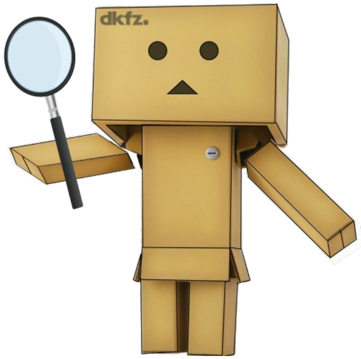
Generative Modeling by Estimating Gradients of the Data Distribution

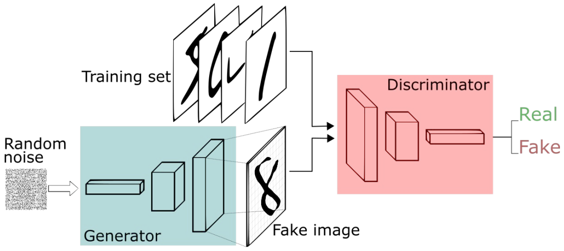
Yang Song, Stefano Ermon

David Zimmerer
Medical Image Analysis (#MIA-san-mia)
DKFZ



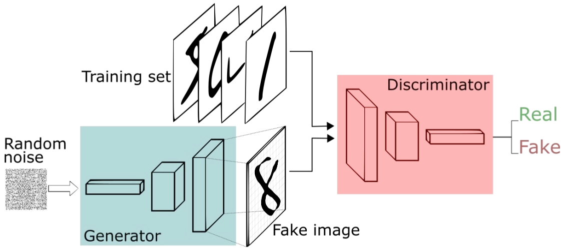
Generative modeling

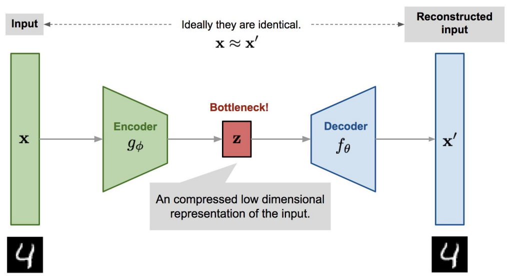
Generative modeling



GANs

Generative modeling

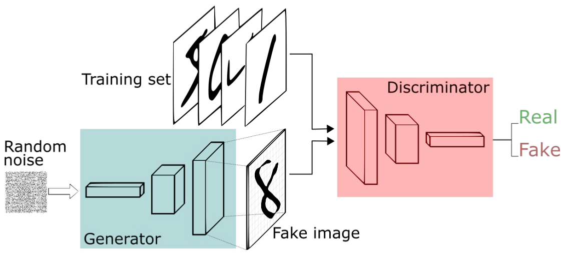




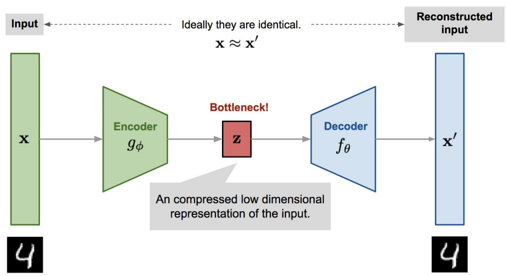
GANs

VAEs

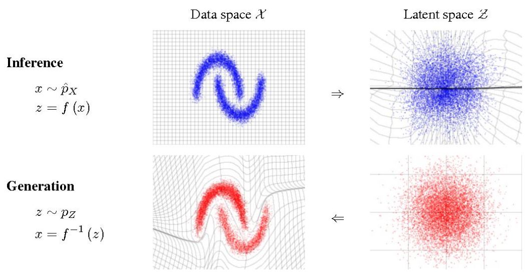
Generative modeling



GANs

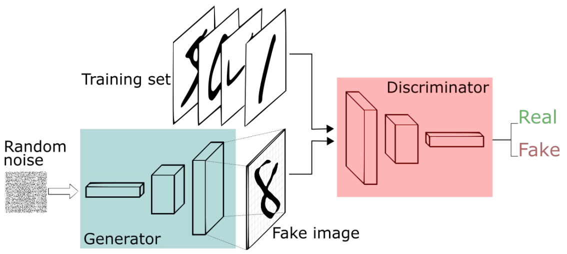


VAEs

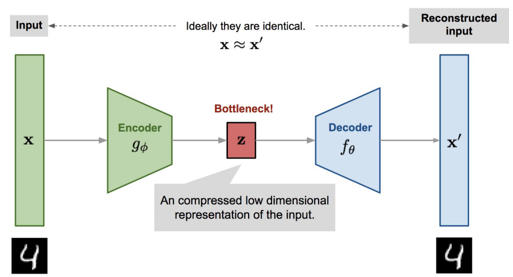


Flows

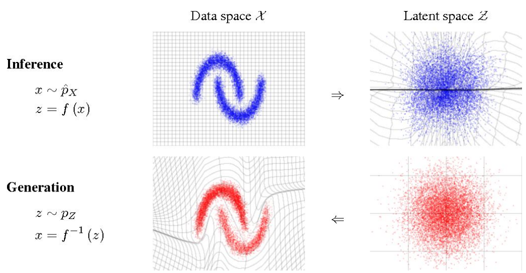
Generative modeling



GANs



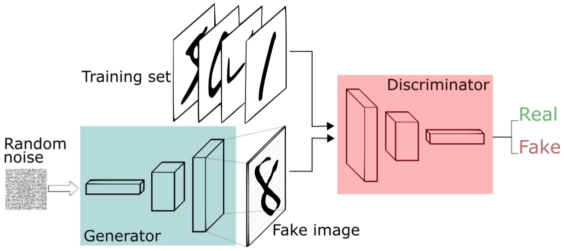
VAEs

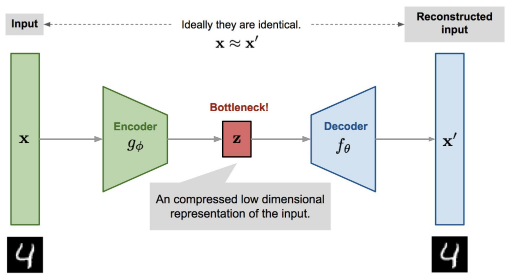


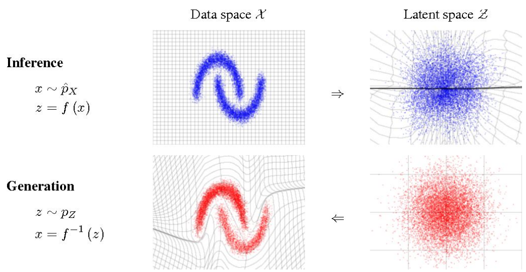
Flows

...

