

对抗生成网络

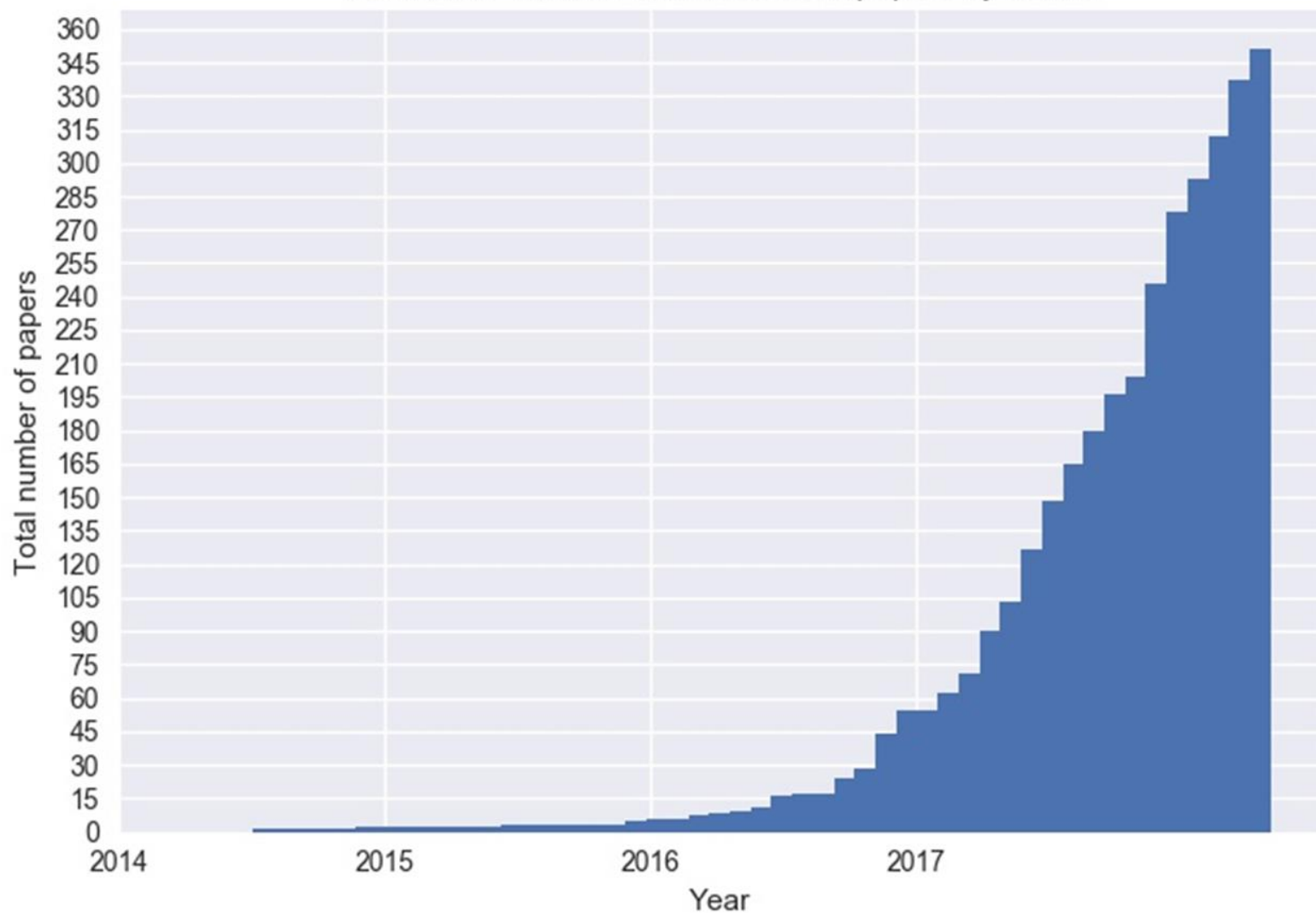
Generative Adversarial Network

七月在线 张雨石

2018年9月16日

<http://blog.csdn.net/stdcoutzyx>

Cumulative number of named GAN papers by month



对抗生成网络

- 对抗生成网络原理
- 深度卷积对抗生成网络 (DCGAN)
- 图像翻译 (Pix2Pix)
- 无配对图像翻译 (CycleGAN)
- 多领域图像翻译 (StarGAN)
- 文本生成图像 (Text2Img)



对抗生成网络原理

☐ Discriminative Problem

- 分类

- 手段

 - ☐ Dropout

 - ☐ Batch Normalization

 - ☐ ResNet

 - ☐



对抗生成网络原理

□ Generative Problem

- 无中生有
- 图像高清化
- 图像修复
- 线描图上色
-



对抗生成网络原理



[Goodfellow et al., 2014]
University of Montreal



[Roth et al., 2017]
Microsoft and ETHZ



[Karras et al., 2018]
NVIDIA



对抗生成网络原理



(a)



(b)



(c)



对抗生成网络原理



对抗生成网络原理

□ 发展状况

- Discriminative Model
- Generative Model

□ 如何能借用Discriminative优势？

- GAN



对抗生成网络原理

□ 对抗哲学

- 道高一尺魔高一丈

□ 警察和小偷

- 小偷不停的更新偷盗技术以避免被抓
- 警察不停的发现新的方法是识别小偷



对抗生成网络原理

□ 对抗哲学

- 逐步递进

□ 郭靖和他的对手们

- 黄河四鬼

- 沙通天

- 欧阳克

- 梅超风

- 西毒



对抗生成网络原理

□ 生成器和判别器

- 交替训练——道高一尺魔高一丈
- G的目标：生成一个仿真数据
- D的目标：判别一个样本是否真实
- 最终目标：得到一个好的G



对抗生成网络原理

☐ 生成器和判别器

■ 交替训练——道高一尺魔高一丈

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

■ 训练D

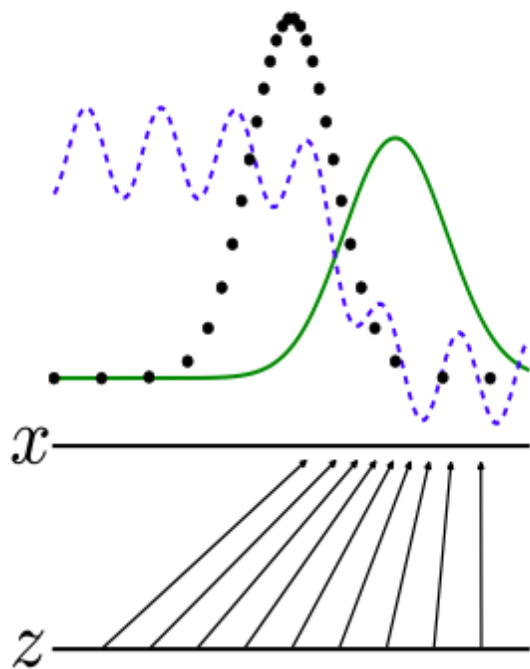
☐ 真为真，假为假

■ 训练G

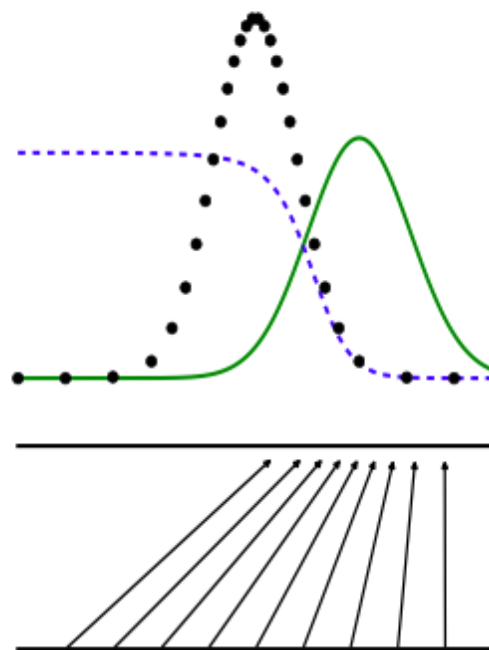
☐ 假为真



对抗生成网络原理



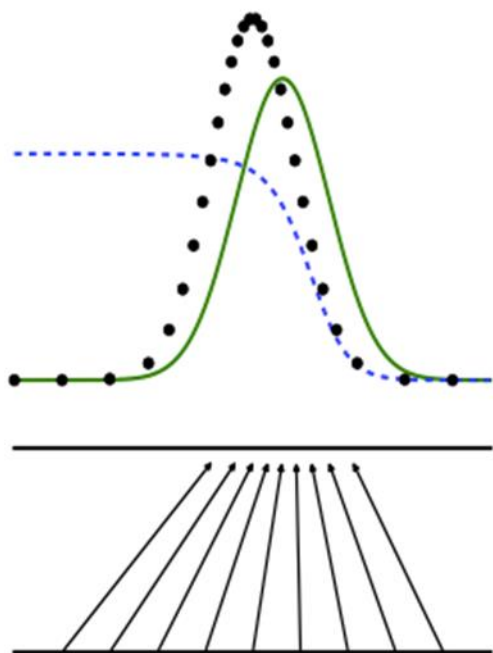
(a)



(b)

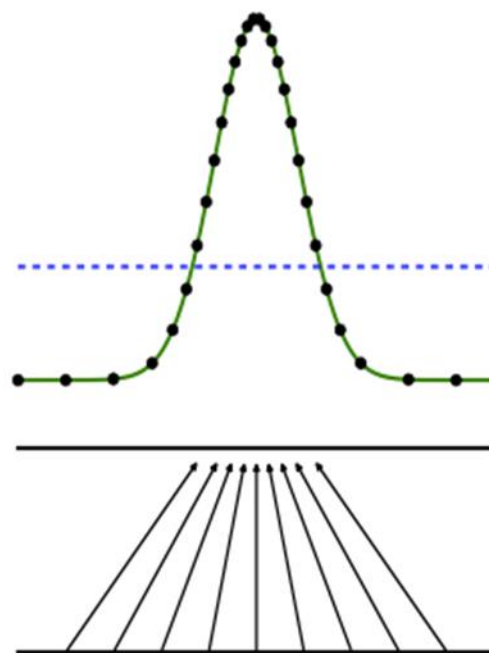


对抗生成网络原理



(c)

...



