**深 圳 大 学**

本 科 毕 业 论 文（设计）

题目:  **基于深度学习的图像补全技术**

姓名:  **陈宜威**

专业:  **软件工程**

学院:  **计算机与软件学院**

学号:  **2014150001**

指导教师:  **李炎然**

职称: **教授**

2018年4月25日

**深圳大学本科毕业论文（设计）诚信声明**

本人郑重声明：所呈交的毕业论文（设计），题目《基于深度学习的情感分类技术研究》是本人在指导教师的指导下，独立进行研究工作所取得的成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式注明。除此之外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。本人完全意识到本声明的法律结果。

毕业论文（设计）作者签名：

日期： 年 月 日

基于深度学习的图像补全技术

计算机与软件学院软件工程专业

学号：2014150001

**【摘要】**图像修补技术，指的是用图中已知的部分区域来填充修补图像中的受损区域，受损的区域可大可小。广义的说，图像修补技术有着非常悠久的历史，早在古代就已经有对书画文物书画等进行修复。而数字图像修复这一专业术语直到2000年被Bertalmio【1】等人提出，在这将近20期间，得到了巨大的发展，也取得了很多的突破。传统的图像修复方法有很多的局限性，它们通过图像已有的信息来进行修复，而深度学习则提出了基于一种大量数据进行训练的网络，通过卷积的方法来感知图像的信息和特征，利用神经元进行信息和特征的保存，通过数以十万次的训练，使得网络能够达到良好的高层语义的表达能力，达到图像缺失区域的补全。本文在深度卷积和对抗卷积神经网络的基础上，提出了局部和全局鉴别器的方法，两者结合起来感知缺失区域与整张图像的修复效果，还提出了多种误差函数来训练三个不同的网络-完成网络，局部鉴别网络，全局鉴别网络；我们使用开源深度学习引擎Tensorflow实现模型，使用CeleBA和Imagenet数据集进行训练和检验，最后分析实验结果，比较图片修复的效果，检验模型的有效性。最后，对深度学习在图像修复领域的存在的一些问题进行简要的总结和讨论，并展望未来发展的方向。

【**关键词**】图像补全；深度学习；深度卷积神经网络；对抗卷积神经网络；局部和全局鉴别网络

1. 引言
   1. 研究背景及意义

随着计算机技术的发展和进步，各种图片采集工具譬如手机，相机走进了千家万户，人们的生活出现了大量的图像。然而，这些图像在获取，处理，压缩，传输，解压缩的过程中，难免会出现图像的损坏，导致图像局部的信息丢失。

一般而言，图像修复技术在以下方面有比较大的应用场景。1.图片缺失区域的修复：一些老旧照片以及历史书画等文物在存储过程中就出现损坏或变质的情况，这时图像可以利用数码设备将原始的文物扫描为数字图像，应用图像修复技术对其进行修复，可以达到既不损坏原始的图片，又能获得修复后的图片效果。2.图片内容的移除：在日常生活中，常在图片或者视频里嵌入文字，譬如一些广告词和新闻标题，而当我们需要一张没有文字水印的图片的时候，文字就会影响我们对图片的观察和提取。这时，我们可以先进行图像文字的移除，然后应用图像修复技术来对图像中文字所在的区域进行修补，从而达到文字从图像中移除的目的。3.图片失真的修复：在图片进行放大以及编码压缩解压后存在的图片失真情况，应用图像修复技术可以获得高质量的还原。4.在摄影，影像制作等方面，为了达到一些想要的视觉效果，通常会应用数字图像补全技术对图像的某些特定的物体进行隐藏或者移除，以达到带给人们视觉上想要的冲击以及新鲜感的目的，因而数字图像修补技术在特效制作等领域发挥出了极其重要的作用。

数字图像修复，简而言之，就是利用图像中的已知信息，修补图像中丢失信息，从而达到完整图像的过程。在这个过程中，为了使得图像的完整性不被破坏，需要对受损的图像进行信息的填充，目的是不要求修复的结果能完全还原图像的缺失区域，相对的，它能够达到很好的视觉观赏性，在不观察原图像的情况下，不能察觉出图像的内容改动。也由于其解的不确定性，吸引了越来越多的学者投入更多的精力进行研究。

* 1. 国内外研究现状

图像修补技术，指的是用图中已知的部分区域来填充修补图像中的受损区域，受损的区域可大可小。广义的说，图像修补技术有着非常悠久的历史，早在古代就已经有对书画文物书画等进行修复。而数字图像修复这一专业术语直到2000年被Bertalmio【1】等人提出，在这将近20期间，得到了巨大的发展，也取得了很多的突破。

图像修复(image inpainting)是图像处理的一个较为前沿的领域，也是当前计算机视觉研究方向的热点问题。相对于国外研究的蓬勃发展，而国内在图像和视频修复等领域起步比较晚，理论基础和技术水平也亟待提高，很多研究方法还停留在传统的修补方法上[2]。

图像修复问题已经研究了将近二十年，但是对于图像任意区域分割出大于1/8且小于1/4面积进行修复的问题，由于缺失的区域比较的大，传统的修补方法，包括基于变分PDE[3]和纹理合成[4]等技术，在修复效果都比较的差。

深度学习（Deep learning）[4]作为近年来快速变化的技术，在图像分类、图像分割等计算机视觉问题上展示了其强大的特征提取能力，各种其他领域的研究者看见了深度学习的巨大潜力。人们开始尝试将深度学习方法运用到图像修补的研究中，发现卷积神经网络在特征提取等问题上效果突出。这项技术被认为可以提取图像的高层语义特征[4]。这一项技术在近几年被应用到图像修复领域，得到了越来越喜人的研究效果，相比于传统的方法，利用GPU的强大计算能力，结合TensorFlow这样能够在各种设备运行的强大的人工智能平台，通过深度学习方法从训练集中大量图片数据提取所需要的高频信息，从而对未知的测试样品的填充区域进行预测，从而达到修补图像区域的目的，能达到比传统方法更好的修复效果。

* 1. 本文主要工作

本选题主要针对过去传统图像修复方法对大面积的图像缺失区域进行修补所出现的修复效果不佳，基于深度学习以及TensorFlow平台进一步提高图像修复的效果开展技术研究。深度学习在图像修复领域在近年来由于GPU计算能力的提升，再加上深度学习的诸多研究譬如**卷积神经网络（CNN）**[5]，**生成对抗神经网络（GAN）**[6]的发展，到现在**深度卷积生成对抗神经网络（DCGAN）**[7]，该网络利用卷积层强大的局部特征以及信息的提取能力来进一步提高了图像特征提取的能力，从而达到图像修复效果的进一步提升。

