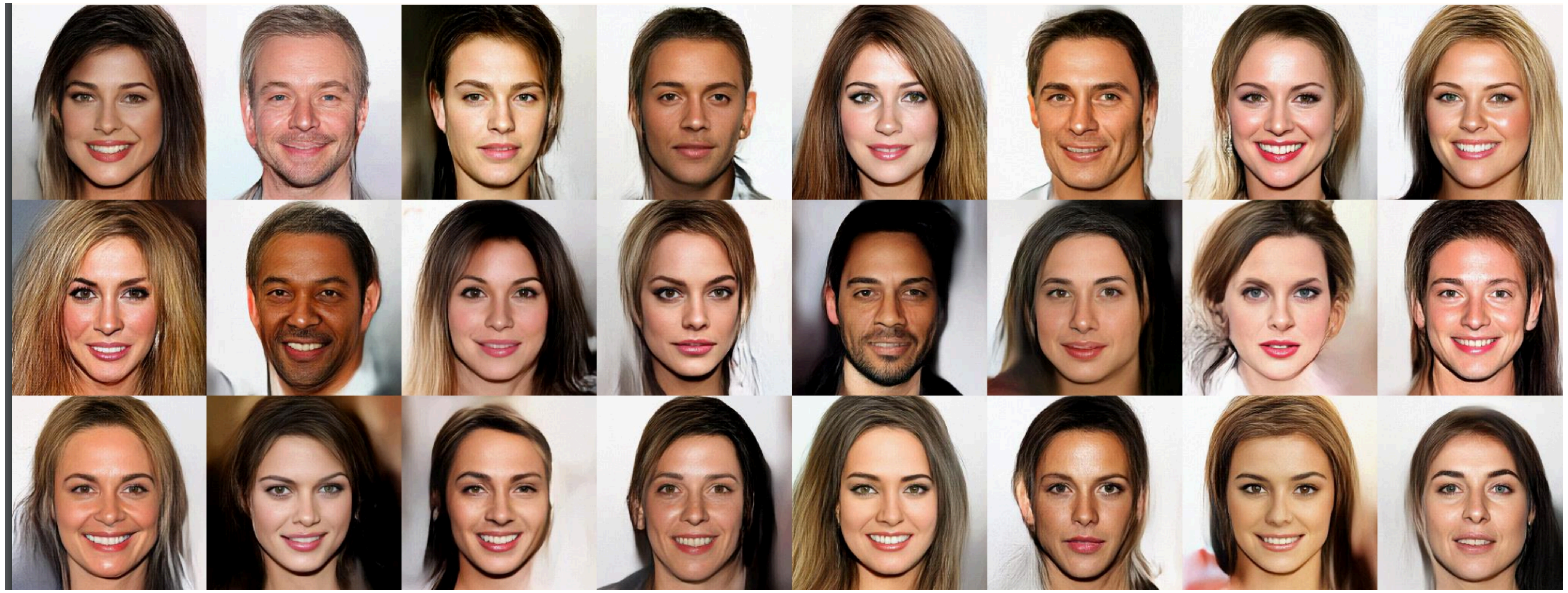
Flow model

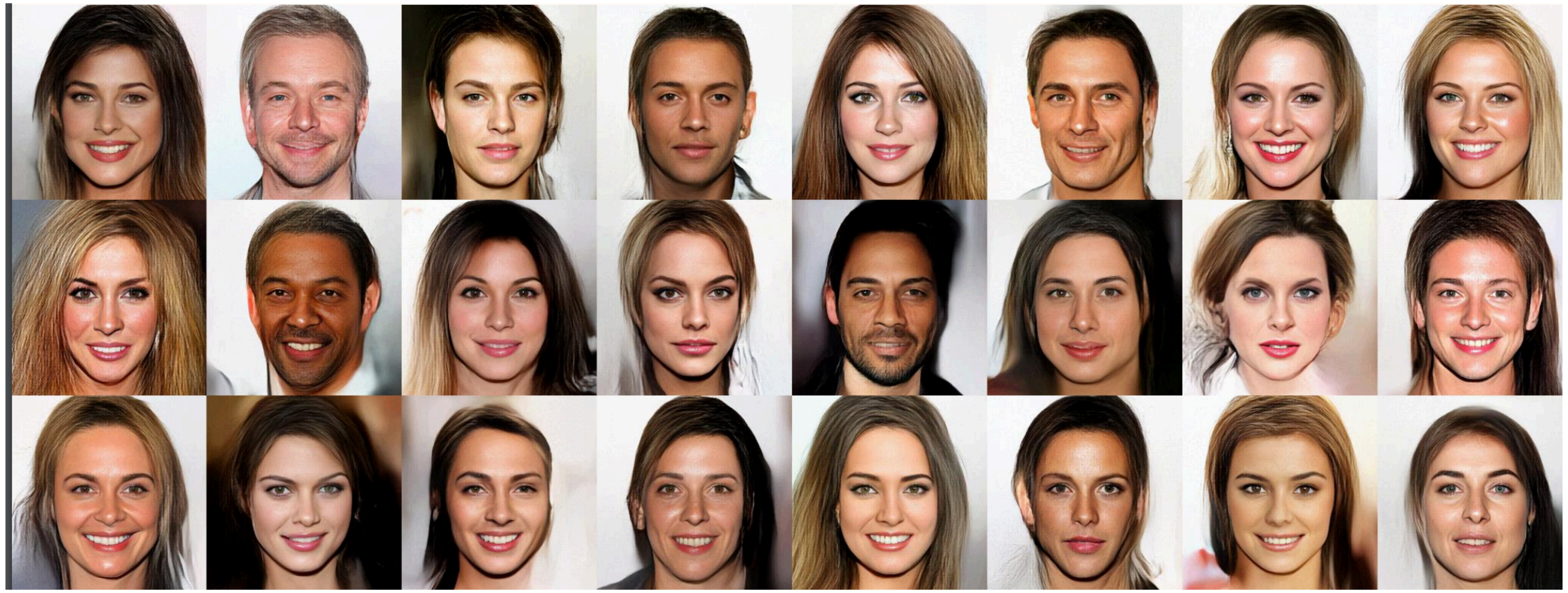
Probabilistic generative models

- p_θ가 데이터의 확률 분포를 학습
- $p_\theta \approx p_{true}$를 통해 $X \sim p_\theta$ 샘플링
- Flow, VAE, GAN



Probabilistic generative models

