**基于卷积神经网络的城市卫星图像解释**

摘 要

一般的城市卫星图像是包含多个波段信息的卫星图像，但是数据相对不易获取，本课题主要研究对象是RGB三通道的谷歌卫星地图。通过调研多种图像识别模型，最终选定U-Net模型，对卫星图像中的对象划分成10种类别，为每个类别单独训练。并且开发在线平台，提供在线截图识别以及得到相应的统计信息的功能。

**关键词：**卷积神经网络，卫星图像，U-Net网络，图像识别，城市建成区

**Urban Satellite Image Interpretation Based on Convolutional Neural Network**

**ABSTRACT**

The general urban satellite image is a satellite image that contains multiple bands of information, but the data is relatively inaccessible. The main research object of this topic is the RGB three-channel Google satellite map. Through studying a variety of image recognition models, the U-Net model was finally selected, and the objects in the satellite image were divided into 10 categories and trained separately for each category. And develop an online platform to provide online screenshots to identify and obtain the appropriate statistical information.

**Key words：**convolutional neural network, satellite image, U-Net, image identification, Urban built-up area

目 录

[1 引 言 1](#_Toc513469733)

[1.1城市卫星图像识别、解释的意义 1](#_Toc513469734)

[1.2研究历史与现状 1](#_Toc513469735)

[1.3图像识别中深度学习网络常用方法 2](#_Toc513469736)

[1.3.1前馈深度学习网络基本概况 2](#_Toc513469737)

[1.3.2 反馈深度学习网络基本概况 3](#_Toc513469738)

[1.3.3双向深度网络基本概况 3](#_Toc513469739)

[1.4 本文所作的工作 4](#_Toc513469740)

[2 理论部分 5](#_Toc513469741)

[2.1计算机视觉理论 5](#_Toc513469742)

[2.1.1计算机视觉理论的发展 5](#_Toc513469743)

[2.1.2计算机视觉理论的技术基础 5](#_Toc513469744)

[2.2卷积神经网络CNN 6](#_Toc513469745)

[2.3全卷积网络FCN的基本概况 6](#_Toc513469746)

[2.4 U-Net卷积网络的基本概况 7](#_Toc513469747)

[2.5 过拟合问题及其常用解决方案 8](#_Toc513469748)

[2.5.1过拟合问题定义 8](#_Toc513469749)

[2.5.2过拟合出现的原因 8](#_Toc513469750)

[2.5.3过拟合常用解决方案 8](#_Toc513469751)

[2.6 权值初始化 8](#_Toc513469752)

[2.6.1权值初始化的意义 8](#_Toc513469753)

[2.6.2权值初始化的常见方法 9](#_Toc513469754)

[2.7 多类别分类函数softmax 9](#_Toc513469755)

[2.8 样本选择 9](#_Toc513469756)

[2.8.1样本选择的概念 9](#_Toc513469757)

[2.8.2常用抽样方法 10](#_Toc513469758)

[3 实验内容部分 10](#_Toc513469759)

[3.1 研究内容以及课题目标 10](#_Toc513469760)

[3.1.1 研究背景 10](#_Toc513469761)

[3.1.2 课题目标 10](#_Toc513469762)

[3.2 数据预处理 11](#_Toc513469763)

[3.2.1 图像预处理 11](#_Toc513469764)

[3.2.2 数据采集筛选 11](#_Toc513469765)

[3.3 评估指标 11](#_Toc513469766)

[3.3 建立网络模型 12](#_Toc513469767)

[3.3.1 模型整体结构 12](#_Toc513469768)

[3.3.2 卷积层 12](#_Toc513469769)

[3.3.3 批归一化处理Batch Normalization 13](#_Toc513469770)

[3.3.4 激活函数 13](#_Toc513469771)

[3.3.5 池化层 13](#_Toc513469772)

[3.3.6 Dropout 层 13](#_Toc513469773)

[3.3.7 反卷积层 13](#_Toc513469774)

[3.3.8 连接层 13](#_Toc513469775)

[3.3.9 剪切层 13](#_Toc513469776)

[3.4 后处理过程 13](#_Toc513469777)

[3.4.1 去噪点 13](#_Toc513469778)

[3.4.2 确定图像边界 13](#_Toc513469779)

[3.5 实验步骤 13](#_Toc513469780)

[4 结果和讨论 14](#_Toc513469781)

[4.1模型训练结果和讨论 14](#_Toc513469782)

[4.1.1深度为5的U-Net网络 14](#_Toc513469783)

[4.1.2深度为6的U-Net网络 15](#_Toc513469784)

[4.1.3一般建筑预测结果 16](#_Toc513469785)

[4.2 多种城市用地预测结果和讨论 17](#_Toc513469786)

[4.2.1 合并单一模型 17](#_Toc513469787)

[4.2.2 二分类多模型 18](#_Toc513469788)

[4.2.3 预测结果讨论 19](#_Toc513469789)

[4.3 在线标注平台开发 19](#_Toc513469790)

[5 结论和展望 19](#_Toc513469791)

[5.1 结论 19](#_Toc513469792)

[5.2 展望 19](#_Toc513469793)

[参考文献 19](#_Toc513469794)

[谢 辞 20](#_Toc513469795)

# 1 引 言

## 1.1城市卫星图像识别、解释的意义

图像识别是随着计算机技术的发展而兴起的一门技术科学，它是从图像的信息和数据出发，在专家经验和已有认识的基础上，利用计算机和数学推理的方法对图像内容进行描述、辨认、分类和解释的过程，是信息科学和人工智能的重要组成部分。传统的图形识别相关研究关注于图像特征的设计，近年来，卷积神经网络的特征学习策略在计算机视觉的一系列领域中取得了广泛的成功，逐步取代了传统的人工设计特征的研究，相比于传统的人工设计特征具有更强的判别和泛化性能。本课题应用卷积神经网络，对城市卫星图像提取特征，进行解释，辅助城规研究。

城市卫星图像包含一个城市的城市建成区、功能区等信息，通过对一个区域进行一定时间段内的连续遥感观测，应用机器学习的方法，提取图像有关特征，并分析其变化过程与发展规模，对这些大规模稳定的识别时间序列的城市图像，检测城市时空变化，可以得到一个城市的发展变化，从而在市政规划工作的时候提出一些预见性建议。

广义上的图像识别涉及图像处理、图像分类和图像解释三个范畴，本课题将对城市及其周围区域的卫星图像进行处理、分类、解释，得出诸如城市房屋建设规模、工厂规模、绿地规模等多种的城市统计信息，希望能够实现确定城市建成区以及城市功能区的目标，为今后城市诊断提供科学依据。

## 1.2研究历史与现状

传统的计算机识别主要是关注的图像的一些底层信息特征，常用的方法有基于频谱特征的分析方法、共生矩阵、灰度差分统计等。但是，纹理、色彩等底层次的信息判别性能很弱，具有一定的统计特性，对于平移、旋转不敏感，表达高级语义信息较弱。而且使用这种传统的图像识别方法，泛化性能有严重的缺陷，需要不断的定义信息特征。在具体的特定领域应用当中，如果想要对这些底层特征进行整理、抽象，方法更加复杂，难以复用。

