Training Generative Adversarial Networks with Limited Data

NeurlPS 2020

Catalogue

- Background
- Related work
- Introduction
- Model
- Conclusion

Background

Limited Data

- It remains challenging to collect a large enough set of images for a specific application that places constraints on subject type, image quality, geographical location, time period, privacy, copyright status…
- Specifically, in some application about medicine…

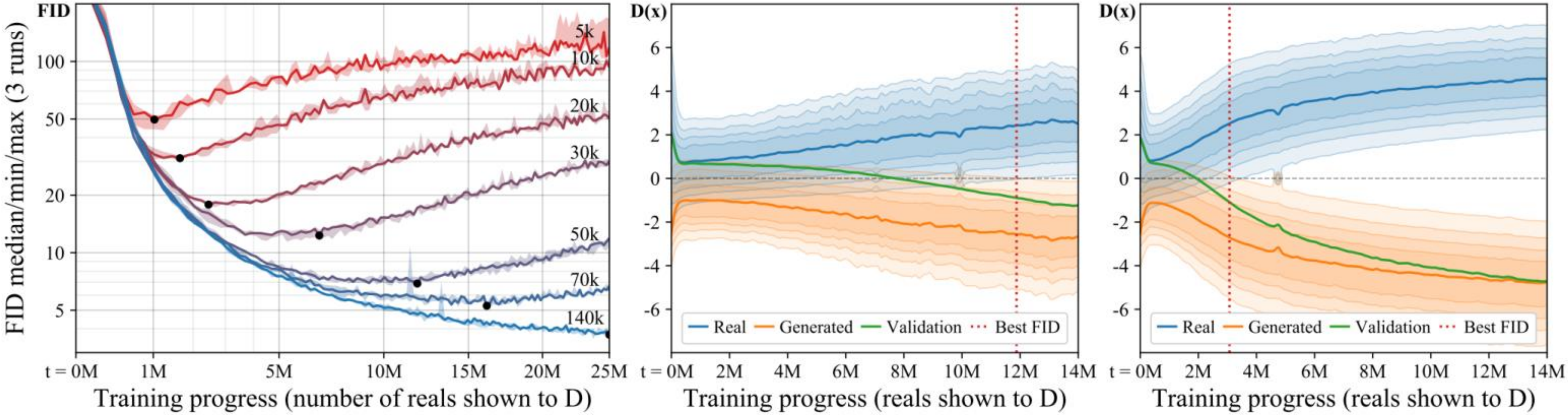
GAN

- Based on large dataset
- The key problem of training on small dataset is the discriminator overfits to the training examples which will lead to mode collapse…

