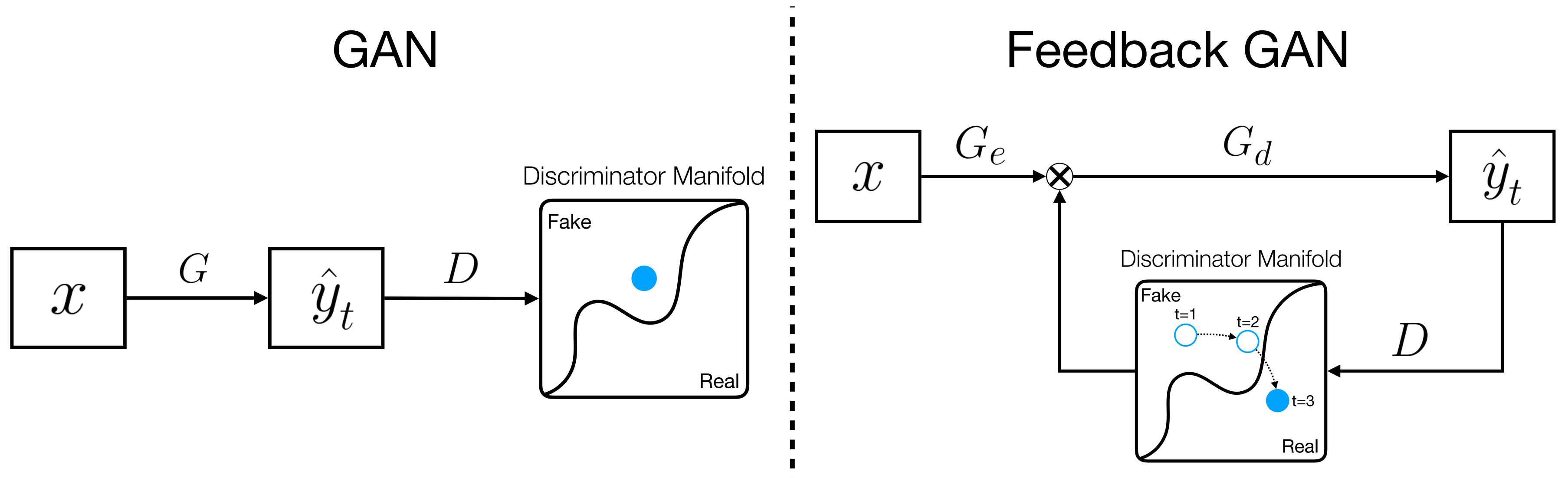
Feedback Adversarial Learning: Spatial Feedback for Improving Generative Adversarial Networks

Minyoung Huh* Shao-Hua Sun* Ning Zhang

Motivation

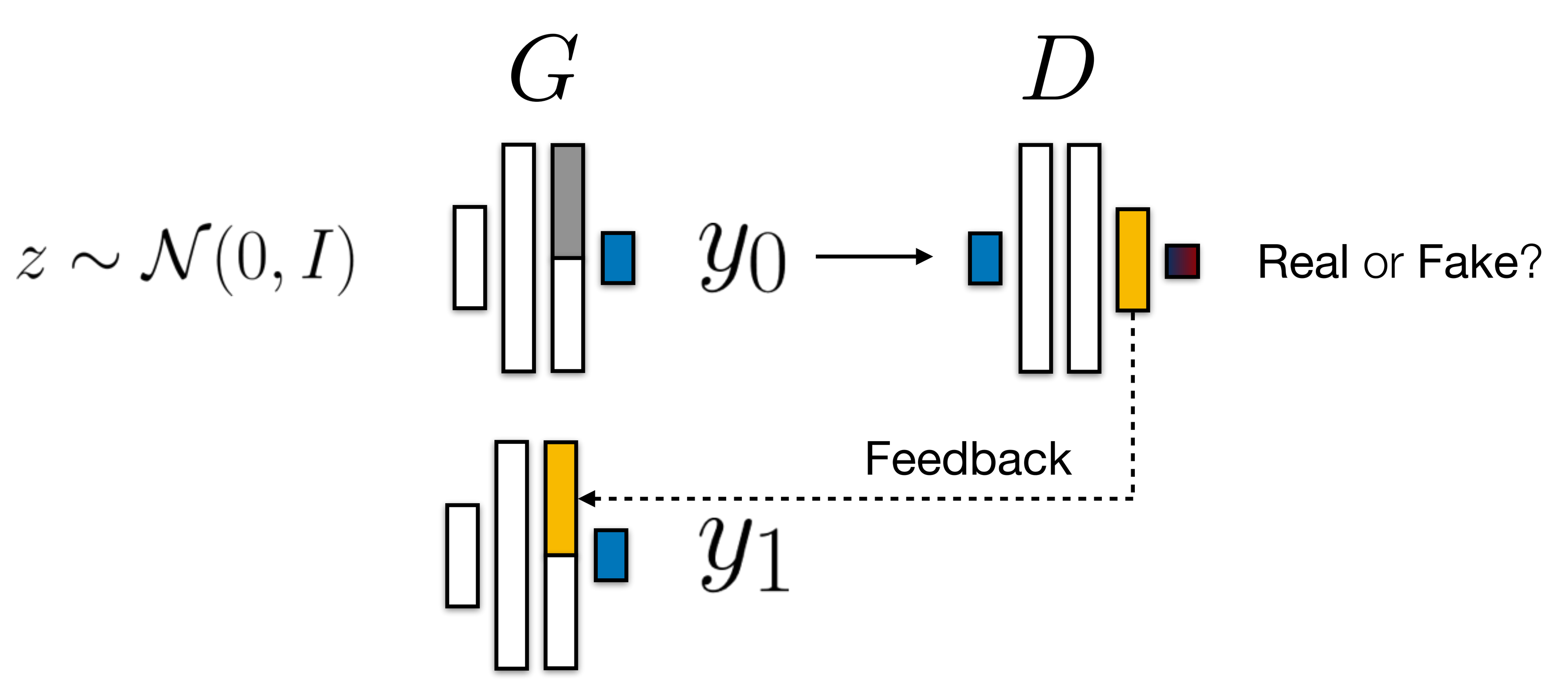
Leverage discriminator's **feedback signals** to improve samples generated by Generative Adversarial Networks (GANs)