(d)



深度卷积对抗生成网络(DCGAN)

□ 如何构建一个新的GAN网络

- 定义目标
- 定义G的输入输出
- 定义D的输入输出
- 定义G和D的结构



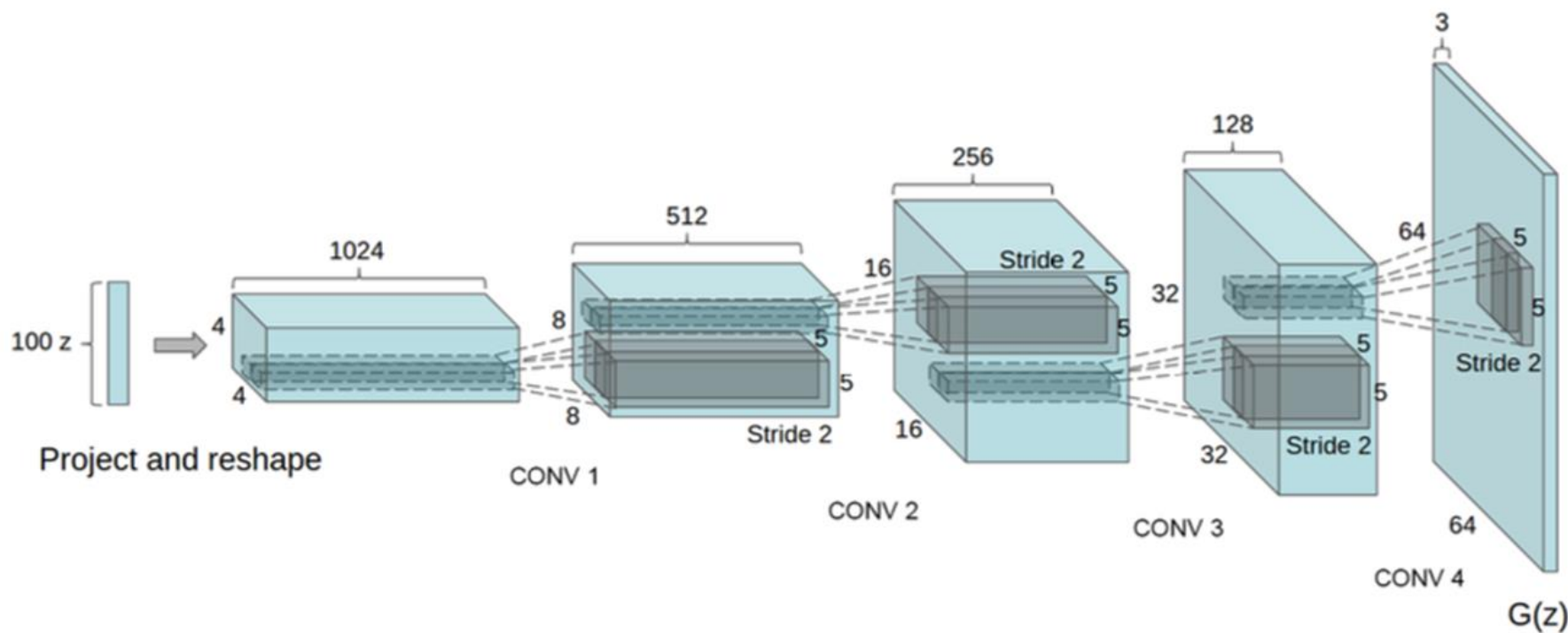
深度卷积对抗生成网络(DCGAN)

□ 目标

- 从随机向量生成真实图像

深度卷积对抗生成网络(DCGAN)

□ 生成器G



深度卷积对抗生成网络(DCGAN)

☐ 生成器G

- 输入：随机向量

- 输出：图像

- 反卷积

 - ☐ Deconvolution

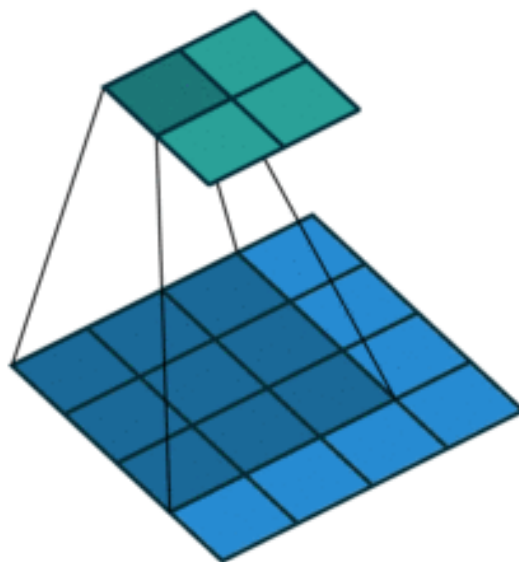
 - ☐ Transposed Convolution

 - ☐ Fractional strided convolution



反卷积

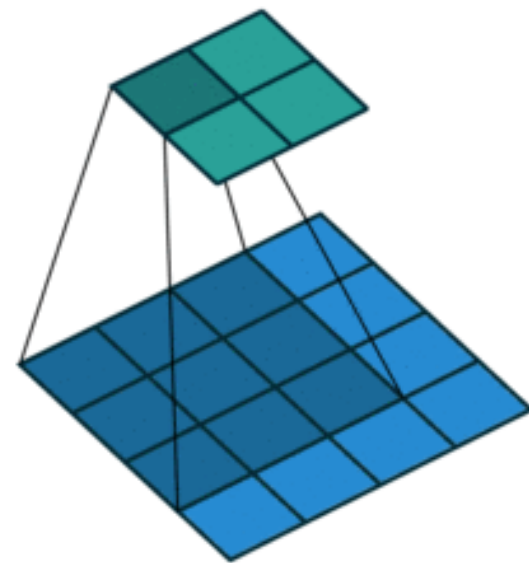
□ 卷积操作的展开



反卷积

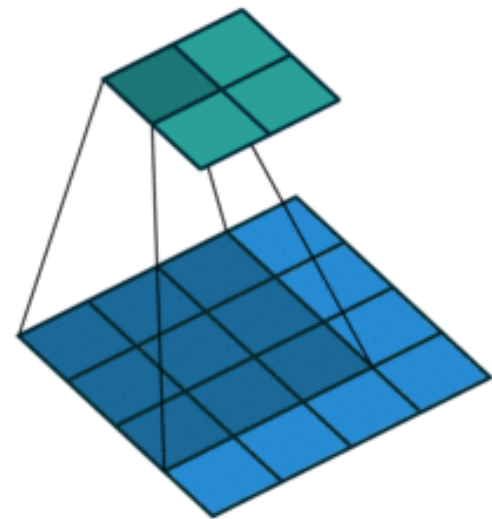
A00	A01	A02	A03
A10	A11	A12	A13
A20	A21	A22	A23
A30	A31	A32	A33

w0,0	w0,1	w0,2
w1,0	w1,1	w1,2
w2,0	w2,1	w2,2



$$B_{00} = a_{00} * w_{0,0} + a_{01} * w_{0,1} + a_{02} * w_{0,2} + a_{10} * w_{1,0} \\ + a_{11} * w_{1,1} + a_{12} * w_{1,2} + a_{20} * w_{2,0} + a_{21} * w_{2,1} + a_{22} * w_{2,2}$$

反卷积



$$B_{00} = a_{00} * w_{0,0} + a_{01} * w_{0,1} + a_{02} * w_{0,2} + a_{10} * w_{1,0} + a_{11} * w_{1,1} + a_{12} * w_{1,2} + a_{20} * w_{2,0} + a_{21} * w_{2,1} + a_{22} * w_{2,2}$$

[a₀₀, a₀₁, a₀₂, a₀₃, a₁₀, a₁₁, a₁₂, a₁₃, a₂₀, a₂₁, a₂₂, a₂₃, a₃₀, a₃₁, a₃₂, a₃₃]

$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix}$$



反卷积

□ $Y = CX \rightarrow X$ 是 (16,1), C 是 (4,16)

□ $Y = C^T X \rightarrow x$ 是 (4,1), C^T 是 (16,4)

[b00, b01, b10, b11]

