

计算机视觉
期末大作业实验报告

卫星云层图像的理解与识别

学院：软件学院

姓名：公岩松 学号：2120190505

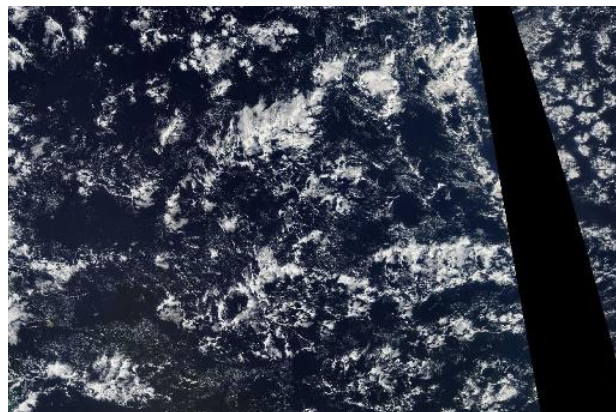
姓名：龚士博 学号：2120190506

姓名：甘润东 学号：2120190511

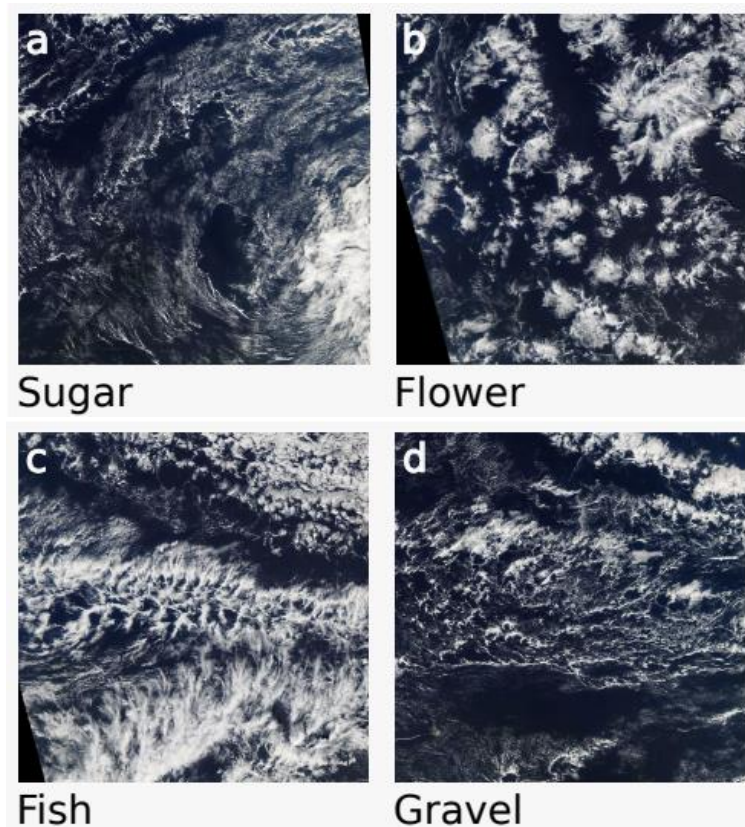
一、问题描述

在分析地球的气候时，浅云起着巨大的作用。它们也很难理解，无法在气候模型中表示。本次实验的目标是改善对卫星云层图像的理解，从而帮助建立更优秀可靠的气候模型。云可以通过多种方式进行组织，但是不同组织形式之间的分界线是极其模糊的。这使得构建传统的基于规则的算法来分离不同种类的云具有很大的挑战性。

本次实验中，数据图像来自于 NASA Worldview，是两颗极地轨道卫星 TERRA 和 AQUA 拍摄的，由于这些卫星上成像器较小，因此两轨道卫星的图像可能会缝合在一起，在缝合时会有一些不可见部分，这部分标记为黑色。下面是一幅卫星云层图像的示例：

