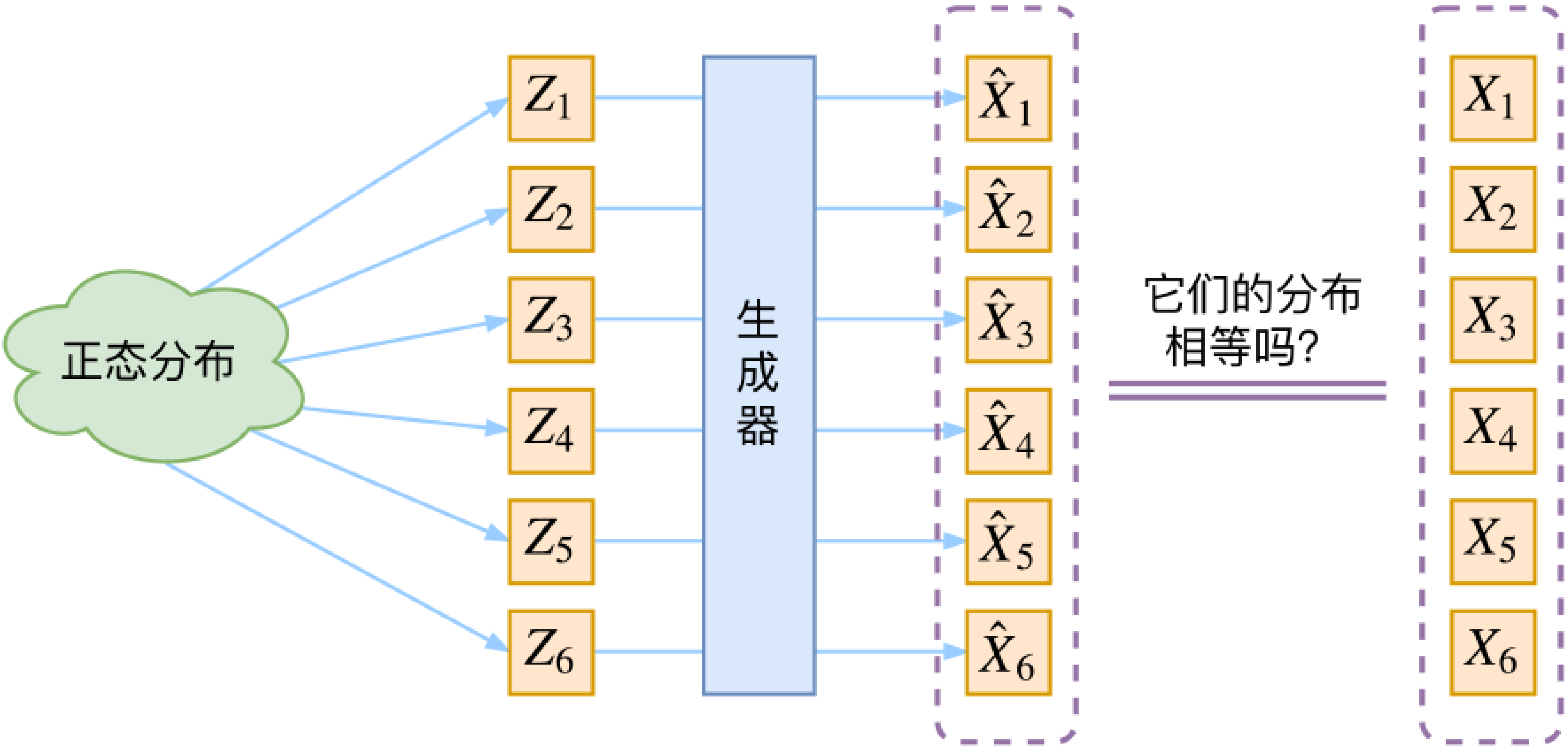
生成模型

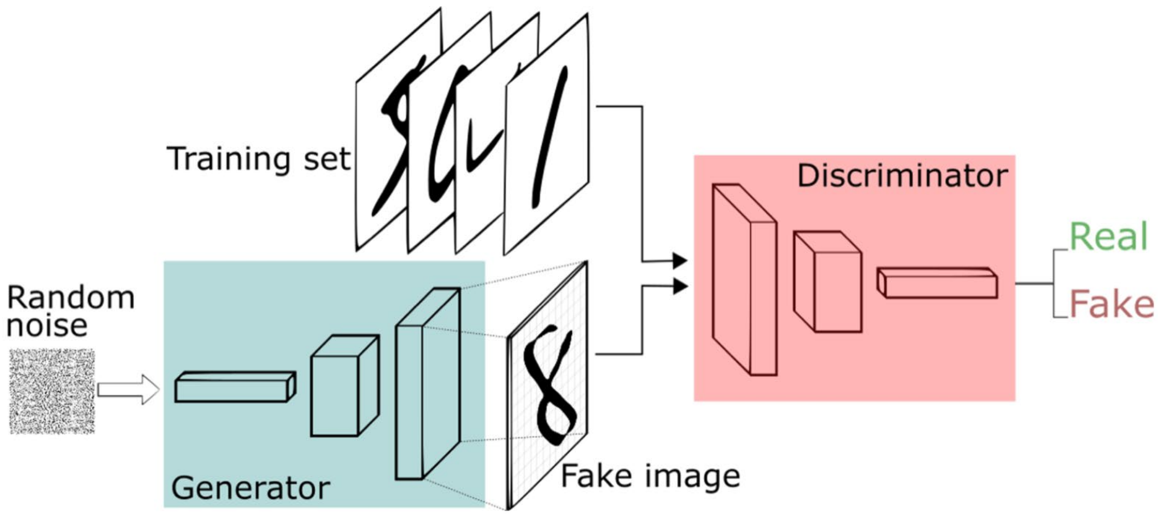
- **目标：** 构建一个从隐变量Z恢复原始数据X的模型$X=g(Z)$，Z服从常见分布（方便采样）。



生成对抗网络

东南大学 SOUTHEAST UNIVERSITY

□ 是什么：



➢ 训练步骤：

- 先固定生成器，将m个噪声（任意分布）输入生成器得到m个假图像，再将m个真图像与m个假图像一同输入判别器**训练**，判别器（二分类模型）**损失函数**为：正确分类真假图像。训练k步（k个minibatch）判别器。
- 再固定判别器，将m个噪声（任意分布）输入生成器进行**训练**，**损失函数**为：让判别器将假图像判为真图像。训练1步生成器。
- 重复以上epoch次。最终**理想情况下**，生成器生成的假图像与真图像在判别器下输出均为50%

GAN伪代码

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k, is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

for number of training iterations **do**

  for k steps **do**

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Sample minibatch of m examples $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.
  - Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)\right].$$

  end for

- Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

GAN优缺点及作用

东南大学 SOUTHEAST UNIVERSITY

- **优：**
  - 不像一般的AE生成模型死板的学习真实图像分布（输入为真实图像），GAN通过对抗训练的思想，只需输入噪声便能产生以假乱真的图像（可产生从未有过的图像）
  - 对复杂分布具有强大的拟合能力，**生成能力强**，一般用作**图像生成**
- **缺：**
  - 基于KL和JS散度的损失函数不易训练。因为当两个概率分布相距很远时，无法产生有意义的梯度。
- **作用：**
  - 广义上：图像生成、风格迁移、黑白老照片上色修复、AI艺术、黑白老照片上色修复。可实现照片转成油画、野马转成斑马、黑夜转成白天，简笔画的猫转成真猫，模糊图像转成高清图像等酷炫好玩的应用。