Balanced Training for Sparse GANs

Yite Wang[1,*], Jing Wu[1,*], Naira Hovakimyan[1], Ruoyu Sun[2,3]

[1]University of Illinois Urbana-Champaign, USA

[2]School of Data Science, The Chinese University of Hong Kong, Shenzhen, China

[3] Shenzhen Research Institute of Big Data

Background: Dynamic sparse training (DST)

1. Start with a sparse network parameterized by $\boldsymbol{\theta} \odot \boldsymbol{m}$, where mask $\boldsymbol{m} \in \{0,1\}^p$ and weights $\boldsymbol{\theta}$ are **randomly initialized**.
2. After every constant time interval ΔT, update mask:
   - Remove a fraction of connections.
   - Activate the same number of connections.
3. The total number of active parameters $\|\boldsymbol{m}\|_0$ is kept under a certain threshold: $\|\boldsymbol{m}\|_0/p \le d$.

Motivation

We test two baseline methods:

1. `Static`: Fix masks of both generators and discriminators.
2. Single dynamic sparse training (`SDST`): apply DST to the generator; fix mask of the discriminator.

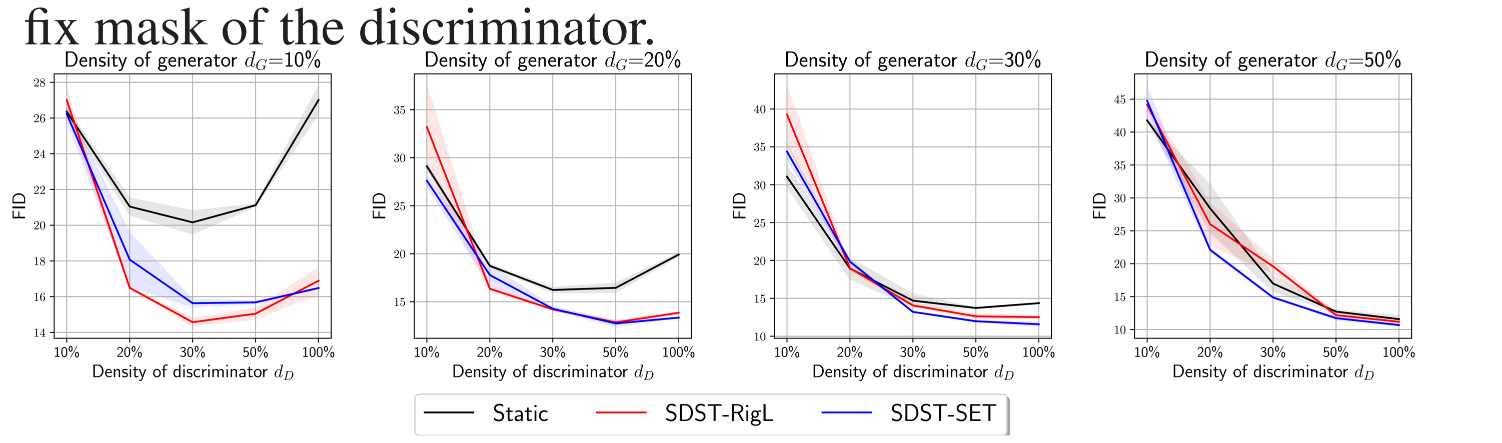
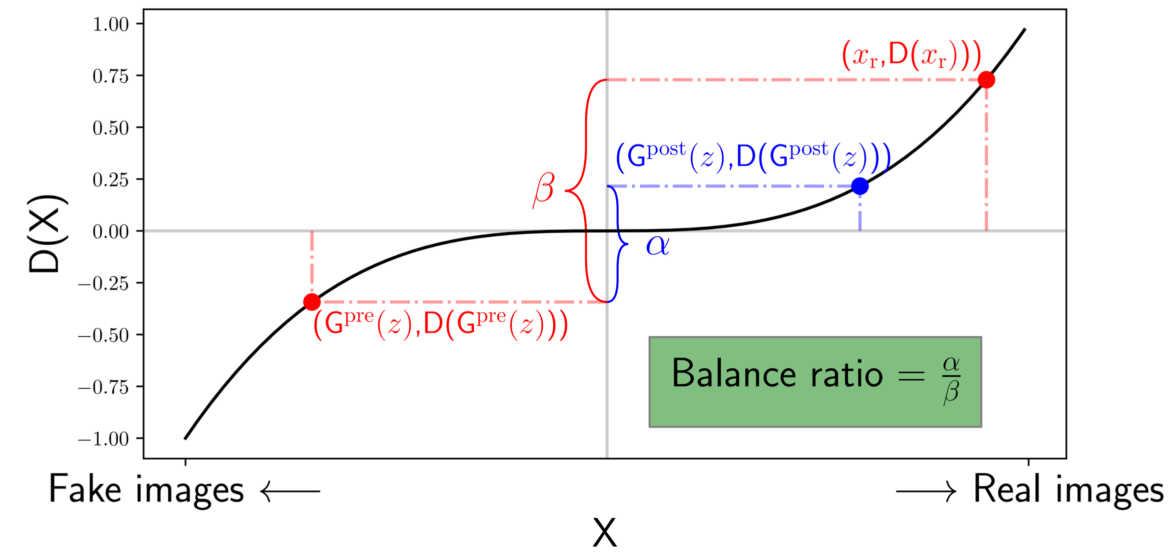


