\text{LD}}$, a latent-direction-based relation distillation loss, aims to align the student model's feature similarity matrix to the teachers. More details on these objectives can be found in StyleKD [37]. As a result, our final training objective function, $\mathcal{L}_{\text{final}}$, is given as follows:

$$\mathcal{L}_{\text{final}} = \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}} + \lambda_{\text{rgb}}\mathcal{L}_{\text{rgb}} + \lambda_{\text{lpips}}\mathcal{L}_{\text{lpips}} + \lambda_{\text{LD}}\mathcal{L}_{\text{LD}}, \quad (5)$$

where λ_{GAN}, λ_{rgb}, λ_{lpips}, and λ_{LD} balance loss terms.

4. Experiment

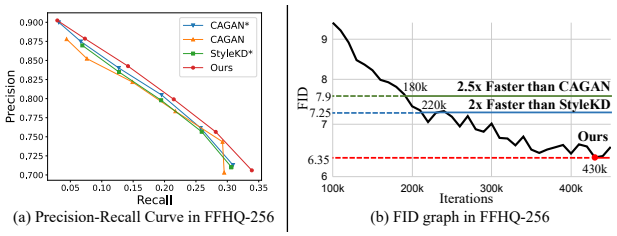
4.1. Experimental Setup

Baselines. Our method is compared with recent StyleGAN compression methods, CAGAN [24] and StyleKD [37]. For a fair comparison, we further train CAGAN with the advanced objective function ($\mathcal{L}_{\text{LD}}$) introduced in StyleKD, which we denote as CAGAN*. Since our method employs StyleKD's training scheme, comparing CAGAN*, StyleKD, and our method is equitable. Since the pretrained StyleKD models are not available, we retrain StyleKD using the official repository for unreported metrics and datasets, referring to this retrained model as StyleKD*. For training details of CAGAN, we apply a facial mask as a content mask when training CAGAN* on the FFHQ dataset. However, on other datasets where content definition is challenging, we use a uniform mask, assigning a value of one to all pixels.

Evaluation Metrics. We validate our pruning method using various quantitative metrics. In addition to the most popular Fréchet Inception Distance (FID) [8], we also measure Precision and Recall (P&R) [17] to assess the quality and diversity of generated samples separately. For FID calculation, we use 50K real and 50K fake samples for each dataset. For P&R, we use all real samples in the dataset and 50K fake samples for FFHQ and LSUN Church. For the LSUN Horse dataset, we use 200K real samples due to the high computational cost associated with complete 1M real samples.

To evaluate the projection capability of the compressed model, we employ the projection method of StyleGAN2, which projects real images into the $\mathcal{W}$ space using noise regularization and reconstruction losses such as MSE and LPIPS, over 1,000 iterations. We use the Helen-Set55 dataset for this evaluation, following the practice of CAGAN. We report the averaged MSE and LPIPS for all pairs of real and projected images in the dataset.

Implementation Details. We follow the implementation details provided in StyleKD [37]. For instance, all experiments are conducted with a pruning ratio of 0.7 (i.e. $p_r = 0.7$) and trained for 450K iterations. Also, in the default setting, we set the hyperparameters N and α as 10 and 5, except for LSUN Horse dataset where we set α as 10. We employ non-saturating loss [4] with R_1 gradient penalty [13] and update a model through Adam optimizer [16] with $\beta_1 = 0.0$ and $\beta_2 = 0.99$. The learning rate is set to 2e-3 for both the generator and discriminator. We use a batch size of 16 and consistently employ 4 GPUs for training. Lastly, the loss balancing constants are set to $\lambda_{\text{GAN}} = 1$, $\lambda_{\text{rgb}} = 3$, $\lambda_{\text{lpips}} = 3$, and $\lambda_{\text{LD}} = 30$.