在DCGAN的基础上，我研究了一种由完成网络和两个鉴别器（全局鉴别器以及局部鉴别器）组成的架构[8]。完成网络用于完成缺失图片的修复，而全局和局部鉴别器用于辅助完成网络进行训练，用于鉴别缺失图像是否在全局和局部是否一致完成。在完成网络里面，我除了使用一般的卷积和反卷积以外，还使用了一个空洞卷积[10]的理念，使得卷积核的感受视野变得更大，而不需要通过增加网络的层次来完成。全局鉴别器以整张图片大小作为输入来识别场景的全局一致性，而局部鉴别器只输入缺失区域的图片大小来获得更加高质量的修复效果。这样设计的全局和局部鉴别器对于获得逼真的图像完成起到了非常关键的作用。同时，对于网络的训练过程，我为每一个卷积层添加了批标准化，避免训练过程中可能会出现的梯度消失和梯度爆炸的问题。

1. 深度学习概述

深度学习(Deep Learning)是机器学习(Machine Learning)的一个领域，深度学习通过建立类似于人脑的分层模型结构，对输入的数据进行逐层的提取从底层到高层的特征，从而建立起从底层信号到高层语义的映射关系。2006年，加拿大教授，机器学习的开路者Hinton和他的学生Salakhutdinov在一部顶级刊物《科学》上发表了一篇文章，提出了深度学习的理念，也开启了工业界和学术界的研究热潮。这篇文章主要向人们传达了具有多隐含层的人工神经网络具备很强的特征学习能力，学习到的特征对输入数据有着更加本质的描述，利于人们的可视化。同时，在深度网络里面出现难以训练的问题，可以通过逐层初始化来有效解决，在文章提到，这一困难可以通过无监督学习来解决。深度学习是机器学习中基于对数据进行表征学习的算法，它的好处在于可以使用半监督式的特征学习和分层特征提取高效算法来替代手工获取特征。自深度学习出现以来，它已成为很多领域，尤其是在计算机视觉和语音识别中，成为各种领先系统的一部分。在通用的用于检验的数据集，例如语音识别中的TIMIT和图像识别中的ImageNet, Cifar10上的实验证明，深度学习能够提高语音以及图像的识别精度。

* 1. 人工神经网络

人工神经网络(Artificial Neural Network, ANN)简称神经网络(NN)，是基于生物学中神经网络的基本原理，模拟人脑的神经系统对复杂信息的处理机制的一种数学模型和运算模型。神经网络算法领域最初是被对生物神经系统建模这一目标启发，但随后与其分道扬镳，成为一个工程问题，并在机器学习领域取得良好效果。然而与生物神经网络相比，该模型具有集体并行分布和自学习的能力，能够将输入的信息进行加工和存储结合在一起。总而言之，人工神经网络是一个能够学习，能够总结归纳的系统，也就是说它能够通过已知数据的实验运用来学习和归纳总结，通过把问题表达成神经元单元间的权来解决问题，因此它由大量的神经元相互连接而构成。人工神经网络具有自学习能力，具有联想存储功能，具有高速寻找最优解的能力，也正因为如此，最近十年来，人工神经网络的研究在国内外掀起一股热潮，已取得了非常大的进展，多种神经网络模型如卷积神经网络(Convolutional Neural Network, CNN)，循环神经网络(Recurrent Neural Network, RNN)，在模式识别，计算机视觉，医学，生物等等领域展现了良好的智能特性。

* 1. 人工神经元

人工神经元是一种模仿动物神经网络行为特征的模型。每个神经元代表着一种特定的输出函数，称为激励函数。每个神经元之间的连接则代表着通过该连接信号的加权值，也称之为权重。更通俗的讲，这些就是人工神经网络的记忆。网络的输出会根据网络的连接方式，权重以及激励函数的选择不同而有所差异。单个神经元的作用就是把一个N维向量的空间把一个超平面分成两个部分（称为判断边界），当给定一个输出向量时，神经元可以给出这个向量位于超平面的哪一边的依据。

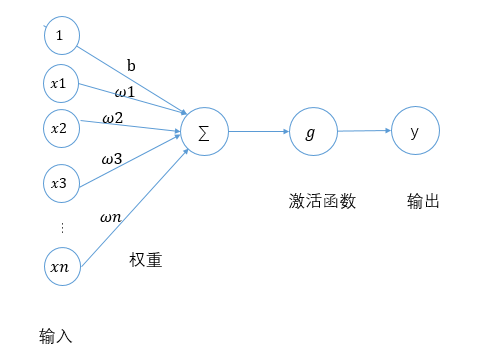
  
图 1：人工神经元结构模型

图 1为神经元结构示意图，我们假设输入向量，权重向量，另外需要一个常数项作为偏置项，则表示激活函数。设表示神经元结构的输出，则我们可以表示该结构的语义函数为：

激活函数(Activation function)

激活函数的主要作用在于提供非线性建模的能力。如果没有激活函数，那么该网络仅能够表达线性映射，此时即便有再多的隐藏层，其整个网络跟单层神经网络也是等价的。因此也可以认为，只有加入了激活函数之后，深度神经网络才具备了分层的非线性映射学习能力。这些函数有如下特性：1）一般可微性，当优化方法是基于梯度的时候，这个性质是必须的；2）单调性， 当激活函数是单调的时候，单层网络能够保证是凸函数。3） 有有限的输出值的范围， 当激活函数输出值是有限的时候，基于梯度的优化方法会更加稳定，因为特征的表示受有限权值的影响更显著;当激活函数的输出是无限的时候，模型的训练会更加高效，不过在这种情况小，一般需要更小的学习率。一般神经网络有如下几种常用的激活函数如表1所示。

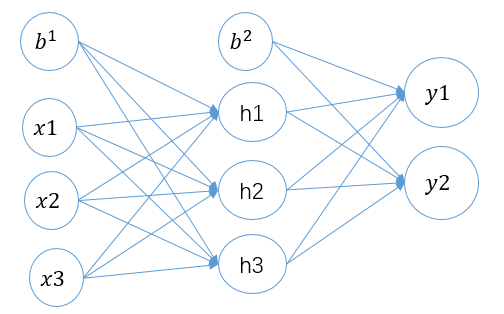
表1 人工神经网络常用的激活函数

|  |  |  |  |
| --- | --- | --- | --- |
| 函数名称 | 表达式 | 形状 | 特性 |
| Sigmoid函数 |  |  | 值域是 [0,1]，当输出接近于0或1，曲线变得很平，导数接近于0。缺点：1）容易出现梯度消失问题；2）输出值不是0均值，导致梯度下降走Z字型下降；3）指数函数计算耗时。 |
| Tanh函数 |  |  | 值域是 [-1,1]，相较于sigmoid输出是零中心的，缺点：1）容易出现梯度消失问题；2）指数函数计算耗时。 |
| ReLU函数 |  |  | 优点：1）对于随机梯度下降的收敛有巨大的加速作用。2）不需要指数计算耗时短。  缺点：1）输出非0均值；2）神经元失活问题。 |
| Leaky ReLU |  |  | 优点:ReLU是将所有的负值都设为零，相反，Leaky ReLU是给所有负值赋予一个非零斜率。一定程度上环节神经元失活问题； |

