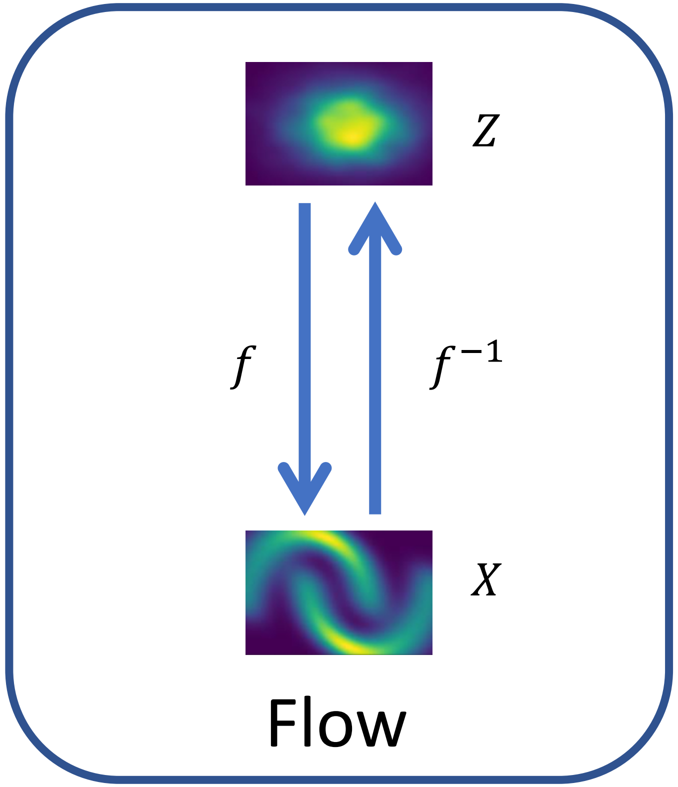
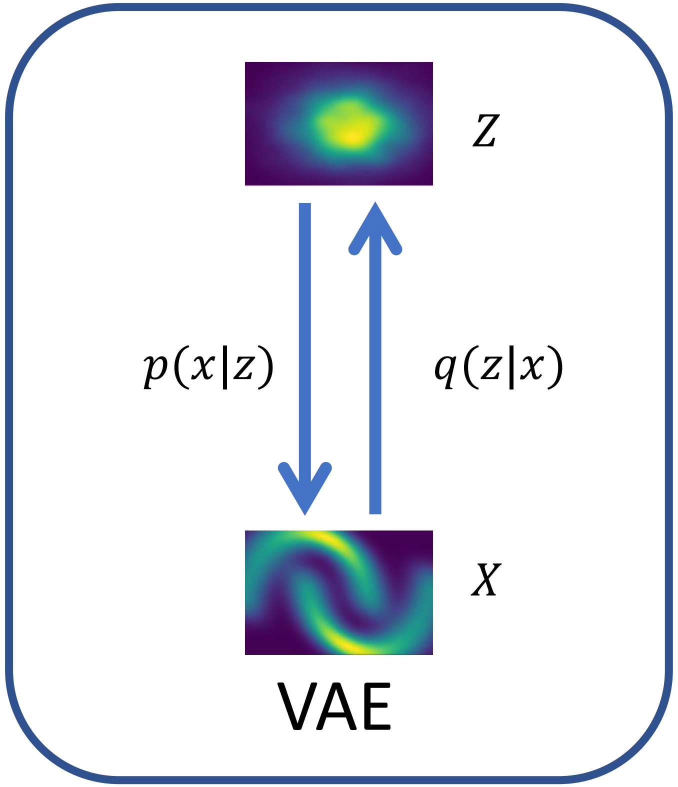
SurVAE Flows: Surjections to Bridge the Gap between VAEs and Flows

Didrik Nielsen[1], **Priyank Jaini**[2], **Emiel Hoogeboom**[2], **Ole Winther**[1], **Max Welling**[2]
Technical University of Denmark[1], UvA-Bosch Delta Lab, University of Amsterdam[2]
didni@dtu.dk, p.jaini@uva.nl, e.hoogeboom@uva.nl
olwi@dtu.dk, m.welling@uva.nl

Recap: Normalizing Flows and VAEs





$$\log p(x) = \log p(z) + \log|\det D_x\, z|$$
$$z = f^{-1}(x)$$

$$\log p(x) \geq E_{q(z|x)}\left[\log p(z) + \log\frac{p(x|z)}{q(z|x)}\right]$$
$$\approx \log p(z) + \log\frac{p(x|z)}{q(z|x)}, with\ z \sim q(z|x)$$

Flows

+ Efficient to sample x $\sim p(x)$

+ Expressive (Villani 2003)