(a) Precision-Recall Curve in FFHQ-256 (b) FID graph in FFHQ-256

Figure 3. **(a) Precision-Recall Curve.** By adjusting the truncation trick parameter ψ within the range [0.5, 1.0] with step size 0.1, we visualize the Precision-Recall curve of the proposed method and baselines. We validate that ours surpasses baseline methods with every range of precision and recall. **(b) FID w.r.t Training Iterations.** We visualize FID curve during training and validate that the proposed method achieves the previous state-of-the-art FIDs only with 2× fewer iterations.

4.2. Quantitative Results

Image Generation Performance. We validate the effectiveness of the proposed method through various generation performance metrics including FID and P&R in diverse datasets such as FFHQ, LSUN Church, and LSUN Horse. In terms of FID, the proposed method consistently outperforms state-of-the-art baselines on all tested datasets, as reported in Tab. 1. For example, our model achieves a FID score of 5.80 in FFHQ-1024, improving 1.8 and 1.39 points over CAGAN and StyleKD, respectively. These consistent FID score improvements demonstrate the superiority of the distribution-matching capability of the proposed method.

For the recall metric directly related to the sample diversity, our method achieves the highest scores in all tested datasets. For instance, our method significantly outperforms CAGAN in terms of recall score in FFHQ-1024 (0.277 vs 0.378), as well as obtains a value of 0.372 in LSUN Church, which is almost comparable with the teacher (0.016 points gap).

For the precision metric, the proposed method shows a slightly lower precision compared to the baseline, due to the precision-recall trade-off. Since the recall can be leveraged to improve precision through a truncation trick, we additionally conduct experiments by adjusting 'ψ' to evaluate precision at the highest recall achieved by baselines. In this setting, we observe that our method also surpasses baselines in terms of precision at the comparable recall. For example, by adopting a truncation parameter of $\psi = 0.95$ in

Table 1. **Quantitative results on StyleGAN2.** Comparison with GAN compression baselines in FFHQ, LSUN Church, and LSUN Horse datasets. 'Param.' refers to the number of parameters in a generator and 'FLOPs' denotes the number of floating-point operations. 'P' and 'R' represent precision and recall metrics, respectively. 'MSE' and 'LPIPS' indicate the distance between pairs of real and inverted samples based on the StyleGAN2 projection method. 'CAGAN*' denotes CAGAN trained with the objective of StyleKD. 'StyleKD*' refers to a retrained model using the official code due to the unavailability of published pretrained models. 'ψ' denotes the parameter for the truncation trick, which adjusts the trade-off between sample diversity and fidelity. We adjust 'ψ' to evaluate precision at the highest recall achieved by baselines. We have omitted 'Param.' and 'FLOPs' in the LSUN Church and LSUN Horse datasets, as they are the same as FFHQ-256. Reported FIDs without ours, CAGAN*, and StyleKD* are taken from StyleKD.

Dataset	Method	Param.	FLOPs	FID↓	P↑	R↑	MSE↓	LPIPS↓
FFHQ-256	Teacher	30.0M	45.1B	4.5	0.603	0.482	0.0292	0.2065
	Scratch	5.6M	4.1B	9.79	N/A	N/A	N/A	N/A
	GS [35]	N/A	5.0B	12.4	N/A	N/A	N/A	N/A
	CAGAN [24]	5.6M	4.1B	7.9	0.703	0.295	0.0316	0.2141
	CAGAN* [24]	5.6M	4.1B	7.41	0.713	0.309	0.0301	0.2113
	StyleKD [37]	5.6M	4.1B	7.25	N/A	N/A	N/A	N/A
	StyleKD* [37]	5.6M	4.1B	7.47	0.710	0.306	0.0307	0.2101
	Ours	5.6M	4.1B	**6.35**	0.706	**0.339**	**0.0293**	**0.2082**
	Ours ($\psi = 0.95$)	5.6M	4.1B	6.86	**0.732**	0.312	N/A	N/A
FFHQ-1024	Teacher	49.1M	74.3B	2.7	0.688	0.492	0.0275	0.1916
	GS [35]	N/A	23.9B	10.1	N/A	N/A	N/A	N/A
	CAGAN [24]	9.2M	7.0B	7.6	**0.721**	0.277	0.0358	0.2099
	StyleKD [37]	9.2M	7.0B	7.19	N/A	N/A	N/A	N/A
	Ours	9.2M	7.0B	**5.80**	0.676	**0.378**	**0.0328**	**0.2082**

