基于深度学习的生成模型的研究

报告提纲

- 生成式模型的定义
- 研究生成式模型的意义
- 经典的生成式模型：VAE 与 GAN
- GAN的应用
- VAE 与 GAN 仍需解决的问题
- 我们的工作
- 总结

生成式模型的定义

- 通常机器学习的任务就是学习一个模型，应用这一模型，对给定的输入预测相应的输出。对于模型的分类有很多种，其中一种分类就是把模型分为：判别模型和生成模型两种。

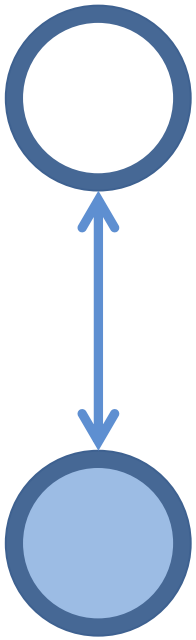
- 判别模型主要是根据输入图像推测图像具备的一些性质，即：已知观察变量x和隐变量z，直接对$p(z|x)$进行建模。它根据输入的观察变量x直接得到隐变量z出现的可能性。
  - 例如：当模型为分类模型时，隐变量z则代表类别变量。

- 生成模型则是要对$p(x,z)$进行建模，然后求出条件概率分布$p(z|x)$作为预测隐变量z的模型，即：

$$p(z|x) = \frac{p(x,z)}{p(x)} = \frac{p(x|z)p(z)}{p(x)}$$

同时，我们也可以根据联合概率分布$p(x,z)$采样生成观测变量x。

生成式模型的定义

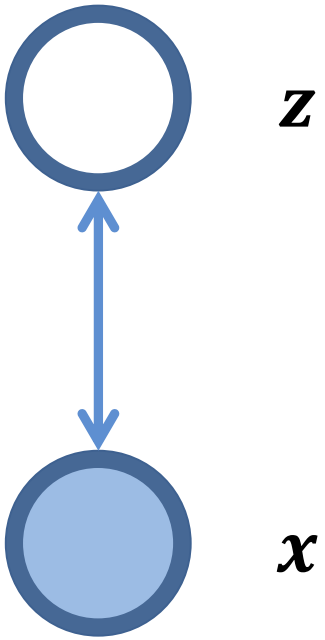


- 通常，我们已知观测变量x服从某固定但未知的分布，与隐变量z构成有向概率图。
- 对于这个概率图，$p(z)$（隐变量的先验）、$p(x|z)$（x相对z的条件概率）及$p(z|x)$（隐变量的后验）三者就可完全描述x和z的关系。两者的联合分布可以表示为：

$$p(x,z) = p(x|z)p(z)$$

- 我们只能观测到x，而z是隐变量，不能被观测。生成任务便是通过一个观察集$\mathcal{X}$，估计观测变量x与隐变量z构成的概率图的相关参数。

生成式模型的定义



- 对于一个模型，如果它能够建模$p(z)$、$p(x|z)$，我们就称之为生成模型，这有如下两层含义：

I. $p(z)$、$p(x|z)$两者决定了观测变量x与隐变量z的联合分布$p(x,z)$。

II. 利用两者可以对观测变量x进行采样。具体做法是：先依隐变量z的先验概率生成样本点$z_i \sim p(z)$，再依观测变量x的条件概率采样$x_i \sim p(x|z)$。

生成式模型的定义

- 借助于深度神经网络强大的建模能力，一些拥有出色生成能力的新的生成模型陆续出现。这其中，最为典型的代表有：变分自编码器(VAE)与对抗生成网络(GAN)。

- 与传统的生成模型PixelCNN相比，VAE与GAN更好地建模了观测变量x与隐变量z的关系，即生成样本由隐变量控制。且隐变量没有过多的约束(例如：nonlinear ICA中对隐变量的维度约束)。

- 与传统的生成模型Boltzmann Machines相比， VAE与GAN不需要高复杂度的马尔科夫链的计算。

报告提纲

- 生成式模型的定义
- 研究生成式模型的意义
- 经典的生成式模型：VAE 与 GAN
- GAN的应用
- VAE 与 GAN 仍需解决的问题
- 我们的工作
- 总结

研究生成式模型的意义

- 人们可能会问：为什么要研究生成式模型？特别是当生成式模型只能**用来产生数据**而不能直接对**数据的概率密度函数**进行估计的情况下。

  - 具体地说，在图像的应用中，这类模型好像仅仅是提供更多的图像，但是我们的世界中并不缺少图像。

研究生成式模型的意义

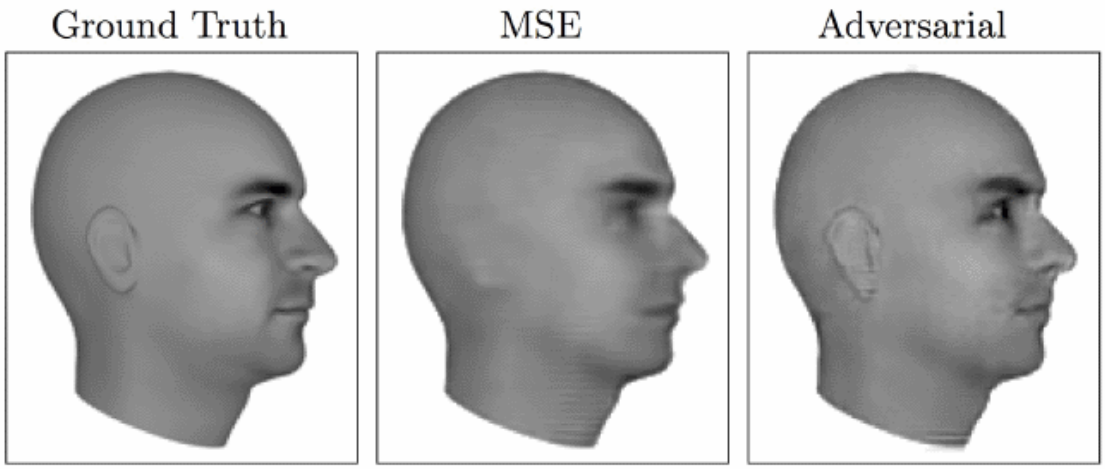
- 具体来说，研究生成式模型有很多原因，其中包括：

  - 生成式模型的训练和采样是一个对我们关于表达和操作高维概率分布问题的能力的非常好的测试。高维的概率分布问题是一个很重要的问题，其在数学和工程领域有很广泛的应用。

  - 生成式模型可以使用有缺失的数据(missing data)来训练，并且可以对缺失的数据进行预测。一个缺失数据训练的例子是半监督学习(Semi-supervised learning)。常见的深度模型都需要大量的标定数据来进行训练才能得到比较好的推广能力。 半监督学习可以使用大量的非标定数据来提高模型的推广能力。生成式模型，特别是GAN可以让半监督学习表现得相当好（GAN应用于半监督学习的例子，我们将在后面的应用中介绍）。

研究生成式模型的意义

- 生成式模型，例如GAN，使机器学习可以用于多模(multi-modal)输出问题。有很多任务，一个输入可能对应多个正确的输出，每一个输出都是可接受的。
  - 视频的未来帧预测

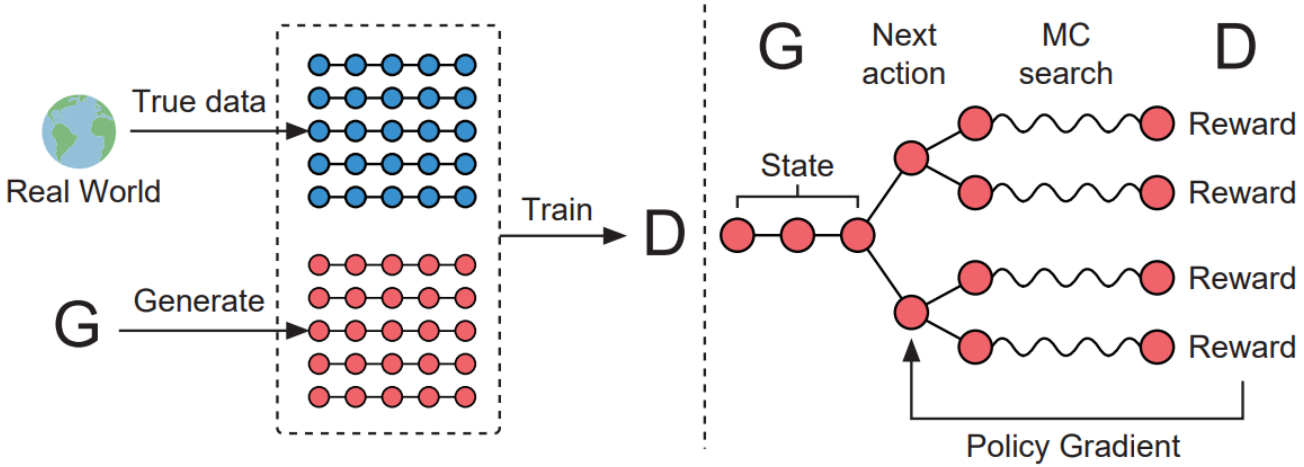


(Lotter et al. 2016)

研究生成式模型的意义

- 生成式模型可以以多种方式被应用到增强学习中(RL)。

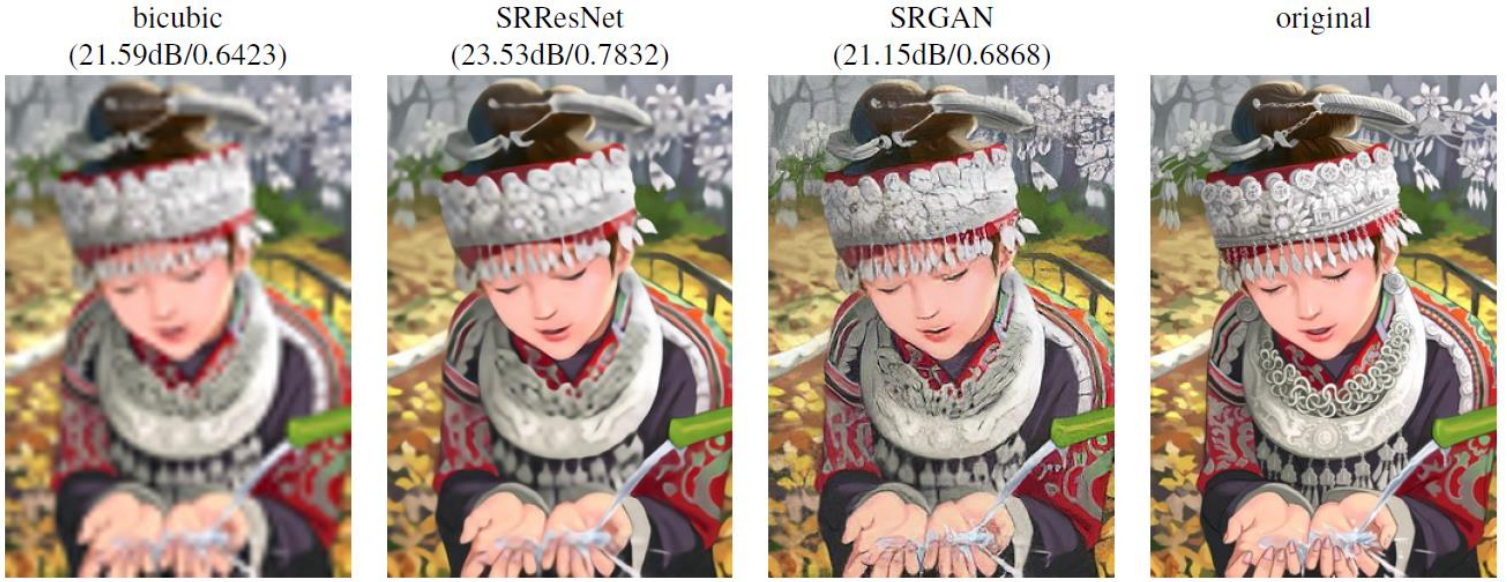
  - 例如：将生成式模型用于对假设环境的增强学习，使其通过使用当前状态以及agent假设的可能行为作为输入，从而学习到未来状态的条件分布。这样即使发生错误行为也不会造成实际的损失。



(Lantao et al. 2017)

研究生成式模型的意义

- 很多任务本身需要实现逼真的数据生成。



(Ledig et al. 2016)

- 图片超分辨率：这种任务是使用一个低分辨率图片产生高分辨率图片。之所以需要使用生成式模型是因为需要模型产生并加入比原始输入的图像更多的信息。一个低分辨率的图像会对应多种可能的高分辨率图像。模型需要根据概率分布从可能的图像中选出一个。通过对所有可能的样本做平均化处理会使图像变模糊。

研究生成式模型的意义

- Image-to-Image 转换应用可以将航空图像转换为地图，或者将素描转换为图像。

(Isola et al. 2016)

报告提纲

- 生成式模型的定义
- 研究生成式模型的意义
- 经典的生成式模型：VAE 与 GAN
- GAN的应用
- VAE 与 GAN 仍需解决的问题
- 我们的工作
- 总结

Variational Auto-Encoder

- 求解概率分布的参数的最经典方法是最大似然估计(MLE)，MLE假设最大化似然的参数为最优的参数估计。具体来说，如果样本集x的概率分布为$p(x)$，如果一次观测中具体观测到的样本分别为$x_1, x_2, \dots, x_n$，并假设它们是相互独立的，那么观测样本的似然为：

$$\mathcal{L} = \prod_{i=1}^{n} p(x_i)$$

- 如果$p(x)$是一个带有参数θ的概率分布$p_\theta(x)$，那么我们应当想办法选择θ，使得$\mathcal{L}$最大化，即：

$$\theta = \arg\max_\theta \prod_{i=1}^{n} p_\theta(x_i)$$

- 对概率取对数，就得到等价形式：

$$\theta = \arg\max_\theta \sum_{i=1}^{n} \log p_\theta(x_i)$$

Variational Auto-Encoder

- 如果右端再除以n，我们就可以得到更精炼的表达形式：

$$\theta = \arg\max_{\theta} \mathbb{E}[\log p_\theta(x_i)]$$

- 对于我们的生成模型，首先我们有一批观测样本$\{x_1, x_2, \dots, x_n\}$，其整体用x来描述，我们借助隐变量z描述x的分布$p_\theta(x)$。我们想要求解概率模型的参数θ：

$$\theta = \arg\max_{\theta} \log p_\theta(x) = \arg\max_{\theta} \log \int p_\theta(x, z)\, dz$$

- 对于包含隐变量的概率模型的求解，我们很自然地想到了EM算法。那么，如果使用EM算法对我们的问题进行求解，我们需要进行如下的计算步骤：

① 初始化参数的初值$\theta^{(0)}$；

② **E**步：记$\theta^{(i)}$为第i次迭代参数θ的估计值，在第$i+1$次迭代的**E**步，计算Q函数：

$$Q(\theta, \theta^{(i)}) = \int p_{\theta^{(i)}}(z|x) \log p_\theta(x, z) dz$$

③ **M**步：求使$Q(\theta, \theta^{(i)})$极大化的θ，确定第$i+1$次迭代的参数估计值$\theta^{(i+1)}$直到收敛：

$$\theta^{(i+1)} = \arg\max_{\theta} Q(\theta, \theta^{(i)})$$

Variational Auto-Encoder

- 在使用EM算法求解参数的迭代计算过程中，涉及到对$p_\theta(z|x)$的积分。但是，实际中由于概率分布的多样性及隐变量的高维等问题，这个 积分一般是难以计算的(intractable)。这就使得Q函数的显式计算变得无法实现。

- 这时，我们不妨换一个思路。我们的最终目的是通过最大似然估计(MLE)求解概率模型的参数，但隐变量的存在以及相关积分项的不可计算使得问题的直接求解变得困难。所以，我们可以求助于另一个“分析工具”——KL散度。

  - 假设有两个分布：$p_1(x)$与$p_2(x)$，我们可以用KL散度来衡量它们的距离：

$$KL(p_1(x)||p_2(x)) = \sum_x p_1(x) \log \frac{p_1(x)}{p_2(x)}$$

  - 当两个分布相同时， KL散度为0，当两个分布不同时， KL散度大于0。

Variational Auto-Encoder

- 当我们希望一个带有参数θ的概率分布$p_\theta(x)$尽可能地接近另一个样本分布$\tilde{p}(x)$时，我们可以用KL散度对参数θ进行求解：

$$\theta = \arg\min_\theta KL(\tilde{p}(x)||p_\theta(x))$$

- 如果x的样本集已经给出，这就意味着$\tilde{p}(x)$是确定的，我们可以得到：

$$\theta = \arg\min_\theta \sum_x \tilde{p}(x)\log\frac{\tilde{p}(x)}{p_\theta(x)}$$

$$= \arg\min_\theta [\sum_x -\tilde{p}(x)\log p_\theta(x) + \boxed{\sum_x \tilde{p}(x)\log\tilde{p}(x)}] \text{ 常量}$$

$$= \arg\max_\theta \sum_x \tilde{p}(x)\log p_\theta(x)$$

$$= \boxed{\arg\max_\theta \mathbb{E}[\log p_\theta(x)]} \text{ 最大似然估计}$$

- 这样，我们从KL散度又得到了最大似然估计的表达式。可以说最大似然估计是KL散度已知$\tilde{p}(x)$的特殊情况。

Variational Auto-Encoder

- 出发点仍然没变，这里再重申一下。首先我们有一批观测样本 $\{x_1, \dots, x_n\}$，其整体用x来描述。借助隐变量z，观测样本与隐变量的联合分布为：$p(x,z)$，由于我们手头上只有x的样本，因此上式可以改写为：

$$p(x,z) = \tilde{p}(x)p(z|x) \tag{1}$$

  - 注意这里的$\tilde{p}(x)$是根据样本$x_1, x_2, \dots, x_n$确定的关于x的先验分布。尽管我们无法准确写出它的形式，但它是确定的、存在的。

- 接下来，直接对$p(x,z)$进行近似。具体来说，我们设想用一个新的联合概率分布$q(x,z)$来逼近$p(x,z)$，那么我们可以用KL散度来计算它们的距离：

$$KL(p(x,z)||q(x,z)) = \iint p(x,z)\log\frac{p(x,z)}{q(x,z)}dzdx \tag{2}$$

- KL散度是我们的最终目标，因为我们希望两个分布越接近越好，所以KL散度越小越好。将(1)式代入(2)式，我们有：

Variational Auto-Encoder

$$KL(p(x,z)||q(x,z)) = \int \tilde{p}(x)[\int p(z|x)\log\frac{\tilde{p}(x)p(z|x)}{q(x,z)}dz]\,dx$$

$$= \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{\tilde{p}(x)p(z|x)}{q(x,z)}dz] \qquad (3)$$

- (3)式还可以进一步简化：

$$KL(p(x,z)||q(x,z)) = \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\tilde{p}(x)\,dz] + \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)]$$

$$= \mathbb{E}_{x\sim\tilde{p}(x)}[\log\tilde{p}(x)\underbrace{\int p(z|x)dz}_{1}] + \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)]$$

$$= \underbrace{\mathbb{E}_{x\sim\tilde{p}(x)}[\log\tilde{p}(x)]}_{\text{常量}C} + \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)] \qquad (4)$$

Variational Auto-Encoder

- 通过移项，我们可以令：

$$\mathcal{L} = KL(p(x,z)||q(x,z)) - C = \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)]$$

- 最小化KL散度也就等价于最小化$\mathcal{L}$。为了得到生成模型，我们把$q(x,z)$写成$q(x|z)q(z)$，于是有：

$$\begin{aligned}\mathcal{L} &= \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x|z)q(z)}dz)] \\ &= \mathbb{E}_{x\sim\tilde{p}(x)}[-\int p(z|x)\log q(x|z)dz + \int p(z|x)\log\frac{p(z|x)}{q(z)}dz] \\ &= \underbrace{\mathbb{E}_{x\sim\tilde{p}(x)}[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))]}_{\text{优化目标}} \quad (5)\end{aligned}$$

Variational Auto-Encoder

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))]$$

- 得到了优化目标，我们就要想办法找到合适的$q(x|z)$和$q(z)$使得$\mathcal{L}$最小化。首先，为了方便采样，我们假设$z\sim N(0,I)$，即标准的多元正态分布，这就解决了$q(z)$。

- 然后，$p(z|x)$也是(各分量独立的)正态分布，其均值与方差由x来决定，这个“决定”，就是一个神经网络的拟合：

$$p(z|x) = \frac{1}{\prod_{k=1}^{d}\sqrt{2\pi\sigma_{(k)}^2(x)}}\exp\left(-\frac{1}{2}\left\|\frac{z-\mu(x)}{\sigma(x)}\right\|^2\right)$$

- 其中，x是神经网络的输入，输出则是均值$\mu(x)$与方差$\sigma^2(x)$。这里的神经网络就起到了类似Encoder的作用。$\mathcal{L}$中的KL散度这一项可以先算出来：

$$KL(p(z|x)||q(z)) = \frac{1}{2}\sum_{k=1}^{d}(\mu_{(k)}^2(x) + \sigma_{(k)}^2(x) - \log\sigma_{(k)}^2(x) - 1)$$

Variational Auto-Encoder

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))]$$

- 现在只剩下生成模型部分$q(x|z)$了，对于分布的选择，原论文给出了两种候选方案：伯努利分布或正态分布。
  - 伯努利分布其实就是一个二值分布：

$$p(\varepsilon) = \begin{cases} \rho, & \varepsilon = 1 \\ 1-\rho, & \varepsilon = 0 \end{cases}$$

- 所以伯努利分布只适用于 x 是一个多元的二值向量的情况，比如：MNIST。这种情况下，我们用神经网络$\rho(z)$来算参数ρ，从而得到：

$$q(x|z) = \prod_{k=1}^{D} \left(\rho_{(k)}(z)\right)^{x_{(k)}} \left(1-\rho_{(k)}(z)\right)^{1-x_{(k)}}$$

- 这时可以算出：

$$-\log q(x|z) = \sum_{k=1}^{D} \left[-x_{(k)} \log \rho_{(k)}(z) - \left(1-x_{(k)}\right) \log\left(1-\rho_{(k)}(z)\right)\right]$$ 交叉熵

- 这里$\rho(z)$就起到了类似Decoder的作用。

Variational Auto-Encoder

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))]$$

- 然后是正态分布，与$p(z|x)$很像，只是x，z交换了位置：

$$q(x|z) = \frac{1}{\prod_{k=1}^{D}\sqrt{2\pi\sigma_{(k)}^2(z)}}\exp\left(-\frac{1}{2}\left\|\frac{x-\mu(z)}{\sigma(z)}\right\|^2\right)$$

- 这里，神经网络的输入是z，输出是$\mu(z)$与$\sigma^2(z)$。于是：

$$-\log q(x|z) = \frac{1}{2}\left\|\frac{x-\mu(z)}{\sigma(z)}\right\|^2 + \frac{D}{2}\log 2\pi + \frac{1}{2}\sum_{k=1}^{D}\log\sigma_{(k)}^2(z)$$