- p_θ가 데이터의 확률 분포를 학습
- $p_\theta \approx p_{true}$를 통해 $X \sim p_\theta$ 샘플링
- Flow, VAE, GAN

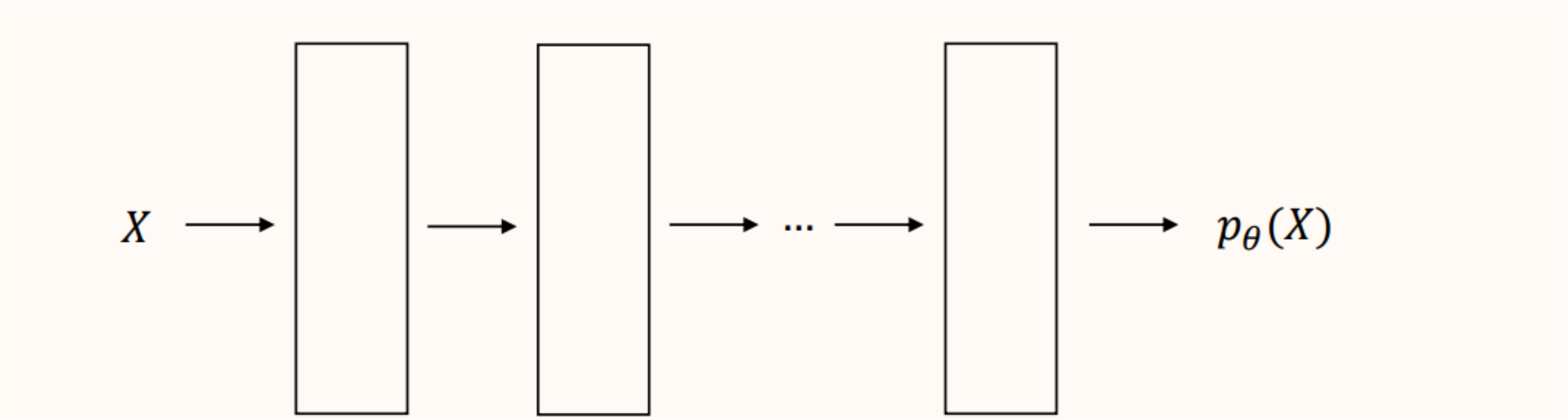


How to learn

$$\underset{\theta \in \mathbb{R}^p}{\text{maximize}} \quad \sum_{i=1}^{N} \log p_\theta\left(X_i\right)$$