$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix}$$

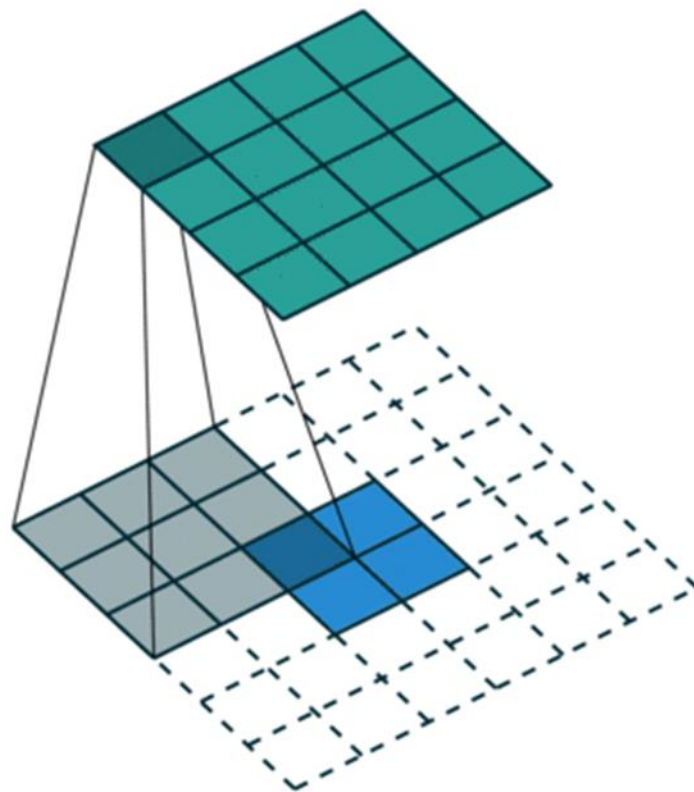


反卷积

B00	B01
B10	b11

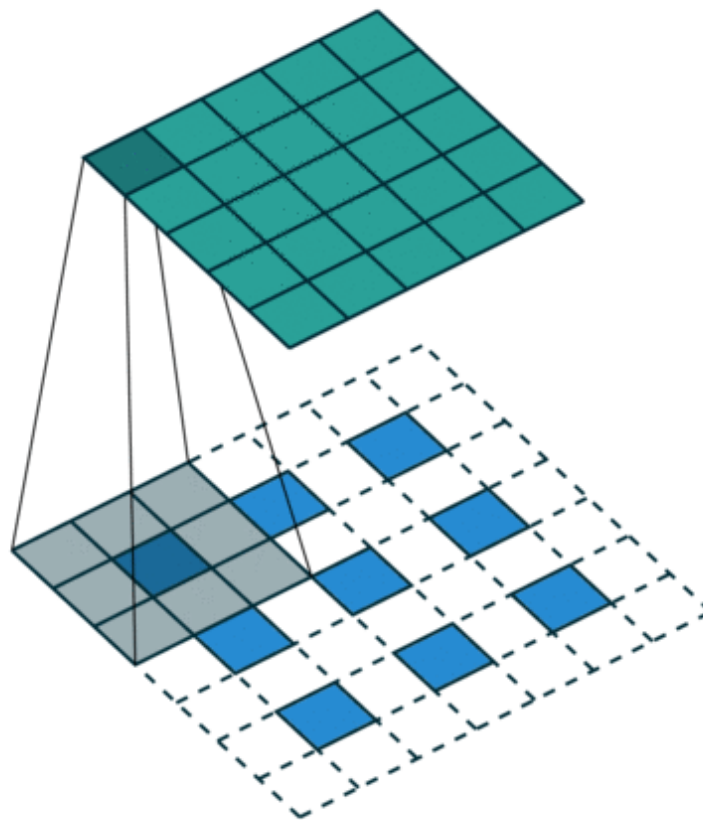
w0,0	w0,1	w0,2
w1,0	w1,1	w1,2
w2,0	w2,1	w2,2

W2,2	0	0	0
W2,1	W2,2	0	0
W2,0	W2,1	0	0
0	W2,0	0	0
...



反卷积

□ Fractional strided



深度卷积对抗生成网络(DCGAN)

□ 模型结构

- Pooling层用convolutional层替代
- D上使用strided convolutions
- G上使用fractional-strided convolutions



深度卷积对抗生成网络(DCGAN)

☐ 模型结构

■ G和D都使用batch normalization

- ☐ 帮助解决初始化差的问题

- ☐ 梯度传播到每一层

- ☐ BN不应用于输入层和输出层

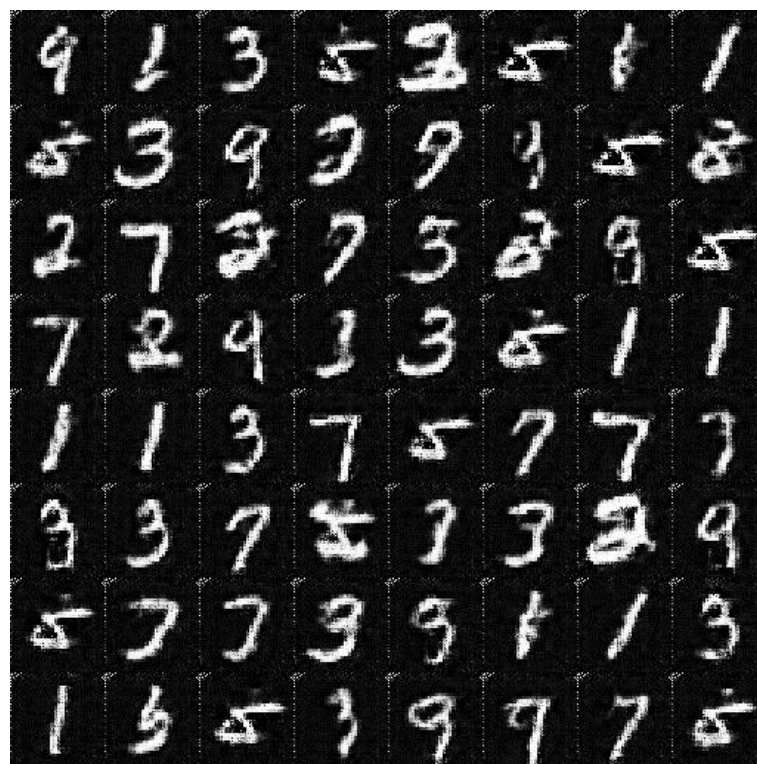
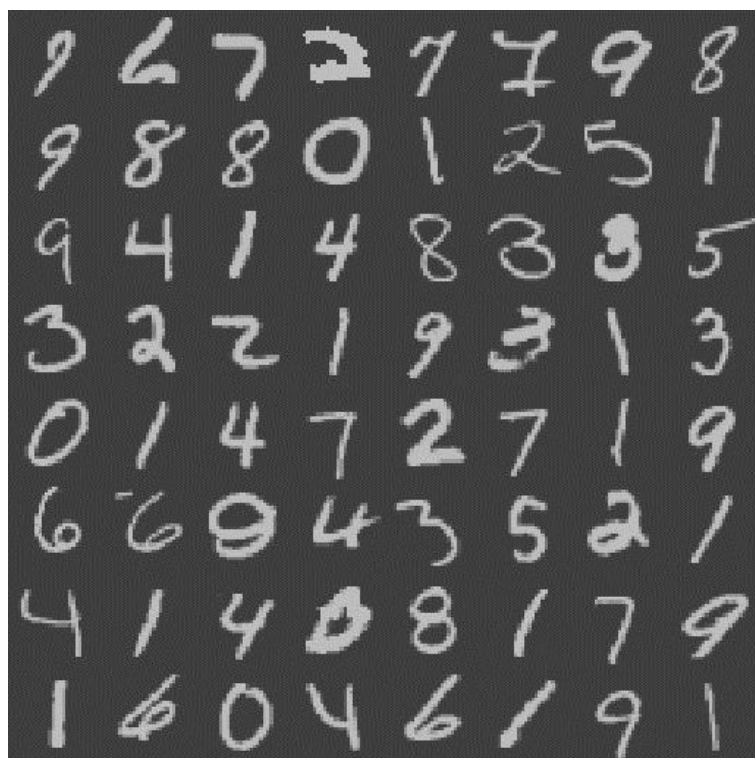
☐ 模型结构