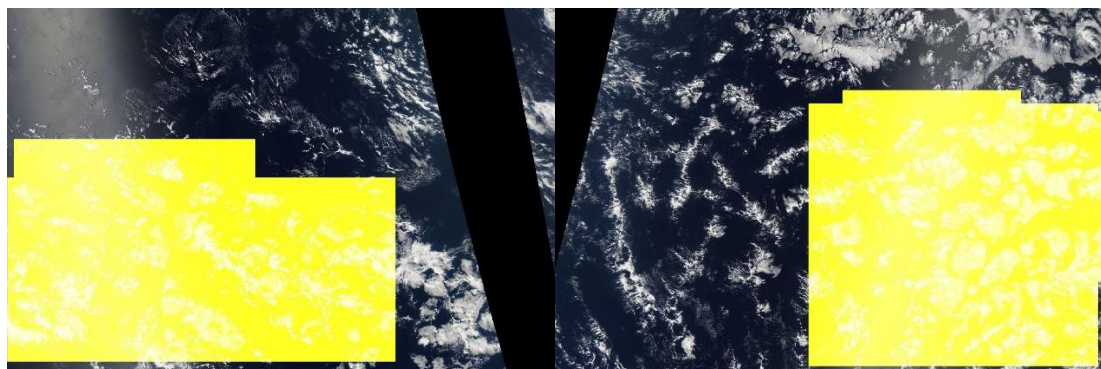


图像集中所有的图像都标有标签，这些标签是在德国汉堡的马克斯-普朗克气象研究所和法国巴黎动力实验室创建的。由 68 位科学家组成的团队在每帧图像中识别出具有特定类型的云层区域，其中云层包含以下四种类别：



其中 Sugar 类型的云形状类似糖粉，几乎没有组织成团。Flower 类型的云，形状类似花束，有大量云汇聚成团，且彼此间具有间隔。Fish 类型的云，形状类似鱼骨，规模较大。Gravel 类型的云，形状类似碎石，具有中等粒度的、随机相互关联的细小云团。

在一帧卫星云层图片中，可能会出现多种类型的云，建立模型的目的是将这些类型的云从图像中分割出来，并且正确预测其类别。下图展示了人工标记的云层分类结果：



其中，黄色透明区域的云层人工标记为 Flower 类型的云层区域，可以看到人

工标记的区域较为合理。由于具有视觉感官和专业知识，人工标记的结果基本符合客观事实。所以建立模型需要根据人工标记的结果进行学习和训练，训练好模型后对云层图像进行分割和预测分类。一般认为，与人工标记区域的“重合度”越高，模型的预测能力越优秀，量化结果采用 Dice 系数进行评价，其公式如下：

$$\text{Dice}(X,Y) = \frac{2 * |X \cap Y|}{|X| + |Y|}$$

其中，X 表示模型预测的像素集合，Y 表示人工标记的像素集合。Dice 系数取值在 0 至 1 之间，且数值越大，证明预测的像素集合与人工标记的像素集合重合度越高，预测模型越准确。特殊地，当 X,Y 几何均为空集时，Dice 系数被设定为 1。

二、问题现状及现阶段主要方法

卫星云层图像理解与分类问题本质上还是图像分割与分类问题，在这个问题上已经有大量实验证明，在数据量足够充足、模型实时性不做要求的情况下，深度网络模型的分割预测结果通常比传统方法建立的模型更加准确和有效。可能基于此原因，现阶段针对这个问题的模型几乎都建立在深度网络模型之上。

目前在 Kaggle 竞赛中，已经有非常多的团队对这个问题进行了研究与探索，并且建立了有效的模型。其中，以 U-Net 框架为基础的模型在这个问题中获得了重要的应用，也取得了十分优秀的结果。

U-Net 发表于 2015 年，属于全卷积网络的一种变体。U-Net 的初衷是为了解决生物医学图像方面的问题，由于效果很好后来也被广泛的应用在语义分割的各个方向，比如卫星图像分割，工业瑕疵检测等。U-Net 框架是 Encoder-Decoder 结构，其中 Encoder 负责特征提取，并且应用了图像增强的方法，在数据集有限