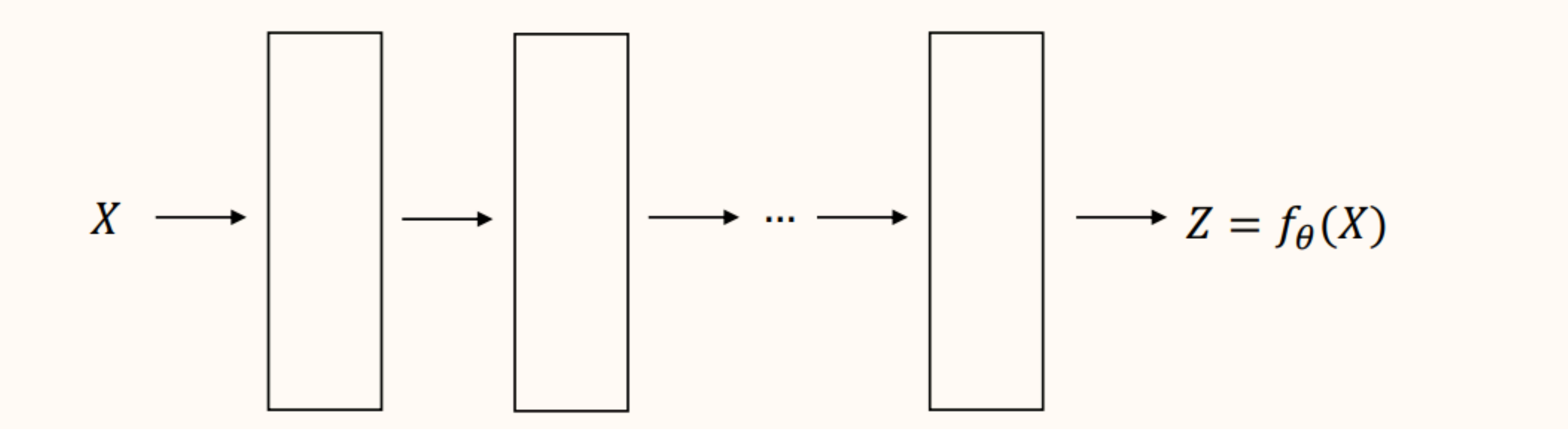
- Maximum Likelihood Estimation을 통해 많이 나온 데이터의 pdf 값을 크게 만듦

Normalization problem



- 문제: $\int_{-\infty}^{\infty} p_\theta(x)\, dx = 1$ 에 맞게 정규화하기 어려움
- 그렇다고 정규화하지 않으면 $p_\theta = \infty$ 라는 자명한 해가 존재

Parameterize Z = $f_\theta(X)$ with DNN



p_θ 자체를 학습시키는 게 아니라,
X를 잘 알려진 분포 Z를 따르는 분포로
변환해주는 함수 f_θ를 학습

Maximum Likelihood Estimation

$$p_X(x) = p_Y(y)\left|\frac{\partial y}{\partial x}\right|$$

$$p_\theta(x) = p_Z(f_\theta(x))\left|\frac{\partial f_\theta(x)}{\partial x}\right|$$

