GitHub

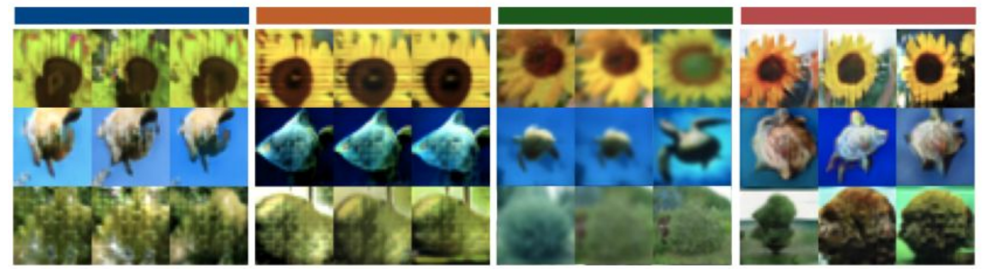
Taming the Tail in Class-Conditional GANs: Knowledge Sharing via Unconditional Training at Lower Resolutions

Saeed Khorram[1], Mingqi Jiang[1], Mohamad Shahbazi[2], Mohamad H. Danesh[3], Li Fuxin[1]

Oregon State University[1] ETH zürich[2] McGill University[3]

Class-Conditional GAN Mode Collapse on Long-tailed Data

- **Problem** Training class-conditional GANs on long-tailed data leads to mode collapse.
- Past work focuses on regularization, normalization, and/or class balancing techniques.
  - These are not helpful when tail classes are highly underrepresented
- We propose knowledge sharing between head and tail classes, agnostic of GAN architectures

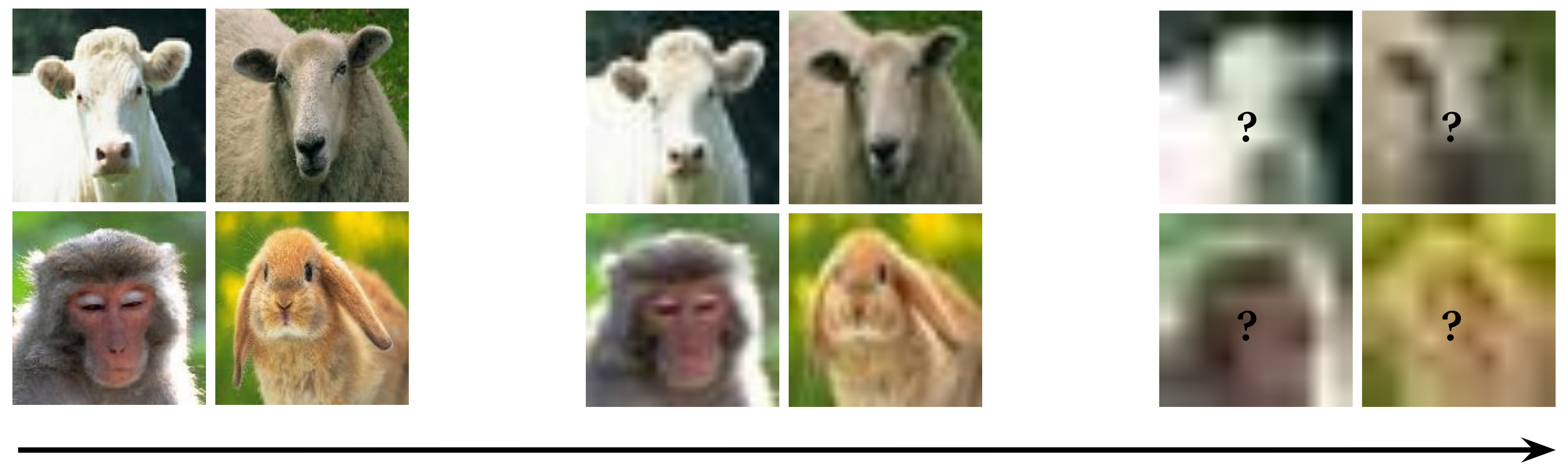


Motivation

- **Observation** head and tail classes are often more similar at lower resolutions

➔ Information at the lower resolutions tends to be more class-independent thus can be shared between the head and tail classes (e.g., background, configuration)

