Training GANs with Stronger Augmentations via Contrastive Discriminator

Jongheon Jeong   Jinwoo Shin

Korea Advanced Institute of Science and Technology (KAIST)

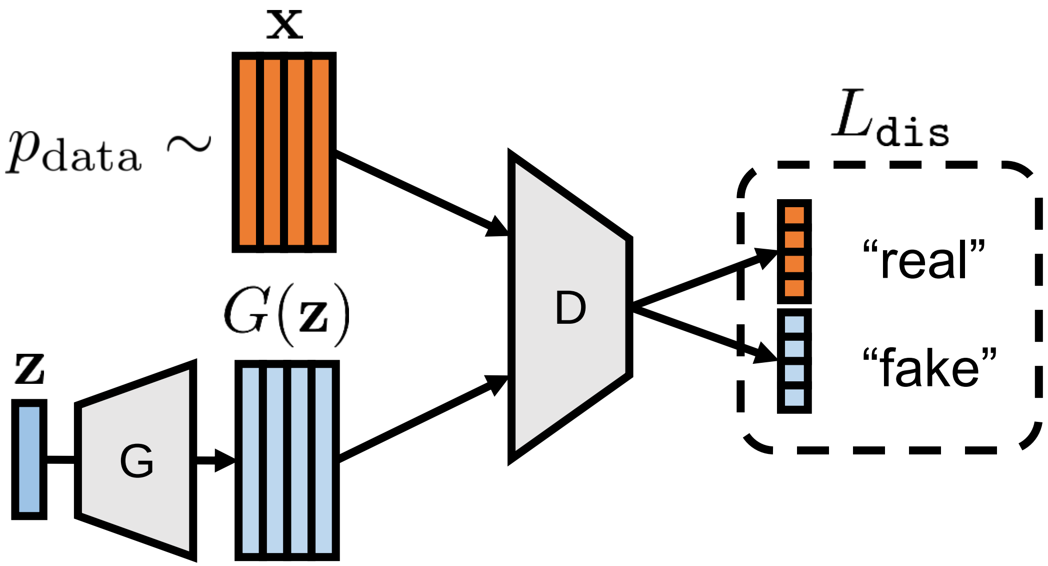
ICLR 2021

Generative Adversarial Nets (GANs) [Goodfellow et al., 2014]

Idea: A generator G vs. a discriminator D for generative modeling of $p_{data}(\mathbf{x})$

- $G(\mathbf{z}) \to p_{\text{data}}$, where $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}),\ \mathcal{U}(-1, 1),\ \ldots$

$$\min_G \max_D L_{\mathtt{dis}}(D; G) := \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}}[\underbrace{\log(D(\mathbf{x}))}_{\text{“real”}}] + \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})}[\underbrace{\log(1 - D(G(\mathbf{z})))}_{\text{“fake”}}]$$



BigGAN [Brock et al., 2019]

StyleGAN2 [Karras et al., 2020a]

[Goodfellow et al., 2014] Generative Adversarial Networks. NeurIPS 2014.
[Brock et al., 2019] Large Scale GAN Training for High Fidelity Natural Image Synthesis. ICLR 2019.
[Karras et al., 2020a] Analyzing and Improving the Image Quality of StyleGAN, CVPR 2020.

Data Augmentation for GAN is Non-trivial

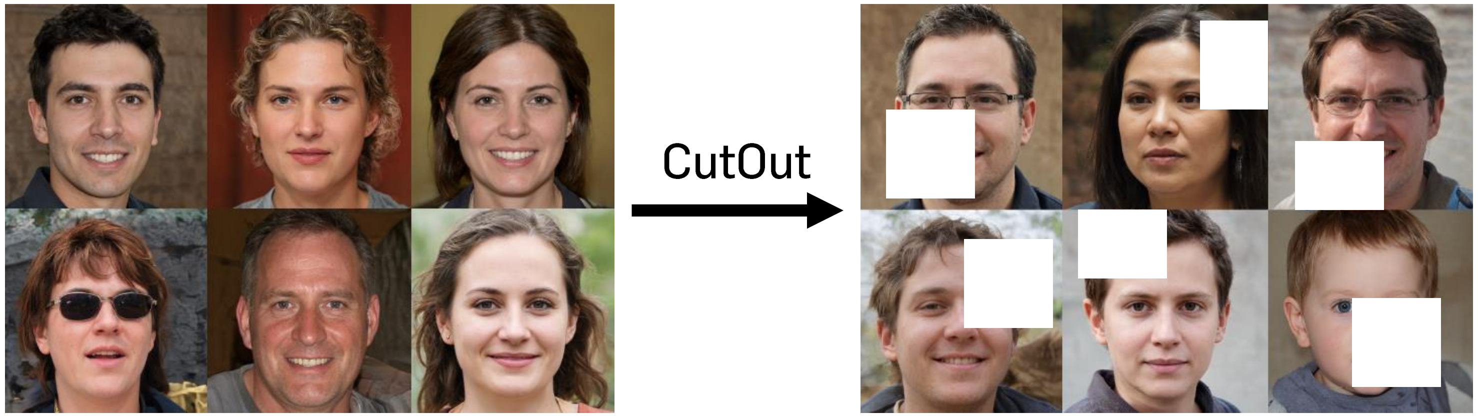
GANs are always data-hungry! → "discriminator overfitting"

- One can try to collect more data [Brock et al., 2019], or to regularize D [Miyato et al., 2018], ...
- ... but how about to simply use "stronger" data augmentation?

