**学生工作日报**

|  |  |  |  |
| --- | --- | --- | --- |
| **姓 名** | **段文静** | **日 期** | **2.14** |
| **工作内容描述 达成情况 预期与解决方案等** | | | |
| 1. **今天的工作内容** 2. **今天学习完了LAPGAN，开始学习DCGAN** 3. **今天的工作总结**   **LAPGAN打破了常规的一次性生成整张图片的方式(global)，转而使用一个带有拉普拉斯金字塔框架的级联卷积网络由粗到细地生成图片，每层独立训练CGAN。之后作者用三种评价方式来评价LAPGAN的生成效果，分别是Parzen window estimate to compute log-likelihoods,结果可视化对比以及人工真伪判断，结果均表明LAPGAN可以起到很好的保真图片还原效果。另外，LAPGAN中的D和G都采用了卷积网络的结果，论文没有细说所以这部分不太清楚。**   1. **存在问题及想法**   **无** | | | |
| **明天的工作计划:**  **学习DCGAN，之后做实验还原。** | | | |

**附件**

参考：

论文：[Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks](https://arxiv.org/abs/1506.05751)

代码：<https://github.com/facebook/eyescream>

补充模型：http://soumith.ch/eyescream

网站：<http://blog.csdn.net/solomon1558/article/details/52562851>

LAPGAN使用一个带有拉普拉斯金字塔框架的级联卷积网络由粗到细地生成图片。LAPGAN将整个图片的学习转化为对样本和生成图像之间的残差学习。将整个训练过程分成多个Laplacian Pyramid level，每个level使用条件GAN方法训练一个独立的产生式卷积网络模型，每个模型抓取拉普拉斯金字塔一个特定尺度的图像结构。

1. 目标函数(损失函数)

LAPGAN的目标函数(损失函数)是带有拉普拉斯金字塔框架的条件式产生对抗网络。目标函数如下：

是信息分布，如类别的鲜艳分布。这个生成模型允许输出被条件l控制。在LAPGAN中，l为另一个CGAN模型生成的图片？

1. Laplacian Pyramid

拉普拉斯金字塔是一个线性可逆的图像表征。由一个带通图像集合和一个低频残差组成。定义d()为图像的下采样函数(下采样为原来一半)，u()为图像的上采样函数(上采样为原来2倍)。首先建立一个高斯金字塔(Gaussian pyramid)，，表示原始图像，为第k次的下采样图像。在构建一个拉普拉斯金字塔(Laplacian pyramid) ，表示高斯金字塔相邻层的残差h。表示第k层和第k+1层的残差。

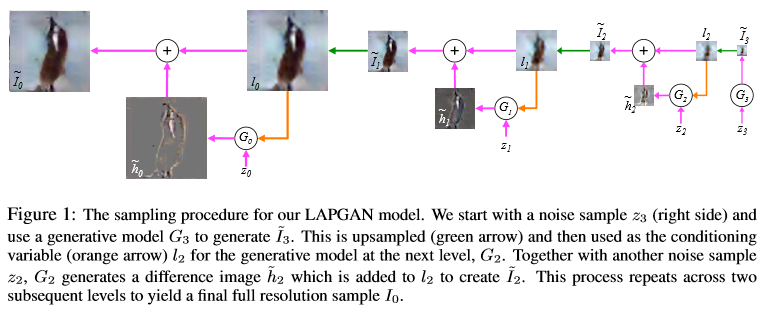
(1)

