

# 第九届“泰迪杯”数据挖掘挑战赛

作品单位：浙江工商大学

作品成员：乔肃 蓝志雄 项琼秋

指导老师：董雪梅

此页信息原作品沒有  
第九届“泰迪杯”数据挖掘挑战赛

# 基于深度学习的岩石样本智能识别研究

## 摘要

岩石样本检测是地质勘探中的重要环节，传统的岩石样本检测算法往往过分依赖复杂的特征工程和地质学的专业知识，且尚不能取得满意的效果。近年来，基于深度学习理论的算法在图像分类、目标检测等方面取得突破性进展，在岩石样本检测任务上也同样拥有巨大潜能。因此，本文将通过图像处理技术和深度学习算法设计有效的模型，识别出岩石样本的岩性类别及含油气情况，实现岩石样本智能识别分类。

针对任务一，本文利用图像的高分辨率特性，分别采取随机裁剪与网格裁剪两种方式对数据进行处理，并应用于不同的模型架构。首先，对于随机裁剪的数据，以 ResNeXt-50 为主干网络，通过搭建“并联式”模块融合图像全局与局部特征，并据此建立“全局-局部模型”，其岩石分类准确率为 81.43%。其次，采用置信学习剔除网格裁剪的错误数据，分别以 EfficientNet-B3 与 MobileNet-V3 为主干网络训练分类模型，在预测阶段结合 TTA 策略进行滑动窗口预测，根据多个预测结果投票得到最终的预测结果，其岩石分类的准确率分别为 78.57% 和 80.00%。最后，利用上述三个模型在不同岩石类别的识别能力差异及各自的优缺点，进行模型融合以再次提升岩性识别能力，最终的岩石分类准确率为 82.86%。

针对任务二，任务目标“计算石油含量”可分解为荧光区域分割、岩石区域分割两个步骤。首先，将图像映射至 HSV 空间，通过可视化方法确定石油荧光的 HSV 取值范围，从而快速、准确地分割出荧光区域。同时为计算岩石所占总面积，本文基于 Canny 边缘检测算法设计了高效的岩石轮廓提取方案，借此重新校准石油荧光的分割结果，从而减少误判。最后，计算荧光区域和岩石区域的像素数量，将二者相除即得所估计的石油含量面积百分比。

**关键词：**岩性识别；含油量计算；图像处理；深度学习

# 目 录

1 前言 .....	1
1.1 研究背景 .....	1
1.2 研究任务 .....	1
2 数据探索 .....	2
2.1 类别不平衡问题 .....	2
2.2 数据质量问题 .....	2
2.3 图像格式问题 .....	5
3 数据预处理 .....	6
4 基于 DCNN 的岩性智能识别与分类 .....	8
4.1 基于随机裁剪数据的图像分类模型 .....	8
4.1.1 ResNeXt 网络构架 .....	8
4.1.2 逐步细化模型 .....	11
4.1.3 全局-局部模型 .....	11
4.2 基于网格裁剪数据的图像分类模型 .....	14
4.2.1 EfficientNet 简介 .....	14
4.2.2 MobileNet-V3 简介 .....	15
4.2.3 置信学习 .....	16
4.2.4 滑动窗口预测 .....	16
4.2.5 TTA .....	17
4.3 训练策略 .....	17
4.3.1 数据集划分 .....	17
4.3.2 Adam .....	18
4.3.3 OneCycleLR .....	18
4.3.4 Focal Loss .....	18
4.4 模型融合 .....	19
5 岩石含油比例计算 .....	20
5.1 荧光区域分割 .....	20
5.1.1 颜色空间转换 .....	20
5.1.2 石油荧光分割 .....	21
5.2 轮廓检测 .....	22
5.2.1 图像降噪 .....	22
5.2.2 Canny 边缘检测 .....	23
5.2.3 形态学处理 .....	24
5.2.4 轮廓检测 .....	24
5.3 计算岩石含油比例 .....	25

6 实验结果分析.....	26
6.1 实验环境.....	26
6.2 评价指标.....	26
6.3 实验结果.....	26
6.3.1 任务一.....	26
6.3.2 任务二.....	31
7 总结与展望.....	32
7.1 总结.....	32
7.2 展望.....	33
参考文献.....	34

第九届“泰迪杯”数据挖掘竞赛

# 1 前言

## 1.1 研究背景

岩石是现代建筑业和制造业的重要原材料，除了作为原材料使用以外，还可以对其进行勘探开发挖掘岩油气藏，若能智能且准确地识别岩石岩性、计算岩石含油量，这将会是一笔巨大的社会财富。

天然岩石有多种来源和用途，根据其不同的特征，可分为火成岩、沉积岩和变质岩三类。岩石的探测与识别是地质调查研究和矿产资源勘查的基础工作，岩石的精准识别与分类对地质的探测与识别极为重要，一般可通过多种方式进行鉴定，例如重磁、测井、地震、遥感、电磁、地球化学、手标本及薄片分析方法等方法。而现场的工作条件通常仅限于目视方法，包括对细粒岩石使用放大镜，视觉检测评估诸如颜色、成分、粒度和结构等属性。

近些年来，卫星和遥感技术的进步推动了多光谱遥感技术的发展，以分类不同类型的地物，包括岩石。但是，利用遥感技术在野外获取超高分辨率岩石图像的成本很高。因此，利用无人机和卫星携带的高光谱技术进行数据采集的高成本往往阻碍了岩石岩性识别的自动化。

而图像采集和计算机图像模式识别技术的迅速发展，使得从野外采集的图像中识别岩石的自动系统得以发展。这些技术为辅助地质学家判断和识别岩石岩性提供了极大的便利，也极大推进了深度学习算法在这一方向的落地可能。

## 1.2 研究任务

利用图像处理技术和深度学习算法解决以下两个任务：

### (1) 岩性识别与分类

以白光环境下拍摄的岩石样本图像数据为基础，设计合适的机器学习或深度学习算法，构建岩石样本岩性智能识别模型。

### (2) 计算岩石含油面积百分含量

以荧光环境下拍摄的岩石样本图像数据为基础，借助石油在紫外线的照射下具有发光的特性，设计合适算法计算图像中岩石的含油面积百分比含量。

## 2 数据探索

### 2.1 类别不平衡问题

所提供的岩石样本数据中，存在着严重的类别不平衡问题，如图 2-1 所示。其中浅灰色细砂岩样本数量最多，占全体数据的 26.98%；深灰色泥岩次之，占比 23.81%；灰色细砂岩的样本数量最少，仅有 5.72%，不足浅灰色细砂岩的四分之一。

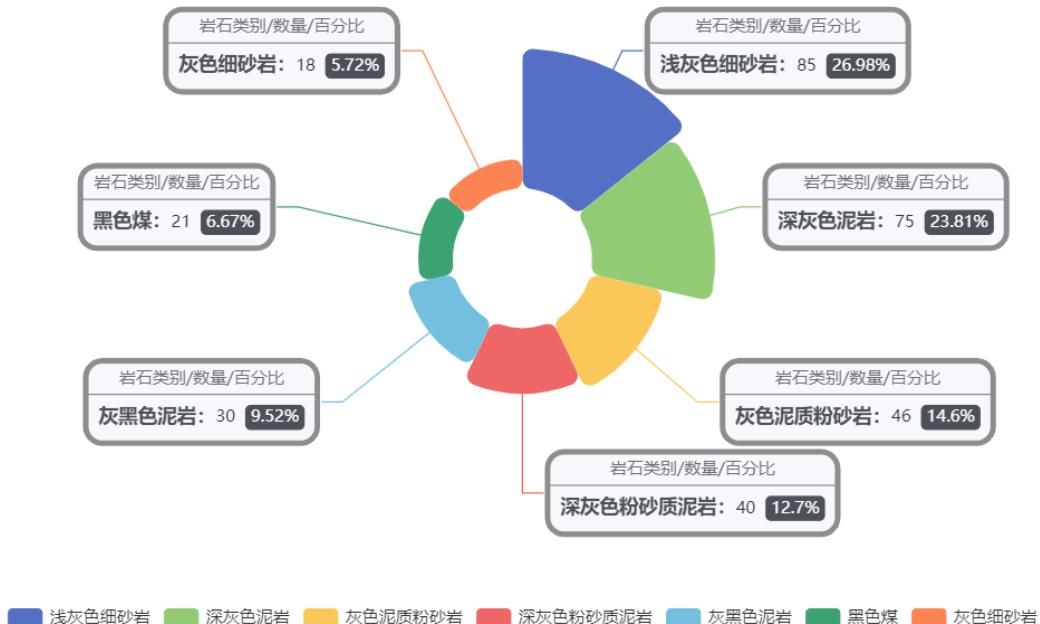


图 2-1 类别分布

如此类别不平衡的数据集，模型很容易过度拟合样本更多的类别，同时忽视对小样本类别的学习。因此类别不平衡问题是模型训练时不可忽视的重要问题。传统的处理类别不平衡问题的方法主要是重采样，重采样可划分为两类，一是对数量多的类别欠采样，二是对数量少的类别过采样。重采样后的数据集具有平衡的类别，但数据集中包含的信息并不会因此增加，故而可能会导致模型欠拟合于数量少的类别而过拟合数量多的类别。因此，可换个角度，从损失函数方面考虑类别不平衡问题的处理。

### 2.2 数据质量问题

经观察发现，样本中提供的七类岩石的类内差异较大，同时也可能混杂有标注错误的样本。以深灰色粉砂质泥岩为例，抽选编号为 73、81、137 的样本，如

图 2-2 第一行所示。从中可见，同一类岩石的颜色、形态可能大相径庭，这无疑加大了分类的难度。图 2-2 按列展示了不同类别岩石可能存在的相似情况，在缺乏专业知识的人眼中，每一组图片可能相差无几（可能是错误标记）。此外，如此巨大的类内差异和微小的类间差异很容易掩盖错误标注的样本，从而误导模型训练时优化的方向，降低模型识别能力。

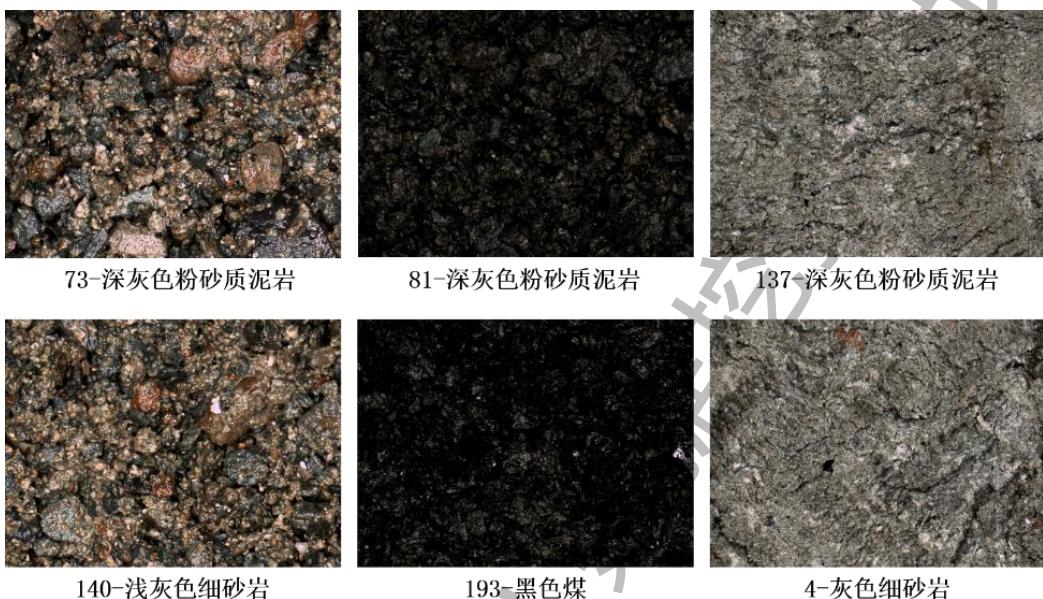


图 2-2 样本差异示例

众所周知，拍摄环境对照片呈现的内容影响极大。如果按干燥程度区分，可将环境简单分为干燥、潮湿、有水三类；按光照情况划分，则可分为光照好和光照差两类。

以灰黑色泥岩为例，选取五个不同环境下拍摄的典型样本进行展示，如图 2-3。对比 40 号和 51 号图像，可见越潮湿的岩石颜色越深。结合 309 号和 198 号图像，可发现当岩石间有水时，水面的反光甚至会掩盖岩石的纹理细节，即使在光照不好的情况下也不例外。整体上看，光照情况越好的图像的纹理细节越清晰，对岩石颜色的反映越真实。