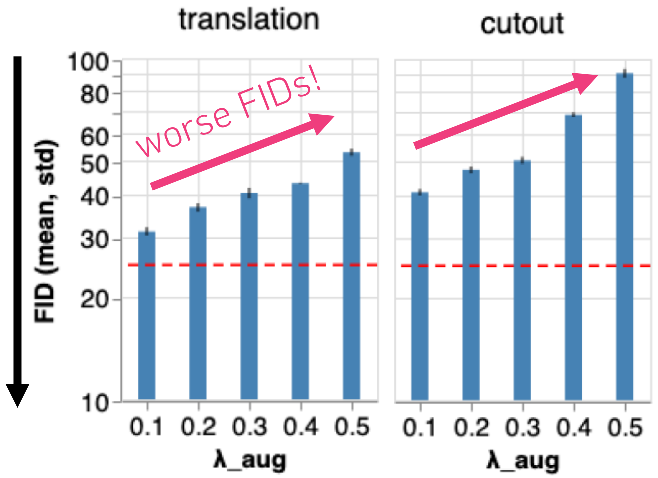
Challenge: How can we safely incorporate data augmentations for GANs?

- "Augmentation leakage": Direct augmentation of $p_{data}(\mathbf{x})$ can significantly shift the distribution

$G(\mathbf{z}) \rightarrow p_{\text{data}}(\mathbf{x})$ $\quad$ $G(\mathbf{z}) \rightarrow p_{\text{data}}(T(\mathbf{x}))$?



lower is better



[ZhaoZ et al., 2020b]

[Miyato et al., 2018] Spectral Normalization for Generative Adversarial Networks, ICLR 2018.
[Brock et al., 2019] Large Scale GAN Training for High Fidelity Natural Image Synthesis, ICLR 2019.
[ZhaoZ et al., 2020b] Image Augmentations for GAN Training, 2020.

Data Augmentation for GAN is Non-trivial

Challenge: How can we safely incorporate data augmentations for GANs?

- "Augmentation leakage": Direct augmentation of $p_{\mathrm{data}}(\mathbf{x})$ can significantly shift the distribution

"GAN-compatible" data augmentations?

- Consistency regularization [Zhang et al., 2020; ZhaoZ et al., 2020a] → "Flip + Translation"
- Differentiable augmentation [ZhaoS et al., 2020] → "Flip + Translation + CutOut"
- Adaptive discriminator augmentation [Karras et al., 2020a] → Dynamic pipelining of augmentations
- AdvAug [Chen et al., 2021] → DiffAug + Adversarial augmentation

🤔 How can we further extend this boundary of "GAN-compatible" augmentations?

💡 **Idea**: Make D to learn a contrastive representation of real + fake!

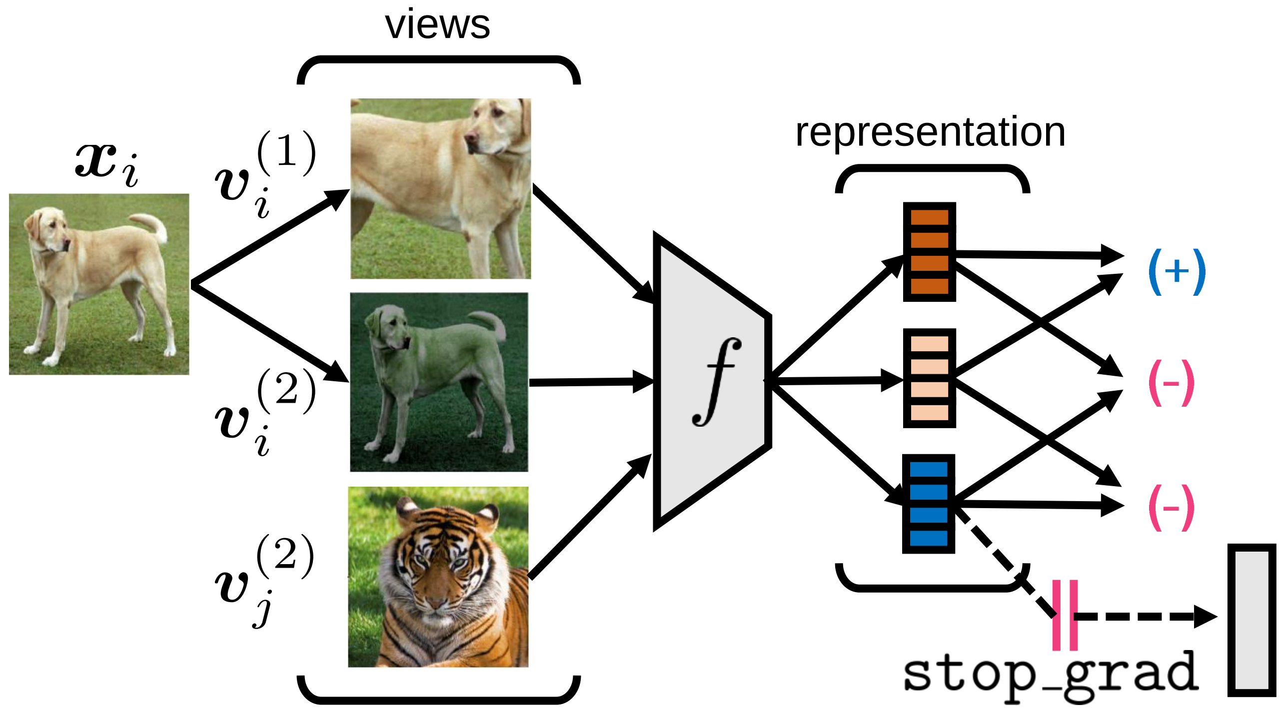
[Zhang et al., 2020] Consistency Regularization for Generative Adversarial Networks. ICLR 2020.
[ZhaoZ et al., 2020a] Improved Consistency Regularization for GANs. 2020.
[ZhaoS et al., 2020] Differentiable Augmentation for Data-Efficient GAN Training, NeurIPS 2020.
[Karras et al., 2020a] Analyzing and Improving the Image Quality of StyleGAN, CVPR 2020.
[Chen et al., 2021] Ultra-Data-Efficient GAN Training: Drawing A Lottery Ticket First, Then Training It Toughly, 2021.

