Mapping x to z (General)

$$z \sim p_\theta(z)$$

$$x = g_\theta(z)$$

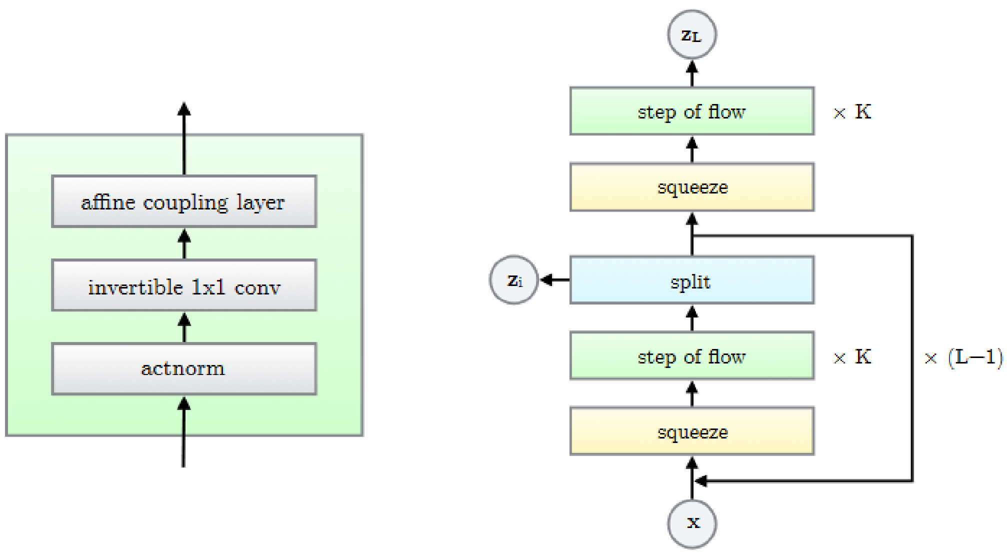
The z is latent variable, sample from probability distribution p_θ, normally normal distribution. g_θ is transformative function, invertible (or *bijective*)

After *change of variables*, probability density function (pdf) given a data-point will be:

$$\log p_\theta(x) = \log p_\theta(z) + \log\left|\det\left(\frac{dz}{dx}\right)\right|$$
$$= \log p_\theta(z) + \sum_{i=1}^{K} \log\left|\det\left(\frac{dh_i}{dh_{i-1}}\right)\right|$$

Jacobian matrix $\frac{dh_i}{dh_{i-1}}$ is a triangular matrix

GLOW



(a) One step of our flow. (b) Multi-scale architecture (Dinh et al., 2016).

Description	Function	Reverse Function	Log-determinant
Actnorm. See Section 3.1.	$\forall i,j : \mathbf{y}_{i,j} = \mathbf{s} \odot \mathbf{x}_{i,j} + \mathbf{b}$	$\forall i,j : \mathbf{x}_{i,j} = (\mathbf{y}_{i,j} - \mathbf{b})/\mathbf{s}$	$h \cdot w \cdot \mathtt{sum}(\log\lvert\mathbf{s}\rvert)$
Invertible 1×1 convolution. $\mathbf{W} : [c \times c]$. See Section 3.2.	$\forall i,j : \mathbf{y}_{i,j} = \mathbf{W}\mathbf{x}_{i,j}$	$\forall i,j : \mathbf{x}_{i,j} = \mathbf{W}^{-1}\mathbf{y}_{i,j}$	$h \cdot w \cdot \log\lvert\det(\mathbf{W})\rvert$ or $h \cdot w \cdot \mathtt{sum}(\log\lvert\mathbf{s}\rvert)$ (see eq. (10))
Affine coupling layer. See Section 3.3 and (Dinh et al., 2014)	$\mathbf{x}_a, \mathbf{x}_b = \mathtt{split}(\mathbf{x})$ $(\log\mathbf{s}, \mathbf{t}) = \mathtt{NN}(\mathbf{x}_b)$ $\mathbf{s} = \exp(\log\mathbf{s})$ $\mathbf{y}_a = \mathbf{s} \odot \mathbf{x}_a + \mathbf{t}$ $\mathbf{y}_b = \mathbf{x}_b$ $\mathbf{y} = \mathtt{concat}(\mathbf{y}_a, \mathbf{y}_b)$	$\mathbf{y}_a, \mathbf{y}_b = \mathtt{split}(\mathbf{y})$ $(\log\mathbf{s}, \mathbf{t}) = \mathtt{NN}(\mathbf{y}_b)$ $\mathbf{s} = \exp(\log\mathbf{s})$ $\mathbf{x}_a = (\mathbf{y}_a - \mathbf{t})/\mathbf{s}$ $\mathbf{x}_b = \mathbf{y}_b$ $\mathbf{x} = \mathtt{concat}(\mathbf{x}_a, \mathbf{x}_b)$	$\mathtt{sum}(\log(\lvert\mathbf{s}\rvert))$

