A High-Quality Robust Diffusion Framework for Corrupted Dataset

The 18th European Conference on Computer Vision ECCV 2024

Quan Dao[1,2]*, Binh Ta[1]*, Tung Pham[1], Anh Tran[1]

* Equal contribution
[1] VinAI Research
[2] Rutgers University

Motivation

- Even though **diffusion models** has strong mode coverage on training dataset, this property could hurt its performance on corrupted training datasets

- We show that diffusion models tend to generate outliers

i.e Denoising Diffusion GAN (DDGAN) generates outlier image of church when there are church outliers in CelebA training dataset.

Motivation

- Even though **diffusion models** has strong mode coverage on training dataset, this property could hurt its performance on corrupted training datasets

- We show that diffusion models tend to generate outliers

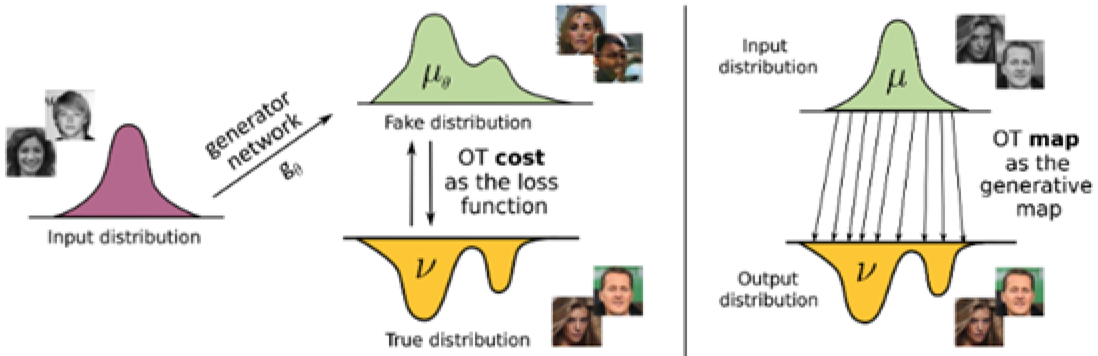
i.e Denoising Diffusion GAN (DDGAN) generates outlier image of church when there are church in CelebA training dataset.

➡ We want to develop a diffusion framework that is robust to corrupted training datasets

Robust Diffusion Unbalanced Optimal Transport

We propose an RDUOT which uses the UOT generative model to learn the backward transitions instead of using GAN as in DDGAN.

- DDGAN learns to minimize distance between fake $p_\theta(x_t|x_{t-1})$ and real $q(x_t|x_{t-1})$. It cannot filter the outliers since real data contains them.
- RDUOT learns map to minimize distance from source $q(x_t)$ to target $q(x_{t-1})$, which could filter outliers as the outlier is often far from clean data.



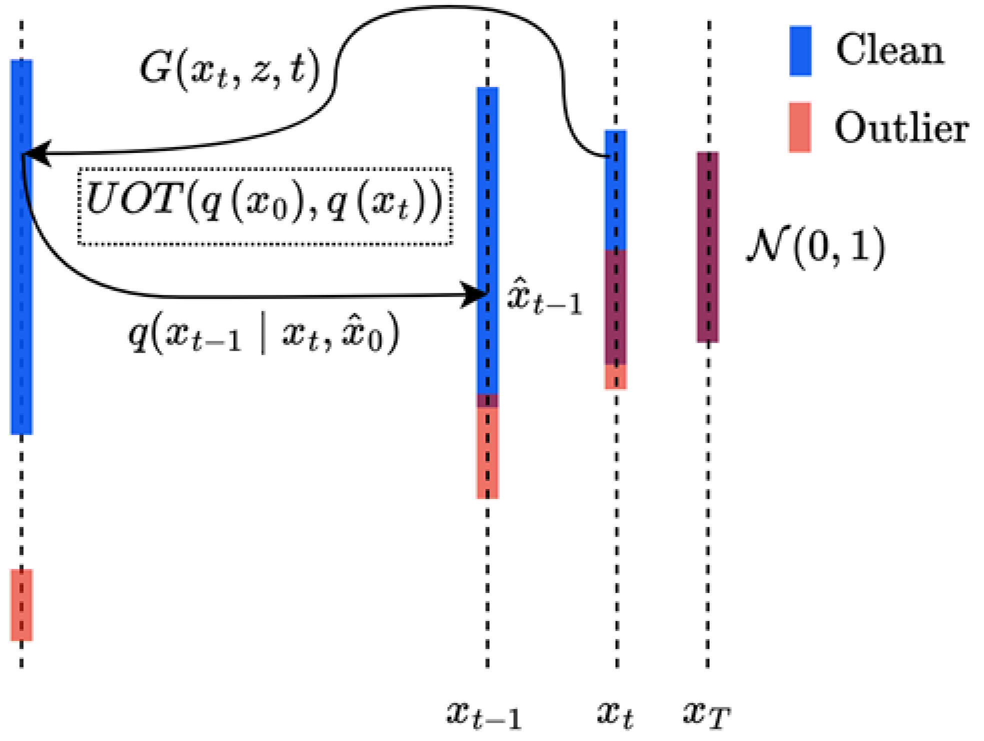
(a) OT cost as the loss for the generative model. (b) OT map as the generative model.