Generative modeling







GANs

VAEs

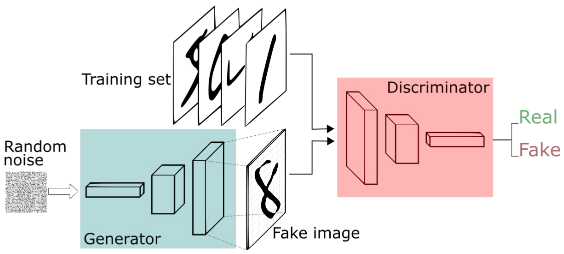
Flows

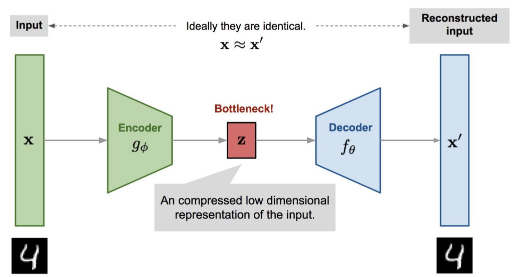
...

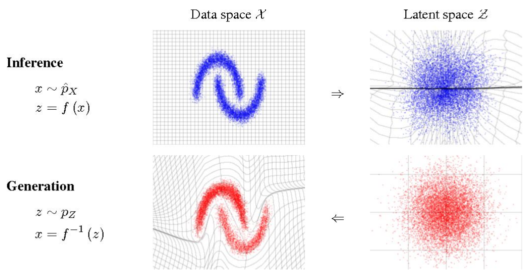
implicit

explicit

Generative modeling







GANs

VAEs

Flows

...

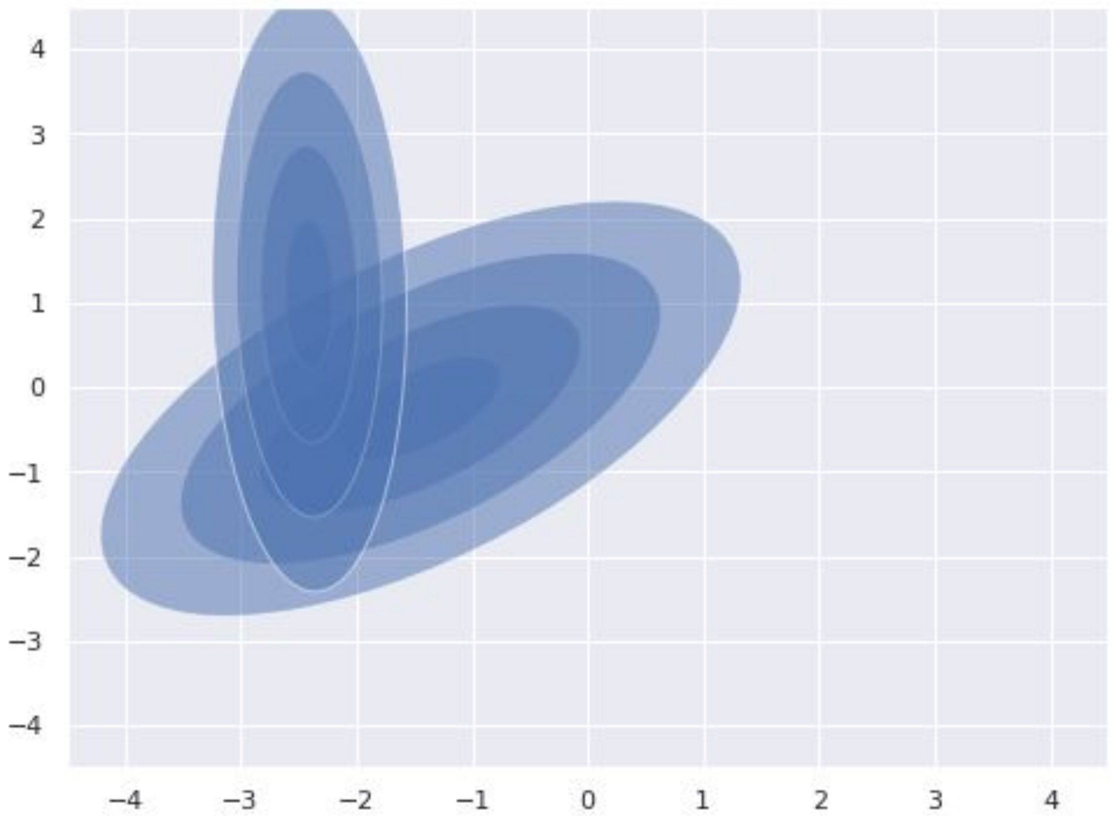
implicitly

explicitly

“learn” p(x)

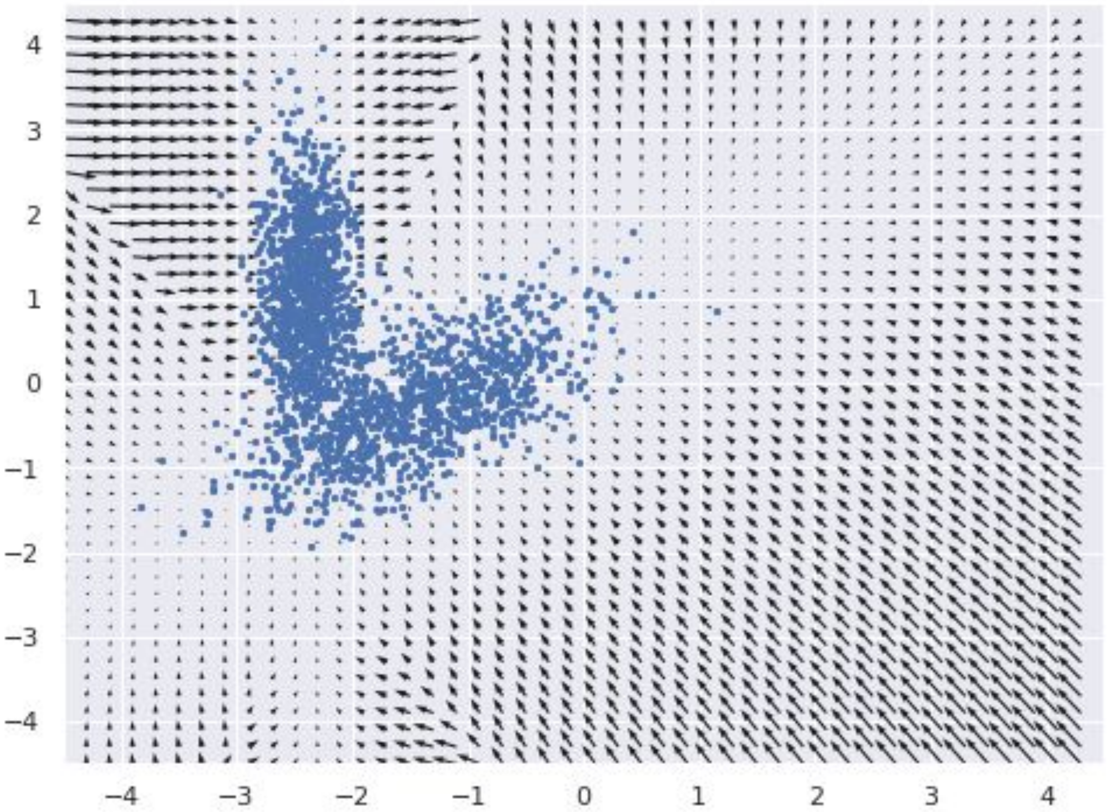
“New” Idea:
Generative Modeling by Estimating Gradients of the Data Distribution

“New” Idea:
Generative Modeling by Estimating Gradients of the Data Distribution



Instead of learning the data distribution directly….