* 1. 多层神经网络

深度学习通过多层的神经网络结构来增强模型的学习容量。在许多问题上，不同抽象层次可以提取出不同的特征表示，多层次的神经网络结构提供了对应于不同抽象层次的特征的学习能力。例如多层卷积神经网络(CNN)在处理图像时，可能会先去识别不同像素、不同颜色组成的局部特征或局部结构，再由这些基本特征组合出更高层次的抽象特征譬如轮廓，纹理等等，每一次抽象都是对上一层特征的总结，不同层次的网络结构可能在对图像的纹理或者宏观结构以及颜色深浅或轮廓等特征信息进行记忆。增加单层网络结构的神经元数量相当于增加神经网络的维度，以增强高级特征的学习能力和线性变换的能力，而增加网络层数则相当于增加非线性变换的次数，以提高了神经网络对输入特征的抽象能力。

我们以一个**三**层前馈网络(Feedforward Network)为例**，**如图2所示，网络从左到右依次可分为输入层，隐藏层和输出层。我们假设输入向量的维度为，隐藏层向量的维度为，输出向量的维度为。输入向量为，第一层网络的连接权重可表示为一个权重矩阵，其中表示第个输入节点到第个隐藏层节点的连接权重，第一层网络的偏置向量为。



**图2 多层神经网络结构**

假设隐藏层均使用ReLu函数作为激活函数，则第一层网络的输出由以下式子给出：

同理，第二层网络的权重矩阵为，偏置向量为。则第二层网络的输出由以下公式给出：

* 1. 反向传播算法

反向传播(Backpropagation)是一种与最优化方法(如梯度下降法)结合使用的，用来训练人工神经网络的常见方法。该方法对网络中所有权值计算损失函数的梯度，这个梯度会反馈给最优化方法，用来更新权值以最小化损失函数。

1. 激励传播 在每次迭代中，传播环节包含两个步骤：
2. 前向传播阶段：将训练输入送进网络以获得网络预测。
3. 反向传播阶段：将网络预测与训练输入的目标求差，并将误差逐层反向传播到网络各层。
4. 权重更新 对于每个突触上的权重和神经元偏置，按照以下步骤进行更新：
5. 根据反向传播的误差项计算出对应于该权重的梯度项。
6. 将该梯度取反并乘上一个比例加到权重上。

控制权重更新的比例会影响到训练过程的速度和训练效果，这个比例称为学习率(Learning Rate)。因梯度的方向指向误差扩大的方向，因此在权重更新的时候需要对其取反，从而让权重向误差变小的方向偏移。第1和第2阶段可以反复循环迭代，直到网络的对输入的响应达到满意的预定的目标范围为止。

1. 基于深度学习的图像特征提取模型
   1. 卷积神经网络(CNN)

受Hubel和Wiesel对猫视觉皮层电生理研究启发，有人提出卷积神经网络（CNN），Yann Lecun 最早将CNN用于手写数字识别并一直保持了其在该问题的霸主地位。近年来卷积神经网络在多个方向持续发力，在语音识别、人脸识别、通用物体识别、运动分析、自然语言处理甚至脑电波分析方面均有突破。

卷积神经网络(Convolutional Neural Network, CNN)是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元，对于图像的处理有着出色的表现。它有一个或者多个卷积层和顶端的全连接层组成，这些卷积层是用来对图像中的某些特征信息进行提取，譬如第一层的卷积层的卷积核用来检测低阶特征，比如边，角以及曲线等。而随着卷积层的增加，对应卷积核检测的特征就更加复杂。在卷积神经网络里，由于没有反馈的过程，所以第二层卷积层的输入实际上就是第一层的输出，这一层的卷积核则用于检测低阶段的特征组合譬如半圆，四边形等，如此累计，以检测越来越复杂的特征。实际上，我们的人脑的视觉信息处理也遵循这样的低阶特征到高阶特征的模式。

卷积神经网络与普通神经网络的区别在于，卷积神经网络包含了一个由卷积层和子采样层构成的特征抽取器。在卷积神经网络的卷积层中，一个神经元只与部分邻层神经元连接。在CNN的一个卷积层中，通常包含若干个特征平面(featureMap)，每个特征平面由一些矩形排列的的神经元组成，同一特征平面的神经元**共享权值**，这一理论是根据Prechelt等人的总结，有两种方式来减少**过拟合**。一种是减少参数空间的维度，如剪枝，权值共享；另一种就是减少每个维度的有效大小，如权值衰减、早停法等。在卷积计算层，我们这里共享的权值就是卷积核。卷积核一般以随机小数矩阵的形式初始化，在网络的训练过程中卷积核将学习得到合理的权值。共享权值（卷积核）带来的直接好处是减少网络各层之间的连接，同时又降低了过拟合的风险。子采样也叫做池化（pooling），通常有均值子采样（mean pooling）和最大值子采样（max pooling）两种形式。子采样可以看作一种特殊的卷积过程。卷积和子采样大大简化了模型复杂度，减少了模型的参数。卷积神经网络的基本结构如图3所示：

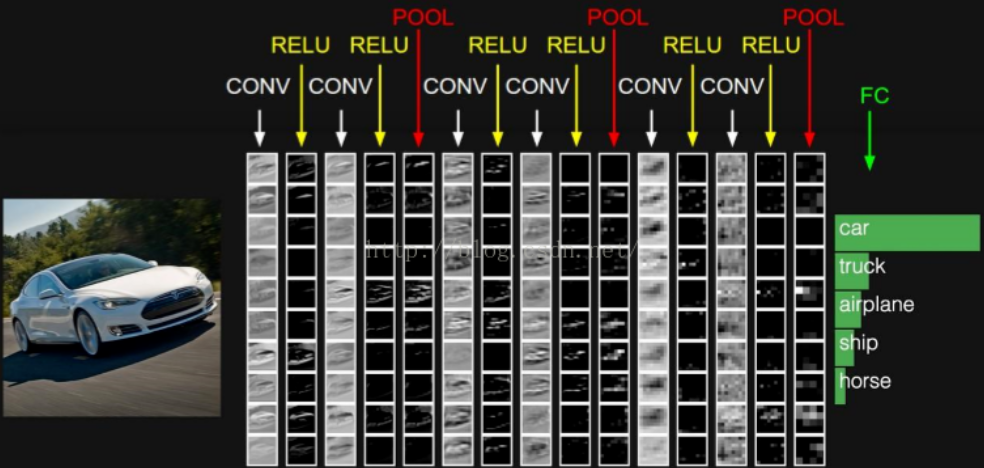


图3 卷积神经网络的一般结构

3.1.1卷积层

在卷积神经网络中，卷积核对局部输入张量进行卷积计算。每计算完一个卷积核大小的局部张量数据后，数据窗口不断平移滑动，直到计算完所有输入数据。这个过程中，需要设定一下几个参数 ：1. 神经元个数：代表卷积核个数，同时对输出的张量大小也有影响；2. 步长：每次卷积核在输入窗口上移动的长度，决定滑动多少步可以到边缘；3. 填充值：在输入数据外围边缘补充若干个0，方便从初始位置以步长为单位可以刚好滑倒末尾位置，为了总长能被步长整除，通常我们只需要指定卷积的方式为Valid或者Same。

