AUDIO LABS

Comparing Normalizing Flow and Generative Adversarial Networks for Super-Resolution

Konstantin Schmidt
International Audio Laboratories Erlangen
konstantin.schmidt@audiolabs-erlangen.de

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

Fraunhofer
IIS

Comparing NF and GANs
Introduction

Original:

Low resolution:

Downsamplingfactor 4

- Deep generative models have been successfully applied to tasks like image super-resolution
- Missing is a comparison of GANs and Normalizing Flow
  - network structures as comparable as possible
  - similar computational complexity
  - same receptive field.

Deep Generative Models
Unsupervised Learning

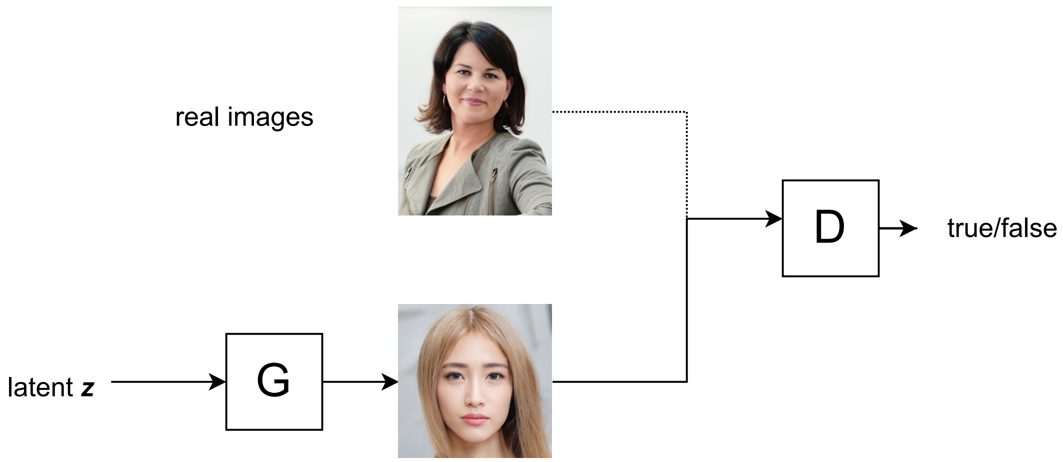
- Variational autoencoders
- Autoregressive models
- Generative adversarial networks
- Normalizing flows
- Diffusion models (!)
- ..hybrid approaches

Deep Generative Models
Unsupervised Learning

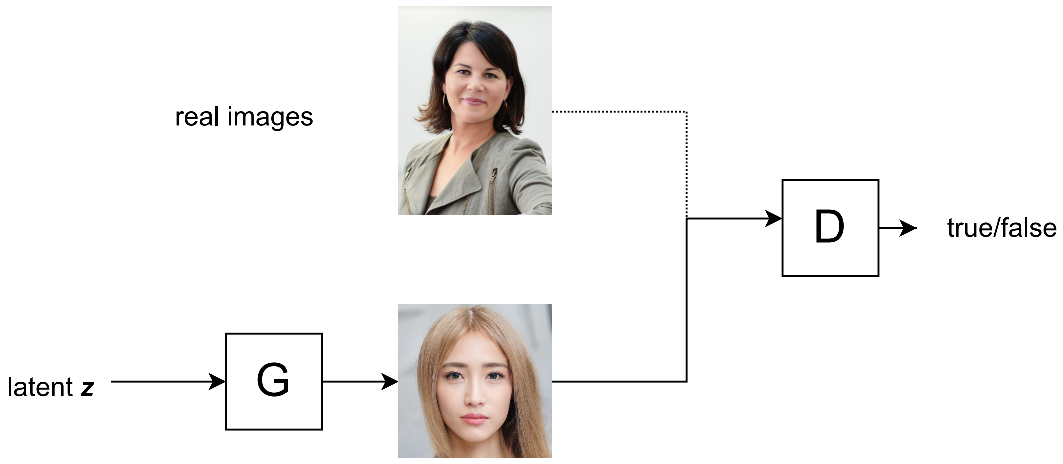
- Variational autoencoders
- Autoregressive models
- Generative adversarial networks
- Normalizing flows
- Diffusion models (!)
- ..hybrid approaches