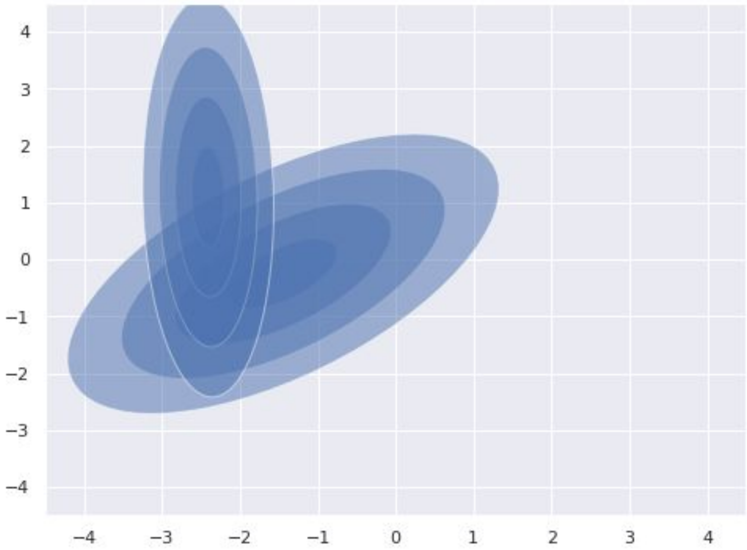
“New” Idea:
Generative Modeling by Estimating Gradients of the Data Distribution

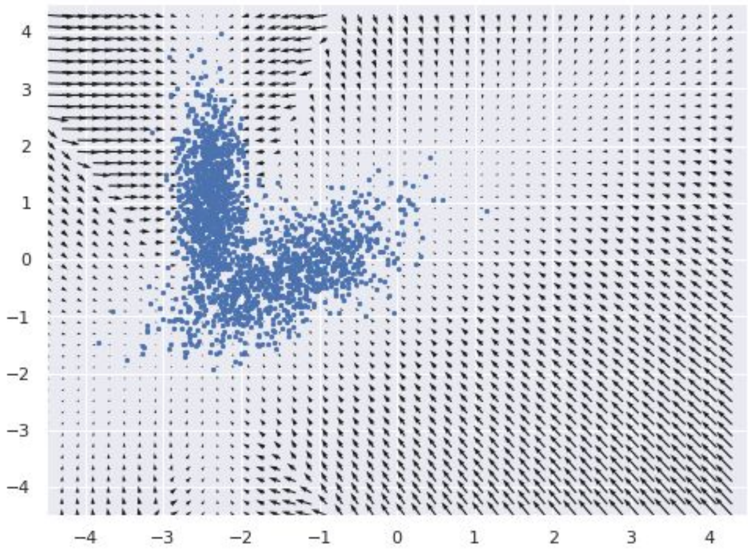


...we learn the gradients of the data distribution

dkfz.

“New” Idea:
Generative Modeling by Estimating Gradients of the Data Distribution





Score matching

[1] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research, 6(Apr):695–709, 2005.

Score matching

$\rightarrow \nabla_{\mathbf{x}} \log p(\mathbf{x})$ i.e. the Gradient of the Data Distribution a.k.a *score*

[1] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research, 6(Apr):695–709, 2005.

Score matching

$\rightarrow \nabla_{\mathbf{x}} \log p(\mathbf{x})$ i.e. the Gradient of the Data Distribution a.k.a *score*

Score matching[1]:

$$\frac{1}{2}\mathbb{E}_{p_{\text{data}}}[\|\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}) - \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})\|_2^2]$$

[1] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research, 6(Apr):695–709, 2005.

Score matching

$\rightarrow \nabla_{\mathbf{x}} \log p(\mathbf{x})$ i.e. the Gradient of the Data Distribution a.k.a *score*

Score matching[1]:

$$\frac{1}{2}\mathbb{E}_{p_{\text{data}}}[||\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}) - \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})||_2^2]$$

equivalent up to a constant to:

$$\mathbb{E}_{p_{\text{data}}(\mathbf{x})}\left[\operatorname{tr}(\nabla_{\mathbf{x}}\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})) + \frac{1}{2}\left\|\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})\right\|_2^2\right]$$

[1] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research, 6(Apr):695–709, 2005.

Score matching

$\rightarrow \nabla_{\mathbf{x}} \log p(\mathbf{x})$ i.e. the Gradient of the Data Distribution a.k.a *score*

Score matching[1]:

$$\frac{1}{2}\mathbb{E}_{p_{\text{data}}}[||\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}) - \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})||_2^2]$$

equivalent up to a constant to:

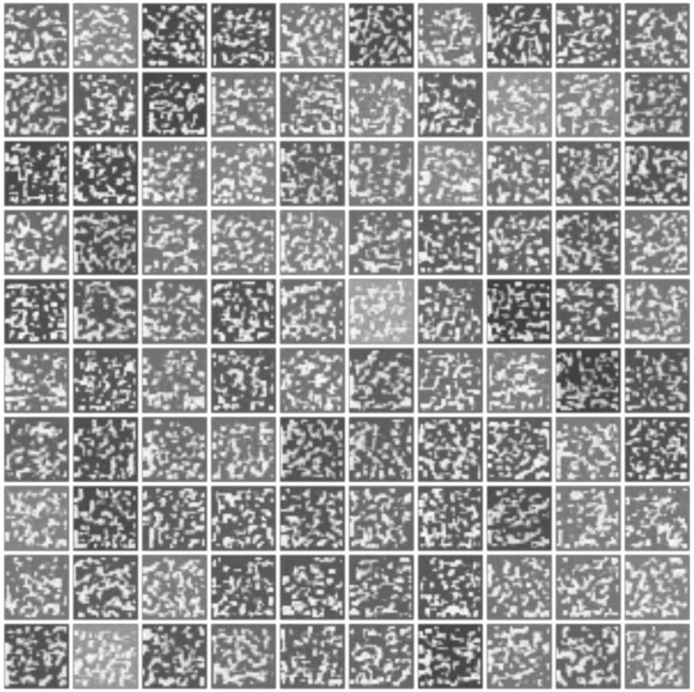
$$\mathbb{E}_{p_{\text{data}}(\mathbf{x})}\left[\operatorname{tr}(\nabla_{\mathbf{x}}\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})) + \frac{1}{2}\,\|\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})\|_2^2\right]$$