的情况下也可以获得不错的精度。提取的特征经过 Decoder 恢复原始分辨率，生成最终的分割与预测结果。

Kaggle 竞赛中在 Public Leaderboard 中排名最高的方法便是在 U-Net 框架基础上进行了改进与提升，最终 Dice 系数取得了 0.68185 的结果。

三、算法主体思路

本文根据前文中的问题现状和现有研究，决定利用两种不同的方式建立模型，一种是在现有研究的基础上，基于 U-Net 网络框架的改进方法。另一种是基于传统方法对图像进行处理、分割、预测的探索。

其中针对前者，主体工作大概包括两部分：首先是对 Encoder 方法的探索，选取不同的 Encoder 方法对于最终的分割预测结果有着较大影响，通过实验和分析，找到了针对这个问题较为优秀的 Encoder 方法。其次，由于数据量不够丰富，实验利用多种图像增强的方式，降低了这一因素对于实验结果的负面影响。

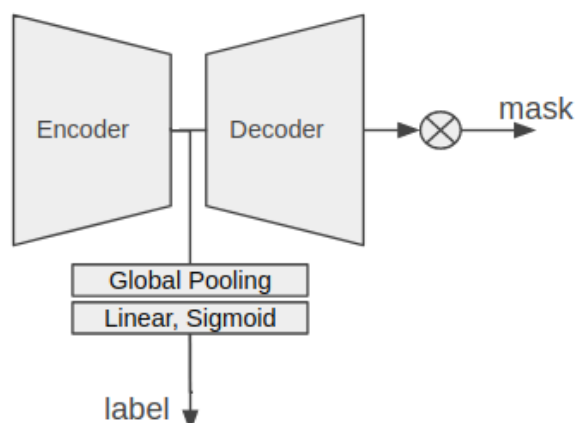
针对后者，主体工作是提出了一套完整的、基于传统方法的分割-预测模型，该模型将图像划分为多尺度窗口，并且基于词袋模型，对每个窗口进行特征提取，利用该特征对分类器进行训练，同时也根据此特征对测试样本的窗口进行预测和分类。最终通过多尺度窗口的融合生成最后的预测像素集合，与人工标注像素集合对比计算 Dice 系数。

四、算法描述与实现

1. 基于 U-Net 框架的网络模型

基于 U-Net 框架的网络模型大致分为两个部分：编码部分和解码分。

其大致结构如下：



首先使用多种 CNN 网络对图像进行特征提取从而获取对应图像的特征向量。实验中使用了三种不同的 CNN 网络模型作为该框架下的编码器(分别为 EfficientNet, ResNet34 和 Resnext101)。

获得的特征向量分别进行两种处理，分别负责对图片在四种云下的四个二分类问题以及像素级分类问题。

因为卫星云图中云的类别是未知，必须对每种云进行一个是否存在的判断，之后根据存在的种类使用 U-Net 的解码器进行独立的像素级分割。

具体的实验步骤如下：

- 1) 为了优化网络模型的训练过程，我们对训练集数据集进行图像强化以丰富其数据丰富度并拖过缩小图像大小以减少模型训练的时间代价；
- 2) 选取多种不同的编码器，分别进行 15 个完整批次，11 组交叉验证的训练；
- 3) 根据不同的编码器在此框架中生成的结果在官网上的数值表现对其进行调整(增加训练迭代次数，提升完整训练批次到 20，可能此种编