- G上除了输出层使用tanh外其他都用Relu

- D上使用LeakyReLU



深度卷积对抗生成网络-效果



深度卷积对抗生成网络-效果



深度卷积对抗生成网络-效果

□ 微调向量使某个物品消失或出现



深度卷积对抗生成网络-效果

□ 微调向量使某个物品消失或出现



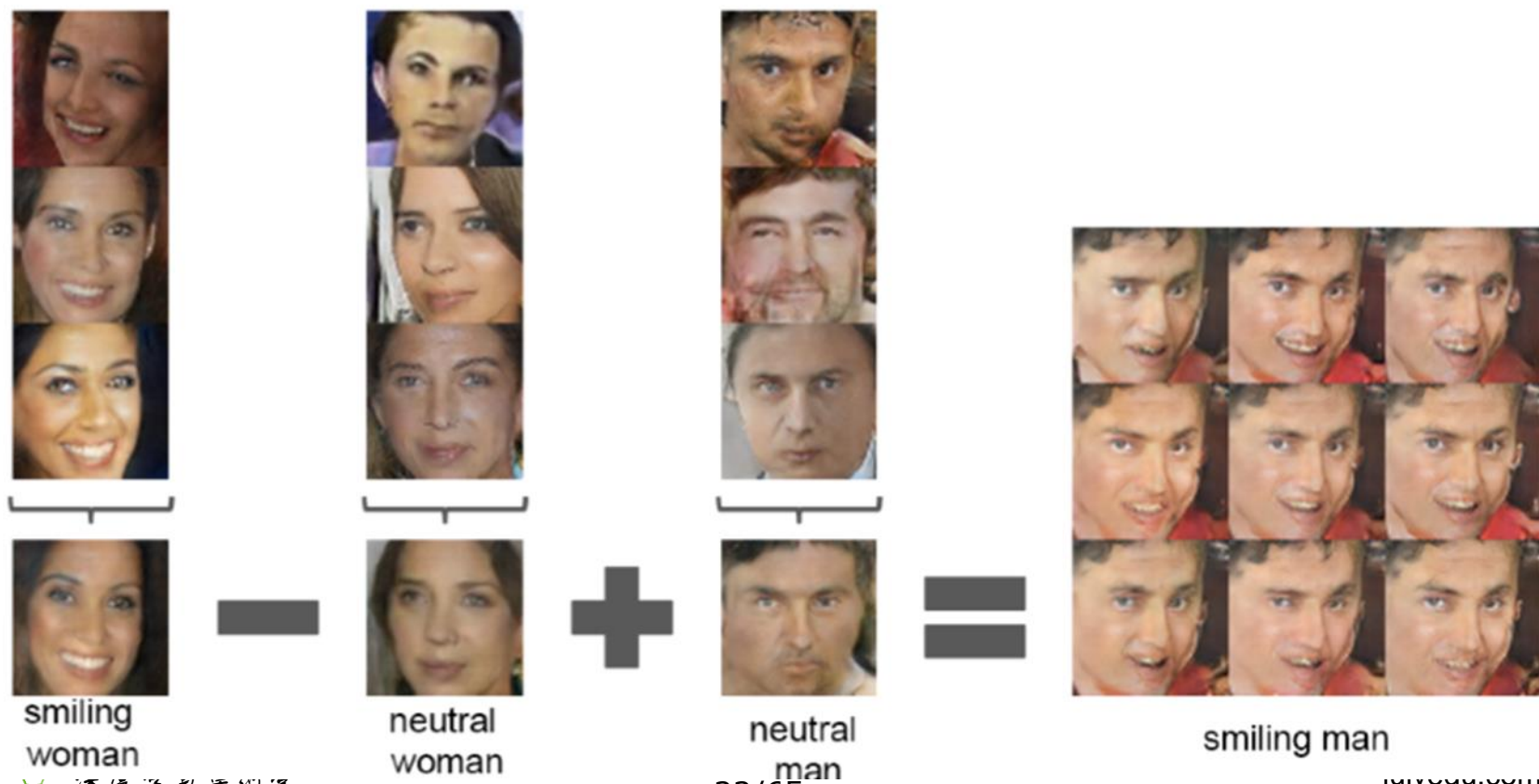
深度卷积对抗生成网络-效果

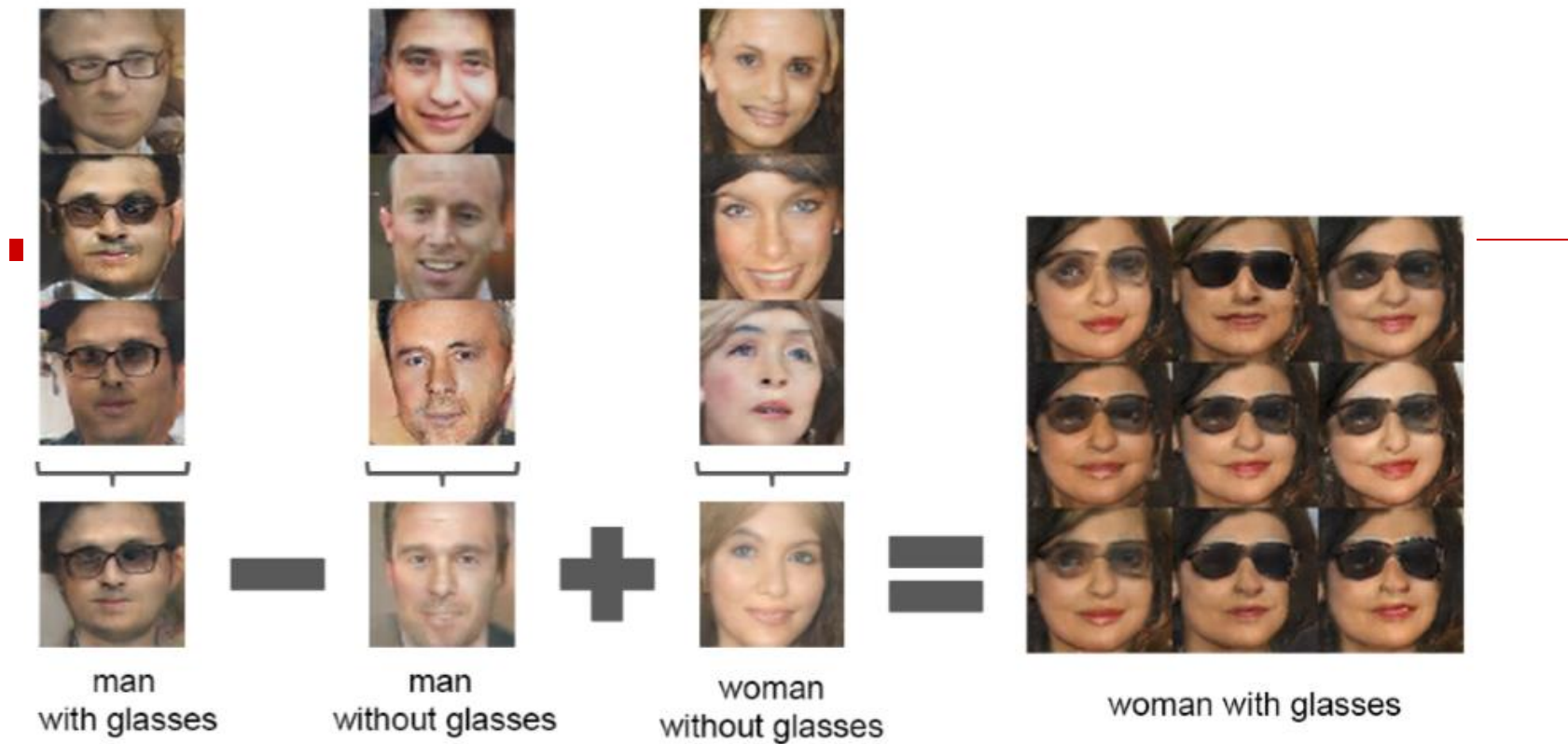
- 使用LR来确认隐空间中哪个和窗户有关
 - 加Mask



深度卷积对抗生成网络-效果

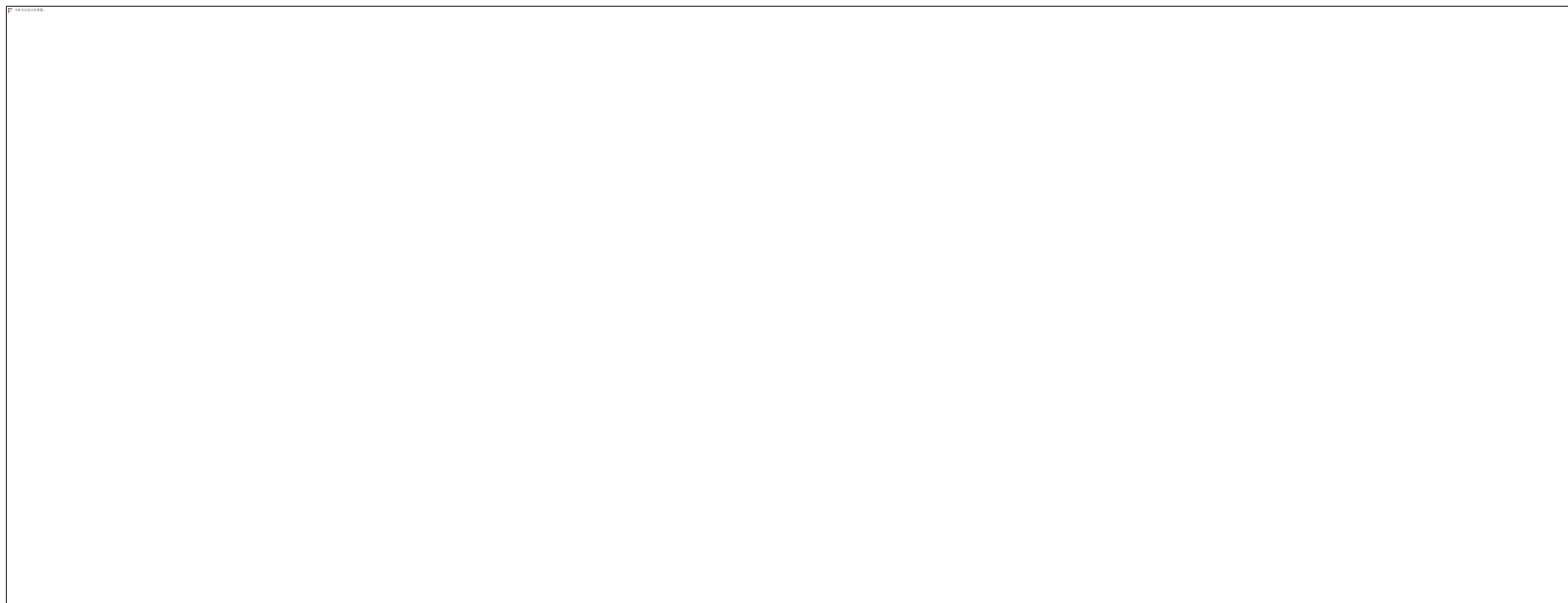
□ 向量加减法



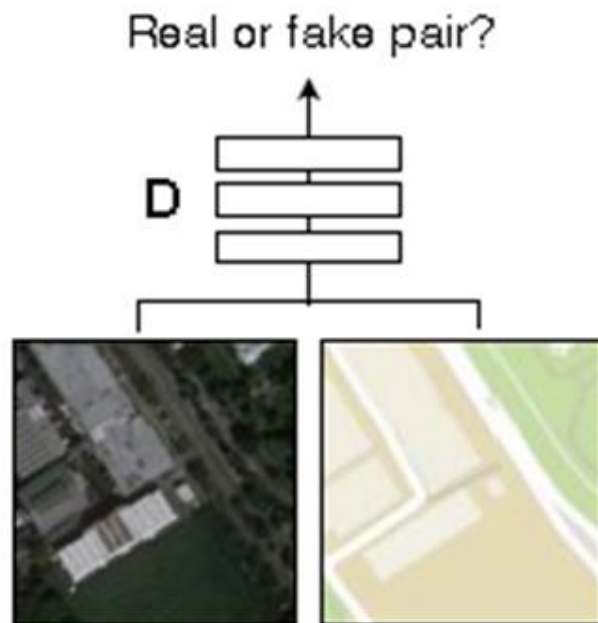


图像翻译(Pix2Pix)