Training Generative Adversarial Networks with Limited Data

- Challenging but useful

Background

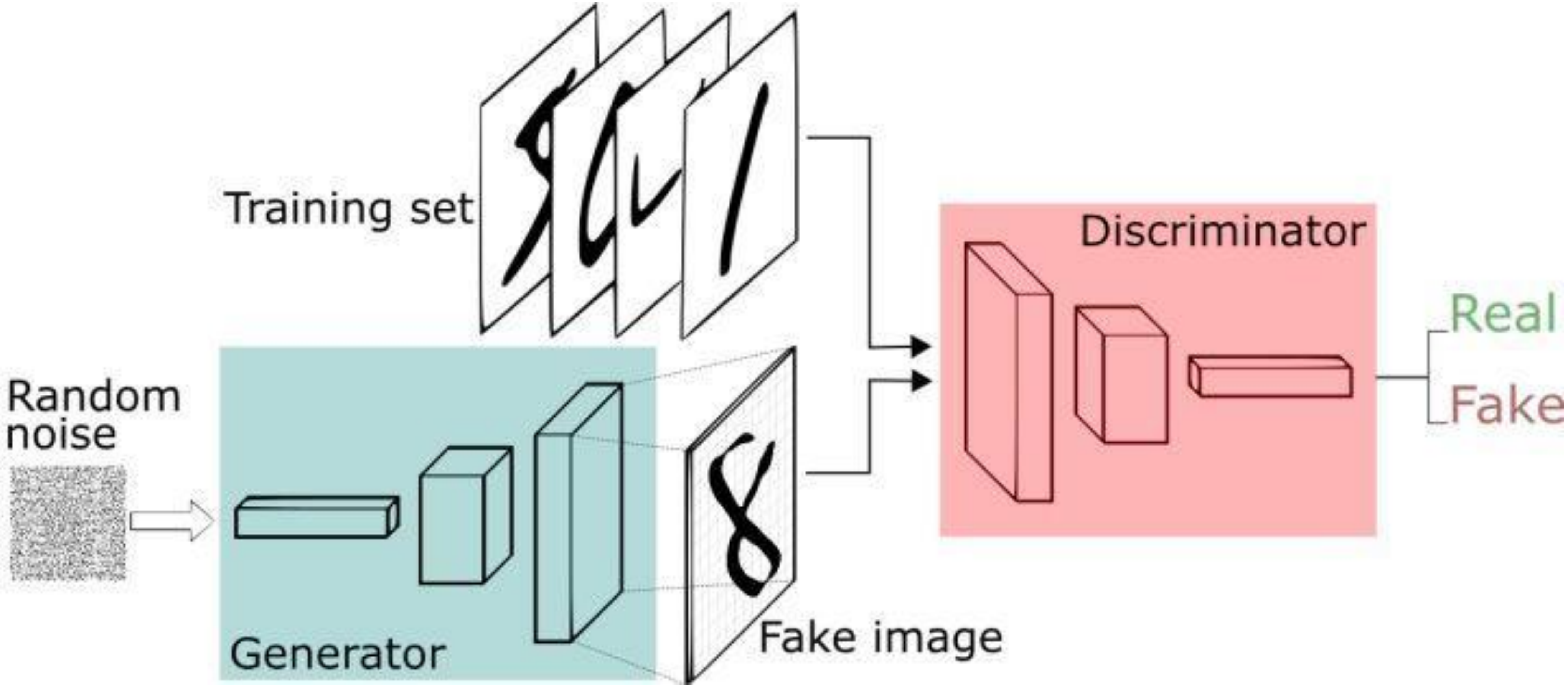


(a) Convergence of FFHQ (256 × 256) (b) Discriminator outputs, 50k (c) Discriminator outputs, 20k

Figure 1: (a) Convergence with different training set sizes. "140k" means that we amplified the 70k dataset by 2× through x-flips; we do not use