Deep Generative Models
Unsupervised Learning

- Variational autoencoders
- Autoregressive models
- Generative adversarial networks
- Normalizing flows
- Diffusion models (!)
- ..hybrid approaches

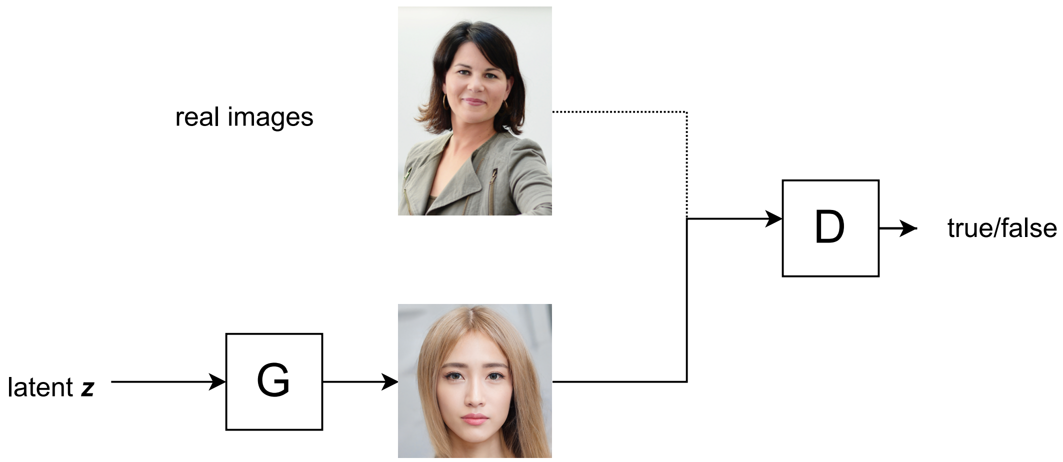


Deep Generative Models
Unsupervised Learning

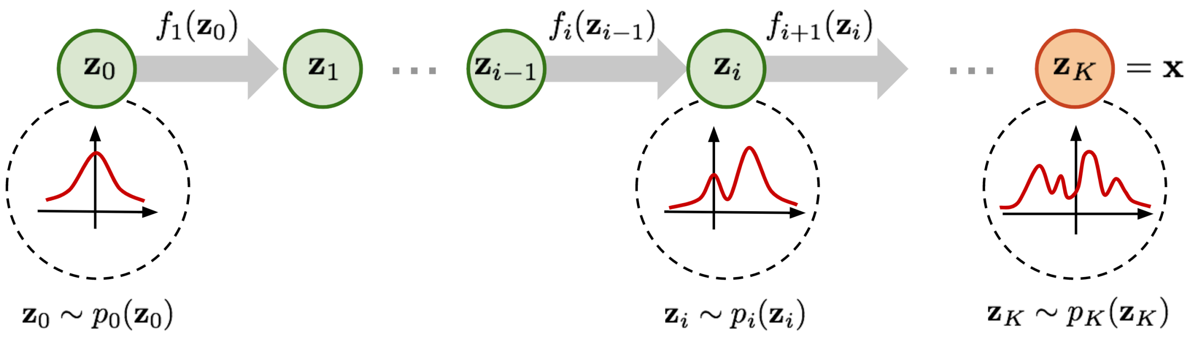


- Variational autoencoders
- Autoregressive models
- Generative adversarial networks
- Normalizing flows
- Diffusion models (!)
- ..hybrid approaches