- 目标：以输入图为条件，生成内容相似但风格不同的新图



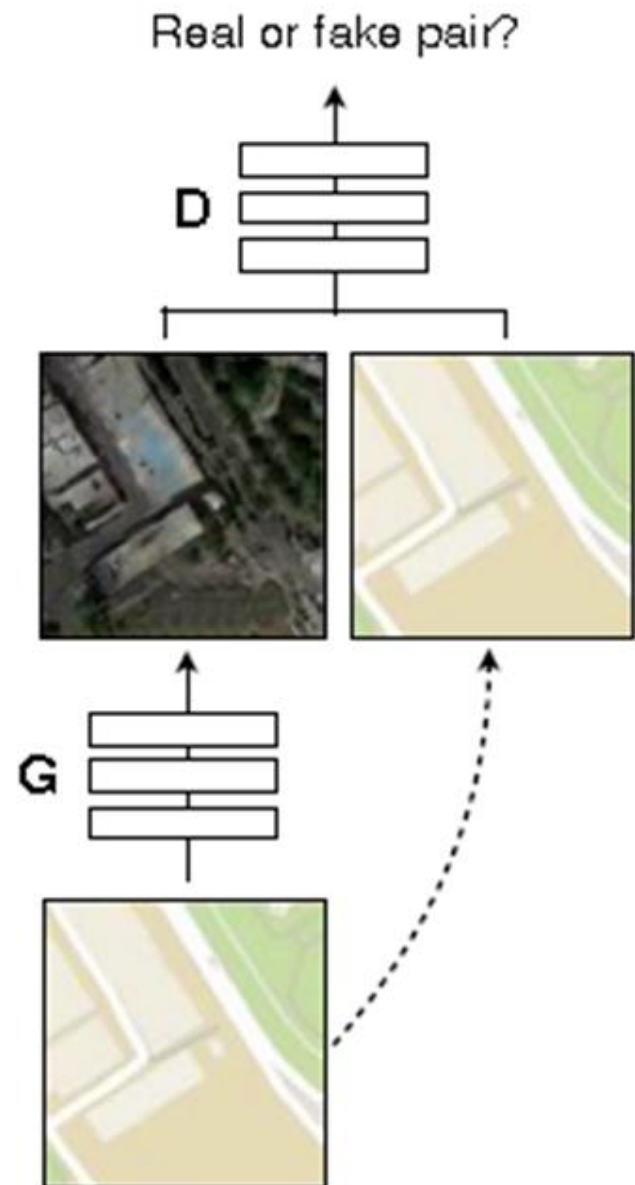
Positive examples



G tries to synthesize fake images that fool **D**

D tries to identify the fakes

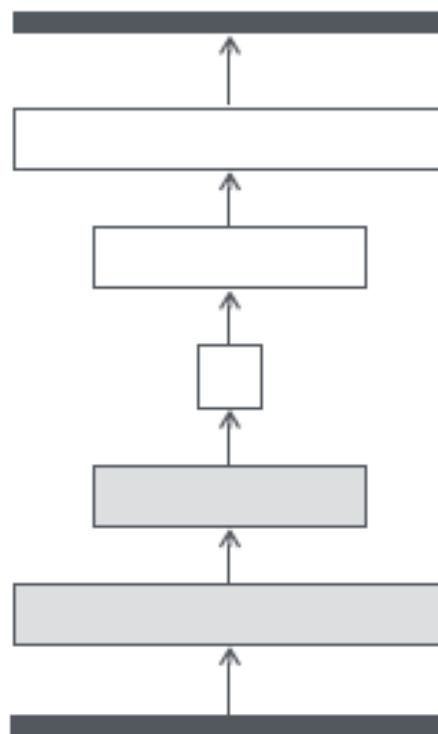
Negative examples



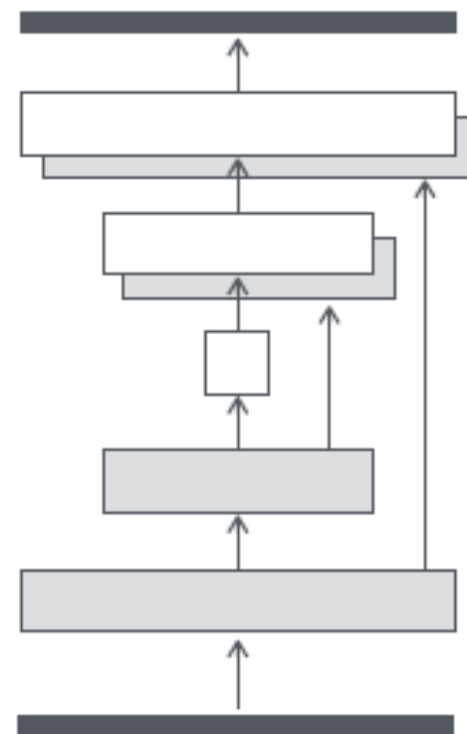
图像翻译(Pix2Pix)

□ 模型结构-G

■ U-Net



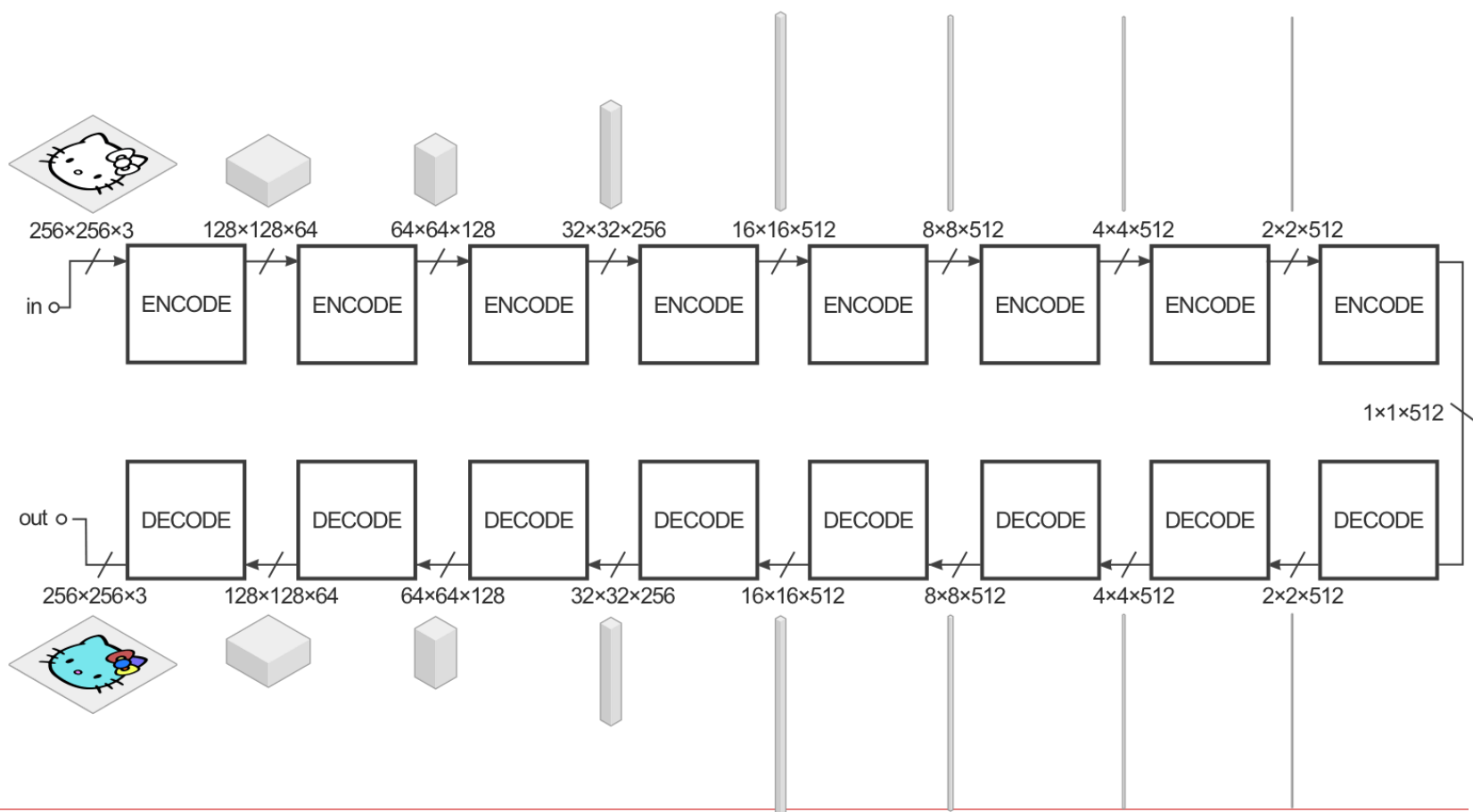
Encoder-decoder



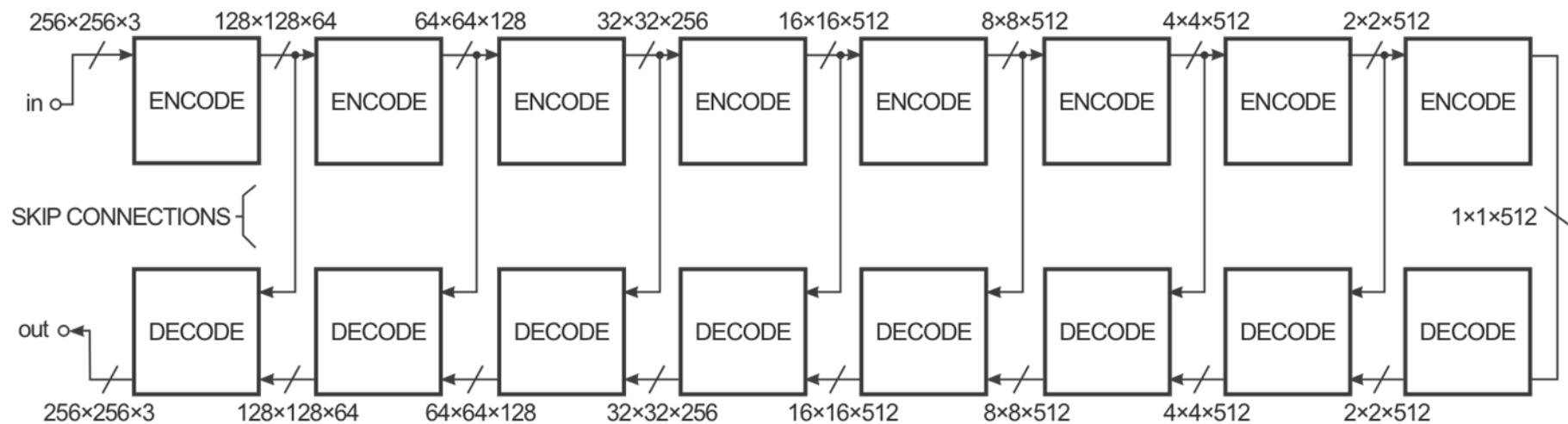
U-Net



图像翻译(Pix2Pix)