与传统计算机图像识别方法不同，深度学习是基于图像特征进行学习的，是机器学习领域一个新的研究方向。深度学习的概念最早由多伦多大学的教授G. E. Hinton提出，主要目标在于利用计算机强大的计算能力模拟人类大脑的神经连接结构。1998 年，Lecun 等人提出了LeNet-5的卷积网络结构，利用监督式的反向传播方法，在字符识别中获得了良好的表现。2012年，Hinton课题组为了证明深度学习的潜力，首次参加ImageNet图像识别比赛，其通过构建的CNN卷积网络AlexNet一举夺得冠军，深度学习领域越加活跃。

深度学习对于图片这种视觉信号的处理流程主要是先进行边缘检测，然后逐步形成更加复杂的视觉形状，通过组合低层特征形成更加抽象的图像表示，从而给出图像数据的分层特征表示。深度学习通过无监督的预训练方法优化网络权重初始值，解决了传统神经网络随机初始化网络初值导致的容易陷入局部最小值的问题。

深度学习领域的城市卫星图像分类方法在国内外有很多专家学者做了大量的研究工作者，并且已经在影像分类和信息提取等研究中得到了广泛的应用，如土地覆盖的分类问题，多时相动态地物的区分、基于多源空间数据的融合分类、模糊分类、融合先验知识的遥感影像分类、影像结构信息提取等。结合人工智能技术和理论，基于图像特征空间分析的非参数型技术分类方法逐步成熟。卫星图像的获取越来越方便，单纯依靠人工目视判读来获取城市中的信息成本高，效率地，而且有一定的主观性，深度学习方法的出现解决了这样的困扰。在深度学习领域中对城市卫星图像识别的常用分类识别方法为前馈网络中的卷积神经网络CNN，及其延伸出来的各种深层卷积网络结构：VGG16、全卷积网络FCN、U-Net等。

## 1.3图像识别中深度学习网络常用方法

### 1.3.1前馈深度学习网络基本概况

F. Rosenblatt提出的感知机是最简单的单层前向人工神经网络，但是无法解决线性不可分问题，1984年日本学者K. Fukushima等人基于感受野的概念，提出的神经认知机可看作卷积神经网络的一种特例, Y. Lecun等提出的卷积神经网络是神经认知机的推广形式。在前馈神经网络中，所有的数据流动方向只沿着一个方向，从输入到输出，中间一般经过多层隐含层，卷积神经网络CNN就是一种前馈深度学习网络。

卷积神经网络是由多个单层卷积神经网络组成的可训练的多层网络结构，一个典型结构的卷积神经网络包含有输入层、卷积层、下采样（池化层）、全连接层、输出层，其中下采样并不是每一层的必须过程，。比如对于输入图像 ， 即一个具有高稀疏度的二维数组，卷积神经网络通过交替连接的卷积层和下采样层对图像的特征进行提取和压缩降维。

在卷积神经网络中，定义第  层的特征为 ，当 时。对于卷积阶段整体为 ，其中 为第 层卷积核的权值， 表示卷积操作，卷积的结果与偏置向量 相加，然后根据所设定的激活函数获得下一层的特征 。

卷积过程可以拆分成为特征提取和特征映射两部分。

特征提取部分：在卷积层的特征提取过程中的局部感受野概念最早是由D．H．Hubel等人对于视觉皮层细胞研究提出，每个卷积核检测输入特征图上所有位置上的特定特征，实现同一个输入特征图的权值共享，同时滤波器还有一个固定值的偏置向量，

特征映射部分：经过上面所述的特征提取之后，使用激活函数 模拟人脑神经细胞接受刺激之后的情况，常用的有sigmoid、relu方法，将其映射到某一范围内。传统的卷积神经网络旺旺采用tanh、softsign或sigmoid等这种饱和非线性函数，但是因为这种饱和非线性激活函数会随着值的增大而出现梯度消失的现象，所以现在包括谷歌的AlphaGo在内，都在使用relu激活函数,。几种激活函数的公式如下所示：

Relu:



Sigmoid:



Tanh:



Softsign:



池化阶段（下采样阶段），通常在卷积层之后，目的是在保证一定程度的不改变特性的条件下，对得到的每一个特征图进行降维。也有些前馈深度学习网络并不采用池化操作，而是通过在卷积过程中设置相对较大的步长达到类似的效果。池化层采样一般分为平均采样和最大采样，对 的区域进行池化操作的公式分别如下所示：

平均下采样：



最大下采样：





### 1.3.2 反馈深度学习网络基本概况

反馈深度学习网络主要以反卷积网络为代表，Mattew D. Zeiler等人结合卷积网络的架构思想以及基于图像稀疏化表示特征，于2010 IEEE会议上提出了反卷积模型。反卷积网络和卷积网络的整体训练思路相似，但是训练过程反过来了。反卷积深度学习网络，是一种自顶而下的解码过程，将通过滤波器组训练出的卷积特征组合、重构输入信号。在图像分解时，与稀疏编码采用矩阵乘积的形式不同，反卷积网络采用矩阵卷积的形式。而卷积神经网络是通过卷积、非线性激活函数和下采样实现的一种自底向上的编码过程，可以得到多层信息。

反卷积网络的核心是通过网络的方式对抽象的多维数据进行优化表达式，从而高效的提取出数据的特征。反卷积网络在训练过程中，和卷积神经网络一样，进行各层之间可以权值共享，能够从图像的边缘处、基本的几何结构曲线中获得到有很强结构的丰富特征。反卷积网络通过对稀疏的结构层学习，能够有效的提取结构特征，并且将特征可视化。目前反卷积网络主要广泛应用于判别式网络以及无监督的生成网络模型中，对于二维的图像，应用反卷积网络，得到包含图像中边缘信息以及物体对象的抽象表示的多种滤波器集合。

### 1.3.3双向深度网络基本概况

双向深度网络是由编码层和解码层相互叠加组成的。编码和解码有可能独立出现在深度网络的某一层中，也有可能在同一层中既有编码过程又有解码过程。

在经典的循环神经网络中，状态的传输是从前往后单向的，或者是从后向前的反卷积网络。但是，在现实应用中有些情况下，一个节点的输出和该节点的前后节点都有关系，此时双向神经更加适用。双向深度网络的方法将前馈网络和反馈网络相结合，既含前馈网络反向传播的特点, 又与反馈网络的预训练方法类似。一般情况下，双向深度网络的训练包含了单层网络的预训练以及每一层之间的反向迭代误差。

预训练：一般情况下，双向深度网络的预训练采用的是贪心算法。在第 层的单层中，输入的信号设为 ，权值为 ， 表示输入信号与权值计算得到信号 传递到下一层，此时信号再与权值 生成信号 作为输入。训练的目标是通过贪心的策略使得相差最小，也就是 。

反向迭代误差：与传统反向迭代方法一样，将训练数据集输入后经过隐藏层，最终到达输出层之后，和实际值之间计算误差，将误差从输入层向隐藏层反向传播，直至传播迭代到输入层。双向深度网络通过这种反向迭代误差的方法，对整个深度网络结构当中的权值进行微调。