- 通常情况下，我们会固定方差为一个常数σ^2，这时候有：

$$-\log q(x|z) \sim \frac{1}{2\sigma^2}\|x-\mu(z)\|^2$$

- 于是，这就变成了我们熟悉的MSE损失函数！$\mu(z)$就起到了Decoder的作用。

Variational Auto-Encoder

- 现在，让我们看回VAE的优化目标：

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))]$$

- 结论已经很明显了：
  - 对应于等号右侧的第二项， $p(z|x)$起到Encoder的作用，同时KL散度将Encoder输出的隐向量约束为标准的多元正态分布；
  - 对应于等号右侧的第一项， $q(x|z)$起到Decoder的作用。对于二值数据(例如：MNIST)，我们可以对Decoder用sigmoid函数激活，然后用交叉熵作为损失函数，这对应于$q(x|z)$为伯努利分布；而对于一般数据，我们用MSE作为损失函数，这对应于$q(x|z)$为固定方差的多元正态分布；
  - 待训练完成后，Decoder就是我们的生成模型。生成过程就是从标准多元正态分布中采样得到隐变量z，再输入Decoder，就可以得到生成样本了。

Variational Auto-Encoder

- 到了这一步，我们仍然有一个问题需要解决。首先让我们看看整个模型的计算过程：

  ① 数据的前向传递过程：观测样本x输入Encoder得到均值$\mu(x)$与方差$\sigma^2(x)$，并计算得到KL散度的loss。再依据均值$\mu(x)$与方差$\sigma^2(x)$采样得到隐变量$z \sim N(\mu(x), \sigma^2(x))$。最后将$z$输入Decoder得到输出$\hat{x}$，并使用交叉熵或者MSE损失函数计算重构loss。

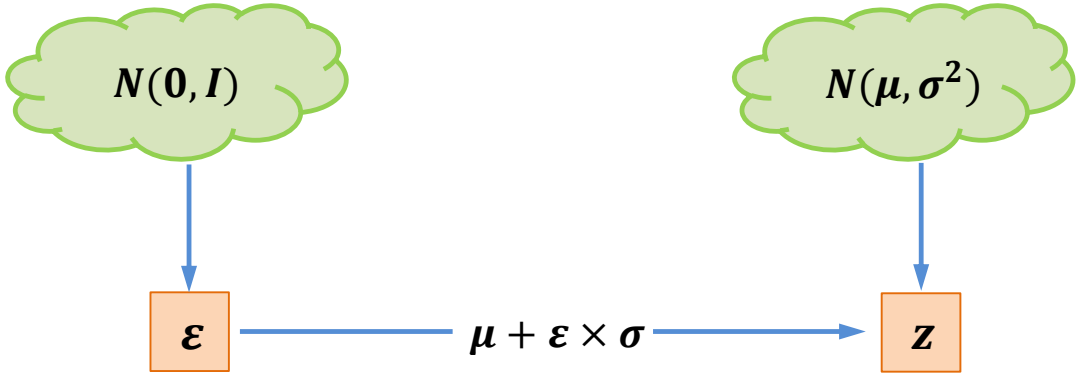
  ② 梯度的反向传递过程：将总的loss反向传递，并通过梯度下降更新Encoder与Decoder的网络参数。

- 但是，在Encoder与Decoder之间的从$N(\mu(x), \sigma^2(x))$中采样出一个z的操作是不可导的。这会使来自Decoder的梯度信息无法传入Encoder中，导致网络的训练无法进行。

Variational Auto-Encoder

- 原论文给出了一个解决方法，英文名是：reparameterization trick，我们这里称它为重参数技巧。其实很简单，它的核心思想是：

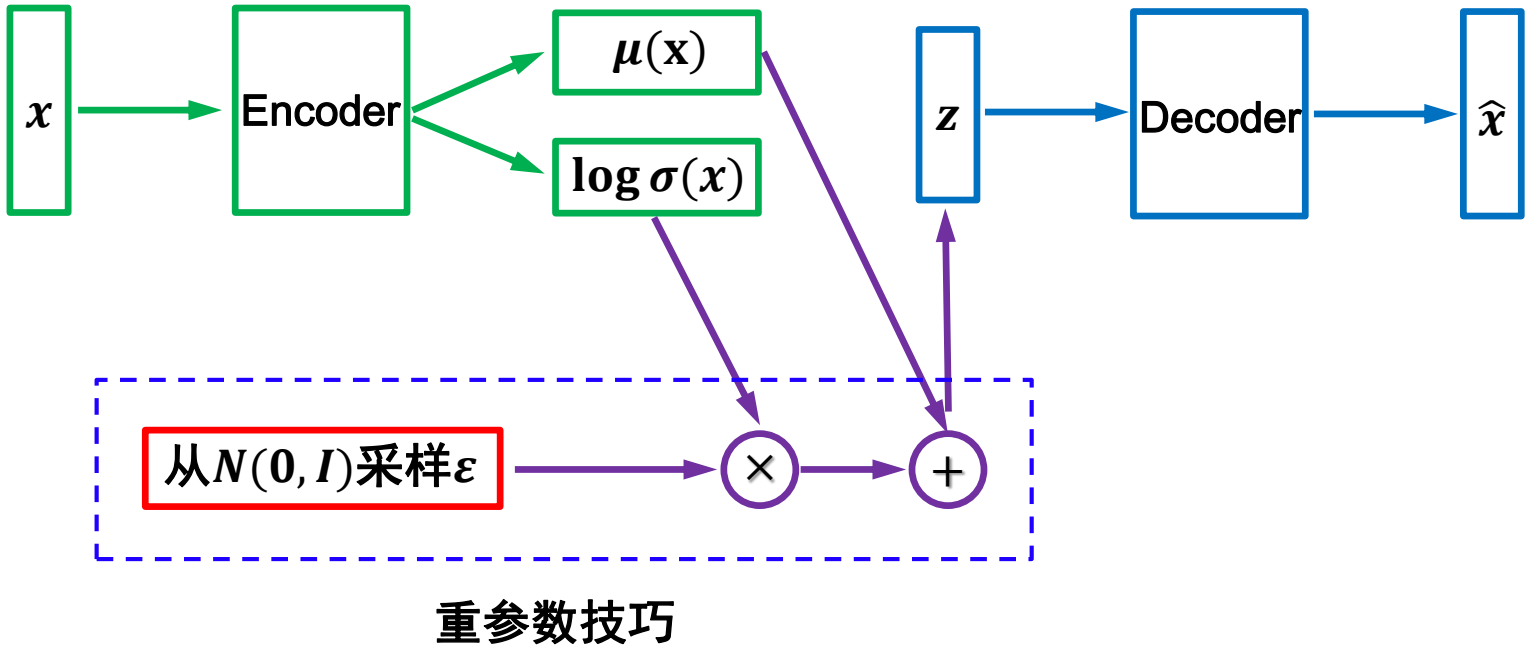
  - 从$N(\mu(x), \sigma^2(x))$中采样一个z，相当于从$N(0, I)$中采样一个ε，然后让$z = \mu(x) + \varepsilon \times \sigma(x)$。



- 于是，我们将从$N(\mu(x), \sigma^2(x))$采样变成了从$N(0, I)$中采样，然后通过参数变换得到想要的结果。这样一来，“采样”这个操作就不用参与梯度下降了，改为采样的结果参与，使得整个模型可训练了。

Variational Auto-Encoder

- VAE的整体模型结构如下：



Variational Auto-Encoder

- 对于$q(z)$为标准正态分布的假设，原论文作者也没有给出详尽的解释，但其实也不难理解，大致的原因有以下两点：

① 首先，因为我们要构造的是一个分布，而不是任意一个函数。既然是分布，就得满足归一化的要求。同时，要使KL散度容易计算，又要方便采样。我们还真没多少选择。

② 其次，同样简单的分布还有均匀分布，为什么我们不用呢？主要还是因为KL散度的计算问题。

$$KL(p(x)||q(x)) = \int p(x)\log\frac{p(x)}{q(x)}dx$$

要是在某个区域中$p(x) \neq 0$而$q(x) = 0$的话，那么KL散度就无穷大了。对于正太分布来说，所有点的概率密度都是正的，因此不存在这个问题。但对于均匀分布来说，只要两个分布不一致，那么就必然存在$p(x) \neq 0$而$q(x) = 0$的区间，因此KL散度会非常大，在训练中会盖过“重构部分”的训练信息，使得训练无法正常进行。但对于GAN，我们将看到，由于没有KL散度的直接计算，隐变量分布是可以使用均匀分布的。

Variational Auto-Encoder

- VAE的本质是什么？

- 抛开极大似然估计的思想，只看训练过程。VAE首先通过Encoder将观测样本x编码（映射）为隐空间中的隐变量z，然后再试图通过MSE损失函数（对于二值图像则为交叉熵损失函数）将隐变量z重构回观测样本x。而这样的训练过程与普通的Auto-Encoder是非常相似的，唯一的不同在于VAE多了作用于隐变量的KL散度约束。

- 再看Auto-Encoder，一个训练好的Auto-Encoder，如果能够在它训练得到的隐空间采样，然后作为Decoder的输入，我们就得到了一个生成模型。但是普通Auto-Encoder的隐空间的分布是复杂且未知的，我们无法进行采样。VAE则对隐空间施加了KL散度约束，使其逼近简单的多元标准正太分布，这就使隐空间的采样变得可能，进而得到我们的生成模型。

- 所以，从训练过程来看，VAE就是隐空间受约束的Auto-Encoder。

Generative Adversarial Networks

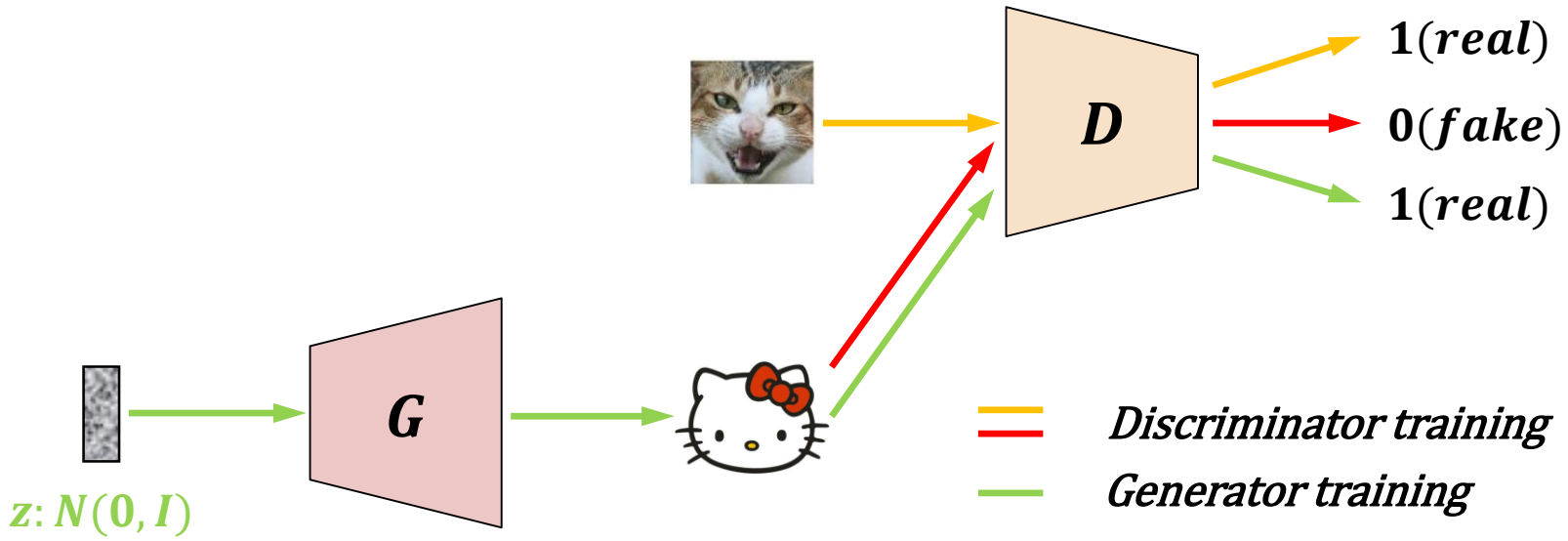
- GAN受博弈论中的零和博弈启发，将生成问题视作判别器D和生成器G这两个网络的对抗和博弈：生成器以随机噪声（一般为均匀分布或者正态分布）为输入，输出生成数据，判别器分辨生成数据和真实数据。前者试图产生更真实的数据，相应地，后者试图更完美地分辨真实数据与生成数据。

- 由此，两个网络在对抗中进步，在进步后继续对抗，由生成式网络得的数据也就越来越完美，越来越逼近真实数据，从而可以生成想要得到的数据。

Generative Adversarial Networks

- 设z为随机噪声，x为真实数据，生成网络和判别网络可以分别用G和D表示，其中D可以看作一个二分类器，那么采用交叉熵表示，GAN的优化目标可以写作：

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



Generative Adversarial Networks

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

- 其中第一项的$D(x)$表示判别器对真实数据的判断，第二项$D(G(z))$则表示对生成数据的判断。
  - 判别器D的目标是最大化这个公式，也就是甄别出哪些数据是来自真实数据分布的。
  - 生成器G的目标是最小化这个公式，也就是让自己生成的数据被判别器判断为来自真实数据分布。
- 通过这样一个极大极小(Max-Min)博弈，循环交替地分别优化G和D来训练所需要的生成式网络与判别式网络，直到到达Nash均衡点。

Generative Adversarial Networks

- 与VAE不同，GAN的优化目标与训练过程来源于博弈的思想，而不需要对隐变量 z 做推断。但实际上继续深入地分析GAN的优化目标就可以发现：它也“暗含着”对分布距离的最小化。

- 固定生成器G的参数，优化更新D的参数时。令$p_r(x)$为真实数据分布，$p_g(x)$为生成数据分布。优化目标可以写为：

$$\max_D \mathcal{V}(D,G) = \mathbb{E}_{x\sim p_r(x)}[\log D(x)] + \mathbb{E}_{x\sim p_g(x)}[\log(1-D(x))]$$
$$= \int p_r(x)\log D(x) + p_g(x)\log(1-D(x))\,dx \qquad (1)$$

- 令(1)式关于$D(x)$的导数为0，可以得到$D(x)$的全局最优解为:

$$D^*(x) = \frac{p_r(x)}{p_r(x)+p_g(x)} \qquad (2)$$

Generative Adversarial Networks

$$D^*(x) = \frac{p_r(x)}{p_r(x) + p_g(x)} \tag{2}$$

- (2)式即为最优判别器的表达式，这是一个很直观的结果。当$p_r(x)$与$p_g(x)$完全相同时，则$D^*(x)$恒为0.5，意味着判别器已无法再区分真实数据分布与生成数据分布。

- 固定判别器D的参数，优化更新G的参数时。优化目标可以写为：

$$\min_G \mathcal{V}(D,G) = \mathbb{E}_{x\sim p_r(x)}[\log D(x)] + \mathbb{E}_{x\sim p_g(x)}[\log(1 - D(x))] \tag{3}$$

- 我们将最优判别器的表达式代入(3)式，有：

$$\mathbb{E}_{x\sim p_r(x)} \log \frac{p_r(x)}{p_r(x) + p_g(x)} + \mathbb{E}_{x\sim p_g(x)} \log \frac{p_g(x)}{p_r(x) + p_g(x)} \tag{4}$$

Generative Adversarial Networks

$$\mathbb{E}_{x\sim p_r(x)}\log\frac{p_r(x)}{p_r(x)+p_g(x)}+\mathbb{E}_{x\sim p_g(x)}\log\frac{p_g(x)}{p_r(x)+p_g(x)} \tag{4}$$

- 对(4)式变形得到：

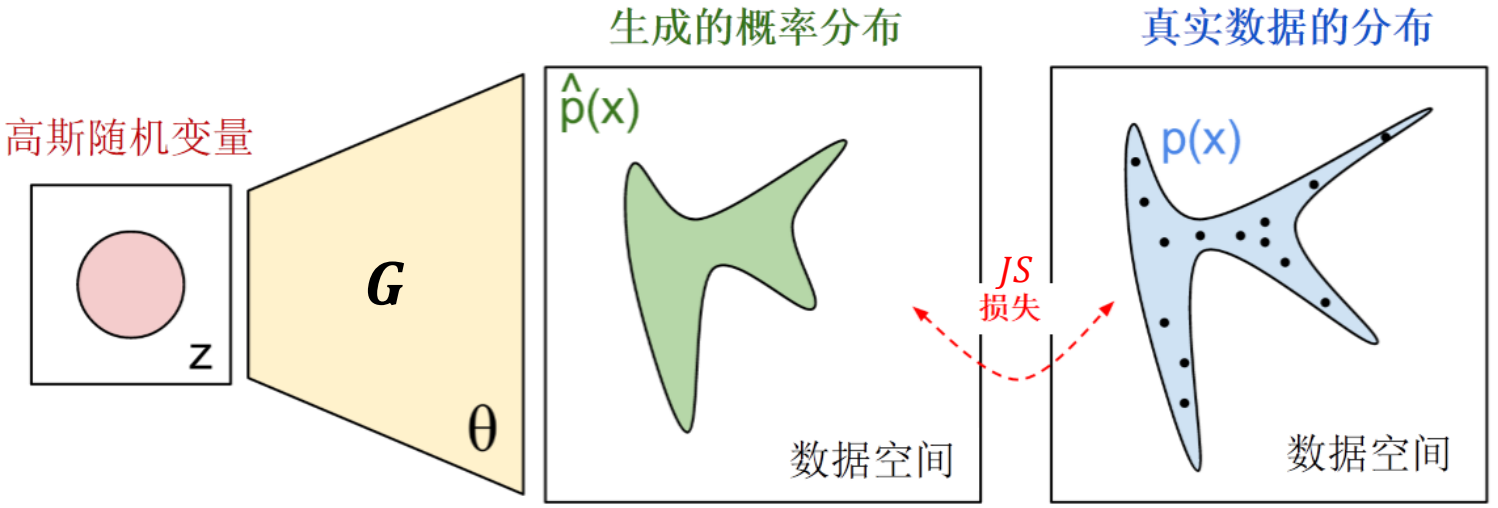
$$\mathbb{E}_{x\sim p_r(x)}\log\frac{0.5\times p_r(x)}{0.5\times(p_r(x)+p_g(x))}+\mathbb{E}_{x\sim p_g(x)}\log\frac{0.5\times p_g(x)}{0.5\times(p_r(x)+p_g(x))}$$

$$=\mathbb{E}_{x\sim p_r(x)}\log\frac{p_r(x)}{0.5\times(p_r(x)+p_g(x))}+\mathbb{E}_{x\sim p_g(x)}\log\frac{p_g(x)}{0.5\times(p_r(x)+p_g(x))}-2\log 2$$

$$=KL(p_r(x)||\frac{p_r(x)+p_g(x)}{2})+KL(p_g(x)||\frac{p_r(x)+p_g(x)}{2})-2\log 2$$

$$=2JS(p_r(x)||p_g(x))-2\log 2 \tag{5}$$

Generative Adversarial Networks



$$2JS(p_r(x)||p_g(x)) - 2\log 2 \tag{5}$$

- 可以看到，随着交替优化的进行，判别器D会逐渐接近最优判别器D^*。当这个近似达到一定程度时，生成器G的loss可以近似等价于最小化真实分布与生成器生成分布之间的$JS(Jensen-Shannon)$散度。也就是说：GAN的思想始于博弈论中的零和博弈启发，而等价于数据概率分布的距离优化。

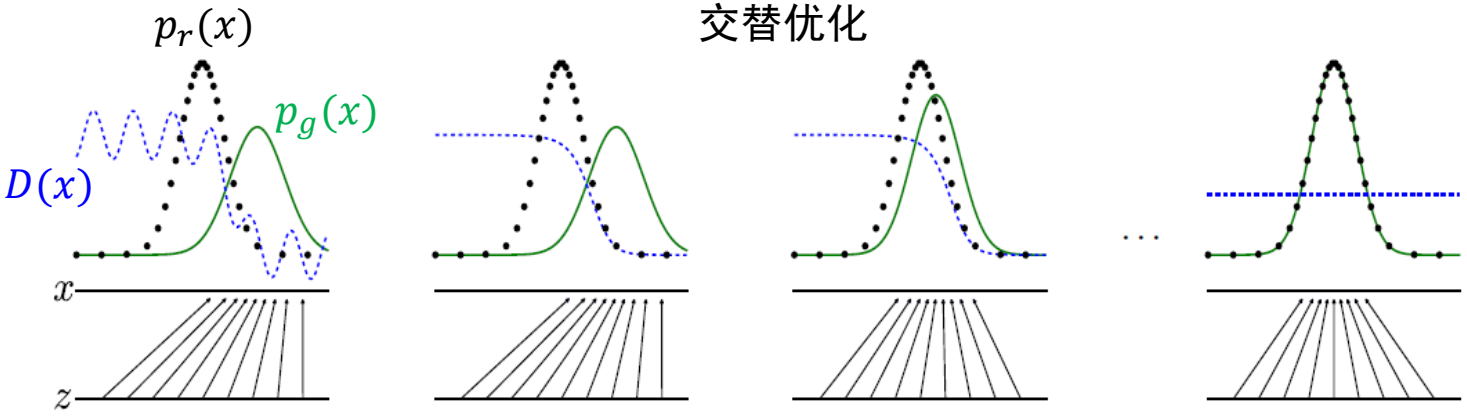
Generative Adversarial Networks

- 从训练的角度看，判别器D起到一个二分类器的作用，每一次对D的更新都在增强D区分真实数据与生成数据的能力（即：为两种数据正确分配两种标签），也就是在两种数据间划分正确的决策边界。而G的更新则试图让生成数据也被分类为真实数据，从而使得新的生成数据更加接近决策边界与真实数据。随着交替迭代的不断进行，生成数据会不断接近真实数据，从而最终能够使得D很难再区分，以很高的真实度来拟合真实数据。

生成器G：将随机向量z映射到数据空间，最大化生成样本被打上真实标签的概率。

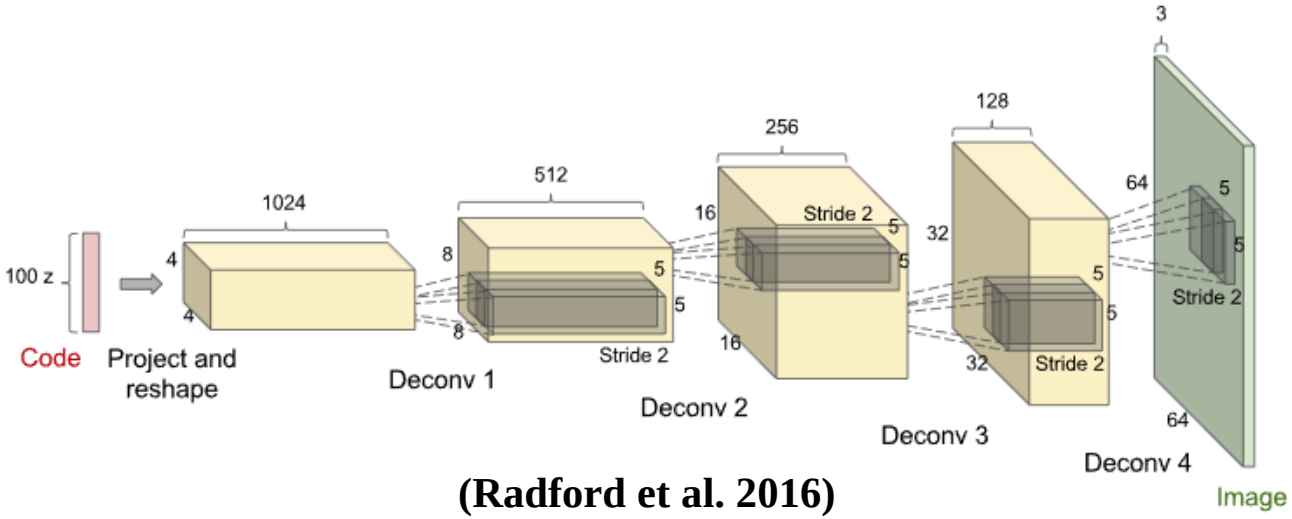
判别器D：最大化分配给真实数据与生成数据正确标签的概率。

交替优化



Generative Adversarial Networks

- 一开始提出的朴素GAN在网络结构上是通过以多层全连接网络为主体的多层感知机(MLP)实现的，然而其调参难度大，训练失败相当常见，生成的图片质量也相当不佳，尤其是对较复杂的数据集而言。



