2019届研究生硕士学位论文

分类号： 学校代码：10269

密 级： 学 号：51164500035



**East China Normal University**

**硕士学位论文**

**MASTER’S DISSERTATION**

**论文题目：****基于生成式模型的不完全数据补全**

**院 系： 计算机科学与软件工程学院**

**专 业： 计算机科学与技术**

**研究方向： 机器学习**

**指导教师： 沈超敏 副教授**

**学位申请人： 余米雪**

2018 年 5月 22日

Thesis for Master's degree in 2019 School Code: 10269

Student NO.: 51164500035

**EAST CHINA NORMAL UNIVERSITY**

Title： Data Inpainting Based on Generative Models

|  |  |
| --- | --- |
| Department: | Computer Science and Software Engineering |
| Major: | Computer Science and Technology |
| Research Area: | Machine Learning |
| Tutor: | Associate Professor Chaomin Shen |
| Author: | Mixue Yu |

**Fulfilled in May, 2019**

**余米雪硕士学位论文答辩委员会成员名单**

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 职称 | 单位 | 备注 |
| 张桂戌 | 教授 | 华东师范大学计算机系 | 主席 |
| 周爱民 | 教授 | 华东师范大学计算机系 | / |
| 张倩 | 副教授 | 华东师范大学计算机系 | / |

# 摘 要

近些年来，随着海量数据的产生和收集模式的进步，国内外工业界和学术界均开始关注如何利用数据更好的服务企业。通过对这些数据的分析和处理，可以更好的产生经济效益。这些数据产生的同时，又难免存在着偏差和噪声。图像数据作为整体数据的一个子集，具有直观性、显著性的特点。本文着重关注的是如何处理图像数据的噪声，在不借助隐空间表达的前提下，利用生成网络去处理有噪声的图像数据。通过学习数据的原始分布来补全未知的内容。

当下，有两种传统的方法来解决这个问题。其一是基于纹理合成的补全方法，该方法从图像边缘递归的填充图像缺失块，这种方法被证明是目前最有效的算法之一。该算法不足之处在于它必须按像素依次填充，补全内容只能是图片中已出现的纹理。另一种是偏微分方程求解模型的方法，它将图像补全问题变成一个带约束的最优化问题。该方法虽然能保存渐变区域的光滑度，并锐化边缘，但是需要长时间的求解迭代，而且容易破坏图片中物体的连通性。这两种传统方法的问题在于，当遇到图片中需要的部分被遮挡时，他们无法解决好这种情况。

生成对抗网络在图像生成应用中效果显著。当数据没有噪声时，它可以拟合数据的真实分布，并产生清晰丰富的样本。但是当数据存在噪声时，生成对抗网络训练的结果会大打折扣，甚至无法产生有效样本。为了解决这个问题，AmbientGAN采用一个约束极强的条件，它通过假设已知样本数据的损失分布，让生成样本服从这个分布，使得该网络可以生成理想样本。

我们将AmbientGAN改成补全网络，并借鉴传统方法中利用已知数据处理未知数据的思想。对于图像中未缺失的部分，我们采取自编码器的架构，使得生成补全网络可以还原出数据原始的分布，并能重构出原始无缺失区域的图像。此外，本文利用特殊的图像补全技巧，将补全方法应用于超分辨任务中，并取得较好的成效。

AmbientGAN补全网络要训练到理想状态，完全依赖于数据的损失度量。一旦损失度量参数出现偏差，补全网络将无法实现。由于现实中并不是所有数据都存在缺失，因此，对于数据集中不完全有噪声情况，我们抛弃AmbientGAN的约束提出了一个新的模型。该模型首先从数据集中挑出部分无噪声的数据，这些数据用来作为判断网络的原始数据分布的输入。该方法避免了直接采用真实缺失标签作为监督信息，也能取得较好的效果。

**关键词**：图像补全，生成对抗网络，自编码器，超分辨，损失测量

# ABSTRACT

With the rapid development of data science and machine learning, in recent years, both industrial and research community have shown a prosperoty in data analyzing and modeling. However, in the real world, unbiased clean data can often be too expensive to be obtained, while the vast majority of the data comes with bias, loss and noise. In this paper, we address the issue of modeling the image data with different types of lossy measures, including random block, noise, compression loss, etc.

Traditionally, there are two approaches to address the issue: the texture based approach and the partial derivative equation (PDE) approach. The former adopts an inpaiting algorithm to recursively restore the pixel from the loss margin, which is proved to be the most effective method up to date. Whereas the algorithm requires to iterate through all pixels in the images, the information of the restored pixels can only comes from the original image. On the contrary, the latter approach formulate the problem as a constraint optimization of image PDEs, which can combine many different priors of the images, as well as significantly preserve the smoothing area and high frequency edges. The major deficiency is that it requires a heavy computational cost to solve the optimization of PDEs, and the connectivity in the original image might be altered during the optimization. Both of the two approaches can not generalize well when a large area in the images are biased or lossed.

As one of the most important framework in generative deep learning, the generative adversarial networks (GAN) are shown to be highly efficient and effective in modeling the data distribution and generating high quality images when unbiased and clean data are observable and accessible. When the images are biased and noisy, the model is prone to mode collapse. The AmbientGAN solves the issue with a relatively strong constraint. When the clean data is unavailable for both the generator and the discriminator, the AmbientGAN shows that the true underlying distribution can still be recovered if the random projection from the original distribution and the lossy measure converge in distribution. The AmbientGAN is able to generate unbiased samples under several lossy measures even if the clean data is completely unobservable.

In this work, we adopt the AmbientGAN framework on inpainting tasks. Taking advantage of the idea of modeling the lossy measure with observed data in previous works, we adopt an encoder-decoder framework to learn the lossy measure in the data, so as to recover the underlying true data distribution. Our model is able to restore the images with large area of losses. In addition, a specifically designed trick is adopeted to extend the idea to image super-resolution task. Both of the two approaches show promising results.

In AmbientGAN, the lossy measure is pressumed to be known and differentiable. In real world environment, the assumption might be too strong to hold. In our work, by removing the consraint, we propopsed a novel model using only unpaired data. We show that the ditributional matching in the latent space helps the model to distinguish the lossy measure. Extensive empirical evaluations show that our method can achieve desiring performance under the unsupervised framework.

**Keywords:**  Inpainting, GAN,Autoencoder, lossy measure,super-solution

**目 录**

[摘 要 V](#_Toc2867632)

[ABSTRACT VI](#_Toc2867633)

[第一章 绪论 1](#_Toc2867634)

[1.1 研究背景 1](#_Toc2867635)

[1.2 国内外研究现状 2](#_Toc2867636)

[1.3 论文主要工作 3](#_Toc2867637)

[1.4 本文的主要贡献 4](#_Toc2867638)

[1.5 本文的组织结构 5](#_Toc2867639)

[第二章 图像补全技术介绍 7](#_Toc2867640)

[2.1 图像补全相关技术介绍 7](#_Toc2867641)

[2.1.1 Inpainting定义 7](#_Toc2867642)

[2.2 基于纹理合成的inpainting算法 8](#_Toc2867643)

[2.3 偏微分方程求解模型（PDE） 10](#_Toc2867644)

[2.3.1 全变分模型（Total Variation） 10](#_Toc2867645)

[2.3.2Mumford-Shah补全模型 11](#_Toc2867646)

[2.4修复技术的度量标准 11](#_Toc2867647)

[2.4.1 结构相似性（SSIM） 11](#_Toc2867648)

[2.4.2 峰值信噪比（PSNR） 12](#_Toc2867649)

[2.5 传统方法实验结果 12](#_Toc2867650)

[2.6 本章小结 13](#_Toc2867651)

[第三章 生成模型 14](#_Toc2867652)

[3.1 生成模型定义 14](#_Toc2867653)

[3.2 变分自动编码器 15](#_Toc2867654)

[3.2.1 自动编码器 15](#_Toc2867655)

[3.2.2 变分自动编码器（VAE） 15](#_Toc2867656)

[3.3 生成对抗网络 16](#_Toc2867657)

[3.3.1 GAN的优缺点 17](#_Toc2867658)

[3.3.2 GAN的改进 18](#_Toc2867659)

[3.4 GAN的度量函数 19](#_Toc2867660)

[3.4.1 散度 19](#_Toc2867661)

[3.4.2 Inception Score 19](#_Toc2867662)

[3.4 Context Encoder 20](#_Toc2867663)

[3.5 AmbientGAN 21](#_Toc2867664)

[3.6 本章小结 22](#_Toc2867665)

[第四章 基于GAN的缺失数据补全算法 23](#_Toc2867666)

[4.1 问题的提出 23](#_Toc2867667)

[4.2 基于GAN的修复模型框架 24](#_Toc2867668)

[4.2.1 CoordConv Layer 25](#_Toc2867669)

[4.2.2 Mask网络架构 26](#_Toc2867670)

[4.2.3网络建模 27](#_Toc2867671)

[4.2.4 损失函数 29](#_Toc2867672)

[4.3 算法流程 31](#_Toc2867673)

[4.4 实验结果与分析 32](#_Toc2867674)

[4.4.1 实验设备配置 32](#_Toc2867675)

[4.4.2 MNIST 实验 32](#_Toc2867676)

[4.4.3 CelebA实验 35](#_Toc2867677)

[第五章 基于GAN的部分缺失补全算法 39](#_Toc2867678)

[第六章 总结和展望 41](#_Toc2867679)

[6.1 总结 41](#_Toc2867680)

[6.2 展望 41](#_Toc2867681)

[参考文献 43](#_Toc2867682)

[致谢 45](#_Toc2867683)

[附录 攻读学位期间发表的学术论文 47](#_Toc2867684)

**插图和附表清单**

[表 1 状态管理类功能说明 **错误!未定义书签。**](#_Toc511003597)

[表 2代码覆盖率实验结果 **错误!未定义书签。**](#_Toc511003598)

[表 3 性能监测 **错误!未定义书签。**](#_Toc511003599)

[图 2.二.1 Inpainting技术流程示意图。输入图I为计算机领域常用的Lena图，输入I被区域L遮挡后导致脸部关键信息丢失，利用Inpainting技术可以恢复出另一个Lena（该图是Lena 61岁重拍的） 7](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867685)

[图 2.2 Efros的算法概观（图片取自[14]） 9](#_Toc2867686)

[图2.3 基于纹理合成inpainting算法实验效果 9](#_Toc2867687)

[图 2.4 Betalmios的实验效果[] 12](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867688)

[图 3.1 自编码器架构 15](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867689)

[图 3.2 VAE的采样过程 16](#_Toc2867690)

[图 3.3 GAN的计算流程与结构 16](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867691)

[图 3.4 各种改进GAN的框图 18](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867692)

[图 3.5 Context Encoder 网络架构 20](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867693)

[图 3.6 AmbientGAN结构图 21](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867694)

[图 3.7 AmbientGAN实验结果 22](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867695)

[图 4.1 AmbientGAN在不同损失度量下的效果 23](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867696)

[图 4.2 坐标卷积的架构 26](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867697)

[图 4.3 自注意力机制 26](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867698)

[图 4.4 改进版AmbientGAN补全网络 28](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867699)

[图 4.5 小型数据集所用的生成补全网络 33](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867700)

[图 4.6 MNIST 补全实验 34](file:///D:\毕业论文\good_luck\motify_lalala_v1.docx#_Toc2867701)

# 绪论

## 研究背景

近年来随着存储技术的高速发展和计算能力大幅提升，数据的产生和存储获取变得越来越容易。基于这些数据，我们可以分析潜在的数据因子，并进行建模决策研究，挖掘潜在的经济利益。由于大数据能带来无穷的经济效益，无疑成了工业界最热门的主题词。目前互联网企业通过庞大的用户群体和订单业务以及用户行为，收集海量的数据，并对其进行数据建模和分析，以此来预测用户潜在的行为。企业根据这些数据分析的结果，做出决策来达到收益的最大化。

但是在实际的数据收集过程中，因受到硬件、时间、天气、人员等各种因素的影响，数据并不是全量完整的，大部分数据存在缺失以及不同程度的噪声。目前很多学者提出了关于数据缺失问题的解决方案，大多数方法都是基于一些很强的假设与约束条件，或者只能应用于单一的数据源或者有清晰无误的数据标签。然而在这些海量数据中，我们不能保证我们用来建模的数据没有偏差和噪声，也无法对每个数据进行定制化的筛选和建模。