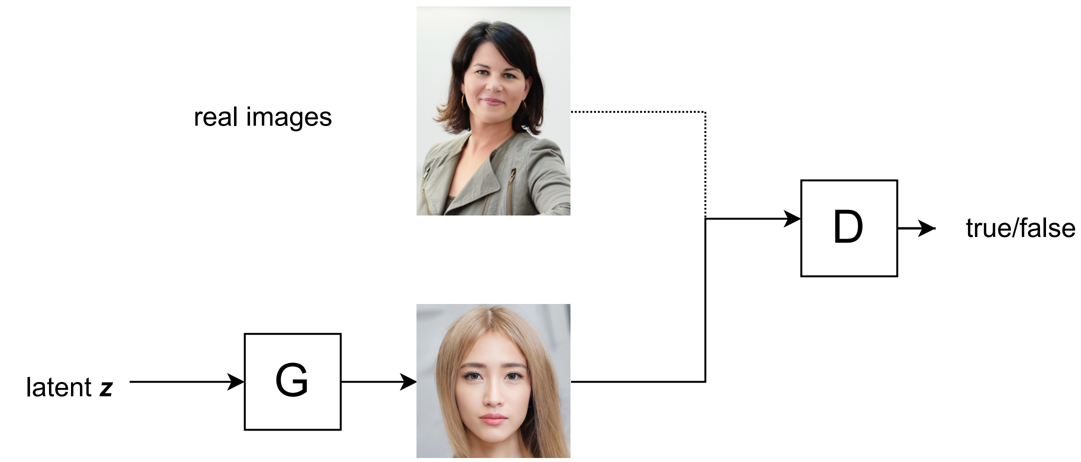
Deep Generative Models
Unsupervised Learning



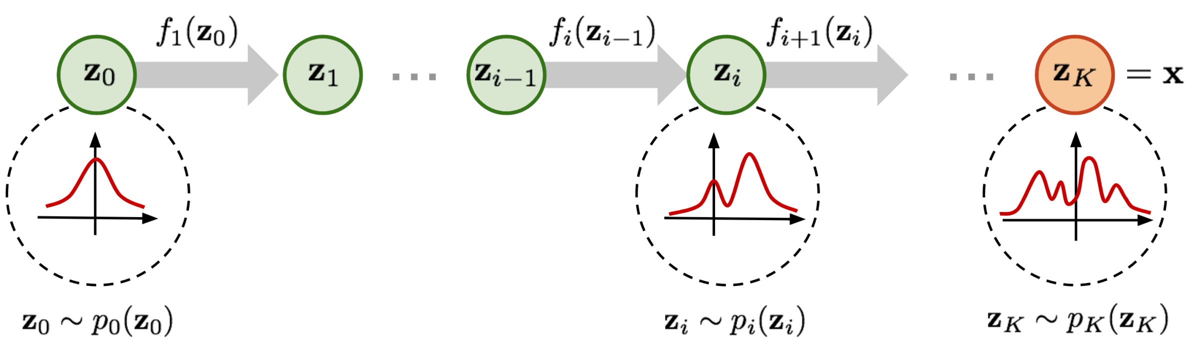
- Variational autoencoders
- Autoregressive models
- Generative adversarial networks
- Normalizing flows
- Diffusion models (!)
- ..hybrid approaches



Deep Generative Models
Unsupervised Learning



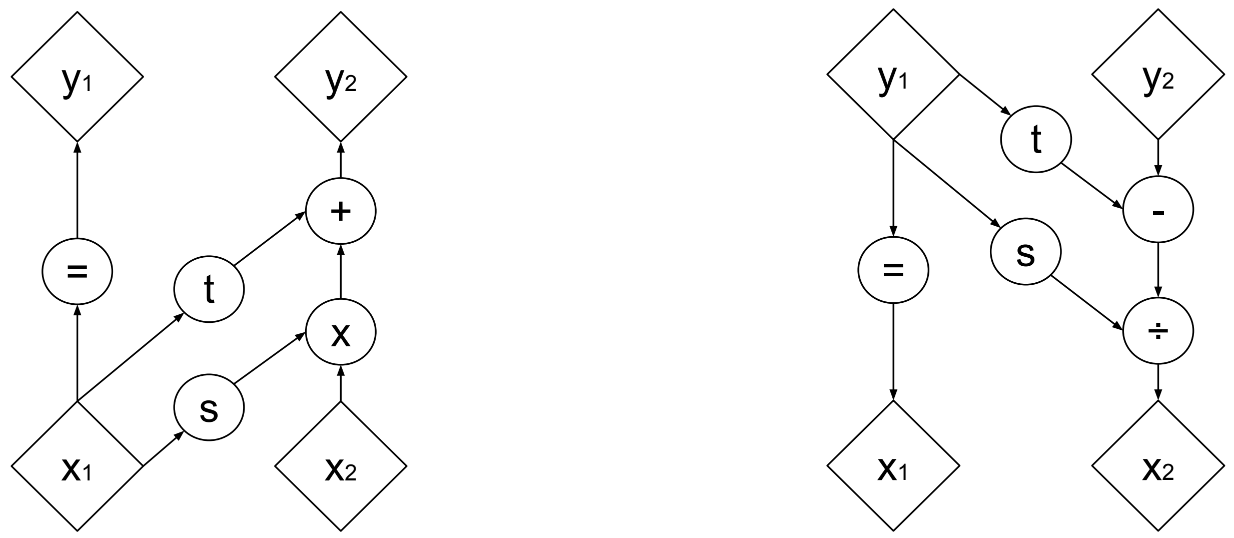
- Variational autoencoders
- Autoregressive models
- Generative adversarial networks
- Normalizing flows
- Diffusion models (!)
- ..hybrid approaches