我们以步长为1时Valid卷积方式计算过程为例，以一个的输入矩阵和一个的卷积核进行卷积。首先：卷积核的元素和输入矩阵左上角区域的元素对应相乘，然后相加，得到输出矩阵左上角的4这个元素，卷积核与输入数据窗口做内积，其具体计算过程则是逐个元素相乘然后相加，然后，卷积核在输入矩阵上向右移动1（设定的步长值）方格，与输入矩阵的区域元素对应相乘后相加，得到输出矩阵第一行3元素；当窗口移动到末尾时，窗口向下移动1（步长值）格，从左到右依次与卷积核进行卷积操作。后面依次类推，最终得到的输出矩阵。

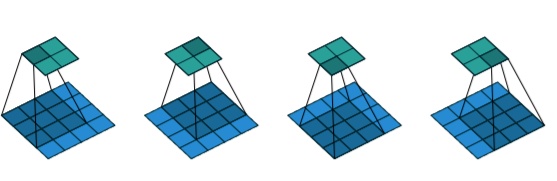


图4 输入矩阵大小为，卷积核大小为，步长为，通过Valid方式的卷积过程

一般地，我们假设输入的数据矩阵大小为，卷积核的矩阵大小为，卷积设定的步长为s，那么通过Valid方式卷积得到的输出矩阵大小为。而通过Same方式，则需要补齐行或者列元素，假设卷积核大小为 ，输入矩阵经过补0后得到的大小为，步长为s，得到的输出矩阵大小为 （注：floor 表示对结果进行向下取整）。

3.1.2池化层

池化层包括平均池化以及最大值池化。池化有点类似卷积，区别在于，池化不需要学习任何参数，只要取池化区域的最大值或者平均值。一般地，我们假设输入的数据矩阵大小为，经过补个0的大小为，池化区域的矩阵大小为，池化的步长为s，那么通过池化得到的输出矩阵大小为 （注：floor为取整数的较小值，如3/2则取整数的1）。

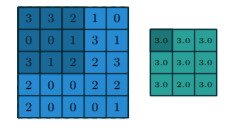


图6 输入矩阵大小为，池化核大小为，步长为，通过平均值方式池化的过程

3.2生成对抗网络(GAN)

生成对抗网络（Generative Adversarial Nets,GAN）是根据2014年Goodfellow等人提出的一种无监督学习的机器学习的训练方法。生成对抗网络通常由两部分组成，一个是网络的生成器，一个是网络的鉴别器，生成器类似于生活中的罪犯，而鉴别器好比如一个警察，罪犯的目的就是尽可能的去欺骗警察，而警察的目的就是想方设法的不被欺骗，罪犯和警察都在不断的提升自身的能力来达到自己的目的，两者之间的对抗构成了一个动态博弈的过程，同时彼此之间的对抗也使得自身得到不断的提升。

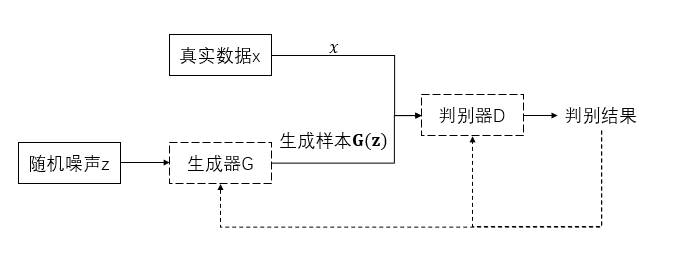


图7 GAN流程图

假设我们的生成模型是，其中是一个随机噪声，而将这个随机噪声转化为数据类型，以图片问题举例，这里的输出就是一张图片。是一个判别模型，对任何输入，的输出是0-1范围内的一个实数，用来判断这个图片是一个真实图片的概率是多大。令和分别代表真实图像的分布与生成图像的分布，我们判别模型的目标函数如下：

类似的生成模型的目标是让判别模型无法区分真实图片与生成图片，那么整个的优化目标函数如下：

可将公式拆分成两个部分进行理解：1) ，其中代表真实样本，表示通过鉴别网络判断结果为真实样本的概率；2) ，表示输入样本的噪声，表示由生成器根据噪声所生成的样本，表示生成样本在鉴别器得到的判断其为真实样本的概率，生成网络的目的是让生成样本越接近真实样本越好，也就是越接近于极大值越好；而鉴别网络的目的则是得到的值为极大值，D()得到的值接近于极小值。

最大最小化目标函数的优化过程其实是一个“二元极小极大”的博弈过程，训练D期望D能够最大概率的区分真实数据样本和通过G生成的数据样本（最大化，训练G以最小化），这需要两者的学习速率基本一致，从而达到从对方学习到知识来提升自己，否则，这两者某一个学习过快，或过慢，将会出现两者的学习到的知识不均衡，一方的损失不能再下降，也就不能为另一方的更新提供学习的知识。

3.3深度卷积生成对抗网络(DCGAN)

深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Networks,DCGAN)是生成对抗网络（Generative Adversarial Networks）的一种延伸，将卷积网络引入到生成式模型当中来做无监督的训练，利用卷积网络强大的特征提取能力来提高生成网络的学习效果。GAN具有无需特定的损失函数，学习过程可以学习到很好的特征表示，然后GAN存在训练非常不稳定，会使得生成器产生没有意义的输出。基于这些缺陷，论文为CNN的网络拓扑结构设置了一系列的限制来使得它可以稳定的训练，使用得到的特征表示来进行图像分类，得到比较好的效果来验证生成的图像特征表示的表达能力，同时对GAN学习到的核进行了定性的分析，展示了生成的特征表示的向量计算特性。

