春节过得差不多了，研究还得继续。前几天出了一篇引起不小轰动的GAN文章，介绍了一种新的GAN——WGAN（Wasserstein GAN）。文章对GAN做了较深入的研究，理论、实践做得都不错。据文章介绍，WGAN效果很好，能够很好地解决GAN不稳定的问题，discriminator的输出不再是是否属于自然图像的概率，而是尽可能地使得自然图像的响应高，而generator产生的图像响应低。文章一出，在reddit上引起了热议。下期我们将对WGAN作深入研读。

本期的主题是LAPGAN，将拉普拉斯金字塔框架引入到GAN中，与一般的GAN（general GAN）直接产生图像不同，generator产生不同层次（level）上的残差图像，以此构造最终的图像。LAPGAN基于CGAN（条件生成对抗式网络），在介绍LAPGAN之前，我们先来介绍CGAN以及Laplacian Pyramid。

PS: 上期预告中提到了本期会介绍GAN的一种应用：根据文字描述产生图像。本期暂不介绍这部分内容（我偷懒了...），我们将它与其他GAN应用一起留到后面，作为GAN专题的一个结尾，预计在下下期推出，给自己立个flag，防偷懒...

**CGAN**

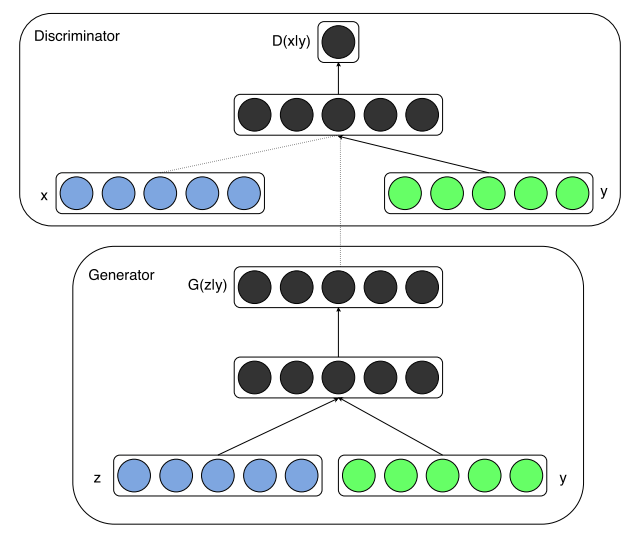
Conditional generative adversarial network，条件生成对抗式网络与general GAN的区别在于，CGAN引入了额外的信息，比如，在MNIST上可以将label的信息作为额外信息提供给generator和discriminator。

GAN很难训练的一个原因可能是搜索空间太大，容易被引导“误入歧途”。给GAN提供多一些信息，能够大大地缩小搜索空间，或许能够得到更好的结果。当然，还得看额外信息怎么利用。

不难理解，CGAN的loss function如下：



下图是CGAN的一个简单版本的示例：



CGAN的论文在MNIST数据集上，将label作为额外信息用于引导产生的是哪个数字；此外，作者还用它产生图像的标签，不过，从展示的效果来看，似乎并不是很好。

**Laplacian Pyramid**

Laplacian Pyramid是一种线性可逆的图像表示方法。下面的介绍将用到两个记号：



对图像I，我们可以构造一个具有K个level的高斯金字塔（Gaussian pyramid）:



其中



根据高斯金字塔，我们可以得到拉普拉斯金字塔（Laplacian pyramid）：



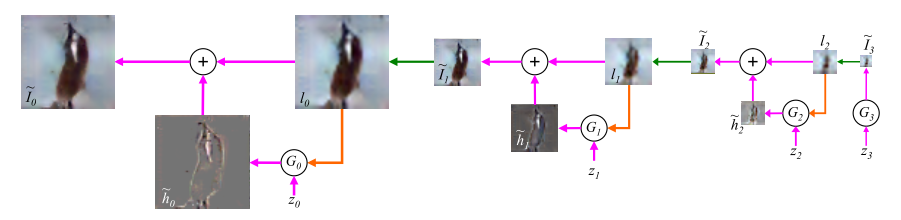
可以将h\_k形象地看成是残差图像，即根据下一层（低分辨率）经过高斯滤波的图像上采样重构图像造成的图像质量损失。

如果你对高斯金字塔和拉普拉斯金字塔不熟悉，建议再深入了解一下，它们在深度学习统治图像领域以前，是应用非常广泛的一类方法。

**LAPGAN**

我们将LAPGAN分为sampling和training两部分分别进行介绍。比较特别的是，LAPGAN不适合将generator和discriminator拆开单独介绍。LAPGAN的generator实际上是多个GAN的generator协作；而discriminator也是多个GAN的discriminator协作。