(Radford et al. 2016)

- 因此判别式模型发展的成果被引入到了生成模型中，称作深度卷积对抗神经网络(DCGAN)。DCGAN结构虽然没有带来理论上的解释性，但其强大的图片生成效果使其成为GAN训练中使用非常广泛的网络结构。

报告提纲

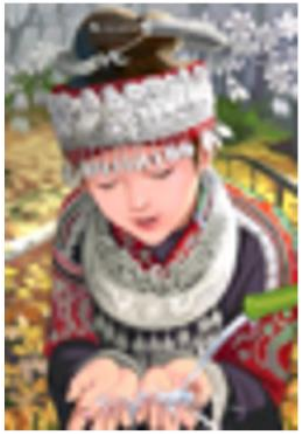
- 生成式模型的定义
- 研究生成式模型的意义
- 经典的生成式模型：VAE 与 GAN
- GAN的应用
- VAE 与 GAN 仍需解决的问题
- 我们的工作
- 总结

GAN的应用

- 相较于VAE在文本等离散数据领域的大放异彩，GAN则在发展更为成熟的CV领域获得了更多的应用机会。一方面的原因是因为GAN生成的图像相比VAE生成的图像有更好的视觉效果。另一方面，还有一个非常重要的原因！

- 为了能够更清楚地说明这个概念，我们举图像超分辨率为例子。

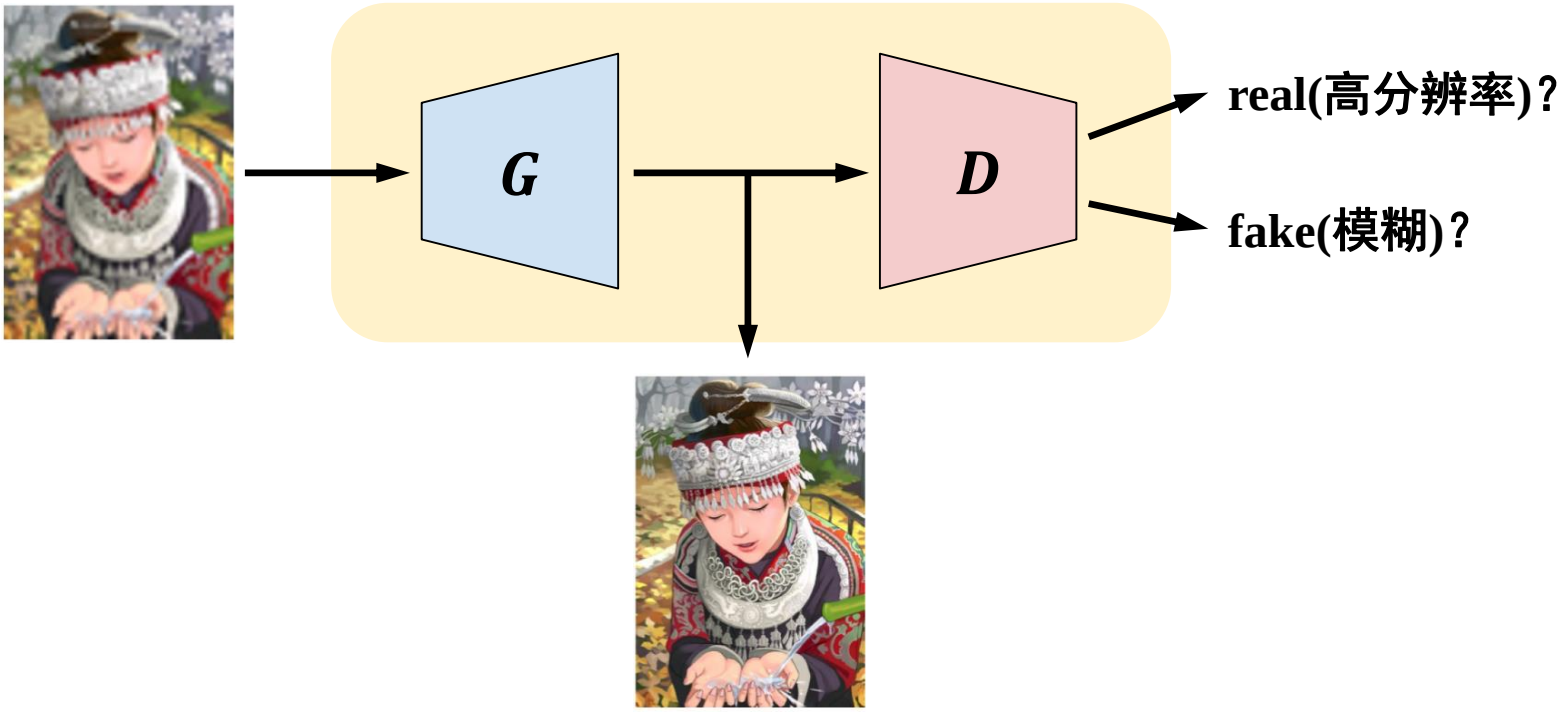
模糊图像



高分辨率图像

GAN的应用

- GAN的判别器可以看作一个二分类器：给真实数据打上“真”标签，给生成数据打上“假”标签。那么，这个所谓的“假”数据的定义真的是一尘不变的吗？

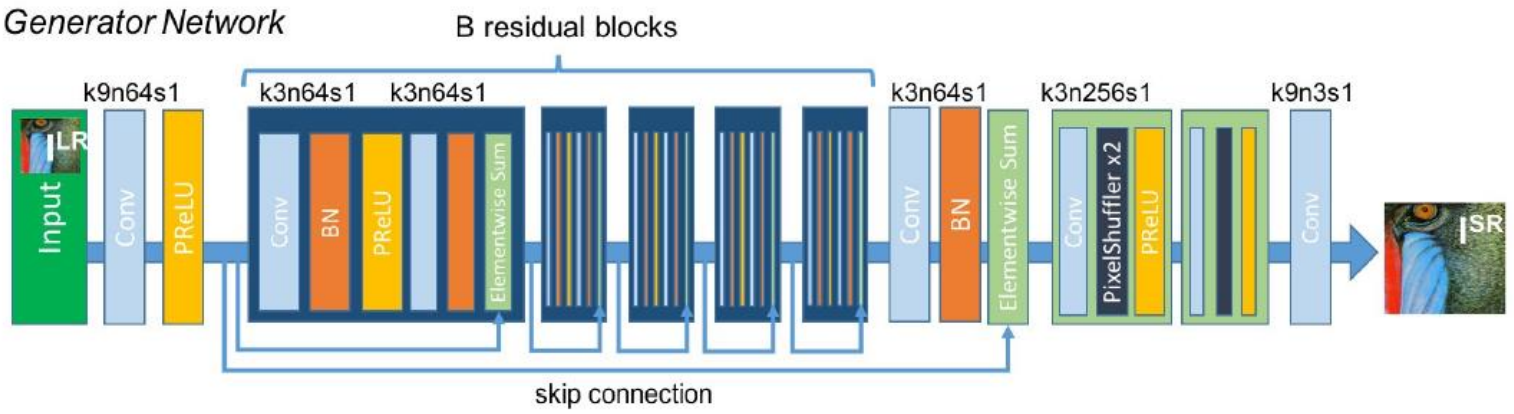


Ledig, Christian, et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network."*IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*.

GAN的应用

- 图像超分辨率
  - **生成网络**---生成低分辨率图片对应的超分辨率图片

低分辨率图片

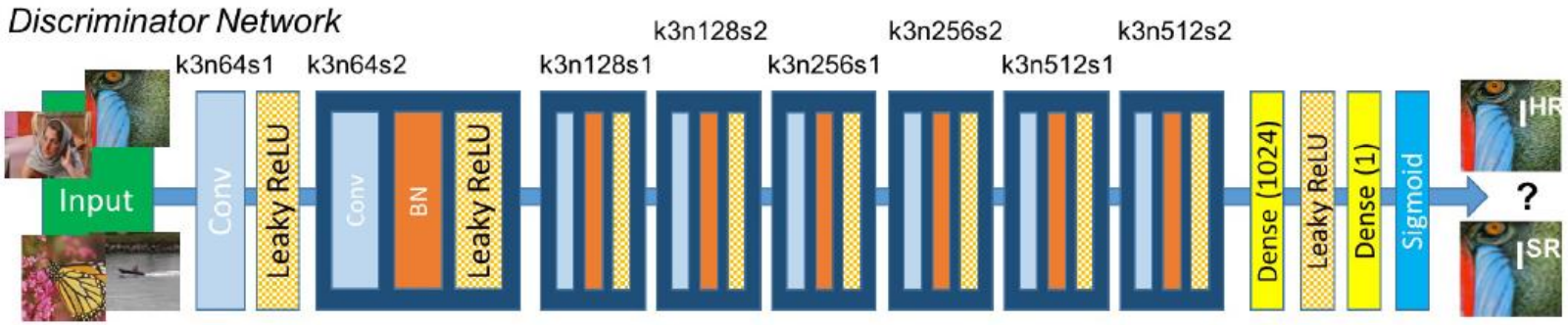


超分辨率图片

Ledig, Christian, et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network."*IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*.

GAN的应用

- **判别网络---发现生成的与真实的高分辨率图片之间的区别**



高分辨率？
超分辨率？

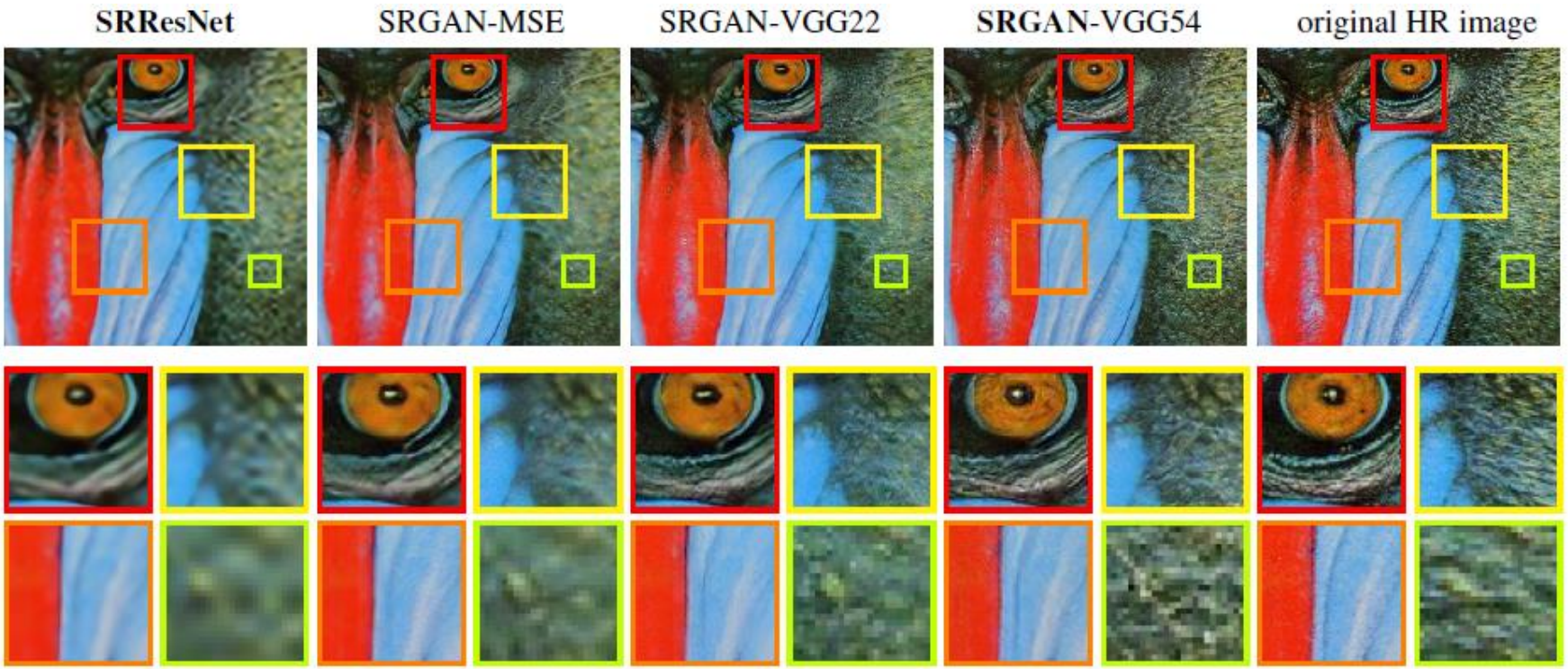
- ✓ content loss:

$$l^{SR}_{VGG/i.j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

- ✓ adversarial loss:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$$

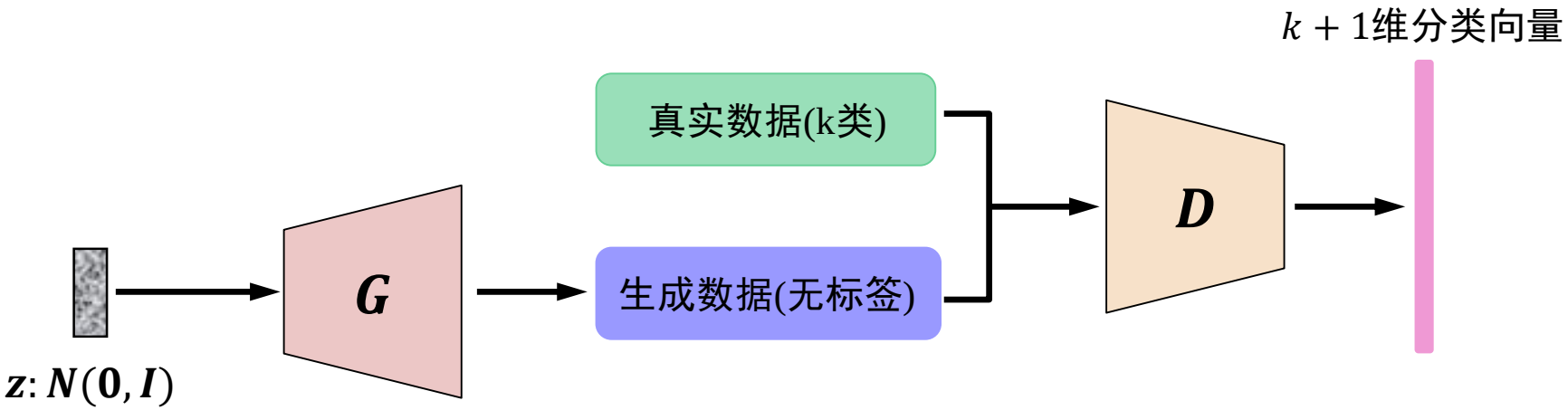
GAN的应用



图片超分辨率效果比较

GAN的应用

- 半监督GAN



$$\mathcal{L} = \mathcal{L}_{supervised} + \mathcal{L}_{unsupervised}$$

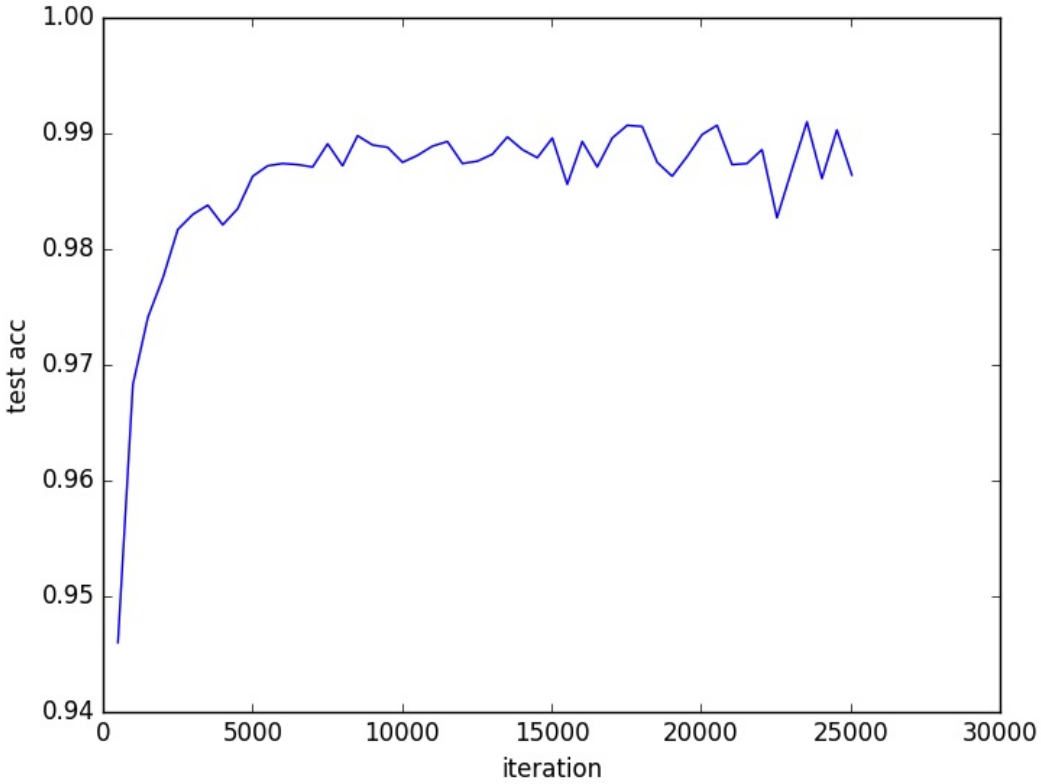
$$\min_D \mathcal{L}_{supervised} = -\mathbb{E}_{(x,y)\sim p_{data}(x,y)} \log p_{model}(y|x, y < k+1)$$

$$\max_G \min_D \mathcal{L}_{unsupervised} = -\mathbb{E}_{x\sim p_{data}(x)} \log[1 - p_{model}(y = k+1|x)] - \mathbb{E}_{x\sim p_g} \log p_{model}(y = k+1|x)$$

Salimans, Tim, et al. “Improved techniques for training gans.” *Advances in Neural Information Processing Systems(NIPS). 2016.*

GAN的应用

- 利用半监督GAN，可以在仅使用200个带标签的MNIST图像的情况下，实现在1万张测试图片上99%的测试精度。这是用其他的半监督方法很难做到的。



半监督GAN在MNIST测试集上的表现

GAN的应用

- 图片文本标注（image caption）

  - 传统的训练方法要求模型预测的每个词向量与训练集的ground truth完全一致，这会导致模型生成的标注过于单一（对同一张图片不能给出不同描述风格的标注；甚至对于相似的图片只给出完全一样的标注）。

Baseline: a man riding skis down a snow covered slope

Shetty, Rakshith, et al. "Speaking the same language: Matching machine to human captions by adversarial training." *IEEE International Conference on Computer Vision (ICCV). 2017.*

GAN的应用

- 引入GAN作对抗训练，让判别网络自己学习文本与图片之间的联系。
- 判别网络：
  - ✓ $$L(D) = -\log\left(D(S_p^r, x)\right) - \log\left(1 - D(S_p^g, x)\right) - \log\left(1 - D(S_p^f, x)\right)$$
- 生成网络：
  - ✓ $$L(G) = -\log\left(D(S_p^g, x)\right) + \|\mathbb{E}\left[dist_x(S_p^g, x)\right] - \mathbb{E}\left[dist_x(S_p^r, x)\right]\|_2$$
- (S_p^r, x)表示训练集中相互对应的数据对(标注，图片)； (S_p^g, x)表示生成的标注与相应的图片构成的数据对； (S_p^f, x)表示从训练集中随机采样得到的没有对应关系的数据对；$dist_x()$表示判别网络中间层的feature map。

GAN的应用

Ours: a person on skis jumping over a ramp

Ours: a skier is making a turn on a course

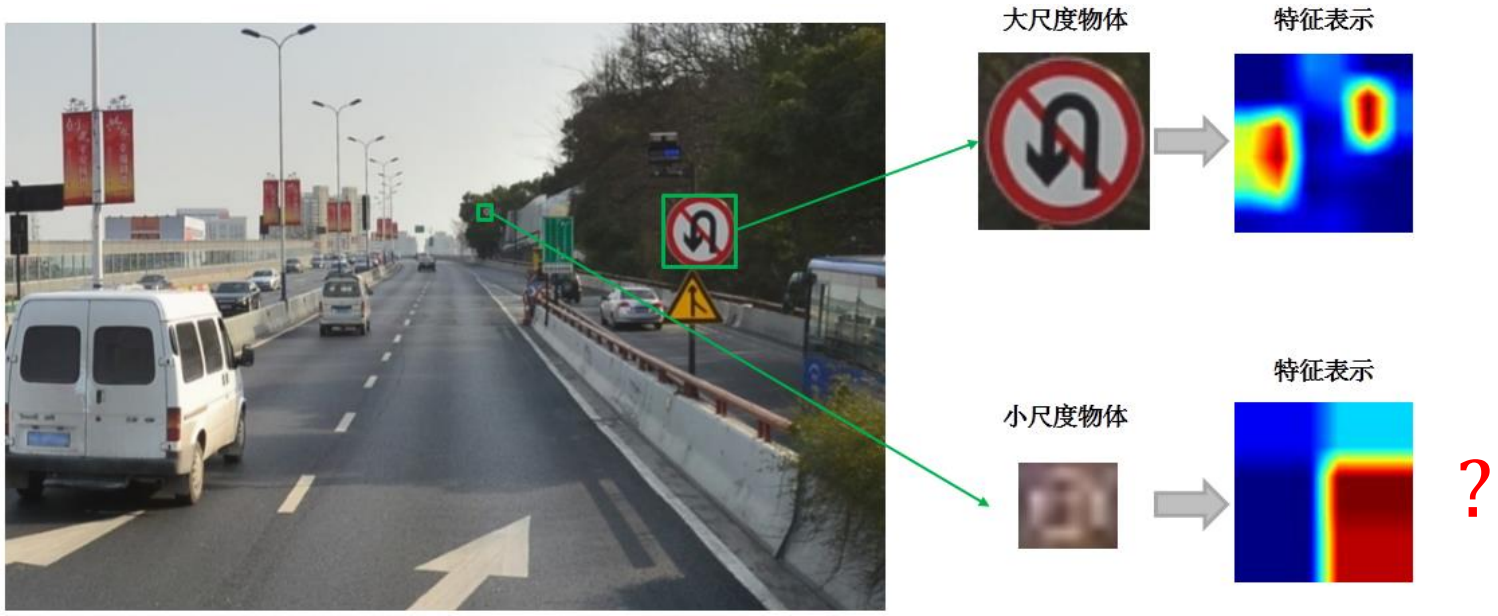
Ours: a cross country skier makes his way through the snow

Ours: a skier is headed down a steep slope

Ours	a bus that has pulled into the side of the street a bus is parked at the side of the road a white bus is parked near a curb with people walking by	a group of people standing outside in a old museum an airplane show where people stand around a line of planes parked at an airport show
Base line	* a bus is parked on the side of the road * a bus that is parked in the street a bus is parked in the street next to a bus	a group of people standing around a plane a group of people standing around a plane a group of people standing around a plane

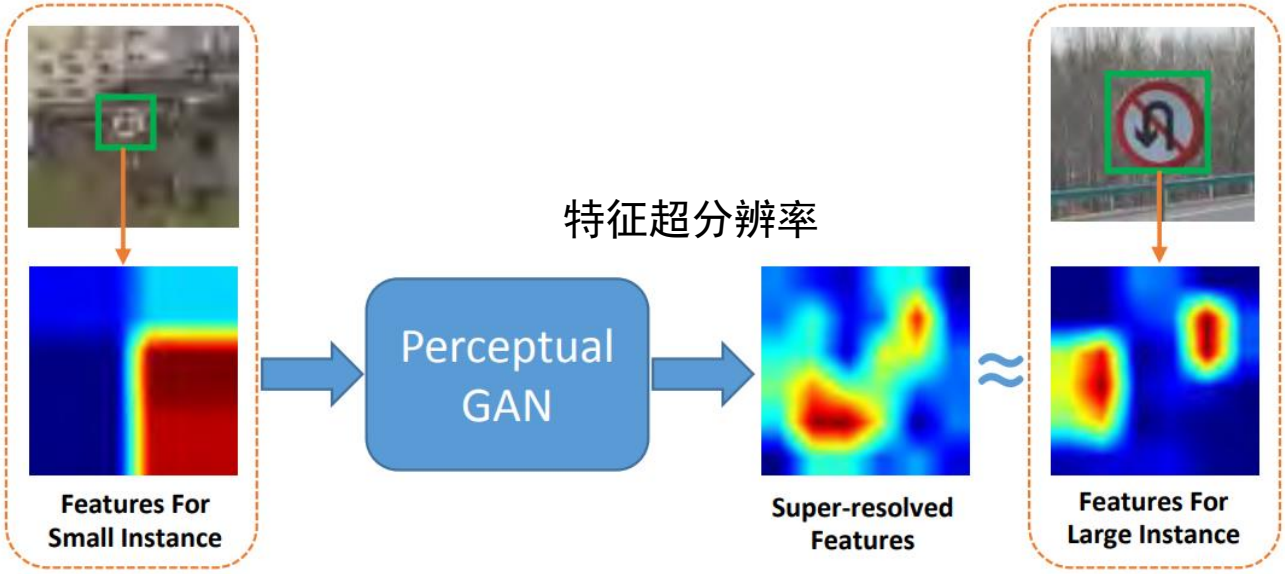
GAN的应用

- 小物体检测



Li, Jianan, et al. "Perceptual generative adversarial networks for small object detection."*IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*.