data amplification in any other case. (b,c) Evolution of discriminator outputs during training. Each vertical slice shows a histogram of $D(x)$, i.e., raw logits.

Background

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$



Related work

- Data Augmentation

leaking augmentations

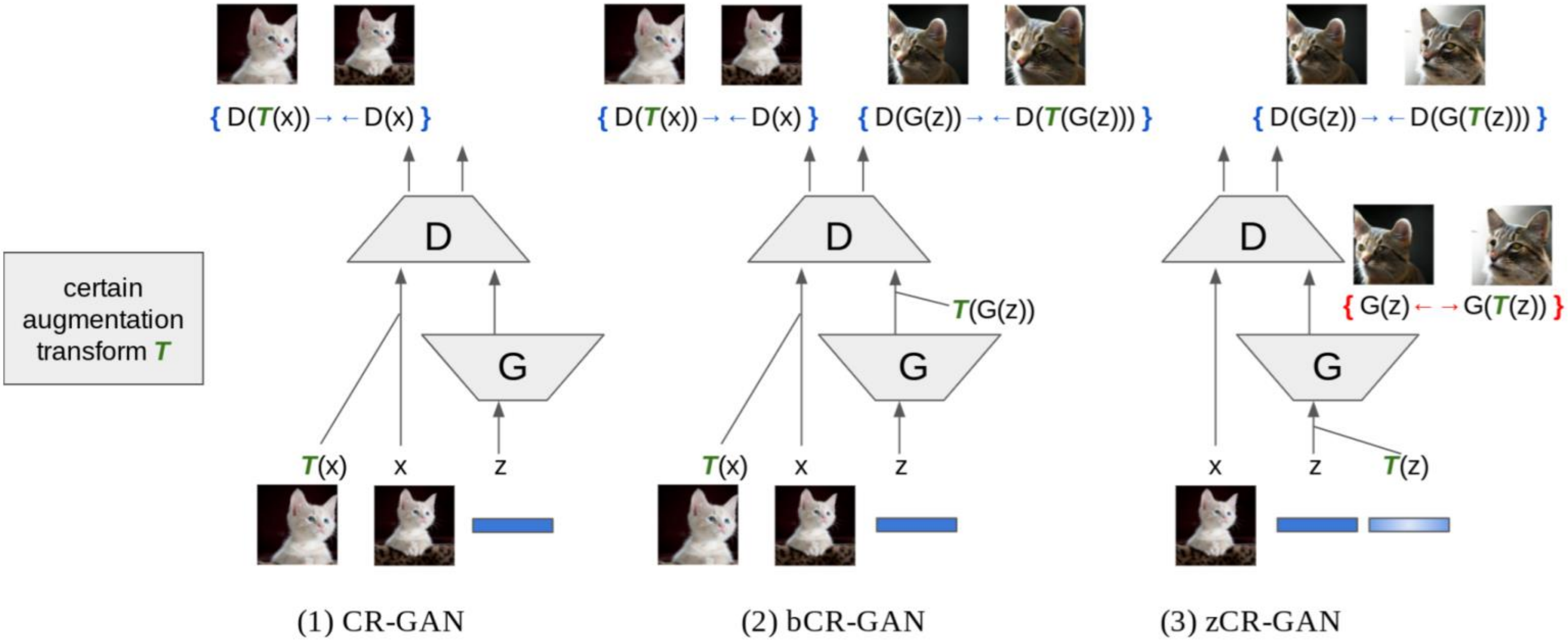
- Consistency Regularization (CR-GAN 2020,bCR-GAN 2020)