码器模型在 15 批次并没有完全收敛)。如果数值结果并没有显著的提升, 则更换编码器, 继续迭代训练。

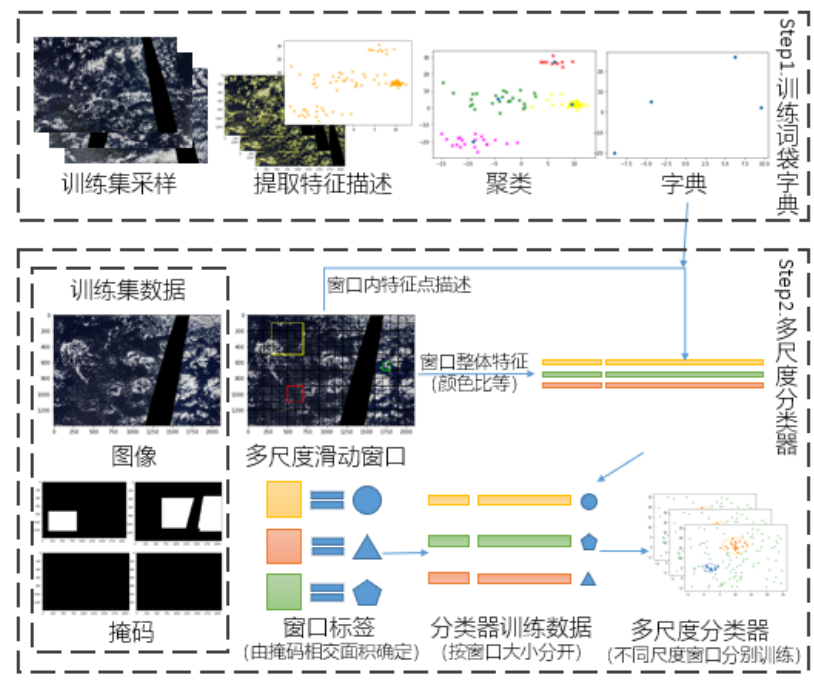
2, 基于传统方法的分割-预测模型

这个分割-预测的传统方法中, 分割方法为多尺度滑动窗口, 预测方法使用单隐含层神经网络作为分类器, 而分类器使用的特征向量的提取来自图像颜色特征与基于 SIFT 特征的视觉词袋模型。

为使用词袋模型, 首先需要训练词袋模型所使用的字典。考虑原始训练数据量其实很大, 首先随机抽取一部分图像样本, 对从中提取的 SIFT 特征描述进行聚类, 选取这些聚类中心作为字典。

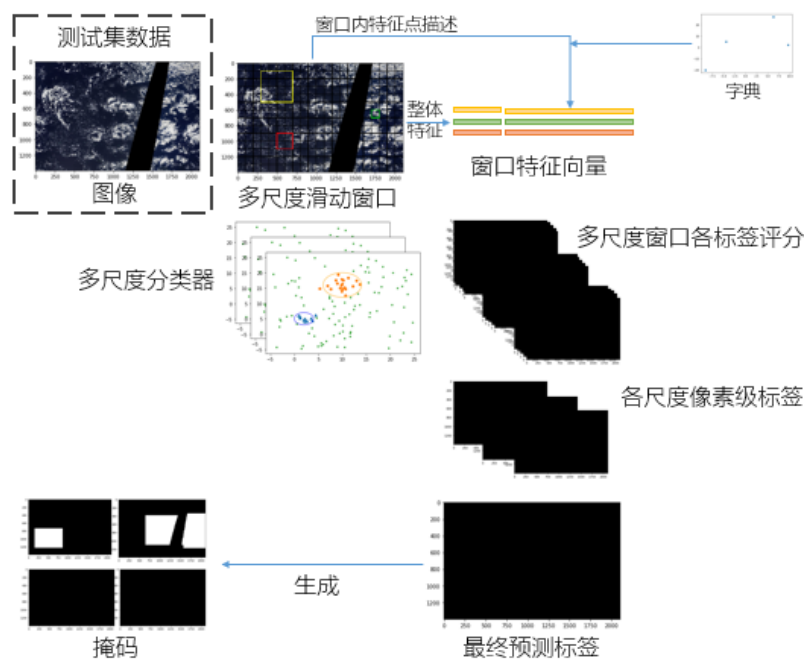
接下来是图像的分割和特征提取的过程。使用不同大小的滑动窗口以一定的步长在原图像上滑动, 依据这个窗口内的图像块的信息提取特征向量。根据对数据的分析, 我们发现二值化结果的颜色比在不同标签下分布差异显著, 且不同标签对腐蚀和开运算有不同敏感度, 因此选择二值图像、腐蚀结果、开运算结果的颜色比作为图像窗口的整体特征; 然后对这个窗口提取 SIFT 特征, 通过词袋模型的字典转化成词袋向量。相拼接后构成窗口的特征向量。