GAN的应用

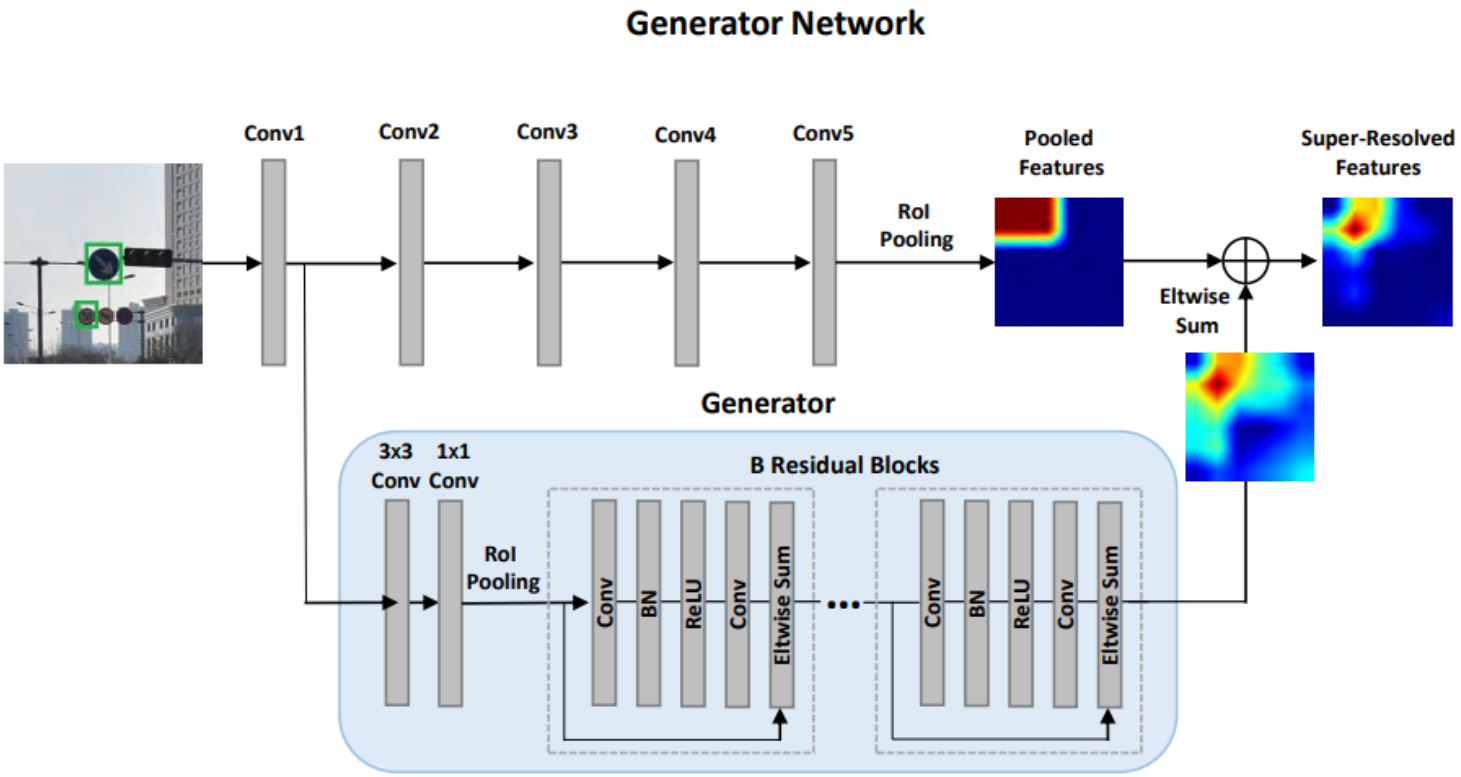


小尺度物体提取到的特征

大尺度物体提取到的特征

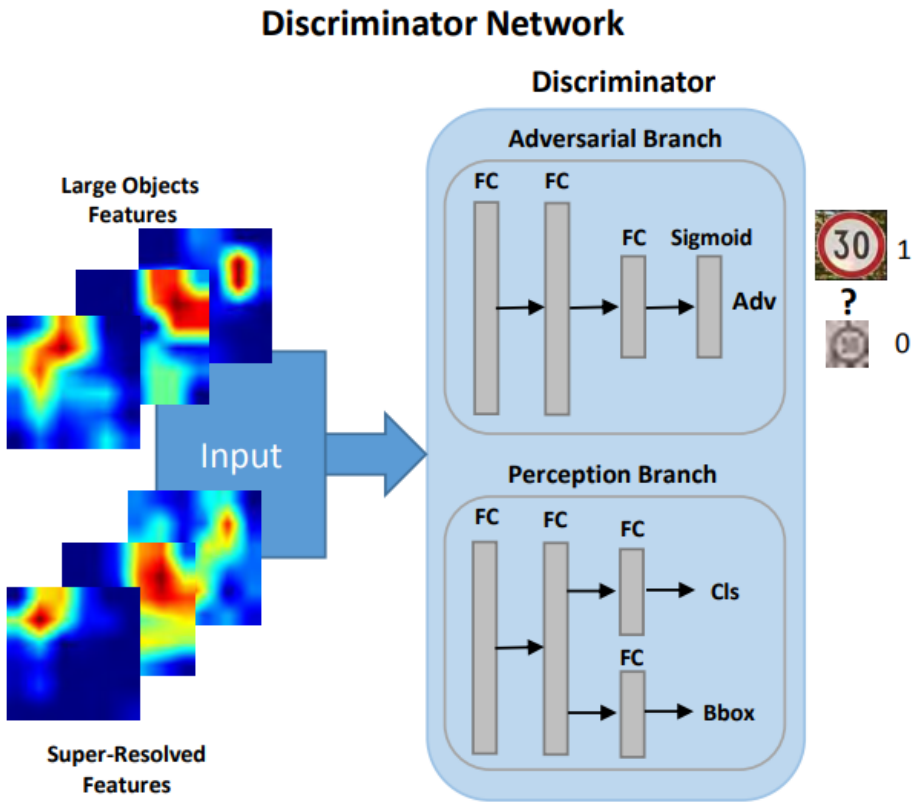
GAN的应用

- 生成网络结构：



GAN的应用

- 判别网络结构：



GAN的应用

- 优化目标

  - ✓ Adversarial loss: $\min_G \max_D L(D,G) \triangleq \mathbb{E}_{F_l \sim p_{\text{data}(F_l)}} \log D(F_l) + \mathbb{E}_{F_s \sim p_{F_s}(F_s|f)}[\log(1 - D(\underbrace{F_s + G(F_s|f)}_{\text{residual learning}}))]$.

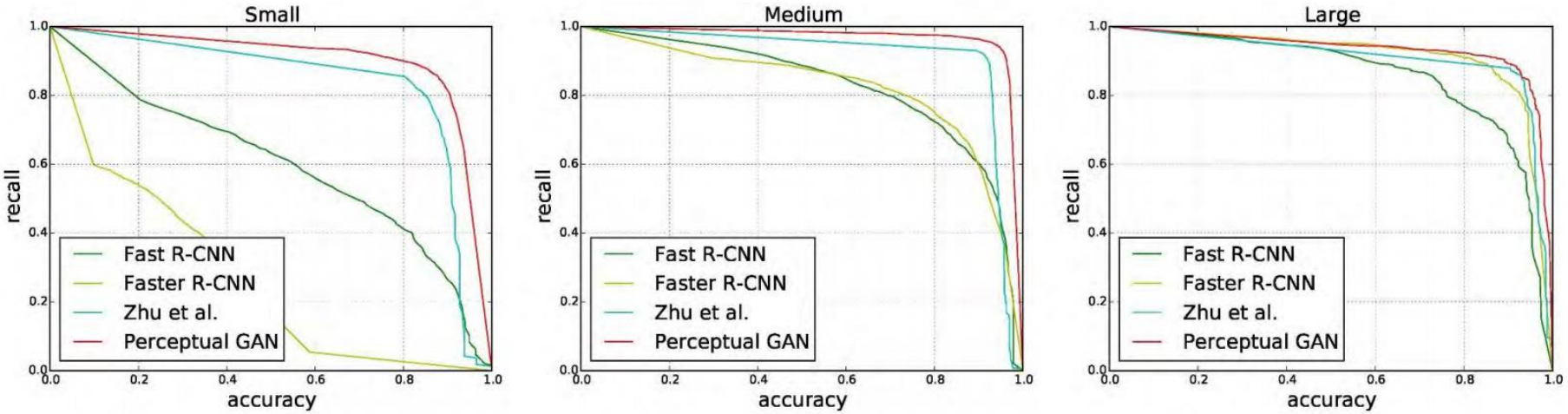
    - ➢ 其中，F_l为大尺度物体的特征，F_s为小尺度物体的特征，$F_s|f$为小尺度物体输入特征提取网络得到的浅层细粒度特征。

  - ✓ perceptual loss: $L_{dis\_p} = L_{cls}(p,g) + \mathbf{1}[g \geq 1] L_{loc}(r_g, r^*)$

    - ➢ 其中，L_{cls}为检测区域分类的loss，L_{loc}为检测区域中物体的bounding-box坐标计算loss，$1[g \geq 1]$表示坐标计算的loss只对前景目标有效，g代表类别标签，对背景而言$g = 0$。

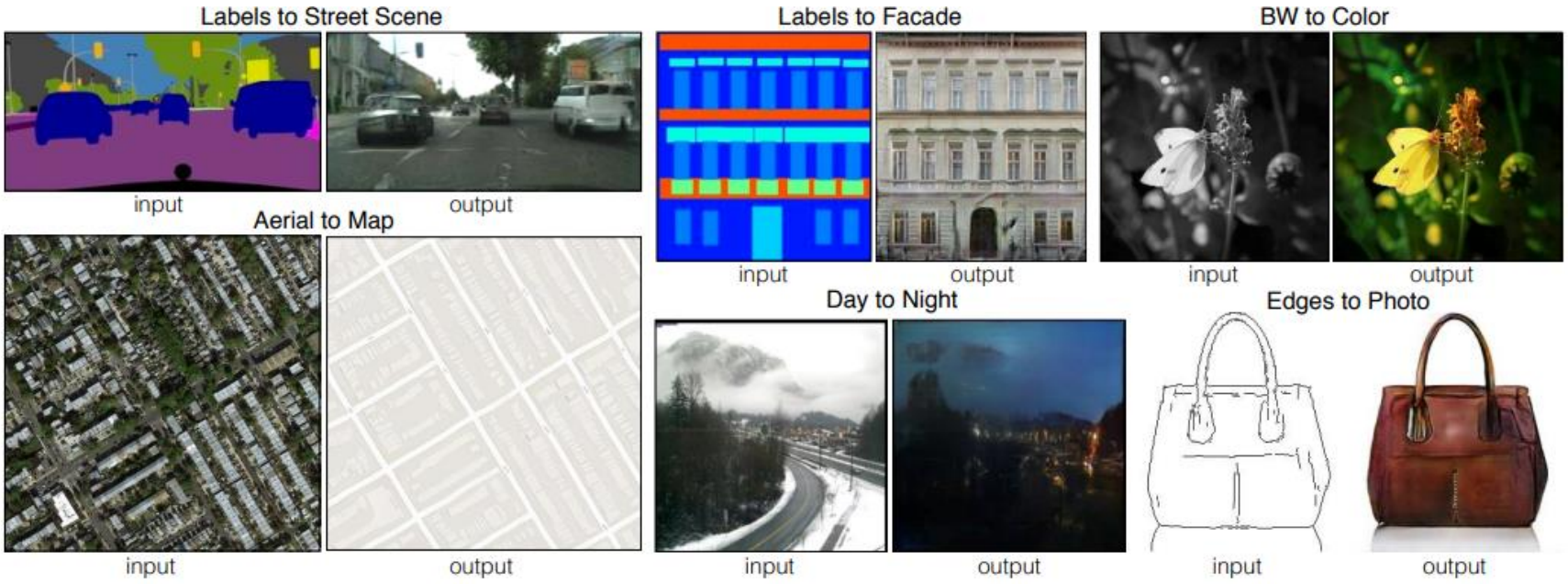
GAN的应用

- 针对交通标志的检测任务，在Tsinghus-Tencent 100K数据集上，Perceptual GAN大幅提升了小物体的检测性能



GAN的应用

- 图像→图像的翻译



Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks."*IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*.

GAN的应用

Zhu, Jun Yan, et al. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks." International *Conference on Computer Vision (ICCV), 2017*.
Wang, Ting-Chun, et al. "High-resolution image synthesis and semantic manipulation with conditional gans." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018*.

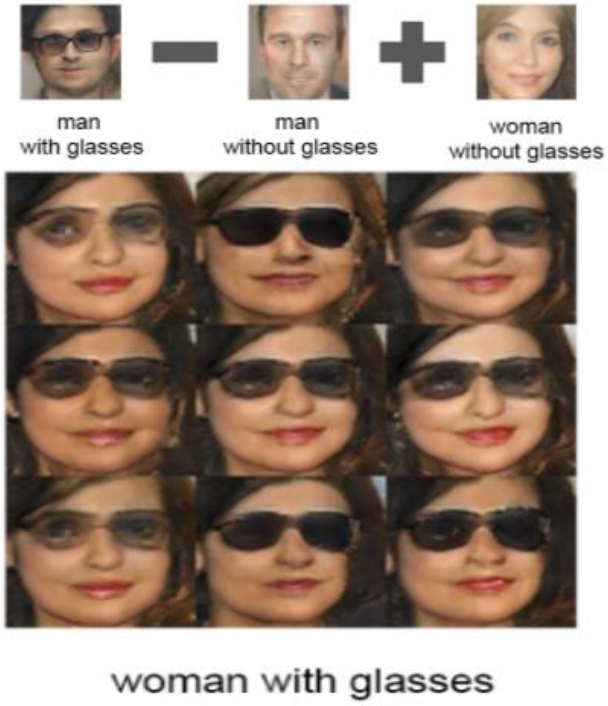
GAN的应用

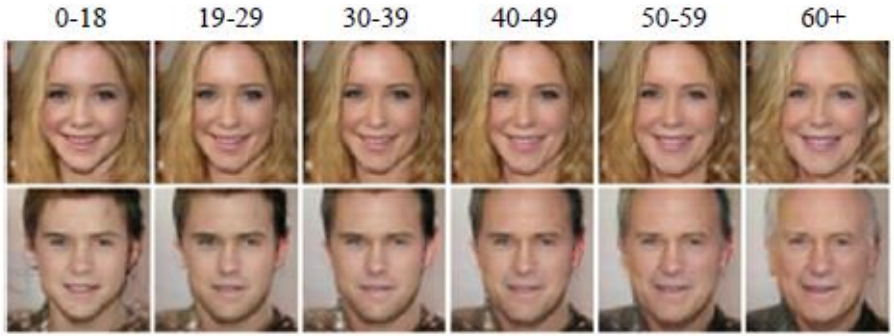
多样化的转化结果

Zhu, Jun Yan, et al. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks." International *Conference on Computer Vision (ICCV), 2017*.
Wang, Ting-Chun, et al. "High-resolution image synthesis and semantic manipulation with conditional gans." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018*.

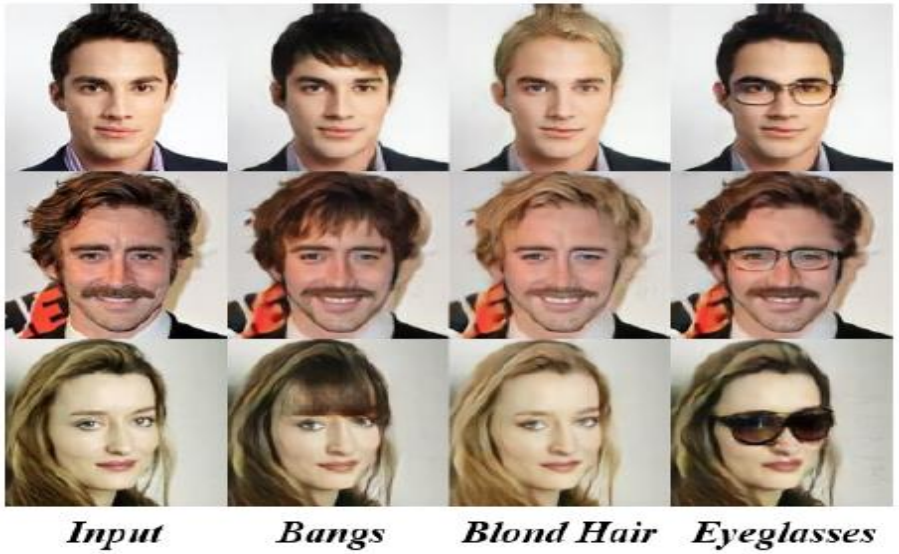
GAN的应用

- 人脸属性变换





Face Aging



Radford, et al. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv, 2016*.

Antipov, et al. "Face aging with conditional generative adversarial networks." *IEEE International Conference on Image Processing (ICIP), 2017*.

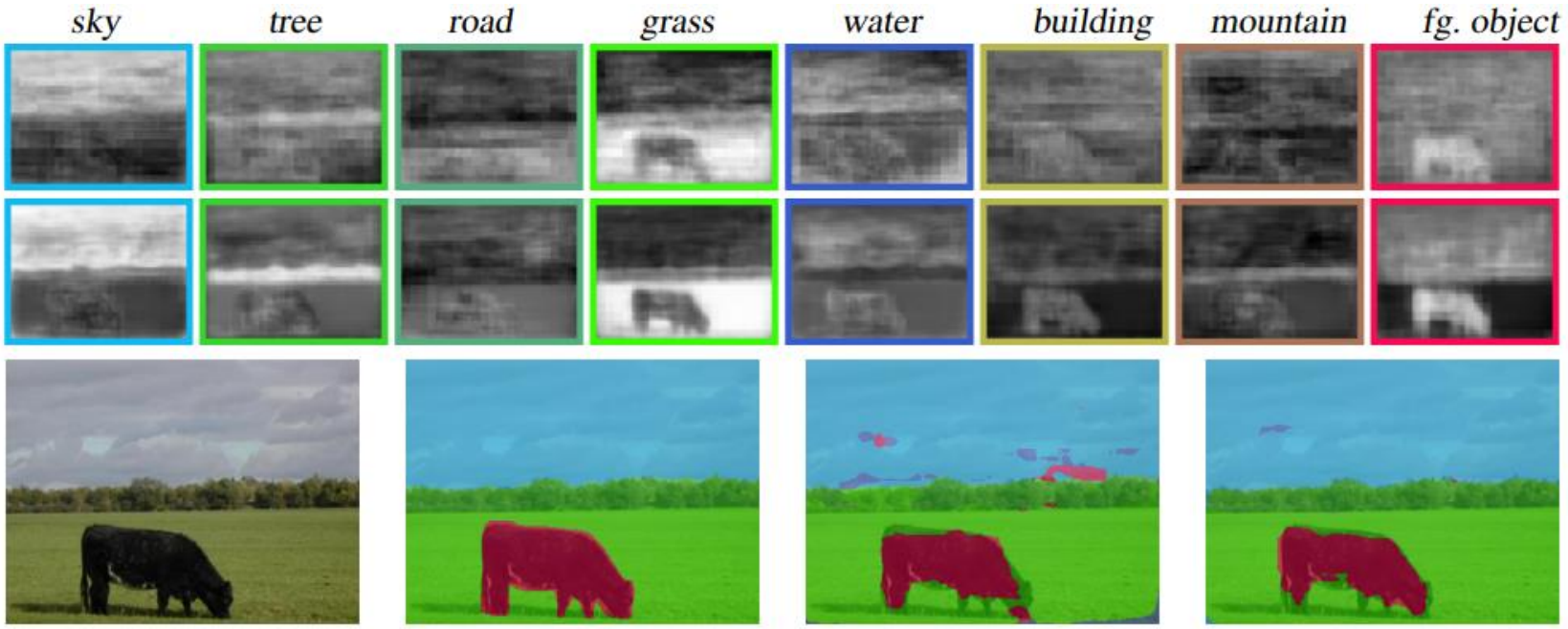
GAN的应用

- 基于文本的图片生成

Zhang, Han, et al. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks."*IEEE International Conference on Computer Vision (ICCV). 2017.*

GAN的应用

- 图像语义分割



Luc, Pauline, et al. "Semantic Segmentation using Adversarial Networks.". *Neural Information Processing Systems (NIPS), 2012*.

GAN的应用

- 视频未来帧预测

Ground truth

ℓ_2 result

Adversarial result

Ground truth

ℓ_2 result

Adversarial result

Adversarial result

l_2 result

Mathieu, Michael, Camille Couprie, and Yann LeCun. “Deep multi-scale video prediction beyond mean square error.” *arXiv*, 2015.