Additional conditional information can be used to steer the network

Normalizing Flow

Affine Coupling Layer



(a) Forward propagation (b) Inverse propagation

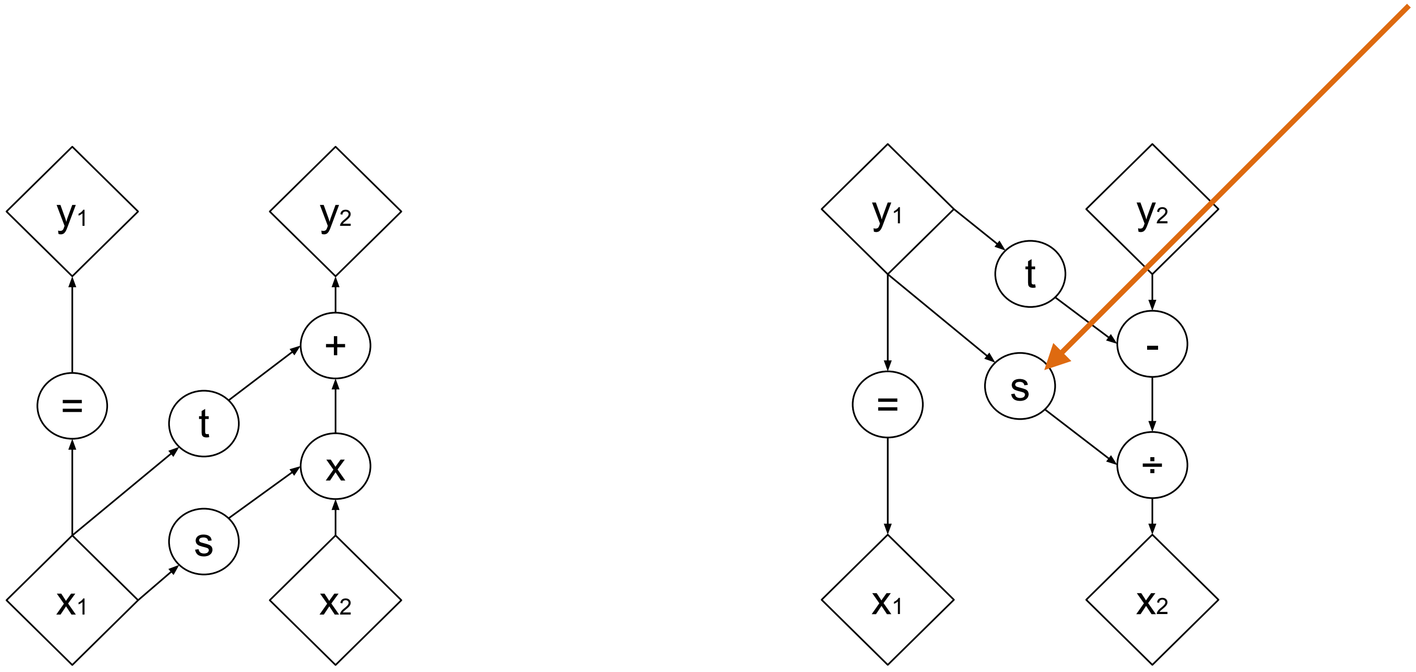
$$y_{1:d} = x_{1:d}$$
$$y_{d+1:D} = x_{d+1:D} \odot \exp\big(s(x_{1:d})\big) + t(x_{1:d})$$