深度卷积生成对抗网络相比于传统的GAN和CNN，有如下特点：1）在鉴别器D模型当中使用带步长的卷积来替代CNN当中的池化层，而在生成器G模型里面使用反卷积来完成对噪声生成图片的过程；2）除了生成器模型的输出层和鉴别器的输入层之外，在网络的其它层上都使用批量归一化(Batch normalization)，加入批量归一化可以稳定学习，有助于处理初始化不准确导致的训练问题，帮助梯度传播到每一层，防止生成器把所有的样本都收敛到同一个点。由于直接将批量归一化应用到所有层会导致样本震荡和模型不稳定，通过在输出层和输入层不采用批量归一化从而防止出现这种现象。3）去除了全连接层，而直接使用卷积层连接生成器和判别器的输入层以及输出层。4）在生成器的输出层使用Tanh作为激活函数，而在其它层使用ReLU激活函数；在判别器上使用leaky ReLU激活函数。

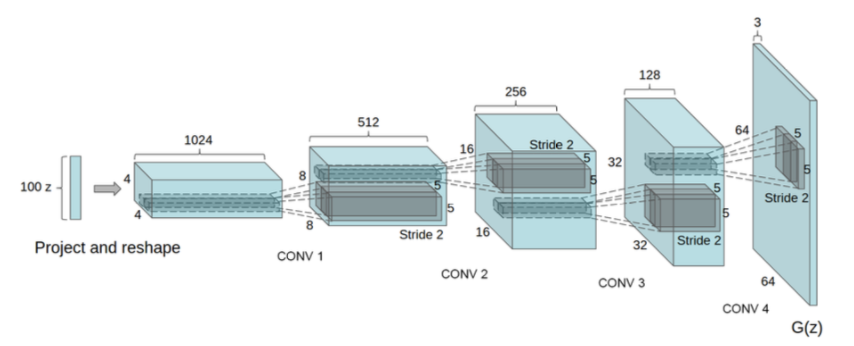


图8 DCGAN的生成网络拓扑结构

3.3.1 反卷积

反卷积(Transposed Convolution)的提出是在Zeiler发表的一篇论文Deconvolutional networks中，当时并没有出现反卷积这个名字，反卷积这个术语正确使用是在其之后的工作中(Adaptive deconvolutional networks for mid and high level feature learning)。这一概念的在很多与神经网络可视化相关的领域利用起来，譬如场景分割和生成模型。反卷积与卷积有着密切关系，可以得知卷积层的前向传播就是反卷积层的反向传播过程。图片在经过卷积和池化之后，图片的原始像素就会丢失了，下一层的卷积学习的是上一层的特征图，所以神经网络的层级越高的神经元与输入图片的连接就变得没有那么强了，也会导致最终学习的结果不好，因而这一模型应用在图像生成领域会导致学习效果不佳，而应用反卷积可视化生成目标图像同样大小的矩阵，把输入图像作为目标去训练整个网络。 [Adaptive Deconvolutional Networks for Mid and High Level Feature Learning]；

一般地，我们假设输入的数据矩阵大小为，反卷积核的矩阵大小为，反卷积设定的步长为1，经过补个0的大小为，那么通过Valid方式卷积，得到的输出矩阵大小为。

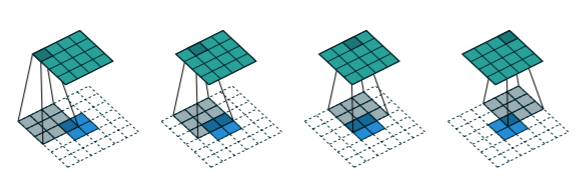
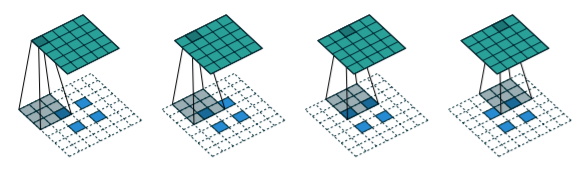


图9 输入矩阵大小为，卷积核大小为，步长为，通过Valid方式的反卷积过程

如果卷积的步长大于1，则反卷积过程需要在输入矩阵的行和列插入0来达到反卷积的小数步长，设卷积过程的步长为,则反卷积的步长相应的变为，假设输入的数据矩阵大小为，反卷积核的矩阵大小为，反卷积设定的步长为，经过补个0的大小为，那么通过Valid方式卷积，得到的输出矩阵大小为。

图10 输入矩阵大小为，卷积核大小为，步长为，通过Valid方式的反卷积过程

1. 基于目标的图像特征提取模型
   1. 图像修复特征提取模型

基于深度卷积生成对抗网络(DCGAN)得到的特征生成的良好效果，我们提出了一种新颖的图像补全算法，用来保持图像在全局和局部的一致性。使用完成神经网络，我们可以通过填充任意区域的缺失区域来完成任意分辨率的图像。生成神经网络由12层卷积层来对原始图片（去除了需要填充的部分），进行encoding，从而得到原始图像16分之一大小的多维矩阵。同时，为了训练该完成网络获得一致性，我们使用了全局和局部的鉴别器。全局鉴别器观察整张带缺失区域的图片以评估它是否整体一致，而局部鉴别器则仅仅观察以缺失区域中心的不大于原始图像4分之一区域以确保生成的图像布丁的局部一致性。然后图像完成网络被训练以欺骗这两个鉴别器，从而达到生成和真实图像无法区分的无论是从整体还是从细节上都保持一致性的图像。