Figure.1: FID (↓) of SNGAN on CIFAR-10 with different sparsity ratio combinations. **(1)** `SDST` does **NOT** show consistent improvement over `Static` sparse training baseline; **(2)** Neither strong nor weak sparse discriminators can provide satisfactory results.

Balance ratio: Towards quantifying the unbalance in sparse GAN training

Formulation of the balance ratio:



At each training iteration/step, we require:

- The random noise $\mathbf{z} \sim p$ drawn for training.
- Real images $\mathbf{x}_r \sim p_{\text{data}}$ drawn for training.
- Discriminator after weight update $D(\cdot; \boldsymbol{\theta}_D)$.
- Generator before and after weight update $G^{\text{pre}}(\cdot; \boldsymbol{\theta}_G)$ and $G^{\text{post}}(\cdot; \boldsymbol{\theta}'_G)$.

The balance ratio (BR) is defined as:

$$\mathrm{BR} = \frac{\mathbb{E}_{\boldsymbol{z}\sim p}\left[D(G^{\text{post}}(\boldsymbol{z})) - D(G^{\text{pre}}(\boldsymbol{z}))\right]}{\mathbb{E}_{\boldsymbol{x}_r\sim p_{\text{data}}}[D(\boldsymbol{x}_r)] - \mathbb{E}_{\boldsymbol{z}\sim p}[D(G^{\text{pre}}(\boldsymbol{z}))]} = \frac{\alpha}{\beta}.$$