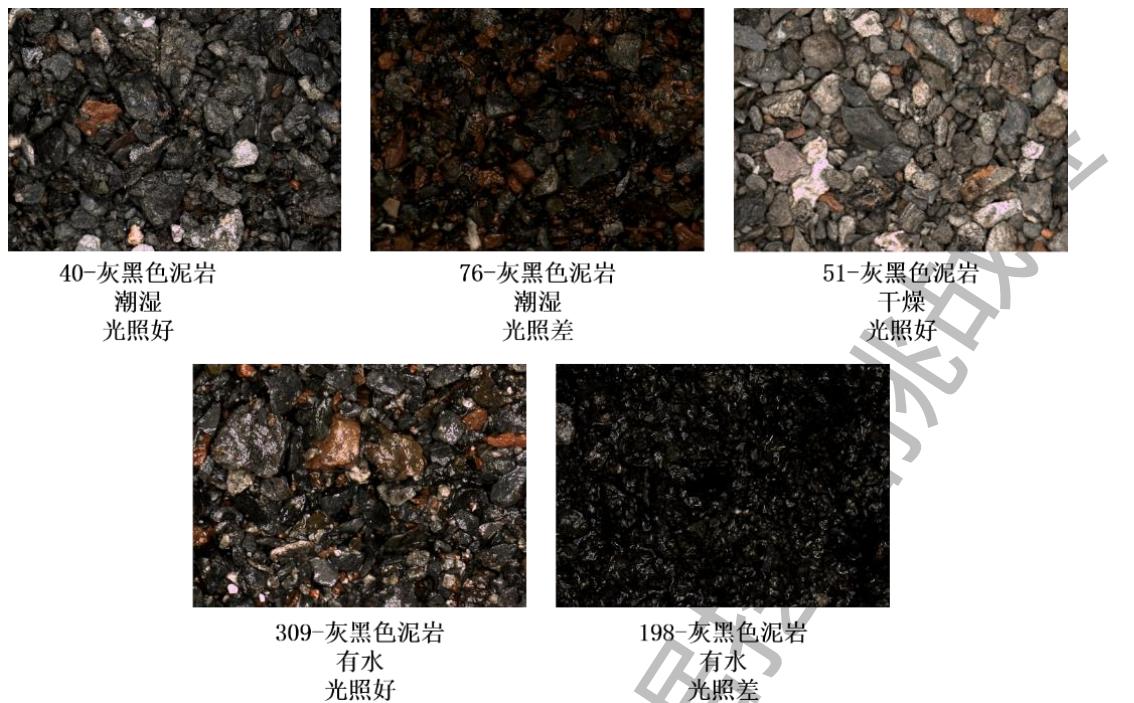


图 2-3 环境问题示例

此外在部分的岩石样本数据中，还存在着掺杂其他类别岩石或杂物的情况。以类别为浅灰色细砂岩的 235 号照片为例，如图 2-4 所示。图像混杂了大量橙色、褐色和黑色的小石块，同时落入一片树叶，遮挡将近八分之一的区域。如此杂乱的数据无疑是对数据预处理和模型提取特征能力的一大挑战。

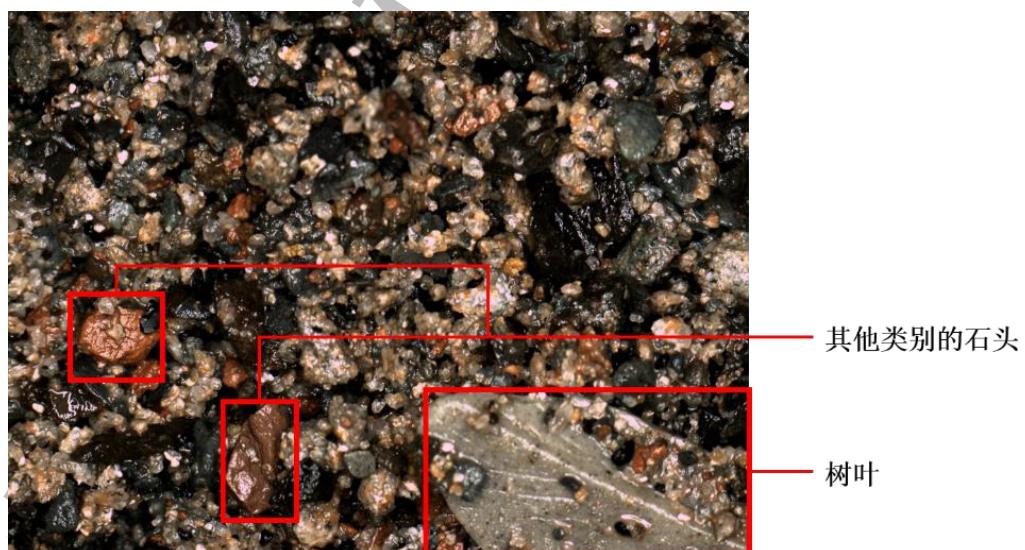


图 2-4 数据质量问题示例

## 2.3 图像格式问题

样本图像中存在  **bmp**  和  **jpg**  两种格式，且两种格式的图像大小和风格均有巨大差异。 **bmp**  格式图像无明显背景，大小为  **3000x4096**  (像素)。此类照片受环境因素影响大，光照、潮湿程度各异，且往往混有其他岩石或物体。 **jpg**  格式图片存在显著蓝色背景、大小为  **2048x2448**  (像素)。此类照片光照差异不大，岩石中的杂物也不多，但其中的岩石多为细小的碎块，难以提取纹理特征。

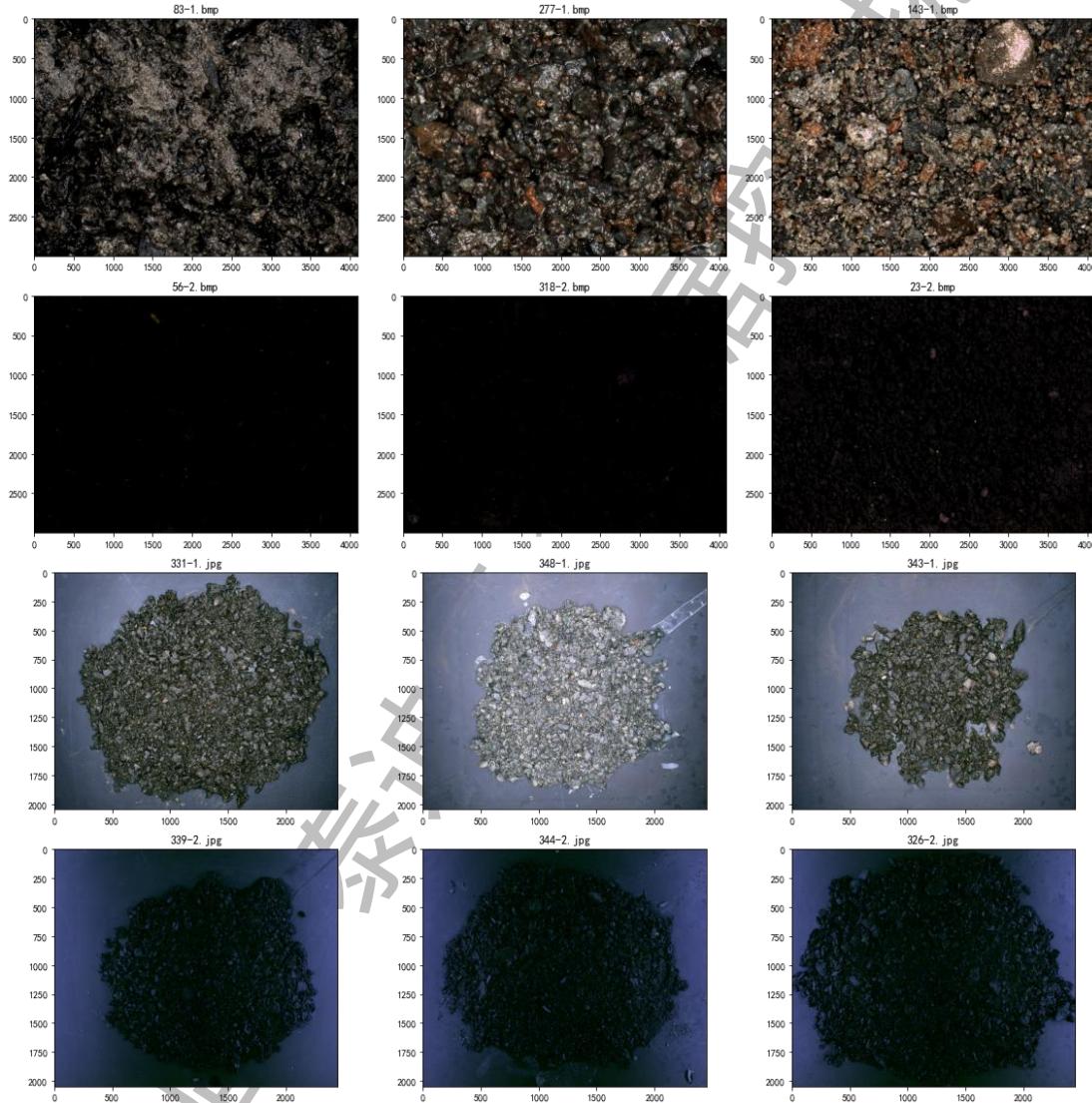


图 2-5 图像格式问题数据示例

### 3 数据预处理

由于数据集中照片的分辨率较高且样本数较少，可选择通过裁切图像的方式增加样本数量。根据不同的模型特点，本文有针对性地使用了随机裁剪和网格裁剪这两种略有不同的增强方式，其示意图如图 3-1。

其中随机裁剪即在原图上随机剪出预设大小的小块图像作为全新的样本，这种方式的优点在于生成的样本数量多，模型有机会学习到原始图像的所有信息。但由于该方式的随机性较强，难以预料裁剪出的内容，因此当数据质量不高时，依靠随机裁剪扩增的样本中可能包含大量的噪声。

而网格裁剪即按照预设的网格，将图像裁剪为相同尺寸的小块。使用网格裁剪扩增的样本具有固定的数量和内容，虽然会损失一部分横跨网格的信息，但更便于进行数据分析和预处理，从而剔除噪声数据。

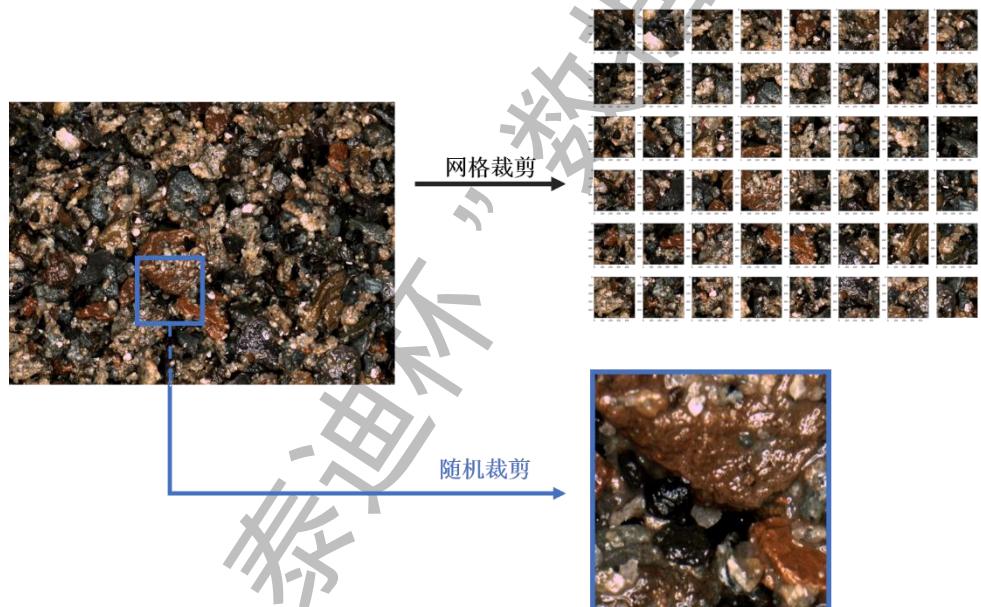


图 3-1 裁剪方案

通过前文的探索性数据分析，常发现某不知名红褐色石头混杂在样本中，且不属于任何一类。因此在裁剪后，选择计算红色和橙色在样本中所占的比例，通过实验确定阈值，从而剔除以红褐色岩石为主体的样本。此外，根据前文可知同一类岩石可能因光照、潮湿程度等环境因素改变颜色、纹理。因此为了增加模型的鲁棒性，选择采用随机增强策略对样本进行随机变换。

图像增强是一项通过简单图像变换使数据特征更明显的技术，现已逐渐推广到数据扩增用途。而随机增强即预先选择多种图像变换方案，并以一定概率随机

组合这些变换，使得样本图像千变万化，从而训练模型提取图像中最关键的特征信息，使模型具有更强的泛化能力。为对抗样本中存在的光照、潮湿等环境因素的影响，选择水平翻转、垂直翻转、随机饱和度调节、随机亮度调节、随机对比度调节五种基础增强方案，并在训练模型时随机组合这些基础方案。以 16 号样本为例，五种基础增强方案的示意图如图 3-2。