Contrastive Representation Learning

An encoder f is learned to extract the shared features between two views $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$
[van den Oord et al., 2018; He et al., 2019; Chen et al., 2020]

SimCLR [Chen et al., 2020] defines the views by "Resize + Crop + Flip + ColorJitter + Gray + GaussianBlur"

- Unlike current GANs, **contrastive learning** can much benefits from stronger augmentations



$$L_{\mathtt{InfoNCE}}(\boldsymbol{v}_i^{(1)}, \boldsymbol{v}_{1:K}^{(2)}) := -\log \frac{\exp s(f(\boldsymbol{v}_i^{(1)}), f(\boldsymbol{v}_i^{(2)}))}{\sum_{j=1}^{K} \exp s(f(\boldsymbol{v}_i^{(1)}), f(\boldsymbol{v}_j^{(2)}))}$$

a "similarity" function

Fine-tuning for downstream tasks
(e.g. classification with labels)

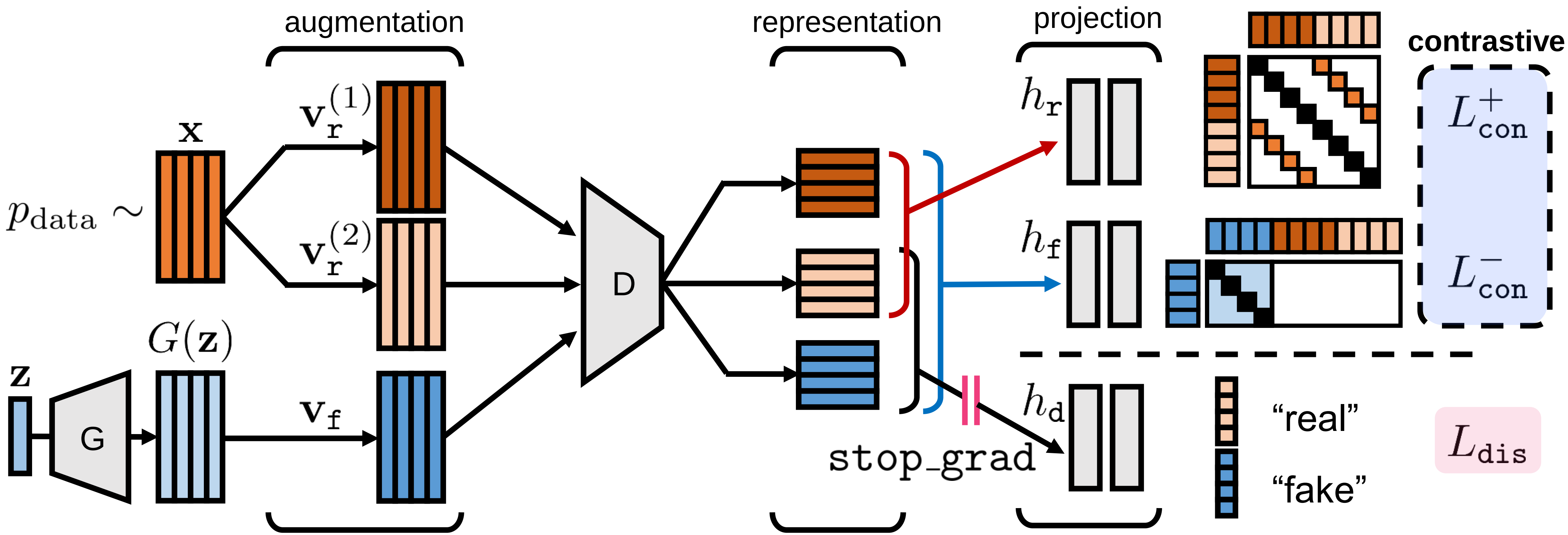
[van den Oord et al., 2018] Representation Learning with Contrastive Predictive Coding, NeurIPS 2018.
[He et al., 2019] Momentum Contrast for Unsupervised Visual Representation Learning. CVPR 2020.
[Chen et al., 2020] A Simple Framework for Contrastive Learning of Visual Representations, ICML 2020.