Jacobian of affine coupling layer:

$$\frac{\partial y}{\partial x^T} = \begin{bmatrix} \mathbb{I}_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial x^T_{1:d}} & \operatorname{diag}\big(\exp\left[s\left(x_{1:d}\right)\right]\big) \end{bmatrix}$$

Normalizing Flow

Affine Coupling Layer



(a) Forward propagation (b) Inverse propagation

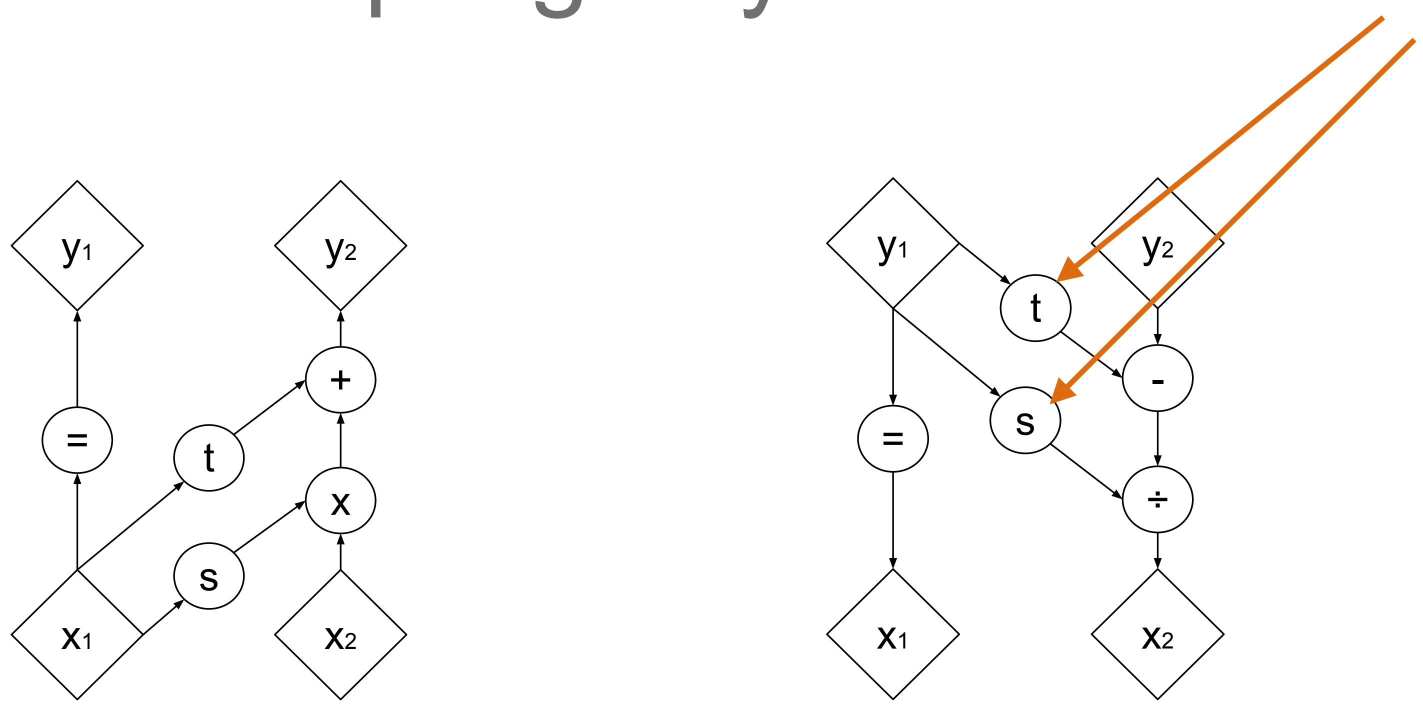
Jacobian of affine coupling layer:

$$y_{1:d} = x_{1:d}$$
$$y_{d+1:D} = x_{d+1:D} \odot \exp\left(s(x_{1:d})\right) + t(x_{1:d})$$

$$\frac{\partial y}{\partial x^T} = \begin{bmatrix} \mathbb{I}_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial x^T_{1:d}} & \operatorname{diag}\left(\exp\left[s\left(x_{1:d}\right)\right]\right) \end{bmatrix}$$