leaking augmentations

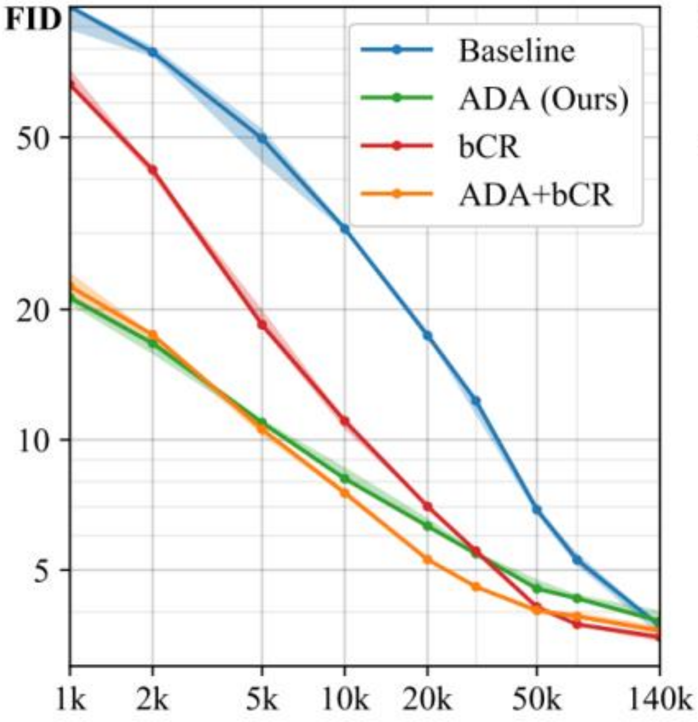
- Stochastic Discriminator Augmentation

(a) 8 × 8 cutout. (b) CR samples. (c) bCR samples.
(d) 16 × 16 cutout. (e) CR samples. (f) bCR samples.
(g) 32 × 32 cutout. (h) CR samples. (i) bCR samples.