Dataset	Method	FID↓	P↑	R↑	Dataset	Method	FID↓	P↑	R↑
LSUN Church	Teacher	3.97	0.599	0.388	LSUN Horse	Teacher	4.50	0.541	0.442
	CAGAN* [24]	5.09	0.598	0.352		CAGAN* [24]	5.73	0.589	0.335
	StyleKD* [37]	6.10	**0.611**	0.302		StyleKD* [37]	5.69	0.589	0.325
	Ours	**4.87**	0.584	**0.372**		**Ours**	**5.11**	0.581	**0.356**
	Ours ($\psi = 0.97$)	5.07	0.610	0.349		**Ours** ($\psi = 0.95$)	5.22	**0.610**	0.337

FFHQ-256 and LSUN Horse datasets, our method provides higher precision, recall, and even FID scores compared to the baselines. In the case of LSUN Church, with a truncation parameter of $\psi = 0.97$, we achieve comparable precision to StyleKD while providing 0.047 points higher recall score. Similarly, compared to CAGAN, ours has comparable scores in the aspect of recall but achieves higher precision with a reasonable margin. Note that, without the truncation trick, the proposed method substantially outperforms the baseline models in terms of diversity. For further validation, we additionally report the precision-recall curve in the FFHQ-256 dataset by adjusting the parameter ψ in Fig. 3-(a) within a range of [0.5, 1.0]. As shown, ours provides higher precision and recall compared to the baselines across the ranges.

To assess the applicability of our method with other network architectures, we conduct evaluations with StyleGAN3. Thus, we reproduce scores of baselines at the same training settings as ours and focus on relative improvements in this experiment. Specifically, we employ official codes of StyleGAN3 and train the generator until the discriminator see 10M of real images. As in Tab 2, our model achieves a FID score of 8.40 and improves 5.85 and 1.17 points over the CAGAN and StyleKD, respectively. In terms of the recall metric, our method also surpasses the baselines. A comparative analysis of metric scores between baselines and our approach further highlights our model's superiority.

Training Converged Speed Ability. Moreover, we visualize the FID curve with respect to the training iteration of our method in Fig. 3-(b). We observe that ours reach the best FID scores of previous state-of-the-art models at much fewer iterations. For instance, ours achieve comparable FID scores only with 2.5× and 2× fewer iterations compared to the CAGAN and StyleKD, respectively. We believe that our enhanced initialization of the synthesis network through our channel pruning strategy significantly improves the generation performance during the early stages of training, and allows the model for further improvement throughout the entire training process.

Table 2. **Quantitative results on StyleGAN3.** Comparison with recent baselines [24, 37] in FFHQ dataset.

Metric	FID ↓	P ↑	R ↑
Teacher	4.76	0.647	0.484
CAGAN [24]	14.25	0.636	0.272
StyleKD [37]	9.57	0.647	0.376
Ours	**8.40**	**0.653**	**0.397**

Image Projection Ability. For evaluation of the image projection ability, we perform experiments on FFHQ dataset using MSE and LPIPS metrics. For both FFHQ-256 and FFHQ-1024 datasets, we validate that the proposed method surpasses baselines in terms of both MSE and LPIPS metrics. Interestingly, we observe that ours reports MSE of 0.0293 which is a comparable score with the teacher model, 0.0292, in FFHQ-256 dataset. We believe that our pruning method brings a significant improvement in the generator's ability to capture the diverse characteristics of real images, and it allows precise projection of them.