GAN的应用

- 图像去雨滴








with raindrop　raindrop removed　with raindrop　raindrop removed

Qian, Rui, et al. "Attentive generative adversarial network for raindrop removal from a single image." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.*

报告提纲

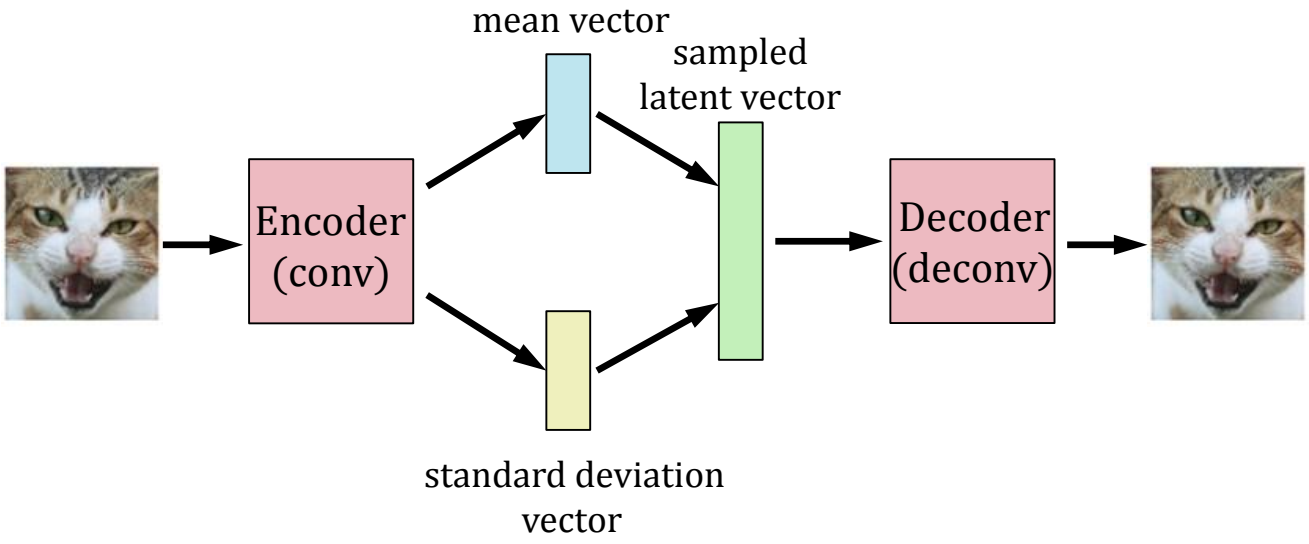
- 生成式模型的定义
- 研究生成式模型的意义
- 经典的生成式模型：VAE 与 GAN
- GAN的应用
- VAE 与 GAN 仍需解决的问题
- 我们的工作
- 总结

VAE的问题

- VAE的优点：

  - VAE的训练过程比较稳定，Loss函数的数值会呈现一个相对稳定的下降趋势；

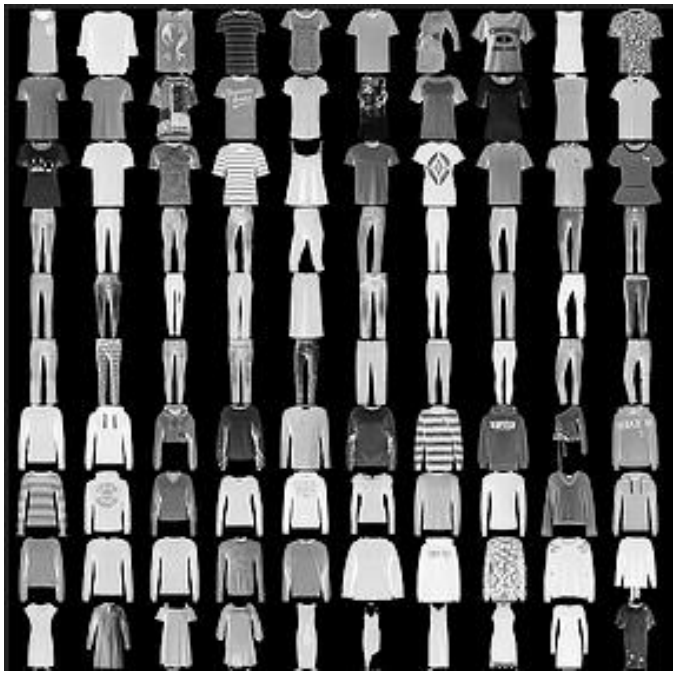
  - 在训练完成后，VAE的编码器*Encoder*与解码器*Decoder*都可以近似看作理想的高维单映射函数（不同的输入得到不同的输出）。这样，只要隐变量的采样足够多样，就可以保证生成样本的多样性。



VAE的问题

- VAE的缺点：

  - VAE所生成的图片会比较模糊。



训练图片

VAE随机生成

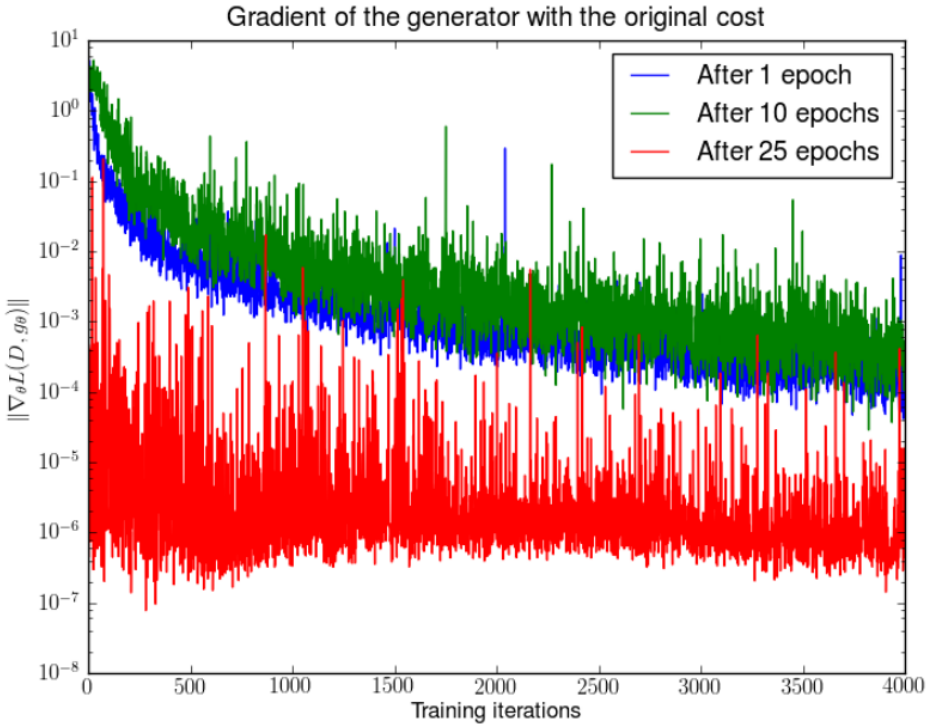
GAN的问题

- GAN的优点：
  - 模型优化只用到了反向传播，而不需要计算马尔可夫链；
  - 模型的训练不需要对隐变量做推断；
  - 在大多数情况下被认为是最好的图片生成模型；
  - 基于博弈的思想，判别器作为可自主学习的度量函数，定义十分灵活，使得GAN很容易与其他任务结合起来做对抗训练。

GAN的问题

- GAN的缺点：

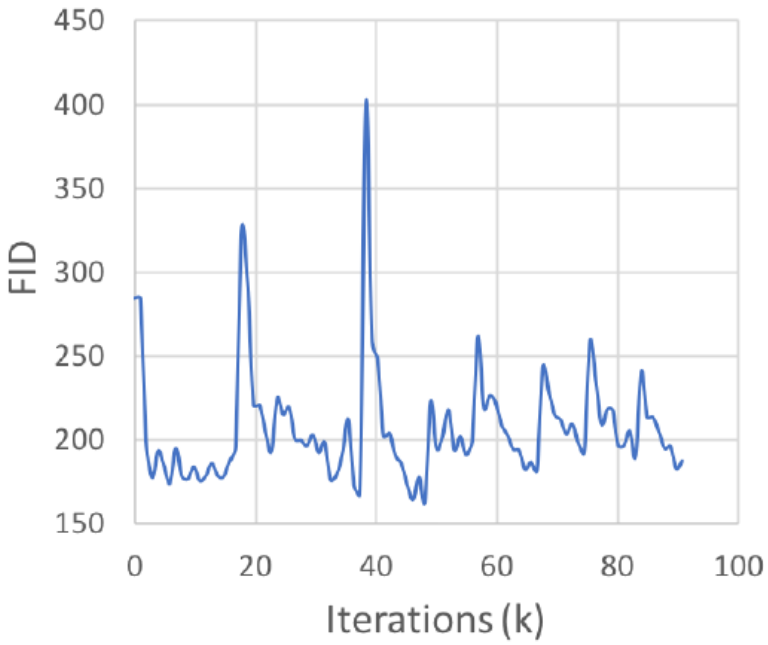
  - 原始GAN训练的过程中会出现梯度消失的现象，使得生成器G几乎无法获得有效的学习信息，因此很难收敛；

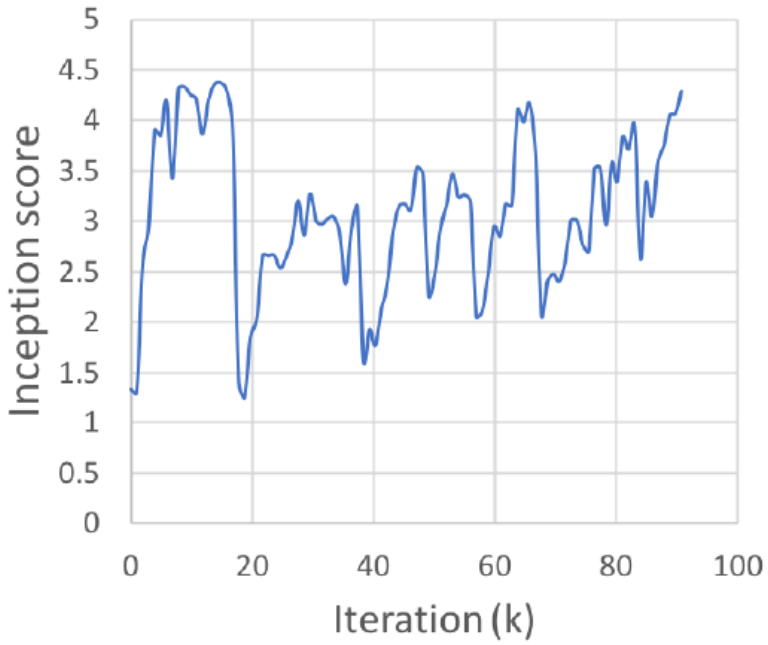


(Arjovsky et al. 2017)

GAN的问题

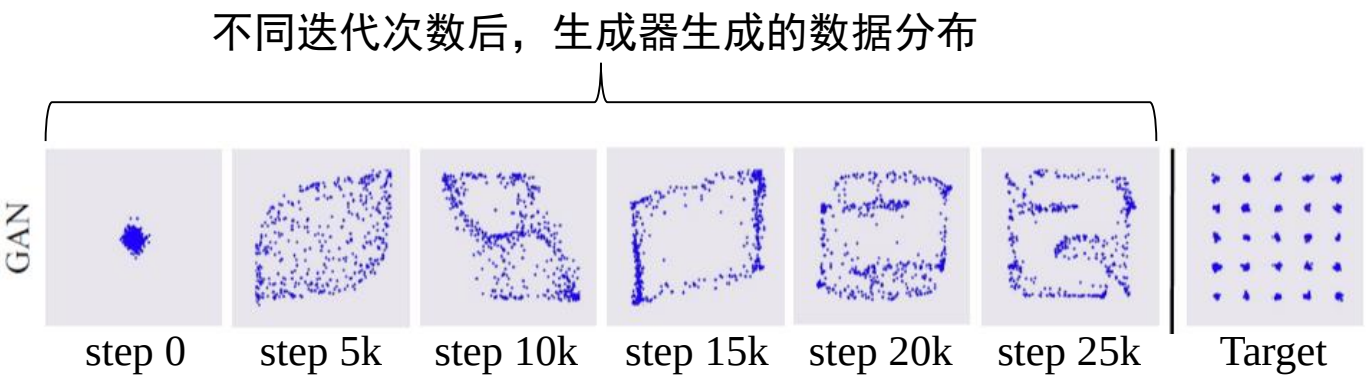
- 原始GAN的训练很不稳定，判别器D与生成器G之间需要很好的同步，例如： D每更新k次，G更新一次；





GAN的问题

- 原始GAN训练完成后，生成的样本会出现模式丢失（**mode collapse**）的问题。

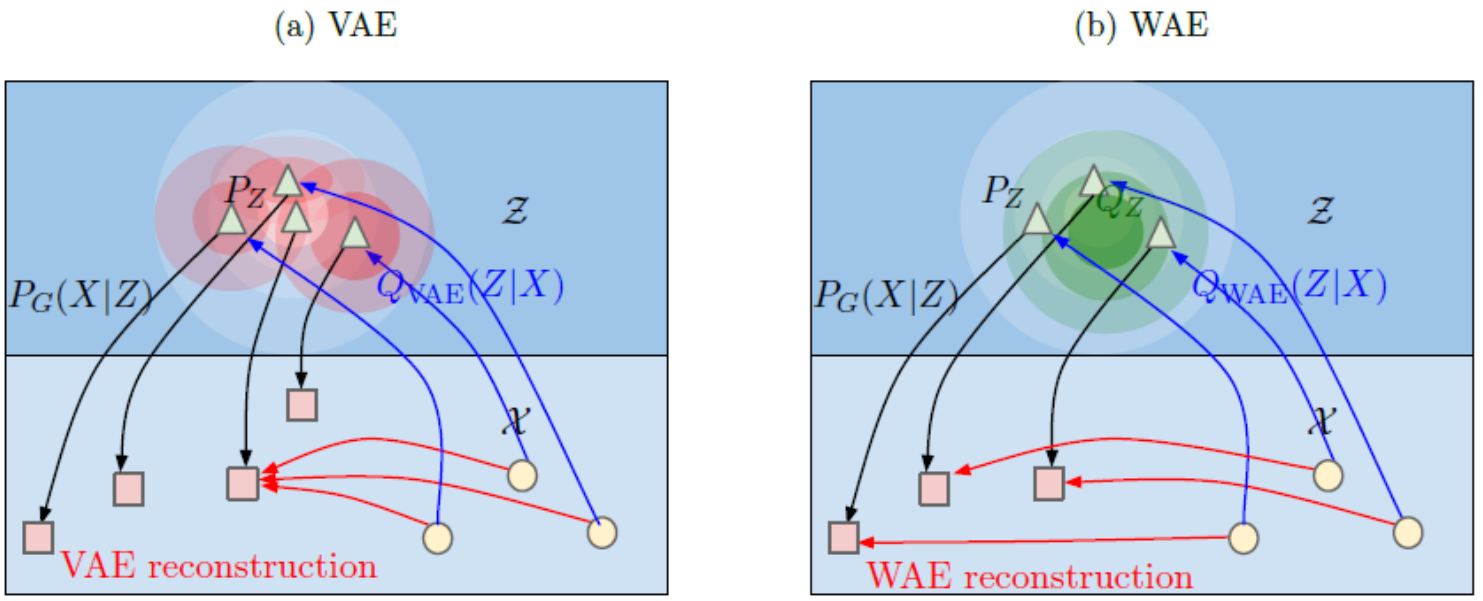


生成图片缺乏多样性

Wasserstein Auto-Encoders

- 通过改变VAE隐空间的约束形式来改善VAE生成模糊的问题。优化目标如下：

$$D_{\mathrm{WAE}}(P_X, P_G) := \inf_{Q(Z|X)\in\mathcal{Q}} \mathbb{E}_{P_X}\mathbb{E}_{Q(Z|X)}\big[\overbrace{c\big(X, G(Z)\big)}^{\text{MSE}}\big] + \lambda \cdot \overbrace{\mathcal{D}_Z(Q_Z, P_Z)}^{\text{GAN度量}}.$$



Tolstikhin, Ilya, et al. "Wasserstein auto-encoders." *International Conference on Learning Representations (ICLR). 2018*.

Wasserstein Auto-Encoders

- VAE与WAE-GAN在CelebA数据集上的生成效果如下：

VAE

WAE-GAN

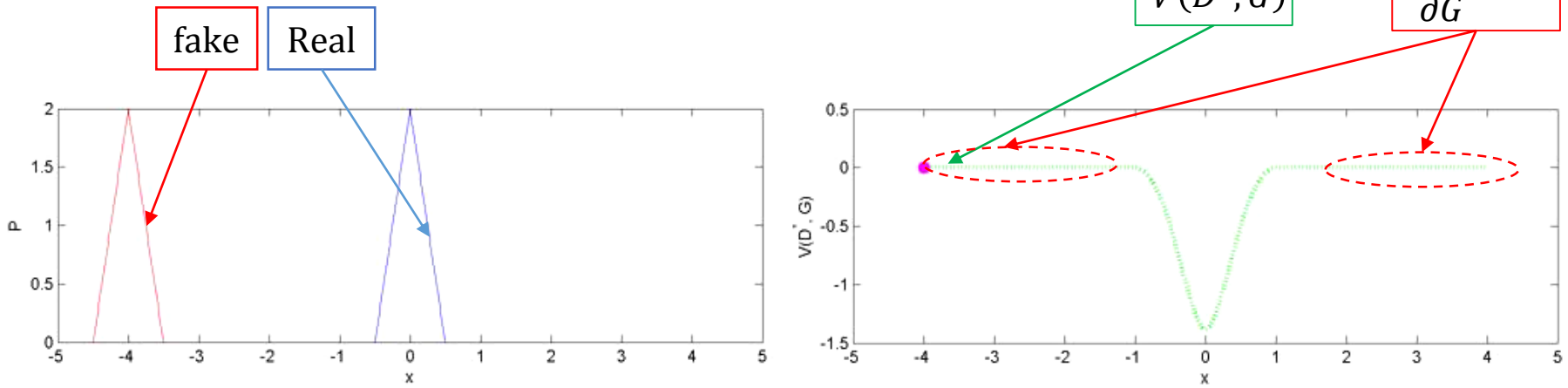
Wasserstein GAN

- 我们在前面的分析提到过，GAN的优化目标等价于最小化生成分布与目标分布之间的JS散度。WGAN这篇论文指明了JS散度优化是导致梯度消失的主要原因，并提出了一个替代选择：Wasserstein距离，用以解决GAN训练不稳定的问题。对于最优判别器D^*，GAN等价于优化：

$$\mathcal{V}(D^*, G) = 2JS\left(\mathbb{P}_r || \mathbb{P}_g\right) - 2log2$$

梯度消失

$$\frac{\partial V}{\partial G} = 0$$



Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein gan." International Conference on Learning Representations (ICLR). *2017*.

Wasserstein GAN

- 作者提出了Wasserstein距离，定义如下：

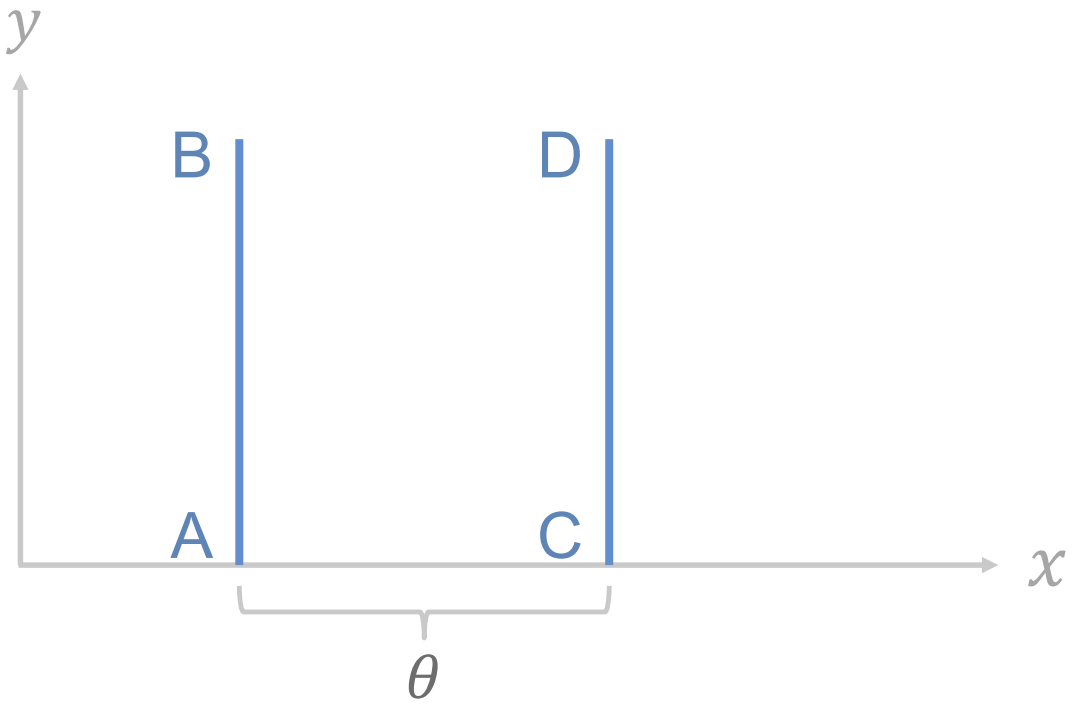
$$\mathcal{W}(P_r, P_g) = \inf \mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|], \gamma \sim (P_r, P_g)$$

  - 其中，(P_r, P_g)是P_r和P_g组合起来的所有可能的联合分布的集合。对于每一个可能的联合分布γ而言，可以从中采样$(x, y)\sim\gamma$得到一个真实样本x和一个生成样本y，并算出这对样本的距离。然后，可以计算该联合分布γ下样本对距离的期望值$\mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|]$，在所有可能的联合分布中对这个期望值取下界，就定义为Wasserstein距离。

- Wasserstein距离相比JS散度的优越性在于，即便两个分布没有重叠，Wasserstein距离仍然能够反映它们的远近。

  - 作者举了一个简单的例子：考虑二维空间中的两个分布P_1和P_2，P_1在线段AB上均匀分布，P_2在线段CD上均匀分布，通过控制参数θ可以控制两个分布的距离远近。

Wasserstein GAN



- 则：

$$\mathcal{W}(P_1, P_2) = |\theta| \quad \text{(平滑的)}$$

$$JS(P_1||P_2) = \begin{cases} \log 2 & if\ \theta \neq 0 \\ 0 & if\ \theta = 0 \end{cases} \quad \text{（突变的）}$$

Wasserstein GAN

- 由于Wasserstein距离难以求解，作者对其进行了变形（证明过程可以看论文的附录）：

$$\mathcal{W}(P_r, P_g) = \max_{\|f\|_L \leq K} \mathbb{E}_{x \sim P_r}[f(x)] - \mathbb{E}_{x \sim P_g}[f(x)]$$

  - 其中，函数f要求满足Lipschitz连续，也就是要求存在一个常数$K \geq 0$使得定义域内的任意两个元素x_1和x_2都满足：

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$$

  - 也就是说，对任意的输入，f的梯度都是有界的。

- 我们可以把f看作一个神经网络，但是，要如何限制其梯度有界呢？

Wasserstein GAN

- 作者给出了一个简单使用的方法：就是限制网络的所有参数ω_i不超过范围$[-c, c]$（weight clipping），这样网络关于输入样本x的梯度就是有界的，就可以满足Lipschitz连续。

- 到此，整个模型的思路就是：构造一个判别器网络f_ω，在限制所有参数ω不超过$[-c, c]$的前提下使得：

$$\mathcal{L} = \mathbb{E}_{x \sim P_r}[f_\omega(x)] - \mathbb{E}_{x \sim P_g}[f_\omega(x)]$$

尽可能取到最大，此时$\mathcal{L}$就会近似等于真实分布于生产分布之间的Wasserstein距离。接下来生成器要最小化近似的Wasserstein距离，也就是最小化$\mathcal{L}$，通过这样的方式，生成器与判别器以对抗学习的方式进行交替训练。

- 这就是WGAN的训练思路与训练过程。

Wasserstein GAN

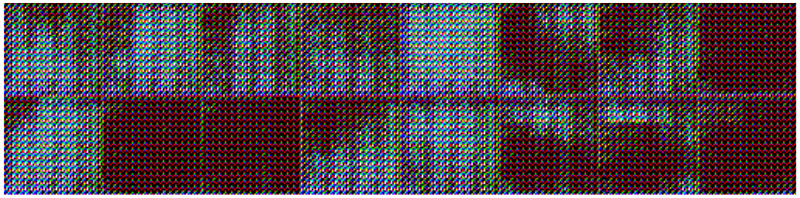
WGAN

with BN

without BN

DCGAN

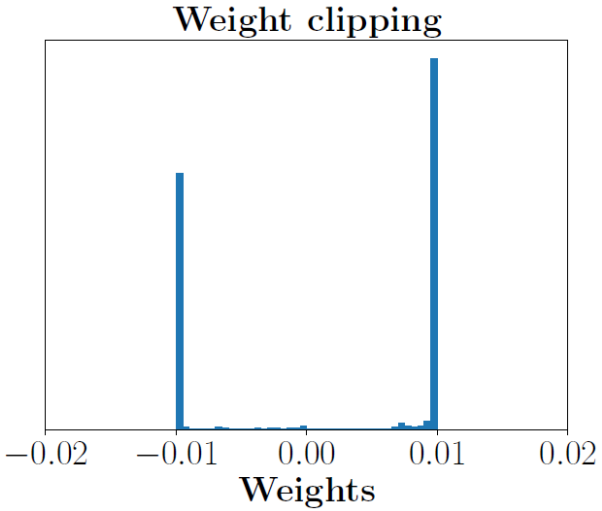
with BN



without BN

WGAN-GP

- WGAN显著地提高了模型训练的稳定性，但它也有缺陷。在WGAN中，为了使神经网络f_ω满足Lipschitz连续而使用了weight clipping的操作，但实际上，这也是有问题的。



- weight clipping会导致网络的权值集中在限制区间$[-c, c]$的边界处，这会导致网络学习到的映射函数过于简单，而限制了整个模型的建模能力与网络的生成能力。

Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." *Advances in Neural Information Processing Systems(NIPS)*. 2017.