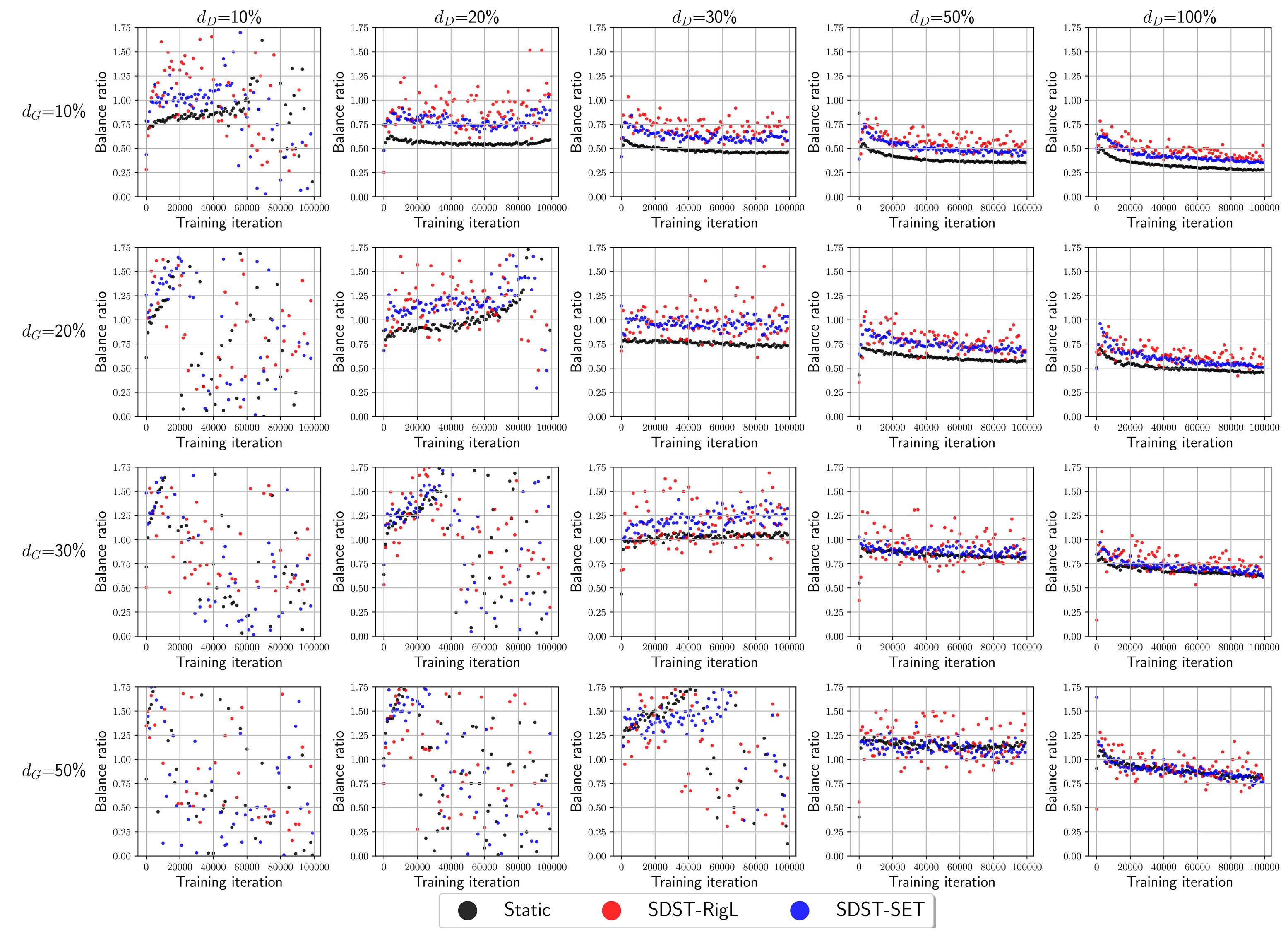


Figure.2: BR visualization of `SDST` against `Static` sparse training with SNGAN on CIFAR-10. **(1)** Stronger the D compared to the G, smaller the BR during training. **(2)** For failure cases (left-bottom), BR increases and then oscillates wildly.

bAlanced DynAmic sParse Training (`ADAPT`)

To enable balanced training, different from `SDST`, `ADAPT` **adjusts the density of the discriminator** during training based on BR.

Discriminator in `SDST`

1. Manually choose budget d_D.
2. Train a **static** discriminator with given budget d_D.

Discriminator in `ADAPT`

Relaxed setting:

$0 < d_D \le 1$

1. Initialize $d_D^{\text{init}} = d_G$.
2. Every ΔT, dynamically adjust d_D s.t. $BR \in [B_-, B_+]$:
   - If $BR < B_-$: decrease d_D to $\max(0, d_D - \Delta d)$.
   - If $BR > B_+$: increase d_D to $\min(1, d_D + \Delta d)$.

Strict setting:

$0 < d_D \le d_{\max} \le 1$

1. Initialize $d_D^{\text{init}} = d_G$.
2. Every ΔT, dynamically adjust d_D s.t. $BR \in [B_-, B_+]$:
   - If $BR < B_-$: Decrease d_D to $\max(0, d_D - \Delta d)$.
   - If $BR > B_+$: Increase d_D to $\min(d_{\max}, d_D + \Delta d)$. If $d_D = d_{\max}$, apply DST to the discriminator.

Results: Quality of generated images

Table.1: FID (↓) of different sparse training methods in the relaxed setting.

Dataset	CIFAR-10 (SNGAN)				STL-10 (SNGAN)				CIFAR-10 (BigGAN)				TinyImageNet (BigGAN)			
Generator density	10%	20%	30%	50%	10%	20%	30%	50%	10%	20%	30%	50%	10%	20%	30%	50%
(Dense Baseline)	10.74				29.71				8.11				15.43			
Static-Balance	26.75	19.04	15.05	12.24	48.18	44.67	41.73	37.68	16.98	12.81	10.33	8.47	28.78	21.67	18.86	17.51
Static-Strong	26.79	19.65	14.38	11.91	52.48	43.85	42.06	37.47	23.48	14.26	11.19	8.64	31.44	22.51	18.22	18.00
◆ SDST-Balance-SET	26.23	17.79	13.21	11.79	56.41	46.58	39.93	30.37	12.41	9.87	9.13	**8.01**	25.39	21.30	21.80	21.20
◆ SDST-Strong-SET	16.49	13.36	**11.68**	10.68	67.37	49.96	37.99	31.08	18.94	9.64	8.75	8.36	22.20	20.56	21.70	18.32
● SDST-Balance-RigL	27.06	16.36	14.00	12.28	43.08	33.90	31.83	30.30	12.45	9.42	8.86	8.03	21.60	19.33	18.57	17.45
● SDST-Strong-RigL	17.02	13.86	12.51	11.35	53.65	33.25	**31.41**	30.18	10.58	9.11	8.69	8.33	21.14	18.95	17.75	16.30
$\text{ADAPT}_{\text{relax}}$ (Ours)	**14.19**	**13.19**	12.38	**10.60**	**35.98**	**33.06**	31.71	**29.96**	**10.19**	**8.56**	**8.36**	8.22	**19.42**	**17.99**	**17.06**	**14.15**

Table.2: FID (↓) of sparse training methods in the strict setting, $d_{\max} = 50\%$.