Intuition

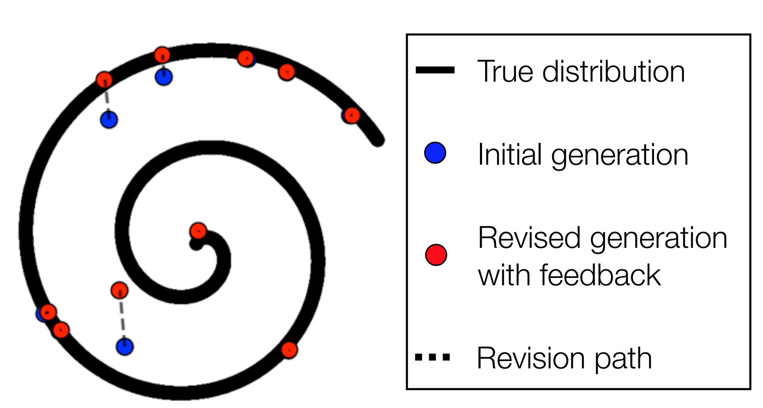
Is **the discriminator's feedback** useful for improving generated samples?

Toy Experiment



Train a GAN to generate points **(x, y)** that are indistinguishable from the samples drawn from the underlying **true** distribution.

The generated samples, the discriminator believes is fake, is **improved** with feedback.



High-dimensional Data

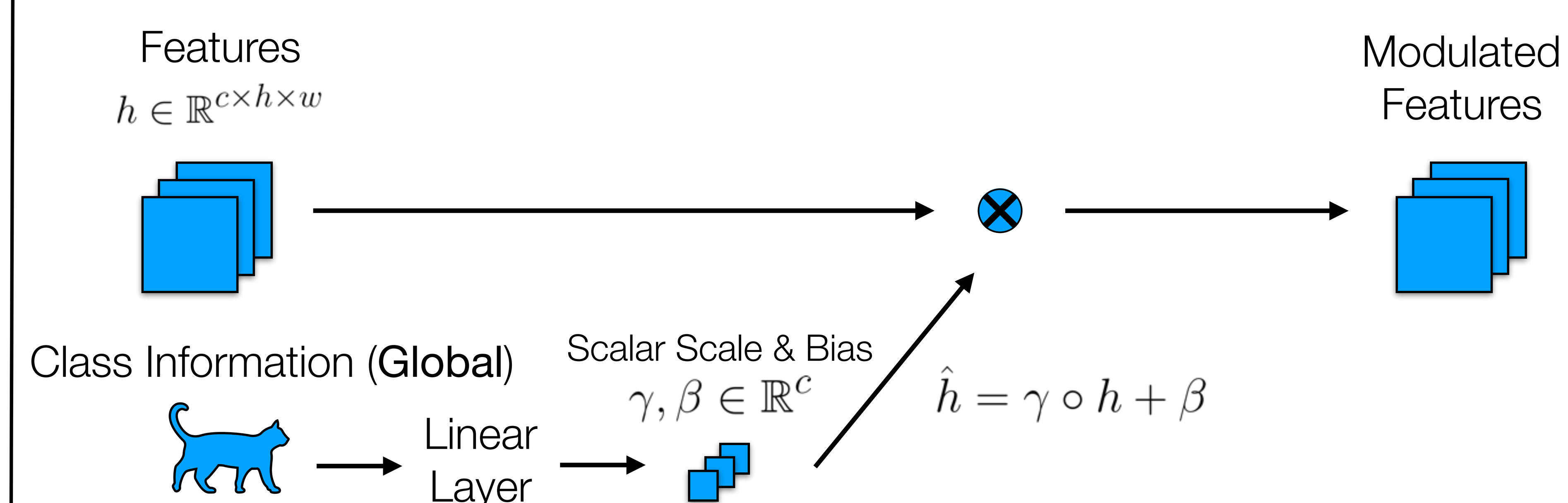
How can we effectively provide **feedback** signals to **high-dimensional data** such as images and voxels?