$\rightarrow$ So what's new ? 🤔

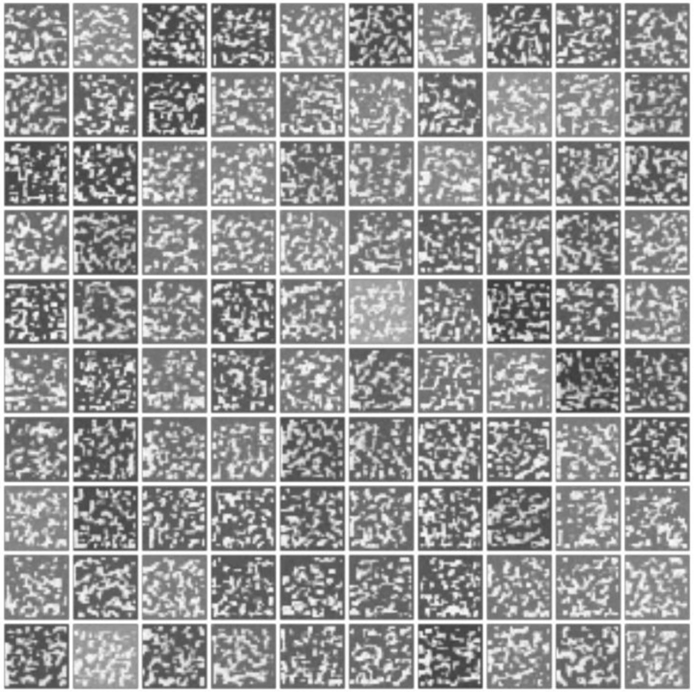
[1] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research, 6(Apr):695–709, 2005.

Trivial Implementation (on MNIST)

Trivial Implementation (on MNIST)

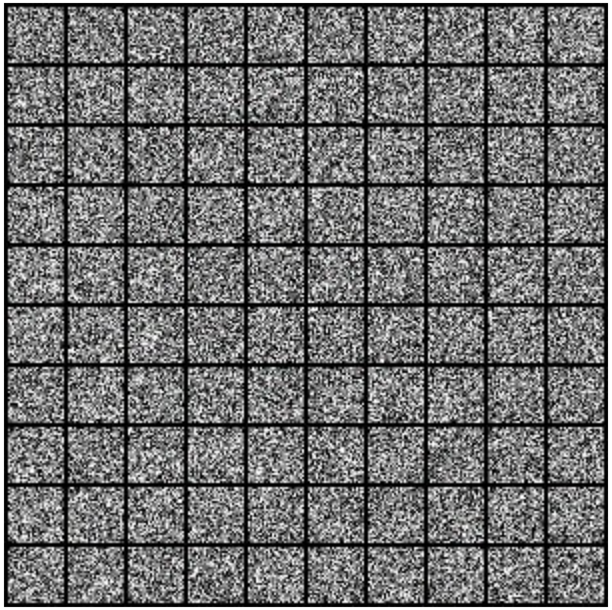


Trivial Implementation (on MNIST)