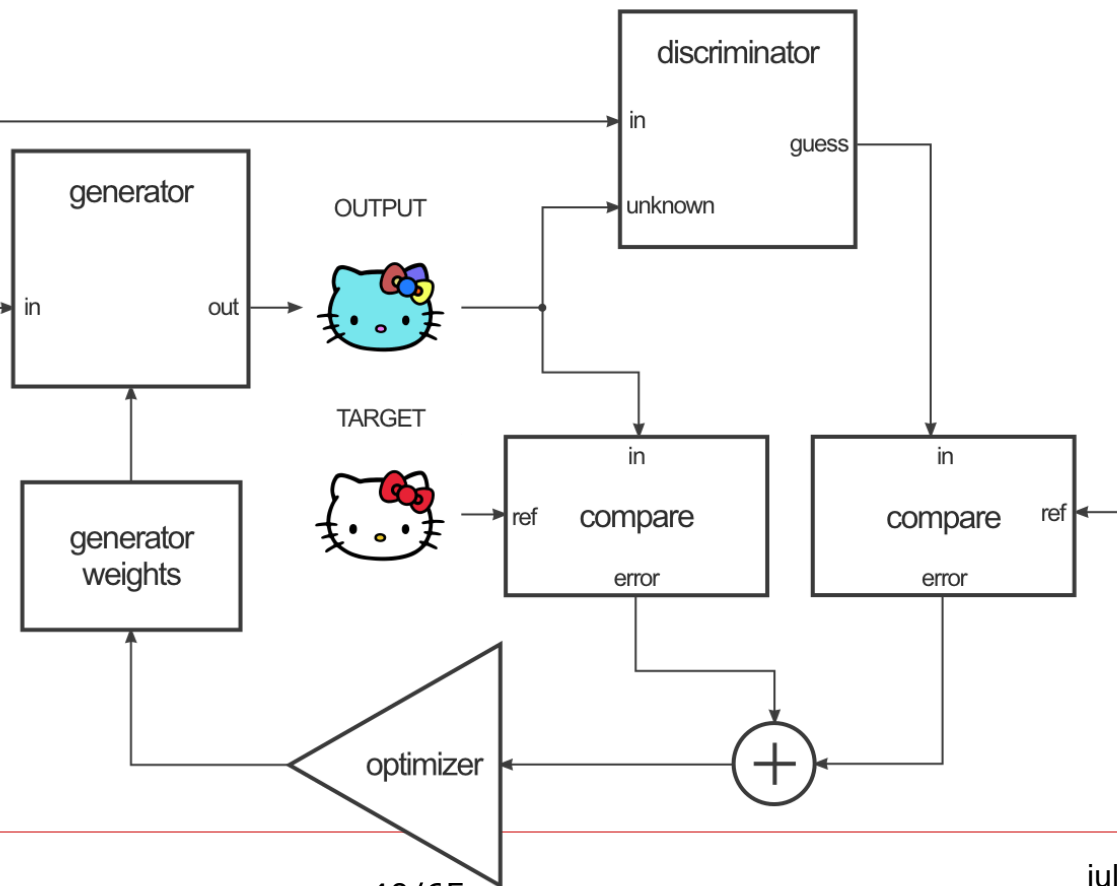
图像翻译(Pix2Pix)



图像翻译(Pix2Pix)

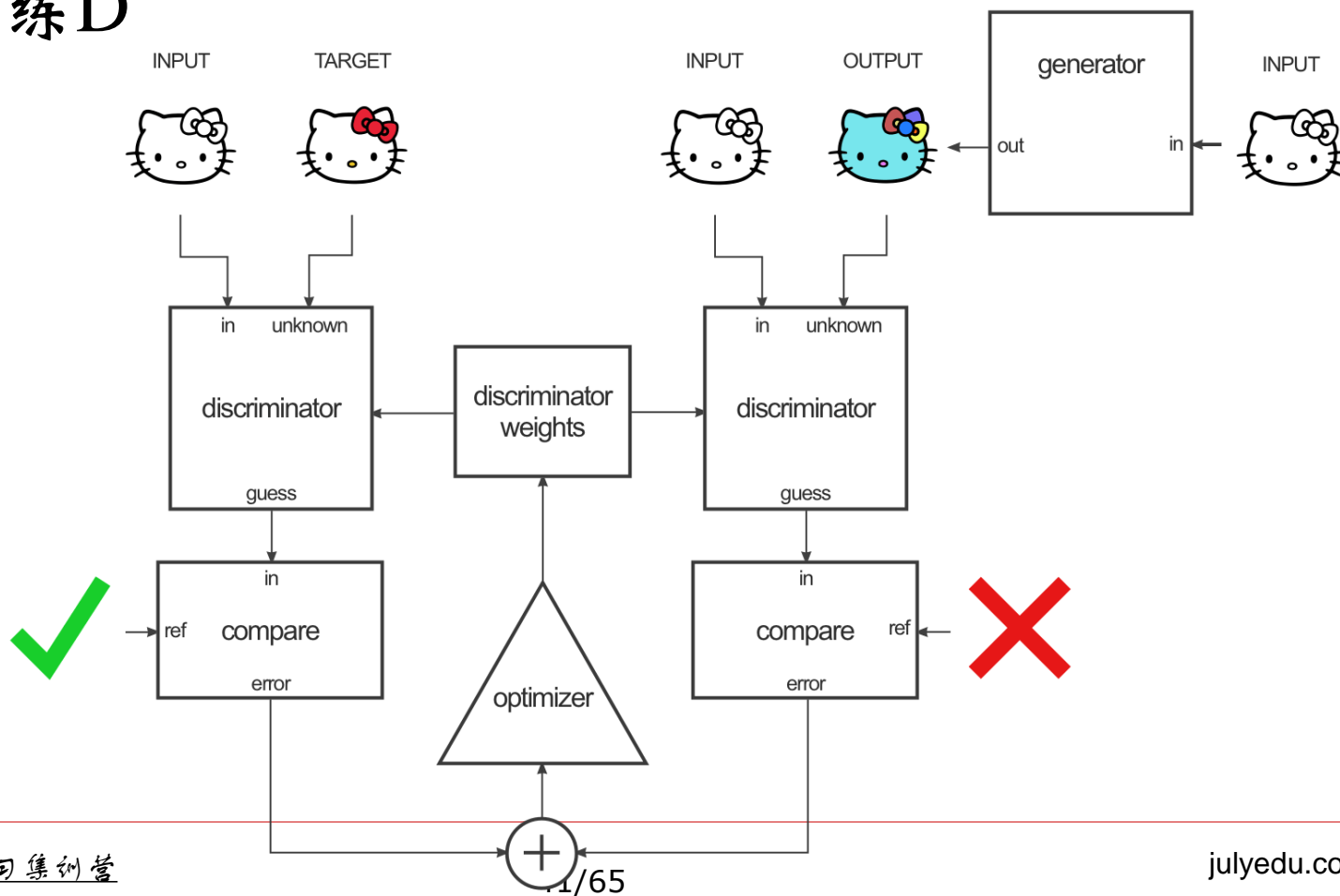
□ 训练G

INPUT



图像翻译(Pix2Pix)

□ 训练D



图像翻译(Pix2Pix)-效果

Aerial photo to map



Map to aerial photo



input

output

input

output



图像翻译(Pix2Pix)-效果



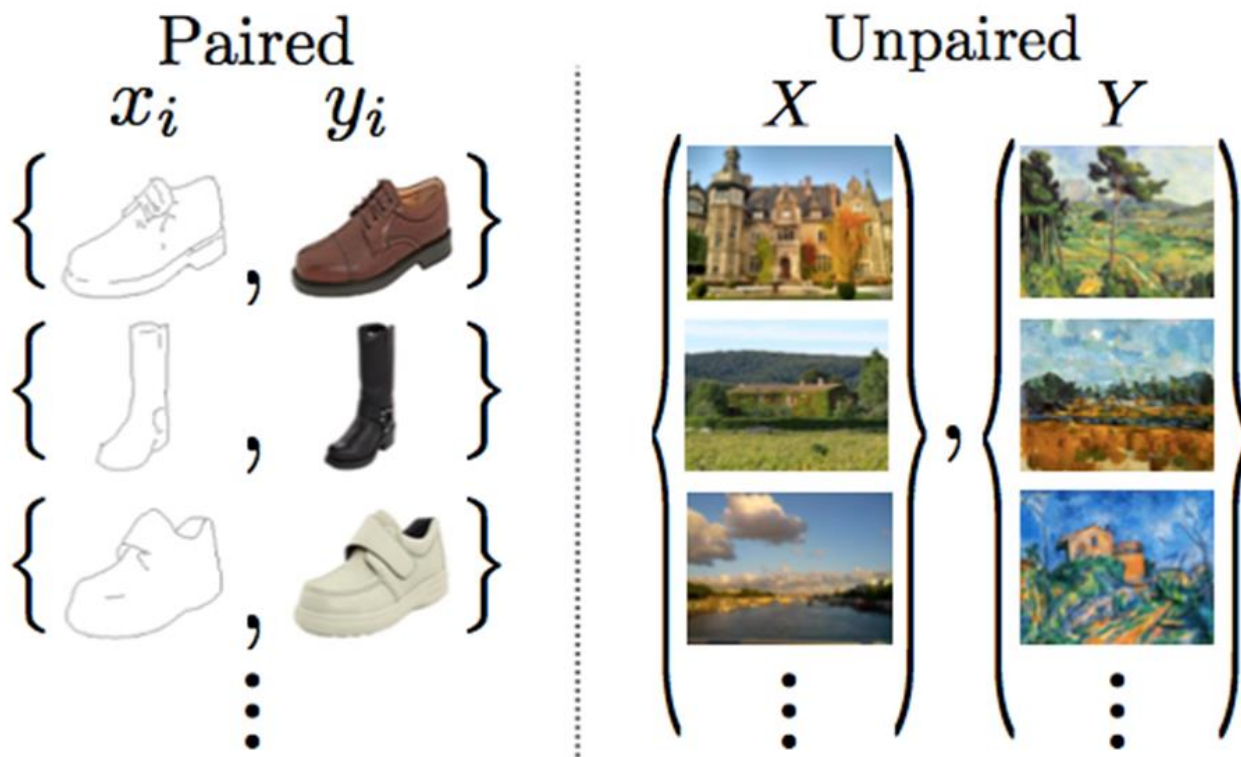
图像翻译(Pix2Pix)-效果

字 種 成 東 字 推
符 利 對 亞 型 斷
到 用 抗 語 進 的
字 條 網 言 行 新
符 件 絡 字 自 方
一 生 對 體 動 法



无配对图像翻译(CycleGAN)

□ Pix2Pix 要求图像必须成对



无配对图像翻译(CycleGAN)

- 目标：从无配对的图像中学到风格的变换
- 想法：类似于语言翻译，翻译过去再翻译回来要保持信息



无配对图像翻译(CycleGAN)

