国际人工智能会议
AAAI 2021论文北京预讲会

Slimmable Generative Adversarial Networks

Liang Hou,[1,2] Zehuan Yuan,[3] Lei Huang,[4,5] Huawei Shen,[1,2] Xueqi Cheng,[1,2] Changhu Wang[3]

[1]CAS Key Laboratory of Network Data Science and Technology,
Institute of Computing Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]ByteDance AI Lab , [4]Inception Institute of Artificial Intelligence , [5]Beihang University

Motivation

For real-time generation tasks, different devices may require models of different sizes due to their different computing power. And even for the same device, it also needs models of different sizes due to different power modes. Consequently, numerous models need to be trained and deployed for a single task.

To combat the above issues, we develop a **width-switchable** generator that can flexibly switch the width (the number of channels in layers) of the generator at runtime to accommodate various quality-efficiency trade-offs.

Methods

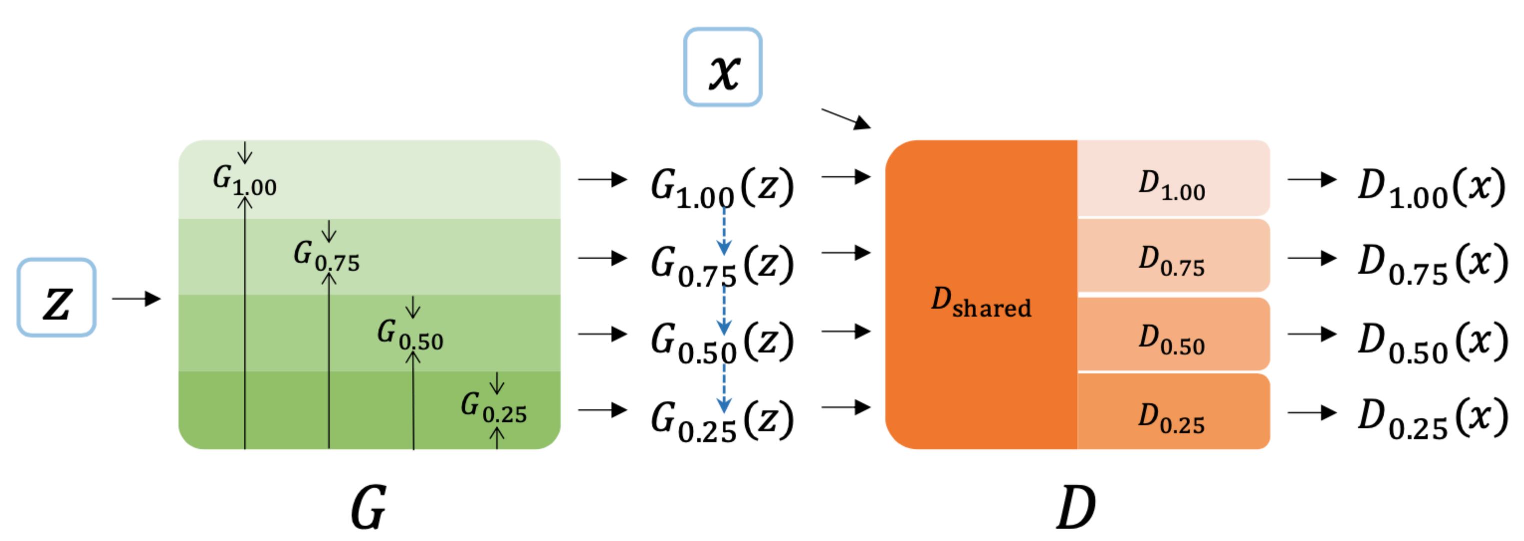


Figure 1: Illustration of SlimGAN with width multiplier list $\mathcal{W} = [0.25, 0.5, 0.75, 1.0]\times$. Wide generators contain the channels of narrow ones. Multiple discriminators share first several layers. Blue dashed lines indicate the stepwise inplace distillation.