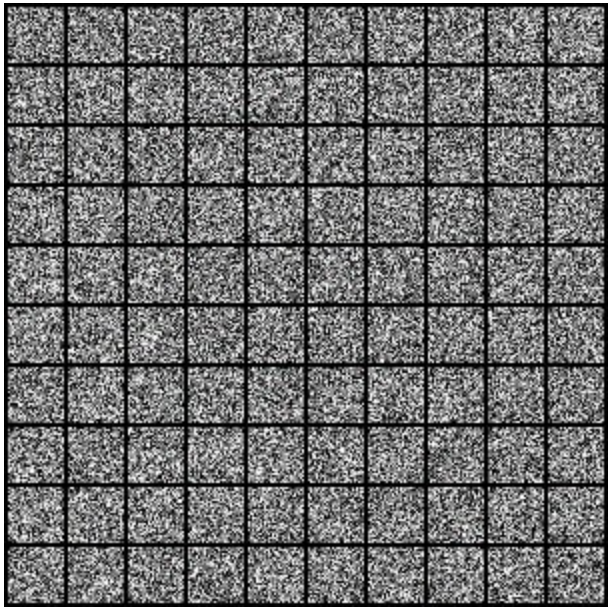


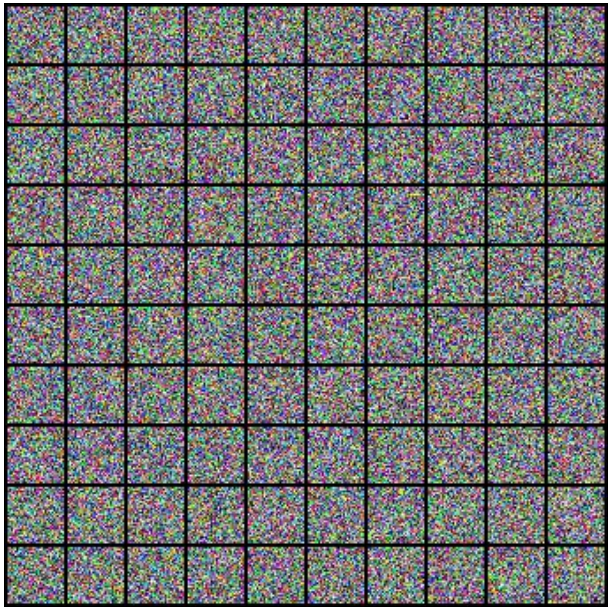
Spoiler: Improved Results

Spoiler: Improved Results

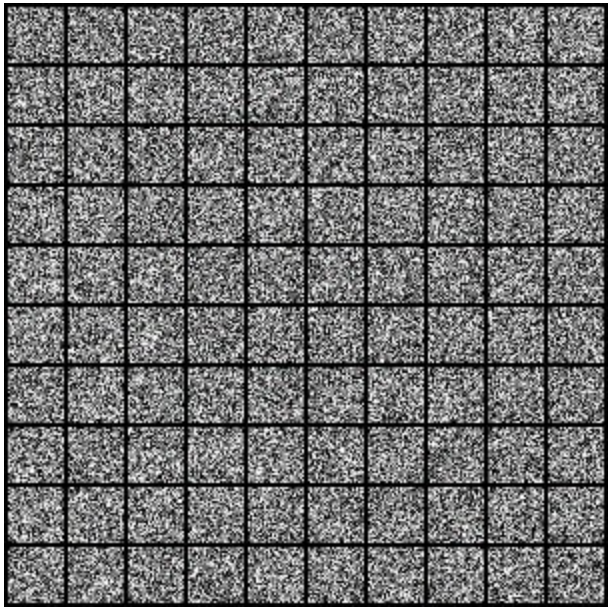
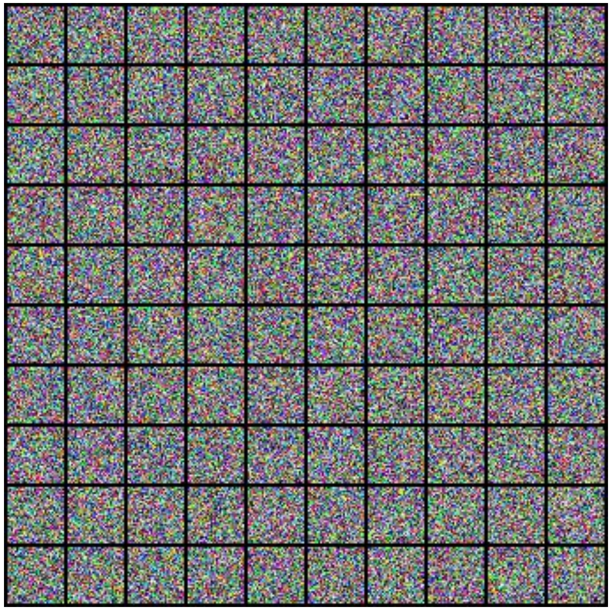
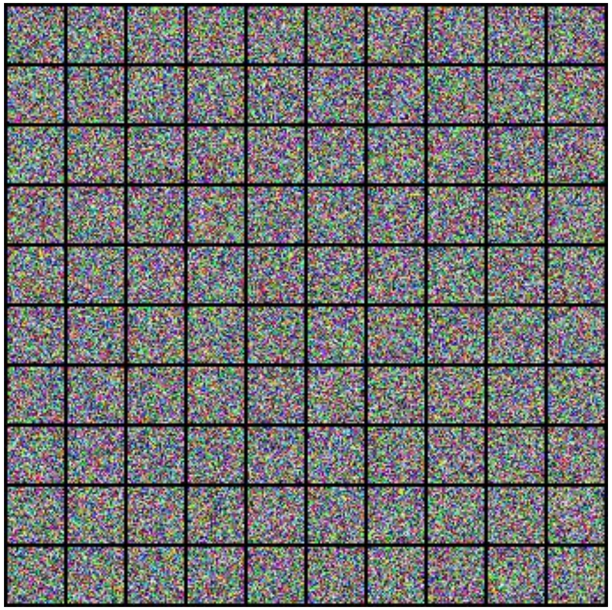


Spoiler: Improved Results





Spoiler: Improved Results


What are the problems ?

What are the problems ?

1. No Support (everywhere)

What are the problems ?

1. No Support (everywhere)
   i.e. p(x) not defined and thus gradient not defined

What are the problems ?

1. No Support (everywhere)
   i.e. $p(x)$ not defined and thus gradient not defined

   → Solution: Add Gaussian Noise

What are the problems ?

1. No Support (everywhere)
   i.e. p(x) not defined and thus gradient not defined

   → Solution: Add Gaussian Noise

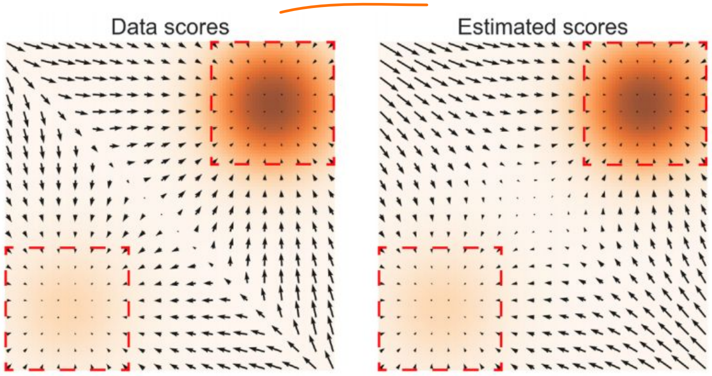
2. Inaccurate score estimation in low density regions

What are the problems ?

1. No Support (everywhere)
   i.e. $p(x)$ not defined and thus gradient not defined

   → Solution: Add Gaussian Noise

2. Inaccurate score estimation in low density regions

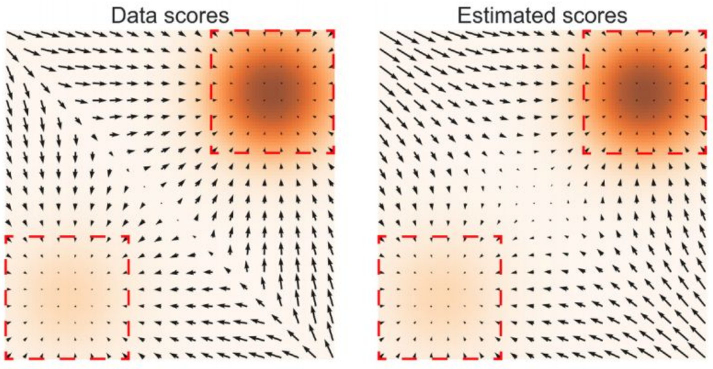


What are the problems ?

1. No Support (everywhere)
   i.e. p(x) not defined and thus gradient not defined

   → Solution: Add Gaussian Noise

2. Inaccurate score estimation in low density regions

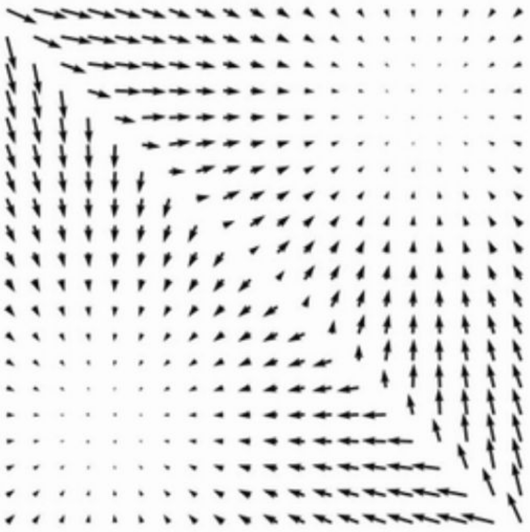


→ Solution add noise at different magnitudes
(large noise: filling low density regions, small noise: fine-adjustments in high density regions)

dkfz.