## 1.4 本文所作的工作

本文结合深度学习中的卷积和反卷积过程，调研后选用U-Net模型，成功应用在了城市卫星图像识别当中，并且根据模型预测效果做成一个完整的网站服务。主要工作如下：

（1）在原有的基础上对U-Net网络模型进行了一些调整，建立模型对城市卫星图像中的一般建筑、工厂、道路、水体、裸地、阴影、运动场、建筑场地、乡村建筑这10类分别进行了训练。实现了对RGB三通道城市图像的像素级别识别。

（2）对于每种类别的预测效果，去除噪点，制定策略合并到一张图像中，确定物体边界、大小。

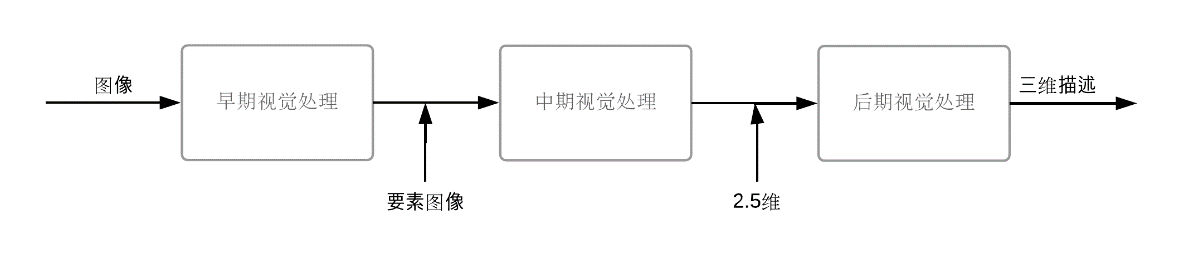
（3）开发平台，通过使用谷歌地图和谷歌静态地图API，完成在线选点预测、查看。

# 2 理论部分

## 2.1计算机视觉理论

### 2.1.1计算机视觉理论的发展

计算机视觉通常利用相机、摄像机等传感器，配合机器视觉算法赋予智能设备人眼的功能，从而进行物体的识别、检测、测量等功能，用机器代替人眼。计算机视觉从20世纪50年代起步，在光电子显微镜成像、识别简单字符等二维平面图像上初步应用。之后在1965年，L. R. Roberts博士提出了 “积木世界” 分析方法，通过计算机视觉系统从数字图像中提取出了诸如立方体、棱柱体，计算机视觉系统第一次将二维图形转化为三维结构，从几何边缘、角点的特征提取，到线条、曲面、平面等常见几何要素，计算机视觉在三维视觉场景下也有了一定的突破。发展到20世纪70年代，计算机视觉理论蓬勃发展，David Marr 教授在1977年提出了新的视觉理论——“Marr视觉理论” ，该理论立足计算机理论，使得计算机视觉研究有了比较明确的体系结构。



到20世纪80年代，研究者结合感知机理论的提出和发展，形成了基于感知机框架的计算视觉，开始出现了基于感知特征群的物体识别理论框架、主动视觉理论框架、视觉集成理论框架等概念，“Marr视觉理论”也得到了继续的完善、补充。

近几年来，计算性能的突飞猛进，信息技术与神经生理学等理论知识结合，计算机视觉与深度学习方法越来越密不可分。计算机视觉已经在工业、生活领域有大量应用，目前已经有大量的运行在手机、电脑的计算机视觉系统，计算机视觉的应用已经成为现代生活的一部分。

人类通过视觉对外部世界的认知，平均有超过80%的信息量，计算机视觉理论是希望通过算法模型，以成像系统取代视觉系统，计算机去取代人脑分析，来对外部世界进行模拟，来达到拓展、超越人类视觉的效果。

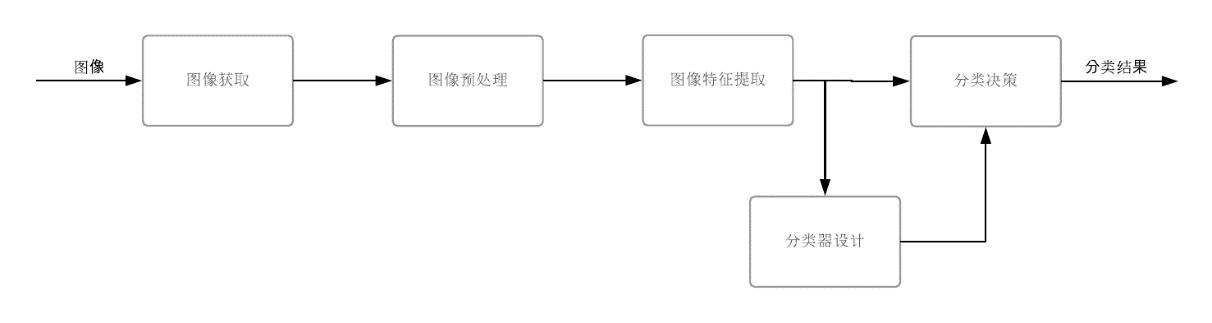
### 2.1.2计算机视觉理论的技术基础

计算机视觉侧重于处理得到的图片，在对图片中的内容进行识别、检测等。常常使用图像处理、计算机图形学、模式识别、神经网络等技术。

图像处理是通过数字图像处理方式，将图像信号转化为多维矩阵形式的数字信号，之后通过计算机算法模型再进行处理。图像处理可以实现图像复原、压缩编码减少传输图片大小、提高图片质量、图片补全和增强。图像中去除噪点的干扰、图像美化都属于是图像处理的范畴，本质是对多维矩阵形式的数字信号进行矩阵处理。

计算机图形学是一种使用数学算法将二维或三维图形转化为计算机显示器的栅格形式的科学，图形学的数学基础包括了拓扑学、曲面理论、集合论、分型几何学等。包含建模、渲染、动画、人机交互等方面。

模式识别发展于20世纪50年代末，图像模式识别将图形处理、分类器、以及分类决策的过程加以应用，实现自动分类出图像中目标的效果。含有模式识别的图像识别系统包括5大模块：图像获取、预处理、特征提取、分类器模型、分类决策。模式识别的图像识别系统具体的组成方式如下所示：



人工神经网络的研究源自对人脑的研究，神经网络是由多个神经元连接组成的，基于对单个神经元的研究提出了感知机模型，感知机模型分为单层感知机和多层感知机模型。单层感知机可以视为含有单一神经元的超平面模型，使用阶跃函数作为单层感知机的传递函数。超平面的感知机只能实现对数据集的线性可分，无法处理线性不可分的数据集。

## 2.2卷积神经网络CNN

F. Rosenblatt提出的感知机是最简单的单层前向人工神经网络，但是无法解决线性不可分问题。1984年日本学者K. Fukushima等基于感受野的概念，提出的神经认知机可看作卷积神经网络的一种特例, Y. Lecun等提出的卷积神经网络是神经认知机的推广形式。

卷积神经网络是逐步兴起的一种前馈人工神经网络结构，因为利用卷积神经网络在图像和语音识别方面能够给出更优预测结果, 这一种技术也被广泛的传播可应用。卷积神经网络的特点在于,采用原始信号（一般为图像）直接作为网络的输入，避免了传统识别算法中复杂的特征提取和图像重建过程。