机器学习及深度学习近年来的迅猛发展，为大数据建模带来了理论性可能。深度学习的火热带动了人工智能的发展。Google通过深度学习和强化学习构建了阿尔法狗，它使得需要人工经验来完成的围棋游戏变成一个可泛化的模型。深度学习在医疗领域的应用也很广泛，比如识别癌细胞、发现新药物的医疗辅助作用等等。最重要的是，深度学习已经融入了人类的日常生活中，虚拟场景建模，图像识别及其美化，风格转化等功能随处可见。

机器学习的主要分支有判别式模型和生成式模型两种。判别式模型主要针对分类问题，该模型主要用于判别某个数据属于哪一类，容易实现。而生成式模型主要用于发掘数据中潜在的关系，寻找出一个可以使得更多的数据服从的模型，从而解决数据不足的难题。判别式模型有感知机，SVM，KNN等基于结构风险最小化的方法，而传统的生成式模型是基于一个可解释的似然估计，诸如HMM，马尔可夫随机场，EM算法等。这些模型通过似然建模出一个分布，再从该分布中抽取样本。但是传统的生成模型，往往是模糊不清的，约束性很强，无法真正的与现实数据拟合，尤其当似然复杂不可解的。幸运的是，生成对抗网络的提出为该问题提供了一个可行解。

生成对抗网络基于零和博弈思想，用一个生成模型产生的样本去骗过判别器。判别器被骗后，会不断的加强。同时，生成器也会努力加强生成逼真的样本来骗过判别器，这样双方不断对抗，最终达到一个平衡态。当判别器分辨不出真假时，我们就会认为生成模型学到了真实样本的分布，同时也避免显式求解似然的问题。

成对抗网络要生成清晰的样本，是需要基于两个必要条件，其一是训练样本都是干净无污染的，其二是基于海量数据的前提下。但现实世界收集到的数据，往往是存在缺失，并且有噪声。因此在大数据时代，我们无法用全部完整清晰的数据来进行分析建模。于是问题就转化为，如何用这些带有噪声的数据，来建立能拟合事物本身的模型。

数据的范围太过宽泛，而图片信息直观简便，有成熟的测量标准，为了方便描述，本文只关注基于生成式模型的图像补全。

## 国内外研究现状

早期生成模型的典型解决方案基于贝叶斯概率图模型，通过引入隐变量，用EM算法、变分贝叶斯、MCMC或变分推理来求解[2]。主要包括最大似然估计法，Gibbs sampling，混合高斯模型等。然而，基于推断的生成式模型往往只在很强的假设条件下才可以简单快速求解，而基于马尔可夫性质的模型在高维空间中的采样效率极低，这些特点限制了现有模型对于高维空间中复杂概率分布的表达能力。

近年来，随着深度学习的兴起，一种新的解决方案是将神经网络作为一个复杂的映射函数，通过一个简单的隐变量和一个复杂的映射来表达高维空间上复杂的随机变量。例如，变分自编码器（VAE）[3]通过对传统自编码器的隐层引入高斯分布作为先验，将原始问题看作一个最大似然问题，最大化对数似然的变分下界，在编码任务与生成任务上获得了很好的效果。另一方面，对抗生成网络（GAN）[1]通过引入对抗训练机制，使得生成样本分布在目标函数的Nash均衡点收敛于真实数据分布。由于原始GAN模型的训练不稳定，时常出现梯度消失或模式坍塌等问题，Arjovsky等人从拓扑与测度的角度做出了详细的理论证明，并指出这些问题的根本原因在于原始GAN中不合理的距离度量[4]；他们进一步提出了基于Wasserstein距离的WGAN模型，通过优化1-Wasserstein距离的Kantorovich-Rubinstein对偶形式来训练WGAN[5]。借助于这种思想，WAE在VAE变分下界的优化问题中引入了Wasserstein距离[6]，并理论性地证明了当自编码器在隐层概率分布收敛时，优化重构误差等价于优化原始形式的Wasserstein距离，从而给自编码器的理论与应用提供了良好的保障。

然而，大部分的生成式模型都只能在数据样本质量比较好的基础上工作。在医疗、航空航天等现实应用中，得到无偏、无损或者是fully observable的数据往往十分困难，而有损、有偏或者是partially observable的数据往往可以以较低的成本获得。如何从带有损失的数据中学习，获得无损数据的概率分布表达是机器学习的重要问题。由于从有损数据恢复无损数据是一个典型的欠定问题，传统的压缩感知领域中，一般假设真实无损数据的分布是稀疏的，从而通过0范数优化来求解[10]。近些年的深度学习研究中，Bora等人抛弃了这个条件，只假设重建后的数据分布在生成样本的流形上，证明了GAN和VAE等深度生成式模型在重构无损真实数据分布任务上的收敛性[7]。Yeh等人提出了基于perceptual loss与contextual loss的图像修复模型[8]，在人脸数据集CelebA上得到了很好的效果。Bora等人进一步提出了AmbientGAN[9]，当lossy measurement已知，且观测得到的有损数据分布可以唯一确定真实无损样本分布时，该模型可以用同一种算法框架解决很多种不同类别的数据损失与压缩，并在几种典型的lossy measurement下有收敛保障。

不难看出，抛弃样本空间的线性条件与稀疏性假设，是近年来深度生成模型在处理高维、复杂且有损失的数据分布上一个颠覆性的进步。本文中作者将继续从这个角度出发，试图解决以下问题：（1）现有的模型中假设lossy measurement是已知的，这是一个非常强的数学条件，我们更希望在lossy measurement未知且随机的情况下做数据修补填充；（2）将CoordConv等用于空间关系的深度学习模型整合到生成式模型的框架下；（3）直接利用有损信息来还原无损信息的分布，而不借助隐空间的表达。

生成式模型在深度学习的研究中方兴未艾。然而，高维空间中的维度灾难，深度神经网络本身的整体非线性与局部线性，以及由之带来的model capacity、极强的表达能力[11]与脆弱性[12]，都使得本领域的研究进展十分困难。现有的生成模型还有很多待完善的地方，本文的研究，对工业应用与深度学习理论两方面有显著的意义。

## 论文主要工作

现实生活中，数据会受到各种各样的噪声污染，需要选择最合适的数据修复技术来对数据进行复原，同时使得复原后的数据符合真实分布。现有的工作大多都是假设噪声分布已知，用传统微分方程纹理合成或者深度模型来复原数据。然而在现实环境下，噪声的分布比较复杂，甚至没有一个闭式解，无法用一个显式的模型来表征。深度学习虽然可以模拟复杂的数据环境，但是目前的深度学习的研究对象是已知噪声分布的数据，不需要对该数据的噪声和扭曲进行处理，一种做法是忽略数据存在噪声，假设他们服从的分布是真实分布，另一种是用模板来标记数据缺失位置，并对缺失部分另外处理。

本文主要关注数据补全任务中的图像修复问题。 由于实际图像被污染的情况比较复杂，人眼不一定可以直接识别出受噪声污染位置及程度，且在数据繁多的情况下，无法一一采用人工方法检测出掩模的情况。所以现有的图像修复方法都是基于一种很理想的状态下进行修复，对不确定性掩模无法进行预测，从而确定应该需要修复的位置和程度，给大批量修复图像带来一定的难度。本文预测掩模的模型是attention模型，无法对图像像素点较多的图片进行建模，因此本文分辨率较高的图片，不采用该模型。

本文的主要工作，实现了对不确定缺失的图像修复模型，该模型能够学习原始图像的分布，并以一个生成式模型表示。我们引入生成对抗网络来作为基本框架，因为这是一个盲操作，实际生活中，图片被某些噪声污染，我们无法知道该图片对应的真实样本的情况，也就是该图片没有对应的标签。而修复后图像，可以与一些被预测出来的没有或者被影响的很小的数据子集进行零和博弈判别，从而使得修复后的图像与干净的数据子集的分布相同，从而骗过判别器。

本文主要提出了两个模型，一个是改进版AmbientGAN模型，保留了AmbientGAN缺失数据的假设约束条件，将网络改成补全网络，并在AmbientGAN补全网络的基础上，增加对网络的约束和对生成图像的约束，并将其应用在超分辨任务上。但是AmbientGAN的假设条件太强，大部分情况下，数据都不符合噪声分布的强约束条件，因此，我们又提出了一个新的模型，去除AmbientGAN的损失测量，利用数据集中的小部分无噪声数据，更加关注生成器生成的样本与输入图片的差异，用自编码器重构原输入图片的无缺失部分，在此基础上用全变分模型来约束补全部分与无噪声部分的光滑性，并对网络进行约束。由于最后的修复图像是补全内容和无噪声的拼接，为了保证修复图像能使得判别器也无法分辨是否属于哪个分布，我们将修复图像输入给判别器判断。

本文为了验证提出的模型的有效性，在matlab上实现做了传统补全算法实验，发现传统算法无法补全自然图像中复杂纹理，或者在未遮挡区域没有出现的元素。在引入生成模型后，我们对GAN结构进行分析，并且用实验解释了AmbientGAN的缺陷。并在三个数据集中实现了两个模型的补全任务，发现与有标签的模型想必比，模型的效果跟他们类似，模型达到了一定的补全效果。

## 本文的主要贡献

本文提出的基于生成式模型的图像补全技术，利用少量的无噪声数据和图片中未被遮挡的像素信息重建数据，并且生成补全内容，避免了传统深度补全算法训练时依赖于一对一的补全标签。本文的主要贡献如下：

* 本文在小数据集上应用了Attention模型来预测图片的缺失掩模，并发现了Attention的局限性
* 本文修改了AmbientGAN模型，将GAN的生成网络改成补全模型，并且应用到超分辨领域。本文结合AmbientGAN模型和自编码器，使得生成网络生成的未遮挡区域能够更清晰的重建原始像素，并用感知损失来约束整个网络，使得生成的结果更加清晰。
* 本文探讨了AmbientGAN的强假设约束条件，发现在实际应用中，AmbientGAN模型中，一旦损失度量有偏差，无法模拟真实的数据分布。因此我们去除了AmbientGAN框架下的损失度量，直接生成的样本来让判别器判断，将补全后的内容与未遮挡区的像素拼接起来，给判别器添加一个新的架构，并用借鉴传统方法中的全变分模型来约束生成网络补全的图片和新拼接的图片的光滑性。
* 本文还对大块遮挡易生成同一种内容提出了一个解决方法，使得生成的图片能够更加符合图片的语义。

## 本文的组织结构

本文结构组织如下：

第一章：介绍基于生成式模型补全问题的当前背景，回顾了各国内外学者的研究现状，并且介绍本文的主要工作和贡献。

第二章：对传统的图像补全技术进行介绍，阐述了什么是补全（Inpainting）技术，以及Inpainting传统的解决方法，并用实验证明了传统方法在解决纹理和大块缺失的效果不尽如人意。

第三章：介绍生成对抗网络的前世今生以及它的优缺点，和两个生成对抗网络在Inpainting任务上的应用，同时通过一个小实验证明了生成网络并不是简单的复制粘贴，而是富有‘创造力’。

第四章：利用AmbientGAN框架对数据进行补全，并结合了自编码器，使得生成的未遮挡区的图像尽可能相似，并用CoordConv来取代卷积以此来表征数据在空间之间的关系，用在视觉上实验证明比纯AmbientGAN好。还发现该模型可以应用在超分辨任务上。

第五章：抛弃AmbientGAN的假设，提出一个新的生成模型框架，并用实验证明了在没有补全内容标签的情况下，也可以补全各种缺失图片。

第六章：对本文工作进行总结，并对下一步工作进行展望。

# 图像补全技术介绍

图像具有简单直观，易于观察的特性，并且是国内外研究的重点，故本文的数数据表达方式均以图像形式呈现。本章简要介绍图像补全的定义，传统补全方法，以及传统补全方法对大面积缺失和物体关键性部分缺失的无效性实验证明。

## 2.1 图像补全相关技术介绍

## 2.1.1 Inpainting定义

图 2.二.1 Inpainting技术流程示意图。输入图I为计算机领域常用的Lena图，输入I被区域L遮挡后导致脸部关键信息丢失，利用Inpainting技术可以恢复出另一个Lena（该图是Lena 61岁重拍的）

修复技术（Inpainting）在现实生活中应用广泛。比如在博物馆里，专门指修复部分损坏文物一类；而在数字世界，我们面对的对象不是一个具体的事物，而是由一堆像素点构成的，也就是计算机视觉所研究的对象图像。对于图像中被遮挡或者有噪声的区域，用一些算法来修复或者替代被污染缺失的部分，从而恢复出原始图像的本来面貌。[13]概括的说，Inpainting就是一种修复图像或者视频的过程。