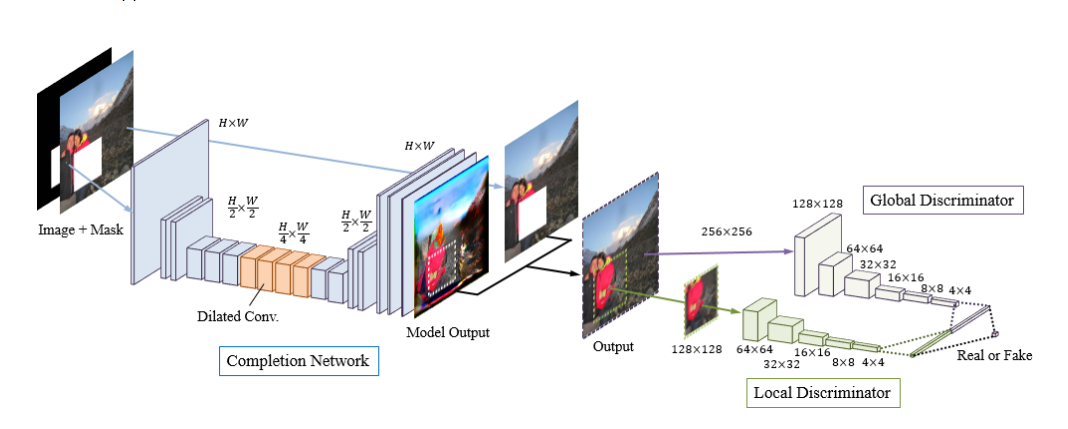


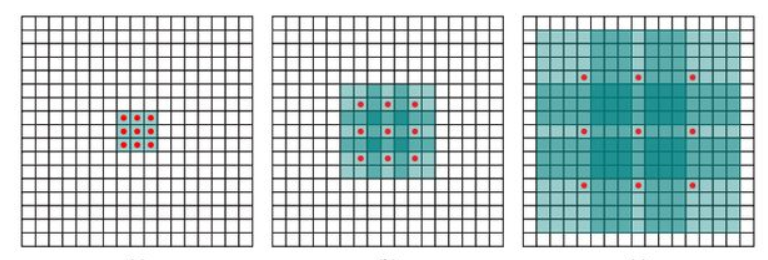
图 11 基于DCGAN改进的完成网络和鉴别网络拓扑结构

* + 1. 完成网络

完成网络采用12层卷积网络对原始图片(去除需要进行填充的部分)进行encoding，得到一张原图16分之一大小的网格。然后再对该网格采用4层反卷积网络进行decoding，从而得到复原图像。我们以一张celebA数据集里的图片resize成大小为的RGB三通道的矩阵，然后随机在图片里面截取一个长度和宽度大于48小于64的空白区域，把带缺失区域的图片输入完成网络。完成网络有如下几个特点：1）去除传统的池化层，通过带步长的卷积来减少信息的丢失；2）引入空洞卷积，在去掉池化下采样操作的同时，而不降低网络的感受视野。3）除输出层以外，其它层以Leaky ReLU作为激活函数，输出层以tanh作为激活函数。4）除了生成器模型的输出层和鉴别器的输入层之外，在网络的其它层上都使用批量归一化(Batch normalization)，加入批量归一化可以稳定学习，有助于处理初始化不准确导致的训练问题，帮助梯度传播到每一层，防止生成器把所有的样本都收敛到同一个点。

4.1．1.1空洞卷积(dilated convolution)

空洞卷积网络(Dilated Convolution network)是在2016年的ICLR(International Conference on Learning Representations)上被提出，当时应用在图像分割领域，随后被应用到语音和自然语言处理领域，且均取得不错的效果。空洞卷积网络的产生是为了解决全卷积网络（FCN）在图像分割领域的问题，图像分割需要输入和输出在像素的形状保持一致，但由于池化层的存在导致全卷积网络(Fully Convolutional Networks)需要通过上采样扩增大小,但是上采样并不能将丢失的信息无损的还原，只能还原成原来的图像大小。空洞卷积网络想法就是把池化层去掉，但是池化层去掉随之带来的是网络各层的感受视野变小，这样会降低整个模型的预测精度。空洞卷积网络的目的在于如何在去掉池化下采样操作的同时，而不降低网络的感受视野。



(a) (b) (c)

我们以的卷积核为例，传统卷积核在做卷积操作时，是将卷积核与输入矩阵中“连续”的的区域逐个“像素”点相乘再求和（如上图a，红色圆点为卷积核对应的输入“像素”，绿色为其在原输入中的感知视野）。而为了不影响原来的数据分布情况，空洞卷积中的卷积核则是将输入张量的的区域隔一定的像素进行卷积运算，这一过程通过在卷积核上补0来完成。如上图b所示，在原来的卷积核的基础上，通过在卷积核每个“像素”周围补0，来达到卷积核的空洞化；在去掉一层池化层后，需要在去掉的池化层后将传统卷积层换做一个扩张系数为2的空洞卷积层，此时卷积核将输入张量每隔一个“像素”的位置作为输入矩阵进行卷积计算，可以发现这时对应到原输入的感知视野已经扩大为；同理，如果再去掉一个池化层，就要将其之后的卷积层换成扩张系数为4的空洞卷积层，如上图c所示。空洞卷积能够在去掉池化层也能保证网络的感受视野，从而确保图像语义的精度。



图 12 完成网络第1到12层参数设置及输入输出张量



图 12 完成网络第13到17层参数设置及输入输出张量

* + 1. 鉴别网络

鉴别网络分为两部分，一部分是局部鉴别网络，一部分是全局鉴别网络，两者联合起来得到整张图片的完整性信息。局部鉴别器以缺失区域为中心的输入矩阵4分之一大小的矩阵区域作为输入，全局鉴别器以整张图片大小作为输入，两者都经过若干层的卷积之后，进行全连接并进行联合过程，如图15所示，输出层经过Sigmoid激活函数输出成一个0到1之间的浮点数；除了输出层以外，其它层以Leaky ReLU作为激活函数。