卷积神经网络的核心思想主要有局部感受野、权值共享和池化技术。

1、局部感受野（Receptive Field）：在机器视觉领域的深度神经网络中有一个概念叫做感受野，用来表示网络内部的不同位置的神经元对原图像的感受范围的大小。普通的多层感知器中，隐层节点会全连接到一个图像的每个像素点上；而在卷积神经网络中，每个隐层节点只连接到图像某个足够小局部的像素点上，从而大大减少需要训练的权值参数。

2、权值共享：在卷积神经网中，同一个卷积核内，所有的神经元的权值是相同的，从而大大减少需要训练的参数。

3、池化：在卷积神经网络中，没有必要一定就要对原图像做处理，而是可以使用某种“压缩”方法，这就是池化，也就是每次将原图像卷积后，都通过一个下采样的过程，来减小图像的规模。

## 2.3全卷积网络FCN的基本概况

UC Berkeley的Jonathan Long等人提出了端到端的Fully Convolutional Networks，用于图像的语义分割，而语义分割的目标就是预测每个像素点的语义标签，该网络目的在于通过对特征的处理，找到每个像素对应的类别，最终实现像素级别的分类效果。

FCN是一个端到端，点对点的全卷积网络，与经典的CNN在卷积层之后使用全连接层得到固定长度的特征向量进行分类不同，FCN可以接受任意尺寸的输入图像。FCN中最主要的三个特点分别是：全卷积化、上采样和跳跃结构。

FCN的整体流程：采用反卷积层对最后一个卷积层的feature map进行上采样, 使它恢复到输入图像相同的尺寸，从而可以对每个像素都产生了一个预测, 同时保留了原始输入图像中的空间信息, 最后在上采样的特征图上进行逐像素分类。最后逐个像素计算它对应的softmax分类的损失, 相当于每一个像素对应一个训练样本。

FCN的优点是相对传统的CNN方法，FCN节省了存储开销，提高了计算效率，解决了像素块大小对感知区域的限制。通常像素块的大小比整幅图像小很多，只能提取一些局部的特征，从而导致分类的性能受到限制。而全卷积网络(FCN)则是从抽象的特征中恢复出每个像素所属的类别。

但是FCN得到的结果还是不够精细。进行8倍上采样虽然比32倍的效果好了很多，但是上采样的结果还是比较模糊和平滑，对图像中的细节不敏感。此外，由于FCN是对各个像素进行分类，没有充分考虑像素与像素之间的关系。忽略了在通常的基于像素分类的分割方法中使用的空间规整（spatial regularization）步骤，缺乏空间一致性。

## 2.4 U-Net卷积网络的基本概况

U-Net属于一种编码器结构，该网络在EM Stacks 的图像数据集中获得了良好的效果，此数据集包含了30个密集注释的医学图像和其他医学图像数据集。虽然U-Net最初是Ronneberger O, Fischer P, Brox T等为了解决生物医学分割问题而提出来的，但是目前U-Net网络展示出了它从少量的数据中学习的能力，同时它已经在诸如卫星图像分割等领域有了很好的应用。

U-Net网络中编码器逐渐减少池化层的空间维度，解码器逐步修复物体的细节和空间维度。在U-Net网络中，卷积层的数量大约在20个左右，4次下采样，4次上采样。U-Net运用了与FCN相同的技巧，将浅层特征图与深层特征图结合，这样可以结合局部 特征以及全局特征，生成更精准的图像。U-Net并不像FCN将特征相加，而是连接生成双倍通道的特征图，再卷积操作。并且采用重叠采样的方式，将图像边缘进行一定的镜像复制生成边缘图像，提高了边缘识别效果。

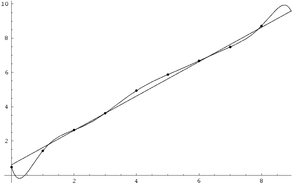


## 2.5 过拟合问题及其常用解决方案

### 2.5.1过拟合问题定义

给定一个假设空间，一个假设属于，如果存在其他的假设属于，使得在训练样例上的错误率比小，但在整个实例分布上比 的错误率小，那么就说假设过度拟合训练数据。过拟合通常发生在模型参数过于复杂时，参数过多导致模型的数据波动性大，在真实数据集上表现的预测能力降低。

过拟合问题在直观上表现为模型在训练数据上表现的很好，找到所有数据的特点，几乎找到了所有点，但是对于未知的数据通常不能预测出来结果。如下图所示，原始数据分布比较符合一元一次的线性函数，而且用线性函数表示也具有良好的泛化能力。如果像图中使用多项式函数表示，虽然将所有的点都拟合在了曲线上，但是不具有普适性。



### 2.5.2过拟合出现的原因

在机器学习的训练过程中，过拟合问题的出现常常是由于以下问题所导致的。

1. 样本抽取问题，如训练样布太少，样本抽样过程中没有对使用场景或者特点进行充分考虑，或者是使用的抽样方法不科学。
2. 噪音数据对样本的干扰过大，导致训练的模型学习到了噪音特征，而忽视了输入数据与输出数据之间的关系。
3. 假设条件在模型具体应用时不再成立。历史和未来数据的假设不成立、建模数据与实际应用的数据不是同一分布的，就违反了模型使用场景。
4. 模型太过复杂，训练参数过多。
5. 神经网络这种迭代模型在训练时迭代次数过多，拟合了所有的噪音数据和没有代表性的训练样例。

### 2.5.3过拟合常用解决方案

1. 在网络模型中衰减权值。在迭代过程中，增加一个和网络的权值的总数相应的惩罚项，保持权值逐渐减小，使得学习的过程与决策面的方向相反。
2. 适时地停止迭代训练，使得模型学习到训练数据的主要特征即可。
3. 验证数据，在训练的数据之外再提供一组验证的数据，最终使用在验证集合上误差最小的模型。
4. 添加dropout 层。具体过程是指在神经网络的训练过程中，对于所有的神经网络单元，按照某一概率将其暂时从网络中丢弃。

## 2.6 权值初始化

### 2.6.1权值初始化的意义

权值初始化在深度学习中十分重要。随着网络层次变深，太小的参数会存在梯度消失的问题，此时深层的网络参数很难得到有效训练，此时，某些卷积层可能会有特别多的激活，而另一些卷积层可能对网络没有贡献。，而太大的权值可能会导致模型收敛到较差的局部最优点或者还可能导致梯度爆炸。因此，权值的初始化好坏程度会影响最终模型效果以及训练时间。理想的权值初始化是使得网络中的每一个特征图方差都接近1。

### 2.6.2权值初始化的常见方法

权值初始化的方法主要有：gaussian Initialization方法、Glorot Initialization方法、 He Initialization方法、Positive\_unit\_ball Initialization等方法。

gaussian Initialization方法：在U-Net中，可以通过高斯分布随机生成权值，高斯分布的均方差（2/N）^0.5，其中N为上一层的参数数量。

Glorot Initialization：这种初始化方法是由Xavier Glorot和Yoshua Bengio于2010年提出的初始化方法，这种方法在神经网络中十分适用。为了综合考虑输入与输出，Glorot Initialization对正向与反向过程提取调和平均数。最终的权值分布满足：