拉普拉斯金字塔的最后一层低频残差，用来代替，。同理，若要从后往前恢复整个图片，则用公式得到。

1. LAPGAN(Laplacian Generative Adversarial Networks)
2. LAPGAN得到完整图片(sample procedure)

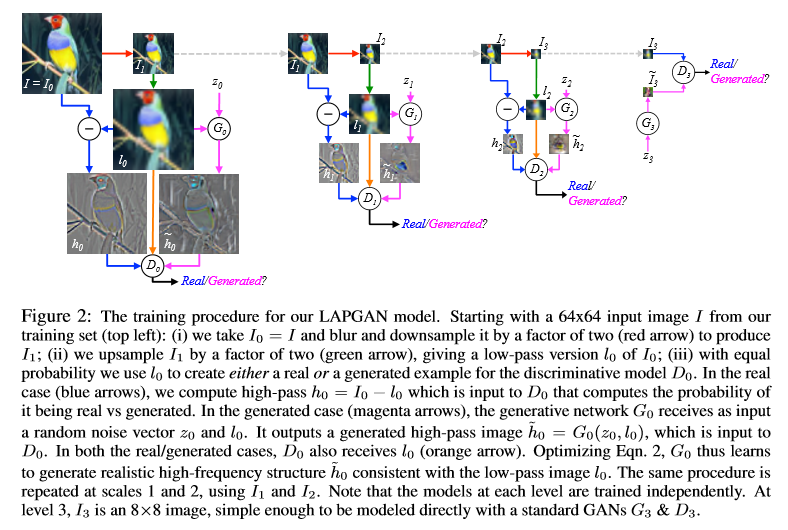
已知有一组生成卷积模型，每个生成模型会输出一个残差，重构图片的过程如下：

这个重构过程以开始，用最后一层的生成模型，输入来生成残差：。除最后一层外，其余层都是条件生成模型，将当前图片的上采样作为条件。



1. 训练过程(training procedure)

对于每一张输入的图片I构造一个拉普拉斯金字塔，令，下采样生成。对上采样，得到的一个低通版本。等概率地为模型选择一个真实样本或生成样本进行判断。若选择真实样本，则计算高通残差；若选择生成样本，则计算高通残差。接下来，将残差和条件共同输入到判别模型中，则输出一个标量代表其属于真实样本的可能性。优化目标函数使能够生成逼真的高频残差。在后边的层中也用同样的方法进行训练。最后一层的下采样样本已经足够小，此时模型变为标准的GAN，。



1. 模型优点

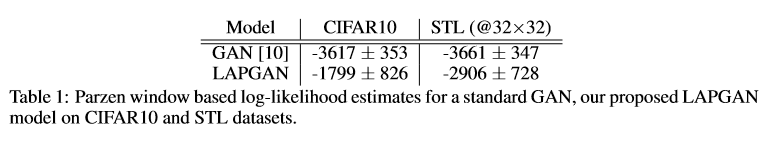
此模型为无监督训练模型，每层的模型独立训练，放弃了传统的一气呵成的思想，生成图片的过程是由粗到细逐渐生成，保真度较高。