4.3. Qualitative Results

Generated samples from same noise input. For qualitative comparison, we show the samples generated by different compressed models from the same noise vector z in Fig. 4. As noted in StyleKD [37], it would be desirable for a compressed network to reconstruct the sample generated by the teacher network. However, we observed that the baseline methods struggle to produce a sample inheriting the characteristics of images from the teacher network. For instance, CAGAN* and StyleKD* generate face images with different shapes of eyes in the third and fourth column of Fig. 4. Similarly, in the first column of LSUN Horse, horses generated by baselines hardly maintain the brown and white color patterns available in the teacher's image, while the proposed method successfully synthesizes those patterns. These results show that the channel pruning with the consideration of the latent space and its effect on images leads to better diversity preservation of generated images.

4.4. Ablation Studies and Analysis

We further discuss the influence of the proposed components and other hyperparameters by comprehensive ablation studies. For this purpose, we train the ablated models with different configurations in FFHQ-256.

Number of directions and strength for perturbations. We investigate the number of directions (N) and the strength of perturbation (α). Due to the computational costs, we report the mean and standard deviations of the best FID scores up to 50K iterations over 3 times trials. As reported in Tab. 3-(a), our method achieves the lowest FID score when $N = 10$. Additionally, we explore the impact of the strength of perturbation (α). While $\alpha = 5$ yields the best FID in this scenario, we consider both $\alpha = 5, 10$ as viable options.

Table 3. **Ablation on the perturbation parameters.** In this table, we report mean and standard deviation of $\text{FID}_{\text{early}}$ measured from the models training until 50K iterations 3 times. We denote N and α as the number of perturbations for each original latent vector and the strength parameter of perturbations, respectively. **Ablation on the direction sampling.** For comparing semantic and random directional vectors, we utilize the GANSpace (PCA) method to obtain the semantic directional vector. The FID score with semantic directional vectors is slightly better than that of the randomly sampled vectors.

N	$\text{FID}_{\text{early}}$ ↓	α	$\text{FID}_{\text{early}}$ ↓
$N = 5$	12.46 ± 0.2	$\alpha = 1$	13.50 ± 0.3
$N = 10$	**12.08 ± 0.3**	$\alpha = 5$	**12.08 ± 0.3**
$N = 20$	12.47 ± 0.2	$\alpha = 10$	12.09 ± 0.2

(a) Ablation on the perturbation parameters

Direction	FID ↓	P ↑	R ↑
Random	6.50	0.692	**0.342**
PCA	**6.35**	**0.706**	0.339

(b) Ablation on the perturbation directions

Table 4. **Ablation on the type of the importance scores.** Comparison of FID score between the average-based importance score (S^{μ}) and the proposed diversity-sensitive importance score (S^{σ}) in FFHQ-256 and LSUN Horse datasets.

Dataset	StyleKD*	Ours (w/ S^{μ})	Ours (w/ S^{σ})
FFHQ-256	7.47	6.71	**6.35**
LSUN Horse	5.69	5.36	**5.11**

Latent perturbations by random vs. PCA-based directions. We conduct the ablation study on directional vectors for latent perturbations. Specifically, we compare the performance of directional vectors derived from GANSpace [7] and vectors sampled from a random normal distribution ($d_{\text{rand}} \sim \mathcal{N}(0, 1)$). Tab. 3-(b) shows a slight degradation in performance when using random directional vectors for the perturbation. However, this degradation is negligible, and our pruned model still outperforms compared to baselines [24, 37] in terms of FID metric.

Thus, this finding highlights the significance of perturbation itself regardless of the type (either random or PCA) of perturbation in capturing sample diversity. Nevertheless, semantic directional vectors from GANSpace yield a lower FID compared to randomly sampled vectors. This result motivates us to select GANSpace-based directions as the default configuration for the latent vector perturbations.