- To train the slimmable generator, we utilize adversarial training with multiple discriminators that share first several layers, and employ the Hinge loss:

$$\max_{D} \mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\min\left(0, -1 + D_{w_i}(x)\right)\right] + \mathbb{E}_{z\sim\mathbb{P}_z}\left[\min\left(0, -1 - D_{w_i}\left(G_{w_i}(z)\right)\right)\right]$$

$$\max_{G} \mathbb{E}_{z\sim\mathbb{P}_z}\left[D_{w_i}\left(G_{w_i}(z)\right)\right], i = 1,2,\dots,N$$

- To achieve consistency between these sub-generators, we propose stepwise inplace distillation that encourage narrows generators learn from wide ones:

$$\min_{G} \frac{\lambda}{N-1}\sum_{i=1}^{N-1}\mathbb{E}_{z\sim\mathbb{P}_z}\left[\left\|G_{w_i}(z) - \mathrm{sg}\left(G_{w_{i+1}}(z)\right)\right\|_2^2\right]$$

- To incorporate class-conditional label information into generators under the width-switchable mechanism, we present sliceable conditional batch normalization:

$$x'_{w_i,c_j} = \gamma_{w_i}\gamma_{c_j}^{:s_i}\frac{x_{w_i,c_j} - \mu(x_{w_i,c_j})}{\sigma(x_{w_i,c_j})} + \beta_{w_i} + \beta_{c_j}^{:s_i}$$

Specifically, we slice cBN vectors γ_{c_j} and β_{c_j} to sub-vectors $\gamma_{c_j}^{:s_i}$ and $\beta_{c_j}^{:s_i}$ with the first s_i elements to match the length of sBN vectors γ_{w_i} and β_{w_i} at current width w_i.

Evaluation Metrics

- **FID** and **IS**: evaluate the quality and diversity of generated samples.
- **mIC**: evaluate the consistency between generators at different widths, and is defined as follows:

$$\mathrm{mIC} = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1, i\neq j}^{N}\mathbb{E}_{z\sim\mathbb{P}_z}\left\|\Phi\left(G_{w_i}(z)\right) - \Phi\left(G_{w_j}(z)\right)\right\|_2^2$$

where $\Phi(\cdot)$ extracts the image features from the Inception v3 network.

Generation Results

Table 1: FID and IS on both unconditional (uncond) and class-conditional (cond) generation. We do not calculate IS on CelebA as it is a face dataset that lacking inter-class diversity, which IS measures. For class-conditional generation, (+) is our proposed sliceable conditional batch normalization while (×) is the naive way that extends each sBN to cBN. Bold numbers indicate our slimmable method outperforms the individually trained models.