How to sample:

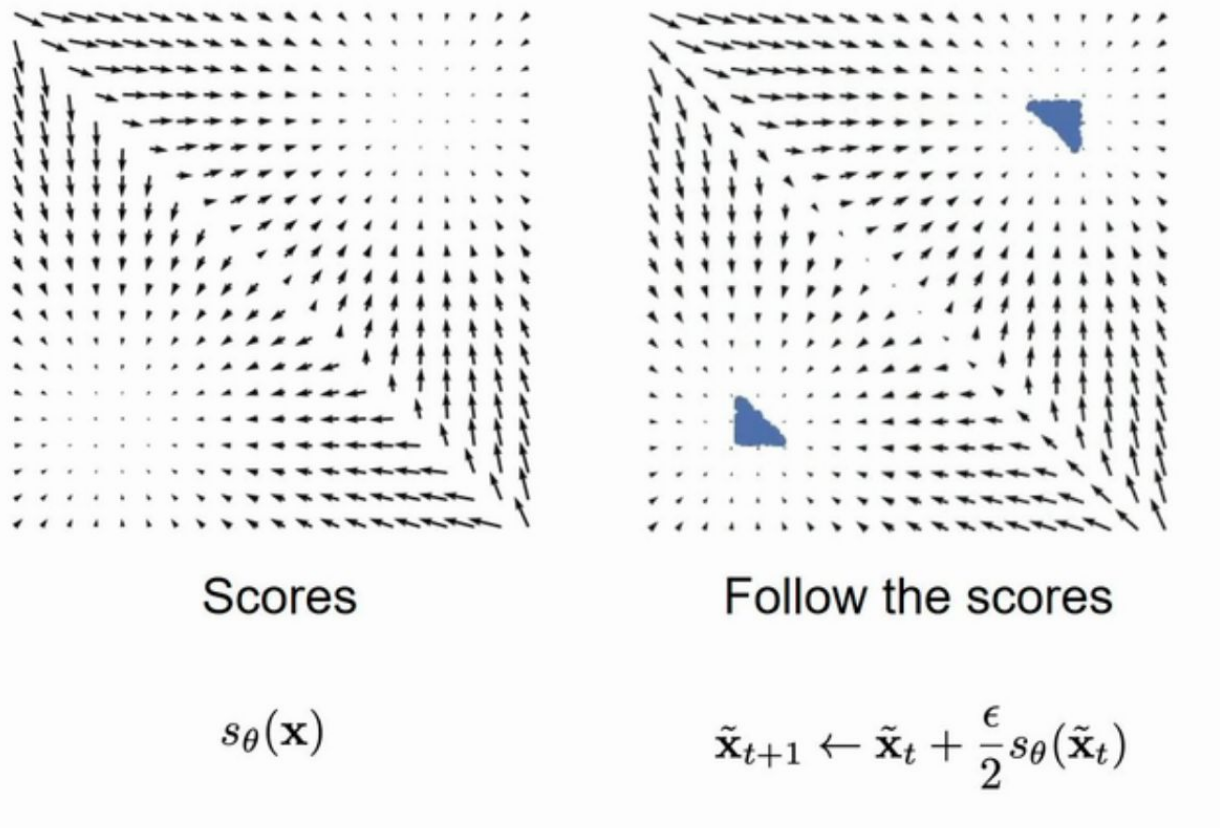
How to sample:



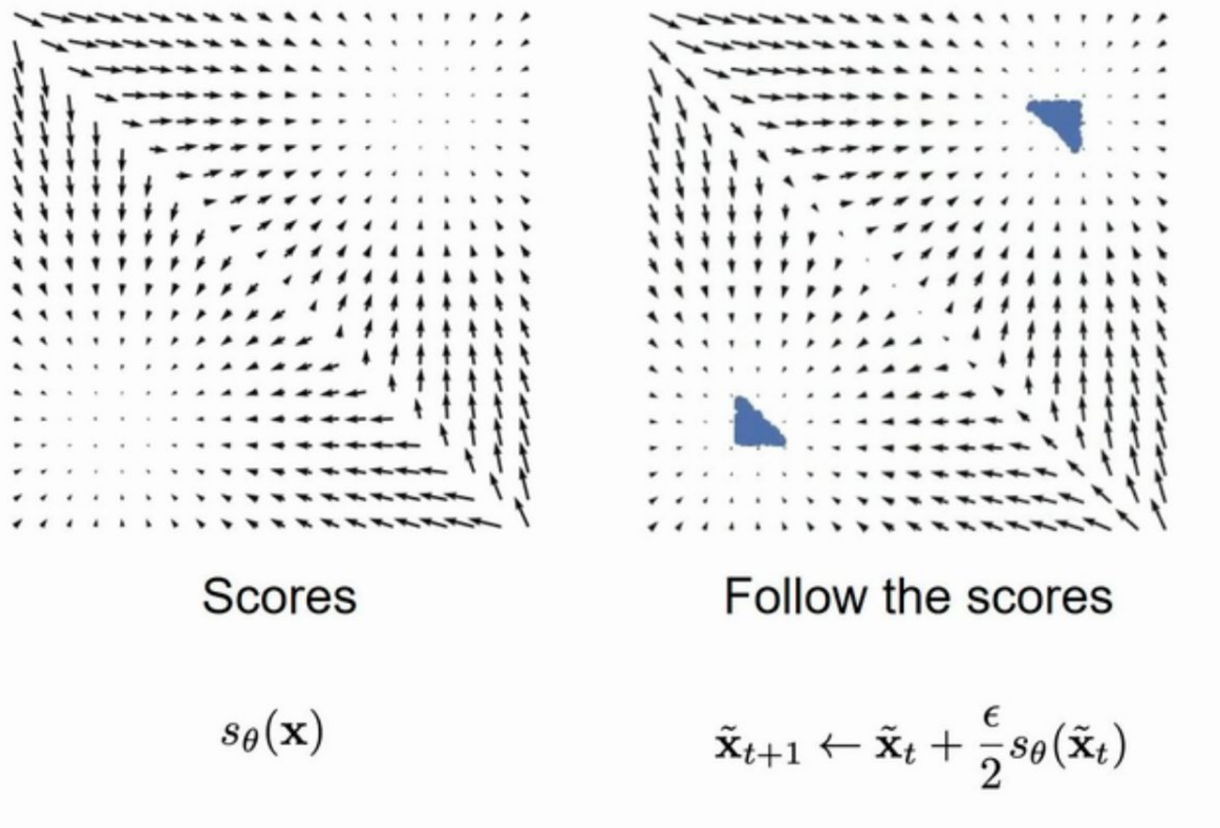
Scores

$s_\theta(\mathbf{x})$

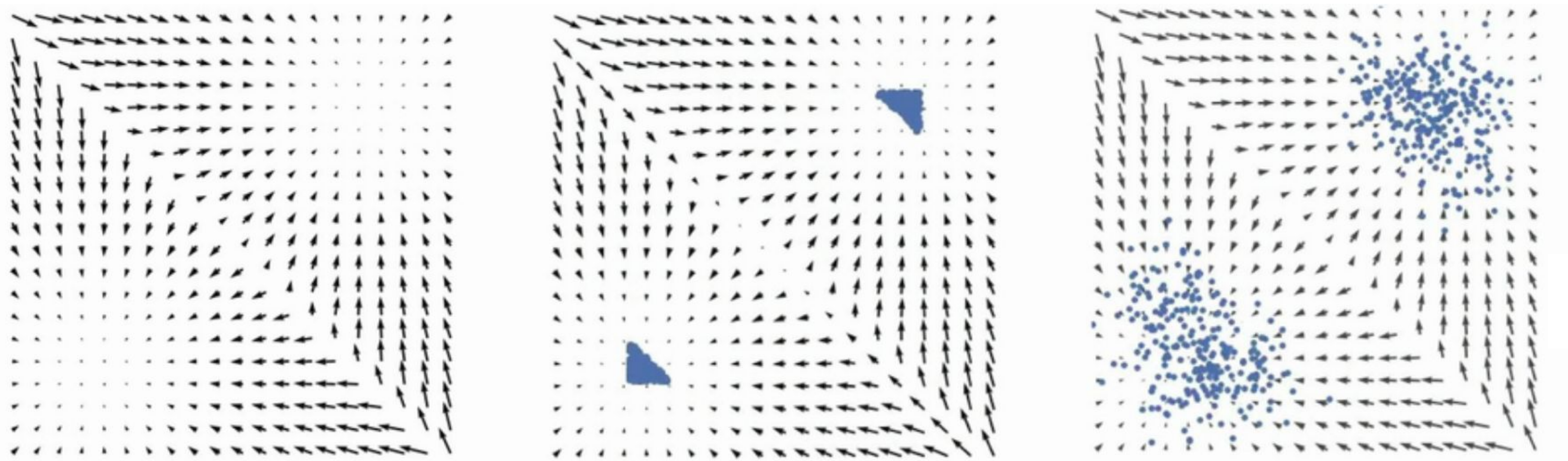
How to sample:



How to sample:



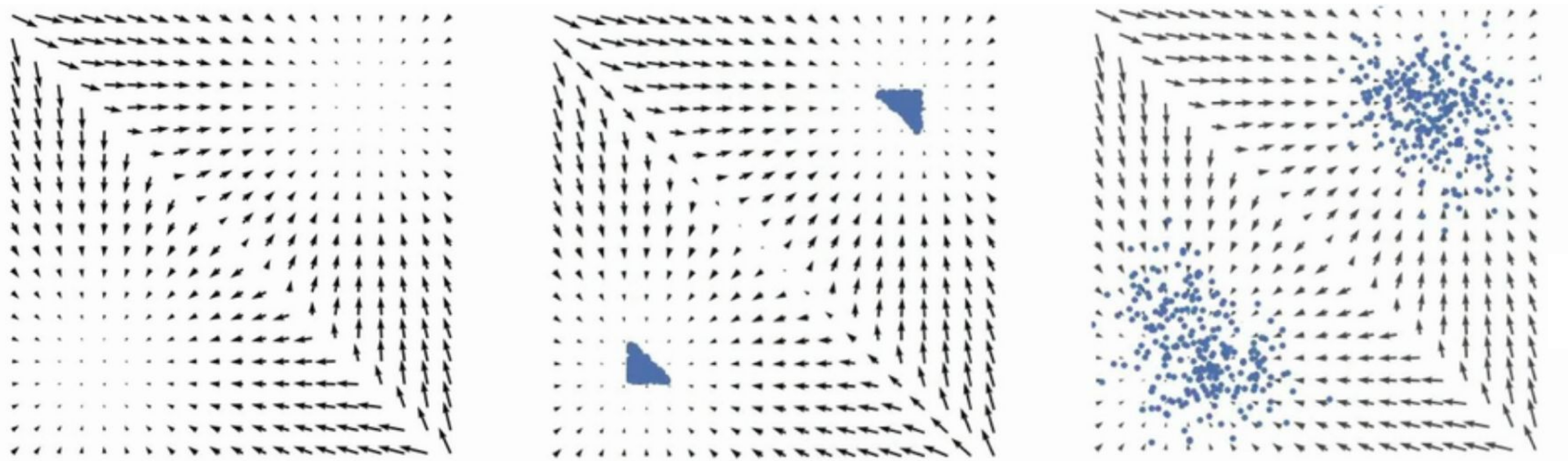
How to sample:



Scores	Follow the scores	Follow noisy scores: Langevin dynamics
$s_\theta(\mathbf{x})$	$\tilde{\mathbf{x}}_{t+1} \leftarrow \tilde{\mathbf{x}}_t + \frac{\epsilon}{2} s_\theta(\tilde{\mathbf{x}}_t)$	$\mathbf{z}_t \sim \mathcal{N}(0, I)$ $\tilde{\mathbf{x}}_{t+1} \leftarrow \tilde{\mathbf{x}}_t + \frac{\epsilon}{2} s_\theta(\tilde{\mathbf{x}}_t) + \sqrt{\epsilon}\mathbf{z}_t$

dkfz.

How to sample:



Scores

$$s_\theta(\mathbf{x})$$

Follow the scores

$$\tilde{\mathbf{x}}_{t+1} \leftarrow \tilde{\mathbf{x}}_t + \frac{\epsilon}{2} s_\theta(\tilde{\mathbf{x}}_t)$$