Type of the importance score measurement. We conduct the ablation study to validate the effectiveness of our diversity-sensitive importance score. As shown in Tab. 4, the proposed diversity-sensitive importance score (S^{σ})

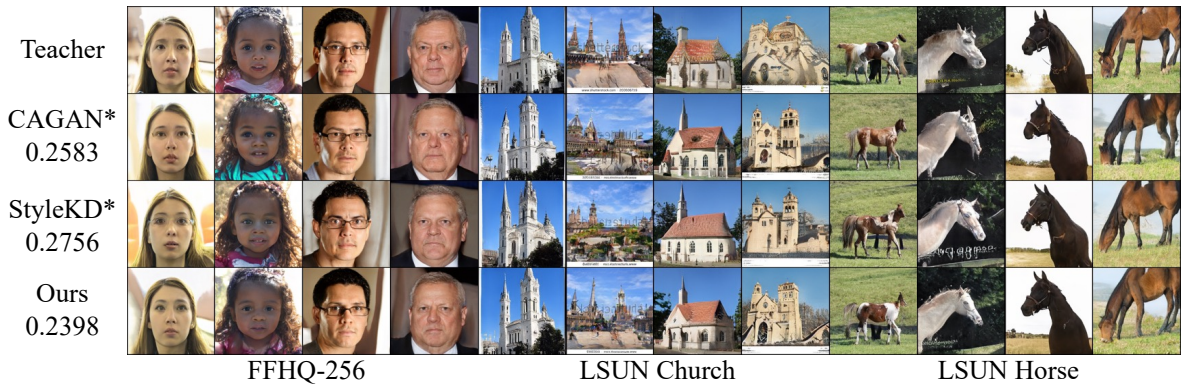
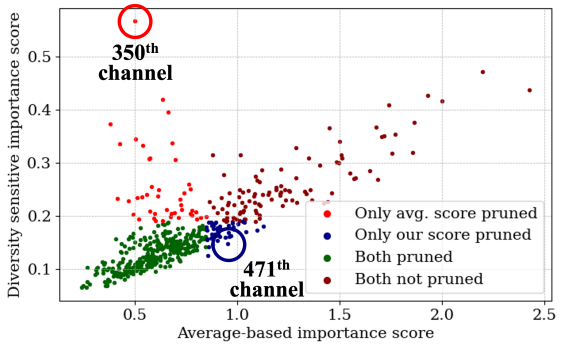
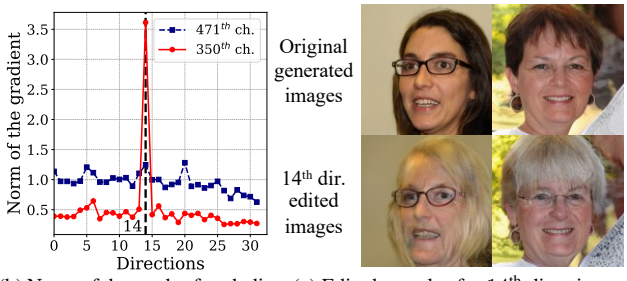


Figure 4. **Qualitative comparison with baselines on various datasets.** For qualitative comparison, we visualize our generated samples and baselines in FFHQ-256, LSUN Church-256, and LSUN Horse-256 datasets. Each column corresponds to samples generated from the same noise vector z. Averaged L1 distances between 10K samples from teacher and student are reported below each method. The lowest distortion of the proposed method validates that ours has enhanced capability to preserve the diversity in the image space.



(a) Channel scatter plot for avg. and our diversity-sensitive scores (S^μ, S^σ)



(b) Norm of the grad. of each dir. (c) Edited samples for 14th direction

Figure 5. **Specific examples for pruned channels** (a) We provide a scatter plot of S^μ and S^σ scores of all channels in 6th layer of teacher generator trained on FFHQ-256. Brown and green dots represent channels that are always pruned and not pruned, respectively. Blue and Red dots indicate channels that only survive in S^μ and S^σ, respectively. (b) The 350th channel exhibits high sensitivity to the 14th direction from PCA. (c) The 14th direction corresponds to an age-related perturbation. The S^μ score prunes the 350th channel, while the S^σ score preserves this channel, which demonstrates high sensitivity to age variation. This result shows that S^σ aims to retain semantic image changes compared to S^μ.

outperforms the average-based importance score (S^μ) in terms of FID. Specifically, the proposed diversity-sensitive method provides 6.35 FID, improving 0.36 points over the average-based score. Similar trends can be found in LSUN Horse dataset. These results confirm that suppressing the importance score derived from latent code itself is crucial.