He Initialization初始化：2015年微软亚洲研究院何凯明等人于正式发表He Initialization 初始化方法，与Glorot方法不同，He Initialization 侧重单向的前进过程，对于均匀分布的he\_uniform 权值分布为：



Positive\_unit\_ball初始化：是在神经网络中，让每一个神经元的输入的权值之和为 1，可以有助于防止权值初始化过大，从而防止激活函数（sigmoid函数）进入饱和区。

## 2.7 多类别分类函数softmax

本课题在整合多个模型预测到一张图像中时，考虑使用softmax 作为最后一层的多分类器。softmax 将二分类拓展到了多分类，输出的结果是分类器对每一个类别的估算概率值，softmax 通过归一化保证了概率和为1。样本属于各个标签的概率计算方法如下所示：



其中代表模型中的全部参数。

## 2.8 样本选择

### 2.8.1样本选择的概念

在机器学习模型的训练当中样本对最终模型的好坏非常关键，样本的信息是否能够代表一般事实情况，标注样本是否正确关系着模型效果以及模型的训练时间。在机器学习的早期阶段，研究人员常常使用人工制作的数据并混杂些许噪音数据的方法来构建数据集，来解决样本不充足的问题。随着技术水平的不断进步，人们建立了很多大数据样本库，比如知名的存储了由250多个人手写的包含60000多个样本的MNIST 标准手写样本库，以及ImageNet 提供的包含22000多个物体的图像数据集。随着数据规模的增加，机器学习模型的性能带来了一定的提升，但是大数据也带来了一定的问题。这些大规模数据集中常常有冗余的样本，消耗模型训练时间，带来“过学习”的问题。所以，我们通常使用一些抽样方法，从中有效的减少训练数据规模，提高学习效率。

### 2.8.2常用抽样方法

1. 简单抽样

简单抽样可以分为放回抽样和不放回抽样。放回抽样是从全体数据集中每次抽取其中一个，抽取之后将其放回，不放回抽样是从全体数据集中每次抽取其中一个，抽取之后不放回，样本不会被二次抽到。

1. 簇抽样

将全体数据集划分成多个不相交的簇，抽样时选取一个或多个簇。簇抽样时，要求簇之间的差异小，簇内数据差异较大从而更好的反应分布特征。

1. 分层抽样

对全体数据划分成多个不相交的层，之后按照一定比例从每个层中进行抽样，与簇抽样相反，分层抽样要求层之间的数据差异大，层之间数据差异小。

1. 系统抽样

系统抽样通常用于全体数据量过大的时候，从将整体样本集N 分为段，每段长度为，设定一个整数， ，在每段中抽取第个，抽样结束之后得到个样本。

1. 自助抽样

在自助抽样中，假设全体数据集含有N 条数据，每次从整体中抽取一条样本，有放回连续抽取样本N 次，此时一部分被抽中0次，一部分至少抽中1次。这样原始数据中有些会出现多次进入训练集，没有被抽中的样本形成验证集。一次没被抽中的概率为 ，当样本很多时，趋近为0.368。

# 3 实验内容部分

## 3.1 研究内容以及课题目标

### 3.1.1 研究背景

同济大学城市规划学院研究发现，卫星图像可以确定一个城市的城市建成区、功能区等信息，结合一些统计数据，可以得到一个城市的发展变化，从而在市政规划工作的时候提出一些预见性建议。本课题希望通过深度学习的方式对城市卫星图像进行解释，来确定城市区域以及发展状况，帮助专家学者进行分析预测。

### 3.1.2 课题目标

本课题拟选取U-Net卷积网络法作为核心深度学习的技术路线，使用长江三角洲附近的谷歌地图数据，选取一部分进行标注构成标注数据集。之后，通过分割、分类的方法对高分辨率的卫星图像进行识别，通过算法识别，判断出城市建成区，进而可以更进一步的找到城市功能区，对U-Net网络结构进行调整，使得该模型在本课题的具体问题中有更好的表现。

## 3.2 数据预处理

### 3.2.1 图像预处理

图像预处理主要研究的内容有以下几个方面，1）各种图像变换的方法，如傅立叶变换、沃尔什变换、离散余弦变换等间接处理技术，将空间域的处理转换为变换域处理，减少计算量，获得更有效的处理。2)图像编码压缩图像编码压缩技术可减少描述图像的数据量。3)图像增强和复原图像增强和复原的目的是为了提高图像的质量，如去除噪声，提高图像的清晰度等。图像增强不考虑图像降质的原因，突出图像中所感兴趣的部分。如强化图像高频分量，可使图像中物体轮廓清晰，细节明显；如强化低频分量可减少图像中噪声影响。图像复原要求对图像降质的原因有一定的了解，一般讲应根据降质过程建立"降质模型"，再采用某种滤波方法，恢复或重建原来的图像。4)图像分割是将图像中有意义的特征部分提取出来，其有意义的特征有图像中的边缘、区域等，这是进一步进行图像识别、分析和理解的基础。

数据预处理，可以说是数据挖掘过程中的一个最重要的步骤。数据预处理常用的方法有去均值、归一化、PAC降维处理，重叠采集等。

1、去均值法：在计算机视觉中，深度学习框架输入图片后一般都会进行均值削减，对所有训练样本图像求均值，然后将每个样本图片减去该均值。测试图片在进行预处理时，也减去该均值

2、归一化处理：归一化主要是为了数据处理方便提出来的，把数据映射到0～1范围之内处理，更加便捷快速。在本课题中，因为RGB通道数据范围一直是0到255，归一化只需要将像素值除以255，因此不需要再进行归一化处理。

3、PCA降维处理：Principal Component Analysis(PCA)，是最常用的线性降维方法，它的目标是通过某种线性投影，将高维的数据映射到低维的空间中表示，并期望在所投影的维度上数据的方差最大，以此使用较少的数据维度，同时保留住较多的原数据点的特性。数据量较大时可以通过PCA进行降维处理，以便减小计算量，提升处理速度。

### 3.2.2 数据采集筛选

数据的筛选对于机器学习模型来说是一个非常重要的过程，有时候是数据维度上的筛选来减少计算难度，有时候是去除掉一些不合适的数据或者多采集一些。本课题所用到的数据采集筛选策略主要是重叠采集和去除空样本数据。

重叠采集：图像分割将多尺度图像块与滑动窗口生成方法相结合，利用区域块重叠技术覆盖到整个图像的边缘。对任意大小的卫星图像，使用重叠采样策略Overlap-tile strategy，将图像边缘进行一定的镜像复制生成边缘图像。最终得到统一大小的RGB三通道的卫星图像。

去除空样本数据：在切割后的卫星图像片中，对于某一类别来说，该类别所对应的像素可能在切分后的小切片中不存在，则将该切片从这一类别中去除出去。通过这样的方法，使得对于某一类的训练时，所有的训练图像中都至少含有一块该类别图像对应数据，训练更有效。

## 3.3 评估指标

本课题目标是识别出卫星图像中的多种类别，实现像素级别的标注，所以模型训练的目标是希望预测出的图像与真实类别图像相似度越高越好。所以考虑使用Jaccard系数。Jaccard系数的计算十分的简单高效，公式如下所示：