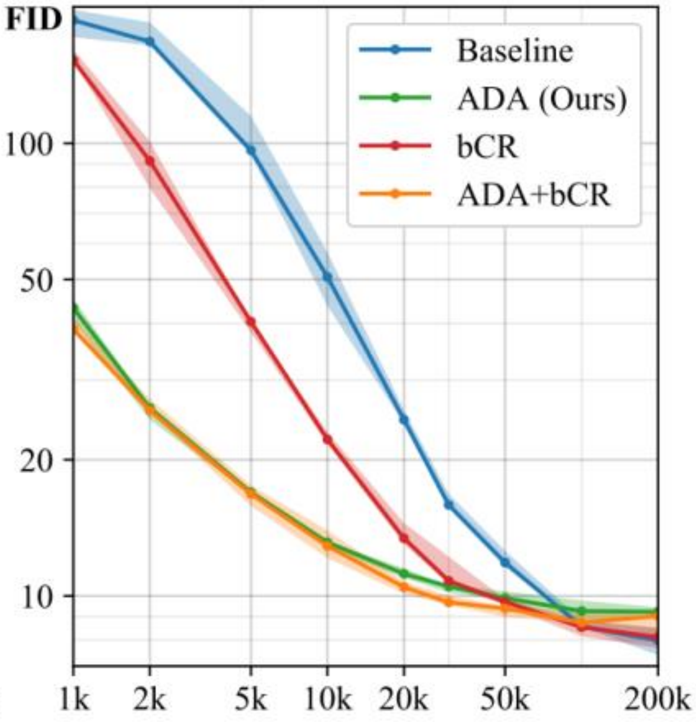
Related work



Introduction



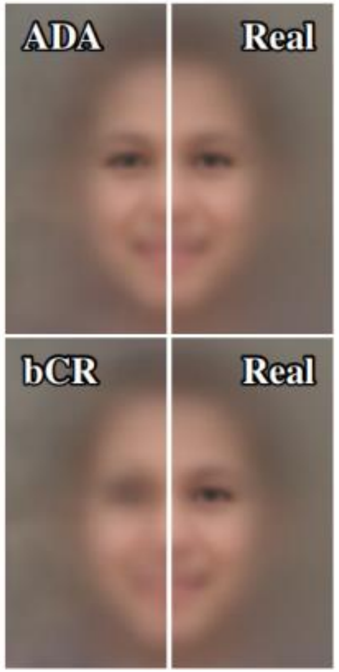
(a) FFHQ (256×256)



(b) LSUN CAT (256×256)

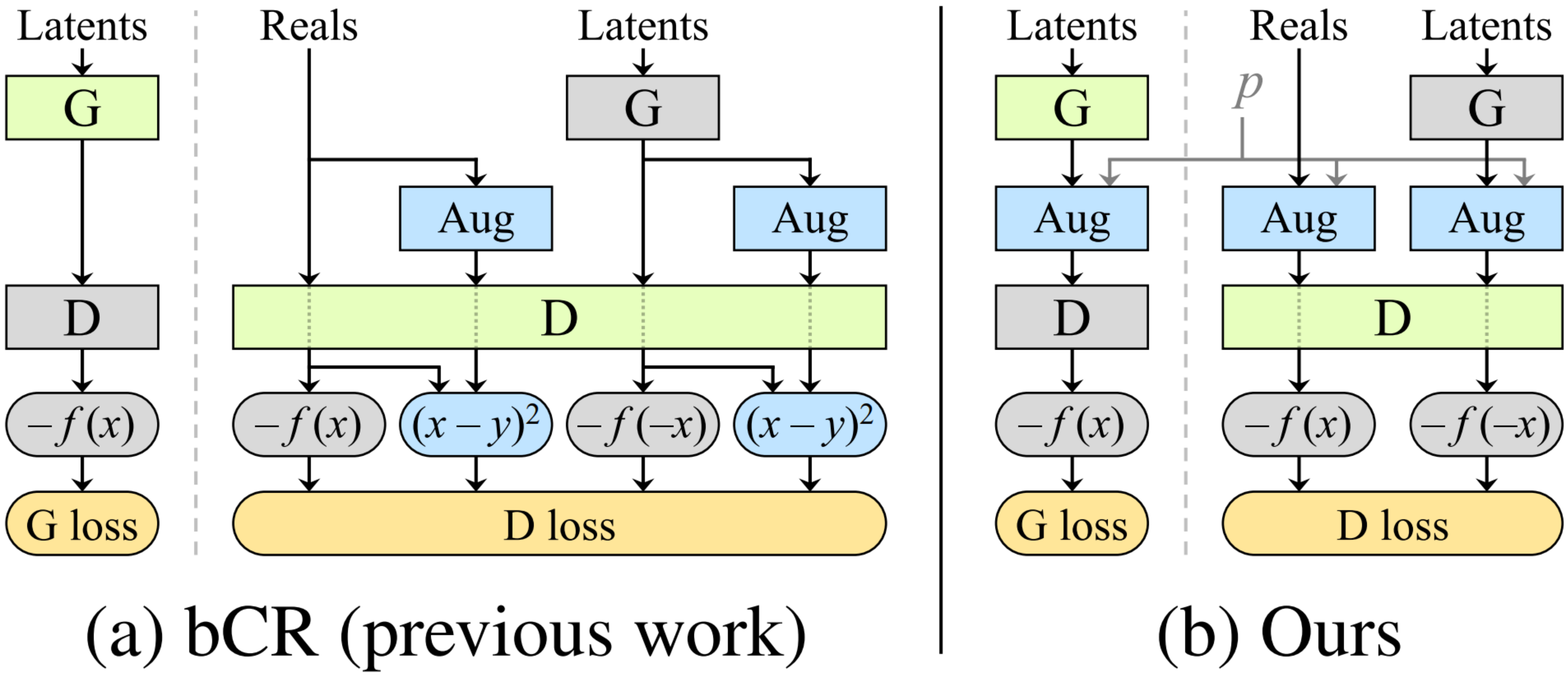
Dataset		Baseline	ADA	+ bCR
FFHQ	1k	100.16	**21.29**	22.61
	5k	49.68	10.96	**10.58**
	10k	30.74	8.13	**7.53**
	30k	12.31	5.46	**4.57**
	70k	5.28	4.30	**3.91**
	140k	3.71	3.81	**3.62**
LSUN CAT	1k	186.91	43.25	**38.82**
	5k	96.44	16.95	**16.80**
	10k	50.66	13.13	**12.90**
	30k	15.90	10.50	**9.68**
	100k	**8.56**	9.26	8.73
	200k	**7.98**	9.22	9.03