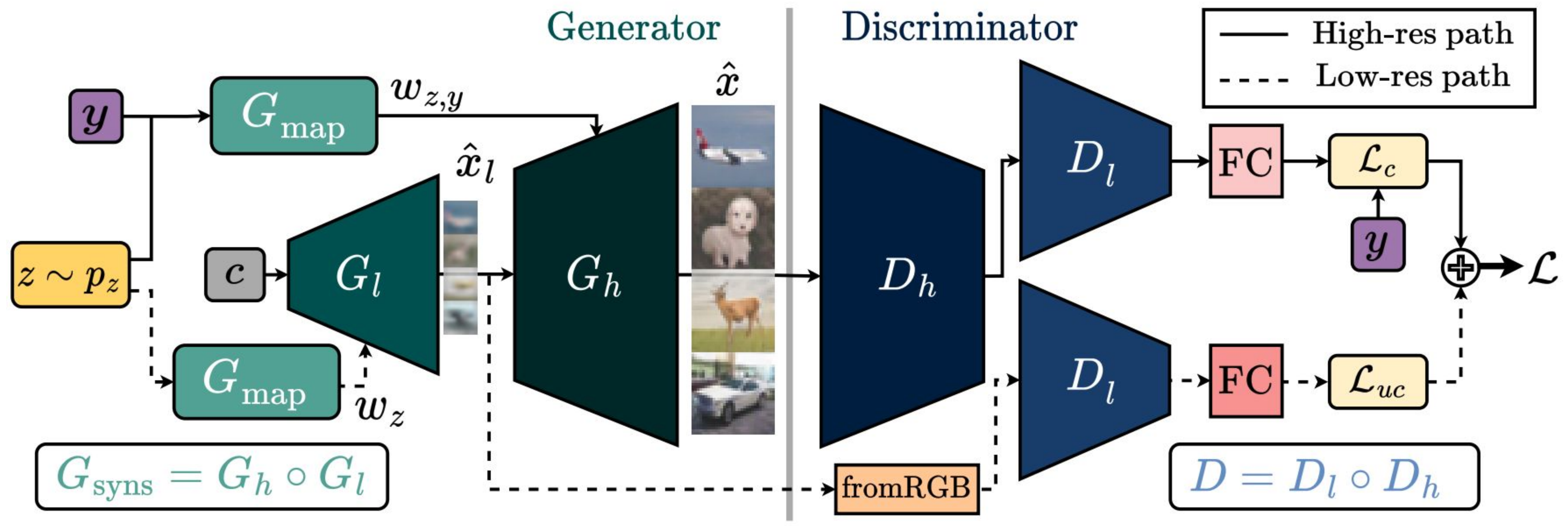
➔ Class-specific features are usually unveiled at higher resolution (e.g., unique texture, fine details)



Lower Resolution

Unconditional Training at Lower Resolutions (UTLO)

- **UTLO's Design** promotes knowledge sharing between head and tail classes via shared unconditional intermediate low resolution $\hat{x}_l$
  - Conditional information are injected at higher resolutions
  - Simultaneously trained with both conditional and unconditional objectives
  - Low resolution images/features primarily inherit rich information from head classes



Adversarial Objective

$$\mathcal{L}_c^D = \mathbb{E}_{\boldsymbol{x},y}[f_D(-D(\boldsymbol{x}|y))] + \mathbb{E}_{\boldsymbol{z},y}[f_D(D(G(\boldsymbol{z},y)))]$$
$$\mathcal{L}_c^G = \mathbb{E}_{\boldsymbol{z},y}[f_G(-D(G(\boldsymbol{z},y)))]$$

UTLO Objective

$$\mathcal{L}^D = \mathcal{L}_c^D + \lambda \cdot \mathcal{L}_{uc}^D$$
$$\mathcal{L}^G = \mathcal{L}_c^G + \lambda \cdot \mathcal{L}_{uc}^G$$

Evaluation Metrics For Long-Tail Datasets

- For **FID/KID**, we sample from the same distribution as the largest available training dataset (e.g. before artificial imbalance)
- We propose **FID/KID-FewShot (FS)**, tailored for evaluating the quality of generated samples in the tail classes
  - We maintain an equal number of real images across all tail classes, emphasizing on the learning quality on tail classes.