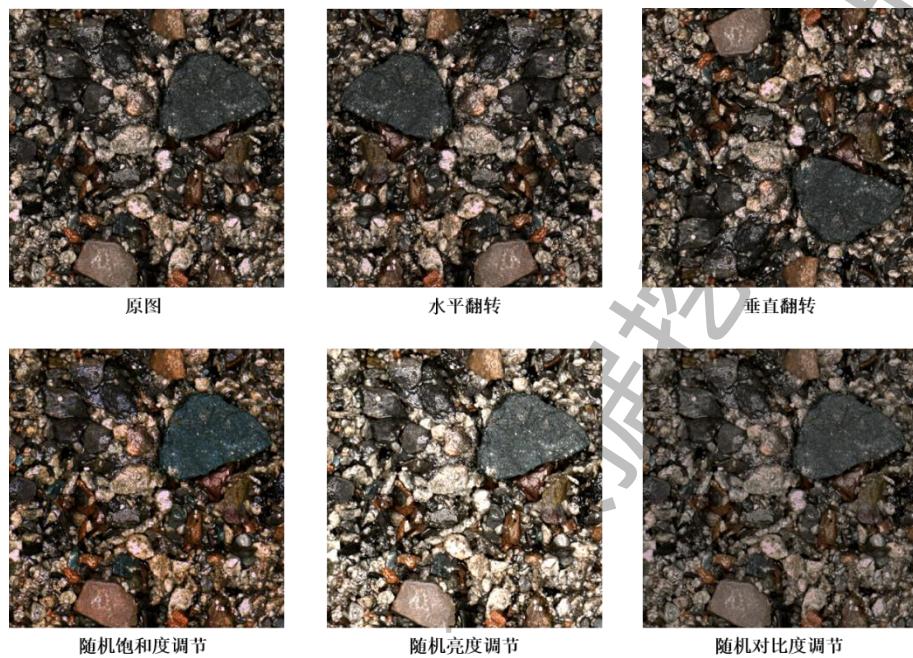


图 3-2 随机增强示意图

综上所述，数据预处理技术路线图如下图 3-3。

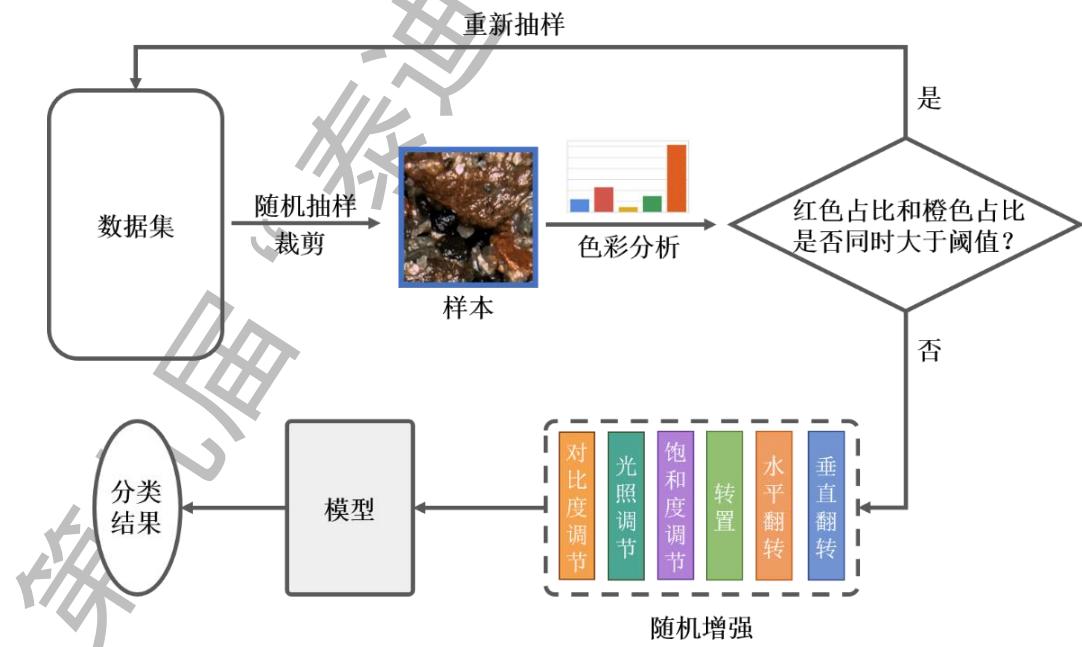


图 3-3 预处理流程图

## 4 基于 DCNN 的岩性智能识别与分类

经过一系列的探索性数据分析，发现除轮廓、色彩等信息对岩石分类有帮助外，岩石的局部纹理特征也对分类效果起着至关重要的作用。因此，本文构建了全局-局部模型，用以提取图像全局和局部的纹理特征，再使用精心设计的融合模块将两种特征合并，馈入前馈神经网络，得到分类结果。

除此之外，注意到所提供的岩石图像中可能混入了其他种类的岩石，这会在一定程度上影响模型分类结果。因此本文基于 EfficientNet-B3 和 MobileNet-V3 构建了两个仅关注局部纹理的岩性识别与分类模型，借助置信学习剔除噪声数据，并提出了滑动窗口预测法以综合考量图像主体所属的类别，进一步削弱由图像数据质量问题带来的负面影响。

最后，再以“硬投票（Hard Voting）”的方式对三个模型的预测结果进行融合，从而得到更为鲁棒且精确的预测结果。

### 4.1 基于随机裁剪数据的图像分类模型

本节基于 ResNeXt-50 进行改进，充分融合图像的全局与局部信息，采用随机裁剪与图像缩放得到的图像作为训练集进行模型训练，得到最终的“全局-局部模型”。

#### 4.1.1 ResNeXt 网络构架

随着计算机硬件设备的不断升级，计算机视觉领域迅猛发展，用于图像识别的网络架构层出不穷。众所周知，网络的深度是决定模型性能的关键因素。理论上可以通过增加网络的层数，让模型获得提取更复杂、更有效的特征的能力，从而使得模型获得更好的性能。但 2016 年何凯明等人在 ResNet[1]一文中指出网络的层数与模型的性能并非成正比的关系，深度高的模型甚至比深度低的模型效果可能更差，即网络退化。因此，何凯明团队提出“残差结构”来解决网络退化问题，如下图 4-1 所示。

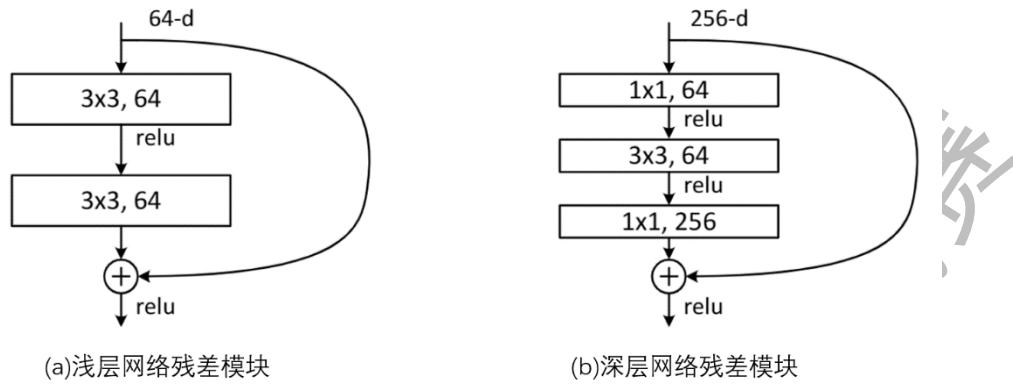


图 4-1 ResNet 残差模块

当输入为 $x$ 时，网络学习到的特征为 $H(x)$ ，由于存在网络退化现象，ResNet [1] 在原来的模块中增加一个旁路(shortcut)，此时模型学习到的原始特征为 $F(x) + x$ ，网络便只需学习 $H(x) - x = F(x)$ 的残差信息，这与直接学习原始特征的网络相比更为简单且有效。此外，当残差为 0 时，该堆叠模块仅做了简单的恒等映射，不会使模型的性能下降，从而解决了模型层数越多，网络性能越差的问题。

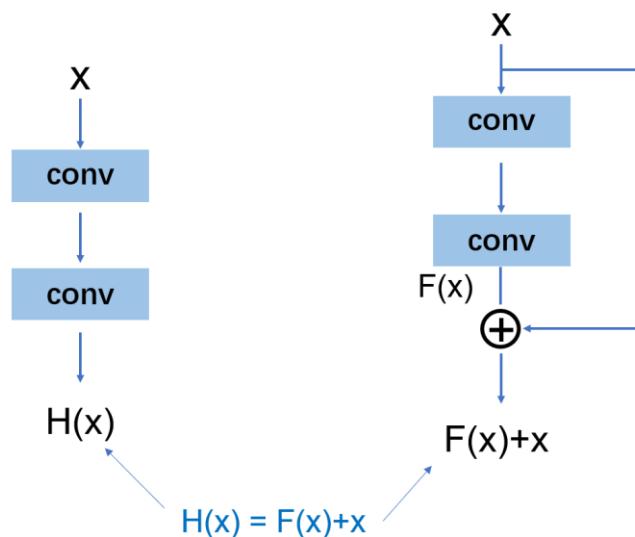


图 4-2 ResNet 残差模块示意图

ResNeXt [2]是由 Facebook 在 2017 年提出的一种基于 ResNet 改进的网络架构。ResNeXt 将 ResNet 网络中的“直线式”残差结构改为多分支的同构结构，其与 ResNet 的区别如图 4-3 所示。

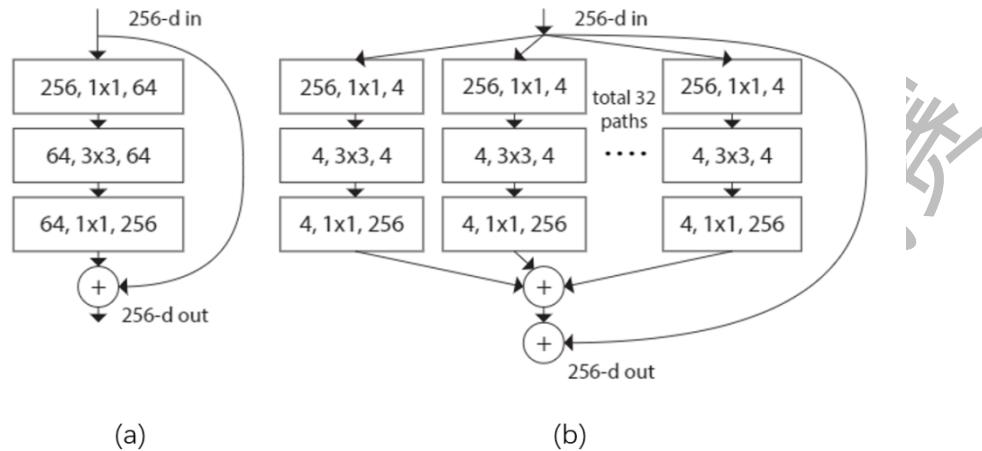


图 4-3 ResNet 与 ResNeXt 残差模块对比图

下图 4-4 为一个没有偏置项的线性激活神经元，它是一个最简单的“split-transform-merge”结构，它将输入的  $m$  个元素，分配到  $m$  个分支进行权重加权，最后对所有的支路求和。由此神经网络的一个单元可以如下式所示：

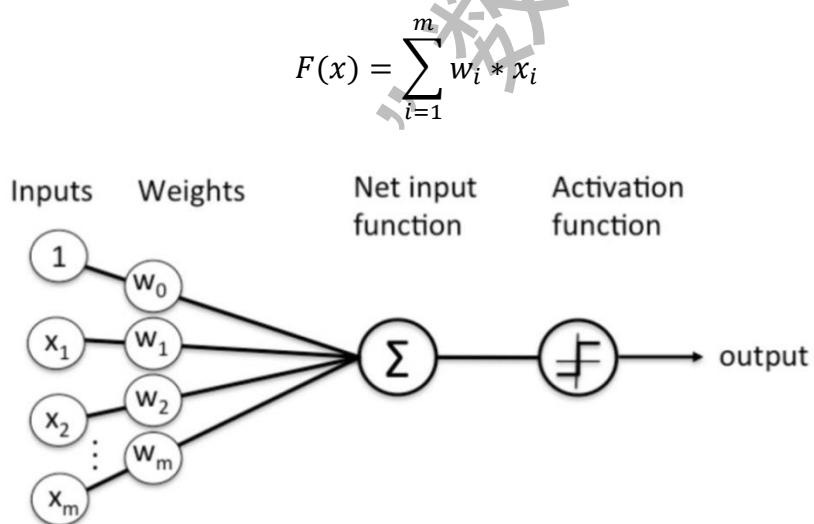


图 4-4 线性激活神经元

而 ResNeXt 就是借鉴了“split-transform-merge”想法，将原本的残差模块分成多个小组（即多种特征）进行独立的同构变换，最后再对其进行特征融合。ResNeXt 的残差结构可以用如下公式表示：

$$y = x + \sum_{i=1}^C T_i(x)$$

其中， $C$  表示每个残差模块中的分组个数， $T_i$  表示特征  $x$  的任意变换。

与 ResNet 相比, ResNeXt 在相同参数的情况下, 错误率更低, 模型性能更好, 故选择 ResNeXt-50 作为以下模型的主干网络。

#### 4.1.2 逐步细化模型

本次岩石总共分为 7 个类别, 分别为黑色煤、深灰色泥岩、深灰色粉砂质泥岩、灰黑色泥岩、浅灰色细砂岩、灰色泥质粉砂岩以及灰色细砂岩。根据岩石特征以及类别, 可将其大致划分为黑色煤、泥岩以及砂岩三类, 然后再针对泥岩和砂岩这两类岩石进行类别细分。因此可针对不同的岩石类型以 ResNeXt-50 为主干网络分别训练三个模型, 并将其组合成一个模型来逐步的对岩石进行分类, 如图 4-5 所示。

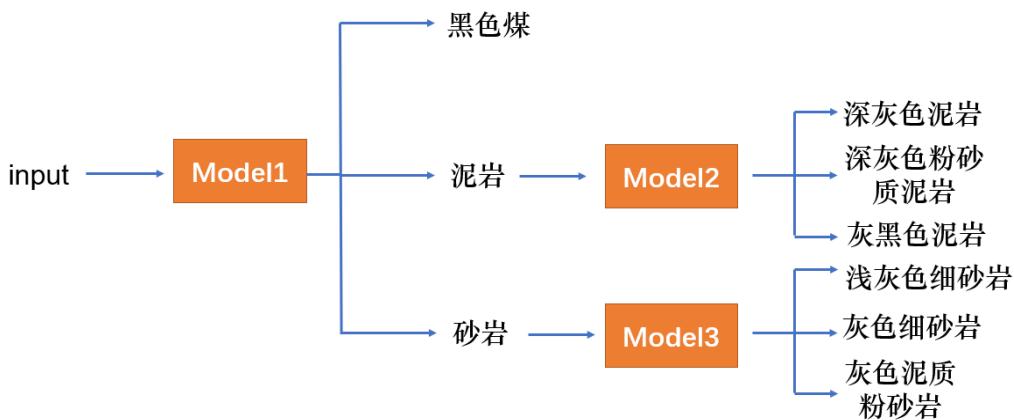


图 4-5 逐步细化模型

逐步细化模型中的三个子模型分别对应三个任务, 期望通过多模型学习不同的特征达到正确分类的目标。但是, 该模型存在一个缺点, 即 Model2 与 Model3 是建立在 Model1 的划分基础上, 一旦 Model1 分类错误, 就意味着 Model2、Model3 的性能无论多好, 都无法改变图像的大类, 因此这张图片的预测类别肯定与真实类别不符。也就是说 Model1 的分类准确度决定了整个模型的性能上限。整个模型的准确率是 Model2、Model3 的准确率与 Model1 准确率的乘积之和。