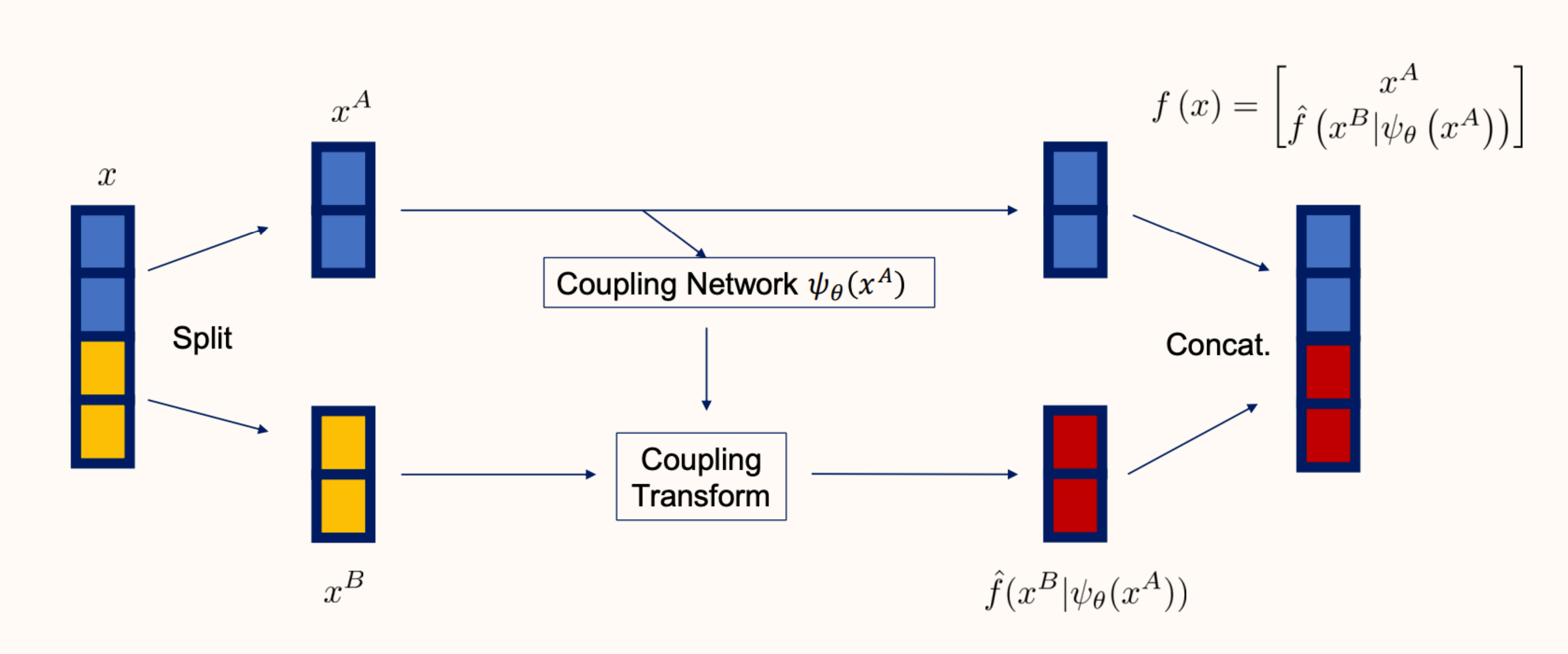
$$\underset{\theta \in \mathbb{R}^p}{\text{maximize}} \quad \sum_{i=1}^{N} \log p_\theta\left(X_i\right) = \underset{\theta \in \mathbb{R}^p}{\text{maximize}} \quad \sum_{i=1}^{N} \log p_Z\left(f_\theta(X_i)\right) + \boxed{\log\left|\frac{\partial f_\theta}{\partial x}(X_i)\right|}$$

다차원
야코비안 행렬

Coupling Flow

$$x = (x_A, x_B)$$

$$f(x) = \left(x^A, \hat{f}(x^B|\psi_\theta(x^A))\right)$$



Coupling Flow Determinant

$$f(x) = \left(x^A, \hat{f}(x^B|\psi_\theta(x^A))\right)$$

$$x = (x_A, x_B)$$

$$\frac{\partial f_\theta}{\partial x}(x) = \begin{bmatrix} I & 0 \\ \cancel{\frac{\partial \hat{f}}{\partial x^A}(x^B|\psi_\theta(x^A))} & \frac{\partial \hat{f}}{\partial x^B}(x^B|\psi_\theta(x^A)) \end{bmatrix}$$

$$\det\left(\frac{\partial f_\theta}{\partial x}(x)\right) = \det\left(\boxed{\frac{\partial \hat{f}}{\partial x^B}(x^B|\psi_\theta(x^A))}\right)$$

NICE Flow

$$\hat{f}(x_B \mid \psi_\theta(x_A)) = x_B + t_\theta(x_A)$$

$$\det \frac{\partial f_\theta}{\partial x}(x) = \det \begin{bmatrix} I & 0 \\ \frac{\partial \hat{f}}{\partial x^A}\left(x^B | \psi_\theta\left(x^A\right)\right) & \frac{\partial \hat{f}}{\partial x^B}\left(x^B | \psi_\theta(x^A)\right) \end{bmatrix} = \det \begin{bmatrix} I & 0 \\ \frac{\partial \hat{f}}{\partial x^A}\left(x^B | \psi_\theta\left(x^A\right)\right) & I \end{bmatrix} = 1$$