Specific examples for pruned channels In Fig 5, we examine the score of each channel in the teacher generator using a scatter plot, with S^μ on the x-axis and S^σ on the y-axis. As visualized in Fig. 5 (a), S^μ and S^σ of most channels are linearly correlated. However, upon closer examination of outliers, a significant portion of these outliers possess low S^μ values but high S^σ values. These outlier channels generally contribute to a smaller number of directions but are highly activated for particular directions. To illustrate this, we visualize the gradient norms for 350th and 471th channels in Fig. 5 (b). The 350th channel exerts a significant influence on the 14th direction. This particular direction corresponds to notable image differences, such as age-related perturbations, as shown in Fig. 5 (c). Thus, our S^σ can assign higher importance scores for these outliers.

5. Conclusion

In this work, we address the issue of a compressed StyleGAN generator struggling to preserve the sample diversity of teacher, observed in previous GAN compression methods. To alleviate this problem, we propose a simple yet effective channel pruning method that calculates channel importance scores based on their sensitivity in the image space according to perturbations in latent space. Extensive experiments demonstrate that our pruning method significantly enhances sample diversity, validating its superior performance in terms of FID.

Acknowledgments

This work was supported in part by MSIT/IITP (No. 2022-0-00680, 2019-0-00421, 2020-0-01821, 2021-0-02068), and MSIT&KNPA/KIPoT (Police Lab 2.0, No. 210121M06).

References

[1] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019. 1

[2] Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5799–5809, 2021. 1

[3] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021. 1

[4] 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*. Curran Associates, Inc., 2014. 5

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 1

[6] Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. StyleneRF: A style-based 3d aware generator for high-resolution image synthesis. In *International Conference on Learning Representations*, 2022. 1

[7] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. *Advances in Neural Information Processing Systems*, 33:9841–9850, 2020. 1, 4, 7

[8] 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. 5

[9] Tie Hu, Mingbao Lin, Lizhou You, Fei Chao, and Rongrong Ji. Discriminator-cooperated feature map distillation for gan compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20351–20360, 2023. 2

[10] Sangeek Hyun, Jaihyun Lew, Jiwoo Chung, Euiyeon Kim, and Jae-Pil Heo. Frequency-based motion representation for video generative adversarial networks. *IEEE Transactions on Image Processing*, 2023. 1

[11] Minsoo Kang, Hyewon Yoo, Eunhee Kang, Sehwan Ki, Hyong Euk Lee, and Bohyung Han. Information-theoretic gan compression with variational energy-based model. *Advances in Neural Information Processing Systems*, 35:18241–18255, 2022. 2

[12] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 1

[13] 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. 3, 5

[14] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34:852–863, 2021. 1

[15] Hyunsu Kim, Yunjey Choi, Junho Kim, Sungjoo Yoo, and Youngjung Uh. Exploiting spatial dimensions of latent in gan for real-time image editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 852–861, 2021. 1

[16] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 5

[17] Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems*, 32, 2019. 5

[18] Gayoung Lee, Hyunsu Kim, Junho Kim, Seonghyeon Kim, Jung-Woo Ha, and Yunjey Choi. Generator knows what discriminator should learn in unconditional gans. In *ECCV*, 2022. 1

[19] Min Kyu Lee, Seunghyun Lee, Sang Hyuk Lee, and Byung Cheol Song. Channel pruning via gradient of mutual information for light-weight convolutional neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1751–1755. IEEE, 2020. 2