+ Exact computation of $\log p(x)$

- No dimensionality reduction

VAEs

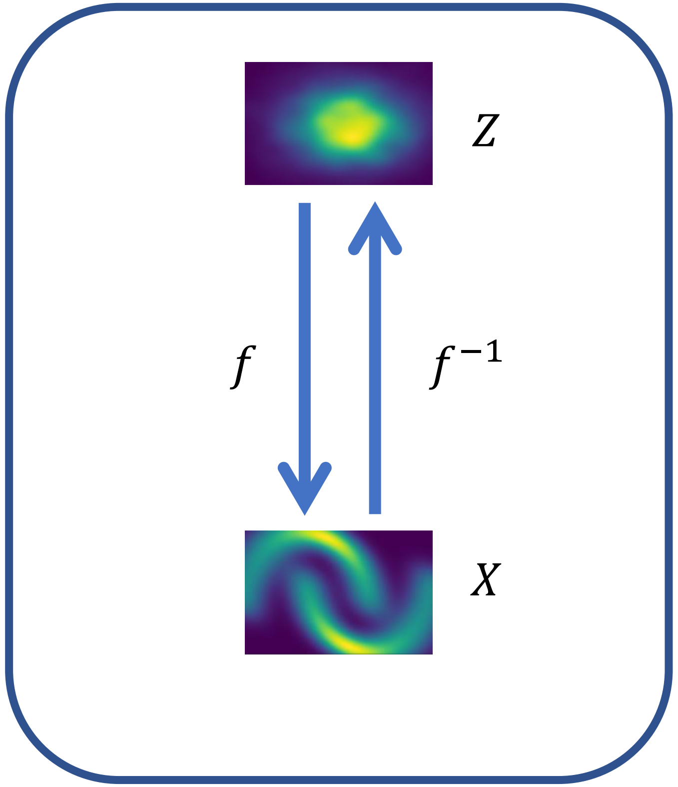
+ Efficient to sample x $\sim p(x)$

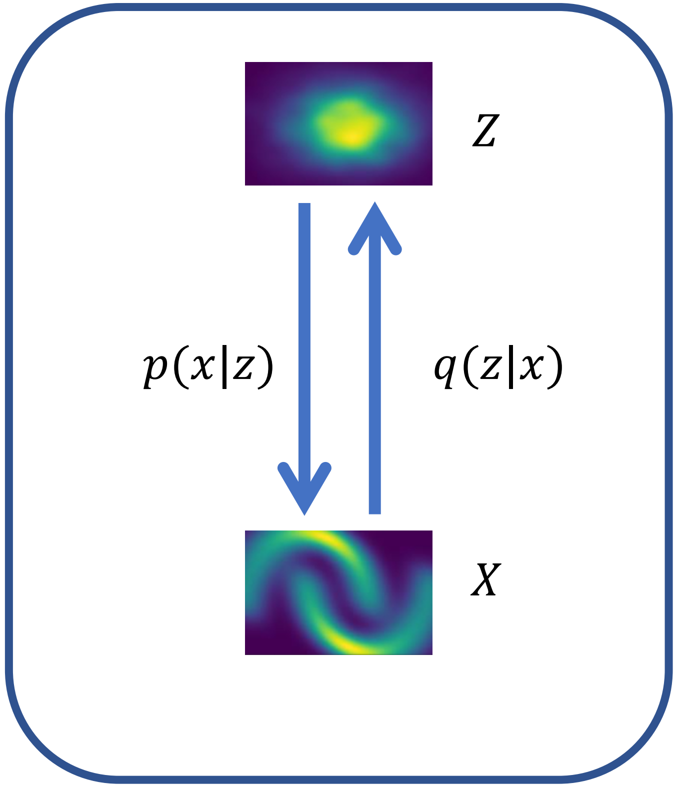
+ Useful lower-dimensional latent representation

- No exact computation of $\log p(x)$

Is it possible to have composable transformations that can alter dimensionality and allow for exact likelihood evaluation?

Combining Flows and VAEs





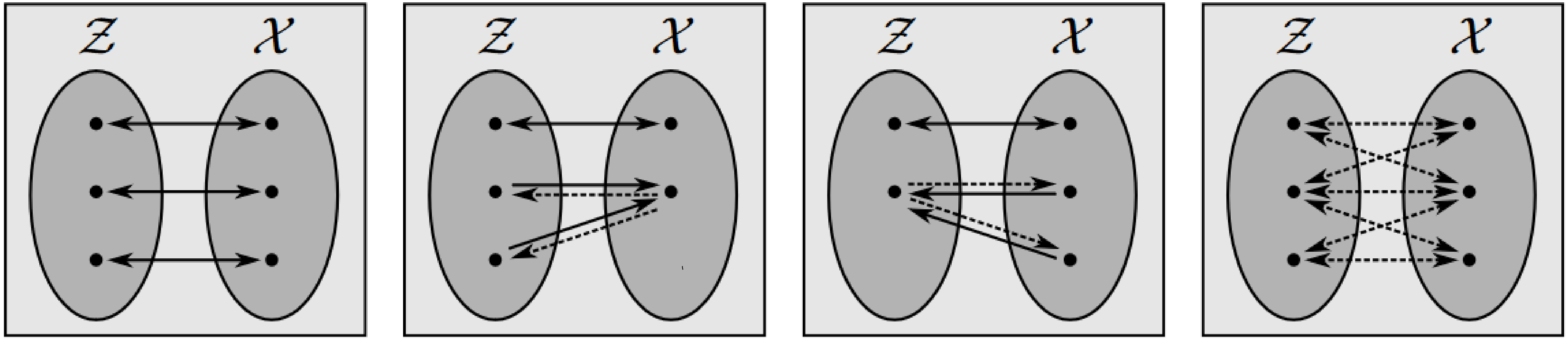
1. Forward Transformation : $p(x|z) = \delta\big(x - f(z)\big)$
2. Inverse Transformation: $q(z|x) = \delta\big(z - f^{-1}(x)\big)$
3. Likelihood Contribution:

$$\log p(x) = \log p(z) + \log|\det D_x z| \qquad z = f^{-1}(x)$$

$$\log p(x) \geq \log p(z) + \log \frac{p(x|z)}{q(z|x)} \qquad z \sim q(z|x)$$

Flow + VAE = SurVAE Flow

Sur(jection)VAE Flow



Flow | Generative Surjection | Inference Surjection | VAE

- $f: \mathcal{Z} \rightarrow \mathcal{X}$ is surjective, if $\forall x \in \mathcal{X}$ there exists $z \in \mathcal{Z}$ such that $x = f(z)$
- Surjections are in general not invertible, but right inverses exists i.e $g: \mathcal{X} \rightarrow \mathcal{Z}$, such that $f \circ g(x) = x$
- g can be stochatic, i.e $q(z|x)$, with support only over the preimage of x, $B(x) \coloneqq \{z | x = f(z)\}$

Composable building blocks of SurVAE Flows

$$\begin{aligned}
\log p(x) &= E_{q(z|x)}[\log p(x|z)] - D_{kl}[q(z|x)||p(z)] + D_{kl}[q(z|x)||p(z|x)] \\
&= E_{q(z|x)}\left[\log p(x|z) - \log\frac{q(z|x)}{p(z)} + \log\frac{q(z|x)}{p(z|x)}\right] \\
&= E_{q(z|x)}\left[\log p(z) + \log\frac{p(x|z)}{q(z|x)} + \log\frac{q(z|x)}{p(z|x)}\right] \\
&\approx \log p(z) + \log\frac{p(x|z)}{q(z|x)} + \log\frac{q(z|x)}{p(z|x)}, \quad \text{where } z \sim q(z|x) \\
&= \log p(z) + \mathcal{V}(x,z) + \mathcal{E}(x,z), \quad \text{where } z \sim q(z|x)
\end{aligned}$$

Transformation	**Forward** $\boldsymbol{x} \leftarrow \boldsymbol{z}$	**Inverse** $\boldsymbol{z} \leftarrow \boldsymbol{x}$	**Likelihood Contribution** $\mathcal{V}(\boldsymbol{x},\boldsymbol{z})$	**Bound Gap** $\mathcal{E}(\boldsymbol{x},\boldsymbol{z})$
Bijective	$\boldsymbol{x} = f(\boldsymbol{z})$	$\boldsymbol{z} = f^{-1}(\boldsymbol{x})$	$\log\lvert\det \nabla_{\boldsymbol{x}} \boldsymbol{z}\rvert$	0
Stochastic	$\boldsymbol{x} \sim p(\boldsymbol{x}\|\boldsymbol{z})$	$\boldsymbol{z} \sim q(\boldsymbol{z}\|\boldsymbol{x})$	$\log \frac{p(\boldsymbol{x}\|\boldsymbol{z})}{q(\boldsymbol{z}\|\boldsymbol{x})}$	$\log \frac{q(\boldsymbol{z}\|\boldsymbol{x})}{p(\boldsymbol{z}\|\boldsymbol{x})}$
Surjective (Gen.)	$\boldsymbol{x} = f(\boldsymbol{z})$	$\boldsymbol{z} \sim q(\boldsymbol{z}\|\boldsymbol{x})$	$\log \frac{p(\boldsymbol{x}\|\boldsymbol{z})}{q(\boldsymbol{z}\|\boldsymbol{x})}$ as $p(\boldsymbol{x}\|\boldsymbol{z}) \rightarrow \delta(\boldsymbol{x} - f(\boldsymbol{z}))$	$\log \frac{q(\boldsymbol{z}\|\boldsymbol{x})}{p(\boldsymbol{z}\|\boldsymbol{x})}$
Surjective (Inf.)	$\boldsymbol{x} \sim p(\boldsymbol{x}\|\boldsymbol{z})$	$\boldsymbol{z} = f^{-1}(\boldsymbol{x})$	$\log \frac{p(\boldsymbol{x}\|\boldsymbol{z})}{q(\boldsymbol{z}\|\boldsymbol{x})}$ as $q(\boldsymbol{z}\|\boldsymbol{x}) \rightarrow \delta(\boldsymbol{z} - f^{-1}(\boldsymbol{x}))$	0