Contrastive Learning for GAN Discriminators?

🤔 Can we leverage the "SimCLR" augmentations for training GAN?

💡 **ContraD**: We propose a contrastive learning scheme for GAN discriminators

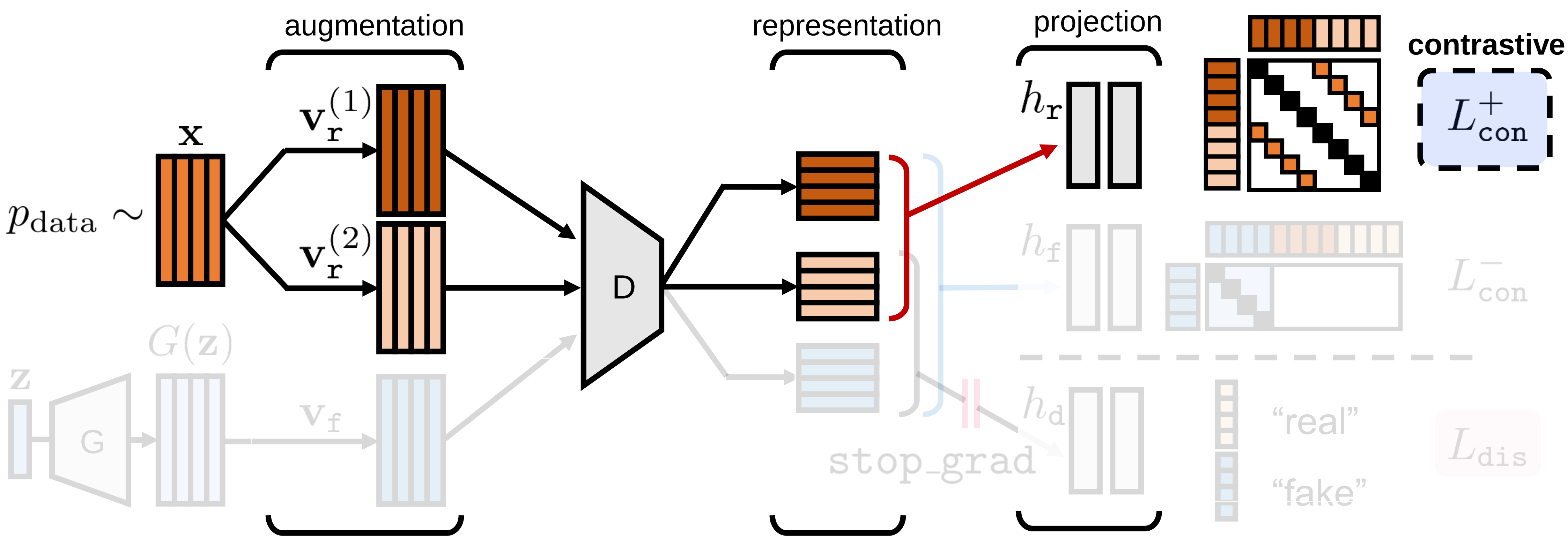
- Modifies only the discriminator objective upon **any GAN training**
- **Idea**: D is NOT directly optimized for the GAN loss $L_{\mathtt{dis}}$, but its contrastive alternative $L^{+}_{\mathtt{con}} + L^{-}_{\mathtt{con}}$



ContraD: the Contrastive Discriminator

1. **ContraD is equivalent to SimCLR** [Chen et al., 2020] **for the "real" samples**
   - We can naturally adopt the strong augmentations from SimCLR to train D

$$L^{+}_{\texttt{con}}(D, h_{\texttt{r}}) := L_{\texttt{SimCLR}}(v^{(1)}_{\texttt{r}}, v^{(2)}_{\texttt{r}}; D, h_{\texttt{r}})$$