Rout, et al. GENERATIVE MODELING WITH OPTIMAL TRANSPORT MAPS (ICLR2022)

Robust Diffusion Unbalanced Optimal Transport

- As t increases, clean and outlier distribution is mixed up, challenge UOT to filter them. Hence, we learn UOT map to minimize distance from $q(x_t)$ to $q(x_0)$ since in $q(x_0)$, outlier and clean are distinguishable.

- Use Softplus as convex conjugate instead of KL and χ^2 due to its Lipschitz property allows fast convergence



Corrupted Dataset

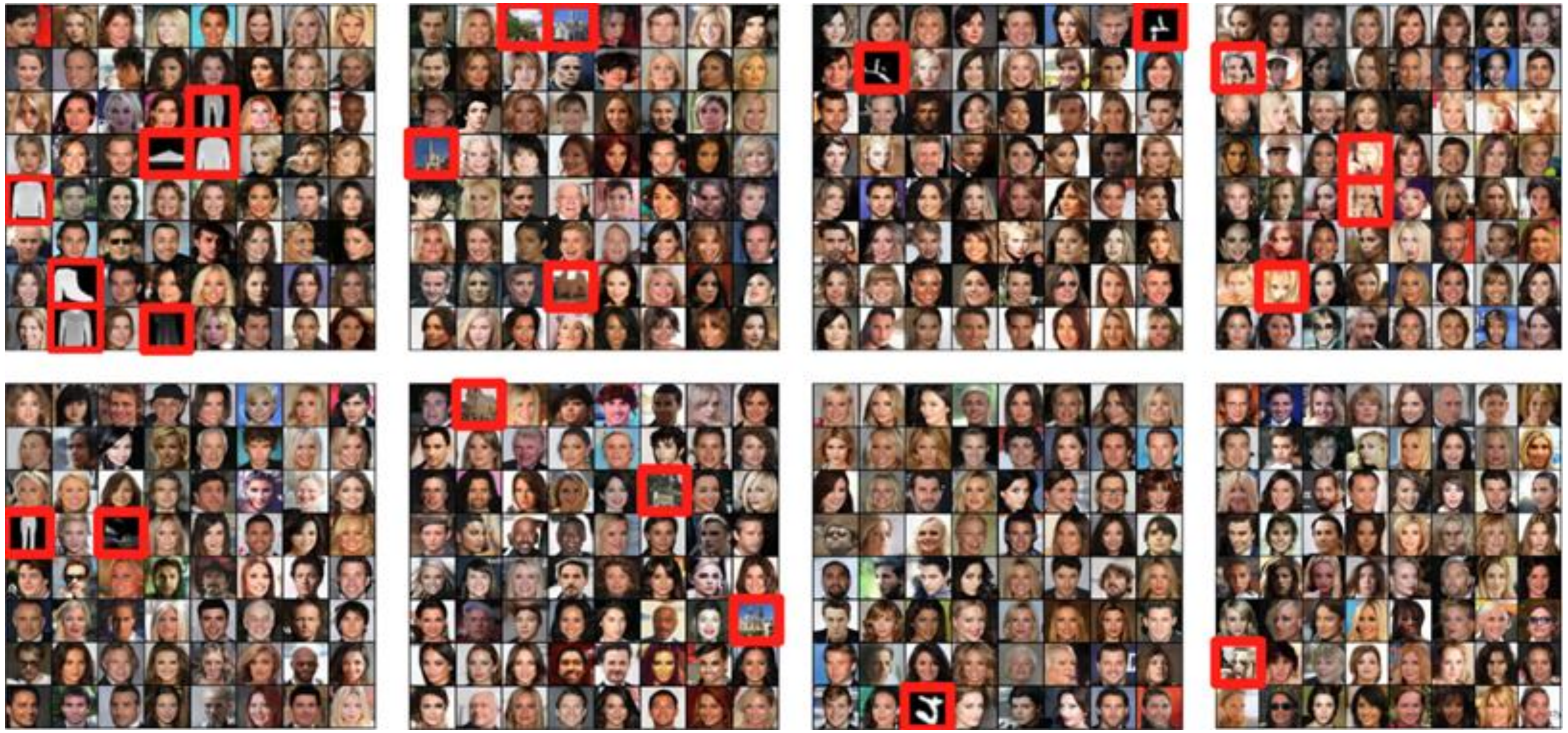
	Synthesized Outlier		FID	
Perturb ratio	DDGAN	RDUOT	DDGAN	RDUOT
3%	3.2%	0.2%	4.76	3.43
5%	4.1%	1.7%	8.81	4.37
7%	6.9%	2.3%	9.55	5.17
10%	9.8%	3.8%	14.77	6.98