1. 实验

对三个数据库进行实验：CIFAR10，STL，LSUN，结果用三种方式进行衡量：

1. Parzen window estimate to compute log-likelihoods，值越大越好。

将LAPGAN和传统GAN效果进行比较，结论是LAPGAN的图片生成效果好

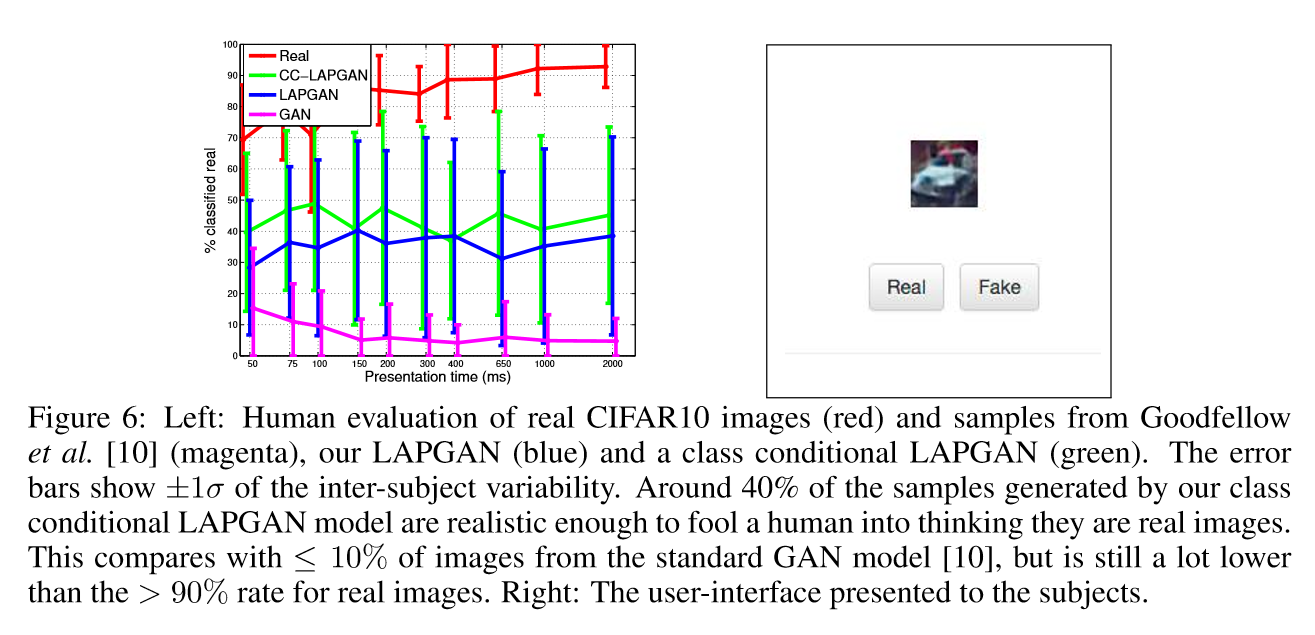


1. 将结果进行可视化比较

试验中采用三种结构，class conditional LAPGAN(CC LAPGAN)，LAPGAN，GAN。可视化结果按照类别归类展示。LAPGAN生成的样本更加目标化，目标的边缘更加清晰，而CC LAPGAN在LAPGAN的基础上使得目标结构更加清晰。最右侧的一列是训练样本，该列样本于其相邻列生成样本的L2像素距离最接近当不相同，也能说明LAPGAN模型没有记忆训练样本，而是独立生成样本。另外LAPGAN还可以合成很复杂的图片(教堂前景，塔楼，卧室)，目前现有的生成模型中没有模型能达到如此复杂度。见论文图3,4,5。

1. 人工评估

给评估志愿者展示四种类型的图片：真实图片；由GAN合成的图片；由LAPGAN合成的图片；由CC LAPGAN合成的图片。志愿者需要在规定时间内判断图片是真实的还是合成的。左侧图为结果统计曲线，横轴是展示时间。结果显示，随着展示时间的延长，GAN生成图片的效果会越来越容易被认出来。CC LAPGAN的仿真效果达到了40%左右，相比于GAN的10%仿真效果来说好了很对，但对比真实图片的高达90%的真实度，CC LAPGAN仍需提高。



|  |  |  |  |
| --- | --- | --- | --- |
| **姓 名** | **段文静** | **日 期** | **2.13** |
| **工作内容描述 达成情况 预期与解决方案等** | | | |
| 1. **今天的工作内容**     * + 1. **学习论文Conditional Generative Adversarial Nets，在GAN的基础上加条件训练，约束GAN使其更易收敛。条件的形式没有限制，在产生模型和判别模型中加入相同的条件即可。**        2. **学习LAPGAN，该模型在生成高分辨率图像上效果很好。** 2. **今天的工作总结**   **今天先学习了条件生成对抗网络，这种网络简单粗暴，条件视情况而定，在后来GAN的发展中几乎都加了条件。在这之后又开始学习LAPGAN，使用拉普拉斯金字塔结构，使得GAN在生成高分辨率图像时可以逐步生成，对于高分辨率图像的生成很有效果。**   1. **存在问题及想法**    * + 1. **条件生成对抗网络中没有具体说怎样将条件和输入（x,z）进行融合，只是大概举了两个例子，一个数字生成例子，一个多模式图像标注例子，输入和条件的融合都是通过隐藏层(maxout, ReLU)来实现，为什么这么用不清楚。** | | | |
| **明天的工作计划:**  **继续学习LAPGAN。** | | | |

**附件**

参考

论文：[Conditional Generative Adversarial Nets](https://arxiv.org/abs/1411.1784)

网站：<http://blog.csdn.net/solomon1558/article/details/52555083>

<http://blog.csdn.net/wspba/article/details/54666907>

GAN不需要预先建模就可进行训练，训练太过自由，对于像素较较多的图片，过于庞大的数据库，该种训练方法不太合适，要通过给GAN中的G和D增加一些约束性条件来解决训练过于自由的问题。