为训练模型, 在训练集图像上进行分割和特征提取, 并将窗口中的主要标签作为窗口标签。这样得到了特征向量和窗口标签, 训练的主要过程是为每一个窗口尺度, 通过这些特征向量与窗口标签训练一个分类器。整个训练过程示意图如下所示:



在测试阶段，使用与训练时相同的分割方式，然后使用分类器为每个窗口设置一个分类标签，通过得分的累计，确定这一尺度下全部像素点的标签。在确定各尺度的逐像素标签后，为每个像素点，选择出现在不同尺度上的标签加权作为这个像素点的标签。由此得到一张图像的预测结果。

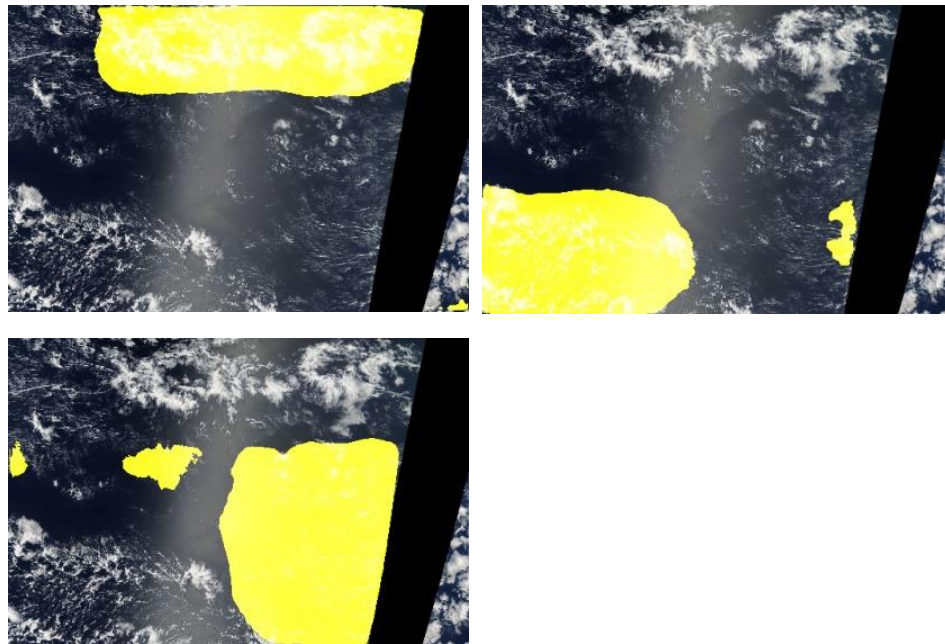
测试流程如下下图所示：



五、实验结果与分析

1、 基于 U-Net 框架的网络模型

1) 模型预测结果



上图自上而下， 从左往右依次为同一图片在 Fish, Gravel 和 Sugar 上的识别结果。

2) Dice 系数及 Kaggle 排名结果

Model_test_1: 编码器为 EfficientNet 的模型

Name	Submitted	Wait time	Execution time	Score
model_test_1.csv	a few seconds ago	0 seconds	22 seconds	0.65797
Complete				
Jump to your position on the leaderboard				

Model_test_3: 编码器为 ResNet34 的模型

Name	Submitted	Wait time	Execution time	Score
model_test_3.csv	a few seconds ago	0 seconds	22 seconds	0.66858
Complete				

