

CGAN 汽车设计效果图生成

张翔，陆轶

车轮互联

November 29, 2017

人类婴儿

- ▶ 1 Month 注视人脸
- ▶ 3 Month 物体识别, 眼球移动
- ▶ 6 Month 空间关系, 抓取
- ▶ 9 Month
轨迹和预测
扔东西
- ▶ 12 month
综合规划
模仿和生成 (语言和造音)

AI

- ▶ Face Recognition
- ▶ Faster R-CNN, Visual Tracking
- ▶ SLAM + TCN
- ▶ convolutional dynamic neural advection (CDNA)
spatial transformer
predictors (STP)
- ▶ Net2Net, Planning
VAE, GAN

模型

- ▶ Variational Autoencoder(VAE):
变分自动编码器
- ▶ Generative Adversarial Network(GAN):
对抗生成网络

方法

在一个先验概率分布中取样(Prior Distribution,一般取N维的高斯分布), 并将取得的随机样本 z 映射到图像的概率分布 X 中。
将 z 映射到 X 的计算 $P(X|z; \theta)$ 就是由神经网络来实现的
(MCMC据说效果不行), 其中 θ 为神经网络的模型参数。

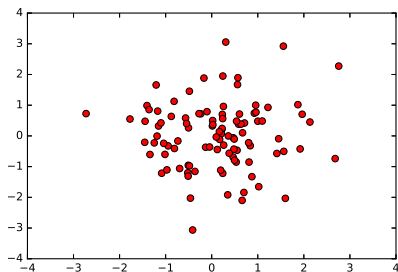


Figure 1: 高斯随机变量 z

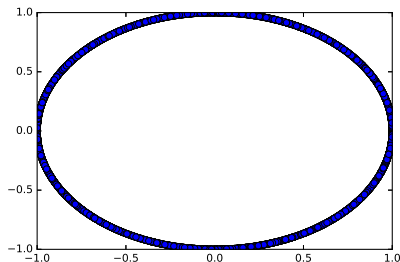


Figure 2: 将 z 映射到 x

其中 $z \sim \mathcal{N}(\mu, \sigma^2)$ 且 $|z| > \epsilon$, $x = z/\|z\| = g(z)$

我们希望可以训练一个神经网络将一个分布近似地映射到另一个分布上, 扮演 $g(z)$ 的角色。当然, 当 $z \rightarrow 0$ 时, 这样的映射也会出现问题。

- ▶ 无监督模型，分为编码Encoder与反编码Decoder两部分。
- ▶ Encoder将输入图片映射到一个隐藏的随机变量 $P(z|X)$, Decoder将隐藏变量重构 $P(X|z)$ 。Encoder与Decoder将以神经网络来拟合。
- ▶ 在生成图片阶段会直接将随机变量 z 映射到图片分布。
- ▶ 训练目标(假设网络的拟合性足够强大)

$$\max \log P(X) = \mathbb{E}_z[\log P(X|z)] - KL[P(z|X)||P(z)]$$

使得解码后生成的图片在概率上接近原图，同时将编码器得到的隐变量接近先验概率 $P(z)$ 。

- ▶ 测试生成图片 $X_{test} \sim P(X|z)$ 。

- ▶ GAN是一个无监督的模型，分为生成网络 G 和对抗网络 D 。
- ▶ 生成网络 $G(z; \theta_g)$ 将一个随机高斯变量映射到图片空间，输出为 x 。 G 的训练目标是生成逼真的图片。
- ▶ 对抗网络 $D(x; \theta_d)$ 判断一个输入图片 x 是否来自真实图片分布，输出一个标量表示 x 是真实图片的概率。 D 的目标是监督由 G 生成的图片是否足够逼真。
- ▶ G 和 D 可为多层感知器，也可以是卷积网络。其中 θ_g 和 θ_d 分别表示神经网络的参数，为了训练目标(G : 作假, D : 打假)不断优化。

构建代价函数

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

▶ G 的代价函数

$$Loss_G = \mathbb{E}_{z \sim \mathcal{N}} - \log(D(G(z)))$$

G 生成的图片越是逼真， $D(G(z))$ 就越大， $Loss_G$ 就越小，其中 z 在高斯分布中随机取样。

▶ D 的代价函数

$$Loss_D = -[\mathbb{E}_{x \sim P_{data}} \log(D(x)) + \mathbb{E}_{z \sim \mathcal{N}} \log(1 - D(G(z)))]$$