Inpainting起源于图像的恢复任务，目前已经有一些很成熟的有理论支持的传统算法。常用的工作有：（1）根据图片的全局信息来填充其中的某一个小空隙，这种方法的目的是用来保证整个图片的统一性。（2）用纹理细节结构来填充。（3）根据缺失空隙的不同区域的不同轮廓来匹配缺失内容。[17]（4）缺失间隙附近的轮廓一定相关，可将外围环境的轮廓能量延续到间隙补全。

现在我们从数学角度来看，Inpainting技术就是根据一张待修补的图片中的完整信息来补全缺失部分信息，从而得到原图的修复图。如图2.1所示，原图为的灰度图片，人物脸部被掩模L遮挡，因此丢掉了许多关键信息，而通过Inpainting算法可以将该图修复成一个正常的人物图。此外，从图2.1可以看出，外围信息并不能提供人物的脸部信息，所以脸部填充后不一定是原来的人物，这是因为Inpainting问题是一个病态（ill-condition）问题，存在无数的可行解，无法得到一个唯一的修复结果。

用矩阵来表示图片的缺失程度，**D**表示未被遮挡的区域，**L**表示被遮挡的区域对图片的每个像素点是否缺失用来表示，所以有

可得到退化（Degradation）后的图片为：

从而可得到修复问题的求解模型为：

其中表示修复模型的参数。

## 2.2 基于纹理合成的inpainting算法

基于纹理合成的Inpainting算法是由Efros[14]等人提出的，被证明是目前最有效的算法之一，其主要思想是从图像缺失块边缘由外向里递归地填充图像缺失块：对于图像缺失块边缘的每个像素点p，和在缺失块以外的像素点q，以及他们的邻接域和，可以被表示为一个优化问题

其中指的是关于块和块之间的误差平方和（SSD）：

其中i和j指的是一个领域（Patch）中的每个像素的横坐标和纵坐标。

一旦当前像素点p填充完毕，该算法将继续往p的领域内进行填充，且不会再重新填充p，这种基于依次填充像素点的方法，无法使用分布式方法异步运行，使得每张图片耗费的填充时间很长。具体纹理合成算法的计算机理可以参照图2.2，给定如图2.2左图的一张纹理图片，和右图正在合成一个像素的新图片。为了合成一个像素， 该算法首先寻找所有类似当前像素点的领域的patch，接着随机选择一个类似领域，并将它的中心像素作为这个新的合成像素。

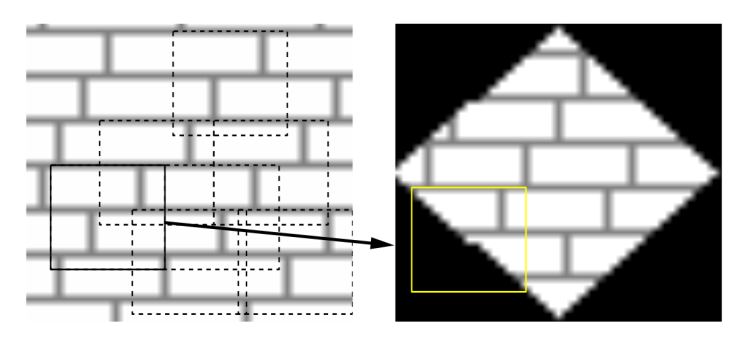
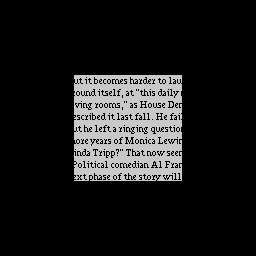
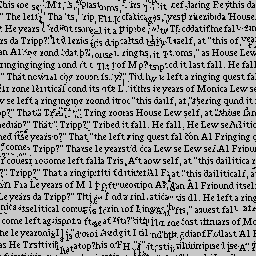


图 2.二.2 Efros的算法概观（图片取自[14]）

为了验证算法的有效性，在Matlab上实现了该算法，两张需要向外补全的图片的使用实验效果如图2.3所示。 我们发现，这个算法在有重复纹理的图片上很有效，如图2.3第一张图所示（来自原作者原文[14]），字母纹理比较重复简单，可以按纹理匹配，但是一旦图像比较复杂，比如人脸中间部分被挡住，基本上很难补全。



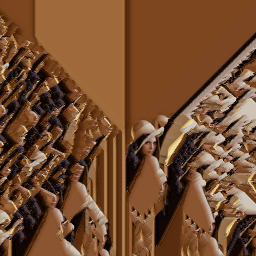


图2.二.3 基于纹理合成inpainting算法实验效果

如图2.3第二个例子Lena图，可以看出该算法很好的找到图片间隙的像素点的领域特征别匹配到图片中存在的patch，生成的图片的纹理基本都是重复的，图片没有增加新的内容。因此这种算法通常对去除目标遮挡，并且遮挡的内容与其他patch较为相似的情况下非常有效，但是很难补全一个复杂的物体或者当关键部分被遮挡的物体。

## 2.3 偏微分方程求解模型（PDE）

图像修复问题还可以从中值滤波、均值滤波、参数估计等方法求解，但是这些方法往往忽略**D**和**L**区之间的边缘信息，从而导致边缘变得模糊，使得修补后的D区的有用信息很少。但是基于能量的变分法可以尽可能的保留图片的几何特征，并且可以解决细节信息丢失的问题。Bertalmio M[13]等人提出偏微分方程求解模型（PDE），利用被遮挡区D周围的边缘信息，并利用热扩散方程将这些周围轮廓纹理等信息传递到被遮挡区D。

用变分的观点来看，图像补全问题可以转变一个约束的最优化问题：

其中表示图像先验模型的能量；表示高斯白噪声的方差；，可以用最小均方误差来表示该能量。这个最优化问题可以用拉格朗日方法求解。

接下来，主要研究基于变分的修补模型技术。

### 2.3.1 全变分模型（Total Variation）

全变分模型（TV）第一次是Rudin等人提出的[15]。TV去噪在恢复分段常数信号上很有效，常用于图像过滤和图像恢复。对于一个由N个点构成的信号有

Rudin巧妙的将一张图片的像素点看作一个分段平滑函数，并在有界变分空间上建模[15]。并且该模型能够对图像边界进行延长，所以常常用于图像的修补任务。Shen等人将TV模型应用在图像修复任务上，建立的模型为：

其中为拉格朗日乘子。用欧拉-拉格朗日法可以求解该方程。

TV模型能保持渐变区域的光滑度，同时能够锐化边缘，不需要任何后处理，但是同时破坏了物体的连通性。

### 2.3.2Mumford-Shah补全模型

Chan[14]和Tsai[18]提出了主动轮廓模型，也是通过分段平滑函数寻找原始图像的最优逼近，主要应用于图像分割和图像补全，而且该方法利用了图像梯度信息，在轮廓内求平均梯度，故对噪声非常敏感，无法用于灰度不均衡的图像上。最开始Mumford和Shah提出的模型的目标函数如下：（该方法不是研究的重点，使用的符号不是本文定义，是作者原文定义的，感兴趣的读者可以查看[14]）

该模型的优点是计算复杂度低，但是它只能修复特有的缺失[19]，Mumford-Shah修补的是边缘最小，修补后的边缘直接连接已有的边缘，导致边缘不平滑，同时它也存在有TV一样的问题。

## 2.4修复技术的度量标准

在图像修补领域，有两个重要的评价修补效果的指标，结构相似性（SSIM）和峰值信噪声比（PSNR）。

### 2.4.1 结构相似性（SSIM）

SSIM是一种早期用来衡量电子图像和视频的感知质量的重要指标，目前被广泛的应用在计算机视觉领域，被用来计算两张图片的相似性或保真度。SSIM目标是研究视觉系统中每个与感知相关的部分的结构的失真程度，而不单单是感知误差。SSIM可以衡量不同规格图像之间的相似度，现有两张不同规格图片和，SSIM距离为：

其中表示的均值，表示的均值，表示的方差，表示的方差，表示和的协方差，和用来防止SSIM出现除零溢出。SSIM的值在0到1之间，当两张图完全相等时，SSIM的值为1。

### 2.4.2 峰值信噪比（PSNR）

PSNR是表示信号最大功率与破坏保真度的噪声功率之间的比值。该指标只考虑对应像素之间的差异，只是基于误差敏感，故需要先计算出两张图的误差：

得到均方误差后，可以很容易求出PSNR指标：

人对频率较低的图像对比度和亮度比较敏感，而对色度的感觉比较弱，但是PSNR却不考虑人眼的感知能力，因此PSNR可能会与直观看到不一样的结果。所以一般情况下，要将PSNR和SSIM同时纳入补全图像的质量指标

## 2.5 传统方法实验结果



图 2.二.4 Betalmios的实验效果[]

关于传统的补全算法，对于小部分缺失，补全区域与周围类似的图片可以很完美的补全，如图2.4所示，左图表示原图，中间的图像将人物身上的绳子遮挡住，右图表示PDE方法处理后的结果。可以看出PDE方法可以去除小面积区域的遮挡，很轻松的将绳子隐藏，用周围的纹理覆盖被遮挡的部分，使得看起来这个人物在天空飞翔。

ssim =0.989 ssim =0.959 ssim=0.991 ssim=0.899

psnr=26.273 psnr=20.265 psnr=28.724 psnr =28.720



Input PDEs补全 Gibbs补全 mumford\_shah 补全 TV补全

图 2.二.5 Lena的传统方式补全结果

然而，如图2.5所示，输入图的人脸被挡住，传统方法根本无法还原出原始图片的内容或者与人脸相关的部分，这是因为单纯基于纹理和能量优化的方法，无法拟合原始图像服从的真实分布，只能从图片剩余的信息来获取内容，所以针对关键部分损失的图像，应该引入生成模型来拟合数据的分布，使得不在未被遮挡区的物体也能被还原出来。

## 2.6 本章小结

Inapinting技术可以去除遮挡，修复图像，发现图片中隐藏的信息，在现实世界中应用广泛，该技术概括的说，就是恢复图像或者视频的一个过程。Inpainting问题是一个欠定问题，存在无数中求解方案，所以不一定恢复的和原图一样。本章主要简要介绍两种类型的非深度方法的传统求解模型，基于纹理和基于PDE的图像补全算法，这两种对于缺失部分有类似的部分，或者去除小部分物体遮挡效果很好，但是对于关键部分的遮挡或者缺失的恢复效果很差。为了验证补全效果，本章不仅从视觉上来衡量，还从SSIM和PSNR这两个指标来评价图片的补全效果。

# 生成模型

本章主要对生成模型涉及到的背景知识进行一个系统的概括，主要内容有：（一）生成模型定义；（二）自编码器以及改进；（三）生成对抗网络算法；（四）度量函数以及一些生成模型在缺失度量下的应用。

## 3.1 生成模型定义

与判别式模型建立一个原始空间到目标空间的映射不同，生成模型则是直接在原始空间上建模，利用现有的观测数据来拟合样本空间的近似分布。在生成模型给定隐变量的前提下，对观测空间和标注建立一个联合分布，如高斯混合模型、隐马尔可夫模型和贝叶斯模型等。

原始空间上有一点和对应目标空间标注y，生成模型给出的是联合概率分布，由联合分布可以计算出条件概率，从而可以推导出原始数据空间的分布。假设原数据集里有n个样本，则训练集记为，将数据原始空间的分布设为，因此训练集可以表示为，生成模型的目的是学出一个近似表示分布，从而可以从近似分布中中采样出新的样本：

生成模型通常分为显式密度估计和隐式密度估计两种，其中PixelRNN、变分自编码器和玻尔兹曼机等是显示密度估计，这些估计都是基于似然求解的，但是生成的图片质量较为模糊。而隐式密度估计的一个典型代表是生成对抗网络（GAN），能产生多样和高质量的样本，被业界公认为是the-state-of-art方法，缺点是很难训练，并且不稳定，并且不能解决原样本空间的推断问题。

由于生成模型可以为超分辨、着色等任务产生大量样本，在一定程度上解决部分数据类型样本不足的问题；在金融行业上，关于时间序列数据的生成式模型还可以用来预测和模拟数据，带来巨大的经济效应；在研究上，训练生成模型还能用于泛化特征的隐层来表示推断，给大量的学者提供了研究材料。