[20] Shaojie Li, Jie Wu, Xuefeng Xiao, Fei Chao, Xudong Mao, and Rongrong Ji. Revisiting discriminator in gan compression: A generator-discriminator cooperative compression scheme. *Advances in Neural Information Processing Systems*, 34:28560–28572, 2021. 2

[21] Shaojie Li, Mingbao Lin, Yan Wang, Chao Fei, Ling Shao, and Rongrong Ji. Learning efficient gans for image translation via differentiable masks and co-attention distillation. *IEEE Transactions on Multimedia*, 2022. 2

[22] Congcong Liu and Huaming Wu. Channel pruning based on mean gradient for accelerating convolutional neural networks. *Signal Processing*, 156:84–91, 2019. 2

[23] Jing Liu, Bohan Zhuang, Zhuangwei Zhuang, Yong Guo, Junzhou Huang, Jinhui Zhu, and Mingkui Tan. Discrimination-aware network pruning for deep model compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4035–4051, 2021. 2

[24] Yuchen Liu, Zhixin Shu, Yijun Li, Zhe Lin, Federico Perazzi, and Sun-Yuan Kung. Content-aware gan compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12156–12166, 2021. 1, 2, 5, 6, 7

[25] Pavlo Molchanov, Stephen Tyree, Tero Karras, Timo Aila, and Jan Kautz. Pruning convolutional neural networks for resource efficient inference. *arXiv preprint arXiv:1611.06440*, 2016. 2

[26] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on*

computer vision and pattern recognition, pages 2337–2346, 2019. 1

[27] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2085–2094, 2021. 1

[28] Yuxi Ren, Jie Wu, Xuefeng Xiao, and Jianchao Yang. Online multi-granularity distillation for gan compression. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6793–6803, 2021. 2

[29] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2287–2296, 2021. 1

[30] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 1

[31] Axel Sauer, Katja Schwarz, and Andreas Geiger. Stylegan-xl: Scaling stylegan to large diverse datasets. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022. 1

[32] Yichun Shi, Divyansh Aggarwal, and Anil K Jain. Lifting 2d stylegan for 3d-aware face generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6258–6266, 2021. 1

[33] Sang-Heon Shim, Jiwoo Chung, and Jae-Pil Heo. Towards squeezing-averse virtual try-on via sequential deformation. *arXiv preprint arXiv:2312.15861*, 2023. 1

[34] Xu Sun, Xuancheng Ren, Shuming Ma, and Houfeng Wang. meprop: Sparsified back propagation for accelerated deep learning with reduced overfitting. In *International Conference on Machine Learning*, pages 3299–3308. PMLR, 2017. 2

[35] Haotao Wang, Shupeng Gui, Haichuan Yang, Ji Liu, and Zhangyang Wang. Gan slimming: All-in-one gan compression by a unified optimization framework. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, pages 54–73. Springer, 2020. 6

[36] Zhendong Wang, Huangjie Zheng, Pengcheng He, Weizhu Chen, and Mingyuan Zhou. Diffusion-gan: Training gans with diffusion. In *The Eleventh International Conference on Learning Representations*, 2022. 1

[37] Guodong Xu, Yuenan Hou, Ziwei Liu, and Chen Change Loy. Mind the gap in distilling stylegans. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*, pages 423–439. Springer, 2022. 1, 2, 3, 4, 5, 6, 7

[38] Linfeng Zhang, Xin Chen, Xiaobing Tu, Pengfei Wan, Ning Xu, and Kaisheng Ma. Wavelet knowledge distillation: Towards efficient image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12464–12474, 2022. 2

[39] Jiapeng Zhu, Yujun Shen, Deli Zhao, and Bolei Zhou. In-domain gan inversion for real image editing. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16*, pages 592–608. Springer, 2020. 1

[40] Zhuangwei Zhuang, Mingkui Tan, Bohan Zhuang, Jing Liu, Yong Guo, Qingyao Wu, Junzhou Huang, and Jinhui Zhu. Discrimination-aware channel pruning for deep neural networks. *Advances in neural information processing systems*, 31, 2018. 2