Loss (Bits per dimension)

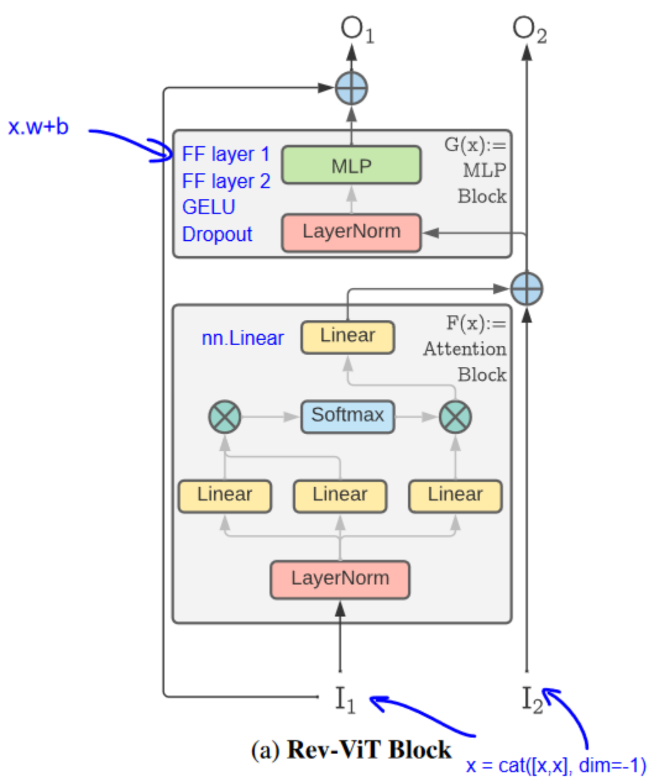
```
n_pixel = image_size * image_size * 3

loss = -log(n_bins) * n_pixel # loss considering noise inputted.
loss = loss + logdet + log_p
loss = (-loss / (log(2) * n_pixel)).mean()
```

$$\frac{1}{n_{pixel}\log 2}\left(\log n_{bins}\times n_{pixel}-\log p_\theta(z)-\sum_{i=1}^{K}\log\left|\det\left(\frac{dh_i}{dh_{i-1}}\right)\right|\right)$$

Reversible ViT

Mangalam_Reversible_Vision_Transformers_CVPR_2022_paper



(a) **Rev-ViT Block**

Idea

$$x = g_\theta(z);\ z = g_\theta^{-1}(x)$$

$$y = h_\theta(z);\ \ z = h_\theta^{-1}(y)$$

$$x = g_\theta\left(h_\theta^{-1}(y)\right) = f_\theta(y)$$

$$x \overset{g}{\longleftrightarrow} z \overset{h}{\longleftrightarrow} y$$

Image to image:

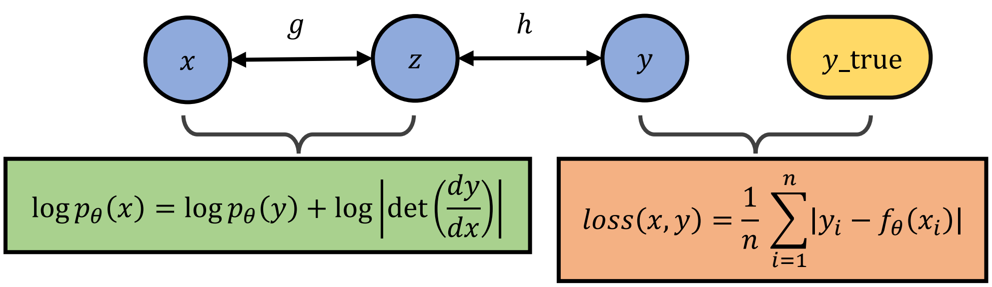
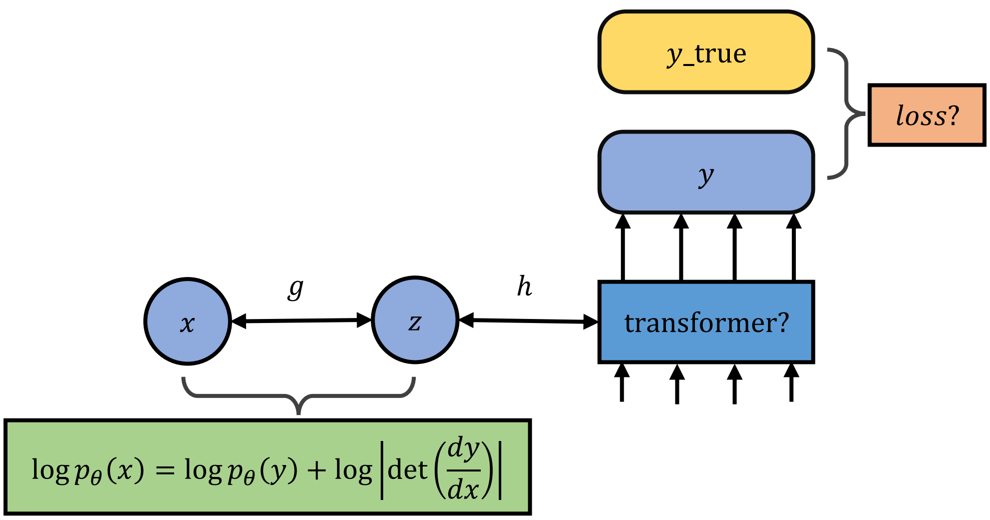
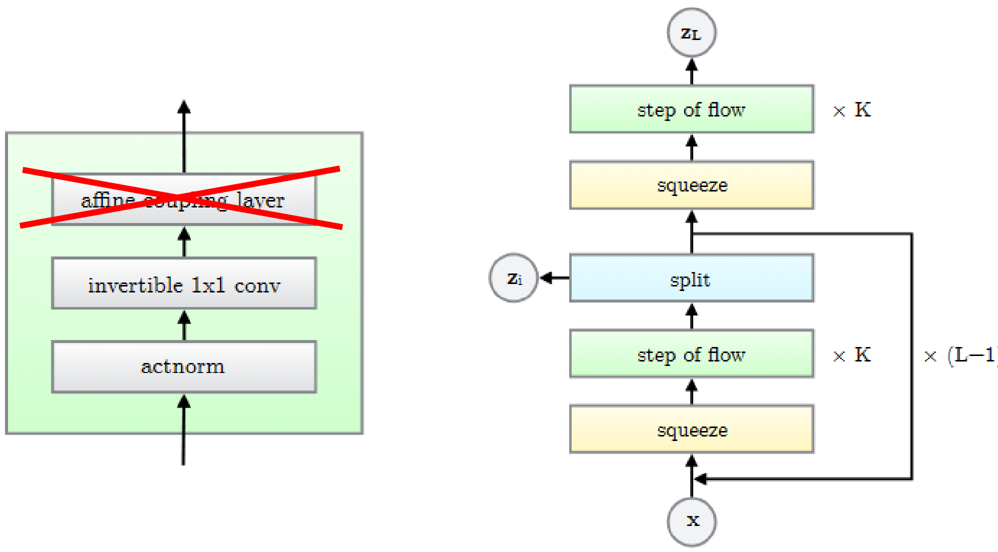


Image to audio:



Structure change:

Rev ViT Block



(a) One step of our flow. (b) Multi-scale architecture (Dinh et al., 2016).

Problem:

Description	Function	Reverse Function	Log-determinant
Actnorm. See Section 3.1.	$\forall i,j : \mathbf{y}_{i,j} = \mathbf{s} \odot \mathbf{x}_{i,j} + \mathbf{b}$	$\forall i,j : \mathbf{x}_{i,j} = (\mathbf{y}_{i,j} - \mathbf{b})/\mathbf{s}$	$h \cdot w \cdot \mathtt{sum}(\log \lvert\mathbf{s}\rvert)$
Invertible 1×1 convolution. $\mathbf{W} : [c \times c]$. See Section 3.2.	$\forall i,j : \mathbf{y}_{i,j} = \mathbf{W}\mathbf{x}_{i,j}$	$\forall i,j : \mathbf{x}_{i,j} = \mathbf{W}^{-1}\mathbf{y}_{i,j}$	$h \cdot w \cdot \log \lvert\det(\mathbf{W})\rvert$ or $h \cdot w \cdot \mathtt{sum}(\log \lvert\mathbf{s}\rvert)$ (see eq. (10))
Affine coupling layer. See Section 3.3 and (Dinh et al., 2014)	$\mathbf{x}_a, \mathbf{x}_b = \mathtt{split}(\mathbf{x})$ $(\log \mathbf{s}, \mathbf{t}) = \mathtt{NN}(\mathbf{x}_b)$ $\mathbf{s} = \exp(\log \mathbf{s})$ $\mathbf{y}_a = \mathbf{s} \odot \mathbf{x}_a + \mathbf{t}$ $\mathbf{y}_b = \mathbf{x}_b$ $\mathbf{y} = \mathtt{concat}(\mathbf{y}_a, \mathbf{y}_b)$	$\mathbf{y}_a, \mathbf{y}_b = \mathtt{split}(\mathbf{y})$ $(\log \mathbf{s}, \mathbf{t}) = \mathtt{NN}(\mathbf{y}_b)$ $\mathbf{s} = \exp(\log \mathbf{s})$ $\mathbf{x}_a = (\mathbf{y}_a - \mathbf{t})/\mathbf{s}$ $\mathbf{x}_b = \mathbf{y}_b$ $\mathbf{x} = \mathtt{concat}(\mathbf{x}_a, \mathbf{x}_b)$	$\mathtt{sum}(\log(\lvert\mathbf{s}\rvert))$ No way to calculate