RealNVP Flow

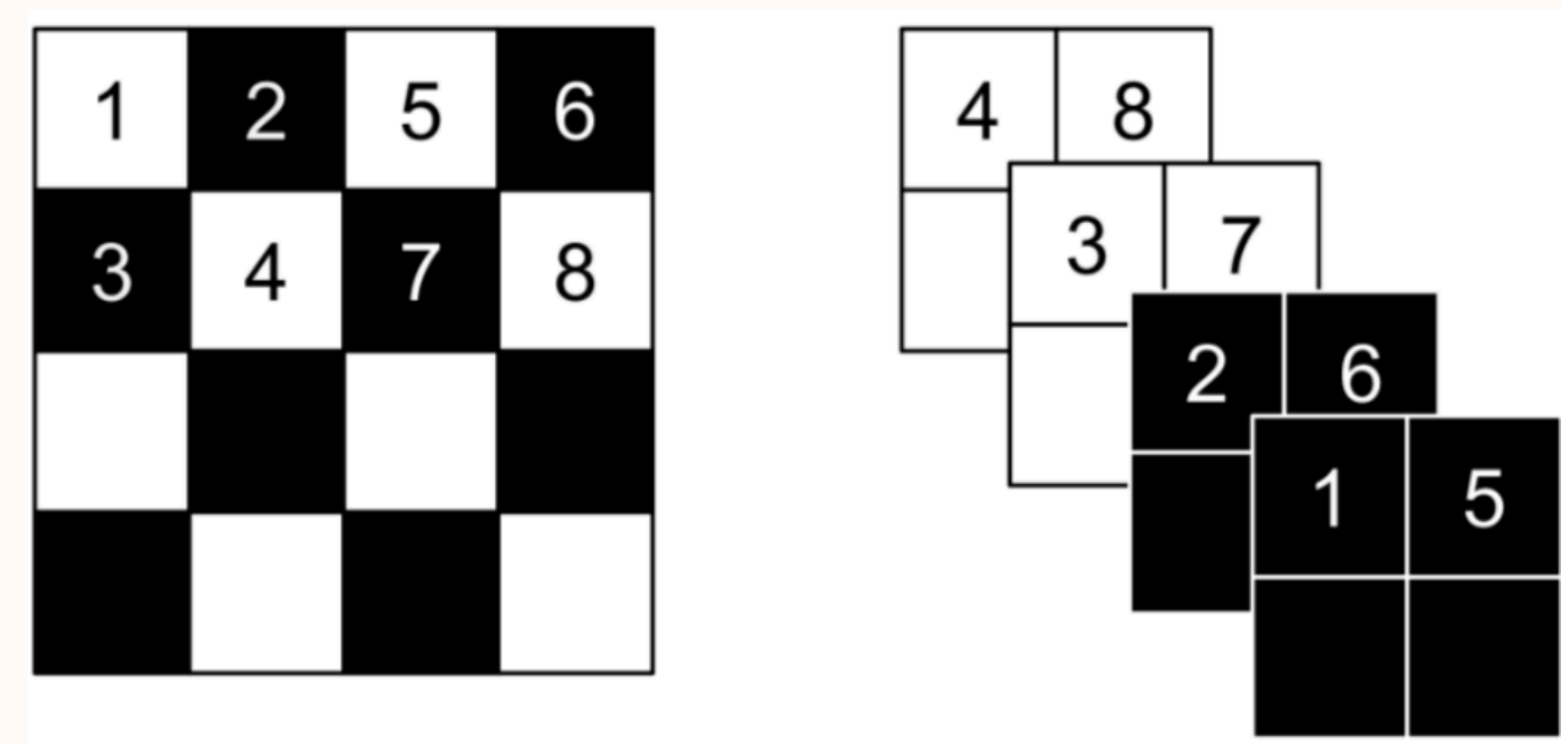
$$\hat{f}(x_B \mid \psi_\theta(x_A)) = e^{s_\theta(x_A)} \odot x_B + t_\theta(x_A)$$

$$\begin{aligned} \det \frac{\partial f_\theta}{\partial x}(x) &= \det \begin{bmatrix} I & 0 \\ \frac{\partial \hat{f}}{\partial x^A}\left(x^B | \psi_\theta\left(x^A\right)\right) & \frac{\partial \hat{f}}{\partial x^B}\left(x^B | \psi_\theta(x^A)\right) \end{bmatrix} \\ &= \det \begin{bmatrix} I & 0 \\ \frac{\partial \hat{f}}{\partial x^A}\left(x^B | \psi_\theta\left(x^A\right)\right) & \operatorname{diag}(e^{s_\theta(x_{1:n/2})}) \end{bmatrix} = \exp(\mathbf{1}_{n/2}^\intercal s_\theta(x_{1:n/2})) \end{aligned}$$