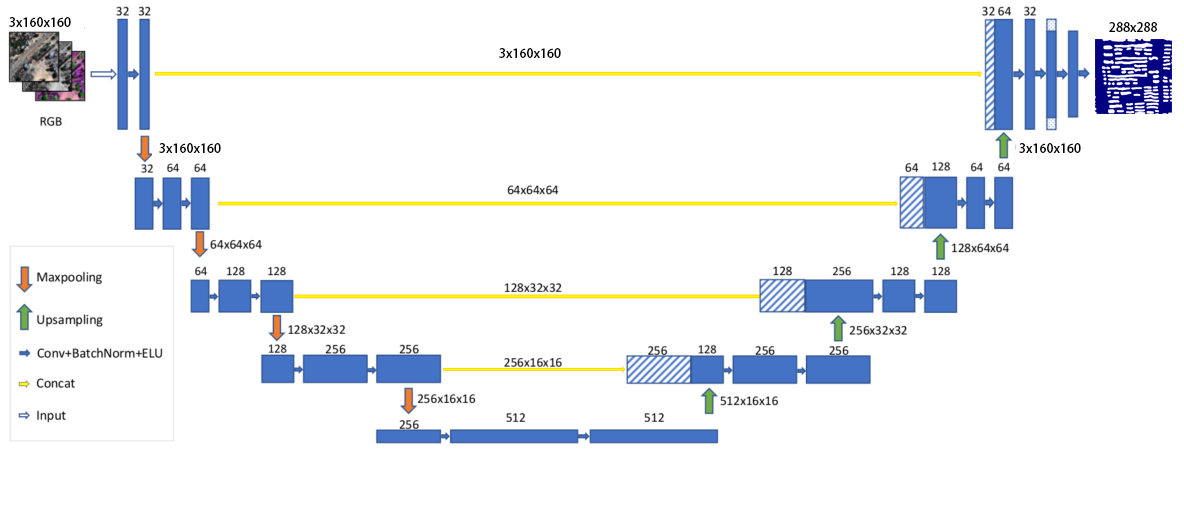


 和 分别代表预测图像的类别集合和真实图像的类别集合， 和交集越大，并集越小，则表示两个集合的相似度越高，也就是预测结果与真实结果越接近。

## 3.3 建立网络模型

在基础的U-Net 网络结构上，本课题做了一些调整。过程中所使用到的相关网络层如下所述。

### 3.3.1 模型整体结构



模型将卷积和反卷积的过程结合起来，属于一种基于像素的编码器网络结构。在网络中编码器逐渐减少池化层的空间维度，解码器逐步修复物体的细节和空间维度。在最终选定的网络结构中，包含了4次下采样，4次上采样，并且还包含了4次连接操作，把编码数据与解码数据结合起来使用，更有利于找到图像特征。

### 3.3.2 卷积层

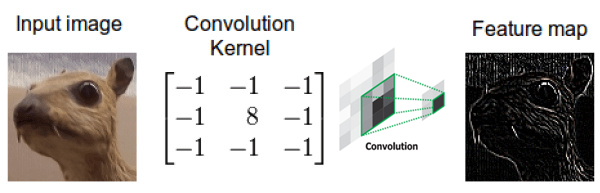
卷积层的目的主要是用来对特征进行压缩、提取的。卷积层的输入数据大多数来自输入图像或者上层的卷积操作。卷积层当中的所有特征图都对应着尺寸相同的卷积核，卷积核的初始化方式有很多种，本课题采用的是he\_uniform 的初始化方法对权值进行初始化的。在卷积层中，每个特征图是对应元素的卷积结果累加之后，加上一个偏置项后，再由特定的激活函数如Relu 激活。

卷积层中具体的特征图的计算公式如下：



 为卷积层 的第 个特征图，输入特征集合表示为 ，对应的偏置项为， 为对应权值， 为卷积操作。

卷积操作是卷积层的核心，可以视作是一种信息混合的方式，本质是一种数学运算，和加减乘除类似，如下图示例，左侧第一张动物图片是输入数据，由RGB三通道构成的矩阵组成，左侧第二张图片为卷积核，经过第三个图片所示的卷积操作，生成的



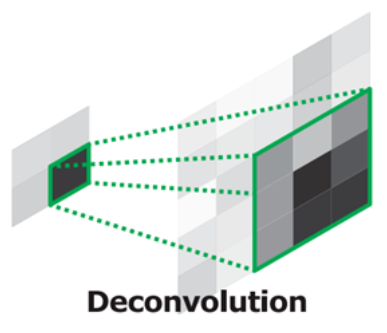
### 3.3.3 批归一化处理Batch Normalization

### 3.3.4 激活函数

### 3.3.5 池化层

### 3.3.6 Dropout 层

### 3.3.7 反卷积层



### 3.3.8 连接层

### 3.3.9 剪切层

## 3.4 后处理过程

后处理（post processing）对于图像识别是非常重要的一环，在训练的模型很难进一步提高精度之后，后处理能通过定义一定规则在一定程度上能够减少误差。

### 3.4.1 去噪点

U-Net 模型训练完成之后，对城市卫星图像预测的结果中，往往产生一些孤立的像素点，或者很小范围的像素簇，因为本课题预测的一般建筑、植被等这种较大对象，所以这个可以认为是噪点。这里使用了OpenCV 工具库的RemoveSmallRegion 方法，删除面积小的区域，将连通区域小于固定值的像素去掉，减少噪音。

### 3.4.2 确定图像边界

为了更加清楚的观察标注类别对象的边界，通过使用第三方skimage 库的find\_contours 方法，来检查二值图像的边缘轮廓，输入模型训练好之后，去除噪点的图像，返回轮廓列表的集合，可以用for 循环得到每一条轮廓，再在地图上展示出来。

## 3.5 实验步骤

U-Net模型能够对输入图像的每个像素做出该点属于目标类的概率，使用Jaccard指数作为模型的评估指标。Jaccard相似指数用来度量两个集合之间的相似性，它被定义为两个集合交集的元素个数除以并集的元素个数，Jaccard距离用来度量两个集合之间的差异性。训练的总目标是以最小化所有像素的二进制交叉熵总和为训练目标。

首先对数据集通过预先计算得到数据的统计值，将图像集归一化为具有零均值和单位方差的数据集。根据欧盟功能区地图、美国土地使用情况地图，对图像按照实际类别进行标注。再将预处理的图像保持不变，或者将图像及相应标签共同调整为一系列正方形图像，以便网络训练使用。对于数据不均衡的问题，可以通过欠采样、过采样、调整数据集的权重等方式解决。对于边界问题，模型采取了扩大每次选取的面积、并且随机进行旋转操作解决。

训练算法的目标就是通过调整函数的权重和偏置来最小化代价函数，也就是最小化所有像素的二进制交叉熵总和∆C，考虑通过梯度下降的方式进行网络训练. 梯度下降算法工作的方式就是重复计算梯度，然后沿着梯度的反方向调整网络中各个节点的参数，以达到找到全局最小值的目标。