Normalizing Flow

Affine Coupling Layer



(a) Forward propagation (b) Inverse propagation

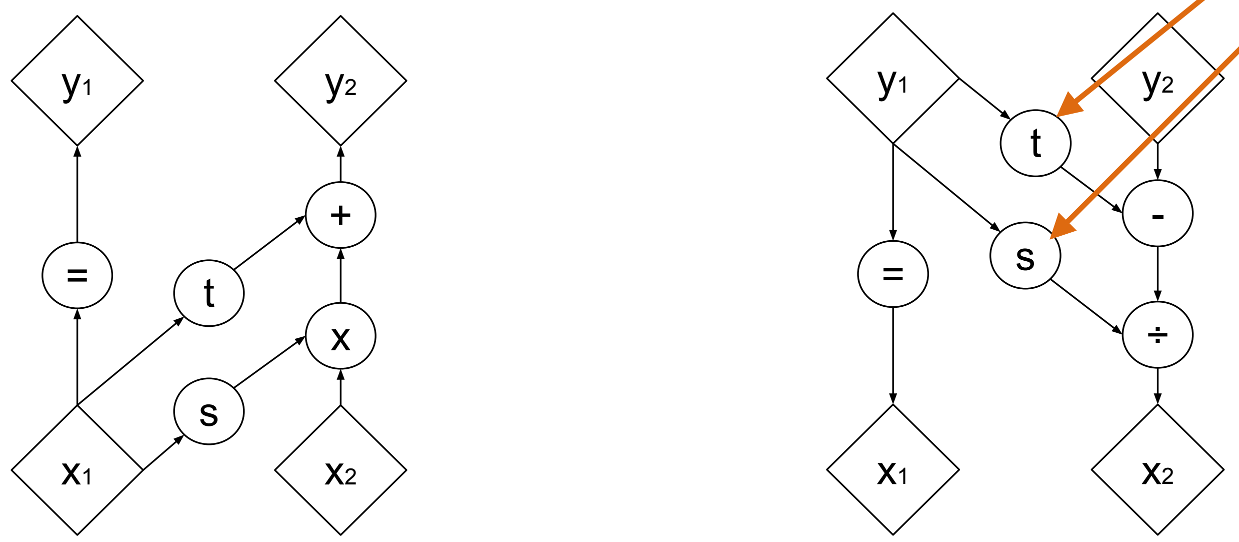
$$
\begin{aligned}
y_{1:d} &= x_{1:d} \\
y_{d+1:D} &= x_{d+1:D} \odot \exp\big(s(x_{1:d})\big) + t(x_{1:d})
\end{aligned}
$$

Jacobian of affine coupling layer:

$$
\frac{\partial y}{\partial x^T} = \begin{bmatrix} \mathbb{I}_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial x^T_{1:d}} & \operatorname{diag}\big(\exp\left[s\left(x_{1:d}\right)\right]\big) \end{bmatrix}
$$

Normalizing Flow
Affine Coupling Layer

(deep) neural networks
Not invertible!



(a) Forward propagation (b) Inverse propagation

$$y_{1:d} = x_{1:d}$$
$$y_{d+1:D} = x_{d+1:D} \odot \exp\left(s(x_{1:d})\right) + t(x_{1:d})$$

Jacobian of affine coupling layer:

$$\frac{\partial y}{\partial x^T} = \begin{bmatrix} \mathbb{I}_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial x^T_{1:d}} & \operatorname{diag}\left(\exp\left[s\left(x_{1:d}\right)\right]\right) \end{bmatrix}$$

Comparing NF and GANs

For Super-resolution

- Generative adversarial networks allow for
  - flexible design of the networks 😃
  - are hard to optimize and need intensive parameter tunings 😠
- Flow-based models are easy to train 😊
  - Networks need to be invertible with a latent domain of the same dimension as the output image 😒

Comparing NF and GANs
For Super-resolution

- SRGAN (arXiv:1609.04802)
  - Rezeptive field:192 pixel
  - 0.15 MOPS per pixel

	Channel dim	Kernel size
Conv	3x64	9x9
17*Conv	64x64	3x3
2*Upsampling Conv	64x256?	3x3
Conv	64x3	9x9