Partition of variable

$$f(x) = \left(x^A, \hat{f}(x^B|\psi_\theta(x^A))\right)$$

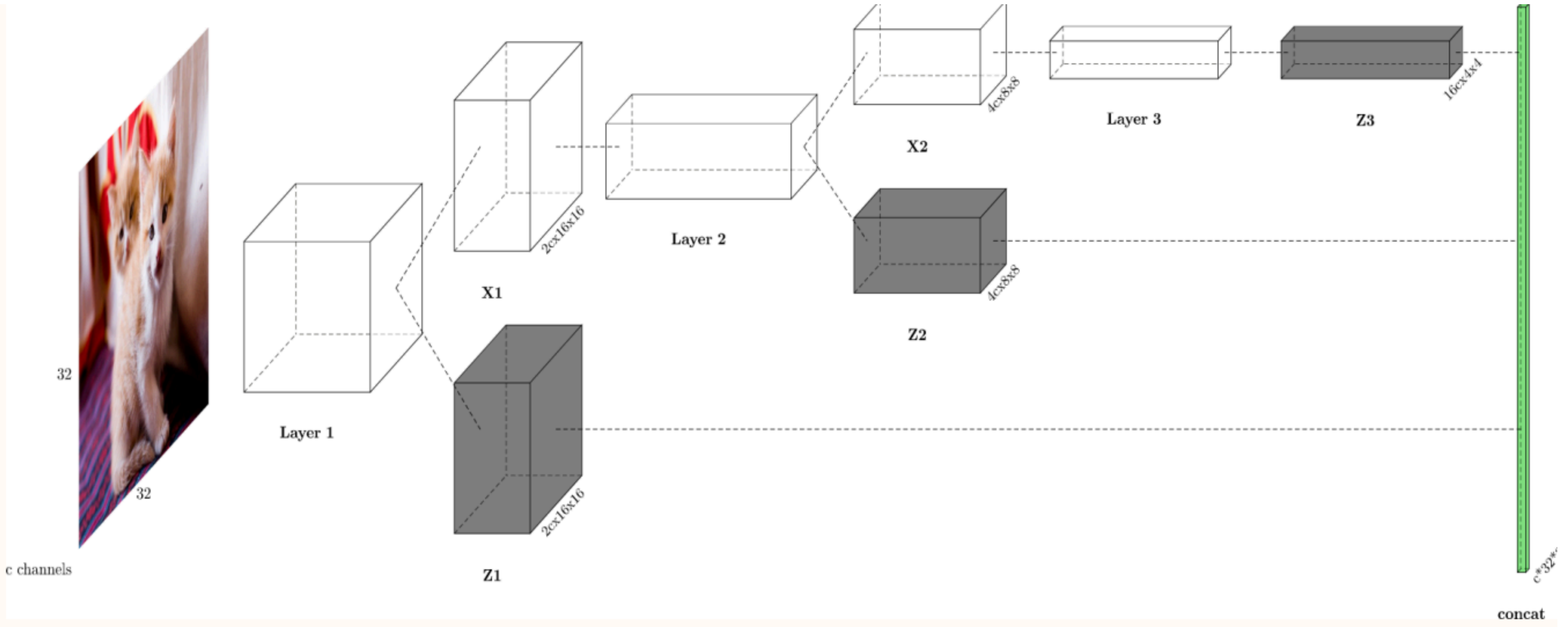
문제: x_A, x_B를 계속 순서를 기준으로 나누면 x_A가 꾸준히 바뀌지 않고 유지되게 됨



Checkerboard　　Channel-wise

RealNVP Architecture

Real NVP Architecture



Input X: $c \times 32 \times 32$ image with $c = 3$

Layer 1: Input X: $c \times 32 \times 32$

- Checkerboard $\times\, 3$, channel reshape into $4c \times 16 \times 16$, channel $\times\, 3$
- Output: Split result to get X_1: $2c \times 16 \times 16$ and Z_1: $2c \times 16 \times 16$ (fine-grained latents)

Layer 2: Input X_1: $2c \times 16 \times 16$ from layer 1

- Checkerboard $\times\, 3$, channel reshape into $8c \times 8 \times 8$, channel $\times\, 3$
- Split result to get X_2: $4c \times 8 \times 8$ and Z_2: $4c \times 8 \times 8$ (coarser latents)

Layer 3: Input X_2: $4c \times 8 \times 8$ from layer 2

- Checkerboard $\times\, 3$, channel reshape into $16c \times 4 \times 4$, channel $\times\, 3$
- Get Z_3: $16c \times 4 \times 4$ (latents for highest-level details)