#### 4.1.3 全局-局部模型

由于逐步细化模型会出现准确率严重受限于以 ResNeXt-50 为主干的 Model1, 因此本文基于逐步细化模型作进一步改进。

逐步细化模型的主要思想是先分大类再细分小类。由于煤、砂岩和泥岩这三个大类别之间的差异较大, 用于区分它们的模型 (Model1) 通常只需要对图像整体的色彩、结构有一定把握即可, 我们把这样大尺度的信息称为全局特征。而

进一步细分泥岩或砂岩的 Model2、Model3 则需要学习图像细节的纹理特征，才能更好地区分差异不大的子类别，我们把这样细粒度的特征称为局部信息。如下图 4-6 所示，逐步细化模型其实是全局信息与局部信息的“串联式”结构连接。

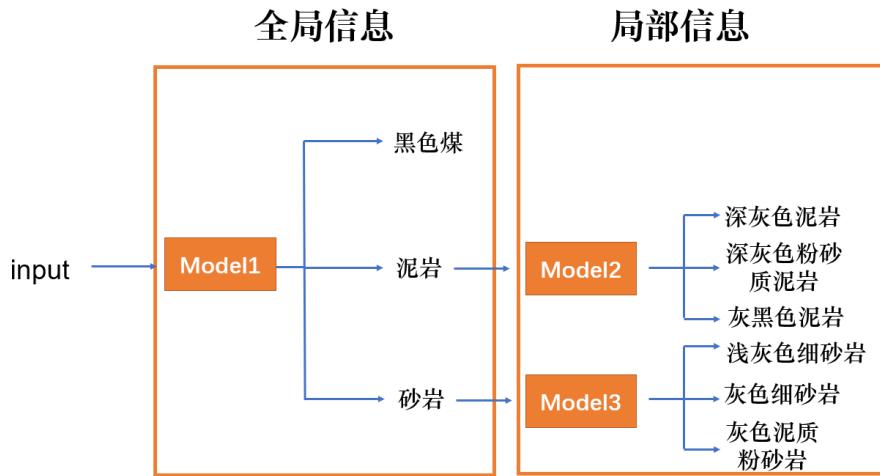


图 4-6 逐步细化模型串联结构

在串联式结构中，Model1 有绝对的话语权，控制了模型的上限。为解决这一问题，本文选择将 Model1 提取的全局特征与 Model2 提取的局部特征进行“并联”，构造全局-局部模型。该模型使得两模块并重，不再受限于单一模型，如图 4-7 所示。



图 4-7 全局-局部模型

其中，Model1 与 Model2 都是以 ResNeXt-50 为主干的模型架构。一般而言，ResNeXt-50 网络只需要尺寸为 224x224 的输入即可。而赛题提供图像的两种分辨率（3000x4096、2048x2448）非常大，这就为多样化的预处理方式提供了基础。为了使全局-局部模型中的子模型各司其职，可以对图片进行不同的处理，从而使模型学到不同尺度的纹理特征。

首先，将整张图片直接缩放到 448x448，再对其进行中心裁剪，得到分辨率 为 224x224 的图像。处理过后的图像与原图相比损失了部分细节的纹理信息，但

是图像的全局信息仍被保留了下来。将这一类图像放入到 Model1 中训练，可以使网络很好地学习到图像的全局信息。

其次，对原图像进行随机裁剪，同样得到一张 448x448 的图像，随后在将其缩放到 224x224。与之前不同，这张图像虽然也是 224x224 的大小，但因其是先在原图上进行裁剪，再进行缩放，所以图像虽然损失了全局信息，但其很好的保存了局部的纹理细节，因此 Model2 能够很好的学习图像的局部信息。

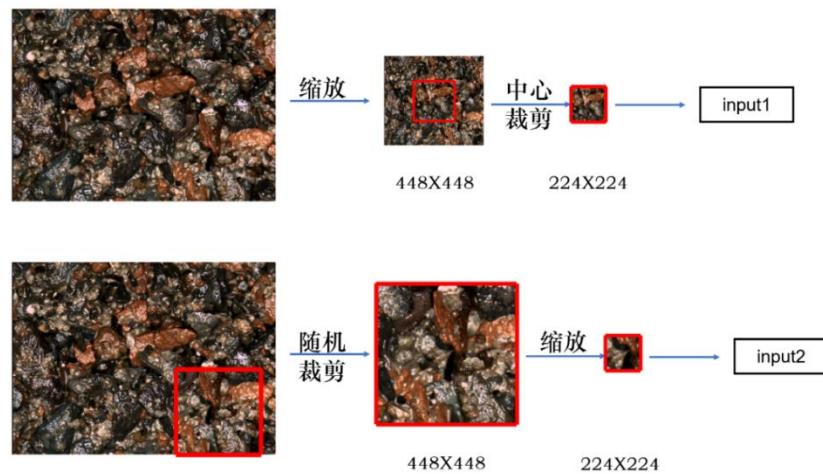


图 4-7 不同模块输入详解

最后，本文将 Model1 与 Model2 的最后一层全连接层移除，即移除类别输出层，此时的模型便成为了特征提取器。再将提取到的全局信息的特征与局部信息的特征拼接起来，馈入一个逐点卷积层，从而深度融合全局、局部信息，让全局-局部模型具有更好的性能。此外，全局-局部模型的并联式结构特点类似于 ResNeXt 中的“split-transform-merge”模式，这表明以 ResNeXt-50 为主干构建的全局-局部模型是合理、可行的。

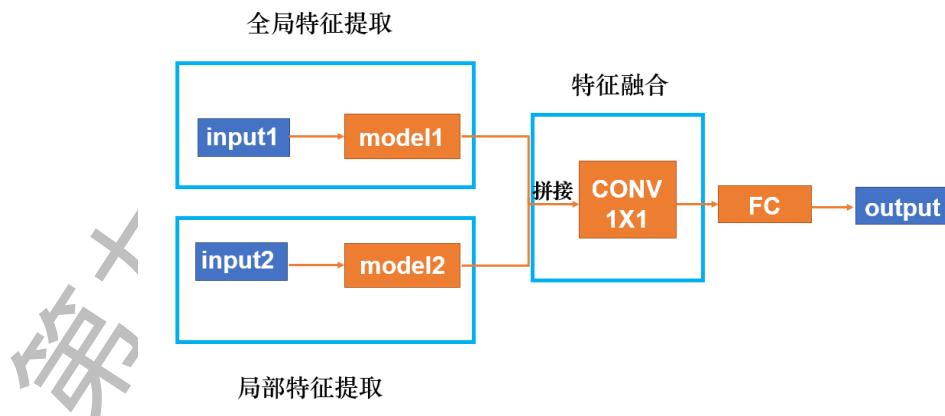


图 4-8 全局-局部模型详解

## 4.2 基于网格裁剪数据的图像分类模型

不同于 4.1 节，本节将使用网格裁剪得到的数据作为训练集进行模型训练，并采用 EfficientNet-B3 和 MobileNet-V3 两个不同的主干网络进行训练，得到两个图像分类模型。

同时，为降低标签错误、数据质量等问题带来的模型分类能力影响，此节通过置信学习对初始数据进行清洗，剔除错误标签的数据。将清洗后的数据划分为训练数据和验证数据，再载入 EfficientNet-B3（MobileNet-V3）进行训练，并以验证数据的分类准确度高低作为模型是否完成训练的标准。完成模型训练后，根据训练得到的权重载入模型，以滑动窗口预测的方式投票得到最终预测结果，以 EfficientNet-B3 为例，其流程如图 4-9 所示（MobileNet-V3 的流程类同）。

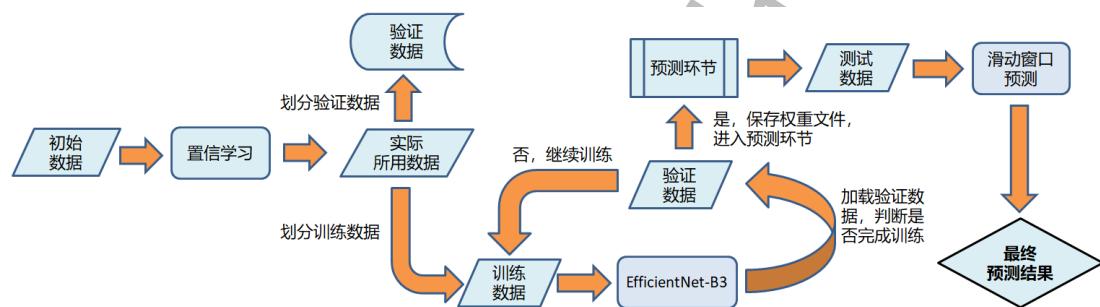


图 4-9 模型训练流程图

### 4.2.1 EfficientNet 简介

EfficientNet [3]由 Google 于 2019 年提出，分为 B0 至 B7 八个版本，其中 EfficientNet-B0 最为基础、也是参数量最少的版本。EfficientNet 的核心在于混合缩放，区别于传统的缩放方式，混合缩放通过定义混合缩放系数 $\phi$ ，根据式 4-2-1 同时实现深度 $d$ 、宽度 $w$ 、分辨率 $r$ 三个维度的动态缩放。

$$d = \alpha^\phi, w = \beta^\phi, r = \lambda^\phi \quad (4-2-1)$$

上式满足约束条件 $\alpha \cdot \beta \cdot \lambda \approx 2$ 且 $\alpha \geq 1, \beta \geq 1, \lambda \geq 1$ 。

混合缩放形式如图 4-10 (e) 所示。

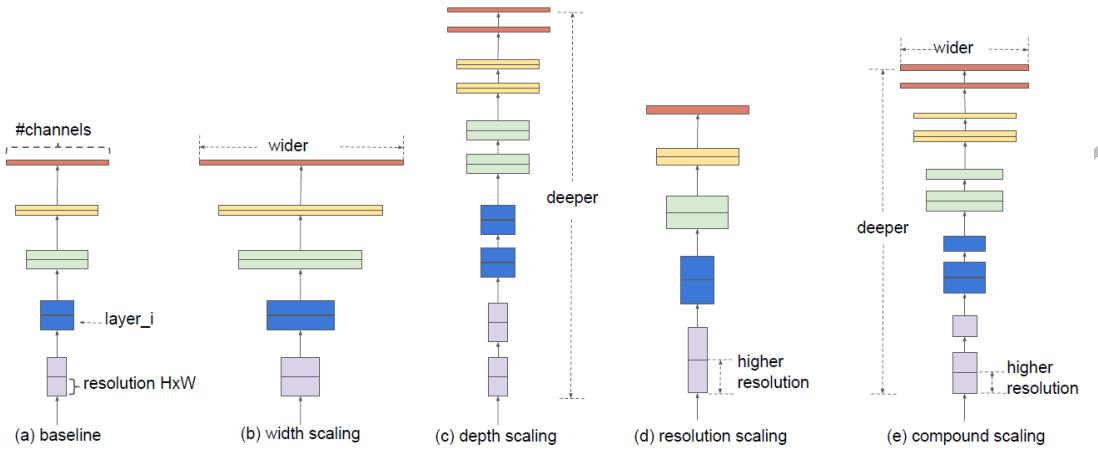


图 4-10 模型缩放方式

#### 4.2.2 MobileNet-V3 简介

MobileNet-V3 [4]是 Google 基于 MobileNet-V1 和 MobileNet-V2 进行改进后的 MobileNet 网络系列又一力作，其使用 NAS 算法搜索得到大体的网络结构，继而使用 NetAdapt 算法确定卷积层的通道数目，最后进行人工的细微调整，得到网络的最终结构，并且为适应不同的算力场景，推出了 Large 和 Small 两个版本以匹配不同算力资源的设备（本文采用 Large 版本）。

此外，MobileNet 系列的网络一直以能够部署在移动端，并能够拥有快速准确的模型预测结果为目的。而 MobileNet-V3 考虑到 Sigmoid 函数在移动端设备的计算成本很高，因而改用 h-swish 激活函数（式 4-2-3）近似 swish 激活函数（式 4-2-2）。

$$\text{swish}(x) = x \cdot \sigma(x) \quad (4-2-2)$$

$$\text{h-swish}[x] = \frac{x(\text{ReLU6}(x+3))}{6} \quad (4-2-3)$$

其中  $\sigma(\cdot)$  为 Sigmoid 函数，ReLU6 为 ReLU6 激活函数。

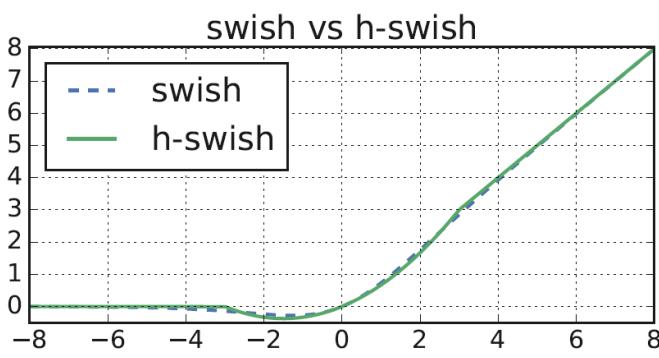


图 4-11 swish 和 h-swish 激活函数比较图

### 4.2.3 置信学习

置信学习是基于剪枝、计数和排序噪声数据，对噪声标签进行表征识别和学习的方法，其利用概率和带噪标签对非标准化的置信联合矩阵进行检验，估计联合分布，并借此对噪声数据进行剪枝。置信学习不需要超参数，只需要输入样本外预测概率和与噪声标签相关的数组，根据标签噪声概率输出有序样本以消除错误标签样本，流程如图 4-12 所示。