左式表示 D 需要能够鉴真($D(x) \rightarrow 1$)，右式表示 D 同时也需要防伪($D(G(z)) \rightarrow 0$)。 x, z 来自于随机取样。

当训练刚开始时 D 还不具备辨识能力，因此

$$D(x) = D(G(z)) = 0.5, Loss_G = \log(2), Loss_D = \log(4)$$

当 D 与 G 训练到完美时，这个结果依旧成立。

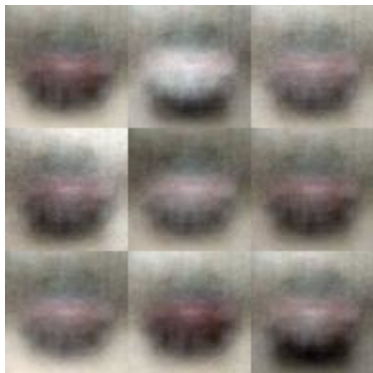


Figure 3: VAE自动生成的图片



Figure 4: GAN自动生成的图片

VAE生成的图片注重最小化像素间的差值，而GAN生成器形成的图片则更加生动形象。



Figure 5: 当GAN生成图片分辨率提高后的失败效果

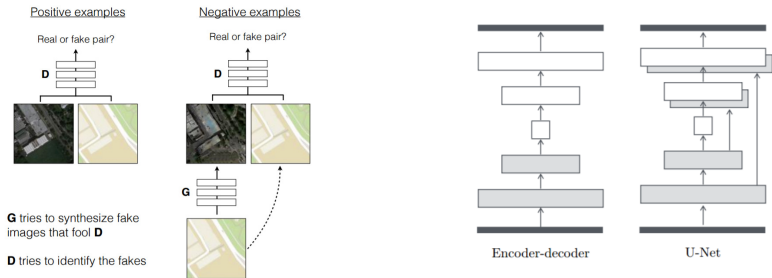
思考：如果对各个角度的图片加上编码，效果是否会提升？



Figure 6: CGAN生成鞋子示例[5]

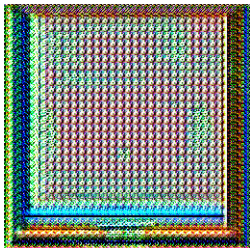
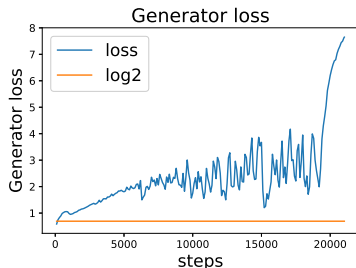
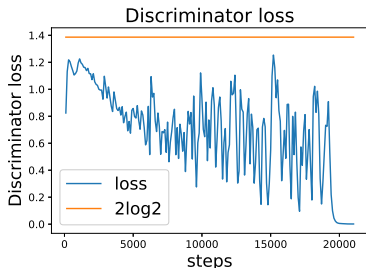
$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}} \log(D(x|y)) + \mathbb{E}_{z \sim \mathcal{N}} \log(1 - D(G(z|y)))$$

在代价函数中加入了 y ,而由于 y 的加入,训练将变得容易很多。
生成模型针对 y 提供的信息进行生成。



- ▶ 其中用户输入的草图就是生成模型中已知的 y 。image2image的作者发现在隐层加噪音效果对输出的填色结果影响不大，于是就使用了dropout代替了 z 使得输出具有一定的随机性(实验发现，输出随机性仍然较小)。
- ▶ 同时生成器U-Net中添加了隐层间链接(skip connection),保留了线条特征对生成至关重要。
- ▶ patchGAN加速了训练。

手绘生成图片训练过程



GAN的训练非常不稳定。如果 D 一旦过强， G 生成的所有结果都会被识破。平衡被打破，训练很容易就失败了。

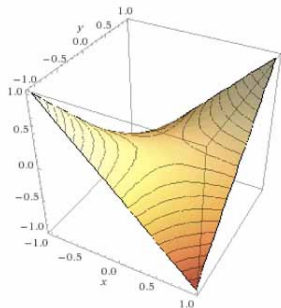


Figure 7: 原先认为训练难度是在寻找纳什平衡 $V(x; y) = xy$

在最优判别器的下，我们可以把原始GAN定义的生成器loss等价变换为最小化真实分布 P_r 与生成分布 P_g 之间的JS散度。而只要它们俩没有一点重叠或者重叠部分可忽略，JS散度就固定是常数 $\log 2$ ，而这对于梯度下降方法意味着——梯度为0

将 G 的参数标注为 θ ,将 D 的参数标注为 ω 。 f 和 g 分别为 D 与 G 的线性输出。 c 为一个常数。

while θ has not converged **do**

for $t \leftarrow 0$ to n_{critic} **do**

$$\quad \quad \text{grad}_{\omega} \leftarrow \nabla_{\omega} [\mathbb{E}_{x \sim P_{data}} [f_{\omega}(x)] - \mathbb{E}_{z \sim \mathcal{N}} [f_{\omega}(g_{\theta}(z))]]$$

$$\quad \quad \omega \leftarrow \omega + \alpha * RMSProp(\text{grad}_{\omega})$$

$$\quad \quad \omega = clip(-c, c)$$

end

$$\quad \text{grad}_{\theta} \leftarrow \nabla_{\theta} \mathbb{E}_{z \sim \mathcal{N}} (-f_{\omega}(g_{\theta}(z)))$$

$$\quad \theta \leftarrow \theta - \alpha * RMSProp(\text{grad}_{\theta})$$

end

Algorithm 1: Wasserstein loss[1]

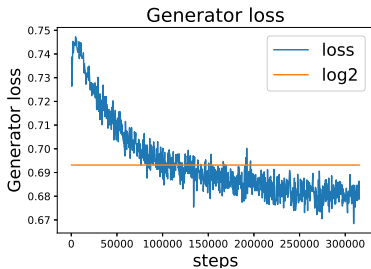
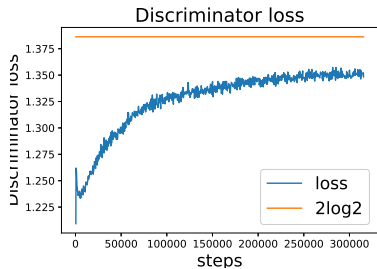




Figure 8: 轮廓生成使用Holistically-Nested Edge Detection

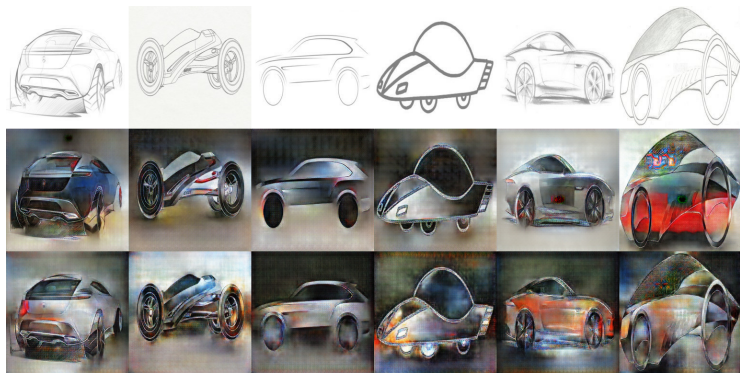


Figure 9: CGAN生成图片示例:第一行用户输入, 第二行生成器生成图片,第三行生成器训练时加入了L1 regularization。

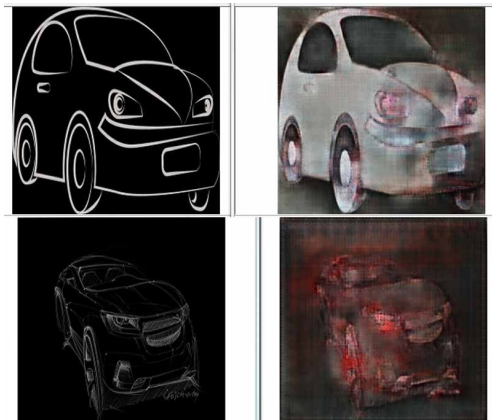


Figure 10: 轮廓的输入风格没有鲁棒性

思考: CGAN条件输入 y 有什么约束? 能否换成其他形式?

- ▶ 3D-GAN
- ▶ AdaGAN
- ▶ ArtGAN
- ▶ Bayesian GAN
- ▶ CatGAN
- ▶ DCGAN
- ▶ EBGAN
- ▶ FF-GAN
- ▶ GoGAN
- ▶ iGAN
- ▶ LAGAN
- ▶ MV-BiGAN
- ▶ PrGAN
- ▶ SGAN
- ▶ TGAN
- ▶ VGAN

LAPGAN[2]

- ▶ 高斯金字塔下采样+拉普拉斯金字塔上采样
- ▶ 输入带条件CGAN
- ▶ 针对残差学习

PG-GAN[6]

- ▶ `resize+conv`同为金字塔结构
- ▶ 加权fade in
- ▶ 直接生成图像
- ▶ PyTorch实现

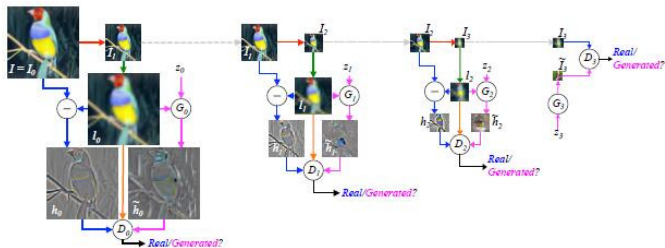


Figure 11: The training procedure for our LAPGAN model. Starting with a 64x64 input image I from our training set (top left)

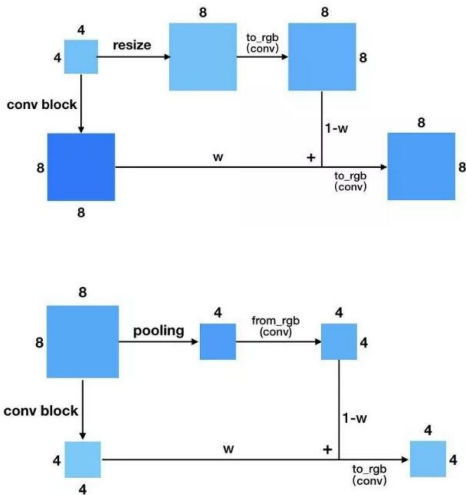
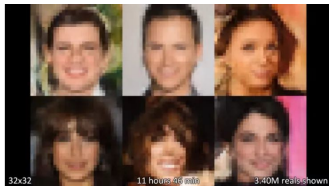
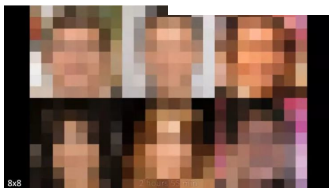








Figure 12: PGGAN的Generator (上) 和Discriminator (下)



-  Martin Arjovsky, Soumith Chintala, and Léon Bottou.
Wasserstein gan.
arXiv preprint arXiv:1701.07875, 2017.
-  Emily Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus.
Deep generative image models using a laplacian pyramid of adversarial networks.
arXiv preprint arXiv:1506.05751, 2015.
-  Carl Doersch.
Tutorial on variational autoencoders.
arXiv preprint arXiv:1606.05908, 2016.

-  Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio.
Generative adversarial nets.
In Advances in neural information processing systems, pages 2672–2680, 2014.
-  Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros.
Image-to-image translation with conditional adversarial networks.
arXiv preprint arXiv:1611.07004, 2016.
-  Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen.
Progressive growing of gans for improved quality, stability, and variation.
arXiv preprint arXiv:1710.10196, 2017.

-  Diederik P Kingma and Max Welling.
Auto-encoding variational bayes.
arXiv preprint arXiv:1312.6114, 2013.
-  Mehdi Mirza and Simon Osindero.
Conditional generative adversarial nets.
arXiv preprint arXiv:1411.1784, 2014.
-  Alec Radford, Luke Metz, and Soumith Chintala.
Unsupervised representation learning with deep convolutional generative adversarial networks.
arXiv preprint arXiv:1511.06434, 2015.

<https://github.com/carpedm20/DCGAN-tensorflow>

<https://github.com/affinelayer/pix2pix-tensorflow>

<https://github.com/luyi666/VAE>

<https://github.com/luyi666/Car-Image-Generation>

Questions

提问?