Likelihood Contribution

Likelihood Contribution $\mathcal{V}(\boldsymbol{x}, \boldsymbol{z})$
$\log \lvert \det \nabla_{\boldsymbol{x}} \boldsymbol{z} \rvert$
$\log \frac{p(\boldsymbol{x}\mid\boldsymbol{z})}{q(\boldsymbol{z}\mid\boldsymbol{x})}$
$\log \frac{p(\boldsymbol{x}\mid\boldsymbol{z})}{q(\boldsymbol{z}\mid\boldsymbol{x})}$ as $p(\boldsymbol{x}\mid\boldsymbol{z}) \to \delta(\boldsymbol{x} - f(\boldsymbol{z}))$
$\log \frac{p(\boldsymbol{x}\mid\boldsymbol{z})}{q(\boldsymbol{z}\mid\boldsymbol{x})}$ as $q(\boldsymbol{z}\mid\boldsymbol{x}) \to \delta(\boldsymbol{z} - f^{-1}(\boldsymbol{x}))$

$$f = f_1 \circ f_2 \circ \cdots \circ f_T$$

Algorithm 1: $\log - \text{likelihood}(\boldsymbol{x})$

Data: $\boldsymbol{x}$, $p(\boldsymbol{z})$ & $\{f_t\}_{t=1}^{T}$

Result: $\mathcal{L}(\boldsymbol{x})$

for t *in* range(T), **do**

  if f_t *is bijective* **then**

    $\boldsymbol{z} = f_t^{-1}(\boldsymbol{x})$;

    $\mathcal{V}_t = \log \left|\det \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\right|$;

  else if f_t *is stochastic* **then**

    $\boldsymbol{z} \sim q_t(\boldsymbol{z}\mid\boldsymbol{x})$;

    $\mathcal{V}_t = \log \frac{p_t(\boldsymbol{x}\mid\boldsymbol{z})}{q_t(\boldsymbol{z}\mid\boldsymbol{x})}$;

  $\boldsymbol{x} = \boldsymbol{z}$;

end

return $\log p(\boldsymbol{z}) + \sum_{t=1}^{T} \mathcal{V}_t$

Tensor Slicing

Let $x = (x_1,\ x_2) \in\ \mathbb{R}^d,\ x_1 =\ f^{-1}(x) = z$

$$p(x|z) = p(x_2|z)N(x_1|z, \sigma I)$$

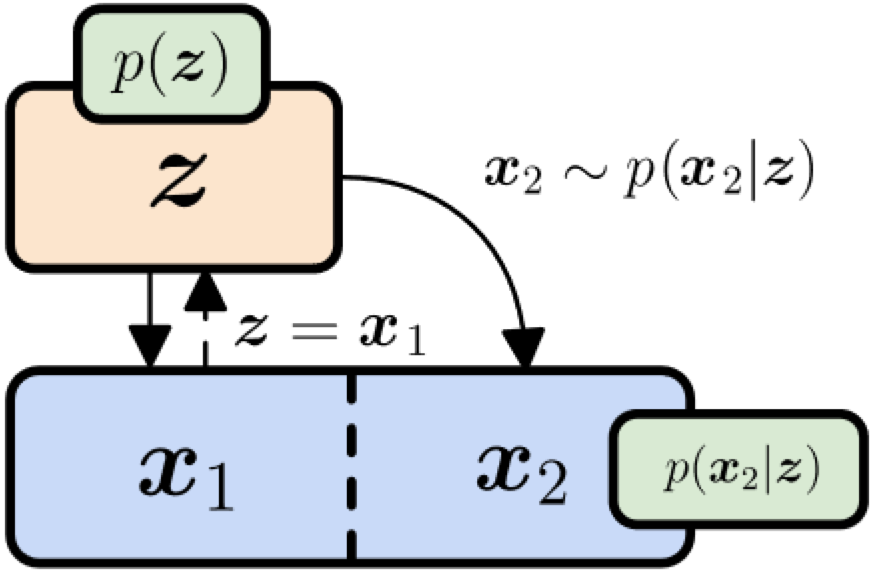
$$q(z|x) = N(z|x_1, \sigma I) \xrightarrow{\sigma \to 0} \delta(z - f^{-1}(x))$$

$$\mathcal{V} = \lim_{\sigma \to 0} E_{q(z|x)}\left[\log \frac{p(x|z)}{q(z|x)}\right]$$

$$= \lim_{\sigma \to 0} E_{q(z|x)}\left[\log \frac{p(x_2|z)N(x_1|z, \sigma I)}{N(z|x_1, \sigma I)}\right]$$

$$= \lim_{\sigma \to 0} E_{q(z|x)}[\log p(x_2|z)]$$

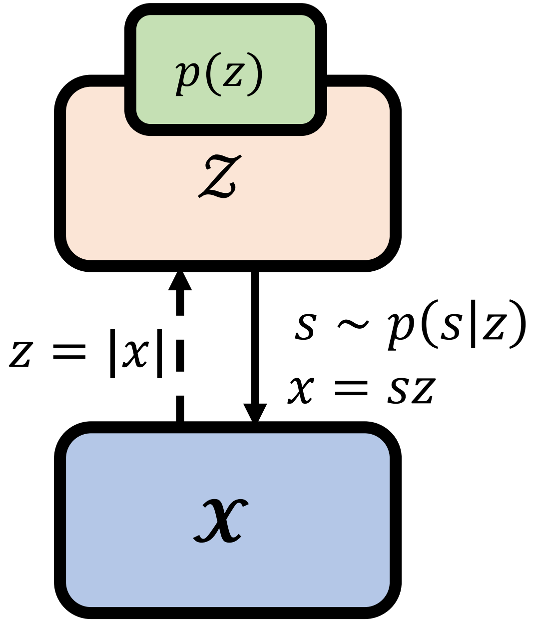
$\approx \log(\mathrm{p}(\mathrm{x}_2|\mathrm{z}))$, where $z \sim \delta(z - f^{-1}(x))$