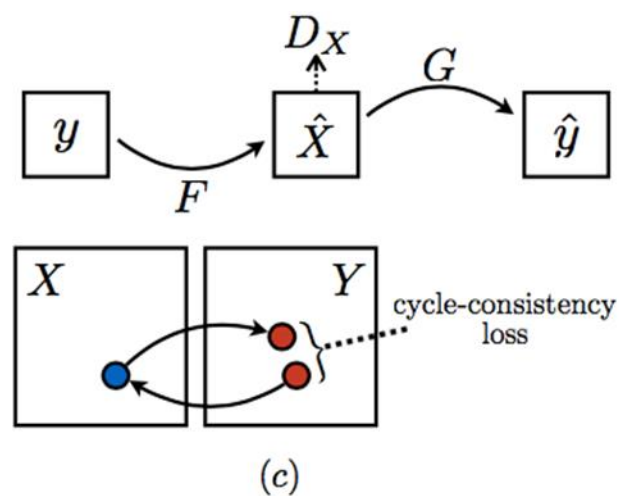
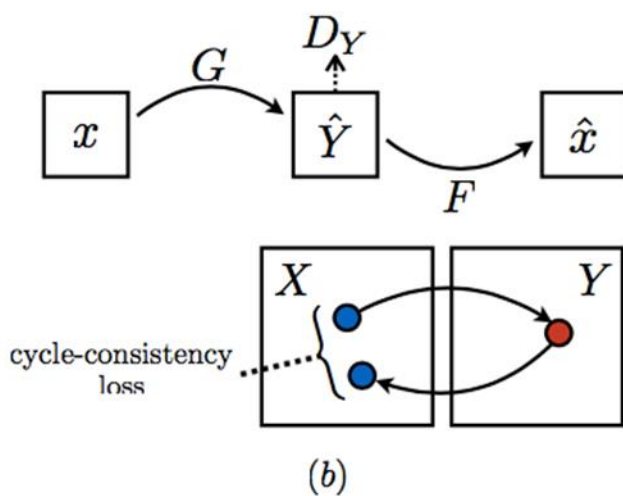
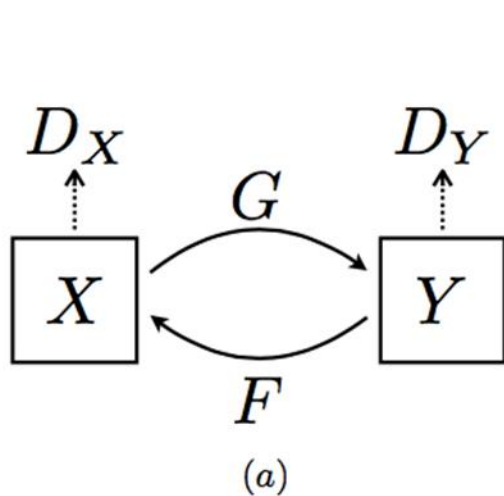
□ 建模:

- 两个领域 x, y
- 映射 $G: x \rightarrow y$
- 映射 $F: y \rightarrow x$
- 判别器 D_x : 区分 x 和 $F(y)$
- 判别器 D_y : 区分 y 和 $G(x)$



无配对图像翻译(CycleGAN)

□ Cycle一致性



无配对图像翻译(CycleGAN)

□ CycleGAN

■ 两套GAN

□ G和Dy

□ F和Dx

■ 两个一致性损失

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))]$$

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ + \lambda \mathcal{L}_{\text{cyc}}(G, F),$$

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] \\ + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1].$$

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y)$$



无配对图像翻译(CycleGAN)

□ 为何会有效？

■ 一致性约束的存在降低了搜索空间



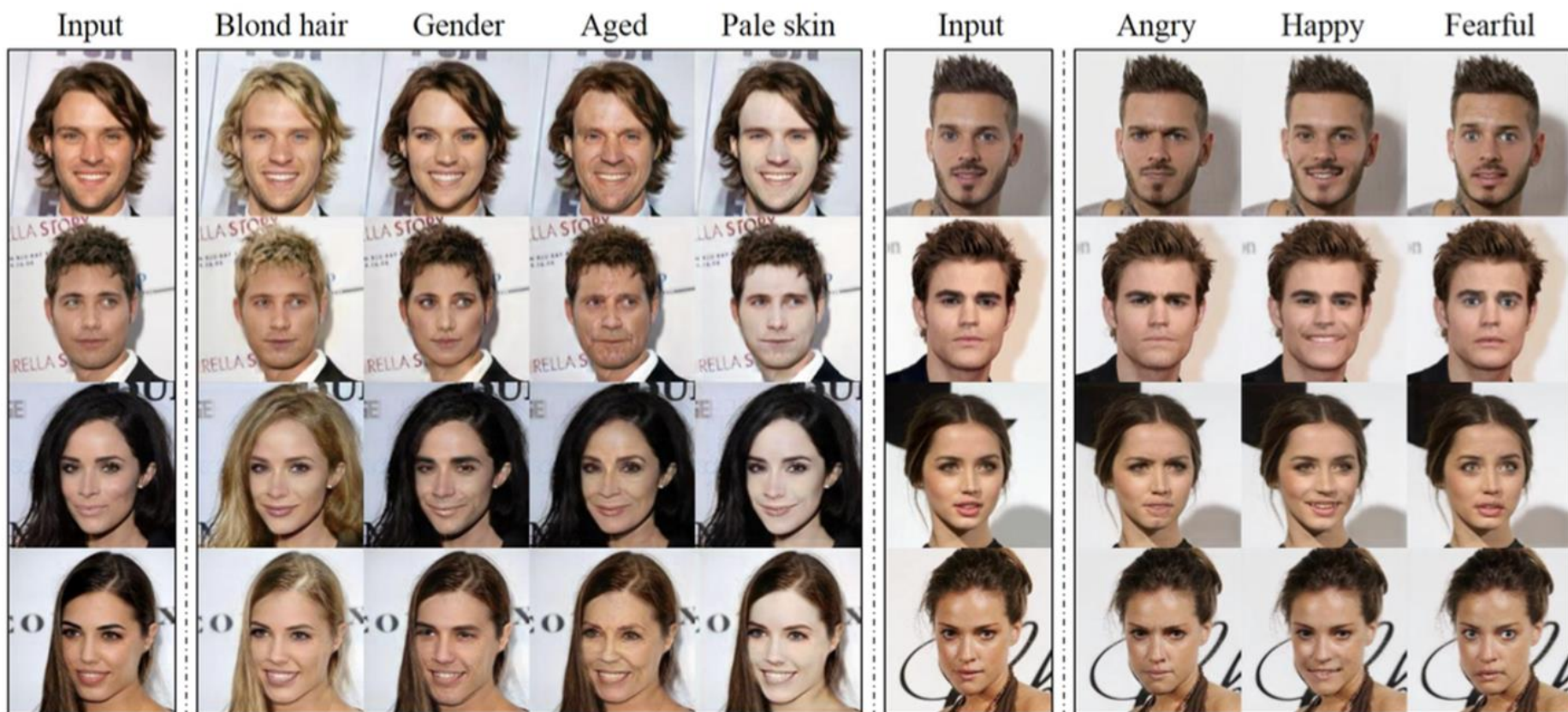
多领域图像翻译(StarGAN)

□ 之前模型的问题

- 每种转换训练一个新模型
- 多个任务：更换头发颜色，表情，年龄等
 - 低效
 - 模型中学到的信息不能共享，效果有限



多领域图像翻译(StarGAN)

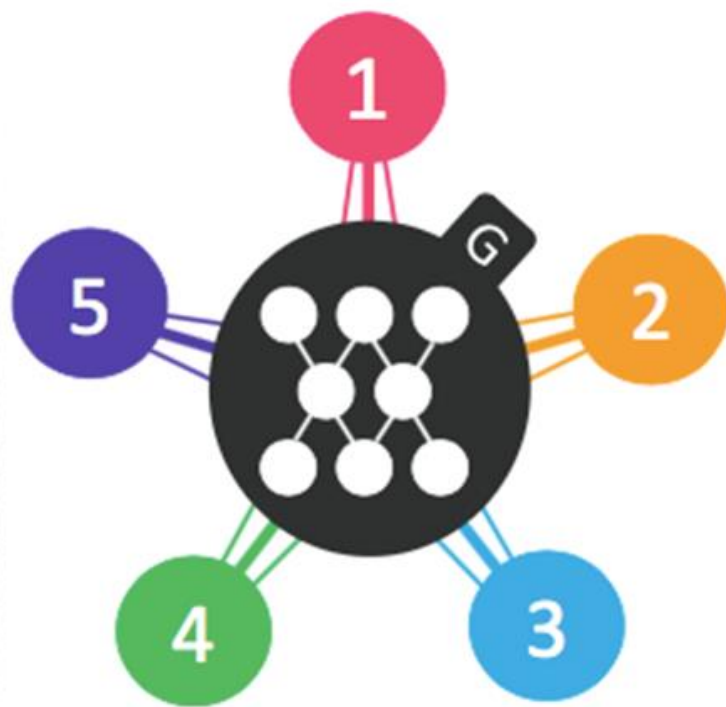


多领域图像翻译(StarGAN)

(a) Cross-domain models



(b) StarGAN



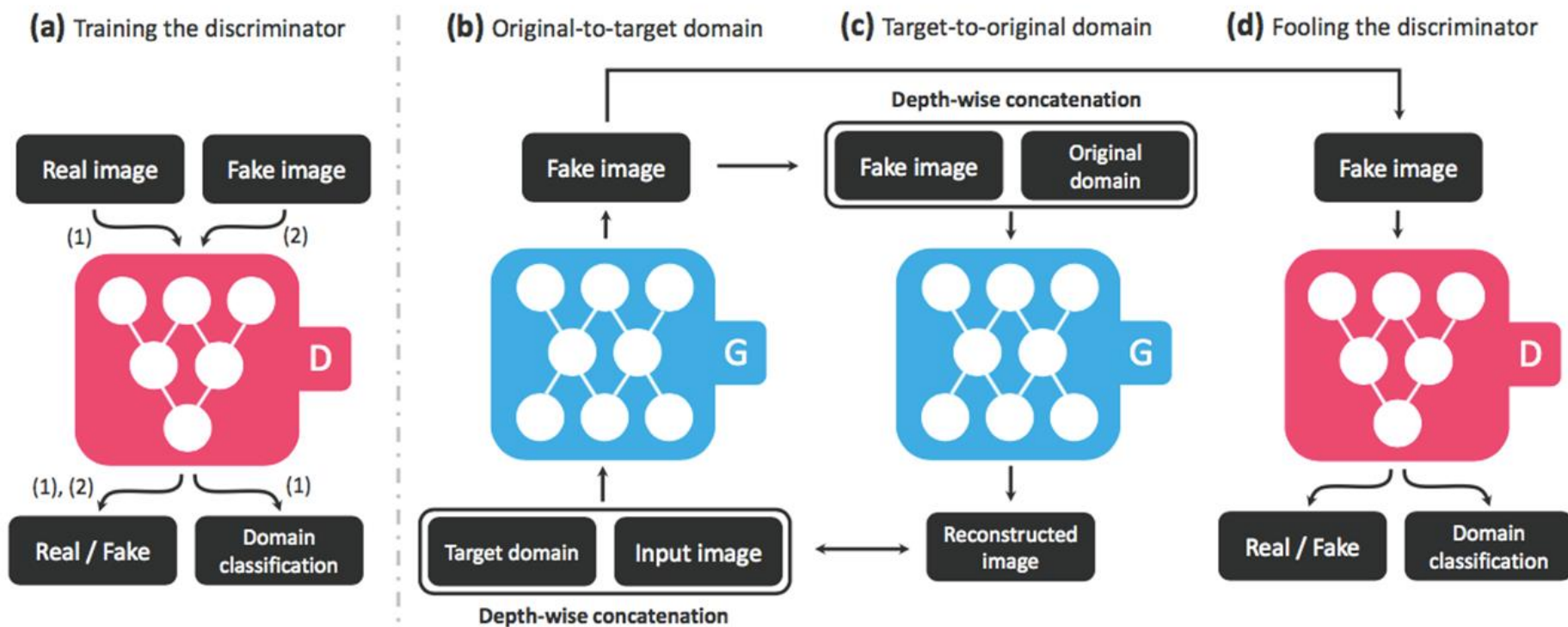
多领域图像翻译(StarGAN)