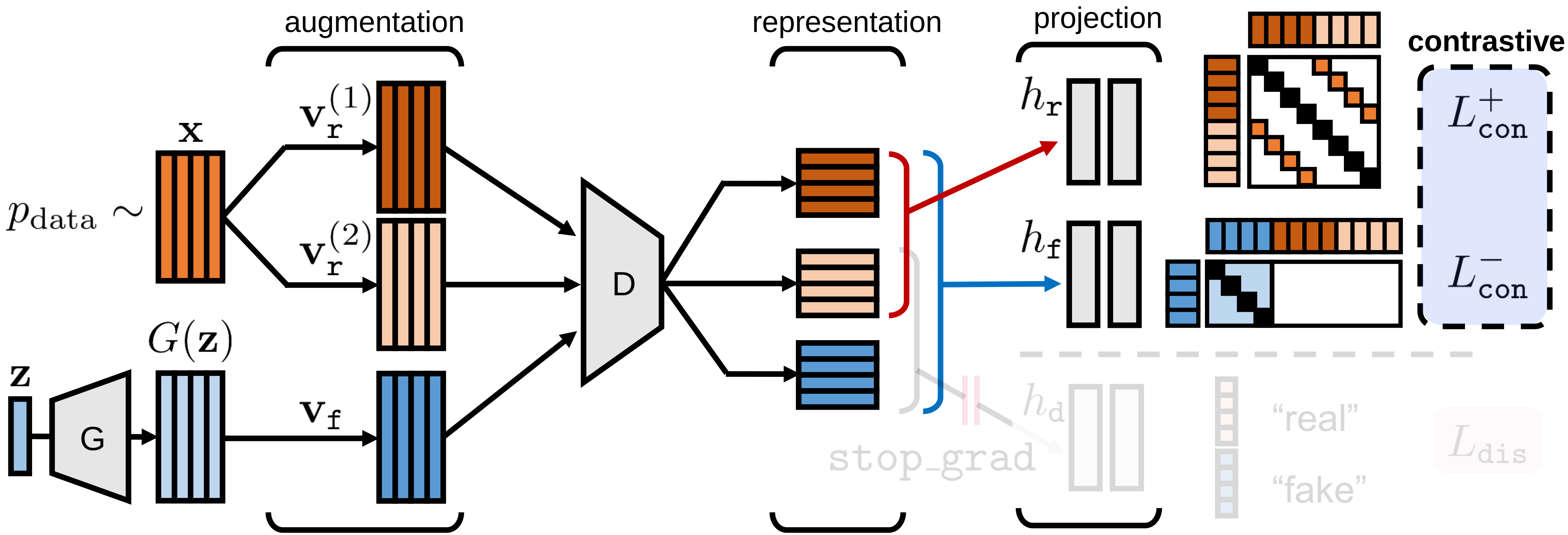


[Chen et al., 2020] A Simple Framework for Contrastive Learning of Visual Representations, ICML 2020.

ContraD: the Contrastive Discriminator

2. $L^{+}_{\texttt{con}}$ may not be enough to discriminate real vs. fake
   - Supervised Contrastive Learning [Khosla et al., 2020] for the "fake" samples

$$L^{-}_{\texttt{con}}(\boldsymbol{v}_{\texttt{f},i}; D, h_{\texttt{f}}) := -\frac{1}{N-1}\sum_{j\neq i}^{N} \log \frac{\exp(s(\boldsymbol{v}_{\texttt{f},i}, \boldsymbol{v}_{\texttt{f},j}))}{\sum_{v^{(2)}} \exp(s(\boldsymbol{v}_{\texttt{f},i}, \boldsymbol{v}^{(2)}))}.$$

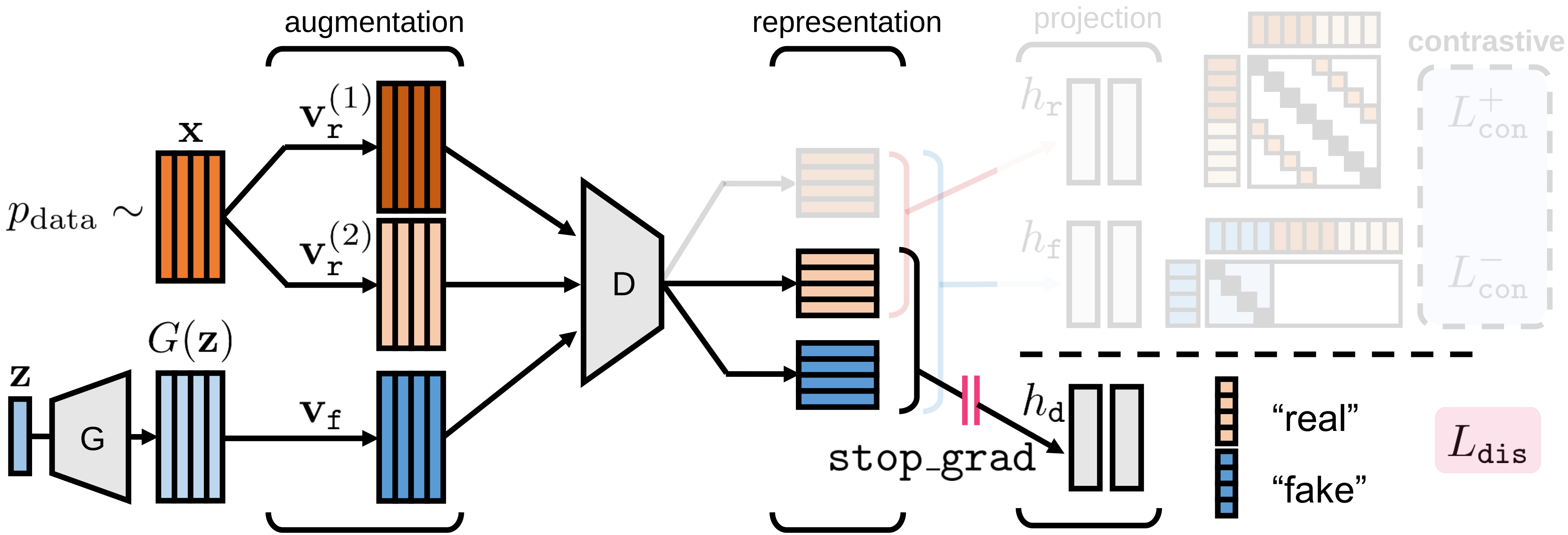


[Khosla et al., 2020] Supervised Contrastive Learning, NeurIPS 2020.