(c) 中位数FID



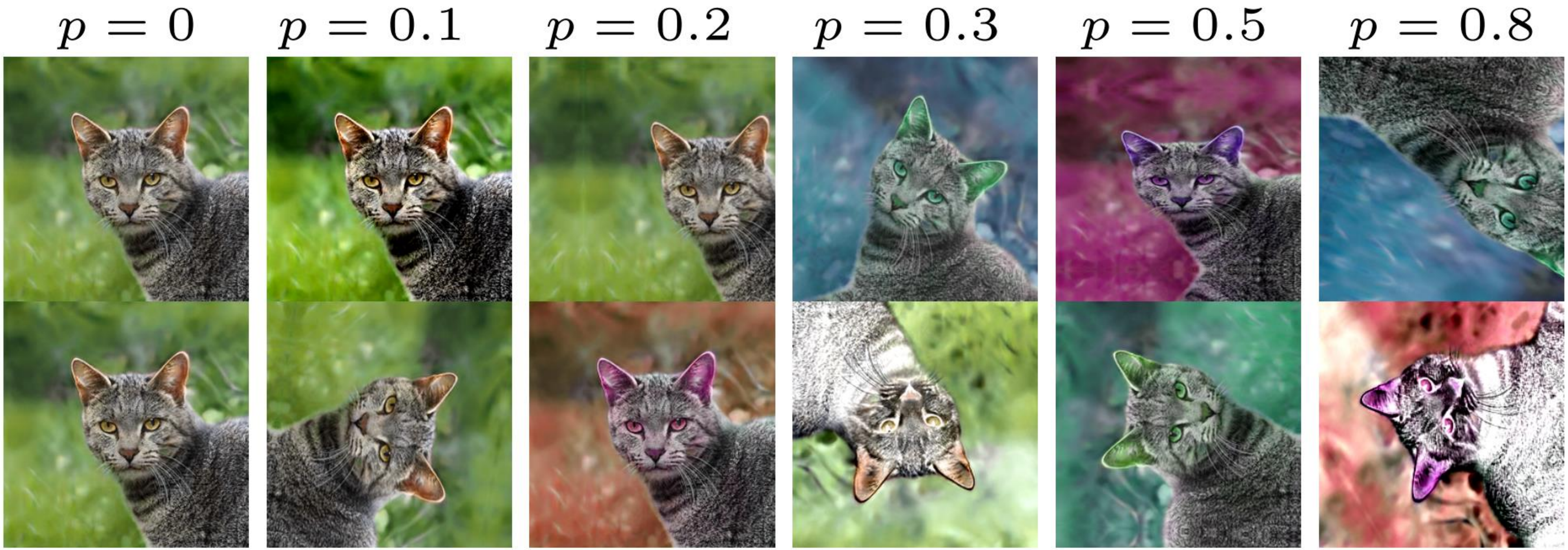
(d) 均值图像

Model



Latents
G
D
$-f(x)$
G loss
Reals
Latents
G
Aug
Aug
D
$-f(x)$
$(x-y)^2$
$-f(-x)$
$(x-y)^2$
D loss
(a) bCR (previous work)
Latents
G
Aug
D
$-f(x)$
G loss
p
Reals
Latents
G
Aug
Aug
D
$-f(x)$
$-f(-x)$
D loss
(b) Ours