对分类目标的10种类别分别建立如上的模型，再通过后处理去除掉图像中噪音，最后将多个结果按照一定次序叠加，最终实现标注出卫星图像中10种类别的效果

# 4 结果和讨论

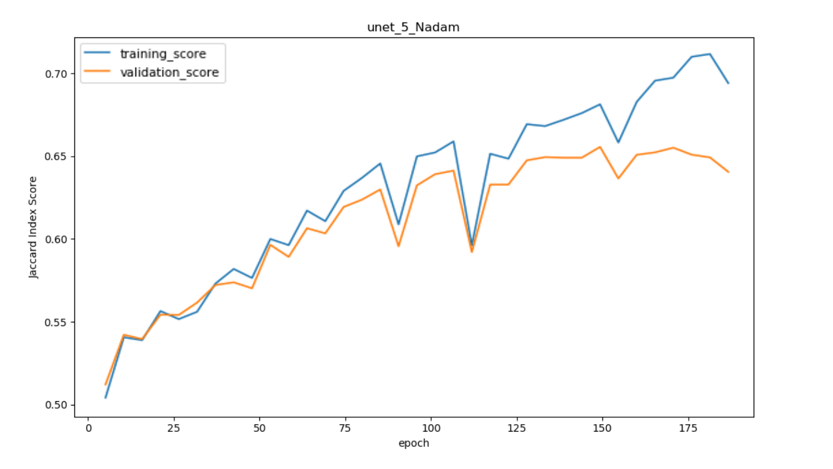
## 4.1模型训练结果和讨论

### 4.1.1深度为5的U-Net网络

以一般建筑的模型效果为例，损失率、正确率、验证集损失率、验证集正确率如下列表所示。

表4.4 一般建筑5层模型部分数据列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Epoch* | loss | accuracy | val\_loss | val\_acc |
| …… | …… | …… | …… | …… |
| 5 | 0.2375 | 0.5113 | 0.2327 | 0.5131 |
| 25 | 0.2159 | 0.5522 | 0.1984 | 0.5547 |
| 45 | 0.1708 | 0.5646 | 0.1859 | 0.5601 |
| 65 | 0.1527 | 0.6013 | 0.1784 | 0.6042 |
| 85 | 0.1154 | 0.6449 | 0.1725 | 0.6413 |
| 105 | 0.1075 | 0.6512 | 0.1727 | 0.6469 |
| 125 | 0.1311 | 0.6505 | 0.1627 | 0.6553 |
| 145 | 0.1471 | 0.6640 | 0.1227 | 0.6511 |
| 165 | 0.0714 | 0.6981 | 0.2344 | 0.6502 |
| 185 | 0.1267 | 0.7213 | 0.1867 | 0.6431 |



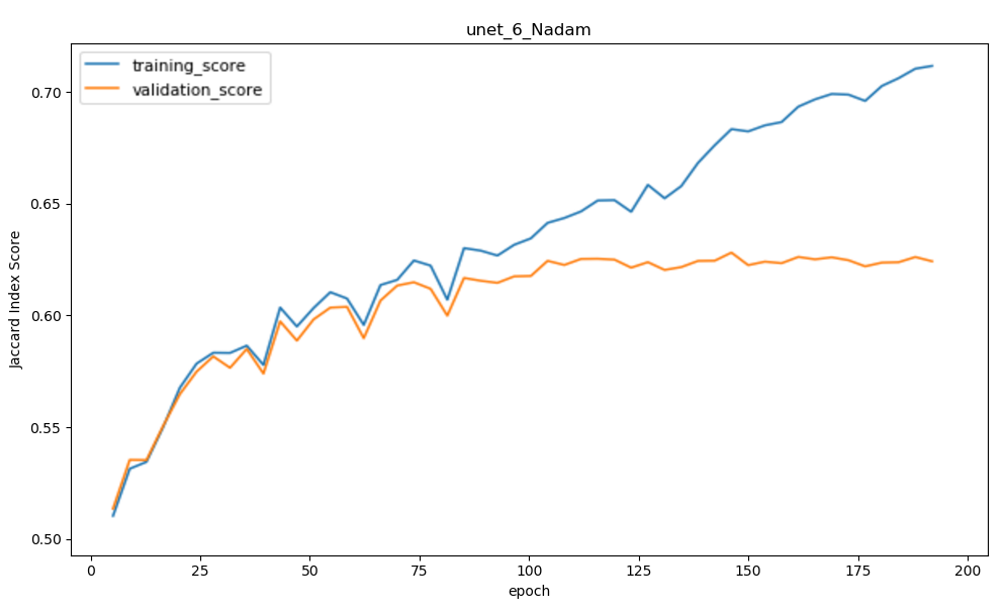
可以看到模型的收敛速度还是比较快的，当epoch 为100左右时，模型在验证集上的效果基本就稳定在65%左右。

### 4.1.2深度为6的U-Net网络

继续以一般建筑的模型效果为例，损失率、正确率、验证集损失率、验证集正确率如下列表所示。

表4.4 一般建筑6层模型部分数据列表

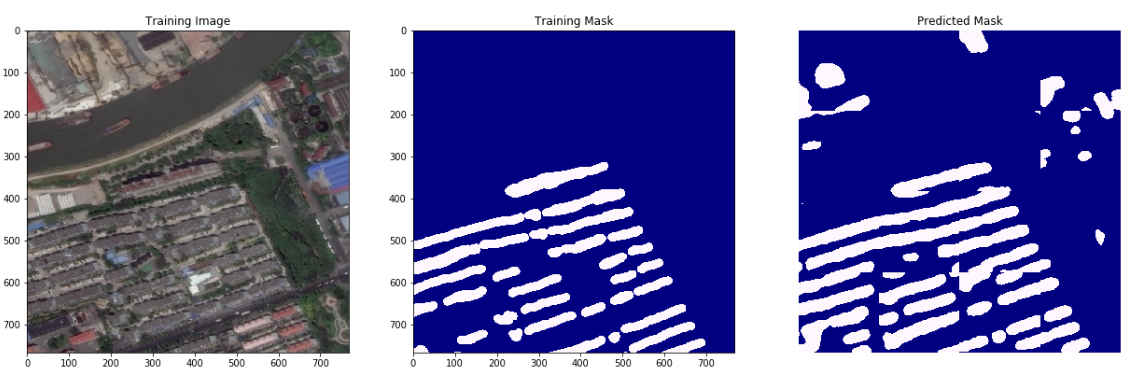
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Epoch* | loss | accuracy | val\_loss | val\_acc |
| …… | …… | …… | …… | …… |
| 5 | 0.2215 | 0.5381 | 0.2371 | 0.5342 |
| 25 | 0.2319 | 0.5897 | 0.1741 | 0.5784 |
| 45 | 0.1218 | 0.5973 | 0.1729 | 0.5982 |
| 65 | 0.1517 | 0.5813 | 0.1124 | 0.5804 |
| 85 | 0.1314 | 0.6056 | 0.1715 | 0.5840 |
| 105 | 0.1139 | 0.6153 | 0.1182 | 0.6182 |
| 125 | 0.1381 | 0.6297 | 0.1621 | 0.6201 |
| 145 | 0.1311 | 0.6812 | 0.1321 | 0.6221 |
| 165 | 0.0914 | 0.7031 | 0.1917 | 0.6238 |
| 185 | 0.1017 | 0.7203 | 0.1761 | 0.6213 |