ContraD: the Contrastive Discriminator

3. The "actual" discriminator = 2-layer NN h_d upon the contrastive representation
   - $L_{\mathtt{dis}}$ is minimized only at h_d to maintain the GAN dynamics
   - $L_{\mathtt{dis}}$ does not affect D, due to the `stop_grad` operation in between

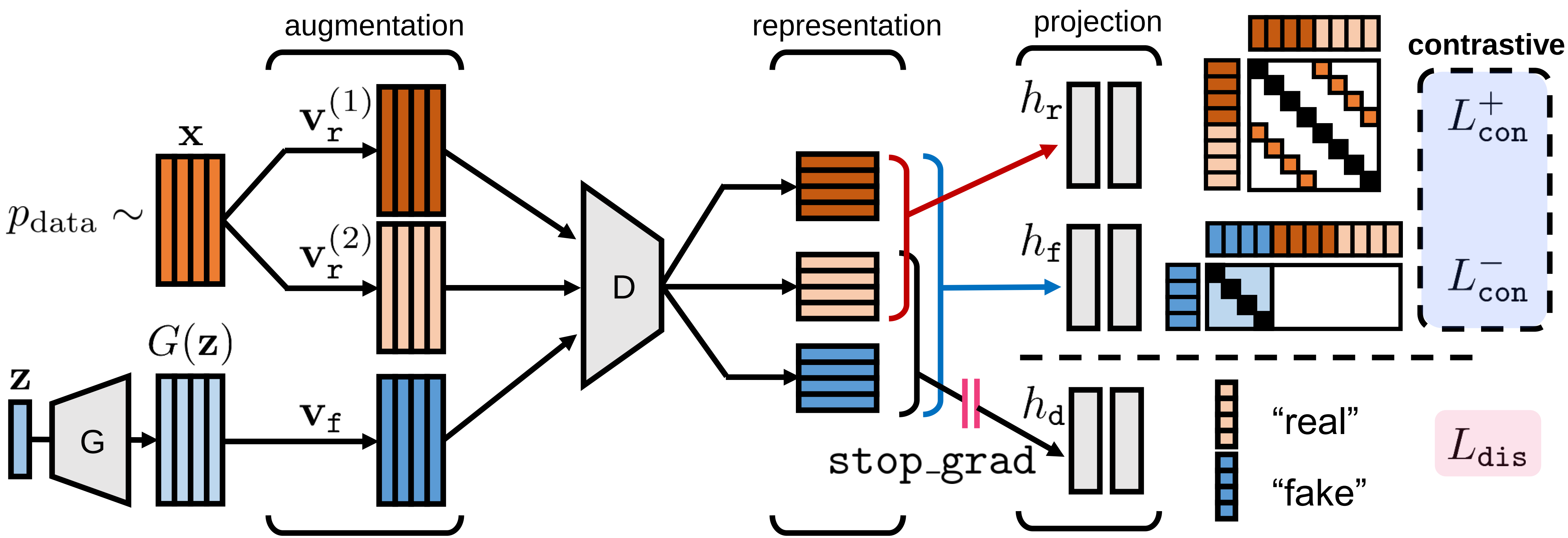
$$L_{\mathtt{dis}}(h_{\mathtt{d}}) := -\mathbb{E}[\log h_{\mathtt{d}}(\mathtt{sg}(D(\mathbf{v}_{\mathtt{r}})))] - \mathbb{E}[\log(1 - h_{\mathtt{d}}(\mathtt{sg}(D(\mathbf{v}_{\mathtt{f}}))))]$$



ContraD: the Contrastive Discriminator

The full ContraD training is an alternating minimization of L_D and L_G, like other GANs

$$L_D := L^+_{\texttt{con}} + L^-_{\texttt{con}} + L_{\texttt{dis}}$$
$$L_G := -\mathbb{E}[\log h_{\texttt{d}}(D(v_{\texttt{f}}))]$$



Experiments: ContraD improves GAN

ContraD significantly improves GANs by successfully incorporating the SimCLR augmentations

- SimCLR = “Resize + Crop + Flip + ColorJitter + Gray + GaussianBlur”

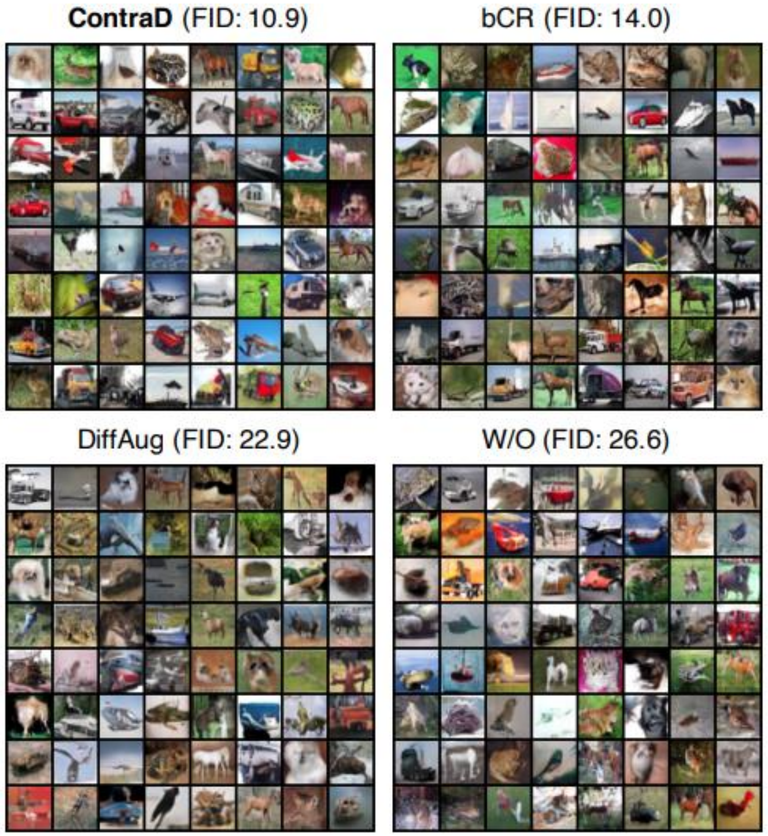


Table 1: Comparison of the best FID score and IS on unconditional image generation of CIFAR-10 and CIFAR-100. Values in the rows marked by * are from those reported in its reference.

Architecture	Method	Augment.	CIFAR-10 FID ↓	CIFAR-10 IS ↑	CIFAR-100 FID ↓	CIFAR-100 IS ↑
G: SNDCGAN D: SNDCGAN	-	-	26.6	7.38	28.5	7.25
	CR (Zhang et al., 2020)	HFlip, Trans	19.5	7.87	22.2	7.91
	bCR (Zhao et al., 2020c)	HFlip, Trans	14.0	8.35	19.2	8.46
	DiffAug (Zhao et al., 2020a)	Trans, CutOut	22.9	7.64	27.0	7.47
	ContraD (ours)	**SimCLR**	**10.9**	**8.78**	**15.2**	**9.09**
G: SNDCGAN D: SNResNet-18	-	-	41.3	6.33	52.3	5.24
	CR (Zhang et al., 2020)	HFlip, Trans	32.1	7.08	36.5	6.55
	bCR (Zhao et al., 2020c)	HFlip, Trans	22.8	7.29	28.2	7.30
	DiffAug (Zhao et al., 2020a)	Trans, CutOut	59.5	5.62	58.7	5.39
	ContraD (ours)	**SimCLR**	**9.86**	**9.09**	**15.0**	**9.56**
G: StyleGAN2 D: StyleGAN2	-	-	11.1	9.18	16.5	9.51
	DiffAug* (Zhao et al., 2020a)	Trans, CutOut	9.89	9.40	15.2	**10.0**
	ContraD (ours)	**SimCLR**	**9.80**	**9.47**	**14.1**	**10.0**

[Zhang et al., 2020] Consistency Regularization for Generative Adversarial Networks. ICLR 2020.
[Zhao et al., 2020c] Improved Consistency Regularization for GANs. 2020.
[Zhao et al., 2020a] Differentiable Augmentation for Data-Efficient GAN Training, NeurIPS 2020.

Experiments: ContraD improves GAN

ContraD significantly improves GANs by successfully incorporating the SimCLR augmentations

- **Less sensitive to architecture**: ContraD could offer a stable training even when **G** (= DCGAN) ≪ **D** (= ResNet-18)

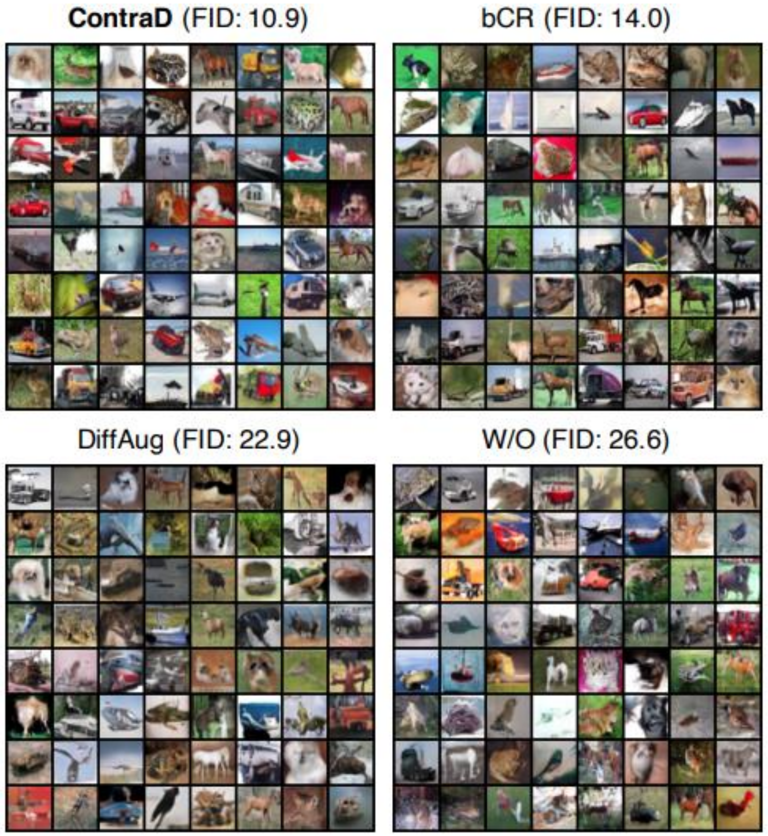


Table 1: Comparison of the best FID score and IS on unconditional image generation of CIFAR-10 and CIFAR-100. Values in the rows marked by * are from those reported in its reference.

Architecture	Method	Augment.	CIFAR-10 FID ↓	CIFAR-10 IS ↑	CIFAR-100 FID ↓	CIFAR-100 IS ↑
G: SNDCGAN D: SNDCGAN	-	-	26.6	7.38	28.5	7.25
	CR (Zhang et al., 2020)	HFlip, Trans	19.5	7.87	22.2	7.91
	bCR (Zhao et al., 2020c)	HFlip, Trans	14.0	8.35	19.2	8.46
	DiffAug (Zhao et al., 2020a)	Trans, CutOut	22.9	7.64	27.0	7.47
	ContraD (ours)	**SimCLR**	**10.9**	**8.78**	**15.2**	**9.09**
G: SNDCGAN D: SNResNet-18	-	-	41.3	6.33	52.3	5.24
	CR (Zhang et al., 2020)	HFlip, Trans	32.1	7.08	36.5	6.55
	bCR (Zhao et al., 2020c)	HFlip, Trans	22.8	7.29	28.2	7.30
	DiffAug (Zhao et al., 2020a)	Trans, CutOut	59.5	5.62	58.7	5.39
	ContraD (ours)	**SimCLR**	**9.86**	**9.09**	**15.0**	**9.56**
G: StyleGAN2 D: StyleGAN2	-	-	11.1	9.18	16.5	9.51
	DiffAug* (Zhao et al., 2020a)	Trans, CutOut	9.89	9.40	15.2	**10.0**
	ContraD (ours)	**SimCLR**	**9.80**	**9.47**	**14.1**	**10.0**

[Zhang et al., 2020] Consistency Regularization for Generative Adversarial Networks. ICLR 2020.
[Zhao et al., 2020c] Improved Consistency Regularization for GANs. 2020.
[Zhao et al., 2020a] Differentiable Augmentation for Data-Efficient GAN Training, NeurIPS 2020.

Experiments: ContraD improves SimCLR

Interestingly, ContraD could also improve the underlying SimCLR as well

- Better linear evaluation and transfer learning performance than only $L^{+}_{\texttt{con}}$ is minimized (= SimCLR)
  - **"Linear evaluation"?** - Train a linear classifier w/ labels upon the frozen D-representation
- Tested on **CIFAR-10/100 (top)** and **ImageNet (bottom)** datasets

Table 3: Comparison of classification accuracy under linear evaluation protocol on CIFAR-10 and CIFAR-100. We report the mean and standard deviation across 3 runs of the evaluation.

Dataset	Training	SNDCGAN	SNResNet-18	StyleGAN2
CIFAR-10	SimCLR ($\lambda_{\texttt{con}} = \lambda_{\texttt{dis}} = 0$)	72.9$\pm$0.02	80.3$\pm$0.05	86.2$\pm$0.06
	ContraD (ours)	**77.5**$\pm$0.20	**85.7**$\pm$0.10	**88.6**$\pm$0.06
CIFAR-100	SimCLR ($\lambda_{\texttt{con}} = \lambda_{\texttt{dis}} = 0$)	30.8$\pm$0.11	41.2$\pm$0.06	61.1$\pm$0.06
	ContraD (ours)	**37.4**$\pm$0.06	**51.1**$\pm$0.18	**68.1**$\pm$0.07

Table 9: Comparison linear evaluation and transfer learning performance across 6 natural image classification datasets for BigGAN discriminators pretrained on ImageNet (64 × 64). We report the top-1 accuracy except for ImageNet and SUN397, which we instead report the top-5 accuracy.

Training (BigGAN)	ImageNet	CIFAR10	CIFAR100	DTD	SUN397	Flowers	Food
Supervised (ImageNet)	63.5	76.1	55.2	45.4	31.7	78.1	44.5
SimCLR ($\lambda_{\texttt{con}} = \lambda_{\texttt{dis}} = 0$)	43.4	81.2	55.3	43.9	37.6	69.8	38.8
ContraD (ours)	**51.5**	**84.5**	**61.1**	**50.6**	**44.4**	**78.6**	**44.5**

Experiments

ContraD significantly improves GANs by successfully incorporating the SimCLR augmentations

- **Less sensitive to architecture**: ContraD could offer a stable training even when **G** (= DCGAN) ≪ **D** (= ResNet-18)

Interestingly, ContraD could also improve the underlying SimCLR as well

- Better linear evaluation and transfer learning performance than only $L^{+}_{\mathtt{con}}$ is minimized (= SimCLR)

… And many more results can be found in the full paper!

- **More challenging datasets:** CelebA-HQ (128 × 128), AnimalFaces-HQ (512 × 512), and ImageNet w/ BigGAN
  - ContraD works for a wide range of datasets, especially under regime of limited data
- **Application of ContraD:** Self-conditional sampling
  - ContraD can induce many cGANs leveraging the learned contrastive representation
- **Detailed ablation study**

Summary

TL;DR: GAN and SimCLR benefit each other when they are jointly trained

We propose ContraD = Contrastive learning for GAN discriminators

1. Enables stronger data augmentation → improved, data-efficient GAN training
2. Can improve the underlying contrastive learning as well
3. Still maintains “contrastive representation” → other downstream tasks

More details can be found:

- **Paper**: https://arxiv.org/abs/2103.09742
- **Code**: https://github.com/jh-jeong/ContraD

Please drop by our poster session for more information!