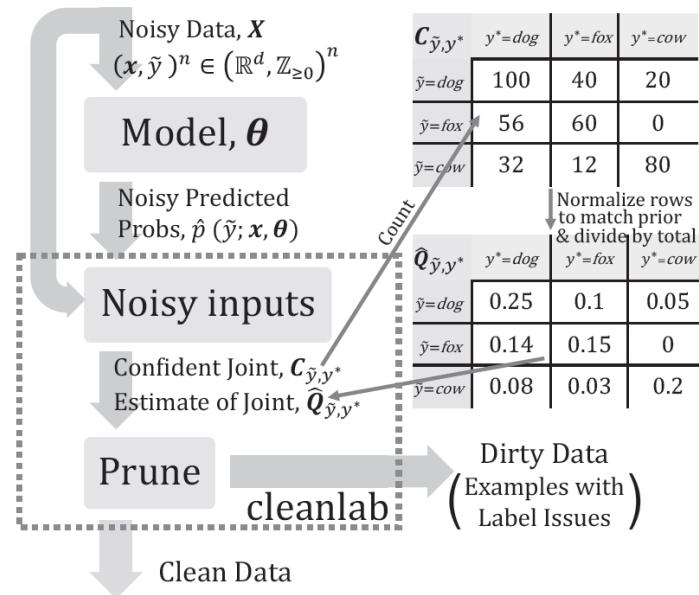


图 4-12 置信学习

### 4.2.4 滑动窗口预测

考虑到岩石图像的数据质量问题，对于单标签的图像，其图像内部可能掺杂了其他类别的信息，但是经过综合考量，多数图像数据中的大部分内容能够对应标签信息，因此可视作图像内部存在“噪声”。

为尽可能降低此类数据质量带来的“噪声”问题，本文提出单张图像以滑动窗口方式进行预测的方法，从而得到单张图像不同位置的多个预测结果，这也正对应了模型训练时所采取的数据处理策略，最终取预测结果中的众数作为实际输入单张图像的最终预测结果，如图 4-13 所示。

若选用滑动距离 $s$ 为窗口大小，则退化为对图像进行切块预测。

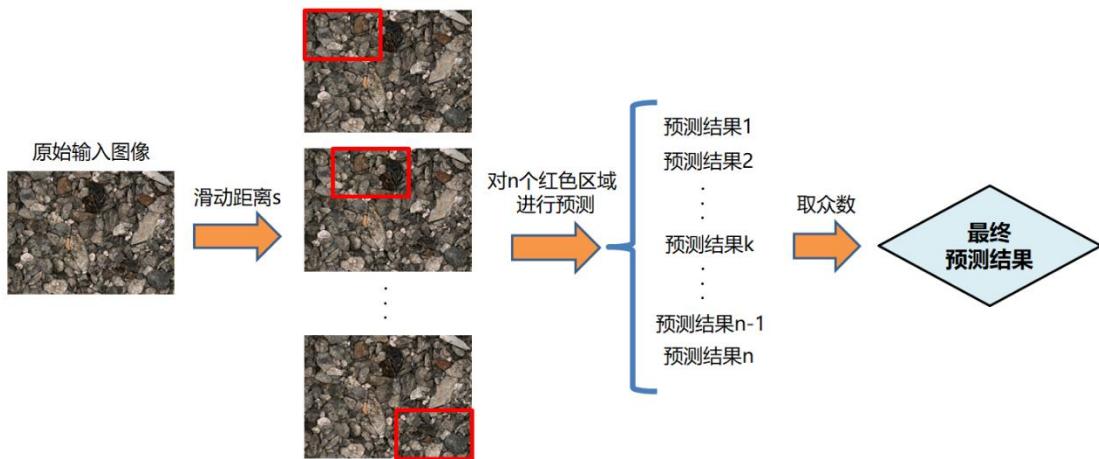


图 4-13 滑动窗口预测示例

#### 4.2.5 TTA

TTA [5]全称 Test Time Augmentation，即测试时增强，旨在模型预测阶段，将待预测的原始图片进行水平翻转、垂直翻转、旋转等数据增强操作，得到多张图并分别预测，再对多个结果进行综合，进而得出最终输出结果，如图 4-14 所示。在测试时，利用 TTA 能够更好地增强模型预测效果，提高模型预测稳定性。因此，本文将采用 TTA 预测策略提高模型的岩性识别与分类能力。

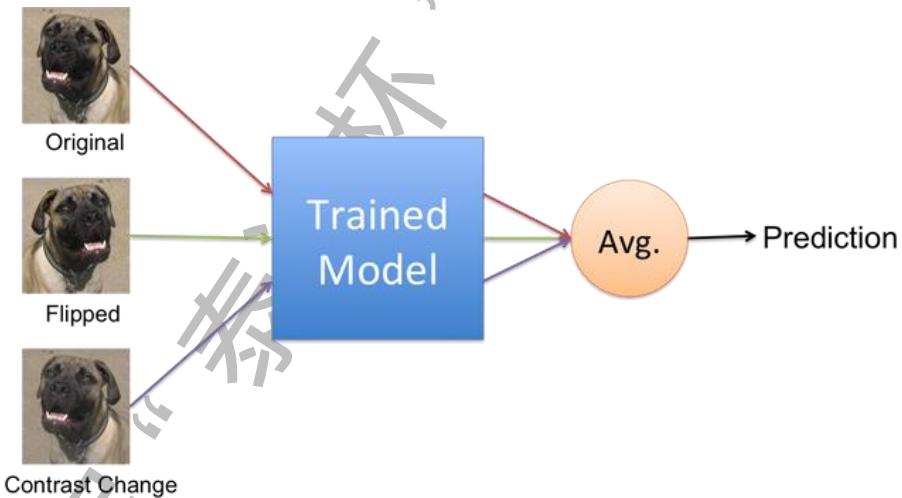


图 4-14 TTA 示意图

### 4.3 训练策略

#### 4.3.1 数据集划分

针对本章的岩石岩性识别与分类任务，将利用所提供的数据集中的白光环境图像数据进行模型训练、验证及测试。

经过多次模型的反复训练，发现训练数据的多少并不影响模型的识别能力，

即训练数据多时的模型分类准确率与训练数据少时的模型分类准确率并无明显差异，受启于 CBLoss [6]一文，这或许是因为样本中的“有效样本数”较少所导致，即样本数量上的“多”并不对应图像信息上的“多”，因此过多的训练集无法从实质上给模型提供过多的有效信息。

故此减少训练集数量以加快模型训练速度。本文通过分层抽样，将 315 张白光环境图像数据划分为 152 张训练集、93 张验证集和 70 张测试集。其中，训练集用于模型训练，验证集用于判断模型是否完成训练及判断过拟合点，测试集用于说明模型的岩性识别与分类能力。

#### 4.3.2 Adam

在优化器选择方面，选用 Adam [7]优化器，Adam 由 Diederik 等人于 2015 年提出，其结合了 Monument、AdaGrad 和 RMSProp 三种优化算法的优点，通过计算梯度的一阶矩估计和二阶矩估计，为不同的参数设计独立的自适应性学习率，具有实现简单、计算高效、对内存需求少、适用于不稳定目标函数等优点。

#### 4.3.3 OneCycleLR

在学习率调整方面，选择使用 OneCycleLR [8]方法，即在初始训练过程先提高学习率以让模型学习地更快，在训练到某一阶段时缓慢降低学习率以让模型找到全局最优解，学习率曲线示意图如图 4-15 所示。

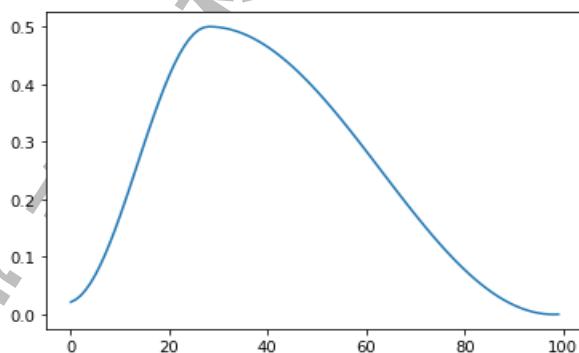


图 4-15 OneCycleLR 学习率曲线

#### 4.3.4 Focal Loss

经数据探索得出数据存在类别不平衡问题的结论，故此使用 Focal loss 作为损失函数，以解决类别不平衡问题。Focal Loss [9]由 Facebook 于 2017 年提出，在交叉熵损失函数（式 4-3-1）基础上引入超参数 $\gamma$ ，以减少易分类样本的权重，使得模型在训练时更专注于难分类的样本（式 4-3-2）。

$$CE(p, y) = CE(p_t) = -\log(p_t) \quad (4-3-1)$$

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \quad (4-3-2)$$

式 4-3-1 和式 4-3-2 中,

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases}$$

## 4.4 模型融合

考虑到不同的模型训练方式、不同的骨干网络会得到不同的模型识别能力，即前文 4.1 节和 4.2 节得到的三个不同模型在进行岩性识别与分类时，所关注的侧重点（图像区域）会有所不同，进而对于不同岩性类别的岩石分类能力也有所差异、各有所长。

因此为得到更为鲁棒、准确的岩性分类结果，本文基于取长补短、群策群力的原则，借助集成学习的思想，在结果测试时对三个模型的预测结果采用“硬投票（Hard Voting）”的方式输出得到最终的预测结果，其示意图如图 4-16 所示。

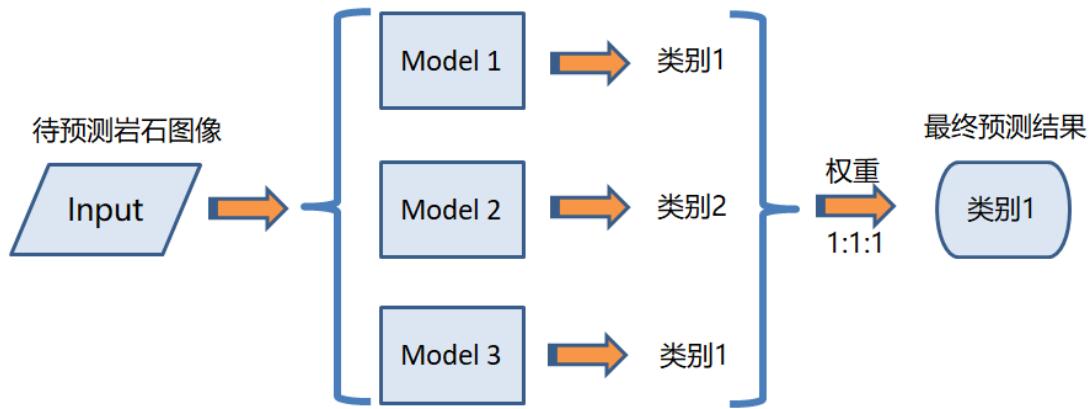


图 4-16 模型融合示意图

若是出现三个模型的预测类别完全不一致的情况，即三个模型预测出三种岩石类别，则采取随机原则，随机选择其中一个模型的预测结果作为最终的集成结果（此类情况极少，在所用 70 张测试集数据中仅出现一例）。

## 5 岩石含油比例计算

在紫外线的照射下，石油会发出绿色或黄色的光芒，这是石油的发光特征。因此可以借助荧光环境下的图像数据，计算图像中黄绿色光点面积占岩石总面积的百分比，从而估计岩石的含油量。

由 2.3 节知，提供的数据中以  **bmp**  格式存储的照片无明显背景，岩石能够占满整幅画面，可以近似地把照片面积看作岩石面积；以  **jpg**  存储的照片有明显的蓝色背景，需要设计算法将蓝色背景剔除。因此对两种格式图像数据分别设计算法计算含油比例，技术路线图如图 5-1 所示。

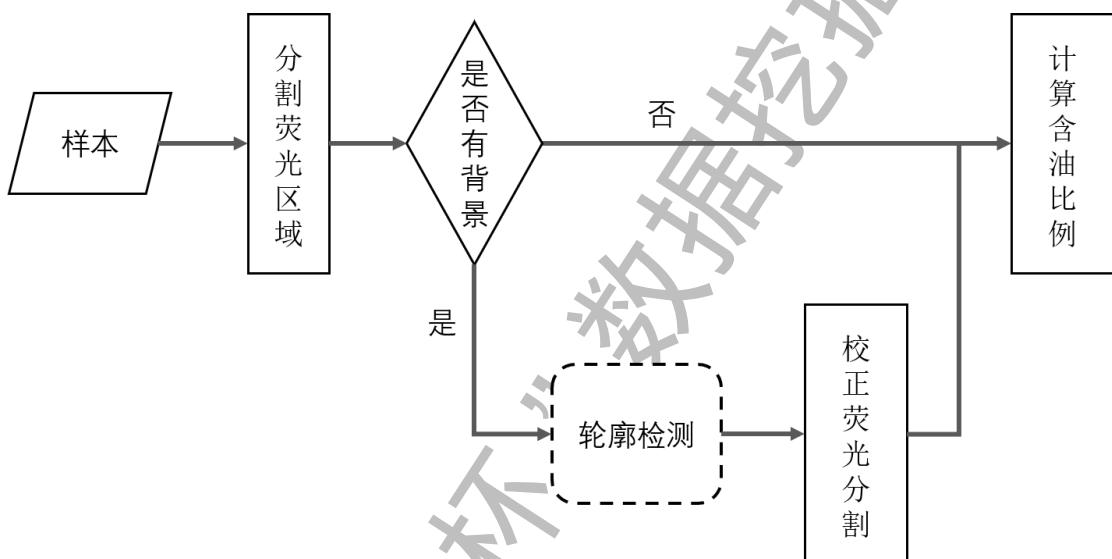


图 5-1 含油比例计算流程图

### 5.1 荧光区域分割

#### 5.1.1 颜色空间转换

在计算机中，照片以矩阵的形式存储，其尺寸可记为  $(W, H, C)$ ，其中  $W$  和  $H$  即照片的宽和高，以像素为单位，而  $C$  表示通道数。使用目前应用最广泛的 RGB 颜色模式存储的照片的通道数  $C = 3$ 。这是因为 RGB 颜色模式通过混合红、绿、蓝三原色模拟其他的颜色，故需要三个通道分别表示对应三原色之一的强度。RGB 颜色模式示意如图 5-2 所示。