Absolute Value Surjection in inference direction

$$p(x|z) = \sum_{s\in\{1,-1\}} p(x|z,s)p(s|z) = \sum_{s\in\{1,-1\}} \delta(x - sz)p(s|z)$$

$$q(z|x) = \sum_{s\in\{1,-1\}} q(z|x,s)p(s|x) = \sum_{s\in\{1,-1\}} \delta(z - sx)\delta_{s,sign(x)}$$



$$\mathcal{V} = E_{q(z|x,\mathrm{s})\mathrm{q}(\mathrm{s}|\mathrm{x})}\left[\log \frac{p(x|\mathrm{Z},S)\mathrm{p}(\mathrm{S}|\mathrm{Z})}{q(z|x,s)q(s|x)}\right]$$

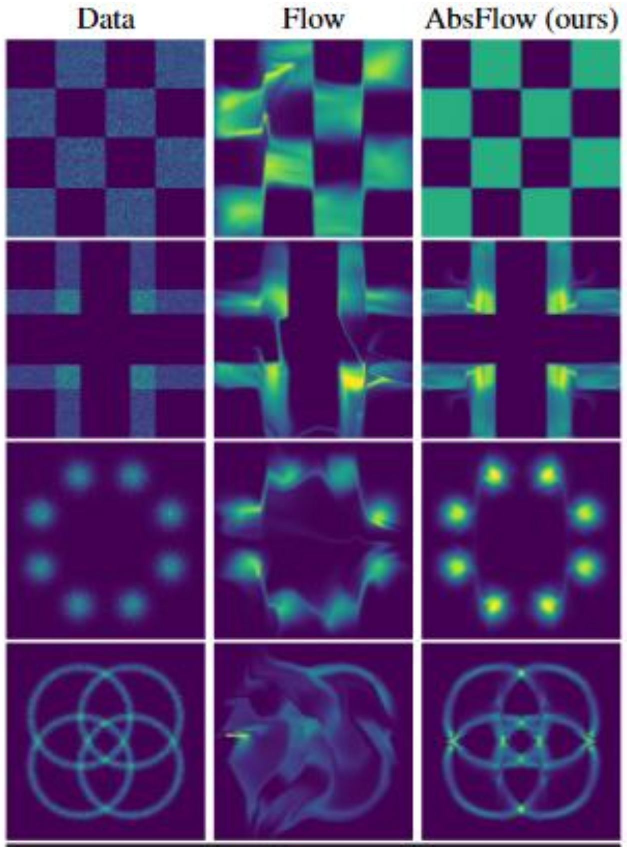
$$= E_{\delta(z-sx)\delta_{sign(x),s}}\left[\log \frac{\delta(x-sz)p(S|Z)}{\delta(z-sx)\delta_{s,sign(x)}}\right]$$

$$\approx \log p(s|z), \text{ where } z = |x|, s = sign(x)$$

Summary of generative surjection layers

Surjection	Forward	Inverse	$\mathcal{V}(\boldsymbol{x}, \boldsymbol{z})$
Rounding	$x = \lfloor z \rfloor$	$z \sim q(z\|x)$ where $z \in [x, x+1)$	$-\log q(z\|x)$
Slicing	$\boldsymbol{x} = \boldsymbol{z}_1$	$\boldsymbol{z}_1 = \boldsymbol{x}, \boldsymbol{z}_2 \sim q(\boldsymbol{z}_2\|\boldsymbol{x})$	$-\log q(\boldsymbol{z}_2\|\boldsymbol{x})$
Abs	$s = \operatorname{sign} z$ $x = \|z\|$	$s \sim \operatorname{Bern}(\pi(x))$ $z = s \cdot x, \quad s \in \{1, -1\}$	$-\log q(s\|x)$
Max	$k = \arg\max \boldsymbol{z}$ $x = \max \boldsymbol{z}$	$k \sim \operatorname{Cat}(\boldsymbol{\pi}(x))$ $z_k = x, \boldsymbol{z}_{-k} \sim q(\boldsymbol{z}_{-k}\|x, k)$	$-\log q(k\|x) - \log q(\boldsymbol{z}_{-k}\|x, k)$
Sort	$\mathcal{I} = \operatorname{argsort} \boldsymbol{z}$ $\boldsymbol{x} = \operatorname{sort} \boldsymbol{z}$	$\mathcal{I} \sim \operatorname{Cat}(\boldsymbol{\pi}(\boldsymbol{x}))$ $\boldsymbol{z} = \boldsymbol{x}_{\mathcal{I}}$	$-\log q(\mathcal{I}\|\boldsymbol{x})$
ReLU	$x = \max(z, 0)$	if $x = 0 : z \sim q(z)$, else $: z = x$	$\mathbb{I}(x = 0)[-\log q(z)]$