Model

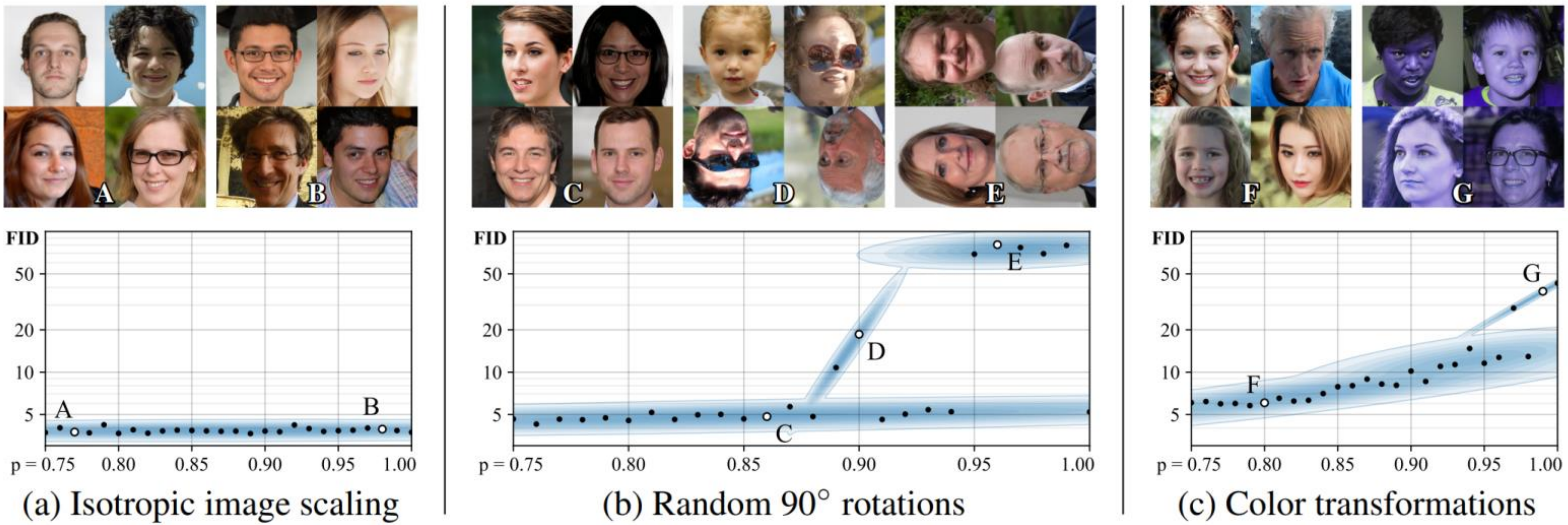


(c) Effect of augmentation probability p

Model

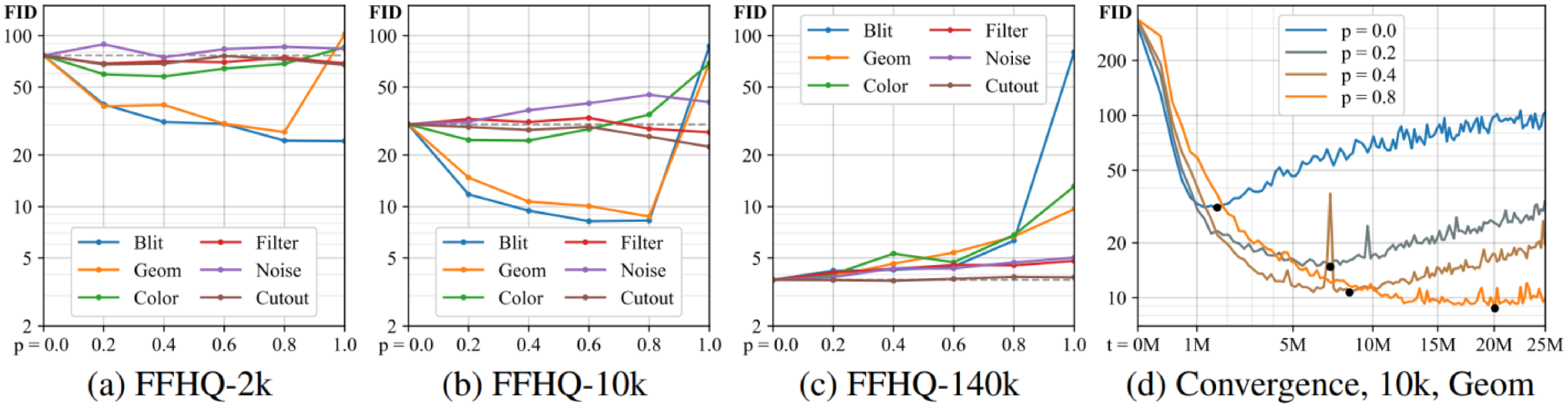
- the training implicitly undoes the corruptions and finds the correct distribution, as long as the corruption process is represented by an invertible transformation of probability distributions over the data space;
- this does not mean that augmentations performed on individual images would need to be undoable;
- this rotation is only executed at a probability $p < 1$: this increases the relative occurrence of 0;

Model



This experiment suggests that as long as p remains below 0.8, leaks are unlikely to happen in practice.