因为传统的生成模型生成的样本质量较差，且约束较多，本文不对此做过多的讨论，神经网络理论上可以拟合大部分复杂的函数，网络的表达能力也很强，因此本文只针对深度生成模型进行讨论。

## 3.2 变分自动编码器

### 3.2.1 自动编码器

自动编码器（autoencoder）是生成模型中的一个重要模型。它是一种学习高维数据的低维表示的无监督机器学习方法。如图3.1所示，编码器先把每个样本映射到低维空间，得到特征向量，该特征向量被解码器解码还原成样本，同时自编码器的输入数据和重构数据应该尽可能相似。从而可以通过限定的维数来将数据映射到低维空间。自编码器通过最小化目标函数来训练解码器和编码器。模型一旦训练完成，它可以单独提取解码器，在解码器输入一个码后，可以生成对应的图片。通过这个方式，可以轻易采样出大量的图片。该模型优点是在无监督样本的条件下，将一个高维数据直接压缩成低维特征，并能产生大量的样本。它的缺点是模型对的分布是未知状态，因此模型无法给低维表示限定范围并产生新的样本。为了解决这个模型的生成样本分布问题，VAE在自编码器的基础上引入概率模型，从而使该问题变得可解。

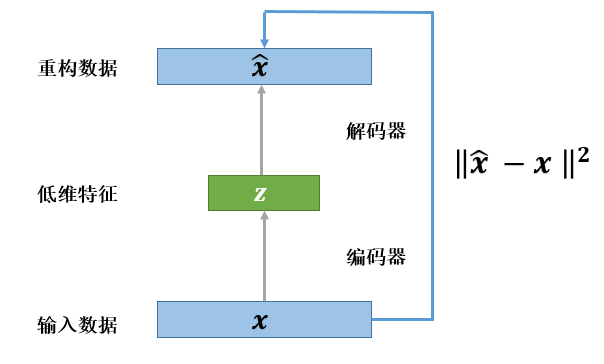


图 3.三.1 自编码器架构

### 3.2.2 变分自动编码器（VAE）

假设z的先验分布是，且训练数据是从隐变量中生成得到的，因此如何估计出参数成了解决这个问题的关键。为了简化运算，通常将高斯分布选作特征的先验分布。如图3.2所示，从先验分布中采样出，然后通过生成模型生成样本，然而生成样本的分布是很复杂的图片分布，所以生成模型的结构一定不是简单的模型。由于神经网络结构理论上可以拟合任何复杂的函数，所以生成网络用神经网络来搭建。

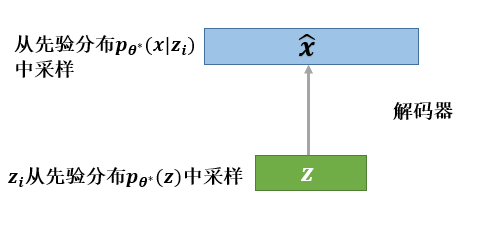


图 3..三.2 VAE的采样过程

由于该模型中的后验很难求解，因此避免直接求解后验，用近似推断近似分布，经过变分推断后，原先的最大化似然转变为最小化：

其中表示Kullback-Leibler Divergence（KL）距离。从目标函数可以看出，优化目标转化为重建输入数据和缩小后验分布和的距离。

## 3.3 生成对抗网络

由于VAE的优化误差是基于点对点的差值，然而这样的度量标准没有考虑到人眼的感受能力。以生成一个手写体数字7为例，假设一个误差点在图像的边角，一个在7这个数字附近，对VAE的误差来说，这两者的结果相等。但从人的直观感觉来说，显然后者看起来更加真实。而生成对抗网络（GAN）优化的是数据整体信息熵，GAN可以很容易分辨上文的两种情况，同时会使生成的图片更加清晰丰富。

GAN起源于博弈论中纳什均衡[20]理论，由互相竞争并不妥协并期望自己利益最大化的两方构成。即使最后达到纳什均衡的优化策略对整体来说不是最优，但自己这方来说一定不是最坏的结果。GAN中的两方分别为生成器G网络和判别器D网络。其中生成器G的作用是将一个随机输入通过生成网络来生成原始空间的数据样本，从而能让D网络认为该样本是属于原始数据空间分布，而不是生成器造出来的‘假数据’。判别网络D目的是尽可能地判别数据属于生成分布还是真实分布。双方为了达到目的，不断进行零和博弈，直到找到两者之间的非合作平衡才停止。

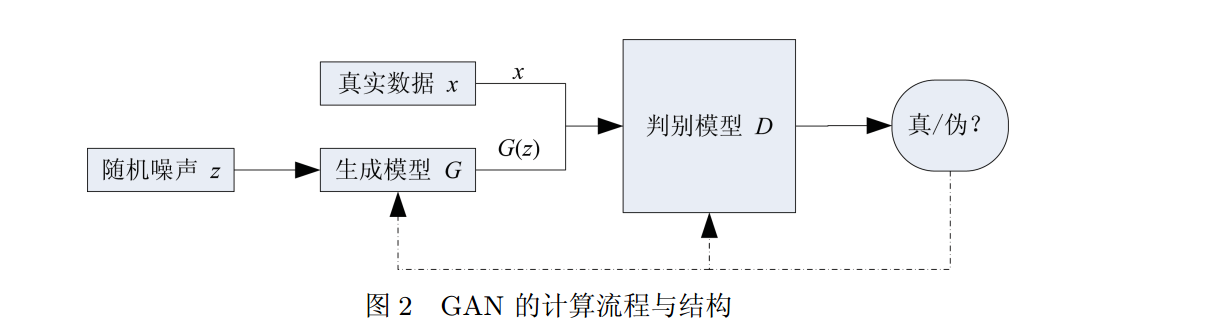


图 3.三.3 GAN的计算流程与结构

GAN的基本框架如图3.3所示，任何可导函数都可以作为G网络和D网络的组成的。假定真实数据，随机变量z是高斯噪声，G(z)表示G网络生成的样本，服从生成分布，接着将x和G(z)交给D网络来判断真伪。GAN的级小极大目标函数为：

从目标函数可知，当D网络达到最优时：

目标函数就转化为G网络的优化，它的目的是最小化真实分布和生成分布之间的JS距离：

GAN通过G和D不断的交替优化，最终达到两个网络之间的非合作平衡。

### 3.3.1 GAN的优缺点

GAN不止生成的图片更加锐利清晰，还有以下几个优点：不需要直接计算似然，也无需要用马尔可夫链采样，更不需要在过程中进行近似推断；任何网络架构的生成器和判别器都可以进行训练，不需要设计任何有关因式分解的模型；如果判别器能够完美适合，生成器就可以完美的拟合数据分布

然而GAN目前存在很多棘手的问题，由于G网络在刚开始训练的时候，生成的质量很差，而D网络的能力很强，容易导致目标函数无法优化，以致于G网络拟合数据的真实分布。此外GAN在实际操作下，很难收敛；产生数据坍塌问题，可能不断生成相同的数据；不需要提前建模，所以整个模型很难控制。

|  |
| --- |
| 算法1 生成对抗网络算法 |
| 输入：数据集，设置超参数判别网络迭代次数  输出：生成新的数据样本   1. While 达到一定的迭代次数 do: 2. for i in range(k): 3. 从先验噪声分布P(z)中随机采样噪声输入 4. 从数据集中随机抽取m个样本 5. 用梯度下降法更新D网络： 7. End for 8. 从先验噪声分布P(z)中随机采样噪声输入 9. 用梯度下降法更新G网络： 11. End while 12. 从先验噪声分布P(z)中随机采样噪声输入 13. 得到生成样本 |

### 3.3.2 GAN的改进

Arjovsky等人从拓扑与测度的角度做出了详细的理论证明，并指出GAN存在的这些问题的根本原因在于原始GAN中不合理的距离度量[4]；他们进一步提出了基于Wasserstein距离的WGAN模型，通过优化1-Wasserstein距离的Kantorovich-Rubinstein对偶形式来训练WGAN[5]。借助于这种思想，WAE在VAE变分下界的优化问题中引入了Wasserstein距离[6]，并理论证明了当自编码器在隐层概率分布收敛时，优化重构误差等价于优化原始形式的Wasserstein距离。

不断有学者对GAN的理论和实践进行改进[]，以下列举出几个重要的改进网络

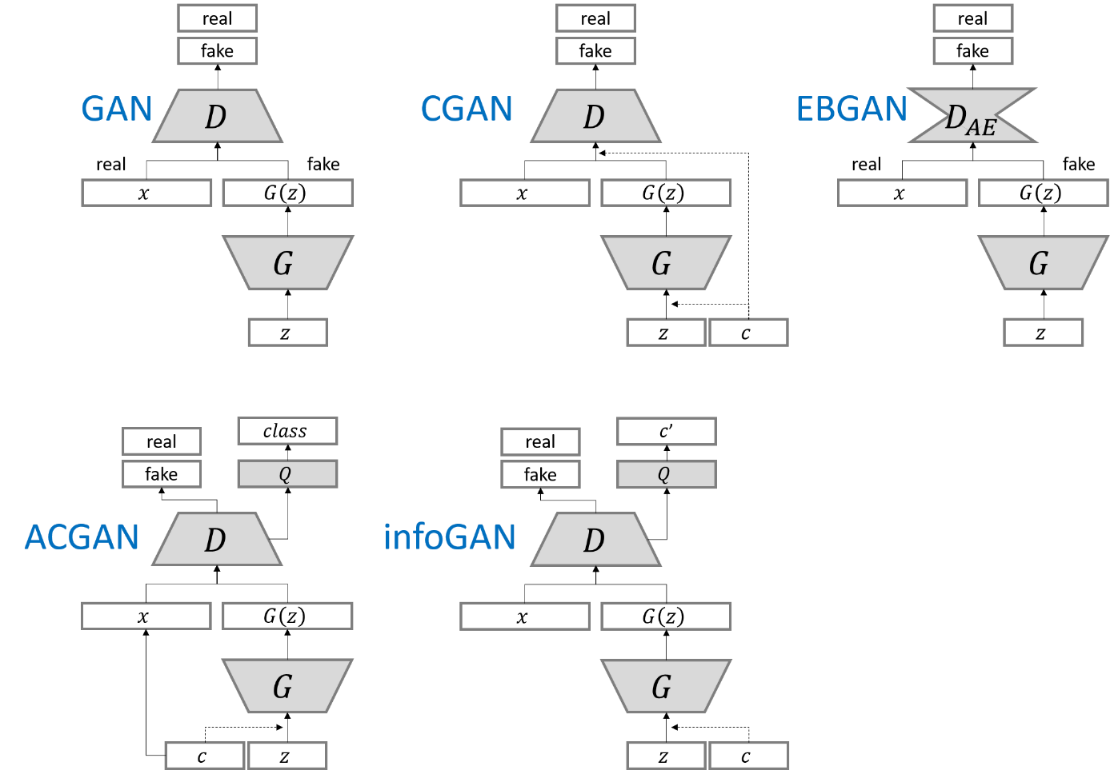


图 3.三.4 各种改进GAN的框图

## 3.4 GAN的度量函数

### 3.4.1 散度

由于真实数据的分布是未知的，GAN的拟合分布是已知的，普通的距离是很难衡量这两个分布。F散度可以很好的解决这个问题。现假设在同一个空间中的两组分布的连续概率函数（PDF）是和：有：

当f函数必须满足以下两个条件：

1. f是凸函数
2. f(1) = 0。

特别地，当时，为KL散度；当时，为逆KL散度；当时，为TV距离。

由于f散度易于计算，所以用来衡量真实分布与生成分布的距离，劣势在于f散度是不对称的，即。

### 3.4.2 Inception Score

在许多关于GAN的论文中，通常采用的是Improve GAN[]提出的评价模型的重要指标Inception Score（IS）来度量GAN的生成效果。IS使用两个度量标准来衡量GAN的表现：生成图像的质量和它们的多样性。GAN可以用熵来衡量无序性的程度，如果一个随机变量具有很强的预测能力意味着它的熵偏小。

在GAN中，我们希望条件概率可预测，也就是说它具有很小的熵。例如，给定一张图像，从人眼看来，我们能直观的判断物体的类型。对于网络来说，需要借助一些工具来辅助判断，故我们使用一个能反映图片生成质量的Inception 网络，用来分类生成图片和预测和类标签。