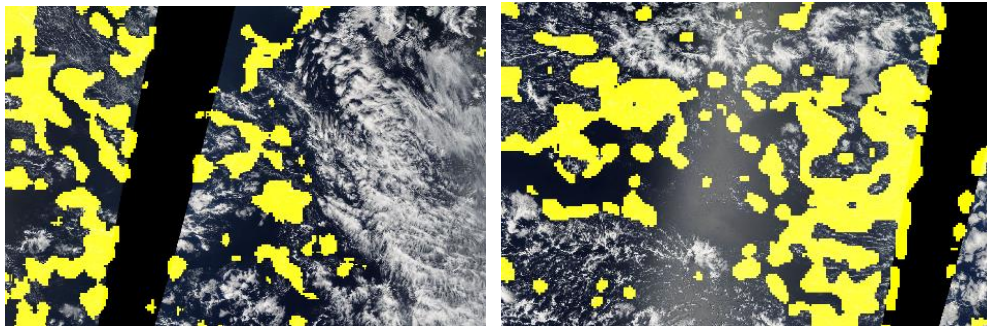
Model_test_4: 编码器为 resnest101 的模型

Name	Submitted	Wait time	Execution time	Score
model_test_4.csv	a few seconds ago	0 seconds	22 seconds	0.67843
Complete				

其中最高分数 0.67843 在 Public Leaderboard 中进行了比对，排名介于第六名到第七名之间。

2. 基于传统方法的分割-预测模型

1) 模型预测结果



以上为不同图片上对 Sugar 分类的预测结果。

2) Dice 系数及 Kaggle 排名结果

Name	Submitted	Wait time	Execution time	Score
test.csv	just now	0 seconds	4 seconds	0.44857
Complete				

3. 实验结果分析

本实验中进行了两种方法的探索，其中传统方法是一种直观、朴素、容易解释的方法，也是深度学习出现前的常见方法。主要的判断依据常集中于某种标签时出现的图案中的关键点，这种方法对局部形状较为敏感。从结果上观察得分可以判断这是一个能产生有效结果的方法。但在云图识别这个具体问题上，传统方法有着不小的局限性。深度学习出现前，对可变窗口的常见处理方法是选择性搜索，但云图依纹理而非色彩等划分，原理

上难以通过图像分割确定候选窗口，同时逐像素预测又容易产生不连贯的标签区域；传统方法也不擅长结合不同尺度的预测结果，各个尺度上实际都过于关注窗口中的细节内容而较轻视图像窗口的整体。

深度学习方法通过使用更大的资源和更多的计算克服了这些缺点。通过多层可调整卷积核的卷积，从图像中提取了更加适合云图的特征；通过对卷积结果在四个种类上分别进行的二分类，拟合了云图中可能出现的特征与标签的联系，并为上采样提供特征支撑；通过上采样的方法，生成了连续且较准确的标签区域。通过不同的编码器来采集不同的底层特征汇总，能考虑更多不同种类的特征，得到一个更加准确的结果。

从实验结果看，编码器的选择对深度学习的方法的结果也有不小的影响，与之类比可以推知，影响窗口特征向量的选择方式的因素，如选择的特征的种类、词袋模型的字典抽样、聚类数目等，并非刻意设计的其他因素，可能也局限了传统方法的性能。

附录：

- ✓ 本次作业共提交文件如下：

文件：‘实验报告.pdf’，其内容是实验报告的具体内容。

文件夹：‘U-Net 代码’，其中包括这一部分的 Python 源代码和提交文件。

文件夹：‘传统方法代码’，其中包括传统方法部分的 Python 源代码和提交文件。

- ✓ 小组 Kaggle 账号与小组成员联系方式：

分组人员：

2120190505 公岩松

2120190506 龚士博

2120190511 甘润东

联系方式：

电子邮件：

raineast666@163.com

1511420@mail.nankai.edu.cn

1511419@mail.nankai.edu.cn

联系电话：

15122391779

13513616850

Kaggle 账户：

RaineastGam