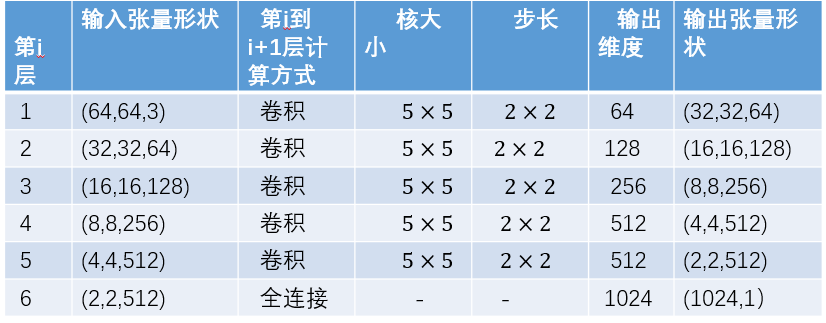


图13 局部鉴别器各层参数设置和输入输出张量

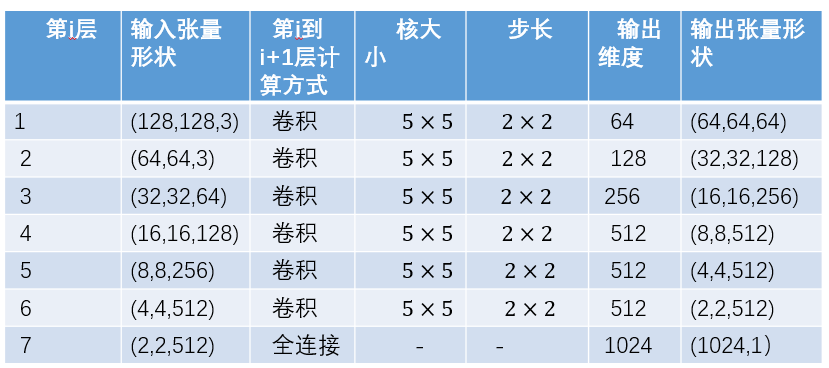


图13 全局鉴别器各层参数设置和输入输出张量

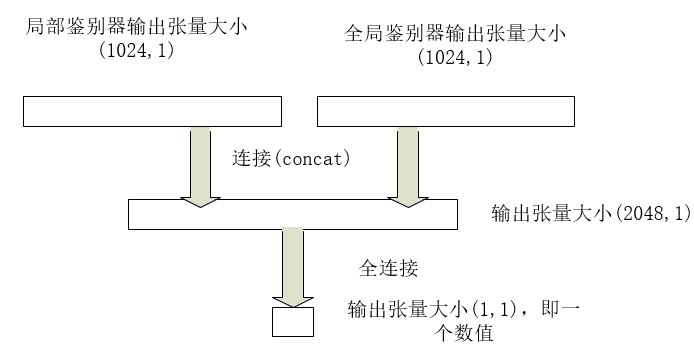


图14 局部鉴别器与全局鉴别器的联合过程

* 1. 模型训练

为了能够逼真的还原图片的缺失区域，我们以GAN误差和均方差误差作为联合的误差函数，为缺失区域，为通过完成网络生成的图片,©代表像素乘法，那么均方误差就是原来的残缺区域与新生成图片的残缺区域的像素乘法然后计算欧几里得距离，其中均方误差公式如下:

由于我们使用了GAN进行训练，我们还需要考虑GAN的目标函数，由于我们的GAN是由两个鉴别网络连接而成得到的概率，当全局鉴别器的输入真实样本图片，局部鉴别器的生成器输入为随机的残缺区域，那么经过整个鉴别网络得到的联合概率为；而带缺失区域的图片为作为全局鉴别器的输入，为局部鉴别器的输入，那么联合的概率为，两者分别作交叉熵得到输入为真实图片和真实缺失区域的概率，输入为真实图片和真实缺失区域的概率，由于两个鉴别器联合过程后经过sigmoid函数非线性化，所以GAN的优化目标是最大化使得其接近于sigmoid函数的最大值1，最小化使其接近于sigmoid函数的最小值0，所以GAN的目标函数由如下公式给出：

考虑到两者的权重比均方误差比GAN误差要占到比重更大，被设置成了一个较小的常数，两者结合起来构成了我们的目标函数：

完成网络和鉴别网络的训练我们都使用梯度下降和反向传播。当我们使用训练集进行训练时，可以每次向神经网络传播单个训练样本，这样的训练方法称之为**随机梯度下降**(Stochastic Gradient Descent)。这样不需要每次遍历所有的数据集，迭代速度会很快但会增加迭代次数，因为每次选取的方向不一定是最优的方向，所以并不能保证算法收敛到最小值，最终可能是在最小值附近来回走动。我们还可以选择每一次迭代的过程选择遍历所有的样本，由所有样本来决定最优的方向，这种方式被称为**批梯度下降** (Batch Gradient Descent)，这种更新参数的方式能够保证算法收敛到最小值，但对于数据集非常大的情况，运行速度会非常的慢。

我们选择了一种介于随机梯度下降和批梯度下降的方法，每次选取大小为b的Mini-Batch(b<样本的总数m)，这种方法称之为**小批量梯度下降法**（Mini-batch Gradient Descent），这样既可以节省整个批量的时间，也使得Min-Batch的计算方向变得更加准确。

为了加快网络的训练速度，我们使用了2台tesla p100处理器进行多GPU版本的训练。实验表明，这能够显著提升我们训练的速度。同时针对不同的误差项，我们均使用了Adam优化器，能够很好的适用大规模的数据和参数的场景。

整个模型的训练以最小化均方误差损失函数和GAN损失函数为目标，首先，我们由生成网络生成缺失区域的图片，与真实图片计算均方误差损失函数，以Tc次迭代训练完成网络最小化均方误差函数；然后我们以Td次迭代最小化GAN损失函数，直到其结果接近于0；最后我们以Tj次迭代最小化联合损失函数，其函数公式由以下公式给出：

1. 实验设计

基于前文提到的理论基础和网络设计，我们选择通过深度学习引擎TensorFlow来实现和训练该模型，测试模型在图片不同相对位置和大小的缺失区域下生成图片的效果，并对比以往的图片生成算法，检验模型的可行性和有效性。

* 1. 实验环境

Tensorflow

模型实现采用Google的开源深度学习框架Tensorflow。TensorFlow是Google开发的一款神经网络的Python外部的结构包, 也是一个采用数据流图来进行数值计算的开源软件库。TensorFlow 让我们可以先绘制计算结构图, 也可以称是一系列可人机交互的计算操作, 然后把编辑好的Python文件转换成更高效的C++, 并在后端进行计算。同时TensorFlow可以运行在多个CPU和GPU上，可以再小到我们日常使用的智能手机，达到数千台数据中心的服务器等各种设备上运行。TensorFlow的计算有有状态的数据流图(Graph)来表示，数据流(Data Flow)在数据流图上进行传递，每个图中的节点代表对数据的计算操作。Tensor被用来装载多维数据和数据的计算，是一个可以被指定维度的张量。在程序运行期间，Tensor仅作为数据张量的容器实际值有Variable来维护。在建立好模型的数据流图之后，TensorFlow使用Session来管理和启动，将数据分发到不同的CPU或者GPU上进行运算。TensorFlow支持自动微分，对反向传播算法提供很好的支持，模型的训练能够更加方便快捷。

* 1. 实验设计

我们使用TensorFlow实现并训练了图片修复模型，设置不同的参数，通过分析模型在测试集上的修复效果。我们采用了celebA人脸数据集和imagenet部分数据集进行训练，其中celebA数据集有11000余张图片，imagenet有25万张图片，其中的90%作为训练数据集，剩下的10%作为测试数据集。

* + 1. 数据预处理

由于数据集里面的图片大小长度和宽度的像素不一致，而我们模型的设计输入是，所以在原来的图片基础上，对数据进行resize，然后我们对数据进行了归一化处理，最终输入模型的张量大小为 ，归一化处理可以在每批训练数据的分布各不相同的情况下，如果数据在不同的分布的情况下，将会大大降低网络的训练速度，所以我们把数据归一化到[-1,1]之间，方便模型学习数据分布。训练过程中Min\_Batch设置成96，迭代次数设置为100000次和200000次，LOCAL\_SIZE代表残缺区域的大小，实验过程HOLE\_MIN和HOLE\_MAX代表的大小，介于40-48之间以及60-64之间，验证不同残缺区域下的修复效果；还要设置为GAN损失函数的学习率设置了1e-4和1e-5，均方误差目标损失函数的学习率设置成1e-3和1e-4。模型每个100次迭代进行一次生成图片结果的校验，校验过程采用与测试数据集一样的数据集。

