

# paper 笔记

Huan Jia Lu

2023-2024

## GAN 系列

### 论文 01

**题目:** Conditional Generative Adversarial Nets (CGAN)

**信息:** Simon Osindero ins:Deepmind,

**摘要:** GAN 最近被引入作为训练生成模型的一种新颖方法。在这项工作中，我们介绍了生成对抗网络的条件版本 (CGAN)，它可以通过简单地输入我们希望以生成器和判别器为条件的数据  $y$  来构建。我们证明该模型可以生成以类别标签为条件的 MNIST 数字。我们还说明了如何使用该模型来学习多模态模型，并提供了图像标记应用的初步示例，其中我们演示了该方法如何生成不属于训练标签的描述性标签。

**关键词:** GAN; Mnist,Conditional GAN

**总结:** 在 CGAN (条件 GAN) 中，标签充当潜在空间  $z$  的扩展，以更好地生成和区分图像。top 图1是常规的 GAN，bottle 图1为生成器和判别器添加了标签，以更好地训练两个网络。

**问题:** Goodfellow 提出了 Generative Adversarial Networks，在论文的最后他指出了 GAN 的优缺点以及未来的研究方向和拓展，其中他提到的第一点拓展就是：A conditional generative model  $p(x|c)$  can be obtained by adding  $c$  as input to both  $G$  and  $D$ 。这是因为这种不需要预先建模的方法缺点是太过自由了，对于较大的图片，较多的 pixel 的情形，基于简单 GAN 的方式就不太可控了。于是我们希望得到一种条件型的生成对抗网络，通过给 GAN 中的  $G$  和

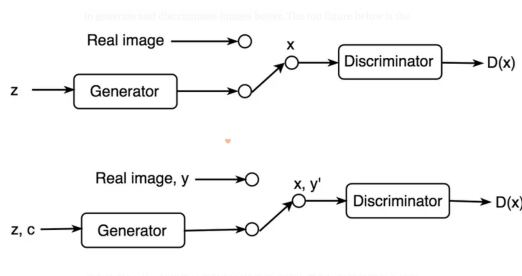


图 1: core

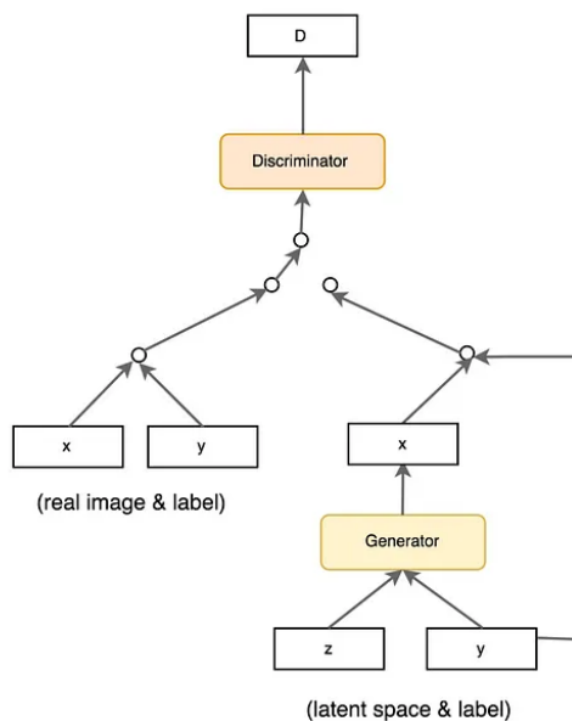


图 2: core

D 增加一些条件性的约束，来解决训练太自由的问题。于是同年，Mirza 等人就提出了一种 Conditional Generative Adversarial Networks .

**方法:** 原始的 GAN 的生成器只能根据随机噪声进行生成图像，至于这个图像是什么（即标签是什么我们无从得知），判别器也只能接收图像输入进行判别是否图像来使生成器。

因此 CGAN 的主要贡献就是在原始 GAN 的生成器与判别器中的输入中加入额外信息  $y$ 。额外信息  $y$  可以是任何信息，例如标签。

CGAN 的提出使得 GAN 可以利用图像与对应的标签进行训练，并在测试阶段利用给定标签生成特定图像。这是一种带条件约束的生成对抗模型 (GAN)，它在生成模型 (G) 和判别模型 (D) 的建模中均引入了条件变量  $y$ ，这里  $y$  可以是 label，可以是 tags，可以是来自不同模态是数据，甚至可以是一张图片，使用**这个额外的条件变量**，对于生成器对数据的生成具有指导作用，因此，Conditional Generative Adversarial Networks 也可以看成是把无监督的 GAN 变成有监督模型的一种改进，这个改进也被证明是非常有效的，为后续的相关工作提供了指导作用。