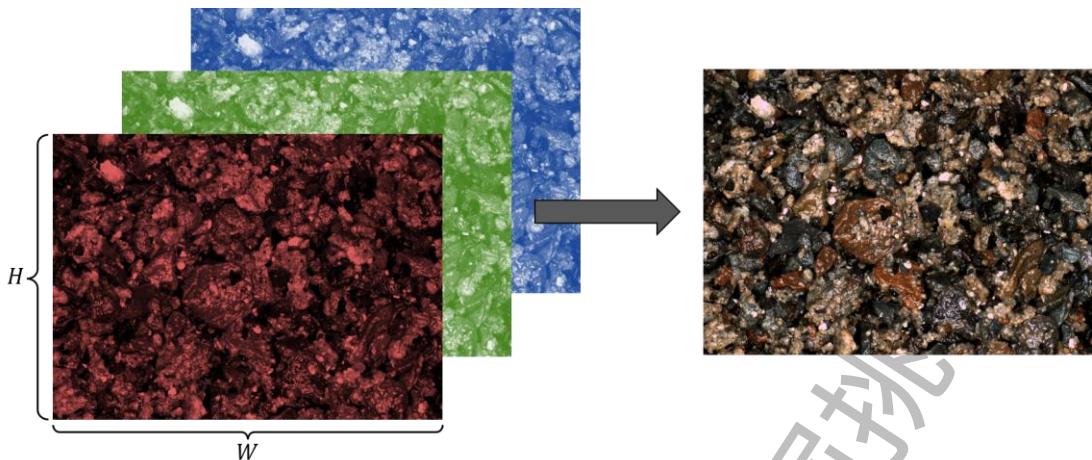


图 5-2 RGB 颜色模式示意图

使用 RGB 标准几乎可以模拟人类肉眼可以感知到的所有颜色，但 RGB 色彩空间内的颜色大多是不连续的，难以通过限定色彩范围提取具有指定颜色的像素。因此，首先将 RGB 图像映射至 HSV 空间——一种依靠色调（Hue）、饱和度（Saturation）、明暗程度（Value）表示颜色的标准。借助 HSV 空间，可轻松地将颜色与色调（Hue）、饱和度（Saturation）、明暗程度（Value）的关系表示为下图 5-3 中的三维图像。可见在 HSV 颜色空间里，相近的颜色往往是相邻的，只需要规定 H、S、V 三轴的取值范围，即可筛选出想要的颜色。

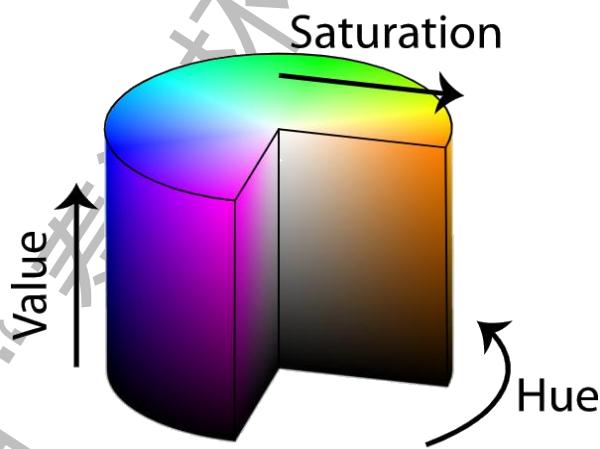


图 5-3 HSV 颜色空间示意图

### 5.1.2 石油荧光分割

根据上述原理编写程序，效果如图 5-4 所示。其中最左侧的窗口中的白点即表示当前色彩范围提取到的石油光斑，中间窗口显示了荧光灯下拍摄的原照片。通过拖动右侧窗口的滑块，可以调节色彩空间范围，从而观察并分析 HSV 色彩

范围对石油光斑分割的效果。为了减少石油的误判和漏判，经过大量实验分析，最终确定石油光斑颜色在 HSV 空间的下限为(26,43,47)，上限为(77,255,255)。

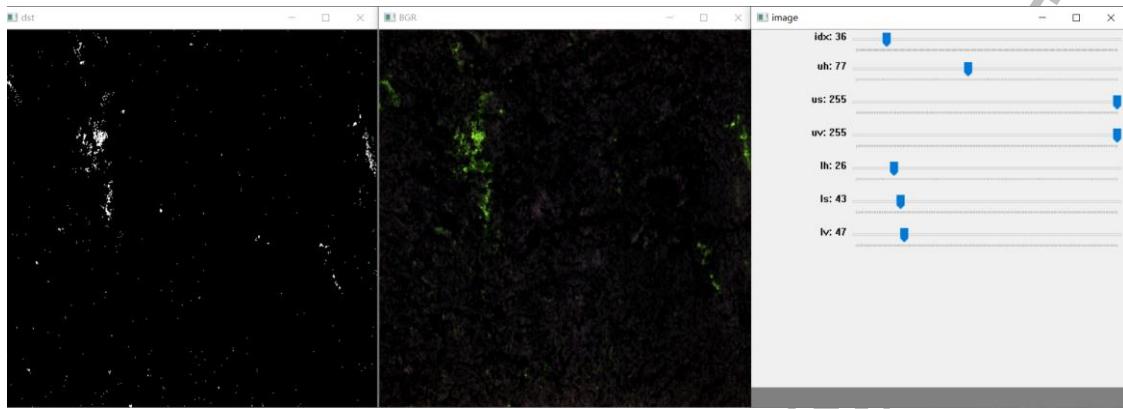


图 5-4 颜色提取示意图

## 5.2 轮廓检测

对于有明显背景的图片，需要设计算法找出岩石所在的区域，从而剔除背景，正确计算岩石面积。针对数据特点，本文设计了一个准确高效的轮廓检测流程。该流程可分为图像降噪、Canny 边缘检测、形态学处理、轮廓检测四个步骤，技术路线图如下图 5-5 所示。

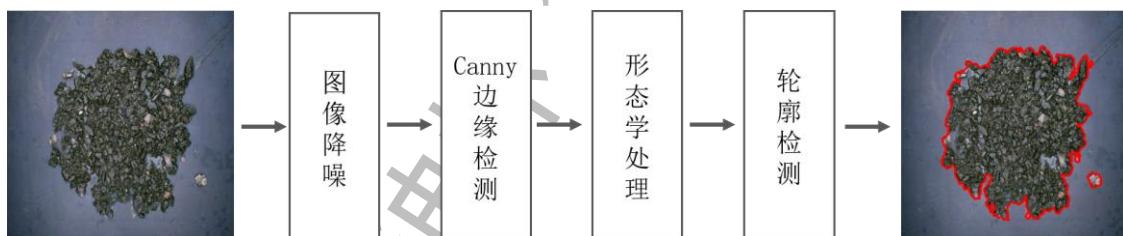


图 5-5 轮廓检测流程图

### 5.2.1 图像降噪

由于所提供的图像数据是高分辨率的，包含许多细节纹理信息，然而任务目标为轮廓检测，只需要关注岩石堆的边缘即可，岩石内部的细节信息只会对轮廓检测带来负面影响。因此使用 $3 \times 3$ 大小的高斯滤波器对图片迭代处理两次，效果示意图如下图 5-6 所示。

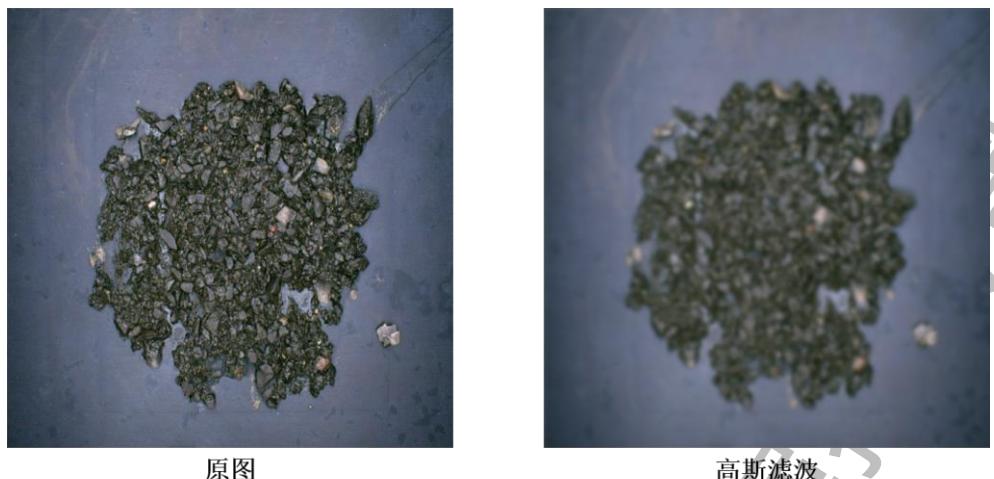


图 5-6 图像降噪示意图

### 5.2.2 Canny 边缘检测

Canny 边缘检测算法 [10]是目前最为常用的边缘检测算法之一。该算法于 1986 年由 John F. Canny 提出，具有低漏判率、低误判率等优良性质。Canny 边缘检测算法需要设定两个阈值超参数 $t_1$ 和 $t_2$ ( $t_1 < t_2$ )，步骤如下：

- STEP1：图像灰度化；
- STEP2：计算图像中每个像素点的梯度强度和方向；
- STEP3：通过非极大值抑制（Non-Maximum Suppression），得到初步的边缘划分；
- STEP4：删除梯度低于 $t_1$ 的边缘，把梯度位于 $t_1$ 和 $t_2$ 之间的边缘记为弱边缘；
- STEP5：抑制孤立的弱边缘。

根据上述五个步骤即可得到图像可能的边缘，效果示意图如下图 5-7 所示：

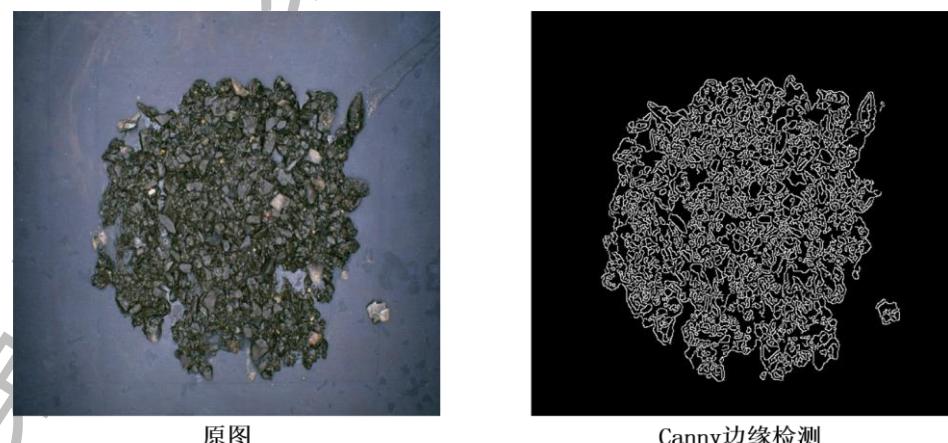


图 5-7 Canny 边缘检测示意图

### 5.2.3 形态学处理

通过 Canny 边缘检测得到粗糙的岩石边缘，并对图像进行放大后仔细观察，发现这样提取到的岩石边缘的一些部分是非闭合的，会影响轮廓检测的准确性。因此需要对提取出的边缘进行额外的形态学处理。

形态学图像处理是一种使用结构元(Structuring Elements)处理图像的技术，类似于滤波或卷积。其中腐蚀、膨胀、开操作、闭操作是形态学图像处理的基本操作。为了封闭岩石的边缘形成轮廓，使用尺寸为 $3 \times 3$ 的全1矩阵作为结构元，对边缘进行形态学处理，效果如下图 5-8 所示。可见经处理后，边缘形成了封闭的轮廓，便于进行进一步的检测。

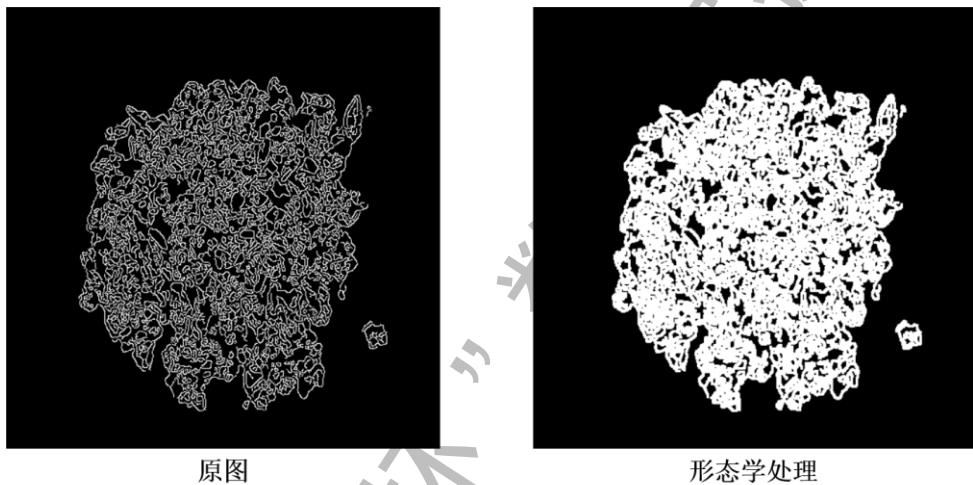


图 5-8 形态学处理效果示意图

### 5.2.4 轮廓检测

选用 Suzuki [11]提出的基于拓扑结构分析的轮廓检测方法，对前文提取的边缘特征进行处理。除了图像最外围的一圈像素以外，每个像素点周围都有 8 个像素点，这被称为该像素的临近像素。Suzuki 方法通过迭代判断像素点与其临近非零像素（边缘）的拓扑学关系，从而推断像素是否属于轮廓的一部分。这一方法不需要设定超参数，简单易用，同时具有快速、稳定、准确等优良特性，是一种广泛应用于工业的轮廓检测算法。该算法的效果示意图如下图 5-9 所示。