Synthetic Data



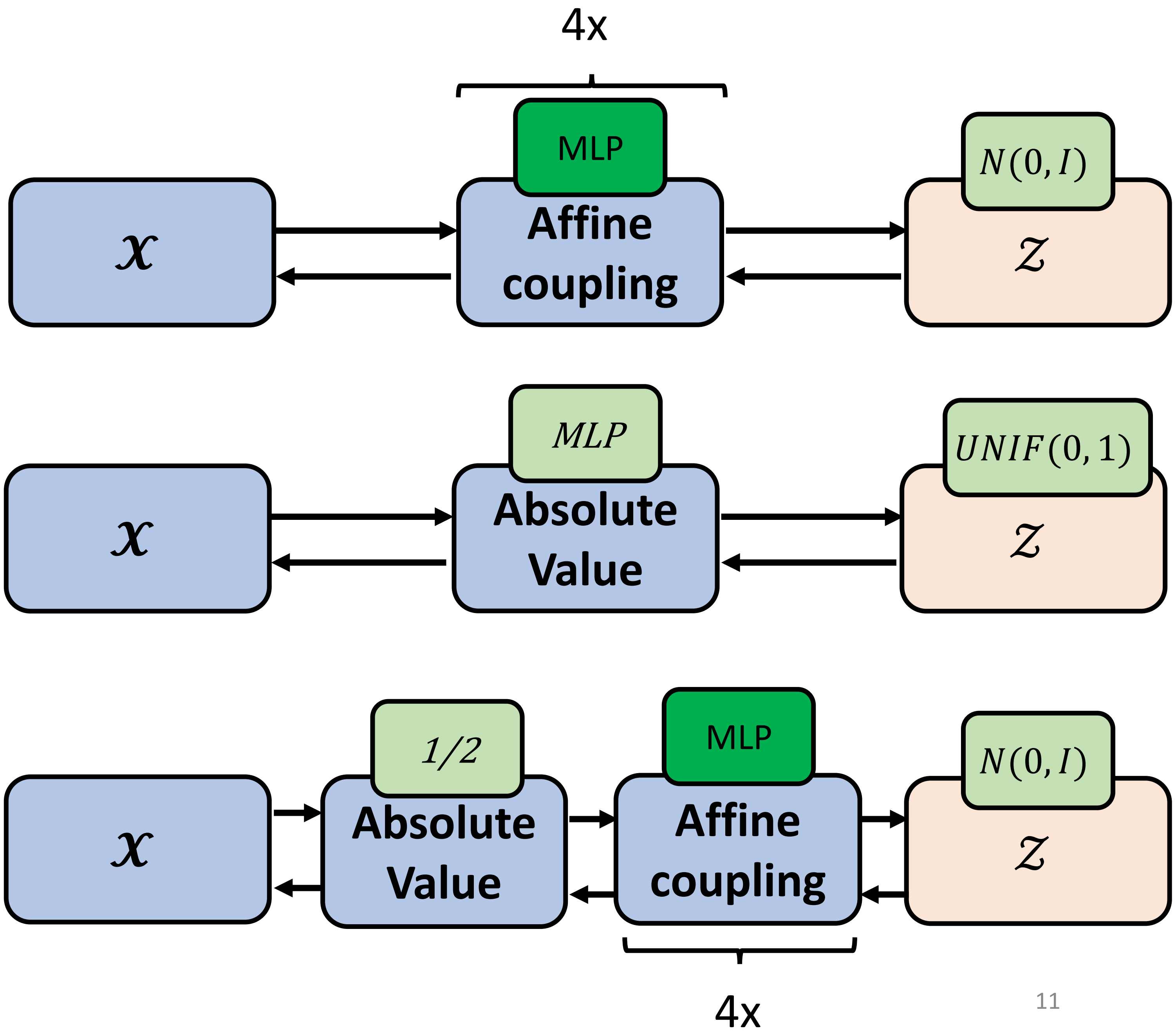
Flow

AbsFlow (Anti-Sym.)

AbsFlow (Symmetric)