WGAN-GP

- 论文的作者提出，我们其实没有必要在整个输入空间上施加Lipschitz限制，只需要重点抓住生成样本集中区域、真实样本集中区域以及夹在它们中间的区域就行了。

- 具体来说，先随机采样一对真假样本，还有一个服从0~1均匀分布的随机数：

$$x_r \sim P_r, x_g \sim P_g, \varepsilon \sim Uniform[0,1]$$

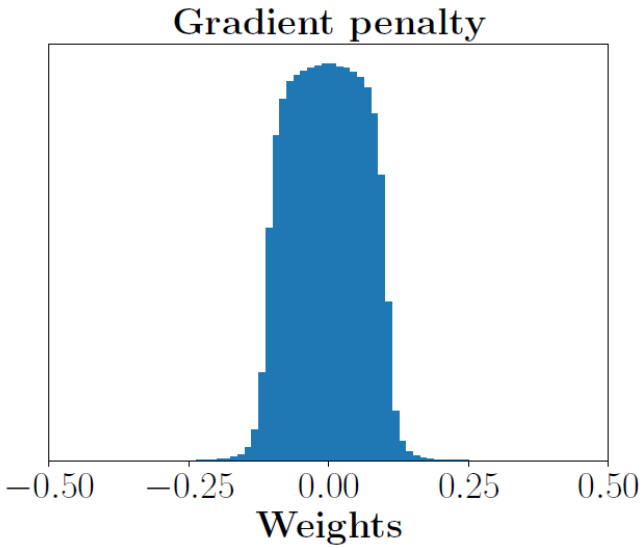
- 然后在x_r和x_g的连线上随机插值采样：

$$\hat{x} = \varepsilon x_r + (1-\varepsilon)x_g$$

- 按照上述流程采样得到的$\hat{x}$所满足的分布记为$P_{\hat{x}}$，就得到最后的目标函数：

$$\mathcal{L} = \mathbb{E}_{x\sim P_g}[D(x)] - \mathbb{E}_{x\sim P_g}[D(x)] + \lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$$

WGAN-GP



Unsupervised

Method	Score
ALI [8] (in [27])	$5.34 \pm .05$
BEGAN [4]	5.62
DCGAN [22] (in [11])	$6.16 \pm .07$
Improved GAN (-L+HA) [23]	$6.86 \pm .06$
EGAN-Ent-VI [7]	$7.07 \pm .10$
DFM [27]	$7.72 \pm .13$
WGAN-GP ResNet (ours)	$7.86 \pm .07$

Supervised

Method	Score
SteinGAN [26]	6.35
DCGAN (with labels, in [26])	6.58
Improved GAN [23]	$8.09 \pm .07$
AC-GAN [20]	$8.25 \pm .07$
SGAN-no-joint [11]	$8.37 \pm .08$
WGAN-GP ResNet (ours)	$8.42 \pm .10$
SGAN [11]	$8.59 \pm .12$

Inception scores on CIFAR-10

报告提纲

- 生成式模型的定义
- 研究生成式模型的意义
- 经典的生成式模型：VAE 与 GAN
- GAN的应用
- VAE 与 GAN 仍需解决的问题
- 我们的工作
- 总结

CVAE-GAN

GAN存在的问题

- 梯度消失
  - 判别器D越好，生成器G梯度消失越严重。
  - 这导致生成式对抗网络非常难以训练。
- 多样性不足
  - 生成器G会在隐空间的不同隐变量上生成相同的结果。

Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua. "CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training." *International Conference on Computer Vision (ICCV)*. 2017.

CVAE-GAN

解决梯度消失的问题

- 我们的方法：非对称训练，对D和G用不同的损失函数
  - ✓ 对判别器D：
    - $\mathcal{L}_D = -\mathbb{E}_{\boldsymbol{x}\sim P_r}[\log D(\boldsymbol{x})] - \mathbb{E}_{\boldsymbol{z}\sim P_{\boldsymbol{z}}}[\log(1 - D(G(\boldsymbol{z}))]$

      -- 和原始GAN相同
  - ✓ 对生成器G：
    - $\mathcal{L}_{GD} = \frac{1}{2}||\mathbb{E}_{\boldsymbol{x}\sim P_r} f_D(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{z}\sim P_{\boldsymbol{z}}} f_D(G(\boldsymbol{z}))||_2^2$

      --特征中心的匹配

CVAE-GAN

特征中心的匹配

- 匹配真实数据和生成的数据的特征中心

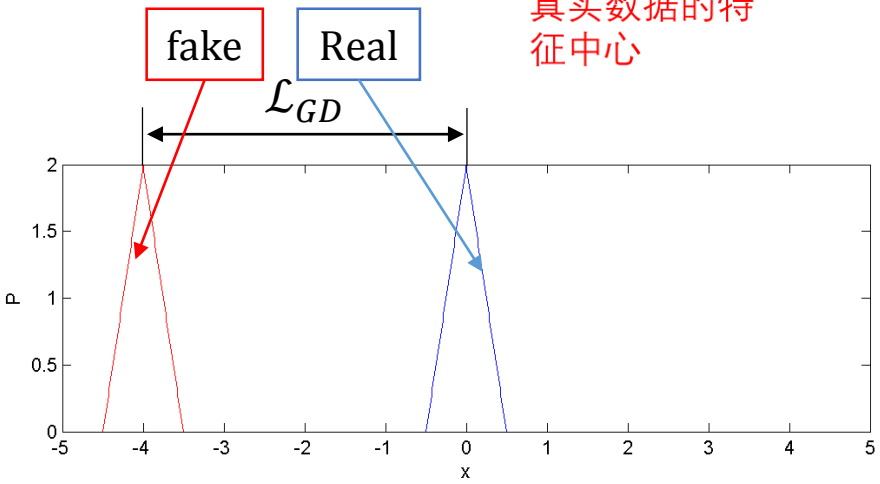
f_D: 网络D中的最后一个全连接层

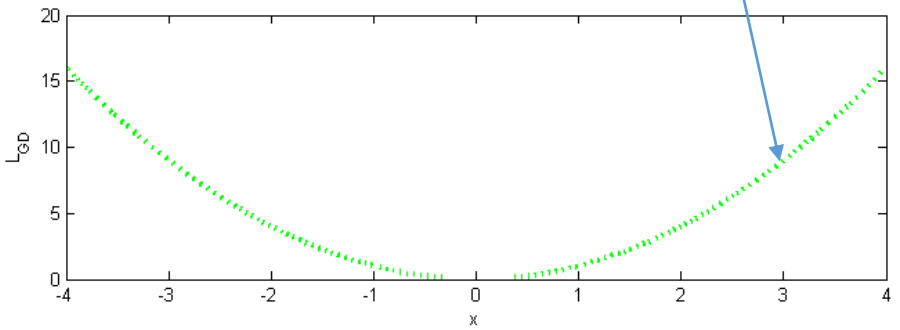
$$\mathcal{L}_{GD} = \frac{1}{2}||\mathbb{E}_{\boldsymbol{x}\sim P_r} f_D(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{z}\sim P_{\boldsymbol{z}}} f_D(G(\boldsymbol{z}))||_2^2$$

真实数据的特征中心

生成数据的特征中心

解决梯度消失的问题



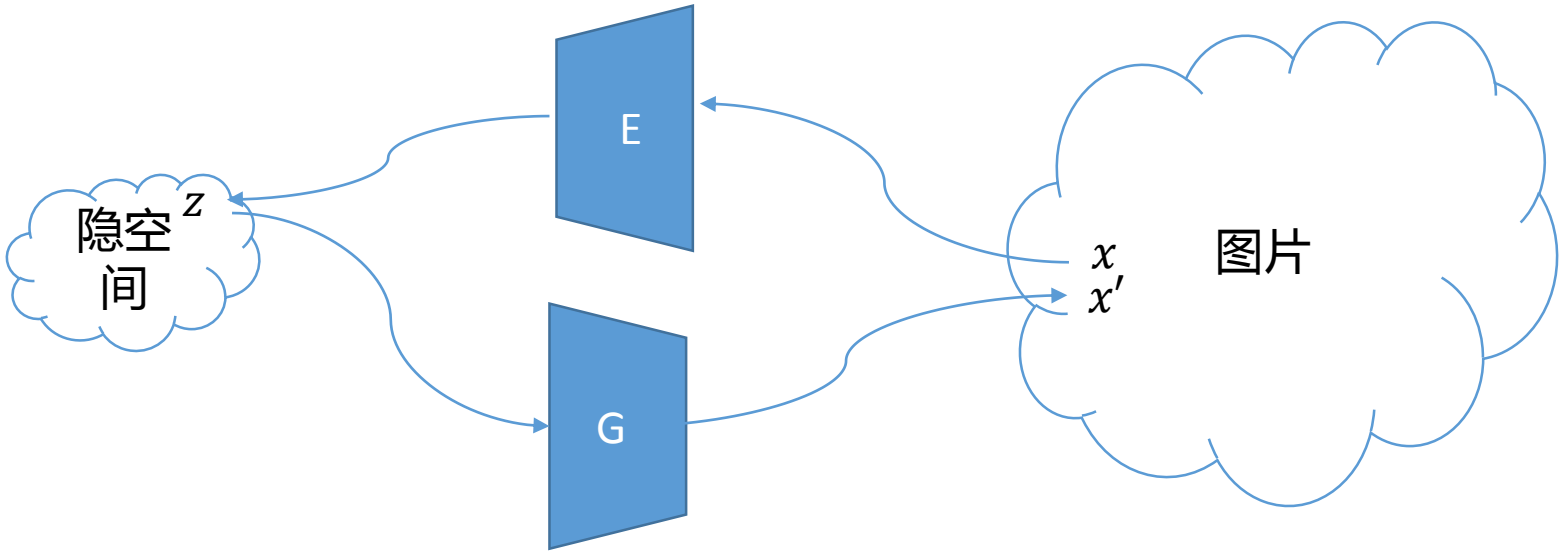


CVAE-GAN

解决多样性不足的问题

- 加入了一个编码器网络E
- 所以我们得到了一个从真实数据到隐空间的映射
- 我们加入了一个成对的像素级别和特征的损失函数：

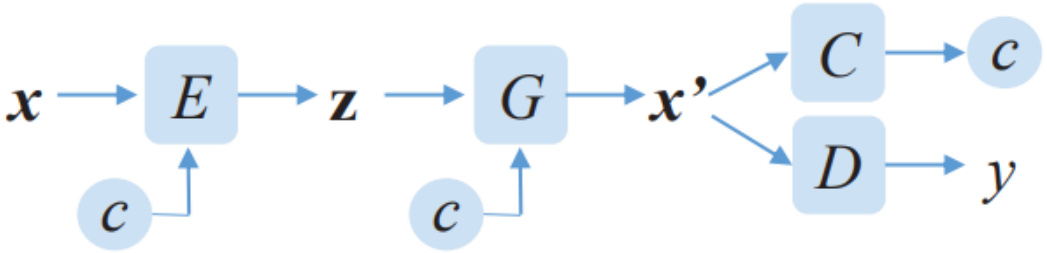
$$\mathcal{L}_G = \frac{1}{2}(\|x - x'\|_2^2 + \|f_D(x) - f_D(x')\|_2^2)$$



CVAE-GAN

细粒度的图片生成

- 框架结构
  - ✓ 引入条件信息
  - ✓ 生成一个具体类别的图片
    - ➢ 某一个人的人脸图片
    - ➢ 某一种鸟或者花的图片



CVAE-GAN

细粒度的图片生成

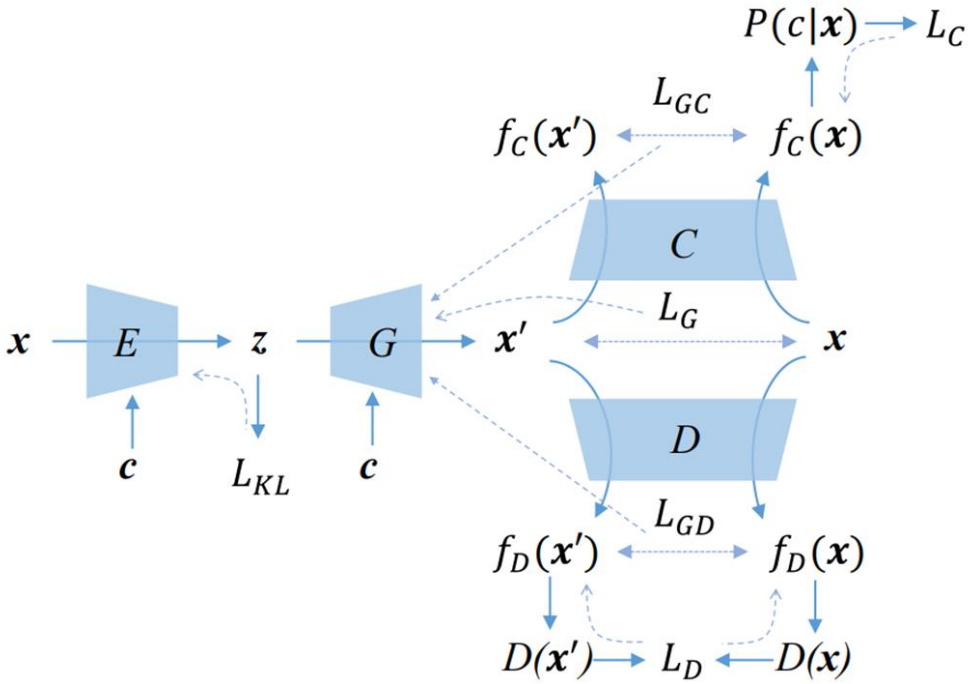
- 加入了一个分类器网络C，实现按照类别标签的图片生成
- 分类器网络C使用softmax的分类损失函数在真实图片上进行训练
  - $\mathcal{L}_C = -\mathbb{E}_{x \sim P_r}[\log P(c|x)]$
- 为了使生成网络G学习到正确的类别信息，我们使用基于分类器网络C的特征中心匹配方法训练G
  - $\mathcal{L}_{GC} = \frac{1}{2}\sum_c ||\mathbb{E}_{x \sim P_r} f_C(x) - \mathbb{E}_{z \sim P_z} f_C(G(z,c))||_2^2$

f_C: 网络C中的最后一个全连接层

CVAE-GAN

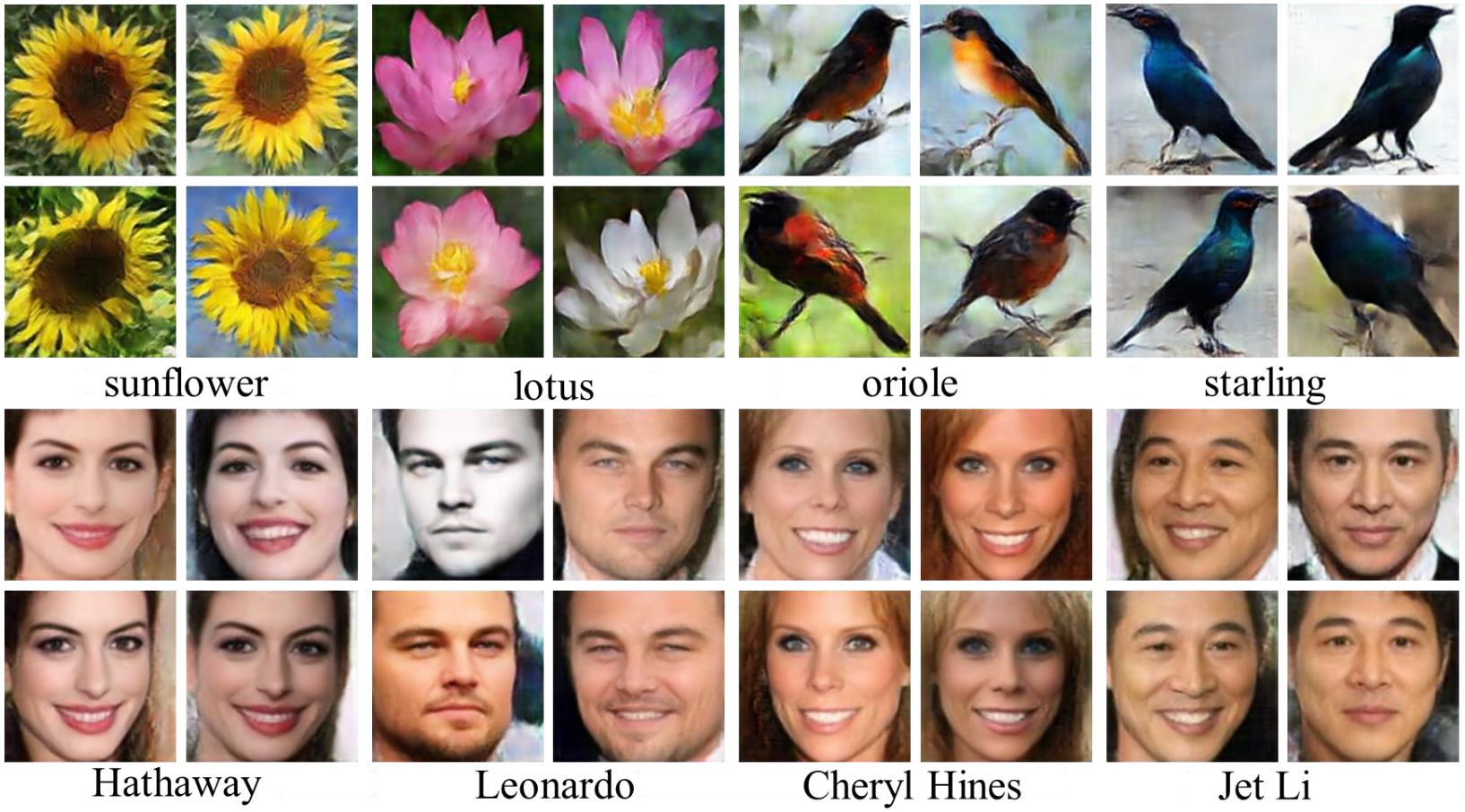
我们的方法

总体的网络结构

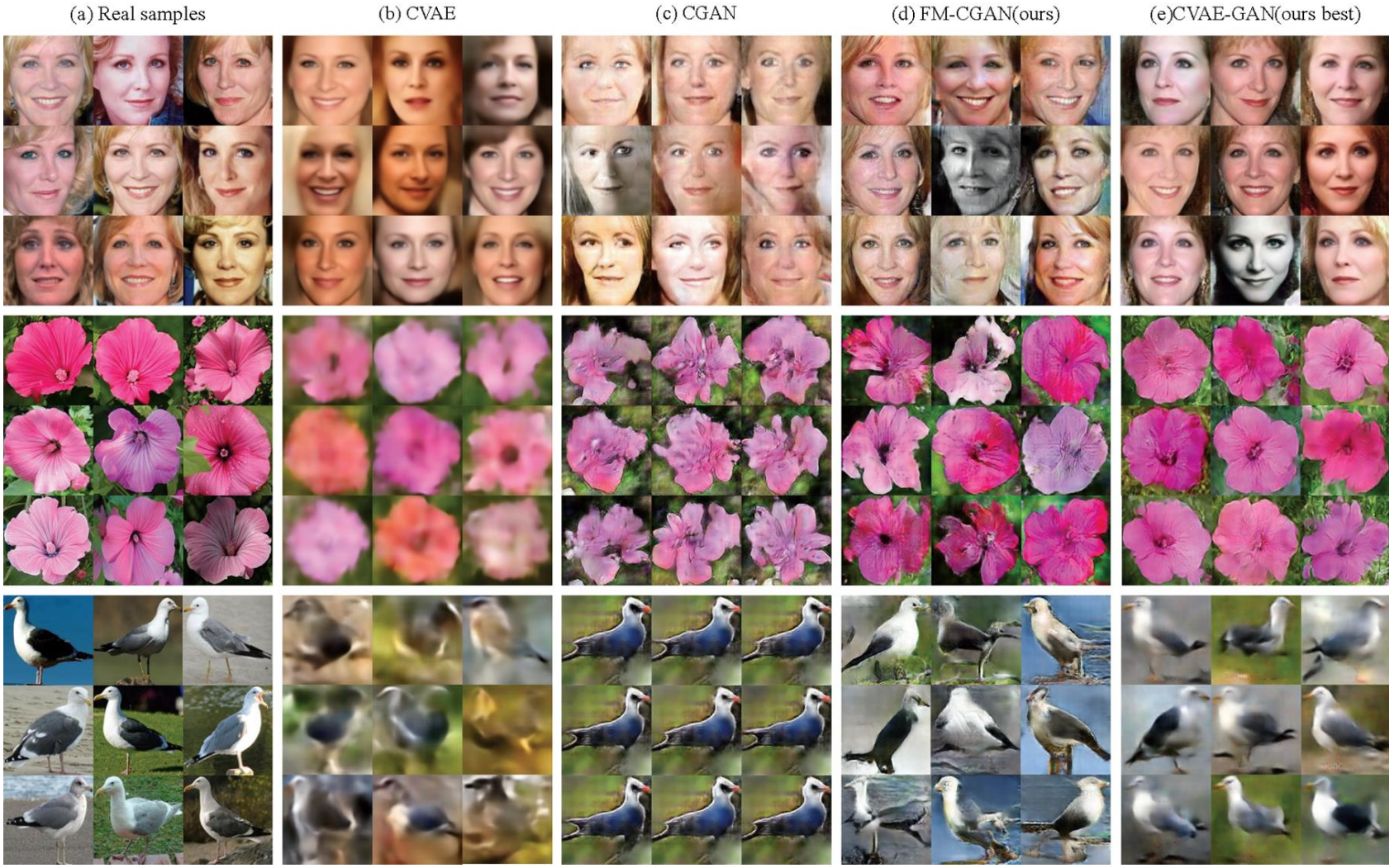


CVAE-GAN

- 用CVAE-GAN生成的图片结果



CVAE-GAN



CVAE-GAN

- 生成图片的质量的数值比较
  - 在CASIA数据集训练

	Real data	CVAE	CGAN	**CVAE-GAN (ours)**
Top-1 acc	99.61%	8.09%	61.97%	97.78%
Inception score	20.85	10.29	15.79	19.03