Adaptive Spatial Transform

Goal: allow the generator to attend and fix local regions based on the discriminator's feedback and its previous generation.

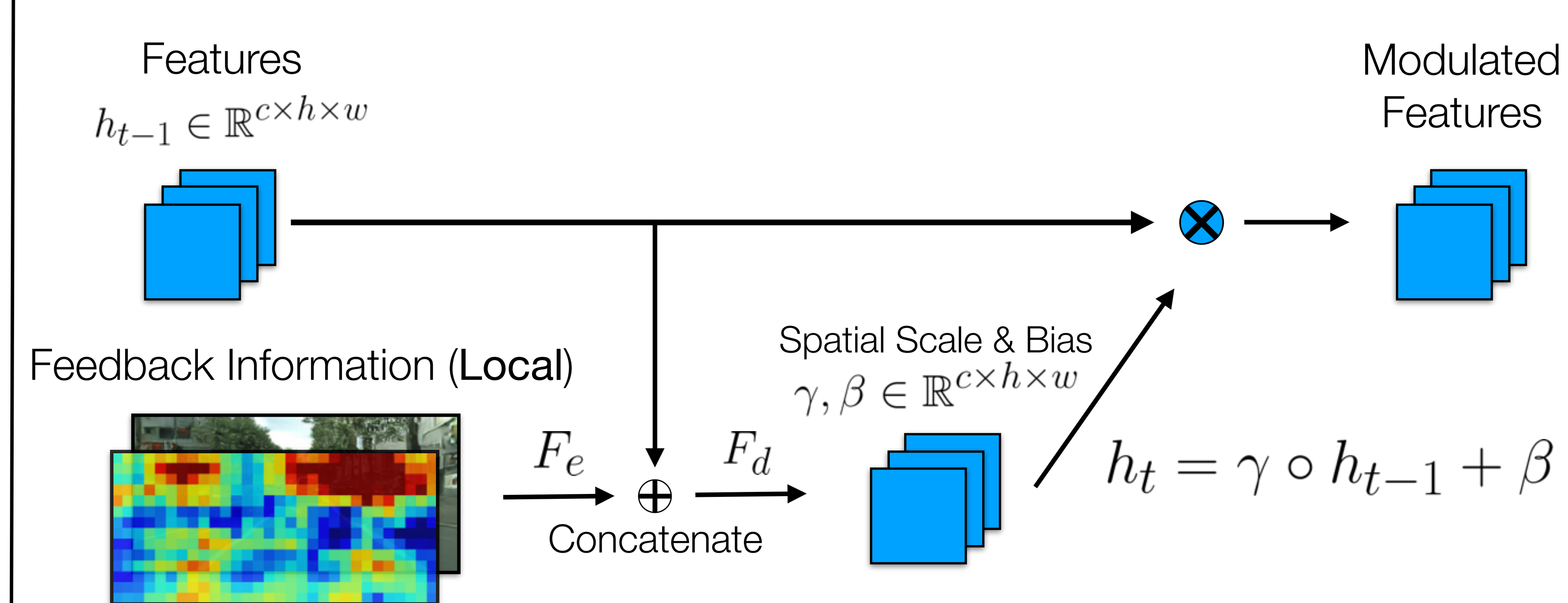
Conditional Normalization

Learn linear layers that predict **global** scalar affine parameters to modulate feature maps using external information such as class information.
(*e.g.* Conditional batch-normalization [1], Adaptive Instance-Norm [2][3])



Adaptive Spatial Transform

Transform feature maps **locally** by predicting affine parameters.