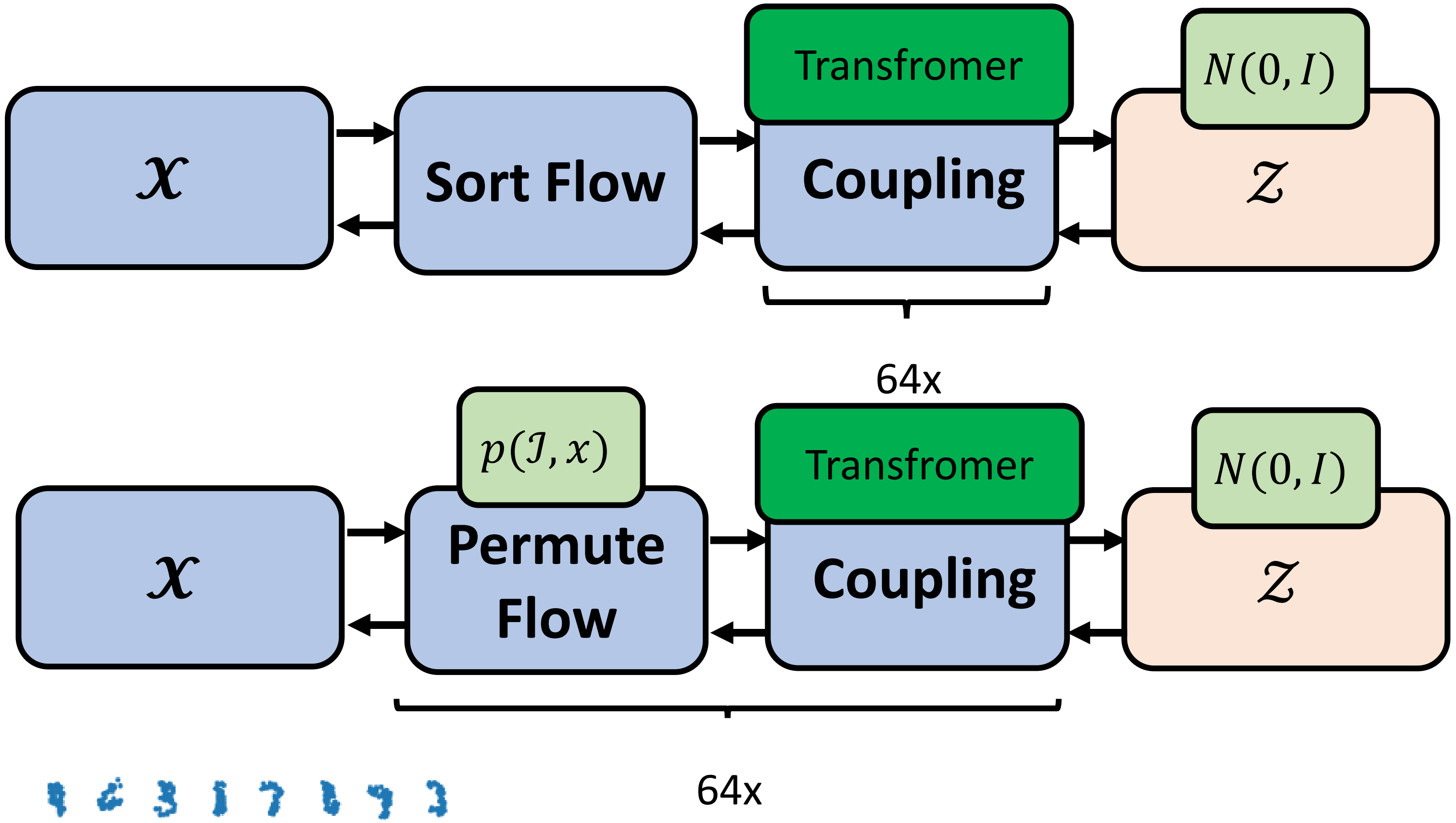
Dataset	Flow	AbsFlow (ours)
Checkerboard	3.65	**3.49**
Corners	3.19	**3.03**
Gaussians	3.01	**2.86**
Circles	3.44	**2.99**

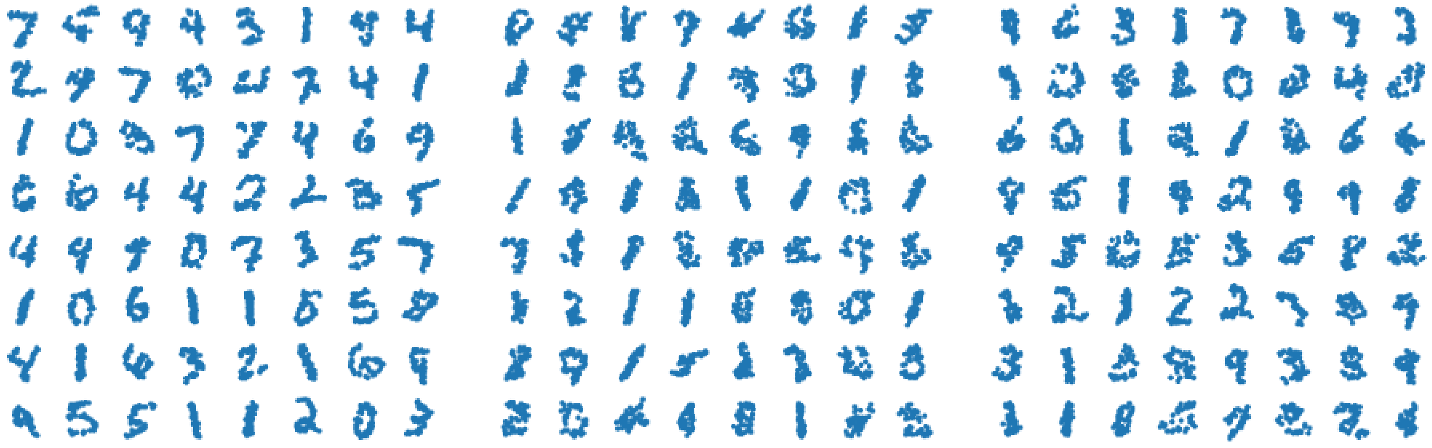
in $-\log p(x)$



Point Cloud Data

Model	PPLL
PermuteFlow	-5.30*
SortFlow	-5.53
Neural Statistican (Edwards and Storkey, 2017)	-5.37
FlowScan (Bender et al., 2020)	-5.26**