* 1. 实验结果和分析
     1. 验证GAN误差和均方误差对结果的影响

为了验证学习率对训练结果的影响，我们设置迭代次数为10000次，HOLE\_MIN和HOLE\_MAX设置为介于40-48之间的测试结果，可以看到网络能够较好的修复缺失的区域，学习到了人脸的五官特征，但对于一些边界的缝合上面处理效果不是很好。针对GAN训练过程遇到的梯度爆炸等问题，我把GAN的优化器学习率调整为1e-5。



图16 完成网络生成的图片

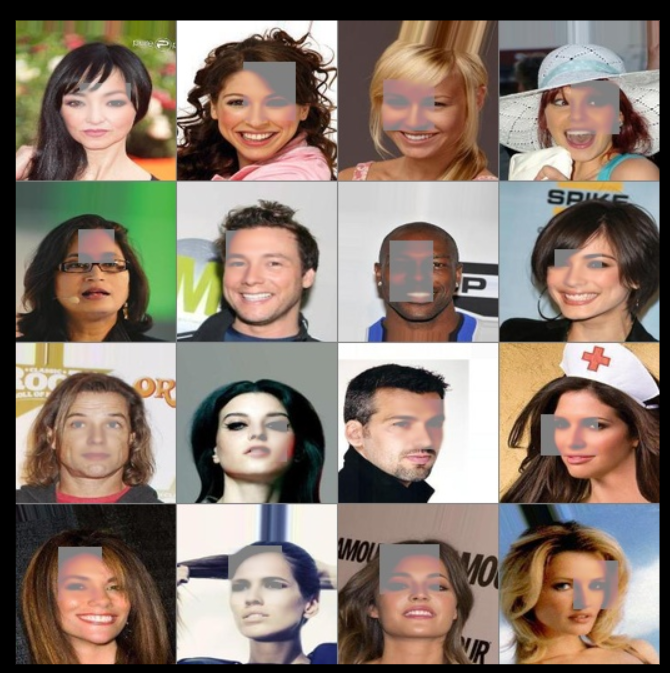
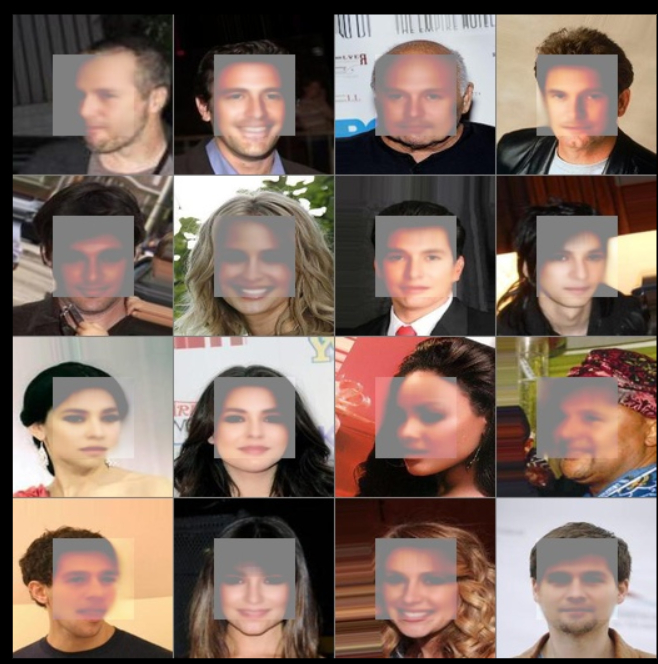


图17 完成网络生成的图片

* + 1. 验证GAN误差和均方误差对结果的影响

实验中我们使用了GAN误差和均方误差，分别训练以降低其差值，已验证其对修复效果的影响。实验表明，使用GAN误差能够提升图像的修复效果，使得生成的图片五官细节更加的丰富。

只使用均方误差的修复效果 使用了GAN误差的修复效果

1. 总结和展望
   1. 本文总结

数字图像修复在数码设备日渐普及的今天，越来越有其研究价值，能够逼真的修复图像，从而能够在生活和工作中给我们带来便利，无疑拥有巨大的应用潜力。

本文基于深度卷积生成对抗网络的良好的特征生成方法，研究了一种新的数字图像修复方法，介绍了网络的结构原理以及训练过程和方法，也得到了相对不错的修复效果。通过详细解析网络各个层级对特征生成方法的研究，在原来的生成网络基础上，引入了全局和局部的鉴别器，通过联合均方误差和GAN误差，进一步提升网络的修复效果。最终，实验在一片空白的残缺区域，能够清晰地生成人脸的五官特征。

* 1. 展望

应用深度学习技术进行图像修复，是近几年来比较新的领域，国内外的研究都处在较为初级的阶段，大量对图像修复模型的改进和尝试不断的出现，其中也有许多修复效果在某一些数据集上能够取得良好修复效果的模型。但是，目前在深度学习的研究都是针对特定的数据集进行特征的学习，对于广泛的图片而言，还无法较为真实的还原出原来图片的效果；同时，在更高分辨率的图片上，由于生成网络需要学习的特征会变得非常多，同时网络的输入矩阵也会变得很大，导致需要的内存变大，参数变多，会出现难以训练的问题。

【参考文献】

1. Zhang Hong ying ,Peng Qi cong. A Survey on Digital Image Inpaiting .School of Comm unication and Information Engineering University of Electionic Science and Technology of China Chengdu 610054.(张红英，彭启琮。数字图像修复技术综述.电子科技大学通信与信息工程学院，成都，610054.2007）
2. BERTALMIO M,SAPIRO G,CASELIES V,et al.Image inpaiting[C].Proceedings of the SIGGRAPH,2000. //PDE
3. DRORI I，DANIEL CO，HEZY Y. Fragment based image completion[J]. ACM Transactions on Graphics,SIGGERAPH,2003,22(3):303 - 312
4. LeCun Y,Bengio Y,Hinton G. Deep learning[J].Nature,2015,521(7553)
5. Kunihiko Fukushima. Neocognitron : A hierarchical neuralnetwork capable of visual pattern recognition. Neural networks 1,2(1998),119-130
6. Goodfellow I , et al. Generative adversarial nets[M]// Advances in Neural Information Processing Systems.MIT Press , 2014.
7. Radford A,Metz L,Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015
8. SATOSHI IIZUKA,EDGAR SIMO-SERRA,HIROSHI ISHIKAWA,Globally and Locally Consistent Image Completion.SIGGRAPH,2017.
9. Yann LeCun,Bernhard,John S Denker,Donnie Henderson,Richard E Howard Wayne Hubbard,and Lawrence D Jackel.Backpropagation applied to hand-written zip code recognition.Neural computation 1,4(1989),541-551.1989.
10. Fisher Yu and VIadlen Koltum.2016.Multi-Scale Context Aggregation by Dialated Convolutions. In International Conference On Learning Representations.

**致谢**

在

**Attention-based LSTM for Sentiment Classification**

**【Key words】**

指导老师：