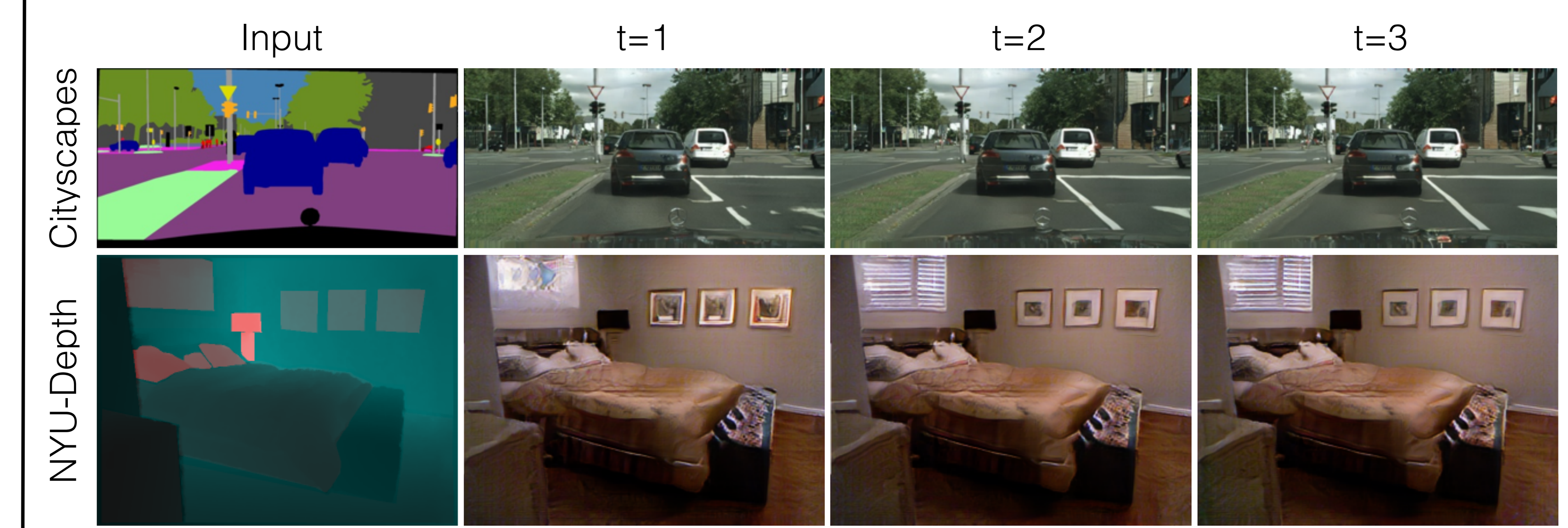
A concurrent work (GauGAN [4]) translates a semantic layout to an image using a similar module: SPatially-Adaptive DEnormalization (SPADE).

Reference

[1] Vries et al., Modulating early visual processing by language, NIPS 2017
[2] Dumoulin et al., A Learned Representation For Artistic Style, ICLR2016
[3] Huang et al., Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization, ICCV2017
[4] Park et al., Semantic Image Synthesis with Spatially-Adaptive Normalization, CVPR2019
[5] Guo et al., Long Text Generation via Adversarial Training with Leaked Information, AAAI 2018