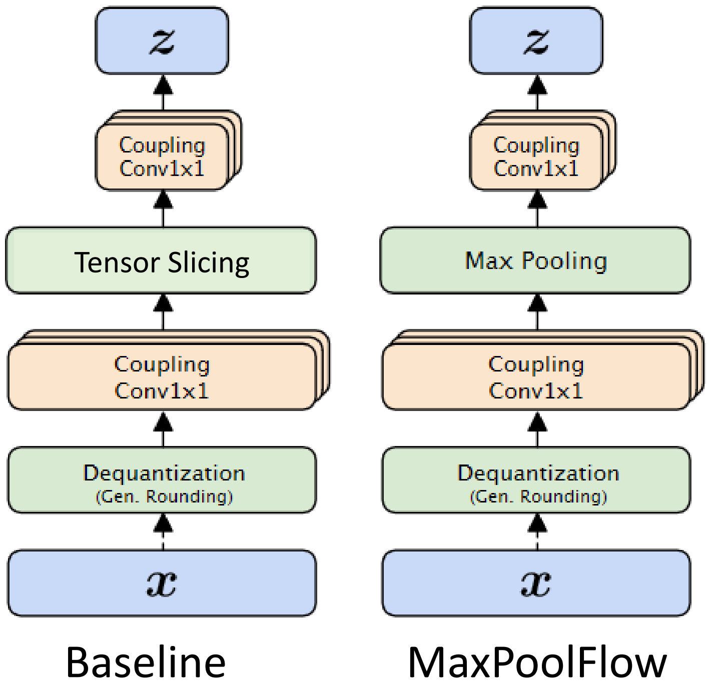
(a) Data (b) SortFlow (c) PermuteFlow

Image Data

bits/dim ↓

Model	CIFAR-10	ImageNet32	ImageNet64
RealNVP (Dinh et al., 2017)	3.49	4.28	-
Glow (Kingma and Dhariwal, 2018)	3.35	4.09	3.81
Flow++ (Ho et al., 2019)	3.08	3.86	3.69
Baseline (Ours)	**3.08**	**4.00**	**3.70**
MaxPoolFlow (Ours)	3.09	4.01	3.74

Model	Inception ↑	FID ↓
DCGAN*	6.4	37.1
WGAN-GP*	6.5	36.4
PixelCNN*	4.60	65.93
PixelIQN*	5.29	49.46
Baseline (Ours)	5.08	49.56
MaxPoolFlow (Ours)	**5.18**	**49.03**



Contributions of the paper

+ Provides a unifying theoretical framework for many models and architectures

+ Shows that dimensionality reduction with exactly tractable log likelihood is possible

+ Offers a full software implementation of SurVAE Flow

- Experiments show only minor improvements or include strong inductive bias

- No Runtime comparison in the experiments

Model	**SurVAE Flow architecture**
Probabilistic PCA (Tipping and Bishop, 1999) VAE (Kingma and Welling, 2014; Rezende et al., 2014) Diffusion Models (Sohl-Dickstein et al., 2015; Ho et al., 2020)	$\mathcal{Z} \xrightarrow{\text{stochastic}} \mathcal{X}$
Dequantization (Uria et al., 2013; Ho et al., 2019)	$\mathcal{Z} \xrightarrow{\text{round}} \mathcal{X}$
ANFs, VFlow (Huang et al., 2020; Chen et al., 2020)	$\mathcal{X} \xrightarrow{\text{augment}} \mathcal{X} \times \mathcal{E} \xrightarrow{\text{bijection}} \mathcal{Z}$
Multi-scale Architectures (Dinh et al., 2017)	$\mathcal{X} \xrightarrow{\text{bijection}} \mathcal{Y} \times \mathcal{E} \xrightarrow{\text{slice}} \mathcal{Y} \xrightarrow{\text{bijection}} \mathcal{Z}$
CIFs, Discretely Indexed Flows, DeepGMMs (Cornish et al., 2019; Duan, 2019; Oord and Dambre, 2015)	$\mathcal{X} \xrightarrow{\text{augment}} \mathcal{X} \times \mathcal{E} \xrightarrow{\text{bijection}} \mathcal{Z} \times \mathcal{E} \xrightarrow{\text{slice}} \mathcal{Z}$
RAD Flows (Dinh et al., 2019)	$\mathcal{X} \xrightarrow{\text{partition}} \mathcal{X}_{\mathcal{E}} \times \mathcal{E} \xrightarrow{\text{bijection}} \mathcal{Z} \times \mathcal{E} \xrightarrow{\text{slice}} \mathcal{Z}$