CGAN是给生成模型G和判别模型D的建模中引入条件变量y，y的类型据情况而定，这个改进被证明是非常有效的。

典型应用是图像的一对多标签问题。一个图像可以有多个描述标签，不同的人标注的结果不同，使用CGAN也可对图像进行多个标签的描述， 以输入的图片作为条件，使标注的词语更加全面，人工标注的结果可视为辅助的tag，不是groundtruth。

1. 目标函数

Y通过一个额外的输入层分别输入到G和D中。在生成模型中，y和z先经过一个隐藏层进行组合后再输入到生成模型中；在判别模型中，x和y也通过一个隐藏层合并后输入到判别模型中，如图所示：

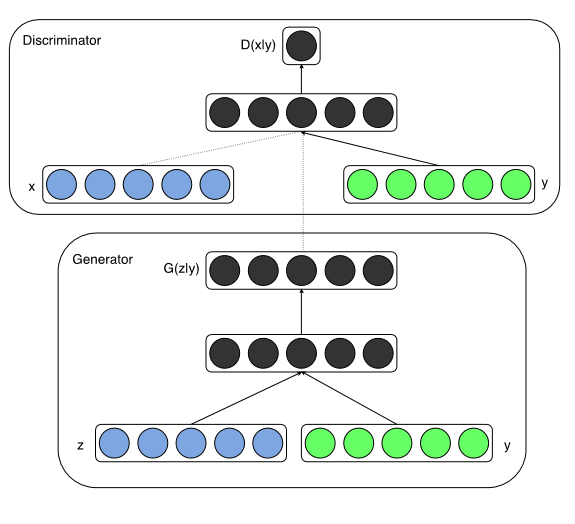


图1 Conditional GAN

1. 两个实验
2. MNIST数据集实验

在MINIST数据集上以类别标签(class labels)作为条件，标签被one-hot编码。根据标签条件信息，生成对应的数字。生成模型的输入为100维服从均匀分布的噪声数据z，z和y分别映射到隐藏层(单元数为200和1000)，用ReLU激活，再将这1200个单元联合，输入到第二层中，经过sigmoid归一化处理后得到784维的生成输出(28\*28单通道图像)。判别模型的输入为784维的输入数据x和one-hot编码的类别标签y。分别映射到隐藏层(240 units and 5 pieces, 50 units and 5 pieces)，maxout层激活。这两个隐藏层再映射到联合maxout层(240 units and 4 pieces)。最后经过sigmoid归一化处理得到输出。

用Parzen window-based log-likelihood estimates做结果评估，数值越大越好。结果表明，条件生成模型的效果与某些模型的效果有可比性，但比GAN要差，后续在超参数空间或条件模型结构上改进后可能会得到更好的结果。

1. 多模态学习(multi-modal)用于图像自动标注

在数据集Flickr 25,000上进行多标签标注，该数据集中包含大量拥有标签的图片，且很多标签是user-generated metadata(UGM, 用户自己产生的标签)，这类标签的特点是更加具有描述性，更加接近自然语言，对于同一幅图片，不同的人会给出同义标记。CGAN以图片特征为条件，产生一个标签向量分布。

对于image representation，用Alexnet提取，以最后一层fcn的4096维输出作为图片特征。对于word representation，先从YFCC100M中获得一个语料库，用它训练一个skip-gram模型，生成200维的语义向量。用卷积模型和skip-gram模型提取Flicker数据集中的图片和tag特征，用来训练条件生成对抗网络。

产生模型的输入为100维的噪声数据和4096维的图片特征数据。噪声数据映射到500维输出的ReLU层，特征数据映射到2000维输出的ReLU层，这两个隐藏层联合映射到一个200维输出的线性层，得到词语向量。

判别模型分别以500维和1200维的ReLU隐藏层作为词语向量和图像特征的隐藏层，并将它们的输出联合输入到一个maxout层(1000 units and 3 pieces)，经过sigmoid单元归一化得到100维的词语输出，从中选出10维最为最终结果。

从实验结果可以看出，如果图片中有很明显的目标，CGAN会更关注于目标的特点来形容，忽略了背景(特征提取导致)。是具象的描述，很少掺杂情感描述词。而图片的人工标注中，既关注了背景信息，又加入了感情色彩。

1. 模型的展望
2. 根据实际应用开发出更加复杂的模型，进行更加细致和全面的分析。
3. 在训练时，以tag为单元进行训练，每回训练一个tag，如果图片有多个tag，就要反复训练多次。作者希望模型可以一次训练多个tag，使训练更加高效。
4. 希望可以用一个统一的模型来做language model，而不是将训练分成中断的几部分。

|  |  |  |  |
| --- | --- | --- | --- |
| **姓 名** | **段文静** | **日 期** | **2.9** |
| **工作内容描述 达成情况 预期与解决方案等** | | | |
| 1. **今天的工作内容** 2. **完成论文Generative Adversarial Nets的初步总结学习，该篇文章首次提出GAN的概念。** 3. **今天的工作总结**   **由于GAN与CNN还是有一定区别的，新上手时学习会比较慢，GAN这篇鼻祖文章主要介绍了GAN中两大模型（生成模型，判别模型）的工作原理，目标函数以及收敛性的理论证明，我在数学推导上仍有一些疑问，可先保留，入门后回过头来再看。**   1. **存在问题及想法**   **1. 在训练过程中，每轮迭代优先保证D在给定当前G下达到最优，再去更新G达到最优，如此循环。可真实训练中每隔k步才训练一回D，他不能保证一回D就训练成功了啊？**  **2. 由于没看代码，网络本身的具体结构还不甚了解，论文在优点处提到了inference和backpropagation，大改表示的意义与神经网络相同，那么Loss function是什么。怎么前推后推我不了解。** | | | |
| **明天的工作计划:**  **明天上午去看病，回来后学习Conditional Generative Adversarial Nets** | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **姓 名** | **段文静** | **日 期** | **2.8** |
| **工作内容描述 达成情况 预期与解决方案等** | | | |
| 1. **今天的工作内容**     * + 1. **参考网站和论文学习生成对抗网络（GAN），先通过网站**[**http://it.sohu.com/20161210/n475485860.shtml简单了解一下GAN**](http://it.sohu.com/20161210/n475485860.shtml简单了解一下GAN)**各部分的概念并总结，再通过14年在arXiv上发表的论文Generative Adversarial Nets系统学习GAN。目前在思考目标函数的优化部分。** 2. **今天的工作总结**   **今天开始学习GAN，GAN也是深度学习的一个分支，DCGAN更是将卷积神经网络和GAN相结合的一个架构。我在学习的过程中发现网站和论文有一些出入，尤其在目标函数的优化部分，希望明天可以将这些出入解决。**   1. **存在问题及想法**    * + 1. 训练部分提到D在训练的内部循环优化的代价很高，数据集有限的情况下会导致过度拟合，因此D每K步优化一次，G每一步优化一次，只要G改变的足够慢，D可长时间保持在最优的状态左右。不太理解这段话的意思？        2. 很容易饱和这段也不太明白，啥叫饱和？ | | | |
| **明天的工作计划:**  **解决上述问题和出入。** | | | |

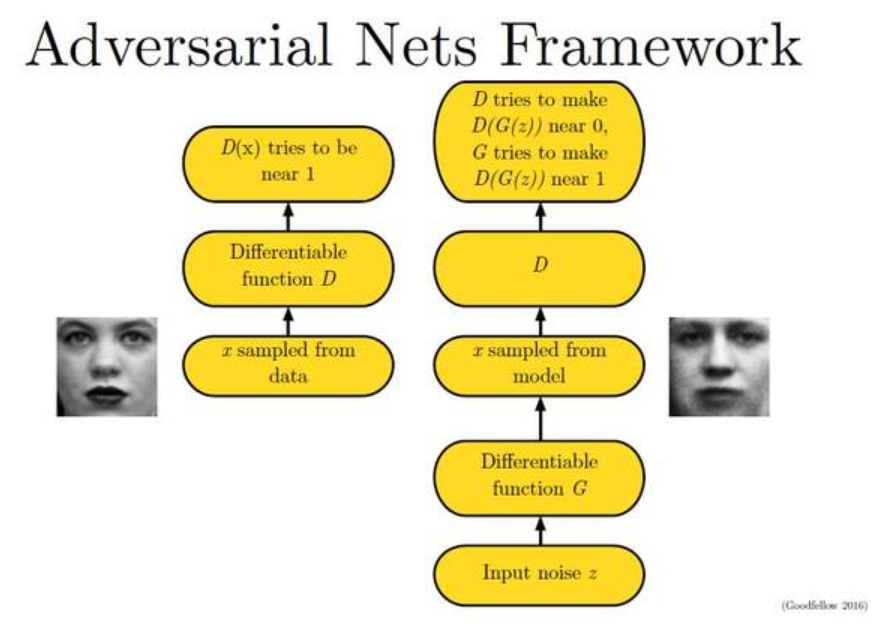
**附件**

参考：

网站：<http://it.sohu.com/20161210/n475485860.shtml>

论文：[Generative Adversarial Nets](https://arxiv.org/abs/1406.2661)

1. 生成对抗网络（GAN）：靠数据和模型自己的内部对抗来实现无监督学习的模型。
2. 生成模型有两种应用：
3. 密度（概率）估计：在不了解事件概率分布的情况下，先假设随机分布，然后通过数据观测来确定真正的概率密度是怎样的。
4. 样本生成：提供训练样本数据分布，根据训练后模型来生成类似的样本分布。
5. 生成对抗模型的架构分成两个部分，分别是判别模型和生成模型。



A是一个生成模型，希望将假数据模仿成真数据；B是一个判别模型，拼命把真数据和假数据分开。对抗学习之后，A的以假乱真技术越来越强，B的鉴别能力也越来越强。

1. 生成模型(Generator Network)

生成模型的目标是将噪音数据z的分布，通过生成模型G，伪装成真实数据x的分布，GAN是一个神经网络，因此需要是可微的(differentiable)。

1. 变量表示
2. 训练数据x，生成模型通过对x data的学习得到分布模型，输入的噪声变量；
3. 将噪声空间通过生成模型映射到data空间的函数，G是一个可微函数，参数为；
4. 判别模型的判别函数输入的值为x输入真实数据data的可能性，而不来自生成器空间；
5. 训练过程(Training Procedure)

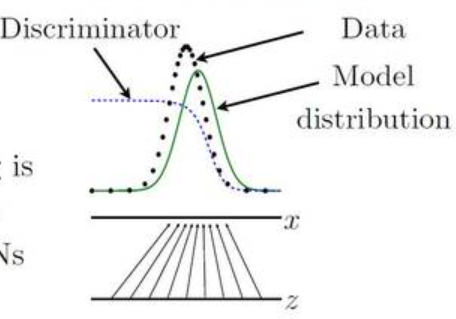
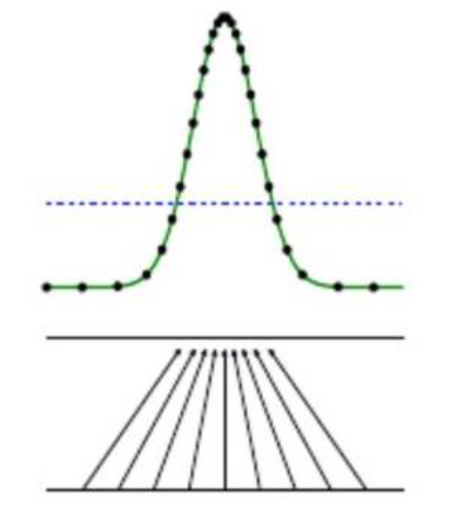
选择任何类SGD(梯度下降)的方法进行训练（A，B两个网络都可微），同时训练两组数据，一组是真实的训练数据(training examples)，一组是由生成模型生成的数据(generated samples)。判别模型的目标是正确区分真实数据和伪造数据；生成模型的目标是最小化判别模型的准确率，使尽可能高，最小化。将训练转化为最大最小博弈问题(minimax game)，得到如下一个统一的优化公式：