第二步，我们需要度量图像的多样。令p(y)是一个边缘分布，可以从如下式子导出：

若y的数据分布具有很高的熵，那么生成的样本具有多样性。为了同时度量GAN的预测能力和生成图片的多样性，可计算它们的KL散度：

然而IS存在两个主要缺点，其一是如果每个类只产生一张的图片，这种情况下即使它的熵很低，IS的分值仍然很好。其二是IS并不能反映过拟合。

## 3.4 Context Encoder

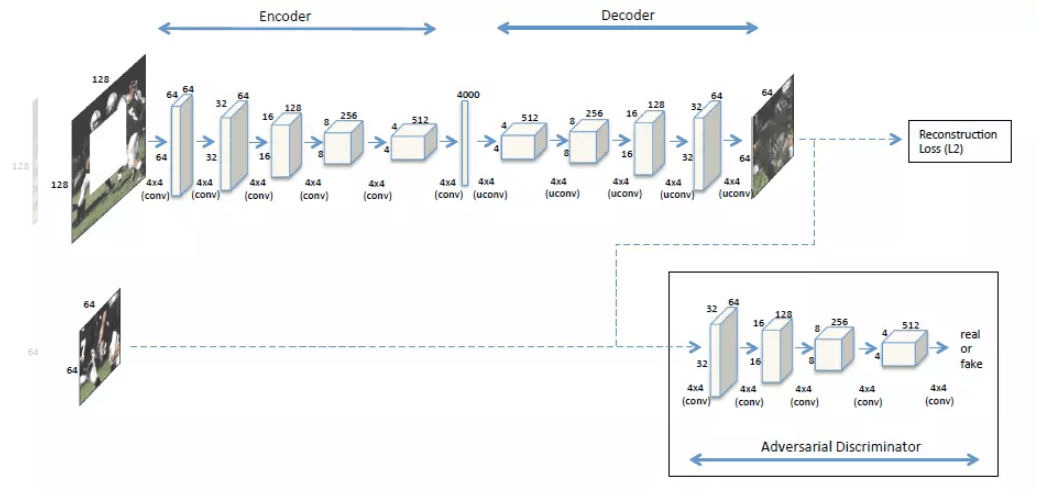
Context Encoder将生成模型用于Inpainting任务，它利用了缺失部分的周围的信息来推测被遮挡区域L的信息。主要的思路是结合自编码器和GAN架构，自编码器学习图像特征并生成图像被遮挡区域的预测结果，而GAN是用来判断预测图的分布来源于真实样本分布的可能性。当生成预测图与真实样本图的内容基本一致时，且GAN的判别器无法区分预测图来自哪个分布的时候，就认为网络达到一种平衡状态。

图 3.三.5 Context Encoder 网络架构

具体的Context Encoder网络架构如图3.5所示

网络训练过程所使用的损失函数由两部分组成：自编码部分的约束为差异重建损失（Reconstruction Loss）和GAN的约束为对抗损失。并且在此基础上，增加一个真实标签与图片的L2损失，使得生成的画面更加清晰。从上面的框架可以看出，该框架对每张图片的真实标签依赖性很强，大部分训练损失都是靠真实标签的重构误差得到。

## 3.5 AmbientGAN

能观测到的数据通常存在扭曲和噪声，换句话说，这些数据是在有损测量下观测到的。为了学习到数据的原始分布，AmbientGAN框架的提出解决了在有损测量下的无损数据分布的学习问题。该方法在数据存在噪声时，通过定义有损测量，给如何在充满扭曲和噪声的数据中学习到原始数据分布问题，提供了前瞻性的思路。

与传统GAN算法的直接辨别数据的来源不同的是，该框架避免了直接判别无噪声的图像，而是让D网络判别有损测量下的生成图像和真实数据。如图3.6所示，随机输入一个噪声，接着通过GAN的G网络生成一张图片，假设缺失分布服从测量，在该测量下可以得到有损图片。对于观测样本，该框架假设这些观测样本的有损度量也是，最后判别器判断每个样本属于哪个分布。让D网络和G网络不断优化对抗，直到收敛，网络训练完成

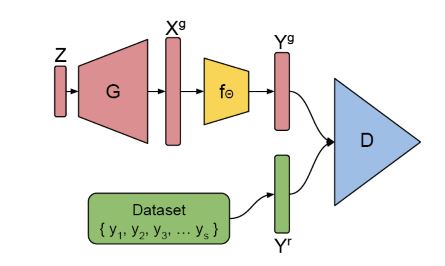
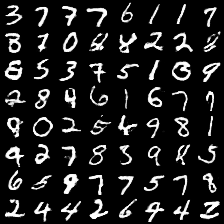
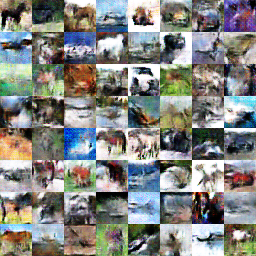
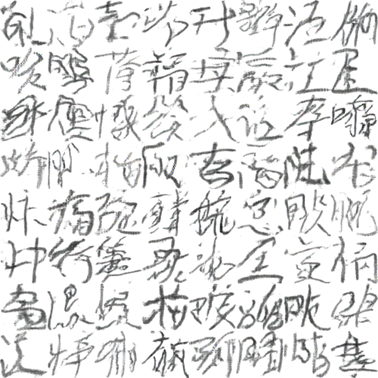


图 3.三.6 AmbientGAN结构图

该框架假设观测样本的损失测量已知，因此目标函数为：

因为Inpainting任务是将缺失图像补全为一个完整的图像，同样图像也是存在缺失和扭曲，所以我们下一章将AmbientGAN框架应用到补全任务中。

为了验证该算法的有效性，我们在分别在CIFAR10，MNIST，中文手写数据集上训练该网络，数据集的具体介绍见后面的章节。



(a)MNIST生成结果 (b)CIFAR10生成结果 (c)手写汉字生成结果

图 3.三.7 AmbientGAN实验结果

三个实验的参数都是用损失测量的方式为黑色block遮挡，遮挡率为百分之25。如图3.7所示，即使GAN的判别器D看到的都是被遮挡后的数据，只要给生成图像加一个遮挡测量，也可以很好的复现出原效果。对于（a）图，可以看到生成的数字非常清晰，基本上符合手写体的形状；对于（b）图也可以看出，AmbientGAN生成的图像基本与原数据集一致，能看出图片的大致形状，因为实际上缺失率较高，原始图片的分辨率也不高，所以恢复成这样效果还是不错的；对于（c）图，该数据集是由不用的人写出来，数据集里很多字体，中国汉字有3千多个，所以AmbientGAN在数据集缺失的条件下，还可以学习出中国人手写的汉字的特点以及笔墨分布，即使我们认为这不是汉字。从实验（c）也从侧面的证明了GAN生成的图片不是复制粘贴，还是在学到的分布情况下，进行自由创作，生成不同的所谓的‘字’，恰恰体现了它的多样性。

## 3.6 本章小结

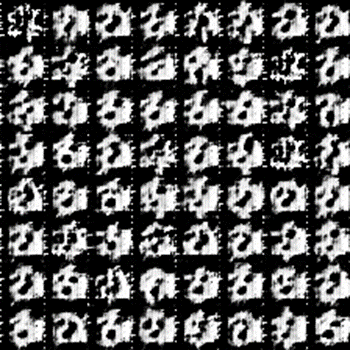
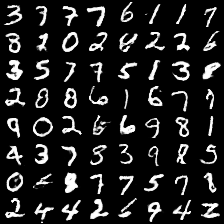
生成模型从观测数据中学习到一个近似数据分布的模型，不仅可以学习到观测数据的分布，而且，可以增加数据的多样性，能够进行高维概率分布的有效检验。传统的生成模型是基于似然推断，由于似然推断只能拟合简单的分布，对于图像这种高维数据，生成图片效果模糊且无法辨认，所以本文不对传统的生成模型进行探讨。而神经网络原则上可以拟合任何复杂的数据分布，利用神经网络搭建生成器，可以取得不错的效果，故本文先讨论了自编码器和生成对抗网络。接着我们讨论了生成对抗网络的度量，表明生成网络的效果可以用IS衡量。最后，我们着重关注生成对抗网络在有损测量下的应用模型。

# 基于GAN的缺失数据补全算法

## 4.1 问题的提出

对结构性强、纹理复杂的图片来说，如果使用第二章提出的传统Inpainting非深度算法，当没有添加先验信息时，图片补全内容通常是模糊不清的。图片被遮挡的纹理一旦没有在非遮挡区出现时，补全效果就差强人意。例如人脸关键部位一旦被挡住，这时不存在人脸的特征先验，因此算法无法捕捉原始数据的分布，导致将无法补全人脸。而如第三章所示，生成模型可以拟合数据的原始分布，并生成清晰丰富的图像。已有很多学者将GAN应用在补全任务上，我们借鉴这种框架，将GAN用于补全缺失数据的任务中。因此我们提出的修复模型是基于生成模型的。

对于基于生成模型的深度学习网络，特征的提取方式往往都是通过卷积神经网络（CNN）来实现的。这种结构的缺点是，一旦某些空间区域出现缺失，CNN会将这些缺失部分一起计算进去，导致生成结果受缺失部分像素影响。所以CNN无法有效的处理空间位置上的关系。



(a) (b) (c)

图 4.四.1 AmbientGAN在不同损失度量下的效果

AmbientGAN为我们提出了一种在缺失测量下的数据无损分布的学习算法。为了衡量AmbientGAN的约束影响，我们在MNIST数据集上做了一个小实验。损失测量选择dropout方式（dropout=表示每个像素以百分之的概率被丢弃，keep\_prob=）。对于D的两个输入，在实际操作上，我们分别在真实数据（加了损失测量后为观测数据）和生成数据中设置keep\_prob相同的参数和不相同的参数。

如图4.1所示，第一组是实验的keep\_prob的真实分布和生成分布都为0.7，得到的结果如图4.1(a)所示：网络能生成正常的数字手写体；第二组实验的真实数据的损失测量的keep\_prob为0.6，真实分布的损失测量的keep\_prob设置0.4，结果如图4.1(b)所示，G网络生成了‘鬼脸’。从生成的像素点个数方面进行分析，对判别器D来说，在像素层面上分辨不出来的前提，就是两个数据源的总像素一定要相等，而如第二组实验所示，生成器丢弃的像素点远比真实手写体丢弃的多。因此为了弥补丢弃后的像素差异，生成器在生成的图片的时候会生成更多的像素，使得两个数据源丢弃后的效果尽可能相似。这样导致的结果是G网络变成‘鬼脸生成器’。第三组实验的真实数据的损失测量的keep\_prob为0.4，真实分布的损失测量的keep\_prob设置0.6，如图4.1(c)所示，G生成的质量不好，原因和第二组实验类似，因此不累述。

‘鬼脸’实验表明AmbientGAN的一个核心约束条件是要已知损失测量，而且该损失测量不能与实际数据的偏差程度有任何偏差，否则无法恢复出原始数据分布。不止如此，这个度量还需要满足一个条件，那就是要求度量函数带有随机性。举个反例，假设损失测量总是对人脸的眼睛遮挡，那么对于判别器来说，无论遮挡部分生成的是什么内容，始终看不到遮挡信息，也就无法分辨出人眼睛的信息。对于更详细的证明，具体可以AmbientGAN参照原文。对于判别器D来说，它需要比较丰富多样的数据输入，因此实际上AmbientGAN需要的训练数据信息是完整的。

由于AmbientGAN的随机损失测量生成数据机制给我们启发，我们在此基础上，提出以下改进：

1. 为了表征像素在空间上的关系，我们受CoordConv[]的启发，给部分卷积层加入像素在空间上的坐标。
2. 利用ambientGAN的框架，用UNET架构将G网络改为补全网络。
3. AmbientGAN在GAN的基础上，只是给生成数据和原始数据加一个测量度量，生成效果一般。对于一张被遮挡的图片，有一些区域的信息是完整，故可以提取完整部分作为监督标签，使得补全的内容细节更加丰富。
4. 我们采用一种特殊的技巧，将该框架应用到补全任务上。

## 4.2 基于GAN的修复模型框架

为了让生成网络尽可能拟合真实数据分布，生成对抗网络通过引入对抗训练机制，当数据质量很高时，生成对抗网络可以达到G网络和D网络的纳什平衡点。受限于各种原因，数据不是完全干净无污染的，很多补全任务中，并没有一对一的补全标签数据。幸运的是，而生成对抗网络判别器在辨别数据属于哪个分布时，也不要求所有的数据有一对一的标签，但是所有的数据必须属于统一分布。因此对D网络来说，它只需判断输入属于哪个分布。实际上，GAN只需要用两个分布的采样，可以避免直接使用缺失数据-真实标签对。