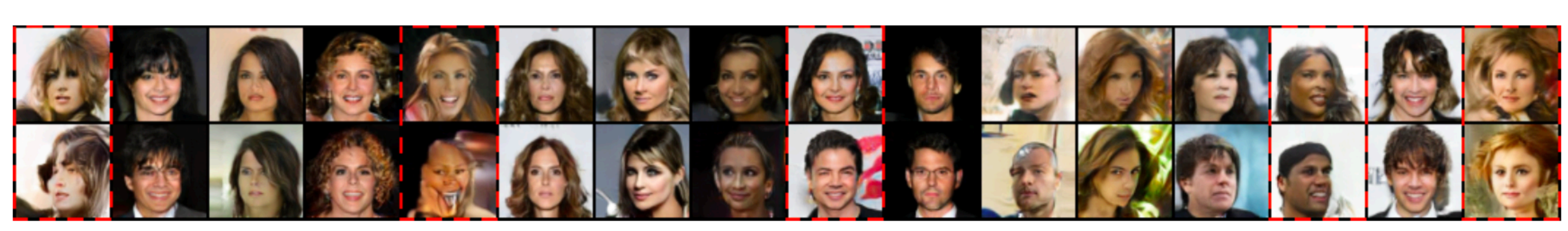
Backbone	Dataset	Method	FID (↓) 0.25×	FID (↓) 0.5×	FID (↓) 0.75×	FID (↓) 1.0×	IS (↑) 0.25×	IS (↑) 0.5×	IS (↑) 0.75×	IS (↑) 1.0×
DCGAN (uncond)	CIFAR-10	Individual	46.9	34.6	30.4	26.7	6.08	6.95	7.39	7.43
		Slimmable	**37.3**	**28.5**	**25.8**	**25.2**	**6.90**	**7.31**	**7.43**	**7.44**
	STL-10	Individual	93.1	69.1	61.8	57.4	6.51	7.82	7.96	8.38
		Slimmable	**68.9**	**60.9**	**56.2**	**55.1**	**7.67**	**8.00**	**8.34**	**8.38**
	CelebA	Individual	24.4	13.2	10.4	9.8	-	-	-	-
		Slimmable	**23.3**	13.3	10.6	**9.4**	-	-	-	-
ResNet (uncond)	CIFAR-10	Individual	41.8	24.1	21.6	20.3	7.36	7.68	7.93	7.91
		Slimmable	**29.9**	**21.6**	**19.6**	**20.0**	7.32	**8.02**	**8.15**	**8.09**
	STL-10	Individual	66.6	58.5	56.3	52.9	7.90	8.52	8.30	8.60
		Slimmable	69.1	59.0	**50.8**	**50.6**	7.60	8.23	**8.83**	**8.81**
	CelebA	Individual	18.0	11.9	9.9	8.9	-	-	-	-
		Slimmable	**13.9**	**10.6**	**9.8**	**8.5**	-	-	-	-
cGAN-pd (cond)	CIFAR-10	Individual	55.1	33.5	16.5	15.5	6.46	7.90	8.22	8.52
		Slimmable (×)	21.7	17.2	16.1	16.2	7.87	8.31	8.49	8.34
		Slimmable (+)	**19.5**	**14.5**	**13.6**	**14.2**	**7.88**	**8.38**	**8.67**	**8.59**
	CIFAR-100	Individual	45.8	23.7	22.5	19.9	7.26	8.49	8.50	9.11
		Slimmable (×)	26.8	19.9	18.9	19.0	8.13	8.90	9.14	9.22
		Slimmable (+)	**23.8**	**18.9**	**18.6**	**17.9**	**8.26**	**9.08**	**9.17**	**9.29**

Ablation Study

Table 4: Ablation Study on CIFAR-10.

DCGAN on CIFAR-10	FID (↓) 0.25×	FID (↓) 0.5×	FID (↓) 0.75×	FID (↓) 1.0×	AVG (↓)	mIC (↓)
Individual	46.9	34.6	30.4	27.4	34.8	-
Individual-f	45.6	33.2	29.4	27.4	33.9	-
Slimmable G	40.0	35.2	34.4	33.4	35.8	264.3
+ shared D	40.9	30.2	27.0	25.2	30.8	282.7
+ shared D + distillation (SlimGAN)	37.3	**28.5**	**25.8**	**25.2**	**29.2**	231.3
+ same D	180.4	136.9	141.3	158.6	154.3	376.8
+ slimmable D	43.6	35.8	31.0	33.0	35.9	269.5
+ distillation (w/o GAN loss for narrows)	87.9	56.2	37.8	28.9	52.7	**204.8**
+ shared D + naive distillation	**36.6**	29.8	26.3	25.5	29.6	232.5

Consistency Results



(a) Slimmable GAN without distillation, showing clear inconsistency (boxed in red dotted lines).

(b) Slimmable GAN with distillation, showing improved consistency.

Figure 2: Qualitative consistency on CelebA.

Table 3: mIC (↓) on CIFAR-10, STL-10, and CelebA.

SlimDCGAN	CIFAR-10	STL-10	CelebA
+ w/o distillation	282.7	277.4	110.2
+ w/ distillation	**231.3**	**243.2**	**96.1**
SlimResGAN	CIFAR-10	STL-10	CelebA
+ w/o distillation	285.7	342.4	116.9
+ w/ distillation	**241.4**	**248.7**	**97.9**

Paper **Code**

主办方：中国中文信息学会青年工作委员会
承办方：智源社区