公式的前半段表示判别模型评估的数据来源于data空间，即真实数据空间，值越大越好；公式后半段表示判别模型评估的数据来源于噪声空间，评估结果越小越好，但右半部分值越大越好；因此优化D，固定G时，向最大化方向发展；优化G，固定D时，向最小化方向发展。D在训练的内部循环优化的代价很高，数据集有限的情况下会导致过度拟合，因此D每K步优化一次，G每一步优化一次，只要G改变的足够慢，D可长时间保持在最优的状态左右。不太理解这段话的意思？

很容易饱和这段也不太明白，啥叫饱和？

真实数据分布（黑点）和模型生成的伪数据分布（绿线，model distribution，横轴表示数据，纵轴表示分布）如图所示，向上的箭头为z到x的映射，，分布差异D（蓝色虚线）表示为

训练网络的目的就是使生成模型伪造的数据分布与真实数据分布尽可能无差别，D会趋于，借此会训练出最优的生成模型。

|  |  |  |  |
| --- | --- | --- | --- |
| **姓 名** | **段文静** | **日 期** | **2017/2/7** |
| **工作内容描述 达成情况 预期与解决方案等** | | | |
| 1. **今天的工作内容**     * + 1. **为车辆检测的两个版本写运行脚本，分别为训练脚本和可视化脚本。存储在服务器端的faster\_rcnn\_1和faster\_rcnn\_2，分别为train.sh和test.sh。** 2. **今天的工作总结**   **今天将车辆检测的两个版本做了整理，写了运行脚本。训练脚本不带参数，因为数据库的路径，数据库名称以及其他信息已经在程序中写死了，若要全部变成变量的形式需要大改，怕越改越乱，如后边有数据库的更改可参照文档中的内容一个个文件来改，也不是很麻烦。测试脚本需要输入的参数为测试图片目录以及测试图片文件列表，如test.txt。写上列表更有助于程序读图片，适用于多个图片和单个图片的情况。如果要对数据库中的测试数据集做检测就更方便了，直接读test.txt即可，不需将测试图片从总数据集中摘出。**   1. **存在问题及想法**   **基于fast\_rcnn的车辆检测已经告一段落，目前没有什么改进的想法。接下来学习“生成对抗网络”和“增强学习”，再想办法提升车辆检测的内容。** | | | |
| **明天的工作计划:**  **学习生成对抗网络** | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **姓 名** | **段文静** | **日 期** | **2017.2.6** |
| **工作内容描述 达成情况 预期与解决方案等** | | | |
| 1. **今天的工作内容**     * + 1. **整理之前做过的《车辆检测》和《去货车后的车辆检测》项目和文档。**        2. **翻阅本科毕业设计指导手册，总结毕设要求。** 2. **今天的工作总结**   **今天主要是将之前做的内容做个收尾。毕设要求中并没有提到一些具体的毕设内容要求，我现在就有点迷茫不知道下一步咋做了。想着是不是再读读文献啥的，找找新思路，或者看看目前车辆检测方面都到哪一步了，具体方法有哪些等等，我怕自己再跑偏就没深看，想明天找您讨论一下下一步的工作。另外学校的申请工作基本上做完了，余留了四所学校由于春节放假有些文件没有办完，节后继续办，可能白天会穿插处理一下这些内容，望老师见谅。**   1. **存在问题及想法** | | | |
| **明天的工作计划:**  **与老师讨论下一步怎么做。** | | | |

毕业设计

1. 题目

中文：基于深度卷积神经网络的车辆检测方法及应用

英文：Deep convolutional neural network based vehicle identification and its application

1. 题目类型：研究设计类

要求：属实际应用型，期在某一个关键技术上有所创新。他不一定要求做出工程产品或可事实规范。

1. 时间节点

2.27～3.5（开学第1周）：收到任务书

2.27～3.26（第1到4周）：学生写开题报告综述，内容包括选题的背景和意义，研究的基本内容和拟解决的主要问题，研究方法及措施，研究工作的步骤和进度，主要参考文献等项目。不少于1500字。

4.10～4.23（第7到8周）：学生写《中期进展检查表》，包括目前已完成的任务，尚需完成的任务。存在问题和拟采取的办法。

5.15～6.11（第12到15周）：答辩