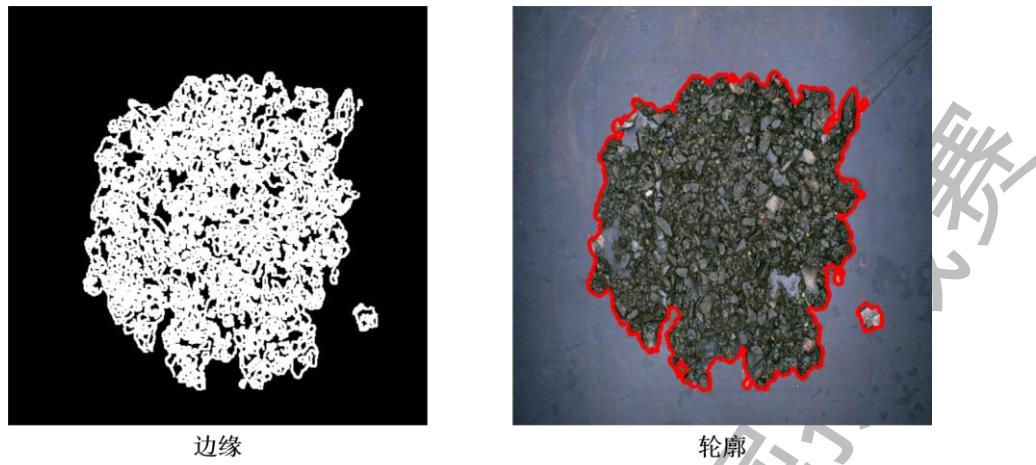


图 5-9 轮廓检测示意图

### 5.3 计算岩石含油比例

如前文所述，将对两种不同类型的照片使用不同的处理方法，从而尽可能准确地计算岩石含油比例。为了方便描述，将图像数据表示为像素点的集合 $\mathbb{G} = \{p_{ij}, p_{12}, \dots, p_{WH}\}$ ，其中 $p_{ij}$ 表示照片中第*i*行*j*列的像素。故经 5.1 节分割得到的荧光区域可表示为集合 $\mathbb{B}$ 。此外，定义运算 $|\mathbb{G}|$ ，表示集合 $\mathbb{G}$ 中包含元素的数量。

对于无明显背景的照片，岩石面积几乎等同于整张图像的面积，岩石含油比例 $p$ 可以通过下式计算：

$$p = \frac{|\mathbb{B}|}{|\mathbb{G}|}$$

对于有明显背景的照片，首先使用 5.2 节描述的技术寻找岩石的外轮廓，得到岩石区域 $\mathbb{C}$ 。可以写出岩石含油比例计算公式：

$$p = \frac{|\mathbb{B} \cap \mathbb{C}|}{|\mathbb{C}|}$$

其中 $\mathbb{B} \cap \mathbb{C}$ 表示对 5.1 节分割得到的荧光区域 $\mathbb{B}$ 与 5.2 节寻找到的岩石区域 $\mathbb{C}$ 取交集，从而剔除轮廓外检测到的荧光内容。此运算对荧光分割结果做出了校正，使得岩石含油比例不再受背景影响，从而得到更加准确且稳定的结果。

## 6 实验结果分析

### 6.1 实验环境

实验的硬件环境、软件环境分别如表 6-1 和表 6-2 所示。

表 6-1 实验硬件环境

Item	描述
GPU	2*GeForce RTX 2080Ti
CPU	Intel(R)Core™i9-9900Kcpu@3.60GHz
显存	22GB
内存	64GB

表 6-2 实验软件环境

Item	描述
操作系统	Ubuntu 18.04
编程语言	Python 3.7
CUDA	Cuda 10.1.243
深度学习框架	PyTorch 1.6.0

### 6.2 评价指标

针对岩石样本的岩性识别与分类任务（即任务一），采用准确率 **accuracy** 进行评价：

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

其中，**TP** 表示预测为正、实际也为正的次数，**TN** 表示预测为负、实际也为负，**FP** 表示预测为正、而实际为负的次数，**FN** 预测为负、而实际为正的次数。

### 6.3 实验结果

#### 6.3.1 任务一

通过第四章一系列的模型训练、验证过程，得到最终的模型岩石岩性识别与分类的结果，并为更详细地比较各模型的性能优劣，按照岩石岩性类别将模型岩石岩性分类准确度汇总至下表 6-3。

表 6-3 按类别的模型分类准确度

类别	全局-局部 模型	EfficientNet -B3	MobileNet -V3	Ensemble -Model
浅灰色细砂岩	100.00%	100.00%	96.15%	100.00%
深灰色泥岩	75.00%	33.33%	66.67%	66.67%
深灰色粉砂质泥岩	33.33%	66.67%	50.00%	66.67%
灰色泥质粉砂岩	33.33%	100.00%	33.33%	66.67%
灰色细砂岩	100.00%	100.00%	100.00%	100.00%
灰黑色泥岩	88.89%	88.89%	100.00%	88.89%
黑色煤	62.50%	50.00%	50.00%	50.00%
合计	81.43%	78.57%	80.00%	82.86%

其中，Ensemble-Model 由全局-局部模型、EfficientNet-B3、MobileNet-V3 三个模型通过“硬投票”方式集成而来（详见 4.5 节）。

通过上表 6-3，可以发现“全局-局部模型”在“浅灰色细砂岩”和“灰色细砂岩”上的识别效果非常好，达到 100.00% 的准确率，但在“深灰色粉砂质泥岩”和“灰色泥质粉砂岩”两个类别上的识别能力较差，仅为 33.33%。而“EfficientNet-B3”在“深灰色粉砂质泥岩”和“灰色泥质粉砂岩”两个类别上的识别能力比较好，分别为 66.67% 和 100.00%，恰好能够一定程度上弥补“全局-局部模型”的识别能力缺陷，因此可在集成后实现“取长补短”的效果，这便是本文选择做模型融合的重要原因。

另一方面，“EfficientNet-B3”和“MobileNet-V3”虽然都是基于网格裁剪数据所训练出来的，但是由于主干网络的差异，其在各岩石类别的识别能力也有所不同（这一点可以通过表 6-3 得到证实），因此能够达到进行模型融合的前提，即单个模型之间存在差异、各有优缺。

也正如我们所设想的一样，集成后的模型结果明显优于单个模型的分类结果，并且在各类别上的识别准确度都有明显优势，并且表现的更为均衡，达到了我们所希望的鲁棒且准确的分类效果。但与此同时，受单个模型的识别能力影响——三个模型在“黑色煤”这一类别上的识别能力都比较差，也造成了集成后的模型对“黑色煤”的识别效果较差。

进一步地，可视化上述模型各自的混淆矩阵，如图 6-1 所示，其中的横纵坐标类别（0,1,…,6）分别对应着上表 6-3 “类别”一栏的顺序。

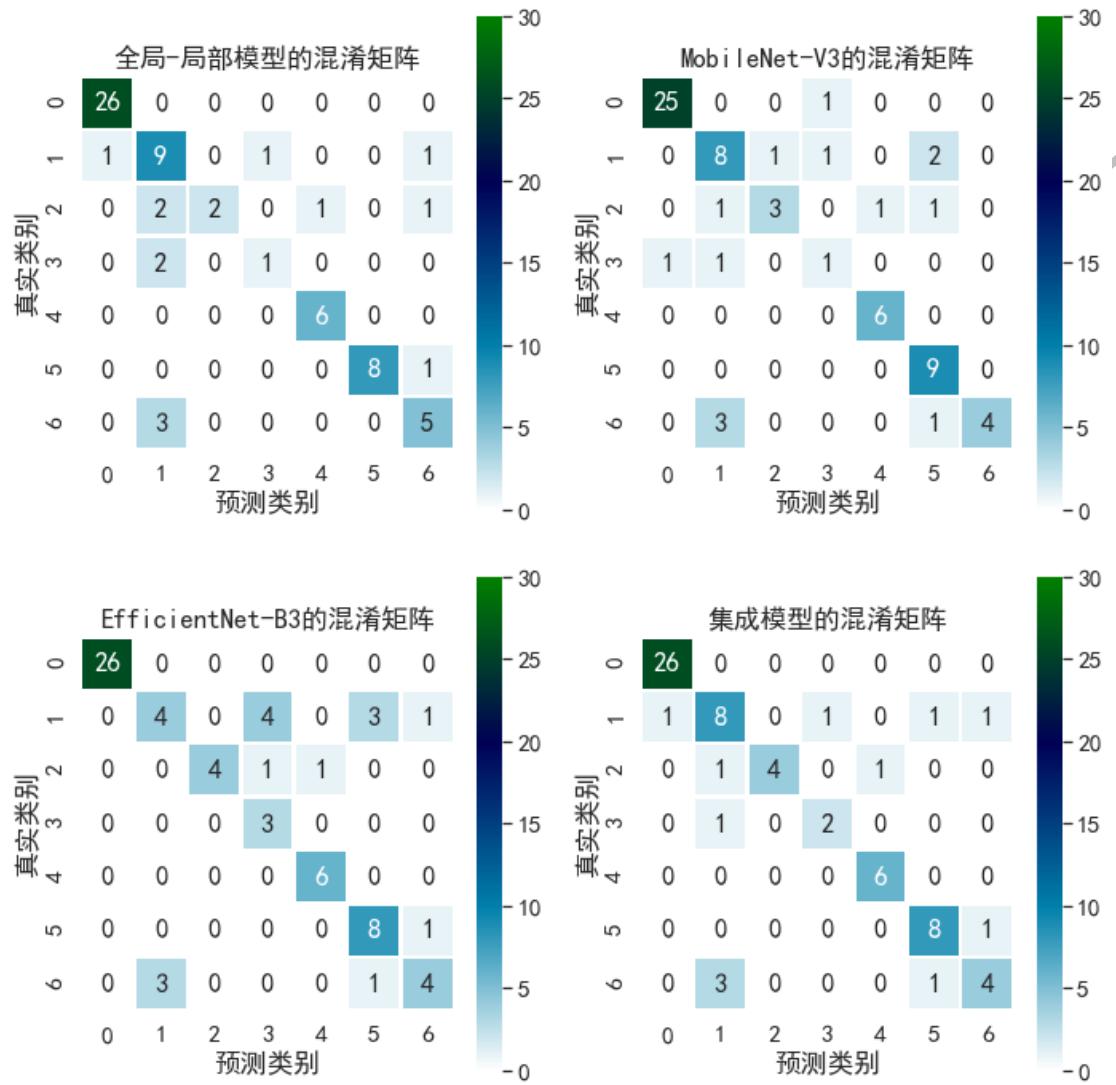


图 6-1 混淆矩阵

通过图 6-1 不难发现四个模型存在一个共性：“黑色煤”经常被错误预测为“深灰色泥岩”。同时，不知是巧合与否，被错误预测的“黑色煤”样本均为 jpg 格式的图像数据（如下图 6-2 第一行所示）。为此，找到同为 jpg 格式但被正确预测的“深灰色泥岩”图像数据（如下图 6-2 第二行所示）以及唯一一张被错误预测为“黑色煤”的“深灰色泥岩”图像数据（如下图 6-2 第三行所示）进行比较，分析其原因。