□ StarGAN

- G的输入添加目标领域信息
- D除了判断是否真实以外，还需要判断类别
- 图像重建loss



多领域图像翻译(StarGAN)



多领域图像翻译(StarGAN)

□ 目标函数

$$\mathcal{L}_{adv} = \mathbb{E}_x [\log D_{src}(x)] + \mathbb{E}_{x,c} [\log (1 - D_{src}(G(x, c)))].$$

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'} [\|x - G(G(x, c), c')\|_1]$$

$$\mathcal{L}_{cls}^f = \mathbb{E}_{x,c} [-\log D_{cls}(c|G(x, c))].$$

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^r,$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^f + \lambda_{rec} \mathcal{L}_{rec},$$

$$\mathcal{L}_{cls}^r = \mathbb{E}_{x,c'} [-\log D_{cls}(c'|x)].$$



多领域图像翻译(StarGAN)

- 解决的是成对图像的问题
- 生成器中有领域信息
- 判别器中也要判断是否是某个领域
- G和D训练的时候目标函数不同



文本生成图像(Text2Img)

□ 问题难点

■ 多样化

- 一个描述对应多张图，搜索空间比图像生成文本更大
- 一个词的变化图就会发生很大的变化

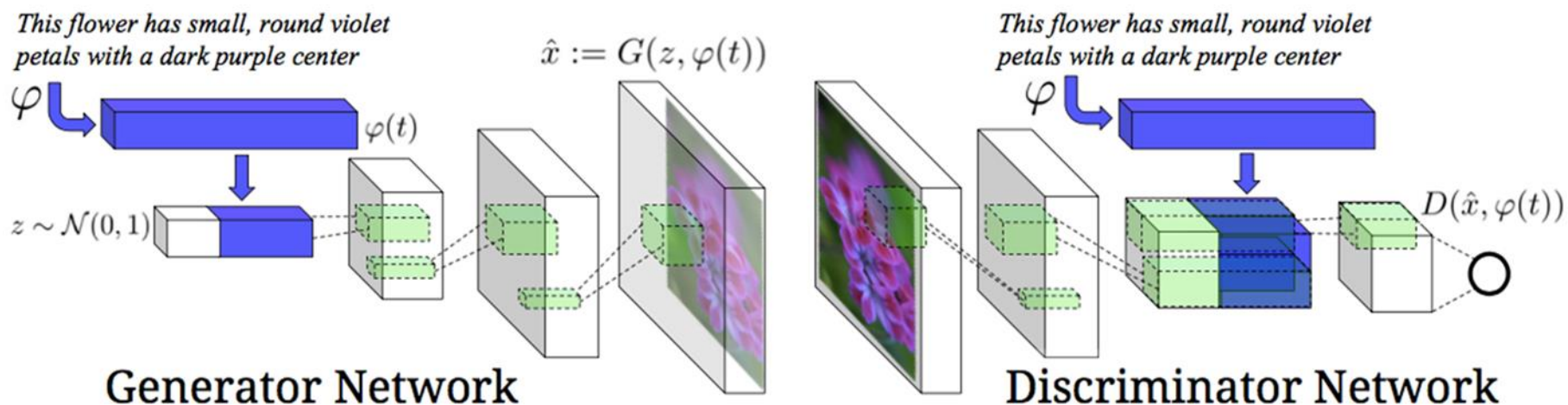
■ 任务

- 生成真实的图像
- 生成与文字匹配的图像



文本生成图像(Text2Img)

□ 模型结构



文本生成图像(Text2Img)

☐ 模型结构

■ G

☐ 输入：随机向量+文本编码

☐ 输出：图像

■ D

☐ 输入：生成图像+文本编码

☐ 输出：图像文本是否匹配+图像是否真实



文本生成图像(Text2Img)

□ 技巧

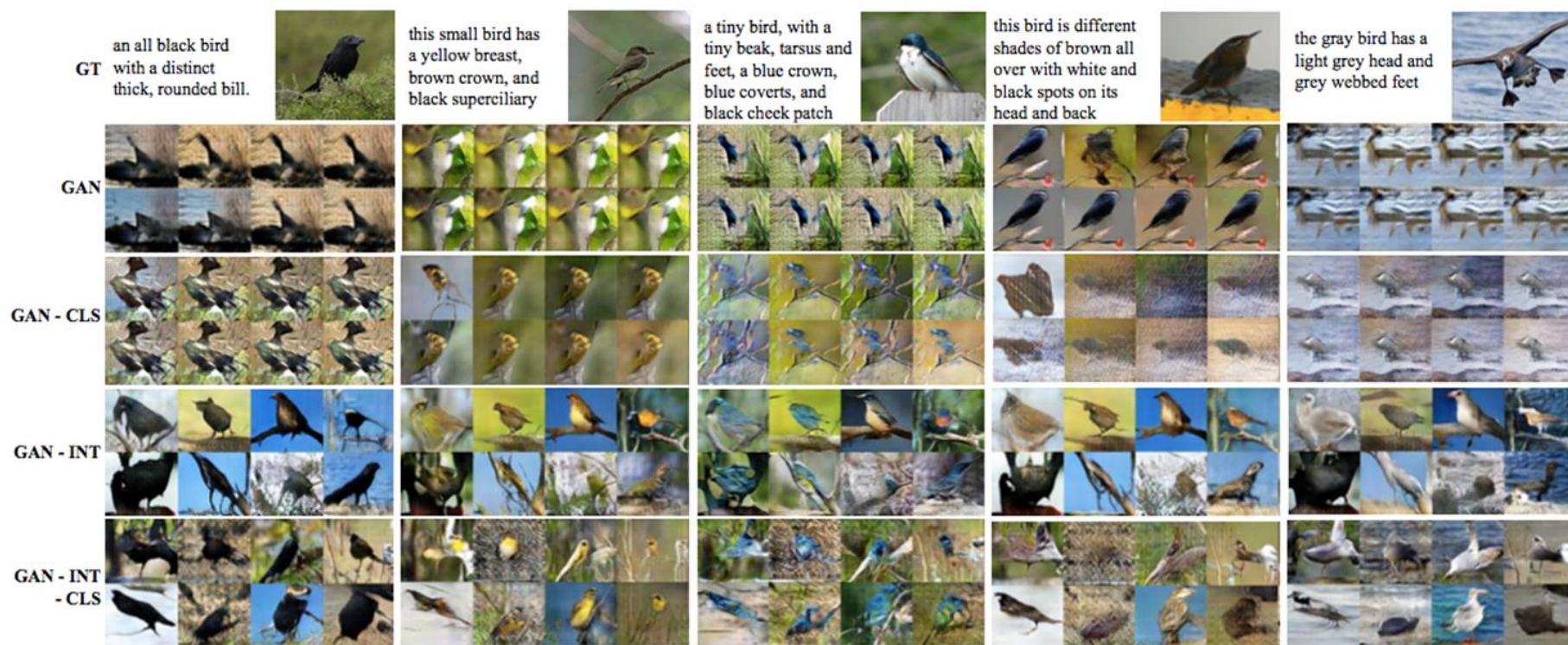
- 先将图像是否合理训练出来，再训练图像文本是否匹配
- 除了<假图，描述>和<真图，描述>之外，添加<真图，不匹配描述>



Algorithm 1 GAN-CLS training algorithm with step size α , using minibatch SGD for simplicity.

- 1: **Input:** minibatch images x , matching text t , mis-matching \hat{t} , number of training batch steps S
- 2: **for** $n = 1$ **to** S **do**
- 3: $h \leftarrow \varphi(t)$ {Encode matching text description}
- 4: $\hat{h} \leftarrow \varphi(\hat{t})$ {Encode mis-matching text description}
- 5: $z \sim \mathcal{N}(0, 1)^Z$ {Draw sample of random noise}
- 6: $\hat{x} \leftarrow G(z, h)$ {Forward through generator}
- 7: $s_r \leftarrow D(x, h)$ {real image, right text}
- 8: $s_w \leftarrow D(x, \hat{h})$ {real image, wrong text}
- 9: $s_f \leftarrow D(\hat{x}, h)$ {fake image, right text}
- 10: $\mathcal{L}_D \leftarrow \log(s_r) + (\log(1 - s_w) + \log(1 - s_f))/2$
- 11: $D \leftarrow D - \alpha \partial \mathcal{L}_D / \partial D$ {Update discriminator}
- 12: $\mathcal{L}_G \leftarrow \log(s_f)$
- 13: $G \leftarrow G - \alpha \partial \mathcal{L}_G / \partial G$ {Update generator}
- 14: **end for**

文本生成图像(Text2Img)-效果



总结

- 对抗生成网络原理
- 深度卷积对抗生成网络 (DCGAN)
- 图像翻译 (Pix2Pix)
- 无配对图像翻译 (CycleGAN)
- 多领域图像翻译 (StarGAN)
- 文本生成图像 (Text2Img)



Thanks!

Q&A