CVAE-GAN

CVAE-GAN的应用

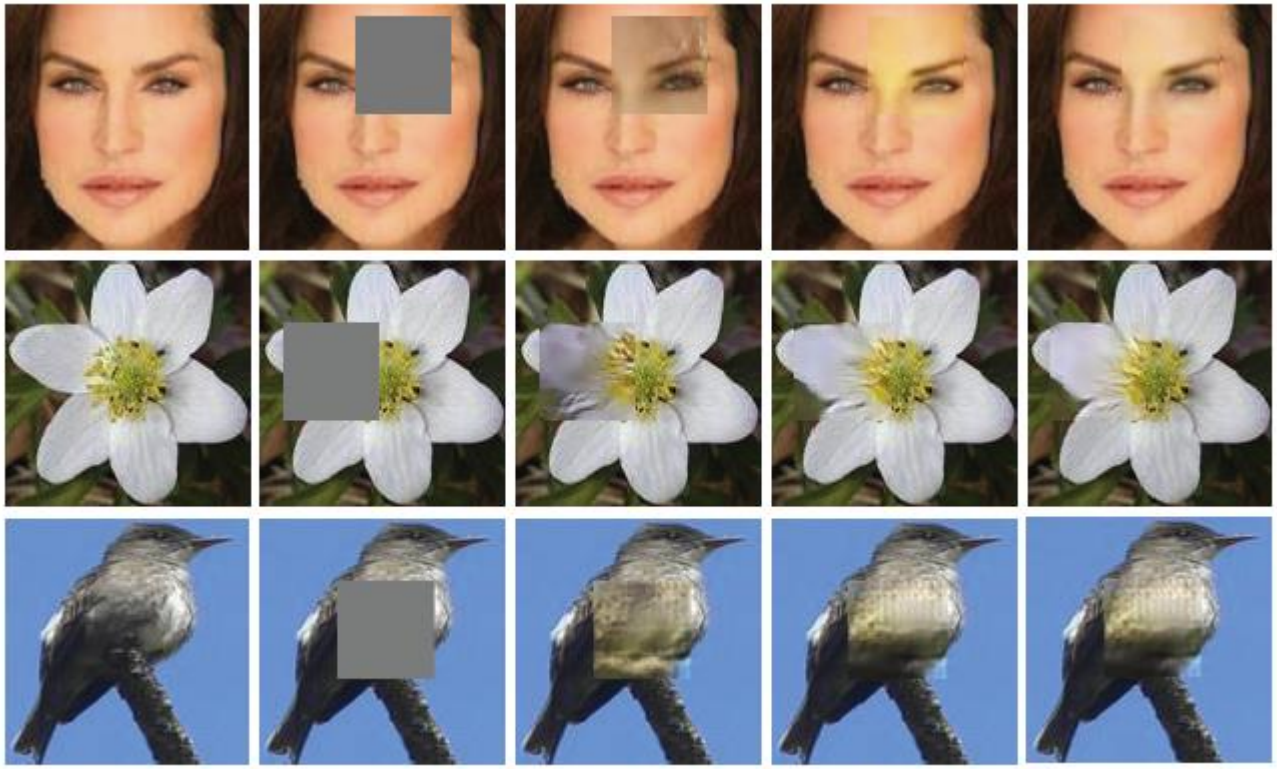
- 图片修复：
  ① 将缺损图片x输入编码器E得到隐向量z；
  ② 通过生成网络G得到生成图片x'；
  ③ 借助对应缺损图片缺损位置的二值mask矩阵M（缺损位置为1，其余位置为0）得到修复图片：
  - $x = M \odot x' + (1 - M) \odot x$

- 可以修复未用于训练的同类型图片。

- 对同一张缺损图片多次迭代重复修复过程可以获得更好的修复效果。

CVAE-GAN

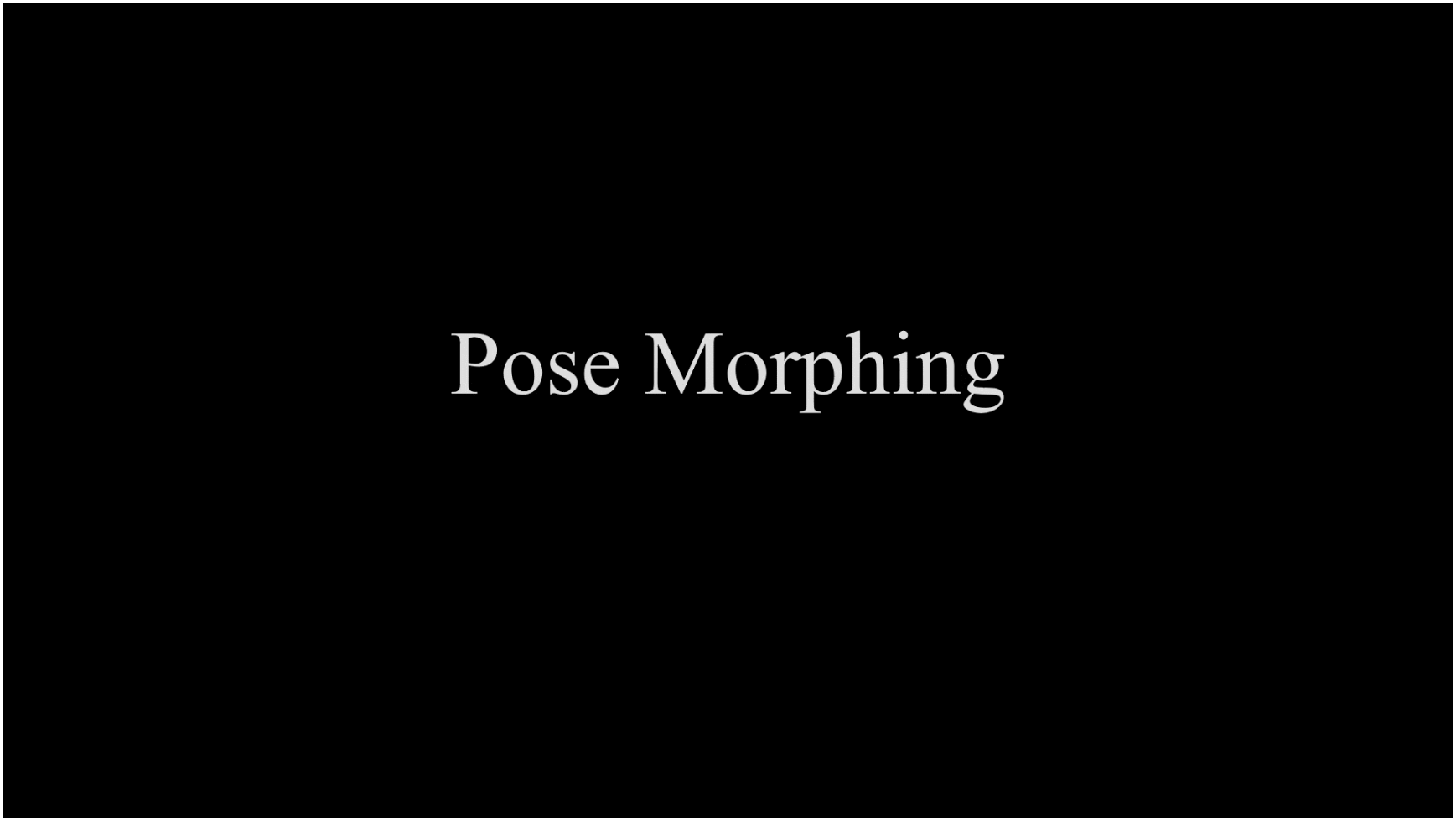
- 图片修复的效果



Results of iteration 1~10

CVAE-GAN

- CVAE-GAN的应用



CVAE-GAN

总结

- 我们提出了一个用于细粒度图片生成的框架。
- 我们提出使用特征中心匹配的方法来解决GAN的梯度消失的问题。
- 我们在网络结构上加入了编码器，利用重构强迫生成器逼近一个理想的单映射函数，以解决多样性不足的问题。
- 我们模型的生成图片的质量取得了非常好的结果。

AEGAN

- VAE存在的问题
  - 生成图像模糊。
- 模糊的原因是什么？
  - 人们认为VAE生成图像模糊的问题来源于重构过程中的MSE损失函数。
  - 我们发现MSE并不是主要的原因。

Jiayu Wang, Wengang Zhou, Jinhui Tang, Zhongqian Fu, Qi Tian, Houqiang Li. "Unregularized Auto-Encoder with Generative Adversarial Networks for Image Generation." *ACM MM*. 2018.

AEGAN

VAE的优化目标

- 目标函数：$\mathcal{L} = \mathbb{E}_{x\sim P_x}[\mathbb{E}_{z\sim Q(z|x)}[-\log p_g(x|z)] + KL(Q(z|x)||P_z)]$

  - 首先如果$KL(Q(z|x)||P_z) = 0$，则意味着样本编码得到的隐变量z的随机性很大，缺乏辨识度，所以$-\log p_g(x|z)$不可能小（重构预测的准确率下降）。

  - 而如果$-\log p_g(x|z)$小则$p_g(x|z)$大，代表重构预测准确，这时候$Q(z|x)$必然不能太随机，即$KL(Q(z|x)||P_z)$不会小。

  - 所以，这两部分的loss其实是相互拮抗的，也就是说，VAE的训练过程包含两个矛盾的优化目标。

AEGAN

VAE的训练过程

- VAE是隐空间受约束的自动编码器，实际训练时需要在目标函数的约束项上加一个额外的参数λ来控制约束项的强弱：

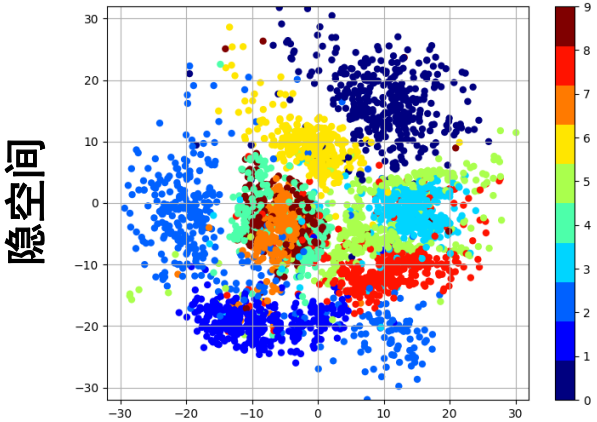
  ✓ $\mathcal{L} = \mathbb{E}_{x\sim P_x}[\mathbb{E}_{z\sim Q(z|x)}[-\log p_g(x|z)] + \lambda \times KL(Q(z|x)||P_z)]$

  重构项　　约束项

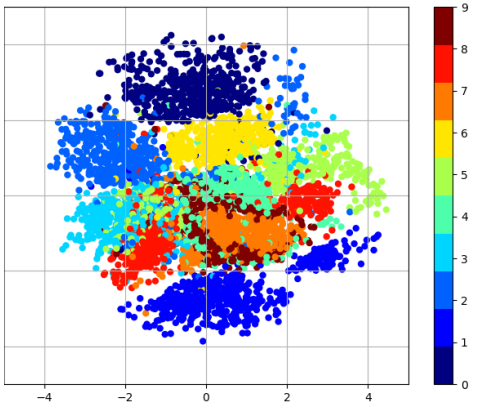
- 实际训练中对于约束项的强弱很难控制：
  - 如果约束太强，隐向量之间的重叠现象会很严重，区分度下降。解码器难以区分不同的隐向量。这导致对于不同的训练样本编码得到的隐向量，解码器只会以极低的重构准确率输出几乎完全相同的重构结果；
  - 如果约束太弱，隐空间不能被足够准确地限制在标准高斯分布内。待训练结束，从高斯分布的采样得到的隐向量不足以提供有意义的样本编码信息，这也会使得生成结果模糊。

AEGAN

- 隐空间的2维可视化



(a) Auto-Encoder.



(b) VAE.

(c) 重构 (VAE).

(d) 随机生成 (VAE).

AEGAN

- 隐空间的约束是生成图像模糊的主要原因
  - 去模糊就要去约束。
  - 为什么要加约束？
    - 作为生成模型，我们需要在隐空间进行有效的采样。而普通自动编码器的隐空间的分布是复杂且未知的，我们无法直接采样。只有通过施加约束的方法让其变得可采样。
    - 所以，要想去除约束，就必须先解决采样问题！

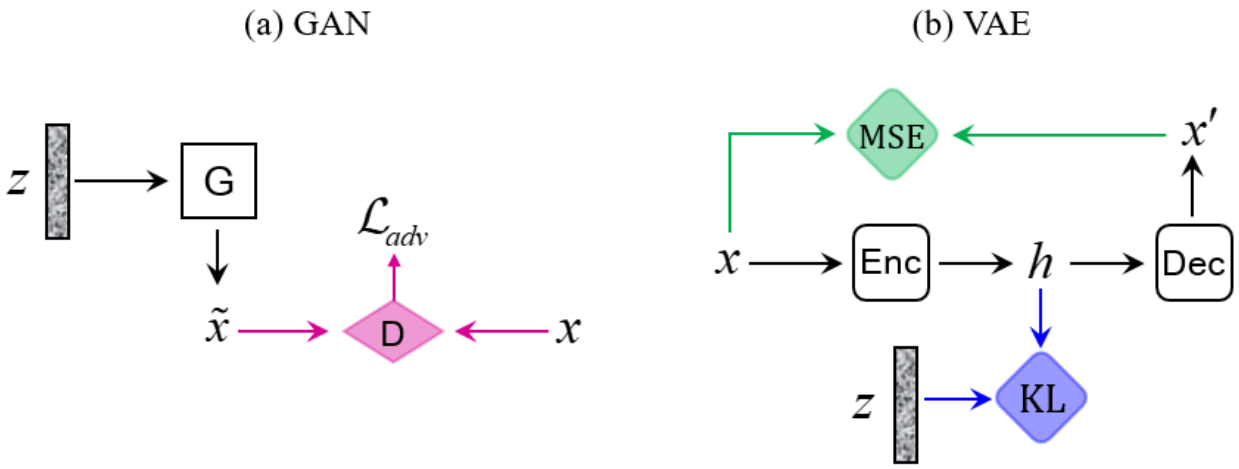
AEGAN

隐空间的采样问题

- 隐空间采样要求：
  - ✓ 能够对空间分布复杂未知的数据进行采样的方法；
  - ✓ 不需要额外的隐变量推断；
    - ➢ 答案已经呼之欲出了，就是使用GAN!

AEGAN

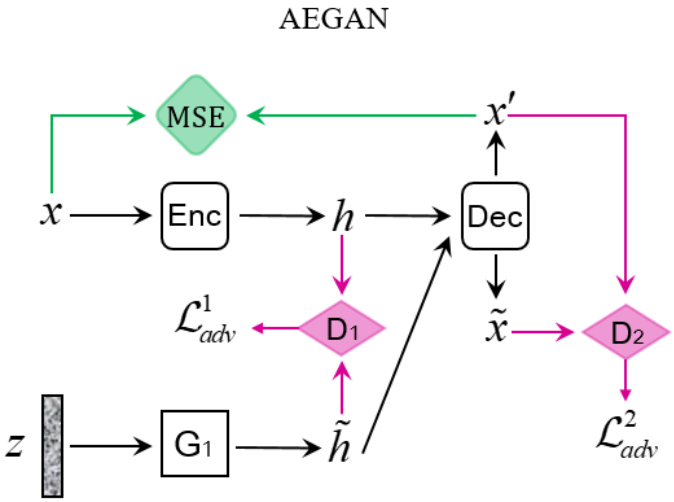
我们的方法



- 去除VAE隐空间的KL散度约束；
- 在隐空间的层面上引入GAN，对隐变量进行采样；
- 将GAN结合进VAE，发挥两种模型各自的优势。

AEGAN

AEGAN



- 我们隐空间的层面引入GAN，通过对正态分布的映射，在隐空间进行采样。判别器D_1用于度量G_1生成的数据分布与真实隐空间分布的距离，G_1生成的隐变量再输入解码器Dec则得到最终的生成图片；
- 为了提高生成图片的质量，我们还引入了一个额外的判别器D_2，用于度量生成样本分布与真实样本分布的区别。通过梯度的反向传播，D_2与D_1共同“指导”G_1的参数更新。

AEGAN

AEGAN优化目标

- 自动编码器:
  - ✓ $$\mathcal{L}_{AE} = \frac{1}{n \times H \times W} \sum_{i=1}^{n} \| x_i - Dec(h_i) \|_2^2$$
- GAN:
  - ✓ D_1: $\mathcal{L}_{D_1} = -\mathbb{E}_{h_i \sim P_{Enc}}[\log D_1(h_i)] - \mathbb{E}_{\tilde{h}_i \sim P_{G_1}}[\log(1 - D_1(\tilde{h}_i))]$
  - ✓ D_2: $\mathcal{L}_{D_2} = -\mathbb{E}_{x_i' \sim P_{x'}}[\log D_2(x_i')] - \mathbb{E}_{\tilde{x}_i \sim P_{\tilde{x}}}[\log(1 - D_2(\tilde{x}_i))]$
  - ✓ G_1: $\mathcal{L}_{G_1} = -\mathbb{E}_{\tilde{h}_i \sim P_{G_1}}[\log D_1(\tilde{h}_i)] - \lambda \mathbb{E}_{\tilde{x}_i \sim P_{\tilde{x}}}[\log D_2(\tilde{x}_i)]$

AEGAN

- AEGAN生成的图片结果

VAE

WAE-GAN

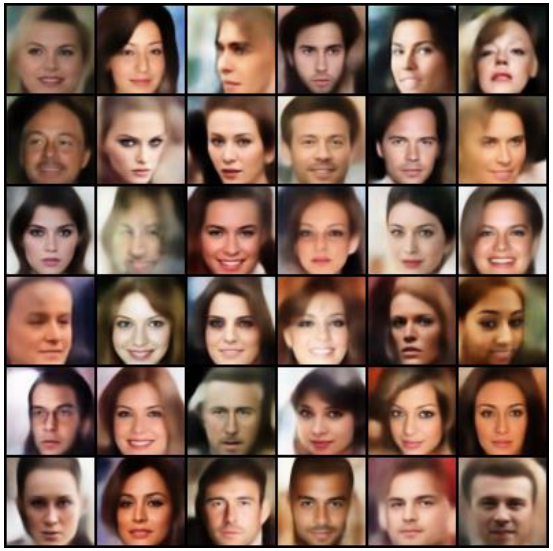
AEGAN

Fashion-MNIST

AEGAN

VAE

WAE-GAN



AEGAN

CelebA

AEGAN

- 生成图片的质量的数值比较（FID score）
  - 在CelebA数据集训练

Model	VAE	WAE-GAN	AEGAN
FID	78	49	40

FID score on CelebA (smaller is better)

AEGAN

隐变量插值

- 一个理想的生成模型应该能够学习到足够紧凑且有意义的隐空间。为了测试我们AEGAN中额外引入的GAN是否学习到了理想的自编码器的隐空间分布，我们对模型进行了隐变量插值测试。

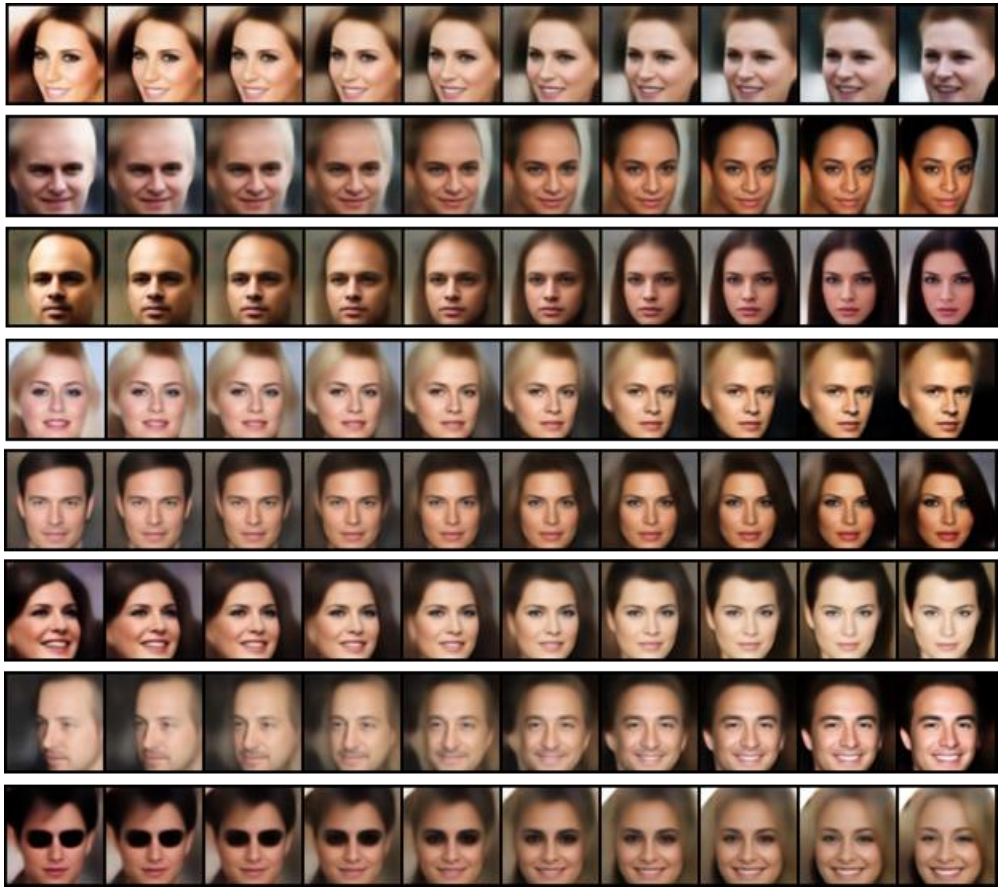
  - 首先我们从训练数据中采样一对样本，记为：(x_1, x_2);
  - 从标准正态分布中随机采样一对随机向量：(z_1, z_2);
  - 通过最小化$\|x_i - Dec(G_1(z_i))\|_2^2$多次迭代更新$(z_1, z_2)$的数值;
  - 通过对z_1，z_2作线性插值得到一系列的z向量：

    $$z = \alpha z_1 + (1 - \alpha) z_2$$

  - 将所有的z向量输入AEGAN得到生成图片。

AEGAN

- 隐变量插值结果



AEGAN

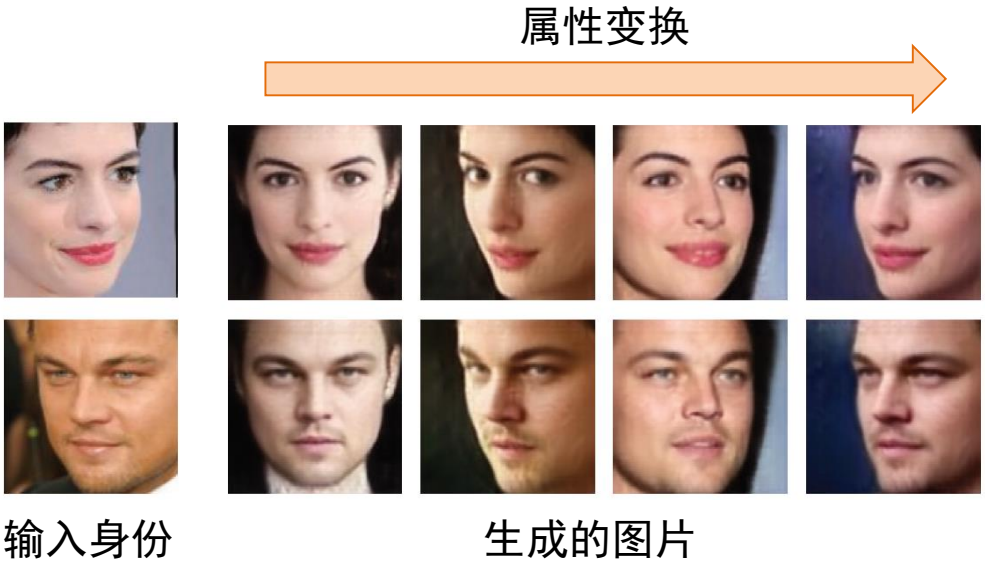
总结

- 我们指出了造成VAE生成图片模糊的原因并不是模型所使用的MSE损失函数，而是施加在模型隐空间的KL散度约束。
- 我们通过引入额外的GAN网络对编码器隐空间进行采样的方式去除了KL散度约束。
- 我们新提出的模型AEGAN很好地解决了VAE生成图片模糊的问题，使生成图片的质量获得了极大的提升。