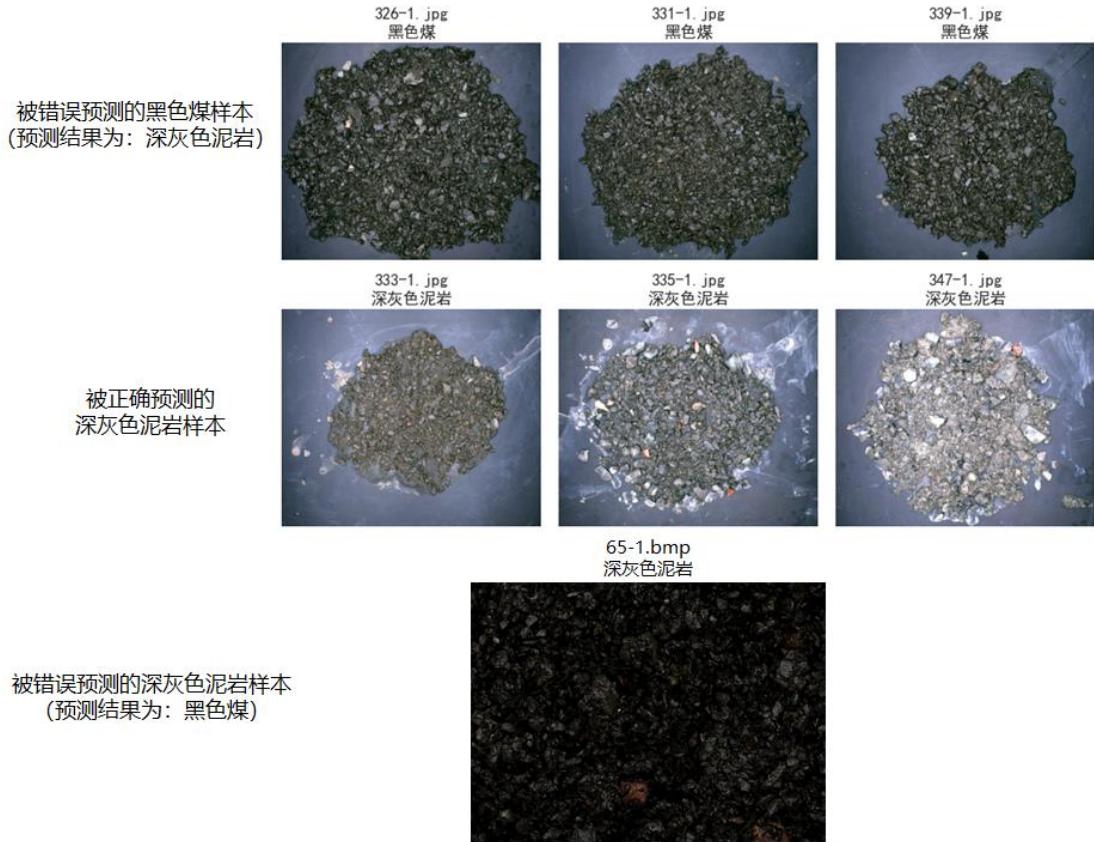


图 6-2 预测数据整体示例

通过上图 6-2 被错误预测的黑色煤样本和被正确预测的深灰色泥岩样本的比较可以发现，被错误预测为深灰色泥岩的黑色煤样本的整体颜色偏深，而被正确为深灰色泥岩的样本的整体颜色偏淡（包括背景颜色），除去视觉上的颜色差异，二者在形状、纹理、大小等特征上的差异极小。

同时，被错误预测为黑色煤的深灰色泥岩样本（上图 6-2 第三行）在肉眼看来，无论是颜色上，还是形状、纹理、大小等方面，的确和黑色煤样本更为相似（这里不排除是因为标签错误的原因），这或许就是模型“错分”65-1.bmp 样本的根本原因。

为更清晰地比较其内在的岩石纹理、形状、大小等特征，选取展示上述图像的[700:1000,700:1000]区域，如下图 6-3 所示。

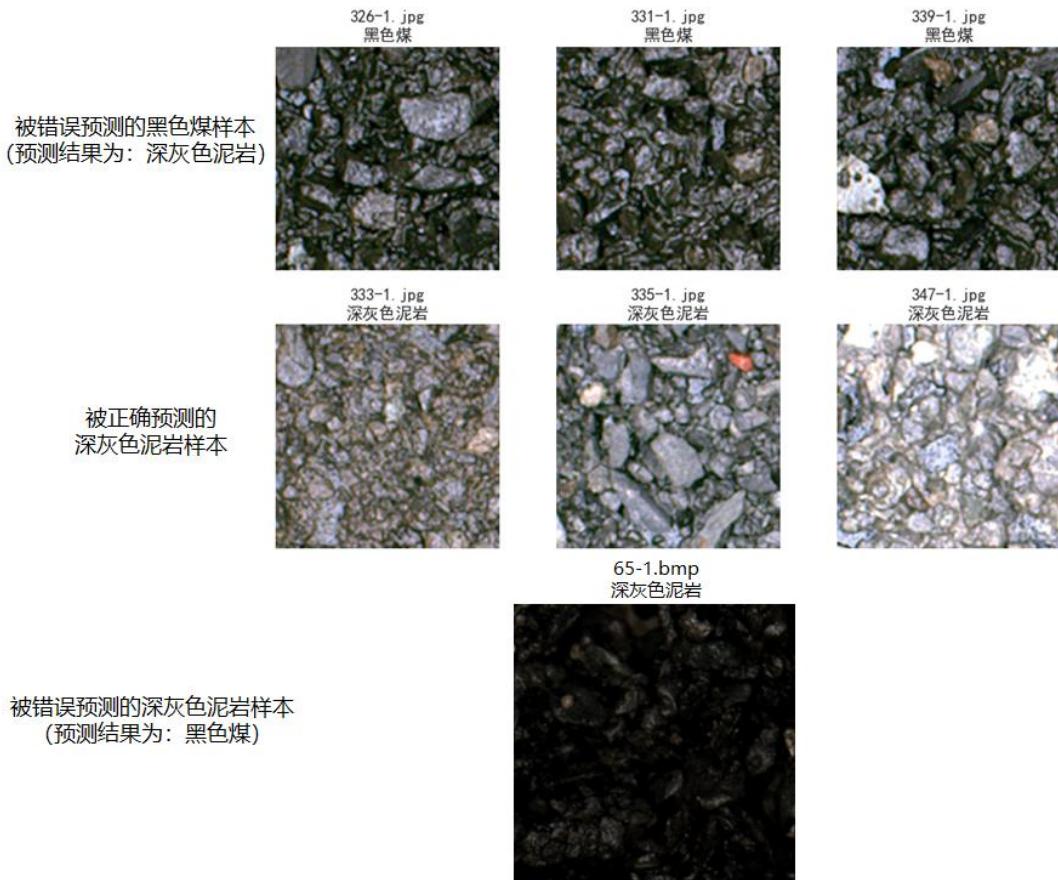


图 6-3 预测数据局部示例

通过上图 6-3 更为局部的细节比较, 可以发现 7 张图像数据的光照条件存在明显差异, 这会在一定程度上影响图像的颜色深浅程度表现, 导致光照条件好的图像看起来颜色偏淡, 而光照条件差的图像看起来颜色偏深。故此时的图像特征“颜色”不应是岩石岩性分类的第一优选项, 或可能成为误导模型分类的干扰项, 因此, 此时模型便不再从“颜色”这一维度切入, 去进行岩性的分类。

若是抛开颜色这一特征, 从其他特征方面去进行分析, 这将进一步证实了我们的猜想——除去视觉上的颜色差异, 二者在形状、纹理、大小等特征上的差异极小。所以, 这将成为模型分类岩石的重大困难, 也因如此, 模型选择将上述三个“黑色煤”样本预测为相较更有把握的“深灰色泥岩”。

综合来看, 最终所获得的模型性能还是比较好的, 其能够在数据质量不佳的情况下, 实现 80%以上的分类准确率, 已然具备一定的岩石岩性识别与分类的能力。饮水思源, 数据便是模型的根本和源头, 脱离数据, 再强大的模型和训练策略都是空谈。若是数据质量得到提高, 模型分类能力想必能够获得进一步提升, 实现质的飞跃。

### 6.3.2 任务二

通过荧光区域分割、演示轮廓检测等步骤（详见第 5 章），得到各样本的含油面积百分比，部分结果如表 6-4（详细结果见上传附件）。

表 6-4 任务二部分结果

样本编号	油含量面积百分比
1	0.94%
2	0.05%
...	...
349	0.30%
350	2.11%

以编号 270 的样本为例，对荧光区域分割和轮廓检测结果进行可视化分析，如图 6-4 所示。可见本文所用方案得到的荧光区域十分准确，即使是微弱的荧光也可以被识别。

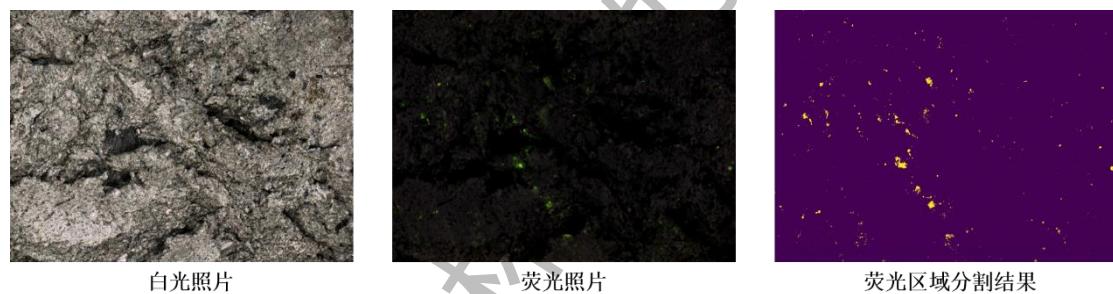


图 6-4 石油荧光分割样例

对于存在明显背景的图像，需要额外进行岩石轮廓检测，以精准确定岩石实际所占面积，保证所计算的含油面积百分比的准确性。以编号为 342 的样本为例（图 6-5），可见所提取岩石区域面积没有受到背景上的白色污渍干扰，能较好地分割出正确的岩石区域，从而保证含油量结果更贴合实际情况。



图 6-5 轮廓检测及荧光分割示例

## 7 总结与展望

### 7.1 总结

本文介绍了当前岩石样本探测的重要意义及其现有的探测手段，阐述了岩石分类智能化的研究价值。通过图像处理技术和深度学习算法，结合数据特点，设计出有效的模型，识别出岩石样本的岩性类别及含油气情况，实现岩石样本智能识别分类。所做工作包括以下方面：

#### 一、任务一

1. 针对图像数据进行探索性分析，发现数据存在类别不平衡、标签错误、图像信息杂乱、数据格式差异等问题；
2. 针对数据探索发现的问题采取针对性策略，包括使用 Focal Loss 损失函数解决类别不平衡问题、利用置信学习剔除标签错误数据、借用裁剪图像的颜色占比减少类别外的干扰信息、通过轮廓检测和闭环处理精确 JPG 格式图像的岩石面积等；
3. 针对训练样本数过少的问题采用多种数据增强方法，并利用图像高分辨率特性进行随机裁剪和网格裁剪，得到实际所用训练数据，并基于两种不同的裁剪策略搭建不同的分类模型，使得不同的模型具备不同的识别能力，以满足模型融合的前提；
4. 针对随机裁剪得到的训练数据，基于 ResNeXt-50 进行改进，分别搭建识别全局信息和局部信息的架构，提出“并联式”模块融合全局信息和局部信息，有效解决“串联式”模块存在的问题，有效提升了岩石分类准确性；
5. 针对网格裁剪得到的数据，通过置信学习剔除错误标签数据以提高数据质量，再将筛选后的数据分别输入到以 EfficientNet-B3 与 MobileNet-V3 为主干的模型中进行训练，在测试阶段结合 TTA 策略进行滑动窗口预测，通过单张图像多个区域的预测结果进行投票得到最终预测结果，有效避免图像信息杂乱等问题带来的类别预测干扰；
6. 针对不同训练策略的三个模型进行“硬投票”集成获得更为鲁棒的岩石分类效果，集成后的模型结果明显优于单个模型的分类结果，并且在各类别上的识别准确度都有明显优势，并且表现的更为均衡，实现所希望的鲁棒且准确的分类效果目标。

## 二、任务二

1. 考虑到数据格式差异问题，将任务目标“计算岩石含油面积百分比”分解为荧光区域分割、岩石区域分割两个步骤；
2. 针对荧光区域的面积计算，选择将图像映射至 HSV 空间，利用交互式的可视化方法确定最佳阈值，由此根据 HSV 空间的色调、饱和度以及明暗程度快速地对荧光区域进行分割，计算分割出的荧光区域面积；
3. 针对岩石区域的面积计算，考虑到 JPG 格式图像的背景问题，即图像大小并非岩石的实际面积大小，故应勾勒出岩石实际所在区域，使得所计算出的岩石面积更贴近于实际。对此，首先采用高斯滤波器对数据进行降噪处理，以减少岩石内部的细节信息对轮廓检测的影响，再者根据 Canny 边缘检测勾勒出岩石的粗糙边界，最后通过形态学图像处理的方法，对 Canny 检测中未封闭的区域进行封闭处理，根据所分割出的荧光面积和岩石面积，计算实际的岩石石油含量面积百分比。

## 7.2 展望

由于缺少地质学有关专业知识，在所提供数据存在标签错误问题的情况下，只能依靠一些简单的观察和带噪学习算法剔除部分错误数据，这不能完全保证所训练数据或测试数据的标签一定正确。

故此，可能导致出现模型实际预测正确，却因数据被错误标记，而使得模型从结果上看起来预测“错误”的情况，致使所计算出的岩石岩性分类准确率比实际偏低（如 6.3.1 节中所示的 65-1.bmp 样本）。

因为时间有限，该任务仍有诸多思路未能够得到一一尝试和比较，仍存在诸多细节需要继续完善和改进。期待未来能有更高质量的数据图像及标签，本文所用模型和方法应该能够达到更好的效果。

## 参考文献

- [1] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [2] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017: 5987-5995.
- [3] Tan, M. and Le, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. ICML, 2019a.
- [4] Andrew Howard, Mark Sandler, Grace Chu, et al. Searching for MobileNetV3. ICCV, 2019.
- [5] Moshkov, N., Mathe, B., Kertesz-Farkas, A. et al. Test-time augmentation for deep learning-based cell segmentation on microscopy images. Sci Rep 10, 5068 (2020).
- [6] Yin Cui, Menglin Jia, Tsung-Yi Lin, et al. Class-Balanced Loss Based on Effective Number of Samples. CVPR, 2019.
- [7] Diederik P., Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization. ICLR, 2015.
- [8] Leslie N. Smith, Nicholay Topin. Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates. arXiv:1708.07120, 2017.
- [9] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár. Focal Loss for Dense Object Detection. ICCV, 2017.
- [10] Canny J. A computational approach to edge detection[J]. IEEE Transactions on pattern analysis and machine intelligence, 1986 (6): 679-698.
- [11] Suzuki S. Topological structural analysis of digitized binary images by border following[J]. Computer vision, graphics, and image processing, 1985, 30(1): 32-46.