Dataset	CIFAR-10 (SNGAN)				STL-10 (SNGAN)				CIFAR-10 (BigGAN)				TinyImageNet (BigGAN)			
Generator density	10%	20%	30%	50%	10%	20%	30%	50%	10%	20%	30%	50%	10%	20%	30%	50%
(Dense Baseline)	10.74				29.71				8.11				15.43			
Static-Balance	26.75	19.04	15.05	12.58	48.18	44.67	41.73	37.68	16.98	12.81	10.33	8.47	28.78	21.67	18.86	17.51
Static-Strong	21.73	16.69	13.48	12.58	50.36	44.06	40.73	37.68	18.91	13.43	10.84	8.47	33.01	23.93	17.90	17.51
◆ SDST-Balance-SET	26.23	17.79	13.21	**11.79**	56.24	44.51	41.23	30.80	12.41	9.87	9.13	8.01	25.39	21.30	21.80	21.20
◆ SDST-Strong-SET	15.68	12.75	**11.98**	**11.79**	57.91	50.05	38.13	30.80	11.85	9.39	8.61	8.01	22.68	20.24	22.00	21.20
● SDST-Balance-RigL	27.06	16.36	14.00	12.28	43.08	33.90	31.64	30.30	12.45	9.42	8.86	8.03	21.60	19.33	18.57	17.45
● SDST-Strong-RigL	15.19	12.93	12.75	12.28	53.74	37.34	31.98	30.30	10.11	9.17	**8.35**	8.03	21.90	20.43	18.29	17.45
$\text{ADAPT}_{\text{strict}}$ (Ours)	**14.53**	**12.73**	12.20	12.11	**41.18**	**31.59**	**31.16**	**29.11**	**9.29**	**8.64**	8.44	**7.90**	**18.89**	**17.37**	**16.93**	**16.02**

(1) In the relaxed setting, ADAPT strikes a balance between high performance and efficient computational cost. It emerged as the top performer in 13 out of 16 test cases. **(2)** In the strict setting, despite the additional constraints on the discriminator, it still performed remarkably well. It ranked in the top two for all test cases and was the best performer in 13 of them.

Results: Computational costs

Table.3: Normalized training FLOPs comparison in the relaxed setting.

Dataset	CIFAR-10 (SNGAN)				STL-10 (SNGAN)				CIFAR-10 (BigGAN)				TinyImageNet (BigGAN)			
Generator density	10%	20%	30%	50%	10%	20%	30%	50%	10%	20%	30%	50%	10%	20%	30%	50%
(Dense Baseline)	100% (2.67×10^{17})				100% (3.94×10^{17})				100% (6.81×10^{17})				100% (9.85×10^{17})			
Static-Balance	8.97%	17.08%	26.25%	47.25%	27.30%	47.14%	59.22%	73.35%	9.79%	19.02%	28.66%	49.03%	23.25%	44.87%	60.91%	79.29%
Static-Strong	58.29%	60.94%	64.53%	74.61%	86.12%	86.94%	87.60%	88.84%	83.78%	84.80%	86.21%	90.15%	48.02%	61.62%	72.79%	85.79%
◆ SDST-Balance-SET	9.78%	18.91%	28.35%	48.44%	27.55%	47.60%	60.17%	75.38%	10.35%	20.12%	29.96%	49.82%	21.13%	37.06%	48.83%	65.58%
◆ SDST-Strong-SET	59.25%	62.94%	66.89%	75.96%	86.36%	87.43%	88.49%	90.82%	84.36%	85.90%	87.52%	90.95%	45.66%	53.91%	60.61%	71.88%
● SDST-Balance-RigL	10.71%	17.43%	25.66%	43.56%	29.51%	50.41%	63.34%	79.03%	9.92%	19.30%	28.90%	48.31%	24.97%	43.86%	57.26%	76.75%
● SDST-Strong-RigL	58.63%	61.35%	64.04%	71.01%	88.51%	90.24%	91.78%	94.57%	83.97%	85.24%	86.59%	89.54%	50.05%	61.02%	69.64%	83.35%
$\text{ADAPT}_{\text{relax}}$ (Ours)	36.67%	57.62%	61.31%	70.11%	46.73%	77.92%	83.62%	90.49%	10.39%	25.90%	40.65%	80.76%	29.75%	51.98%	64.57%	80.81%

Conclusion

- We introduce a metric named **balance ratio (BR)** to study the degree of balance in sparse GAN training.
- We use BR to study the limitations of existing GAN dynamic sparse training (DST) methods.
- We propose `ADAPT`, which dynamically adjusts the discriminator's density, enabling effective control of the BR throughout training.