Model

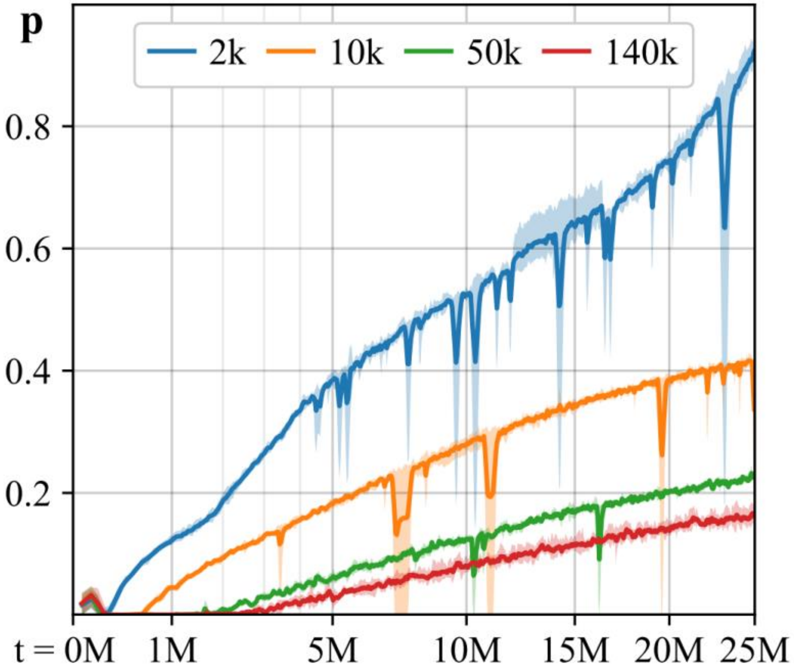


18 transformations that are grouped into 6 categories

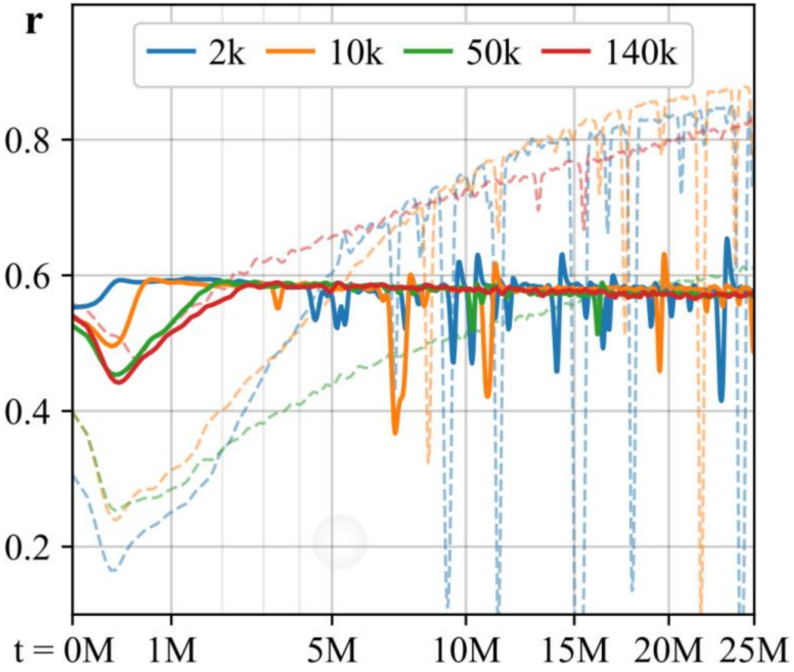
Model-adaptive discriminator augmentation (ADA)

$$r_v = \frac{\mathbb{E}[D_{\text{train}}] - \mathbb{E}[D_{\text{validation}}]}{\mathbb{E}[D_{\text{train}}] - \mathbb{E}[D_{\text{generated}}]}$$

$$r_t = \mathbb{E}[\text{sign}(D_{\text{train}})]$$

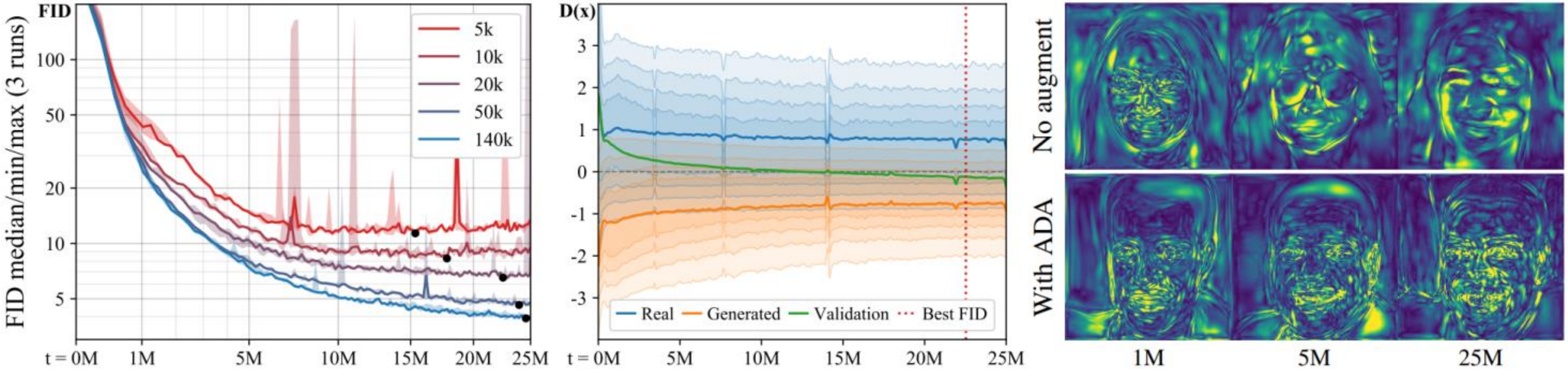


(c) Evolution of p over training



(d) Evolution of r_t

Experiment



(a) With adaptive augmentation (b) Discriminator outputs, 20k (c) Discriminator gradients, 10k

Conclusion

- the optimal augmentation strength depends heavily on the amount of training data, and not all augmentation categories are equally useful in practice
- Compared with Deceive D

Thanks for listening.