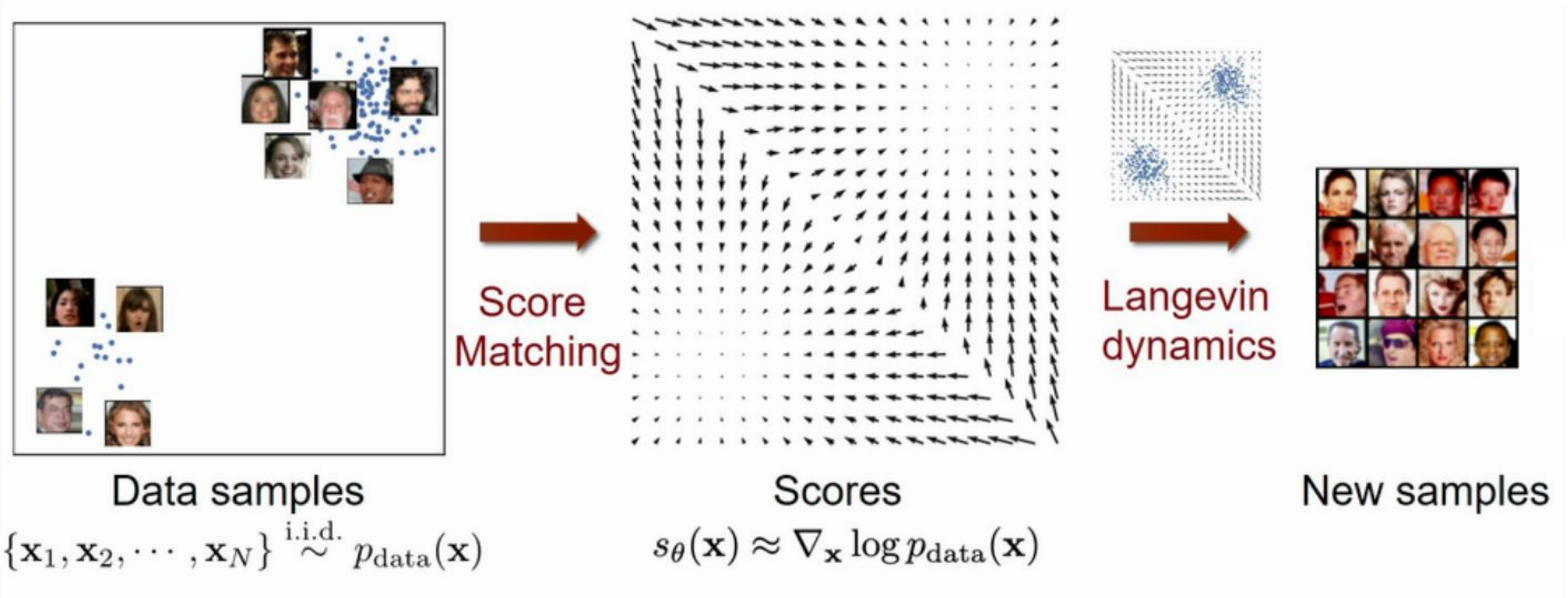
Follow noisy scores:
Langevin dynamics

$$\mathbf{z}_t \sim \mathcal{N}(0, I)$$

$$\tilde{\mathbf{x}}_{t+1} \leftarrow \tilde{\mathbf{x}}_t + \frac{\epsilon}{2} s_\theta(\tilde{\mathbf{x}}_t) + \sqrt{\epsilon}\mathbf{z}_t$$

Approach

Approach

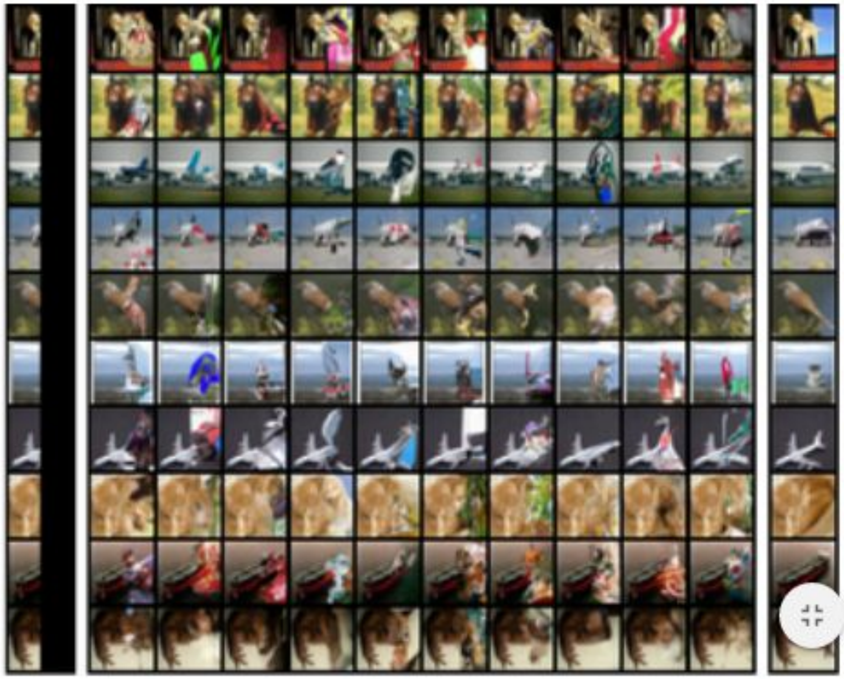

Score
Matching
Langevin
dynamics
Data samples
$\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\} \overset{\text{i.i.d.}}{\sim} p_{\text{data}}(\mathbf{x})$
Scores
$s_\theta(\mathbf{x}) \approx \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})$
New samples

Results

Results: Qualitative

dkfz.

Results: Qualitative



Results: Quantitative

Model	Inception ⇧	FID ⇩
CIFAR-10 Unconditional		
PixelCNN [59]	4.60	65.93
PixelIQN [42]	5.29	49.46
EBM [12]	6.02	40.58
WGAN-GP [18]	$7.86 \pm .07$	36.4
MoLM [45]	$7.90 \pm .10$	**18.9**
SNGAN [36]	$8.22 \pm .05$	21.7
ProgressiveGAN [25]	$8.80 \pm .05$	-
NCSN (Ours)	$\mathbf{8.87} \pm .12$	25.32
CIFAR-10 Conditional		
EBM [12]	8.30	37.9
SNGAN [36]	$8.60 \pm .08$	25.5
BigGAN [6]	**9.22**	**14.73**

Results: Reproducible

→ NeurIPS 2019 - Reproducibility Challenge[2]

[2] A. Matosevic. Reproducibility Challenge – Generative Modeling by Estimating Gradients of the Data Distribution, NeurIPS 2019 Reproducibility Challenge Blind Report, https://openreview.net/forum?id=SkxCSTqG6H

Results: Reproducible

→ NeurIPS 2019 - Reproducibility Challenge[2]

Reproducible ?

[2] A. Matosevic. Reproducibility Challenge – Generative Modeling by Estimating Gradients of the Data Distribution, NeurIPS 2019 Reproducibility Challenge Blind Report, https://openreview.net/forum?id=SkxCSTqG6H

Results: Reproducible

→ NeurIPS 2019 - Reproducibility Challenge[2]

Reproducible ?

[2] A. Matosevic. Reproducibility Challenge – Generative Modeling by Estimating Gradients of the Data Distribution, NeurIPS 2019 Reproducibility Challenge Blind Report, https://openreview.net/forum?id=SkxCSTqG6H

Results: Reproducible

→ NeurIPS 2019 - Reproducibility Challenge[2]

Reproducible ?

Hyper-parameter insensitive ?

[2] A. Matosevic. Reproducibility Challenge – Generative Modeling by Estimating Gradients of the Data Distribution, NeurIPS 2019 Reproducibility Challenge Blind Report, https://openreview.net/forum?id=SkxCSTqG6H

Results: Reproducible

→ NeurIPS 2019 - Reproducibility Challenge[2]

Reproducible ? 👍

Hyper-parameter insensitive ? 👎

[2] A. Matosevic. Reproducibility Challenge – Generative Modeling by Estimating Gradients of the Data Distribution, NeurIPS 2019 Reproducibility Challenge Blind Report, https://openreview.net/forum?id=SkxCSTqG6H

The End