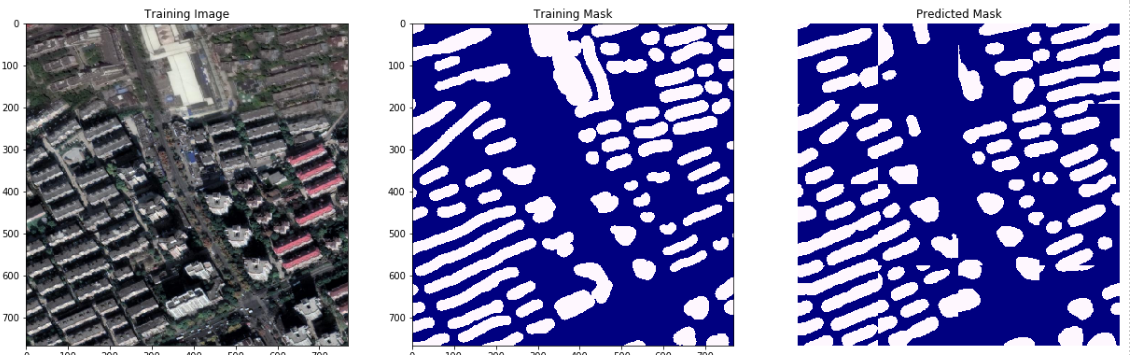


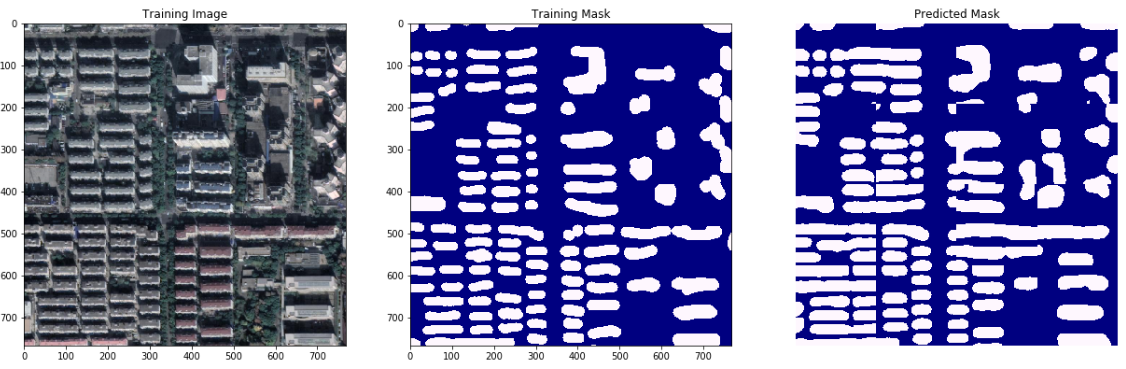
6层网络模型在epoch 为90左右之后，虽然在训练数据集上的Jaccard 系数还在增大，但是在验证数据集上几乎没有什么变化了，趋于平稳。与5层模型相比，6层模型在此数据集上效果并没有五层的U-Net 训练效果好。可见单纯的增加网络层数并不总是能够提高精度。从预测准确度和训练速度角度考虑，最终本课题选用了5层模型。

### 4.1.3一般建筑预测结果

选用效果较好的5层模型，迭代200次（epoch=200），选取验证数据集的数据进行测试，

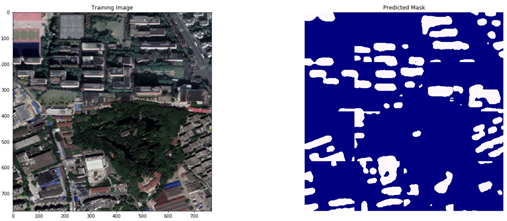






在一般建筑这一类别上，模型预测的正确率为65%左右，但是整体投射到图像中时，几乎覆盖出了图像中一般建筑物。这种差距是因为模型是基于像素级别做训练、预测的，数据标注的时候很难完整的标注的完整的一般建筑对象，此外还存在一些漏标的地方。如第一幅图所示，预测结果图像上半部分预测出了标注数据所遗漏的一般建筑。

但是模型目前也存在一些不足之处：由于是将图像切分后进行预测的，尽管采取了一些扩大采集、镜像、旋转等方式，还是有一些如下图所示的较为明显的分割痕迹。这个问题还可以今后通过提高机器显存空间，减少切割来解决。



## 4.2 多种城市用地预测结果和讨论

### 4.2.1 合并单一模型

通过在模型的最后的输出层添加一层softmax 多类分类器，对于每一个像素位置预测产生一组属于10个类别的评分，最后再归一化。因为标记样本中的类别之间数目相差较大，比如城市卫星图像中绿地的占地面积往往很大，而工厂、农村建筑出现的面积就小很多，所以不能直接使用概率，这样会导致小类别的饥饿现象。本课题目标是尽可能的保证概率分值小的也会取到，softmax 正好能解决这个问题。对于数组，表示中的第个元素，该元素对应的softmax 值计算为：。

使用softmax 的方式对多个类别进行合并，在模型训练时无法收敛，正确率也很低，输出得到结果的样例，如下所示，可见这种方法对于RGB 三通道的城市卫星图像数据效果并不好，杂乱无章。





### 4.2.2 二分类多模型

在二分类多模型方法中，对目标的一般建筑、工厂、道路、水体、裸地、阴影、运动场、建筑场地、乡村建筑这10类，分别训练出10个二分类模型。再按照模型效果，对这10类进行叠加融合。

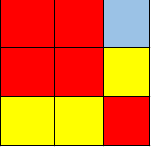
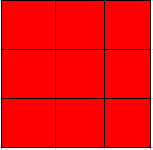
整体的思路是将大面积粗精度的叠放在底层，将较小面积切精度高的叠放在上层。观察各个模型结果，最终的叠放次序从底向上依次为树木、建筑场地、裸地、阴影、水体、运动场、工厂、乡村建筑、一般建筑、道路。

最终合并效果如下图所示：黄色表示一般建筑，蓝色表示水体，紫色为运动场，橙色为裸地，粉紫色为阴影，白色为建筑场地，灰色为道路，绿色为植被覆盖区，红色为乡村建筑。





由于手工标注存在偏差，标注值与预测可能存在一定的位移，所以选取大小的栅格，将栅格中出现频次最高的类别判定为栅格的类别，最终研究栅格级别的相似度。



在验证的长江三角洲地区数据集中，计算得到的平均栅格正确率为65.71%，基本达到预期标准。

### 4.2.3 预测结果讨论

通过以上两种方式的对比，本课题最终采用的方式是第二种多模型分开训练，在人工定义规则合并的策略。虽然速度上比softmax 多分类的方式慢10倍左右，但是精度提升很大，而且今后如果增加了某一个新的类别，只需要针对新类别训练即可，原来的模型们不受影响，有良好的拓展性。

## 4.3 在线标注平台开发

基于上面建立的模型，本课题搭建了一个在线城市用地标注系统，方面用于测试模型效果以及展示成果。

# 5 结论和展望

## 5.1 结论

## 5.2 展望

# 参考文献

1. 李炳穆.理想的图书馆员和信息专家的素质与形象[J].图书情报工作，2000(2):5-8.
2. Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 2528-2535.

# 谢 辞

正文内容