IP-GAN

研究目的

- 生成任意一个人的图片(开放的环境)；
- 可以改变光照，姿势，面部表情等属性(保持身份)。



Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua. "Towards Open-Set Identity Preserving Face Synthesis." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

IP-GAN

我们的方法

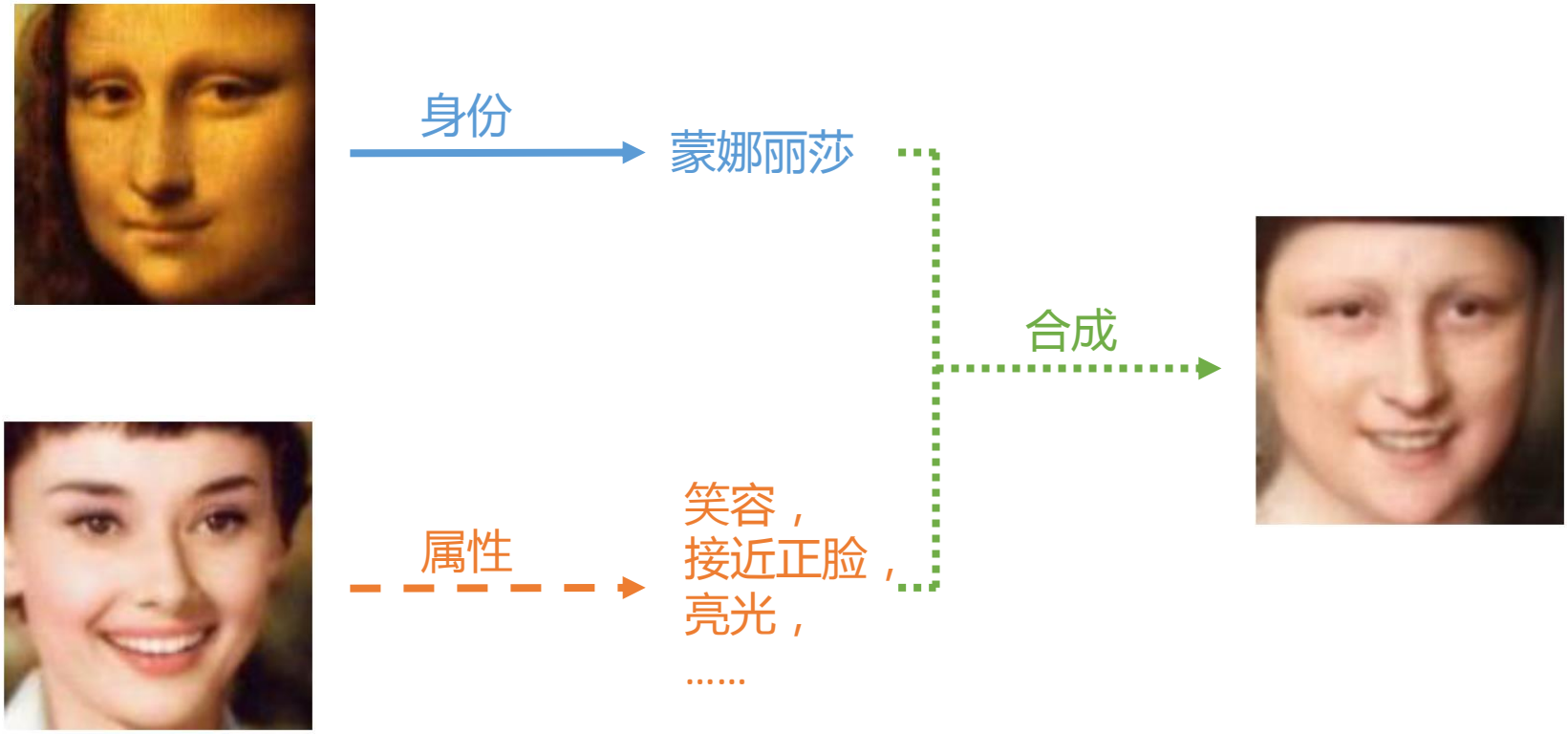
- 分离人脸图片中的身份信息和属性信息

奥黛丽赫本 ← 身份

属性 → 笑容，接近正脸，亮光，……

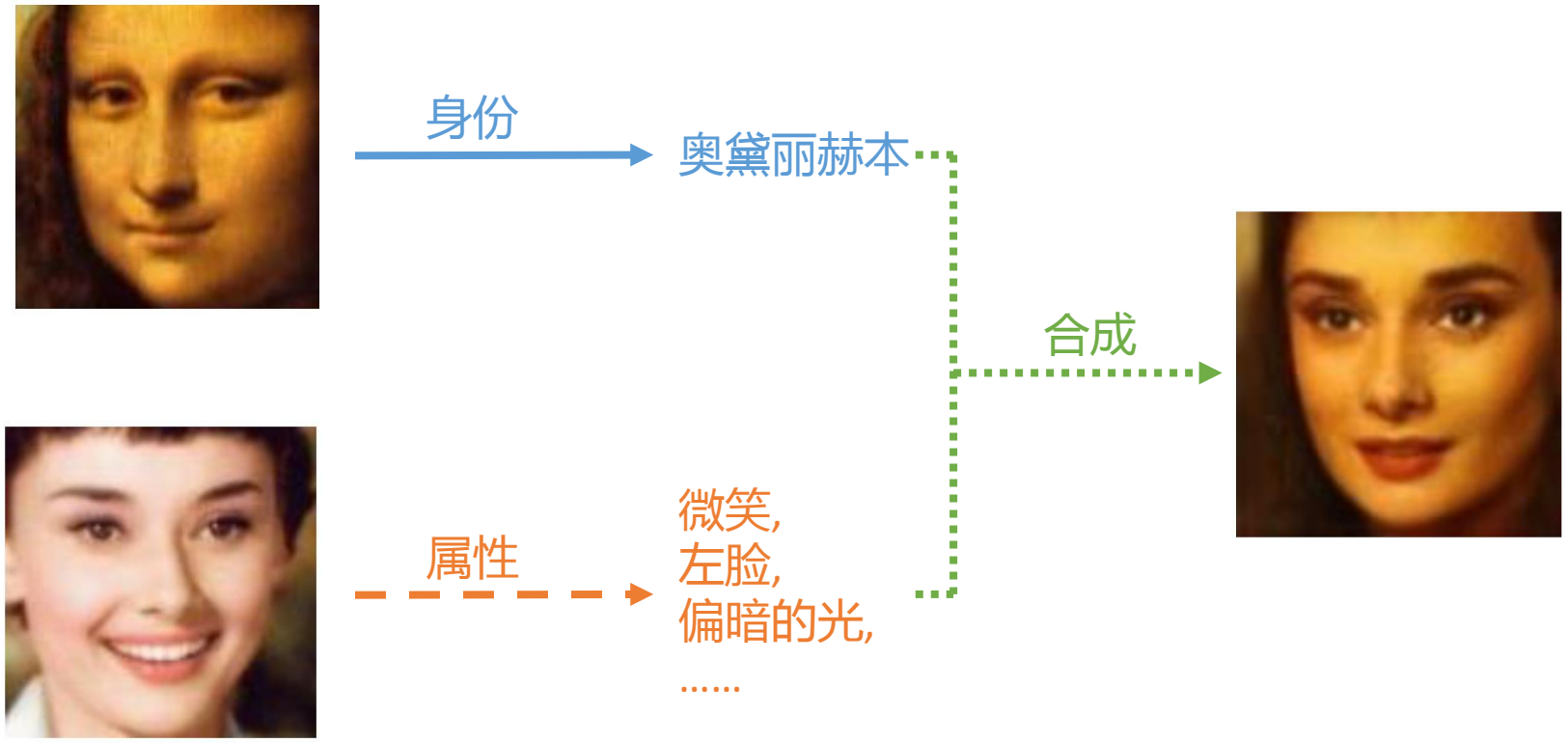
IP-GAN

- 我们的方法
  - 将来自不同人脸的身份信息与属性信息融合，合成新的图片



IP-GAN

- 我们的方法
  - 将来自不同人脸的身份信息与属性信息融合，合成新的图片



IP-GAN

- 我们的方法

IP-GAN

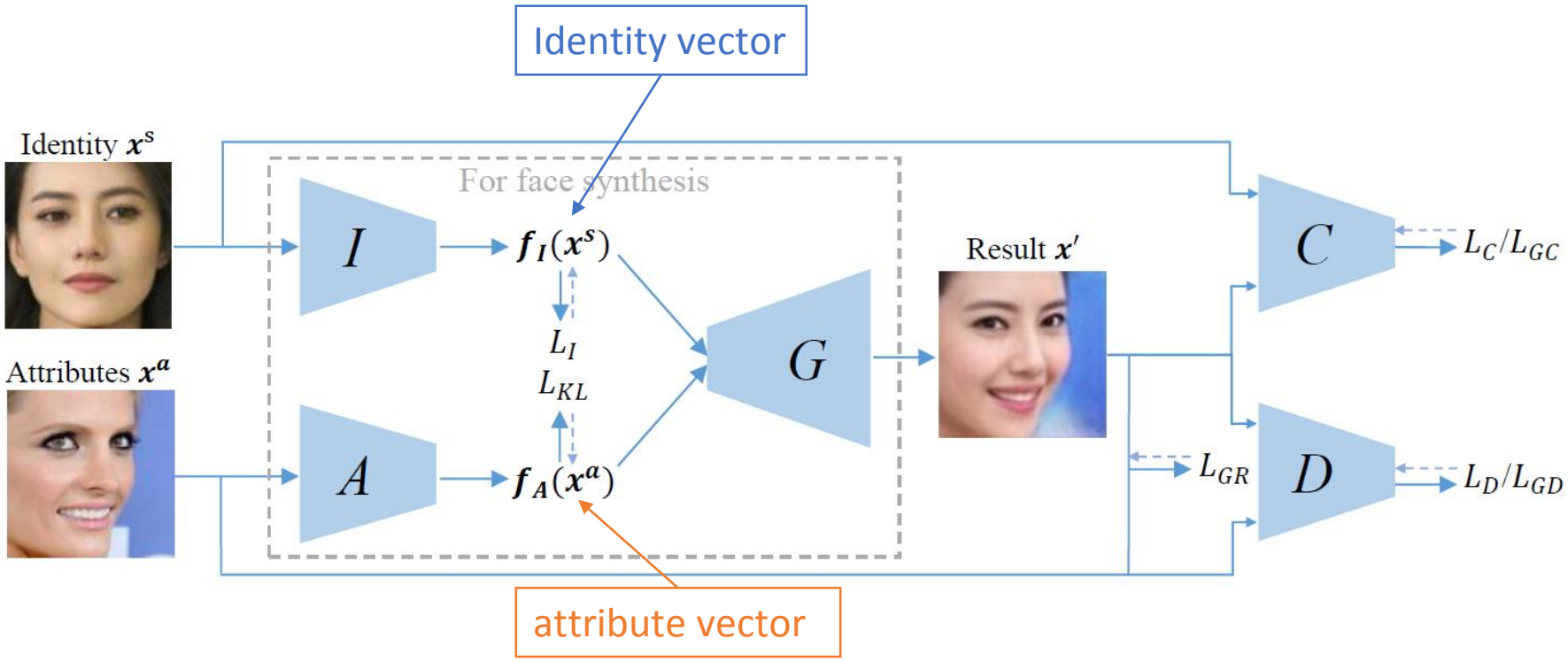
主要的贡献

- 提出一种简单的从图片中解耦出身份信息和属性信息的方法；
- 我们的方法不需要对属性信息进行标注；
- 我们的方法可以用在很多的任务当中，例如，侧脸转正脸，图片属性的转换。

IP-GAN

我们的方法

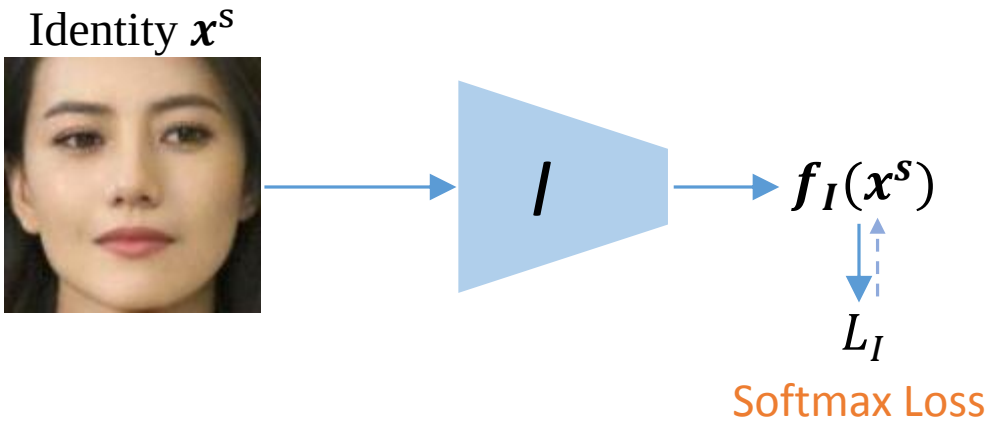
- 整体的网络框架



IP-GAN

提取身份信息的网络

- 网络做人脸的分类的任务
- 相同身份信息的人脸图片有着相同的特征

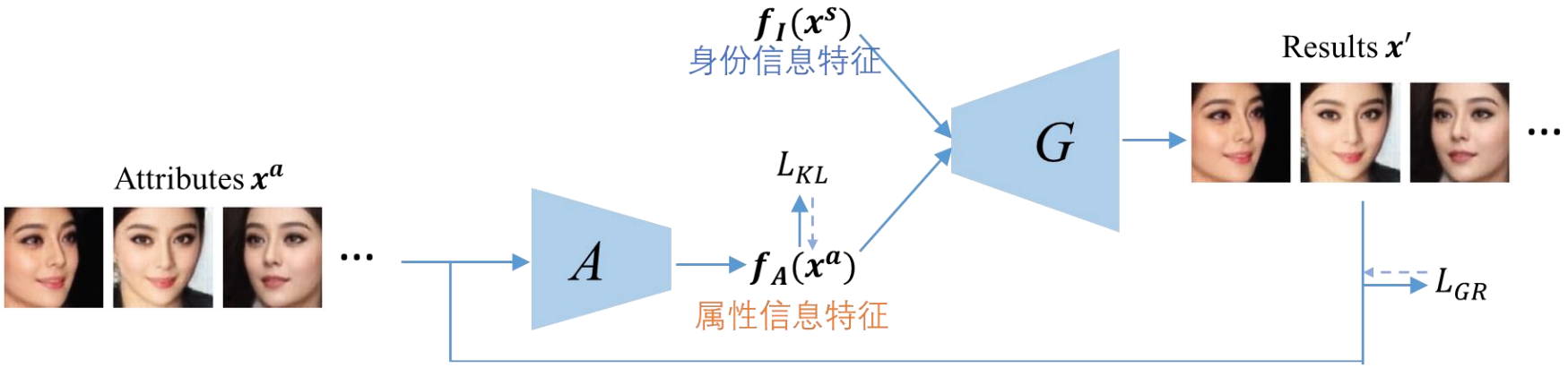


$$\mathcal{L}_{\mathcal{I}} = -\mathbb{E}_{\boldsymbol{x}\sim P_r}[\log P(c|\boldsymbol{x}^{\boldsymbol{s}})]$$

IP-GAN

提取属性信息的网络

重构的损失函数和KL损失函数



KL 损失函数:

$$\mathcal{L}_{KL} = \frac{1}{2}(\mu^T\mu + \sum_{j-1}^{J}(\exp(\epsilon) - \epsilon - 1))$$

重构损失函数:

$$\mathcal{L}_{GR} = \begin{cases} \frac{1}{2}\|x^a - x'\|_2^2 & \text{if } x^s = x^a \\ \frac{\lambda}{2}\|x^a - x'\|_2^2 & \text{otherwise} \end{cases}$$

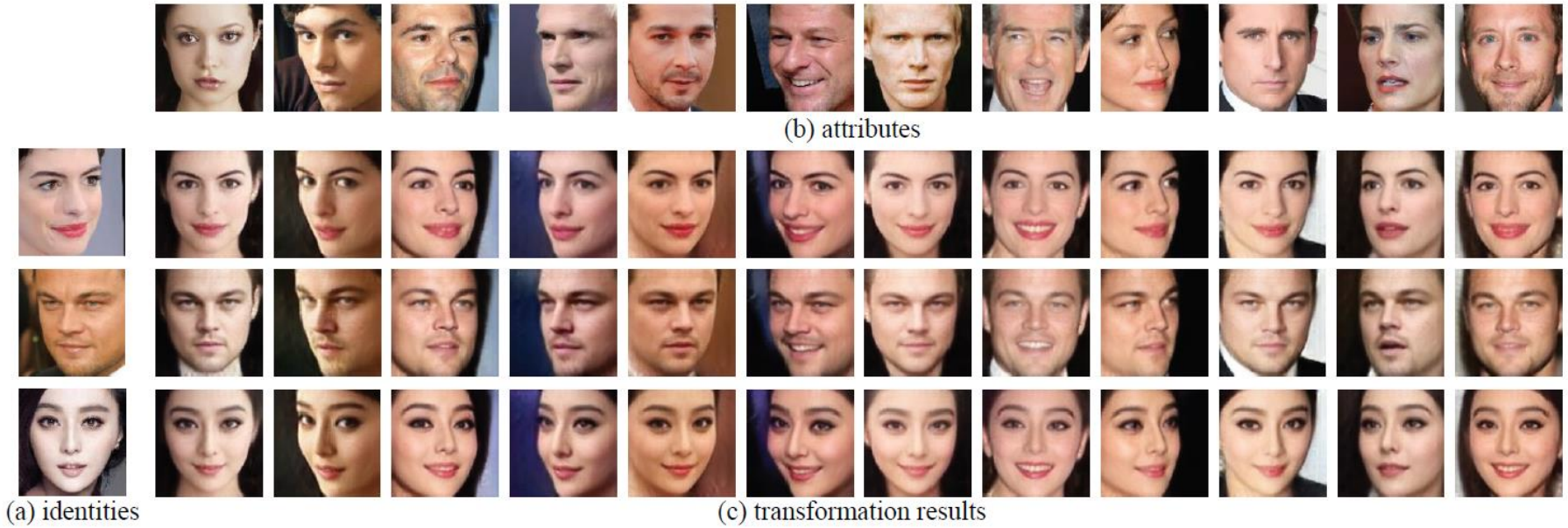
IP-GAN

分类网络与判别网络

- 分类网络C：用于判断生成图片是否对应事先提取到的身份信息
  - Softmax分类： $\mathcal{L}_{\mathcal{C}} = -\mathbb{E}_{\boldsymbol{x}\sim P_r}[\log P(c|\boldsymbol{x}^{\boldsymbol{s}})]$
  - 特征中心匹配：$\mathcal{L}_{GC} = \frac{1}{2}||\boldsymbol{f_C}(\boldsymbol{x}') - \boldsymbol{f_C}(\boldsymbol{x}^{\boldsymbol{s}})||_2^2$
- 判别网络D：用于度量生成图片的真实度
  - 原始GAN loss：$\mathcal{L}_D = -\mathbb{E}_{\boldsymbol{x}\sim P_r}[\log D(\boldsymbol{x}^{\boldsymbol{a}})] - \mathbb{E}_{\boldsymbol{z}\sim P_{\boldsymbol{z}}}[\log(1 - D(G(\boldsymbol{z}))]$
  - 特征中心匹配：$\mathcal{L}_{GD} = \frac{1}{2}||\boldsymbol{f_D}(\boldsymbol{x}') - \boldsymbol{f_D}(\boldsymbol{x}^{\boldsymbol{a}})||_2^2$

IP-GAN

- 人脸属性的转变结果



对特定身份的图片进行属性编辑

IP-GAN

- 侧脸转成正脸的结果

Ours

侧脸转正脸

IP-GAN

总结

- 我们提出了一个可用于面向开放数据集的人脸合成的框架
- 这个框架可以生成保持身份信息并且真实的人脸的图片

报告提纲

- 生成式模型的定义
- 研究生成式模型的意义
- 经典的生成式模型：VAE 与 GAN
- GAN的应用
- VAE 与 GAN 仍需解决的问题
- 我们的工作
- 总结

总结

- VAE与GAN作为优秀的生成模型，可以帮助我们更好地理解周边环境的信息。
- VAE与GAN被广泛地应用到了图像和视频领域。在人机交互、无监督学习和强化学习领域的应用也充满潜力。
- VAE与GAN的训练指标、生成样本多样性以及稳定性正受到广泛的关注，成为下一步要解决的问题。

THANK YOU FOR WATCHING

AND

HAVE A NICE DAY :)