Evaluation

- UTLO improves over baselines across all metrics, datasets, and architectures (FastGAN and StyleGAN2-ADA).

res_{uc}	FID ↓	FID-FS ↓	KID ↓ ×1000	KID-FS ↓ ×1000
8 × 8	**26.2**	**48.4**	**12.6**	**19.6**
16 × 16	27.5	50.3	13.7	20.8
32 × 32	38.0	64.9	23.3	34.3

Ablation: choice of uncond. low resolution

*Importance of FS metrics

Methods	FID ↓	FID-FS ↓	KID ↓ ×1000	KID-FS ↓ ×1000
PGAN (FastGAN)+DA [39]	15.0	60.2	4.6	52.7
+ GSR [35]	15.7	63.7	5.7	58.0
+ UTLO (Ours)	**10.9**	**43.6**	**3.5**	**35.3**

LSUN5-LT dataset | FastGAN

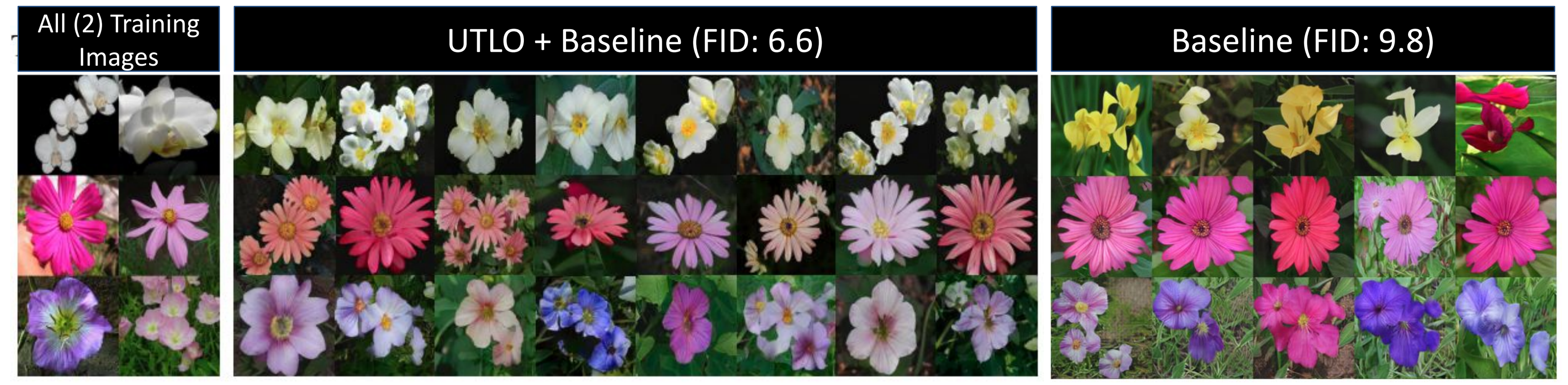
Methods	FID ↓	FID-FS ↓	KID ↓ ×1000	KID-FS ↓ ×1000
StyleGAN2-ADA UnCond. [16]	39.4	104.1	17.3	27.6
StyleGAN2-ADA [16]	51.4	87.1	24.7	35.9
+ Transitional [41]	62.1	99.0	38.5	48.9
+ GSR[35]	39.2	67.2	21.2	32.7
+ NoisyTwins [36]	29.4	50.2	16.7	21.2
+ UTLO (Ours)	**26.2**	**48.4**	**12.6**	**19.6**

AnimalFaces-LT dataset | StyleGAN2-ADA

Visual Analysis

Ablation on the choice of uncond. low resolution

Can any low-res be the starting point for high-res generations?



Conclusion

- *We proposed UTLO for training class-conditional GANs on long-tailed data that addresses mode collapse by promoting knowledge sharing between head and tail classes.*
- *UTLO introduces a new category of class-conditional GANs featuring a partially unconditional generator, trained with both conditional and unconditional objectives.*
- *We introduced FID/KID FewShot metric, enabling a more precise evaluation of the generation quality in the long-tailed setup.*
- *We showed the effectiveness of UTLO through qualitative and quantitative experiments.*

This work is partially supported by National Science Foundation (NSF) under grant 1751412 and 1927564.