先来介绍**sampling**。过程图解如下图：



LAPGAN对给定低分辨率图像，产生高分辨率图像的残差图像，以此得到高分辨率图像。上图给出的是分为3个分辨率恢复的流程。最底层图像通过general的GAN产生：即通过高斯噪声z\_3产生一个低分辨率（如4\*4或者8\*8）的图像I’\_3，然后通过对I’\_3上采样得到I\_2，作为第二层GAN的额外信息，结合随机产生的高斯噪声z\_2，通过CGAN产生第二个level图像的残差图像h\_2，将I\_2与h\_2相加得到最终的第二个level的图像I’\_2。以此类推得到特定较高分辨率的图像。第k个generator完成以下任务：



LAPGAN将图像的生成过程拆分成多个GAN与CGAN协同完成的过程，每个GAN负责产生的信息比general的GAN要少。上图给出的是3个level，4个GAN generator（1个GAN generator+3个CGAN generator）协作产生图像的过程。

了解了sampling的过程，就不难理解training了。下图给出了training的流程：



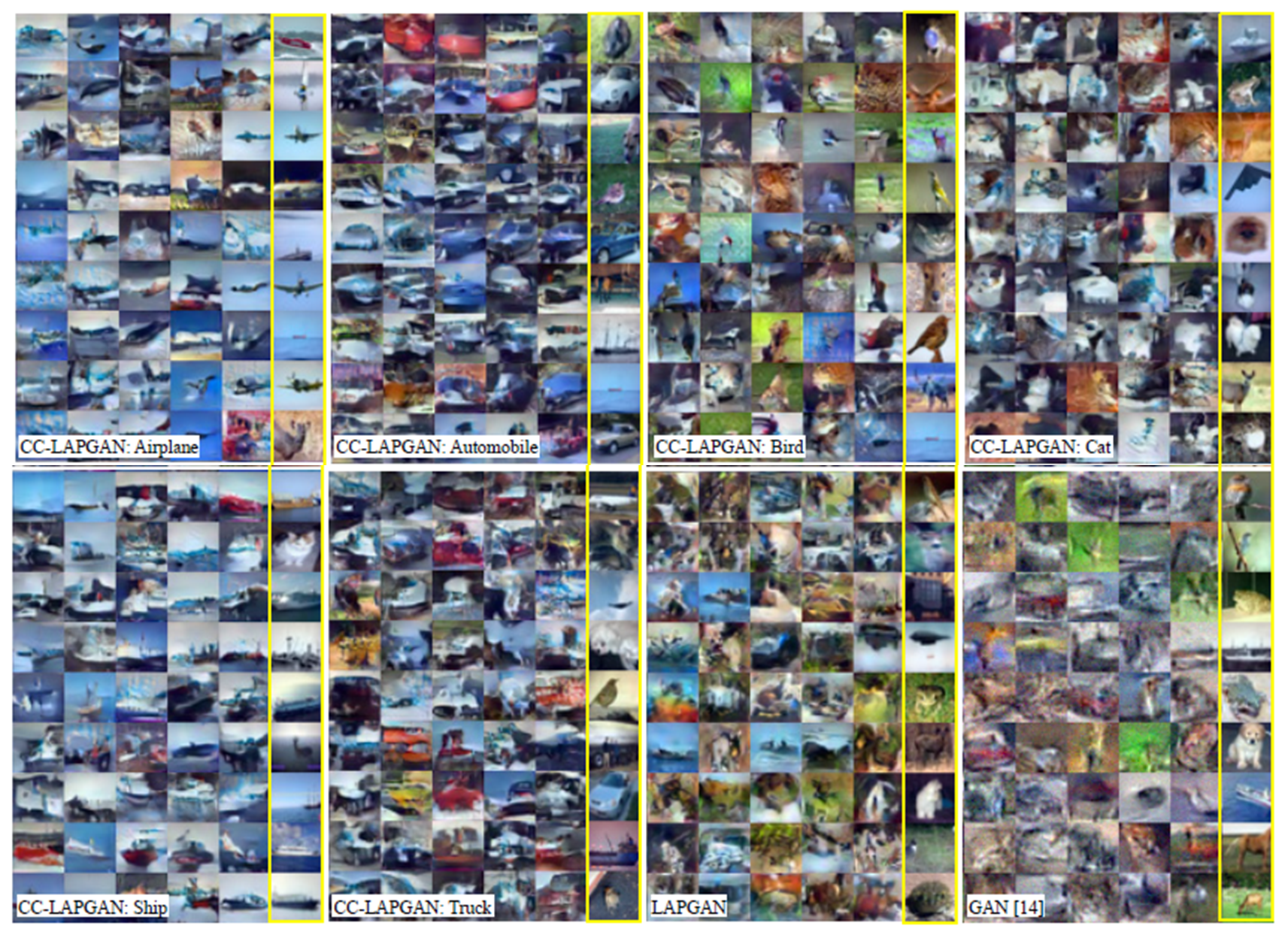
Training的过程融合了sampling的过程。LAPGAN的discriminator并不是直接输入一张图像，经过一个网络判断它是自然图像还是产生的图像，而是从多个level上进行判断。除了最底层的discriminator（D3）的输入是图像之外，其余discriminator（D0, D1, D2）的输入实际上是残差图像。