Various Outlier Ratio: CIFAR-10 perturbed by MNIST

	CI+3%MT	CI+5%MT	CE+5%FT	CE+5%CH	CE+5%FCE
RDUOT	**3.43**	**4.37**	**7.89**	**7.86**	**5.99**
UOTM	4.76	7.89	9.52	8.84	6.72
RobustGAN	10.63	10.68	-	-	-

Various Type of Outlier: CIFAR-10+MNIST and Celeb-A+(FashionMNIST, Church, flipped Celeb-A)

Corrupted Dataset



Top: DDGAN, Bottom: RDUOT
CelebA + 5%(FashionMNIST, Church, MNIST, flipped CelebA)

	RDUOT	DDGAN
CE+FT	7.89	10.68
CE+MT	9.29	12.95
CE+CH	7.86	9.83
CE+FCE	5.99	6.48

FID of RDUOT vs DDGAN

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO

Clean Dataset: Cifar10

Model	FID↓	Recall↑	NFE↓
RDUOT	**2.95**	**0.58**	4
WaveDiff	4.01	0.55	4
DDGAN	3.75	0.57	4
DDPM	3.21	0.57	1000
StyleGAN2	8.32	0.41	1
WGAN-GP	39.40	-	1
RobustGAN	21.57	-	1
RobustGAN*	11.40	-	1
OTM	21.78	-	1
UOTM	2.97	-	1
UOTM$^{\#}$	3.79	-	1

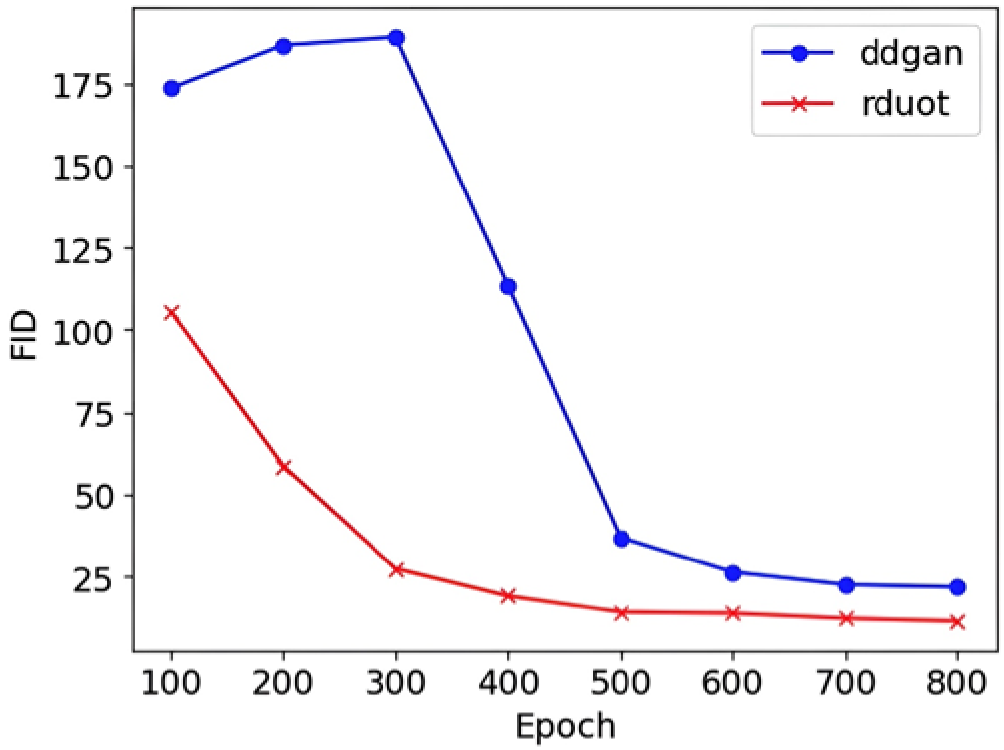
Clean Dataset: STL-10

Model	FID↓	Recall↑
Our	**11.50**	**0.49**
WaveDiff	12.93	0.41
DDGAN	21.79	0.40
StyleFormer	15.17	-
TransGAN	18.28	-
SNGAN	40.1	-
StyleGAN2+ADA	13.72	0.36
StyleGAN2+Aug	12.97	0.39
Diffusion StyleGAN2	11.53	-

Clean Dataset: CelebAHQ-256

Model	FID↓	Recall↑
RDOUT	**5.60**	**0.38**
WaveDiff	5.94	0.37
DDGAN	7.64	0.36
Score SDE	7.23	-
LFM	5.26	-
NVAE	29.7	-
VAEBM	20.4	-
PGGAN	8.03	-
VQ-GAN	10.2	-
UOTM	5.80	-

Convergence & Ablation Study



Training convergence STL-10

Ψ_1^*	Ψ_2^*	FID (clean) ↓	FID (5%) ↓
χ^2	χ^2	3.93	5.04
softplus	softplus	**2.95**	**4.37**

Cifar10+Mnist

Impact of convex conjugate function

Conclusion

A novel diffusion framework based on UOT loss.

Robust, high-fidelity and fast convergence