Experiment

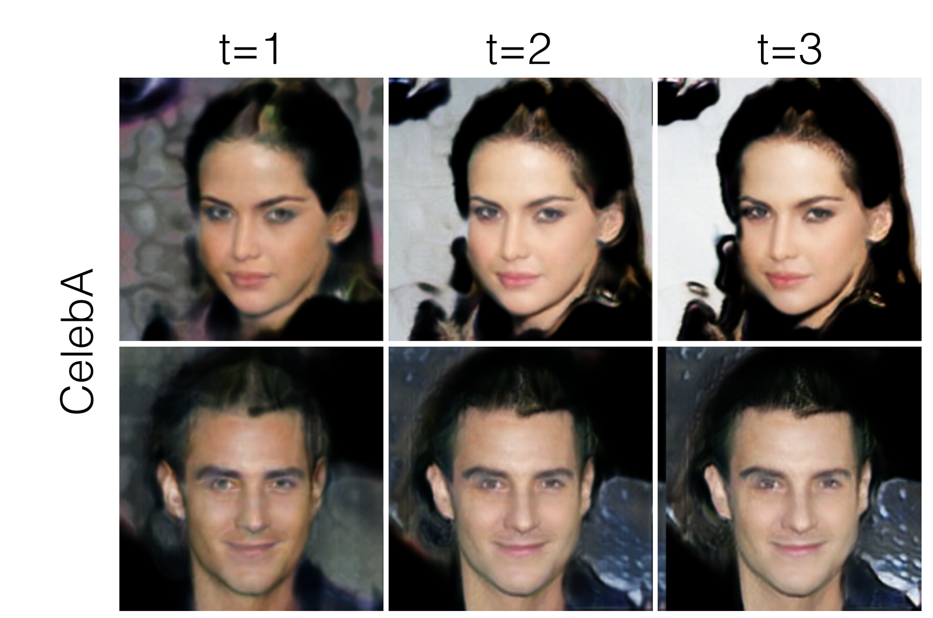
Image-to-image Translation



Cityscapes Model	Val Cat IOU ↑	Val Cls IOU ↑	Val LPIPS ↓	Train LPIPS ↓
Ground Truth	76.2	0.21	0.0	0.0
Pix2Pix	0.380	0.655	**0.428**	0.320
Pix2Pix + Feedback (t=1)	0.383	0.646	0.431	0.265
Pix2Pix + Feedback (t=2)	0.417	0.687	**0.428**	**0.254**
Pix2Pix + Feedback (t=3)	**0.418**	**0.692**	0.429	**0.254**

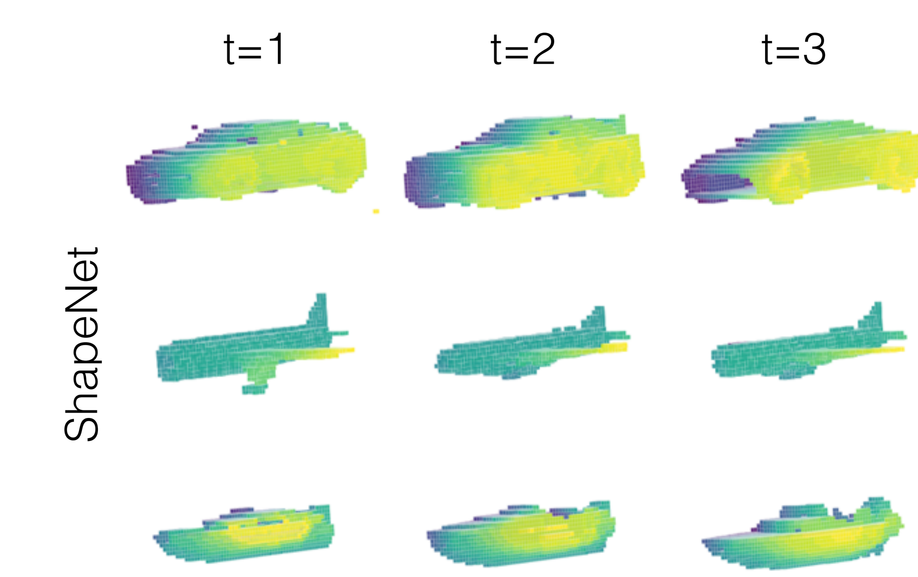
NYU-Depth Model	Val REL↓	Val δ_1 ↑	Val δ_2 ↑	Val LPIPS ↓	Train LPIPS ↓
Ground Truth	0.191	0.846	0.974	0.0	0.0
Pix2Pix	0.191	0.892	0.961	0.483	0.337
Pix2Pix + Feedback ($t = 1$)	0.179	0.702	0.904	0.473	0.281
Pix2Pix + Feedback ($t = 2$)	**0.178**	**0.706**	0.906	**0.469**	**0.275**
Pix2Pix + Feedback ($t = 3$)	0.181	0.701	**0.908**	0.473	0.284

Image Generation



Model	CelebA-FID ↓
GAN	22.56
GAN w/ Feedback (t=1)	26.49
GAN w/ Feedback (t=2)	20.65
GAN w/ Feedback (t=3)	**18.52**

Voxel Generation



Model	Classification accuracy ↑ Airplane	Car	Vessel
Ground Truth	95.9%	99.6%	98.8%
VoxelGAN	93.0%	98.1%	89.2%
VoxelGAN + Feedback ($t = 1$)	93.0%	98.2%	91.0%
VoxelGAN + Feedback ($t = 2$)	94.0%	98.9%	96.2%
VoxelGAN + Feedback ($t = 3$)	**95.6%**	**99.1%**	**97.1%**

Improvements with Feedback