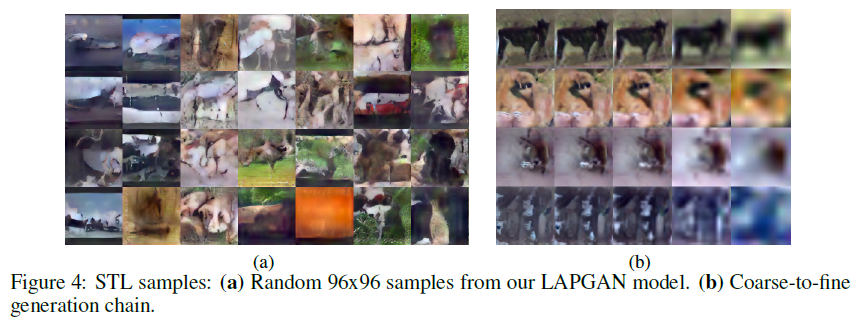
Discriminator的训练集由两部分构成real image和fake image，下面以D0为例进行介绍。对于输入图像I\_0，下采样得到图像I\_1，然后对I\_1上采样得到I’\_0，作为G0（CGAN generator）的条件信息，G0产生残差图像h’\_0，它是D0训练集的fake image部分；而真实的残差图像h\_0 = I\_0 – I’\_0是D0训练的real image部分。

LAPGAN将Laplacian pyramid深度融合到GAN的框架中，上面的图解介绍的整个LAPGAN总共用到了4个GAN generator（G0, G1, G2, G3）以及4个GAN discriminator（D0, D1, D2, D3）。高分辨率图像的产生对于GAN来说还是很大的挑战，LAPGAN能在一定程度上解决这个问题。

**LAPGAN实验**

作者在CIFAR10和STL、LSUN数据集上做了实验，琐碎的实验预处理细节可以参考原文献[2]。上面在介绍LAPGAN时并没有将label信息引入到CGAN中，可以将label作为条件信息作为CGAN的一部分输入（这种用了label信息的LAPGAN作者称之为CC-LAPGAN），下图展示了LAPGAN的实验效果，可以看出，它能够控制产生的图像内容，产生的图像清晰度还不错，比普通的GAN效果要好。





**参考文献**

1. Mirza, M., Osindero, S. (2014). Conditional Generative Adversarial Nets. *Arxiv*.
2. Denton, E., Chintala, S., Szlam, A., & Fergus, R. (2015). Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. *Arxiv*.

**下期预告**

下期我们将研读Wasserstein GAN，这篇文章在reddit上引起了热议，借用rumblestiltsken在reddit上发表的观点，WGAN的论文具有以下极具吸引人的内容：

* GAN 产生的图像质量与loss function相关，并用基于EM（Earth Mover）距离的loss代替JS 散度的loss
* GAN的loss function确实会收敛，因此GAN的超参数可以用其他方法选择
* GAN的训练更稳定，无需batch normalization，即使是MLP（多层感知器）也能产生质量不错的图像
* 提出了减少mode collapse的方法
* 从理论上证明了WGAN为什么能够克服general GAN遇到的多种问题

仅仅看大家的讨论就让我对这篇论文产生了极大的兴趣。GAN在过去一年虽然很火，但是高水平的论文却是少之又少，这篇论文或许就是我一直在期待的高水平论文。下期我们就来精读这篇文章。感兴趣的同学可以跟我一起研读和讨论。参考文献(arxiv 1701.07875v1)：

Arjovsky, M., Soumith, C., & Bottou, L. eon. (n.d.). Wasserstein GAN.

此外，GAN专题很快就要告一段落了，还是很舍不得呢，GAN采用博弈的思想是如此的巧妙，它打开了非监督学习的一扇大门，等待我们去探索无限的可能，如果你也对GAN感兴趣，希望你在GAN专题落幕之后仍能保持对GAN的关注。

以下内容不写进正文：

你可以试试改进GAN，像WGAN一样做深入的理论研究，你也可以做GAN的应用，文字生成图像、草图的具象化...更酷的应用等待着你去开发；甚至，你可以引入协作的思想，创立新的一种方法。我把我的想法都告诉你了，我们来竞赛，下一篇高质量的GAN论文出自谁之手。