但是目前所有的算法都是基于后者，先让GAN判断补全的图片属于哪个分布，然后优化补全后的缺失图片与真实区域（Ground Truth，简称GT）之间的F范数，否则无法补全图像。

如第二章所示一个需要补全的图像，存在遮挡区和非遮挡区，AmbientGAN原始补全网络没有利用非遮挡区的标签信息，所以我们将关注于非遮挡区的训练情况。另外，遮挡区域的划分和遮挡程度都是通过一个指定掩模模板来实现的，有时候人眼并非可以直接分辨出哪些是需要补全区域，且分辨这些区域有时候很费人工。由于[]提出自注意力机制，可以计算出每个元素对结果的注意力权重，所以我们可以轻易将它应用在我们的任务。该方法可以很轻易地可以训练出一个Mask特征层，用来识别输入图片的缺失位置和程度。

与其他研究借鉴图片原有的GT不同，我们只关注于非遮挡区的GT像素。并且利用这些未缺失区域来填补缺失内容。于此同时，利用AmbientGAN网络来判断，重新加损失度量的生成图片是否与观测图片属于同一分布。

### 4.2.1 CoordConv Layer

深度学习常用的提取特征的方式通常是CNN，而大部分的研究都关注怎么架构网络，很少有人注意到提取空间特征的架构的设计。实际上CNN存在一个明显的缺陷，无法将空间表示转换成笛卡尔空间中的坐标和one-hot像素空间的坐标[]。为了解决这个问题，[]提出了CoordConv来处理坐标间的关系，具体做法如图4.2所示，加入两个坐标的位置定位过滤器的空间关系。两个左边是是x轴的坐标和y轴的坐标，加入这两个通道（Channel）后，CoordConv就有了明确的空间位置信息。原作者在原文中，用实验证明了该方法的可靠有效，具体见文献[]。由于Inpainting任务中有个掩模表示了缺失的位置和缺失的程度，对空间关系约束较高，故CoordConv很适合用于提取被遮挡数据的特征。

为了减小大块缺失的区域的丢失像素对结果的影响，我们把部分卷积层的卷积过滤器替换成CoordConv层。

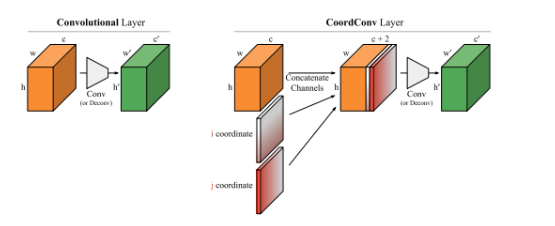


图 4.四.2 坐标卷积的架构

### 4.2.2 Mask网络架构

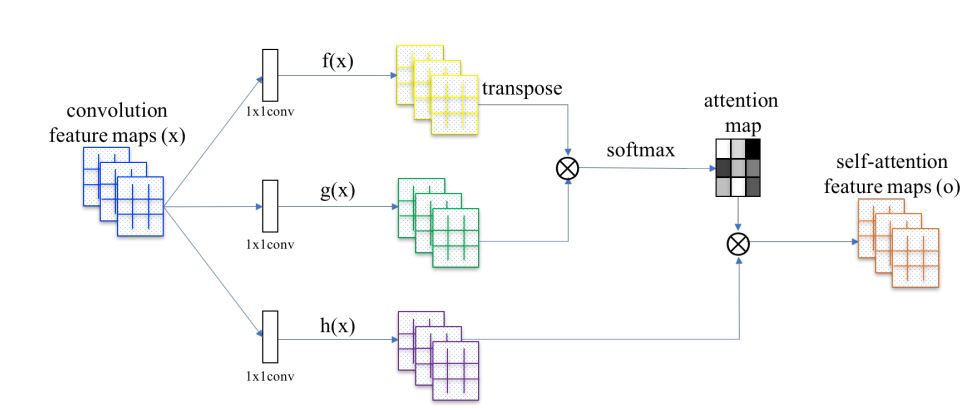
输入的图片通常包含信息完整（未遮挡）和信息丢失部分，对于缺失图片，有些缺失人眼可以很容易察觉缺失的程度。很多种缺失与图片混在一起，人眼可能无法分辨。

图 4.四.3 自注意力机制

自注意力机制目前在计算机区域识别中应用广泛，它可以计算机出每个像素点的重要性程度，使得模型可以按程度对目标进行优化。故我们引入自注意力机制，用来分辨图像缺失的程度，从而判断图片缺失不和谐的区域。自注意力机制主要靠自注意力网络计算出的特征map的权重。如图4.3所示，现假设输入为，把x拉长为一个长向量，然后输入转化到，空间：

其中, ，从而可以计算出第个位置与第个位置之间的相关关系：

从而可以得到最后的Mask：

但是这样的提取Mask方式存在一个问题，当你计算的时候，这个维度，我们知道WH是每个图像的所有像素点总数。被电脑配置的限制，因此这个我们只在MNIST这样大小的数据集上实验。虽然这个方法可以很好的找到缺失的位置，但是对于稍大的图像来说，它非常耗费计算资源。而实际上对于dropout和块遮挡的缺失方式，我们可以预先训练一个神经网络，用这个网络来代替Attention模块。由于我们模型的重点不在预测，且主要的实现方式是为Dropout和块遮挡这类缺失方式，故后续的实验直接采用原有的Mask。我们在这里提供一个想法，未来将着重研究复杂噪声的分布位置和程度。

### 4.2.3网络建模

我们设计的模型框架如图4.4所示，用一个Mask网络来预测图片的缺失特征掩模，实际操作上也可以将原来的Mask直接接入。然后将需要补全的图片和对应的Mask一起输入到G网络中，G网络根据这个图片信息和缺失信息，生成重构出原图片中信息完整区域，并填满缺失的内容。在给生成图片增加损失测量，从而图片重新被遮挡，用D网络判断补全后的图片在给定损失测量下与原始观测图片的分布差异。当模型到达Nash均衡的时候，即G网络的生成能力很强，D网络的判别能力也很强，但D网络判别不出输入的图片是属于加了损失度量的生成分布，还是观测数据的分布时，模型到达一个局部最优。

但是这样生成的图片还是不太清晰，为了利用更多的有效信息，我们给这个框架上加了一个自编码器。该自编码器巧妙的利用了待补全图像的有用信息，通过重构这些区域的像素点来使得生成的图像更加自然。



图 4.四.4 改进版AmbientGAN补全网络

现用表示输入图像，表示补全图像，表示没有被任何物体遮风的图像，则被遮挡的图像，即观测图像为：

则D网络的损失函数为：

训练过程中，当固定生成网络G的参数时，我们希望D网络的判别能力越来越强。当固定判别网络D的参数时，我们希望G网络补全的能力变强。

因此G网络的目标函数为：

U-Net类似于自编码器，左边一侧是编码器，右边一侧是解码器，输入和输出的图片大小应相同。浅层网络容易学到空间结构信息，深层网络容易学到更抽象的语义信息，因此UNET中引入跳层连接，将两种信息有机结合，从而获得较好的效果。而补全任务的输入输出的图片大小相等，故我们采用UNET框架来作为生成网络。

如表4.1所示，生成网络G共有13层卷积层，2层池化层，2层反卷积层，通道数含有加运算符表示跳层连接。由于坐标卷积可以增强空间表达能力，但坐标卷积增加计算量，故我们只在其中四层浅层网络增加坐标信息，以此达到增强空间表达力的目的。对于U-NET激活函数，除了最后一层激活函数是sigmoid函数，其他层都是Relu函数。为了能使U-NET网络能够稳定训练，我们在每一个卷积层的激活函数后增加一个BN层 。而D网络，我们采用的是类似VGG的结构，因为VGG结构迁移性很强，并且比较容易实现。最重要的是，VGG网络可以通过不断加深网络结构可以提升性能。

表4.1 G网络中的U-NET架构

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 模块名 | 过滤器 | 通道数 | 步进数 | BN层 | 激活函数 | 坐标卷积 |
| Conv1 |  | 64 | 1 | 有 | Relu | 是 |
| Conv2 |  | 64 | 1 | 有 | Relu | 是 |
| Pool1 | - | - | 2 | - | - | - |
| Conv3 |  | 128 | 1 | 有 | Relu | 是 |
| Conv4 |  | 128 | 1 | 有 | Relu | 不是 |
| Conv5 |  | 128 | 1 | 有 | Relu | 不是 |
| Pool2 | - | - | 2 | - | - | - |
| Conv6 |  | 256 | 1 | 有 | Relu | 是 |
| Conv7 |  | 256 | 1 | 有 | Relu | 不是 |
| Conv8 |  | 256 | 1 | 有 | Relu | 不是 |
| Deconv1 |  | 128 | 2 | - | Relu | - |
| Concat1 |  | 128+128 | - | - | - | - |
| Conv9 |  | 128 | 1 | 有 | Relu | 不是 |
| Conv10 |  | 128 | 1 | 有 | Relu | 不是 |
| Deconv2 |  | 64 | 2 | - | Relu | - |
| Concat2 |  | 64+64 | - | - | - | - |
| Conv11 |  | 64 | 1 | 有 | Relu | 不是 |
| Conv12 |  | 64 | 1 | 有 | Relu | 不是 |
| Conv13 |  | 3 | 1 | 有 | Sigmoid | 不是 |

### 4.2.4 损失函数

如第二章所示，一张待补全的图片是由遮挡区域***D***区和未被遮挡区域***L***区两部分组成的。而我们用到的AmbientGAN框架，没有对两个区域进行区分，只是将重点放在生成图像上，并对生成图片加损失测量。如果一张图像同时被不同位置遮挡住，我们称为遮挡1和遮挡2 ，被遮挡1遮挡后的图像补全后的图像被遮挡2遮挡住，而D网络只关注遮挡后的生成图像和观测图像。因网络就不会关注那块被遮挡2遮挡的区域，哪怕有这块在没遮挡前有信息有效，GAN也无法优化，使生成的图像更加清晰。

同时许多学者也提出了各种各样的改进框架，许多人认为如果只利用原始的GAN的目标函数，没有借助其他的约束，生成的图像很可能比较模糊。故我们要在模型上增加一些约束，使得图片生成的质量更符合视觉上的观感。

由于我们没有一对一的缺失信息的对应标签，但每张图都是部分缺失，所以我们可以利用***L***区的像素作为生成图片的监督，来使得生成图像更加真实。现令***L***区作为真实值，则可以得到没被遮挡区域的真实标签：

#### 4.2.4.1感知损失

生成对抗网络的其中一个目标函数是要尽可能的缩小生成图片和原始分布的差异，而感知损失将补全网络看成一个风格转换问题，由于生成补全图像是从高层卷积得到，它将与真实图片的卷积得到的特征图与生成卷积得到的特征图进行比较，使得高层卷积信息特征相似，受到风格转化启发，目标函数为：

由于GAN可以利用监督学习训练补全生成网络的效果的原因并不具有可解释性，但是通过感知损失，我们可以理解为GAN的每一层内容生成图都是以一种不直接的方式使生成网络学习到原始数据的分布。

#### 4.2.4.2重构损失

由于UNET的网络结构类似与自编码器，我们目标不止能让D网络分不清出生成图片是属于哪个分布，同样能使得图片经过该网络后，能够重构出原来的**L**区的像素，所以对于**L**区，可以得到目标函数：

该优化目标由两部分构成，第一项是原图的L区和重构图的L区的像素恢复情况，而第二项是用来防止网络崩溃后，产生的全图变为极端值的情况。因为一旦变成极端值，意味着网络出现梯度消失现象。

#### 4.2.4.3 GP Loss

原始GAN存在模式坍塌、训练不稳定等问题，WGAN理论上证明了GAN目标函数中的JSD散度不适合衡量两个分布之间的距离，用Wasseetein距离可以更好的衡量距离，并且解决了模式坍塌的问题，生成的结果也更多样性。目标函数如下：

但是满足WGAN要求权重必须满足lipschitz条件，为了逼近这个条件，WGAN的作者将权重进行裁剪，强制满足这个条件，所以实际上，生成结果并没有理论上的好。为了改变这个条件，有作者提出了WGAN-GP，因为如果把权重限制到一个范围内，最后网络训练出来的权重，可能不是范围上界就是范围下界，且强制裁剪会导致梯度消失或爆炸。因此作者提出了一个GP惩罚项，来防止上述现象的发生：

为了让模型更加稳定，目标函数就变为：

所以我们需要优化的总目标函数为：

我们希望能最大化D的能力，所以对D网络有：

对补全网络G有：

我们可以用梯度下降法来依次求解D网络和G网络。

## 4.3 算法流程

|  |
| --- |
| 算法1 监督生成补全算法(ImproveAGAN) |
| 输入：缺失图片集 ，无污染图片集，梯度下降学习率和超参数*，，，。*  输出：缺失图片对应的Mask， 补全后的图像   1. 预处理： 2. 训练用来预测缺失位置的Mask网络，并保存该模型 3. 载入Mask网络，预测的掩模 4. While 网络收敛： 5. 从缺失图像集中采样出个样本， 6. 固定G网络的参数，更新D网络的参数：  9. 从观测样本图像集中采样出个样本， 10. 固定D网络的参数，更新G网络的参数 11. + 13. End while 14. 保存模型 15. 将输入到G网络中，生成补全的图像 |

## 4.4 实验结果与分析

本节我们将针对AmbientGAN补全网络和impoveAGAN进行实验对比分析。本次实验所用的数据集为MNIST和CelebA。

### 4.4.1 实验设备配置

我们在Python环境下完成我们的实验。设备的操作系统为ubuntu18.04，GPU显存为11G，CPU型号为Inter(R) Core(TM) i7-5930K CPU @ 3.5GHz内存为29G。当图片的规格较大时，计算机的显存较小导致无法容纳Attention网络中中间低秩矩阵。所以我们只在MNIST数据集上预测Mask，其他实验直接用已知Mask接入。

### 4.4.2 MNIST 实验

MNIST是一个在深度学习中很常用的数据集，它是大数据集NIST的一个子集。MNIST数据集是有6000训练样本和1000个测试样本，总共有10个类别样本的手写数字体的集合。每张图片的大小为2828，图片通道只有1。如果把它的所有像素拉长成一个向量，则该向量维度只有784。我们可以在MNIST数据集上进行Attention网络的实验，这是因为Mask Attention网络中间层的大小仅有，这个矩阵内存大大小于GPU显存的容量。

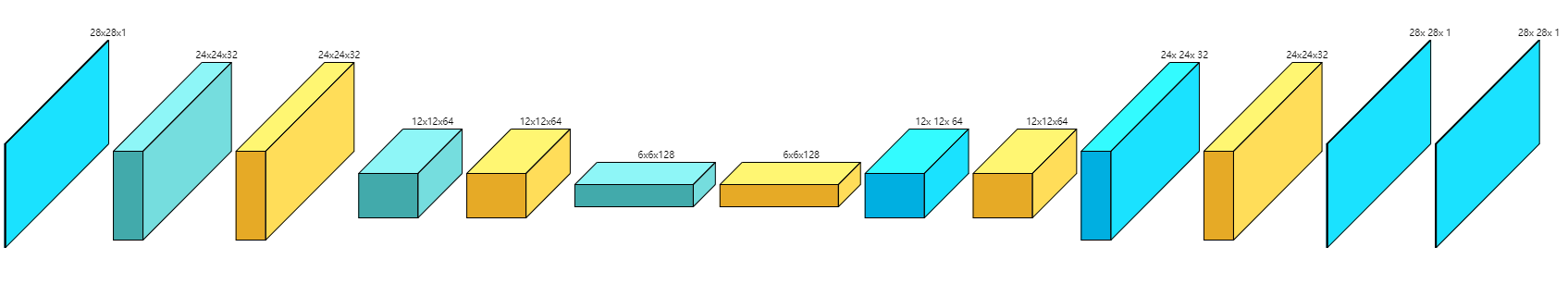
 MNIST图片的输入规格比较小，而我们上面所列出的网络结构对它来说过于复杂，容易过拟合。我们用LeNet-5结果来作为U-NET的基本模块，网络模型大致如图4.5所示：

图 4.四.5 小型数据集所用的生成补全网络

该生成网络先用卷积层提取特征，并且每一卷积层后面都有一个BN层（黄色方块）。从图中可以找到的两个黄色方块，我们对起增加一层跳层连接，使得空间结构信息和抽象的语义信息能够有机结合。

本次实验我们的参数设置如表4.2所示：

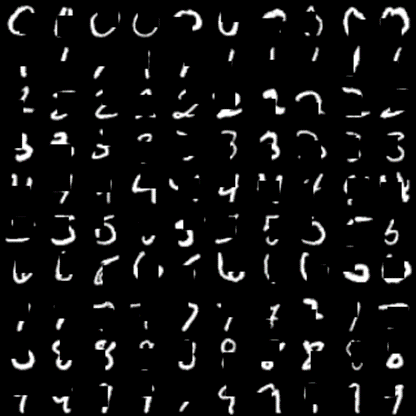
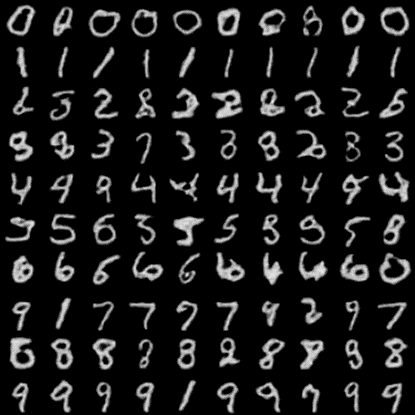
表4.2 MNIST实验的参数设置

|  |  |  |
| --- | --- | --- |
| 参数 | 参数字母 | 数值 |
| 学习率 |  |  |
| GP项参数 |  | 0.005 |
| 感知项参数 |  |  |
| 重构项参数 |  |  |
| 重构项惩罚参数 |  |  |

我们做了两种类型的损失度量下的补全实验，其一是block遮挡，即随机把每张图片的四分之一的区域挡住。另一种类型是dropout度量，即每个像素点有一定的概率丢失。

对于这个实验，我们先用部分数据把Mask网络将缺失模板训练出来，再训练ImproveAGAN。由于大部分论文在MNIST上的实验效果都不错，补全的结果也没有明显的差别。因此本实验只展示imporve AmbientGAN的效果。

如图4.6(a)所示，图片被整块遮挡物遮挡住，且遮挡的位置不固定。因此我们人眼无法看清图片中的真实数字。我们将这些被遮挡的图片用补全网络将其补全，结果如图4.6(b)所示，Improve AmbientGAN补全网络在数据有缺失度量的情况下，学习到了数据的原始分布。它能协调被遮挡的区域和非遮挡区域，使得生成的图片更加自然。为了隐藏边界，生成网络将图片的整体亮度稍微拉低来补偿丢失的区域。由于图片丢失的面积太大，导致补全的结果有多种。比如手写字体2的补全，Improve AmbientGAN根据不被遮挡的部分图片信息，将图像补全为2，8，3，6。实验说明了剩余部分具有相同的特征的数字都有可能出现在补全结果中。这种现象也验证了补全问题是一个病态问题，存在许多可行解，不存在唯一解。



(a) 损失度量为block的手写数字 (b)ImproveAGAN补全效果

图 4.四.6 MNIST 补全实验

如图4.7所示，(a)(c)图以不同的概率将像素丢掉，(b)(d)为补全效果。其中(a)的缺失率为百分之30，从图中可以看出，人眼基本上还能识别，补全的结果也和人眼判断有一致。(c)图的丢失率为百分之80，可以看出人眼基本上已经无法识别图像所表达的数字，但是Improve AmbientGAN依然可以将缺失的信息补全，且补全后的图片很光滑。

|  |  |  |  |
| --- | --- | --- | --- |
| (a)缺失率=0.3 | (b) (a)图补全效果 | (c) 缺失率=0.8 | (d) (c)图补全效果 |

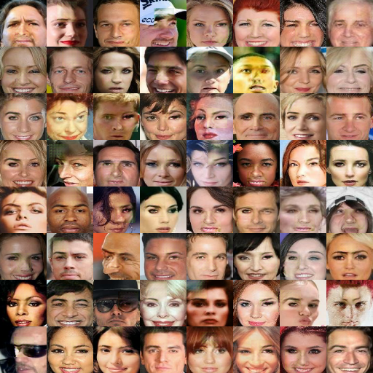
图 4.四.7 Dropout补全实验

我们将MNIST实验作为引子，可以直观的发现在小数据集上，Improve AmbientGAN可以达到很显著的效果。哪怕缺失的面积很大，它也可以将缺失补全，甚至与原图一模一样。接下来，我们将用更复杂的人脸数据对模型来进行进一步探讨。

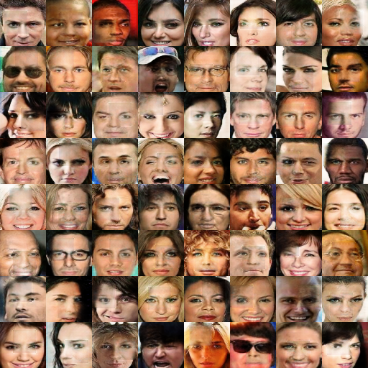
### 14.4.3 CelebA实验

CelebFaces Attribute(CelebA)是一个由香港中文大学实验室开放给公众的人脸数据集。CelebA有202599张人脸，并且每张图片都有标签、特征标记、人脸特征点标记和40个属性。该数据集常被人应用于计算机视觉，人脸风格转移、特征识别甚至数据挖掘任务。由于每张图片的分辨率不一样，故我们把所有的图片都规范成大小。并把CelebA数据集的70%图片划分为训练集，另外百分之30%划分为测试集。为了方便比较，本次实验将着重关注人脸部分，把中心部分裁剪出来以排除图像背景对结果的影响。这样操作后，CelebA图像的大小就变为,。由于处理后的图像规模远远大于MNIST图像，故使用图4.2.3的U-Net架构作为生成补全网络。下面我们将在数据集上分别施加Block和Dropout损失测量。

如图4.8所示，我们将Block 遮挡的大小为。我们对比了AmbientGAN框架下的原始模型和我们更改后的模型。我们发现这两种方法都可以补全人脸，如果没有关键部分被遮挡，补全的图像都比较自然。一旦人脸眼睛鼻子等被遮挡住，虽让能生成人脸的关键要素，但生成的人脸很不自然。且补全内容的色度和亮度，与原有的像素不在一个程度上。



AmbientGAN补全



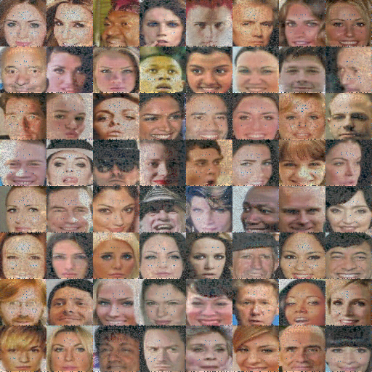
Improve

AmbientAGAN

被遮挡图片 补全结果

图4.四.8 两种方法的Block补全实验

如图4.9所示，给测试样本的每个像素以百分之30的概率丢弃，丢弃后的结果如第一列所示，补全的结果如第二列所示。从实验可以看出原始的AmbientGAN补全网络对于Dropout图片的补全能力有限，这说明了在这种损失度量下，AmbientGAN网络容易受孤立点影响，导致网络不稳定，补全效果不好。而加了约束的Improve AmbientGAN可以解决这个问题，从而让生成的图片更加清晰。



AmbientGAN补全



Improve

AmbientGAN

图 四.9 两种方法的Dropout补全实验

## 4.5 AmbientGAN超分辨任务

我们采用一种特殊的技巧，使得Improve AmbientGAN补全网络能用在超分辨任务上。如图4.10所示，第一步是将低分辨的图像扩展到原来图像的四倍，其中每四个像素点有随机的位置放置原始的像素点，其他三个位置的像素点都置0。这些被置0的区域就是需要补全的区域。我们通过补全这样的一个图片，从而生成高分辨率的图像。

在超分辨任务中，对于输入图像的每个像素点（第行第列），预处理后的图像为，则：

对这个任务来说，损失测量就是每一个的patch，他们中随机三个位置的像素被丢弃。因此这个任务的观测样本是在原图上加一个这样的损失测量，而输入样本是将低分辨图片按式子()处理。然后通过生成补全网络得到一个高分辨的图像，最后让此高分辨图像加上该损失测量，并让D网络判别它的分布。

训练的参数设置与上述实验类似，由于初始化的情况不一样，所以要根据每次训练的loss进行调整，把这些loss调整到一个量级上。否则就会导致只有一个loss对结果有影响。本次实验同样随机抽取百分之70的样本作为训练样本，剩下的样本作为测试。

如图4.10所示，从测试集中随机选出三张图，我们发现，用传统方法处理的图像，放大后许多细节变得模糊。用AmbientGAN补全网络处理的 图片，图片的细节更加的丰富，使得整个图像呈现的更有画面感。但是神经网络普遍存在棋盘效应[]，所以这种方法生成的图片普遍存在颗粒感。即使是这样，但是我们的方法还是产生更多的细节，使得任务更加饱满。未来的研究上，我们将会更加关注这些颗粒感，使得整个图片分辨率程度更高，更接近与真实值。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 输入图片 | 预处理 | **生成超分辨图** | 双线性插值 | GT |
| (a) |  |  |  |  |
| (b) |  |  |  |  |
| (c) |  |  |  |  |

图 四.10 超分辨实验

我们同时也对比了PSNR和SSIM这两个重要的图像质量评价标准。我们发现生成网络生成的图片在这两个指标下，数值偏低。这是因为GAN的只是学到图像的整体分布，而PSNR和SSIM是基于每个对应点的像素值，以及图片的色度亮度计算的。如表4.2可知，我们方法的PSNR和SSIM这两个指标略低于baseline。但我们发现，即使在不改变分辨率的前提下（一个像素变成方块的放大），没有任何细节的增加，PSNR和SSIM的值仍然非常大。由此说明，我们的超分辨应用仍然是有效的，PSNR和SSIM这两个值与直接放大的图片相比，没有显著的降低。这些都足以说明，生成补全网络学到了原始数据的抽象特征，并在对这些特征加以创作，而这些创作使得PSNR和SSIM的值偏低。

表4.2 补全图片PSNR和SSIM

|  |  |  |  |
| --- | --- | --- | --- |
|  | 双线性插值 | 复制像素 | Our |
| 图1 | 0.9549  28.9183 | 0.9491  28.0491 | 0.9079  26.8845 |
| 图2 | 0.9482  31.6745 | 0.9216  29.0291 | 0.8644  27.3243 |
| 图3 | 0.9822  35.7679 | 0.9615  31.2371 | 0.9003  30.2042 |

## 4.6 本章总结

本章将AmbientGAN的生成网络改成补全网络，并将自编码器引入到该框架内，使得生成的图片更加清晰。本章对Attention网络进行探究，发现Attention网络虽然可以轻易得出权重掩模，但是，计算量偏大，不能直接应用于数据集上。为了增强网络的空间性，我们将浅层网络的普通卷积换成坐标卷积，使之可以利用Mask的信息。与此同时，每个需要补全的图像，都存在被遮挡的区域和非遮挡区，非遮挡区具有完整的图片信息，所以我们利用这个信息增加了一些约束，使得生成的图片能更清晰。最后，我们还发现，这个模型可以应用于超分辨任务。

# 基于GAN的部分缺失补全算法

由于现在AmbientGAN大部分在操作的过程，实际上还是要依赖于无缺失的样本，否则生成器的损失测量的强约束条件可能不能满足，故本章将不用AmbientGAN这个网络框架，D网络将直接识别生成器的图片，并只需要使用部分无缺失样本就可以实现。

本章的主要内容如下，第一节为对模型的介绍，第二节为

5.1改进AmbientGAN的问题

AmbientGAN在实际操作上，是需要用真实的数据才能符合损失测量，且约束条件非常严格，生成器的施加损失测量参数一旦与真实样本不一样，生成的图片就无法拟合原始的数据分布。而在现实生活中，我们有真实数据的分布，也有污染分布不属于同一个分布的图片，故，我们去掉AmbientGAN损失度量这个约束条件，让GAN来自己判断数据是属于哪个分布。

我们的目标是在有限真实数据的辅助下，使得GAN补全出来的数据能够符合真实数据的分布，使得图像看起来更真实清晰。

5.2 修复网络方法

由于前一章中生成网络表现出强大的生成能力，本章所有的生成网络架构采用上一章的U-net的结构。由于一张缺失的图片包含着信息清楚的像素点，在实际显示的时候，可以显示原先的像素点，这样可以充分利用图片中已有的信息，来使得整个效果达到最好。故补全后的图像为：

该模型的好处在不知道缺失数据的具体内容的情况下，一方面尽可能用自编码器恢复信息完整区域的重建，另一方面，利用缺失部分和未缺失部分之间的空间关系来预测出补全内容。由于最终的图片是由两部分构成的，对于一张自然图像来说，这两部分分别优化后，不能让边缘部分有明显的边界，故对于生成器G来说，不止要使得生成的样本连续光滑，还要使得修复后的图像变得光滑。从第二章可知，TV模型可以保持渐变区域的光滑度，同时能够锐化边缘，将该模型引入生成对抗修复网络中，在目标函数增加一个TV loss，不妨假设S是修复图像的位置区域，因此有：

整体图片光滑的同时，也应该确确保生成的图片也应该光滑，所以TV项的Loss应为：

由于本章加入了一个真实像素和补全部分的拼接，因此，对判别器D来说，它不只需要判断生成器的真假结果，还需要最后的补全图像，是否能符合数据的真实分布，因此模型架构如下所示：

在这个修复网络模型中，我们保留了原始GAN的结构，并且增加一个修复结果的判别器测试，因此对判别器来说最后的Loss为：

总共的生成器Loss如下

5.3实验结果与分析

本节主要针对改进版修复网络的效果性能进行评判，从视觉和测量标准两方面进行评判，本节使用到的数据集有celeba人脸（不排除背景），和中文手写字数据集

5.3.1 celeba人脸数据实验

5.3.1.1 数据预处理

本次使用的celeba人脸数据集由于每张图片的大小规格，为了方便训练，我们统一将图片resize成，百分之70的图片用来做训练集，百分之30的图片用来做测试集。因为在现实世界中，几乎大部分从自然世界收集到，没有背景的图片少之又少，因此本次实验，我们保留人脸图片的背景信息，不做任何裁剪。

5.3.1.2 Dropout实验

我们令每张图片的每个像素以0.2丢失，部分丢失后的图片如图所示

# 总结和展望

## 6.1 总结

## 6.2 展望

附录

实验中用到的数据集

1. Lena图
2. MINIST
3. Celeba人脸数据集
4. 中国汉字数据集

# 参考文献

1. Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. Advances in Neural Information Processing Systems, 2014.
2. Christopher Bishop, Nasser Nasrabadi. Patter recognition and machine learning. Springer, 461-462. 2006.
3. Diederik Kingma, Max Welling. Auto-encoding variational bayes. Advances in Neural Information Processing Systems, 2013.
4. Martin Arjovsky, Leon Bottou. Towards principaled methods for training generative adversarial networks. In International Conference on Learning Representations, 2017.
5. Martin Arjovsky, Leon Bottou, et al. Wasserstein gan. Arxiv preprint: 1701.07875, 2017.
6. Ilya Tolstikhin, Oliver Bousquet, et al. Wasserstein auto-encoders. In International Conference on Learning Representations, 2018.
7. Ashish Bora, Ajil Jalal, et al. Compressed sensing using generative models. In International conference on machine learning, 2017.
8. Raymond Yeh, Chen Chen, at al. Semantic image inpainting with deep generative models. Arxiv preprint: 1607.07539, 2016.
9. Ashish Bora, Eric Price, et al. AmbientGAN: Gnerative models from lossy measurements. In International Conference on Learning Representations, 2018.
10. Guangliang Chen and Deanna Needell. Compressed sensing and dictionary learning. Proceedings of Symposia in Applied Mathematics, 73, 2016.
11. Chiyuan Zhang, Samy Bengio, et al. Understanding deep learning requires rethinking generalization. In International Conference on Learning Representations, 2017.
12. Ian Goodfellow, Jonathon Shlens, et al. Explaining and harnessing adversarial examples. In International Conference on Learning Representations, 2015.
13. Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]//Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 2000: 417-424.
14. Efros A A, Leung T K. Texture synthesis by non-parametric sampling[C]//iccv. IEEE, 1999: 1033.
15. Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. Physica D: nonlinear phenomena, 1992, 60(1-4): 259-268.
16. Shen J, Chan T F. Mathematical models for local non-texture inpaintings[J]. SIAM Journal on Applied Mathematics, 2002, 62(3): 1019-1043.
17. Darabi S, Shechtman E, Barnes C, et al. Image melding: combining inconsistent images using patch-based synthesis[J]. international conference on computer graphics and interactive techniques, 2012, 31(4)
18. Tsai A, Yezzi J A, Willsky A S, et al. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification[J]. IEEE Transactions on Image Processing, 2001, 10(8): 1169-1186.
19. Esedoglu S, Shen J. Digital inpainting based on the Mumford{Shah{Euler image model[J]. European Journal of Applied Mathematics, 2002, 13(4): 353-370.
20. Nash J F. Equilibrium points in n-person games[J]. Proceedings of the National Academy of Sciences of the United States of America, 1950, 36(1): 48-49.

# 致谢

时光飞逝，岁月如梭，转眼间三年的研究生生涯已接近尾声。在此论文完成之际，我要向所有给予我帮助的老师、同学、家人表示我最诚挚的感谢。

首先，我想感谢我的导师徐立华老师。这三年以来，我的成长和进步都离不开您的付出和培养。这三年以来，徐老师您给我提出所有的宝贵的指导和建议都让我受益匪浅。您对工作的认真负责、对学术的钻研精神以及为人处世之道，都是值得我终生学习的。我所有的收获和成就都离不开您的贡献和努力。希望您在这里能接受我最诚挚的谢意，希望我能在最后给您交上一份满意的答卷。

同时，我还要感谢谢瑾奎老师，徐鸣老师，徐飞老师，窦亮老师。你们这三年的言传身教同样是我成长中的不可或缺的养分。谢老师思维敏捷，为人热心，总是能提出建设性的意见。一针见血的指出问题使我受益匪浅。徐鸣老师平易近人，常常和同学们交流聊天，从这些聊天中我得以一窥一个严谨研究者的所思所为。徐飞老师常常给大家带来了新的知识和思维，拓宽了我的眼界。窦亮老师给予我做助教的机会，学会让我换一个角度看问题，锻炼了我的思辨能力。老师们给予过点点滴滴的指导与帮助将使我受用终身。

我还要感谢所有计算机系的老师，不管是否上过您们的课，我在华东师范大学学到的知识中包括了你们在教育上的辛勤耕耘。也要感谢本论文的各位评审老师以及答辩老师，你们的指导使我多有收益。

同时我也非常感谢一直关心帮助我的同学。感谢和我同一届的杨帅同学、章倩雯同学、郭滨瑜同学和你们共同学习共同生活的时间非常开心，祝愿你们在未来能实现各自的梦想。感谢黄承超师兄、陈森师兄、范玲玲师姐，给我的学术生涯带来的巨大帮助。也感谢卜文奇同学、袁宇杰同学、汪庆顺同学、叶莎莎同学、张雨同学、吴瑜珠同学、蒋欢同学、郑皓月同学等，你们在生活中和学习上都给了我直接或间接的帮助。感谢计算机系的每个同学；感谢辅导员杨文彧老师；感谢每位工作人员。

一路走来，我在华师大收获的不仅仅是专业的知识，更学到了为人处世的经验，历练了我的品格，这里的人和事都对我来说弥足珍贵，再次感谢在这所学校赋予我的一切！在今后的工作和生活中，我会铭记师长们的教诲， 继续不懈努力和追求，来报答所有支持和帮助过我的人！

# 附录 攻读学位期间发表的学术论文

Chaomin Shen, Mixue Yu, Chenxiao Zhao, Yaxin Peng, Guixu Zhang:

Parallel Hashing Using Representative Points in Hyperoctants. CIKM 2018: 813-822 CCF-B

Chaomin Shen, Chenxiao Zhao, Mixue Yu, Yaxin Peng:

Cloud Cover Assessment in Satellite Images Via Deep Ordinal Classification. IGARSS 2018: 3509-3512

Chenxiao Zhao, P. Thomas Fletcher, Mixue Yu, Yaxin Peng, Guixu Zhang, Chaomin Shen:

The Adversarial Attack and Detection under the Fisher Information Metric. CoRR abs/1810.03806 (2018) CCF-A