**内容:** .

**实验:**





	User tags + annotations	Generated tags
	montanha, trem, inverno, frio, people, male, plant life, tree, structures, transport, car	taxi, passenger, line, transportation, railway station, passengers, railways, signals, rail, rails
	food, raspberry, delicious, homemade	chicken, fattening, cooked, peanut, cream, cookie, house made, bread, biscuit, bakes
	water, river	creek, lake, along, near, river, rocky, treeline, valley, woods, waters
	people, portrait, female, baby, indoor	love, people, posing, girl, young, strangers, pretty, women, happy, life

Table 2: Samples of generated tags

图 3: tags

## 0.1 Part A Unimodal

首先测试了 cGAN 生成手写数字，条件信息是数字类别，以 one-hot 的形式输入。

Generator 中：维度为 100 的随机噪声  $z$  从同一分布中采样生成，除此之外还有 label 作为  $y$  一同输入，最后一层 sigmoid 的 784 维的输出作为生成的样本。

Discriminator: 负责将  $x$  映射到 maxout 层，包含 240 个单元共 5 块，中间的隐藏层也包含 240 个单元最终送入 sigmoid 层。

训练：SGD 优化，batch\_size=100，学习率初始值为 0.1，指数衰减。最终的生成图片详见 Fig2，每一行代表不同标签，每一列对应不同标签下的 10 例生成样本。

## 0.2 Part B Multimodal

像 Flickr 等用户产生的元数据不同于较为全为标注过的数据，更贴近于使用自然语言描述图像的语义信息，而不仅仅局限于标注图像内存在的物体。这种 UGM 数据一个很显著的特征就是同义性，用户常用相似的近义词来描述同一概念，因此对这种标签进行归一化就十分有用，概念词嵌入技术 (conceptual word embedding) 是一种很有效的技术手段，相近的概念通常具有相近的特征向量。

本文就利用 cGAN 对图像进行自动打标，产生多模态的分类预测。实验使用的是 MIR Flickr 25000 数据集，跳过了其中未标注的图像。Generator 输入的是随机噪声以及在 100 张样本中出现最多的 10 类标签，输出生成的标签<sup>3</sup>。

**核心：** 在之前的文章中，我们提到了 Generative Adversarial Networks 实际上是对 D 和 G 解决以下极小化极大的二元博弈问题：

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

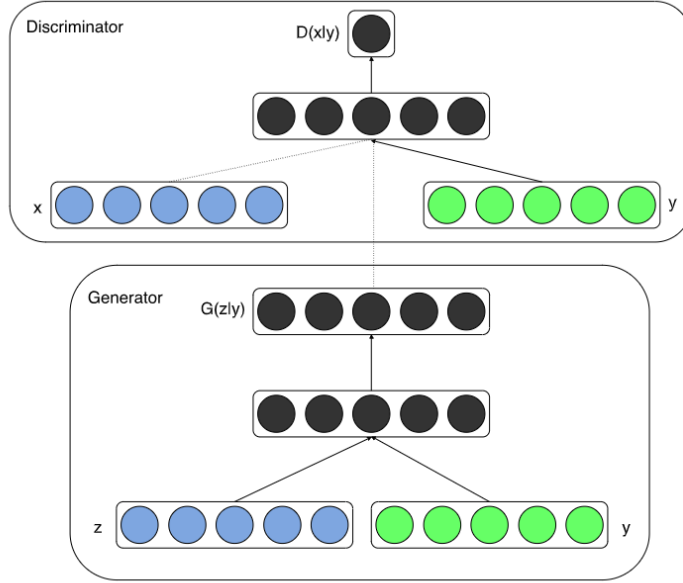


Figure 1: Conditional adversarial net

图 4: CGAN

而在  $D$  和  $G$  中均加入条件约束  $y$  时，实际上就变成了带有条件概率的二元极小化极大问题：

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x | y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z | y)))] \quad (2)$$

在生成器模型中，条件变量  $y$  实际上是作为一个额外的输入层（additional input layer），它与生成器的噪声输入  $p(z)$  组合形成了一个联合的隐层表达；在判别器模型中， $y$  与真实数据  $x$  也是作为输入，并输入到一个判别函数当中。实际上就是将  $z$  和  $x$  分别于  $y$  进行 **concat**，分别作为生成器和判别器的输入，再进行训练。其实在有监督的 DBN 中，也用到了类似的做法。

架构图：见图4

code:

文献:

- [1] Generative Adversarial Networks
- [2] Conditional Generative Adversarial Nets

---

论文 02

题目: 文章标题

信息: 作者, 杂志和发表时间,

.....

$$\sum_{j=1, j \neq i}^n \beta_{i,j} P_j \cdot P_i \tag{3}$$

## 医学影像分割