- Conditional Normalizing Flow
  - Rezeptive field:192 pixel
  - 0.16 MOPS per pixel

	Channel dim	Kernel size
16 Affine Coupling	6x64->64x12	3x3
3 Conv layer per A.Coupling	64x64	3x3

Comparing NF and GANs

Other DNN and training parameters

- PreLu activations
- Trained on Flickr HQ dataset
- BatchNorm for all models (no ActNorm for Flow)
- Evaluated on CelebA
- All networks trained with batch-size of 64
- Random crops of 256 x 256 pixels were used for training

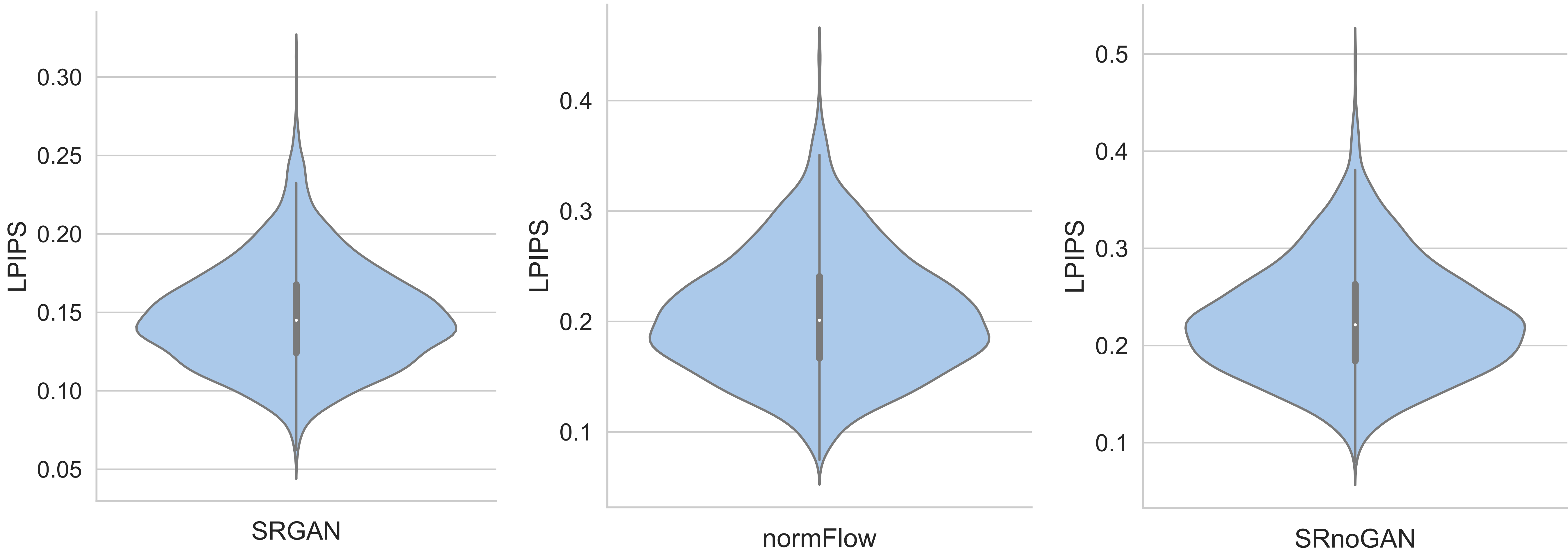
Comparing NF and GANs
Evaluation

	SRGAN	Norm. Flow	SR (no GAN)
FID smaller better	3.858	6.026	5.822
LPIPS smaller better	0.1480	0.2061	0.22674

Frechet Inception Distance (FID): metric based on the *distribution* of activations of *deeper* layers of classification network

Learned Perceptual Image Patch Similarity (LPIPS): metric based on the activations of first layers of classification network

Comparing NF and GANs
Evaluation



Comparing NF and GANs

Evaluation Flow

Comparing NF and GANs
Evaluation SRGAN

Comparing NF and GANs

Evaluation SRGAN (left) - FLOW (right)

Comparing NF and GANs
Evaluation FLOW - SRGAN

Original: 

GAN: 

Bicubic: 

Flow: