

队伍编号	MCB2303548
赛道	A

基于纹理和边缘特征图的道路坑洞分类策略

摘要

道路缺陷的存在不仅对个人和社会的资源支出造成了间接影响，还可能导致严重事故和人员伤亡。因此，坑洼道路的检测和识别成为一项极其重要的计算机视觉任务。该任务旨在利用数字图像技术来识别道路上的坑洼，这一技术具有广泛的应用领域，包括地质勘探、航天科学和自然灾害研究等。近年来，随着深度学习技术的迅速发展，为解决坑洼道路检测提供了全新的解决方案。相较于传统的分类算法，深度学习技术在处理复杂和多变的坑洼图像特征方面更具优势。本文中，我们提出了一种在道路图像中使用边缘检测的坑洼道路分类模型。问题重点在于识别哪些特征及其相应的特征组合适合对图像进行有效分类，以确定图像是否包含凹坑。

针对问题一，我们首先对训练集进行预处理，对大小不一的图片均处理成像素大小相同的图片以及对于标签不平衡问题进行数据增强的处理，在本研究中，我们采取了一种创新的预处理方法，将目标检测模块集成到分类模型的预处理步骤中，对干扰项物体识别并进行将像素值改变为 255 的“背景化”操作，随后进行了去噪处理。为检测凹坑的边缘、形状和纹理特征，我们分别用图像特征提取方法 HOG、LBP 以及 RFC 边缘检测方法对训练集提取相应的特征，随后对这些特征构建了两种坑洼分类模型，一种是特征级融合的集成学习分类模型，另一种是融入了通道注意力机制模块的深度卷积神经网络分类模型。

针对问题二，对于集成学习分类模型，我们可以从子模型各自的参数改进入手对初始模型进行优化。通过网格搜索方法或者绘制学习曲线，以在参数空间中搜索最佳设置。对于深度卷积神经网络模型，以问题一中训练得到的模型架构和图像输入的大小作为初始参数，在初始参数的基础上对关键参数分别进行微调，采用控制变量法进行消融实验，用模型准确率作为模型精度进行调整前后模型效果的纵向比较并定量评估，同时即可得知关键参数的调整对识别精度的影响。另一方面，这两个分类模型进行横向比较，对模型的有效性和坑洼分类结果进行了准确性评价。结果表明：我们的最优模型在验证数据集上的准确率达到了 88.52%。

关键词： 道路坑洼；边缘检测；集成学习；目标检测；InceptionResNetV2

目录

1 问题重述	4
1.1 问题背景	4
1.2 问题重述	4
2 问题分析	4
2.1 问题一的分析	4
2.2 问题二的分析	5
2.3 问题三的分析	6
3 模型假设与符号说明	6
3.1 模型假设	6
4 模型的构建	6
4.1 数据预处理	6
4.1.1 训练集数据增强预处理	6
4.1.2 使用灰度降维	6
4.1.3 使用目标检测的图像预处理	7
4.1.4 去噪	8
4.2 特征提取	9
4.2.1 边缘特征提取	9
4.2.2 HOG 和 LBP 的特征提取	9
4.3 模型的引入	10
4.3.1 集成学习模型	10
4.3.2 KNN	10
4.3.3 支持向量机	13
4.3.4 MLP 算法	13
4.4 深度卷积神经网络模型分类器	13
4.4.1 特征融合	14
4.5 集成学习模型分类器	15
5 模型的求解及优化	15
5.1 深度卷积神经网络分类模型训练及优化	15
5.1.1 模型训练	15
5.1.2 优化	16
5.2 集成学习模型	16
5.2.1 SVM 正则化的选择	16
5.2.2 网络融合策略优化	18
6 模型的评价	19
6.1 模型的优点	19
6.2 模型的缺点	19
6.3 模型的改进	19

6.4 模型的推广	19
7 附录	19

1 问题重述

1.1 问题背景

道路在社会的发展中扮演着至关重要的角色，作为广泛使用的交通媒介，支持着各种业务和人员的流动。然而，道路缺陷，如开裂、车辙、表面磨损和坑洞，对个人、社区、商业组织和国家实体产生了严重的负面影响。这些影响包括机械设备受损、资源支出增加、交通事故风险上升等，因此道路维护变得至关重要。

坑洞的形成通常受到水力作用的影响，通过与高速车辆的接触、雨水和其他杂物的作用，坑洞会不断扩大。这种过程类似于癌症，一个坑洞的形成可能导致同一区域内多个其他故障的产生。修复道路的成本与损坏程度密切相关，随着损坏程度的加剧，维修成本显著增加，这需要国家投入大量资金来改善国内的道路状况。

为了更有效地应对这一问题，引入自动化解决方案变得至关重要。传统的手动报告方法不仅耗时，而且容易出现不一致性。自动化检测和识别过程可以节省大量时间，实现更快的维修响应，从长远来看，有助于实现对道路维护的有效支出。这种自动化方法可以有效减少交通事故风险，降低机械设备受损的可能性，并减少资源支出。

1.2 问题重述

现已提供301幅图像，在这个问题中，我们希望通过分析、提取特征和建模已标记的道路图像，使得我们能够对一张新的道路图像进行自动坑洼状态识别。

问题1：我们需要结合提供的图像文件，提取图像的特征，并建立一个高识别率、快速、准确分类的模型，以便判断图像中的道路是正常的还是有坑洼。

问题2：然后，我们会对问题1中构建的模型进行训练，并从不同的角度进行评估，以确保模型的质量和准确性。

问题3：最后，我们将利用已经训练好的模型来识别测试集中的坑洼图像，并将识别结果记录在“test result.csv”文件中。

2 问题分析

2.1 问题一的分析

针对本问题，我们仔细研究了题目所提供的301幅图像，包括图像的大小、分辨率、颜色空间等特征，并且查看了图像的命名标签，以确定哪些图像中存在坑洼，哪些是正常的道路。对于训练集图像大小不一、目标标签不均衡和数据量偏少的问题，我们需要对这301张图片进行相应的预处理。而接下来特征提取的目标是捕获图像中的有关坑洼和道路状态的信息，以便模型可以进行分类。但是我们发现道路正常的图像里存在很多干扰项物体，包括人、汽车、树木、水印、路灯和建筑物等除坑洼之外的物体，这将使得训练出来的模型假负率往往偏高。

为达到正确分类的要求，我们采用了目标检测的方式识别这些干扰项，在分类前预处理中进行“背景化”，为了提高目标检测的效率，我们使用了YOLOv5模型并进行改进。接着我们把提取到的图像特征采用特征级融合的集成学习策略对道路是否坑洼进行投票分类达到高识别率的要求。同时，我们也提出了另一种深度卷积神经网络分类结构，结合了Inception和ResNet两种架构的思想，用于图像分类任务。对于这两种分类模型，我们在后续章节里进行了详细的对比分析。

综上所述，我们基于 YOLOv5 模型提出了一种经改进后的模型 Advanced-YOLOv5，该模型的网络架构与 YOLOv5 相似，但对 batchsize 和 dropout 这两个参数进行改进。将这目标检测模型与两种分类模型融合，并对各调整模型的精度进行分析，从而达到题目的要求。本问题的探索思路如下图 1a 和图 1b 所示：

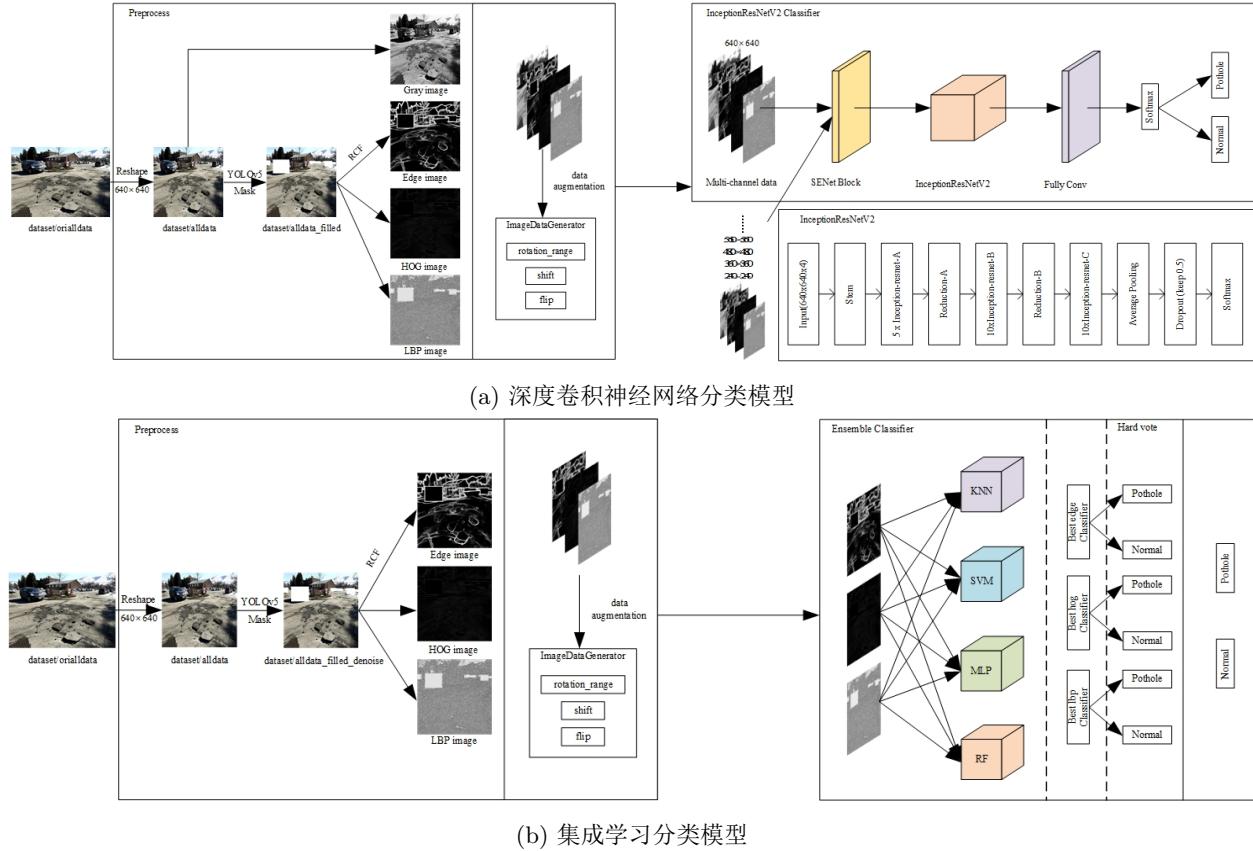


图 1：我们提出的两种分类模型。

2.2 问题二的分析

针对问题一中提出的两种分类模型，两者均运用到了多尺度特征。不同在于，集成学习模型的精度取决于每个子模型的精度。其中，随机森林的树的数目决定这分类的精度，适当的增加树的数目会提升分类的精度。SVM 有一些关键的超参数，如正则化参数（C 值）和核函数参数，会影响模型的非线性拟合能力从而影响分类准确率。MLP 的架构包括层数、每层的神经元数量、激活函数等，选择合适的神经网络架构对模型性能至关重要。K 近邻模型的分类精度受 K 值以及距离度量方法影响。而深度学习算法精度的提高依赖于每次迭代训练的切片数，是否加入通道注意力机制模块等关键参数。

这里我们建立起参数优化的模型，对于集成学习分类模型，我们可以从子模型各自的参数改进入手对初始模型进行优化。通过网格搜索方法或者绘制学习曲线，以在参数空间中搜索最佳设置。对于深度卷积神经网络模型，以问题一中训练得到的模型架构和图像输入的大小作为初始参数，在初始参数的基础上对关键参数分别进行微调，采用控制变量法进行消融实验，用模型准确率作为模型精度进行调整前后模型效果的纵向比较并定量评估，同时即可得知关键参数的调整对识别精度的影响。另一方面，这两个分类模型进行横向比较，在相同数据集上的性能、精确度、准确率等指标进行比较。本问题的分析思路如下图 2-2 所示。

2.3 问题三的分析

使用上述优化后得到的模型对测试集中的坑洼图像进行二分类预测，生成每个图像的预测结果，即判断图像中是否存在坑洼。在填写的过程中确保测试集的标号与模型的分类结果相匹配。

3 模型假设与符号说明

3.1 模型假设

- (1) 假设已知我们人工标记标签图足够准确对目标检测对象划分边界框；
- (2) 假设在神经网络预测中，局部图像特征可代表整体特征；
- (3) 假设面积、像素点误差可以作为对象标签目标检测模型的评价标准；
- (4) 假设灰度、形状大小、分布体现坑洼特性；
- (5) 模型假设提供的样本（301幅图像）是代表性的，即这些图像能够很好地覆盖整个问题领域的特征分布。如果样本不够代表性，模型的泛化能力可能会受到影响；
- (6) 假设数据中没有太多的噪声和错误标签。

4 模型的构建

4.1 数据预处理

如图2展示了我们的数据预处理的过程，我们提取了每幅去噪声后图像的 Edge、LBP 和 HOG 特征，我们就尝试并测试了使用不同特征组合的不同分类模型。

4.1.1 训练集数据增强预处理

鉴于题目所给出的训练集中共 266 张正常非坑洼道路和 35 张坑洼道路的图片尺寸大小不一，我们根据模型的输入大小，首先对训练集进行 640*640 的图像分割预处理。再通过概率翻转、仿射变化、随机区域的像素丢失、部分图片弹性形变、添加高斯噪声、明亮度改变等方式对数据集进行数据增强，将数据集扩展为原来的 3 倍。

4.1.2 使用灰度降维

在道路损坏图像中，存在多种物体，包括人、汽车、树木、水印、路灯和建筑物等除坑洼之外的物体。处理这些多通道图像需要大量的计算资源，而坑洼具有一个没有RGB颜色的特征，当大量的计算资源花费在坑洼之外的多通道图像之上，得到的模型将会失去题目原始的目的。所以为了有效分析这些图像，我们对这些多通道图像进行预处理。

首先我们对两类图像中，红色、绿色和蓝色通道的平均强度是否存在显著差异进行了数据探索，这将决定我们是否可以根据它们的平均颜色通道强度对图像进行分类。

在红色和绿色通道的分布图中，它们的平均强度差异不明显，但在蓝色通道的情况下，坑洞类的图像中蓝色通道的平均强度值较高。坑洞类图像的整体颜色通道强度高于普通类，但差异不够显著，作为分类任务的特征图可能会损坏模型的鲁棒性。

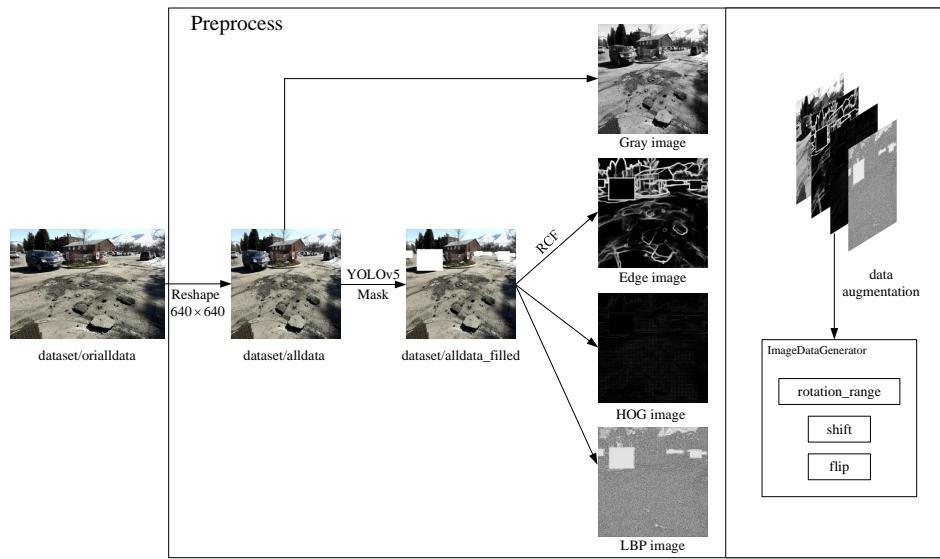


图 2：数据预处理过程。

这里我们选择将RGB三通道图像转换为单个通道，在以最大限度地减少损失的前提下保留坑洼的图像特征。预处理方法包括灰度转换、二值化等。

这里我们选择灰度方法将多通道图像转换为单个通道。在灰度处理后，图像的颜色信息被舍弃，只保留了亮度信息，这样可以减少数据量，同时在不显著损害图像形状的情况下进行转换。对于坑洼的检测和识别，我们只关心图像的亮度信息而忽略颜色信息是可以接受的，这种方法能够简化图像，使得后续的处理更为高效。

4.1.3 使用目标检测的图像预处理

经过上一步对图像灰度处理，此时在路面数据中，坑洼周围有各种各样的物体，如树木和汽车，它们已全部转换为灰色，数据量减少到三分之一但是我们很难分析其特征，容易受到非坑洼物体的干扰。所以我们接着使用目标检测算法来检测图像中的对象 [1]。

这里我们采用 YOLO 算法对图片进行目标检测。YOLO 将图像划分为网格单元，每个网格单元负责检测其内部的目标。每个网格单元负责预测包围目标的边界框（Bounding Box），包括边界框的坐标和置信度分数。它使用边界框坐标来确定目标的位置，并使用置信度分数来衡量模型对目标的置信程度。每个网格单元将输出目标的位置和类别信息。同时 YOLO 可使用多层卷积神经网络来提取不同尺度的特征。这有助于检测不同大小的目标，包括小目标和大目标。为了去除重叠的边界框和重复检测，YOLO 使用非极大值抑制来筛选最终的检测结果。

我们通过 YOLOv5 模型对灰度化的图像进行目标检测，在检测到的对象中，除去坑洼之外的对象，处于该对象的像素值将被改变为 255 并创建背景。这是因为道路和坑洞的颜色是灰色的，将背景替换为 255 可以突出道路和坑洼检测的重要性。因此，除了坑洼之外的其他物体通过预处理过程被去除 [2]。

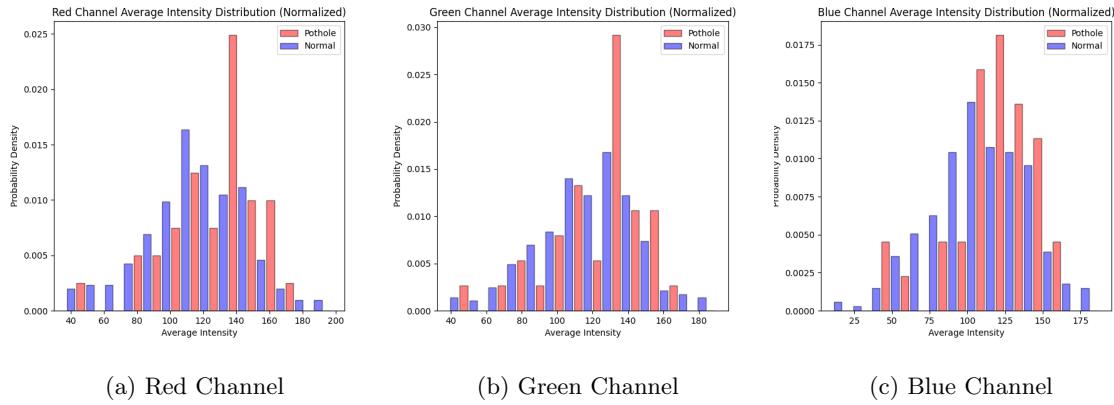


图 3: 不同颜色通道在 Pothole 和 Normal 的强度。原始数据集是不平衡数据, 所以 Normal 类的单通道累积强度会明显高于 Pothole 类, 因此我们计算每个分类下的平均通道强度。

4.1.4 去噪

由于大部分特征提取的方法对微小的纹理变化非常敏感，例如LBP，RCF等，因此噪声可能会导致较大的负面影响提取出与真实图像不一致的特征结果，从而干扰模型，降低模型的分类效果。例如下图是我们采用去噪和未去噪进行特征提取的图像，可以发现图像的LBP特征对于噪声是极其敏感的，特别是在噪声水平较高的情况下。这可能会导致特征提取结果中的虚假性和不稳定性。而具有抗噪声性能的特征提取方法，如局部特征描述符HOG，它们对局部图像区域中的噪声相对不敏感，但是也同样对模型的性能存在一定影响。

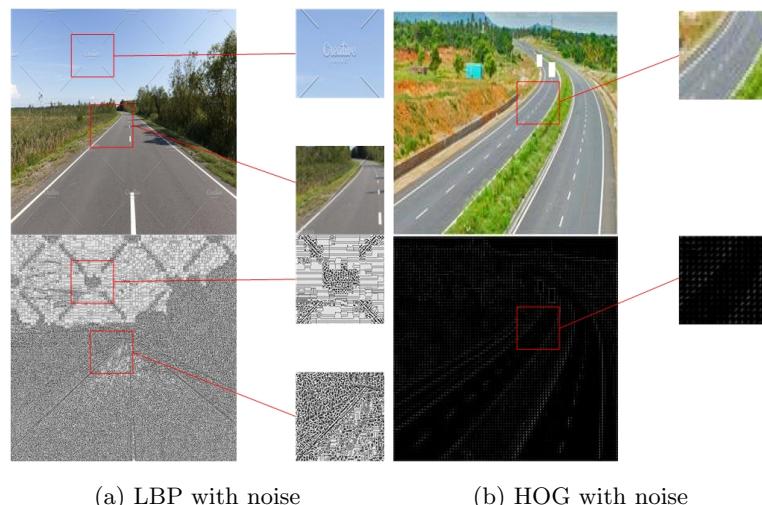


图 4: LBP 图都会受到显著的噪声影响, HOG 图则相对不那么明显。

我们在提取图像特征之前进行去噪，可以确保得到的特征更好地反映图像的本质纹理特征，而不受噪声的干扰。这里我们选择使用标准的中值滤波去噪技术，以减小噪声的影响，同时避免过度处理图像导致信息丢失。

4.2 特征提取

4.2.1 边缘特征提取

在道路损坏检测领域，提取和处理坑洞的特征至关重要。这些特征包括坑洞的边缘类型，如环形边、直线边、阶梯边和渐变边。为了实现对道路损坏的准确检测，边缘检测被广泛采用。边缘检测是一种基于偏微分运算的方法，通过提取与边缘相对应的像素，获取图像中边缘的梯度信息。这些特征和边缘信息通过第一微分和第二差分等图像处理技术进行精细处理，从而为道路损坏的自动检测提供了可靠的基础。

我们采用了RCF (Richer Convolutional Features) 模型对上一节预处理得到的图像进行特征提取。RCF模型是一种基于深度学习的边缘检测模型，它利用深度卷积神经网络 (CNN) 来学习图像的特征，并用于边缘检测任务 [3]。相比于经典的边缘检测算法，例如Sobel算子或Canny边缘检测依赖于图像的局部梯度信息来进行边缘检测，RCF模型通过深度学习技术，特别是使用了深度卷积神经网络，能够学习到更丰富、更抽象的图像特征，而不仅仅局限于梯度信息。深度学习模型可以学习到图像中的各种层次的特征，包括纹理、颜色、形状等，从而提高了对边缘的检测能力 [4]。RCF模型的深度网络结构使得它能够自动学习到适合于边缘检测任务的特征表示，而无需依赖于手工设计的规则和滤波器，相较于传统的边缘检测算法，它在复杂场景下检测和识别坑洼道路通常表现得更加准确和鲁棒。

4.2.2 HOG 和 LBP 的特征提取

方向梯度直方图 (Histogram of Oriented Gradients) 和局部二进制纹理 (Local Binary Pattern) 都是计算机和图像处理中常用的特征信号。它们都被广泛评估物体检测、人脸识别和纹理分类等任务。

HOG 是一个用于局部图像的形状信息和纹理信息的特征控件。它基于图像的梯度方向来构建直方图。HOG特征表示包含图像中对象的形状和边缘信息的向量，它分析了图像中局部梯度方向的分布。由于HOG是在图像的局部方格单元上操作，所以它对图像几何的和光学的形变都能保持很好的不变性。一般而言，HOG特征在物体检测（特别是行人检测）中非常有效，它通常与支持向量机 (SVM) 分类器一起使用。

HOG 主要原理是计算图像的梯度幅值和方向。具体来说先把图像分割成小的连续区域 (Cell)，计算每一个区域梯度方向的直方图，最后合并成更大的区域 (Block) 来对直方图进行归一化，然后合并所有 Block 形成最终的 HOG 特征图。

从图5d和图5e对应的 HOG 特征图可以观察到，HOG特征以梯度的形式捕捉了坑洞的边界，但其边界轮廓并不像它们对应的 LBP 和 Edge 特征图那么明显，相比之下，坑洼的轮廓相对模糊。但是，无论图像中是否存在任何物体，如汽车，行人，HOG 特征图都可以很好的通过其梯度轮廓定位其在图片中的位置。在坑洼图像和非坑洼图像的情况下，我们可以看到坑洼和道路图片捕获的梯度不同，如图5d和图5b。因此，我们可以初步得出结论，HOG 特征图将对我们的分类任务产生影响。

LBP 是一种用来描述图像局部特征的算子，LBP特征具有灰度不变性和旋转不变性等显著优点。LBP 主要工作原理如下：对于图像中的每个像素，考虑其周围的一个领域 (例如， 3×3 的窗口)，然后将邻域中的每个像素与中心像素进行比较。如果领域像素的值大于或等于中心像素的值，则将其标记为 1，否则标记为 0。这样相邻的所有域都会产生一个二进制模式(例如，11100101)，通过计算所有二进制模式的直方图作为图像的 LBP 特征。

从图5d中我们可以清楚地看到，坑洞区域与汽车一起突出显示。而图5a和图5b对应的 LBP 图像，由于没有坑洞，平原图像上的图案保持均匀，我们可以清楚地观察到其与坑洞的图案的差异。基于这一点，我们认为 LBP 特征图中 normal 和 pothole 所体现出来的差异性，可以胜任我们提出的分类方法。此外 LBP 特征图而言，我们认为 LBP 特征图上的车辆对于我们提出的分类一定的挑战。因为，我们发现图5d对应的

Edge 图和 LBP 图也会受到车辆形状边缘的潜在影响。基于这一点，我们提出了一种解决方法，把图像中的车辆通过目标检测模型识别出来，通过白色掩膜遮挡住，使其与坑洞形状有明显区别，以此减低对模型的要求。我们通过实验证明，通过白色掩膜能够有效的提高模型对于验证数据的分类准确率。

LBP 是一种用来描述图像局部特征的算子，LBP 特征具有灰度不变性和旋转不变性等显著优点。LBP 主要工作原理如下：对于图像中的每个像素，考虑其周围的一个领域（例如， 3×3 的窗口），然后将邻域中的每个像素与中心像素进行比较。如果领域像素的值大于或等于中心像素的值，则将其标记为 1，否则标记为 0。这样相邻的所有域都会产生一个二进制模式（例如，11100101），通过计算所有二进制模式的直方图作为图像的 LBP 特征。

从图5d中我们可以清楚地看到，坑洞区域与汽车一起突出显示。而图5a和图5b对应的 LBP 图像，由于没有坑洞，平原图像上的图案保持均匀，我们可以清楚地观察到其与坑洞的图案的差异。基于这一点，我们认为 LBP 特征图中 normal 和 pothole 所体现出来的差异性，可以胜任我们提出的分类方法。此外 LBP 特征图而言，我们认为 LBP 特征图上的车辆对于我们提出的分类一定的挑战。因为，我们发现图5d对应的 Edge 图和 LBP 图也会受到车辆形状边缘的潜在影响。基于这一点，我们提出了一种解决方法，把图像中的车辆通过目标检测模型识别出来（如图6所示），通过白色掩膜遮挡住，使其与坑洞形状有明显区别，以此减低对模型的要求。我们通过实验证明，通过白色掩膜能够有效的提高模型对于验证数据的分类准确率。

4.3 模型的引入

4.3.1 集成学习模型

在前文我们提到我们的特征提取方法，我们通过不同方法提取出图像的灰度，HOG，LBP，边缘特征。我们观察到对于处理多特征数据，集成学习多样化的分类器或许能提供不错的分类效果。

集成学习的有效性依赖于基本模型之间的多样性，旨在通过构建并组合多个预测模型来提升预测准确性和泛化能力。不同的模型可能会从数据中学习到互补的信息，因此，当这些模型被适当地组合时，整体的预测性能可以超越任何单一模型的性能。

集成学习的两个主要类别是 Bagging 和 Boosting：Bagging 涉及独立地训练多个模型，并通过简单的投票或平均来组合它们的预测。Boosting 涉及顺序地构建模型，其中每个模型都试图纠正前一个模型的错误。

集成学习能够综合多个模型的预测，通常能够达到比单个模型更高的准确率和鲁棒性。

对于我们的分类模型，我们采用 Bagging 方式分别独立的训练 KNN，SVM，MLP，RF，以期待获取不同特征图的最佳分类器，如图1b所示。我们通过把 HOG，LBP，Edge，分别应用到四种不同的分类中，共计 12 次分类训练，然后，我们将每个特征的测试集分类结果进行投票，以综合生成最终的分类结果。

硬投票（Hard Voting）和软投票（Soft Voting）是集成学习中常用的两种投票方式，用于组合多个基本模型的预测结果，以获得最终的集成预测。在本问题中，道路坑洼的检测与识别任务作为一个二元分类问题。为了达成最终的分类决策，我们采纳了硬投票策略。硬投票策略的优势在于其决策的明确性：分类结果直接由获得最多投票的类别确定。在我们的集成模型中，每个子模型对于最终投票的贡献是等权重的，即多数票决定最终预测类别。这种方法在处理明确分类任务时尤为有效，因为它简化了决策过程并减少了模型预测的不确定性。

4.3.2 KNN

$$y = \arg \max \sum_{x_i \in N_k(x)} I(y_i, c_j), i = 1, 2, \dots, N; j = 1, 2, \dots, K \quad (1)$$

其中， I 为指示函数：

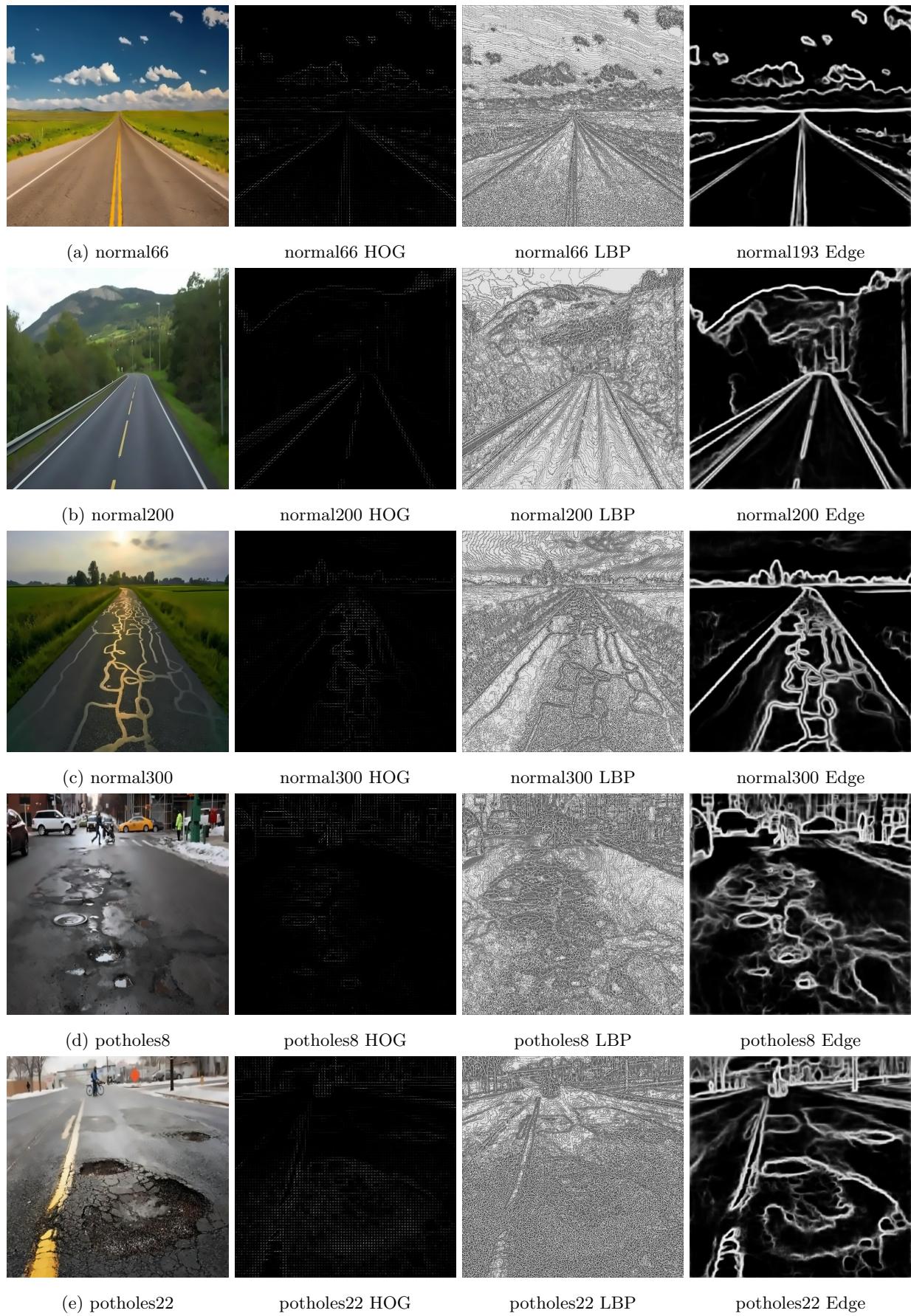


图 5：不同类别图片对应的 HOG, LBP, Edge 特征图。

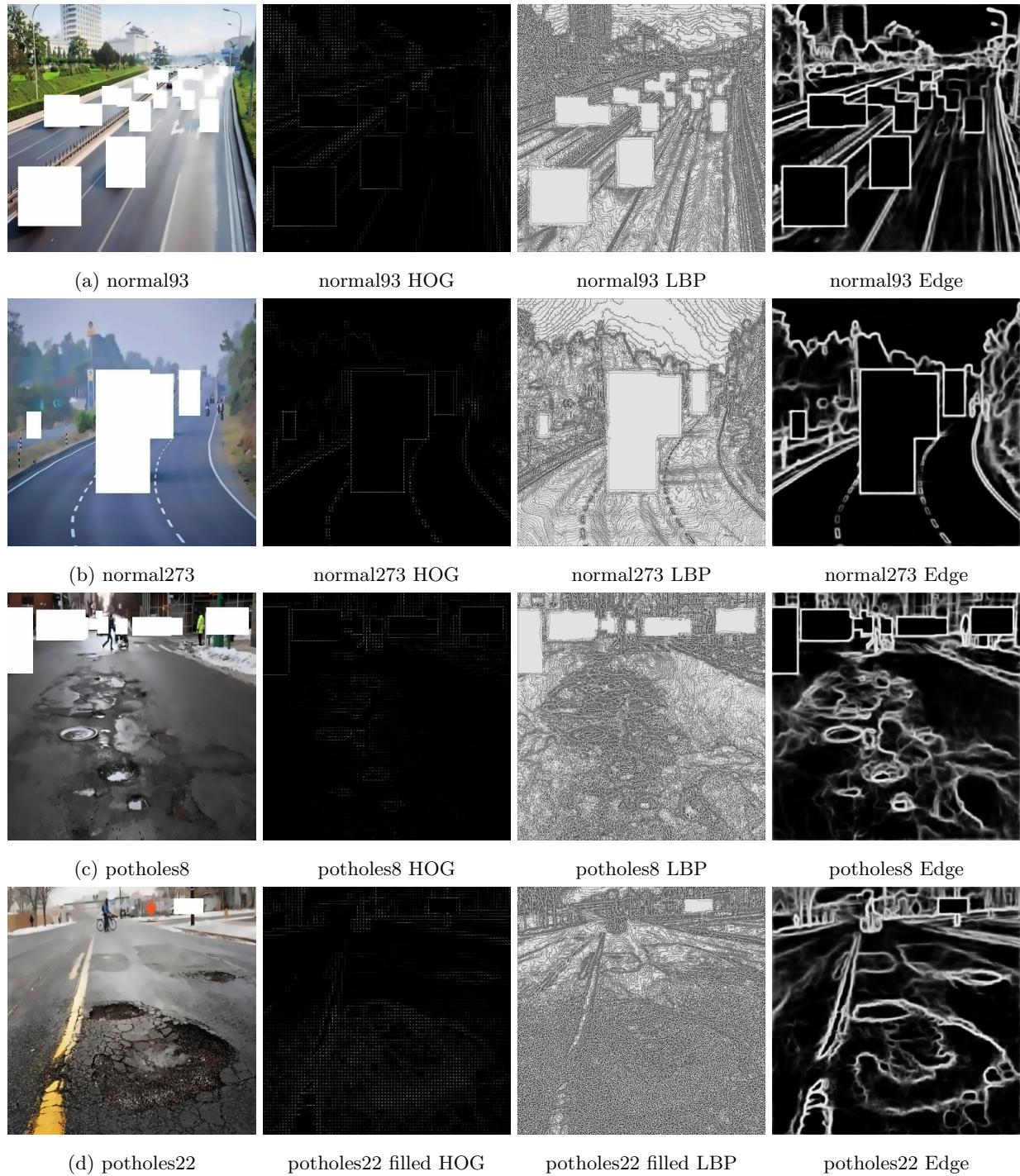


图 6: 为数据集添加白色掩码之后的 HOG, LBP, Edge 特征图。

$$\mathbb{I}(x, y) \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{if } x \neq y. \end{cases} \quad (2)$$

在K最近邻（KNN）算法中，待预测的输入实例 \mathbf{x} 的类别 y 可以通过以下公式确定：

$$y = \arg \max_c \sum_{i=1}^k \mathbb{I}(y_i = c) \quad (3)$$

对于一个待分类的输入实例 \mathbf{x} ，算法的目标是确定实例 \mathbf{x} 的类别标签。根据预先选定的距离度量（通常采用欧几里得距离度量），在训练数据集 \mathcal{T} 中识别出与实例 \mathbf{x} 距离最近的 k 个训练实例。这 k 个实例构成了集合 $N_k(\mathbf{x})$ 。接着，实例 \mathbf{x} 的类别标签 y 由 $N_k(\mathbf{x})$ 中出现频率最高的类别标签决定，这一决策过程基于多数投票规则。

4.3.3 支持向量机

支持向量机（SVM）是一种用于分类的监督学习算法，旨在通过确定一个最优超平面来区分不同类别的数据点，使得两个类别之间的间隔最大化。SVM的核心概念是最大化支持向量到决策边界的距离，其中支持向量是距离分隔超平面最近的数据点。SVM算法试图解决以下优化问题：

$$\max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_{(\mathbf{x}_i, y_i) \in D} y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \right\} \quad (4)$$

\mathbf{w} 是超平面的法向量， b 是超平面的偏置项， D 是训练数据集，其中 \mathbf{x}_i 是特征向量， y_i 是对应的类别标签，取值为+1或-1， $\|\mathbf{w}\|$ 是 \mathbf{w} 的范数，其倒数代表了类别间的间隔 [5]。

4.3.4 MLP算法

多层感知器（Multilayer Perceptron, MLP）是一种深度前馈人工神经网络结构，由多个层次的节点组成，每个节点（或称为“神经元”）与下一层的所有节点全连接。每个节点在接收输入后，通过一个非线性激活函数进行变换，常用的激活函数包括 Sigmoid 函数和修正线性单元（ReLU）。MLP 的训练过程涉及权重 W 和偏置 b 的调整，这是通过反向传播算法实现的。具体来说，MLP 的训练目标是最小化损失函数 L ，函数衡量了网络预测输出 \hat{y} 与真实标签 y 之间的差异 [6]。

4.4 深度卷积神经网络模型分类器

鉴于深度卷积神经网络（CNN）在图像识别问题中通常具有更好的性能，特别是在大规模数据和复杂特征情况下，我们也提出了面向道路坑洼检测识别的二分类问题的深度卷积神经网络。与前述的集成学习模型相比，本模型能够实现端到端的学习，从数据中学习特征表示和分类决策，无需手动进行特征提取，大大简化了模型训练过程。同时，对于具有复杂特征和模式的图像数据，深度 CNN 模型通常能够更好地捕获这些特征，因为它们具有多层的卷积和池化层，可以处理不同尺度和抽象级别的特征。

我们对预处理后得到的 HOG 特征矩阵，LBP 特征矩阵，RFC 特征矩阵与去噪处理后的灰度特征矩阵进行了特征融合。随后，我们采用了特征叠加方法，将各种特征表示层叠成多层通道，通过通道级融合输入到网络中。

首先我们把特征融合后尺寸为 640x640 像素，具有 4 通道的训练集。接着，我们在 Stem 阶段对输入图像进行一些初始的卷积和池化操作，以减小图像的尺寸并提取低级别的特征。整体主要流程由 3 组 Inception 模块和残差连接和 2 组池化层和卷积层组成，最后通过一个全局平均池化层，将最后一个卷积层的

特征图降维为一个固定大小的特征向量。为了降低网络层间的复杂度，我们增加了概率为 0.5 的 Dropout 层，即暂时灭掉一半的神经元防止过拟合。最后，Softmax 层用于将模型的输出映射为类别概率分布，以进行分类任务。图7为本模型的具体网络架构：

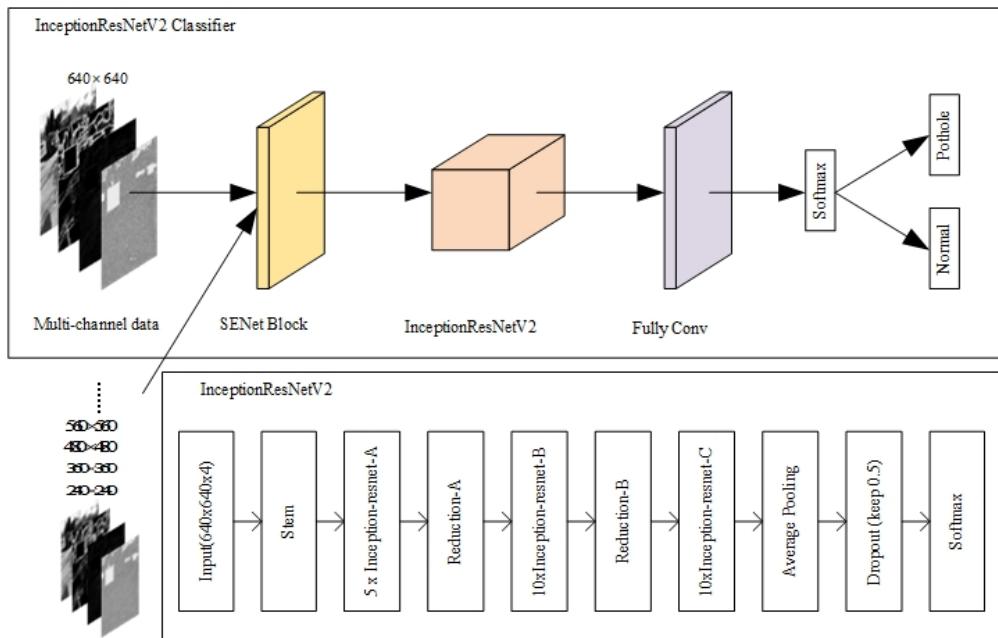


图 7：结合了SENet Block的InceptionResNetV2的架构。对比原生 InceptionResNetV2 模型，原来的思路打算加载预训练模型，利用迁移学习获取卓越的分类结果，但预训练模型的权重尺寸 240×240 的与我们经过处理后的 640×640 的尺寸不符，但过度缩小尺寸损失图像的信息所能带来的对分类器的正向作用，可能会大于迁移学习得到的好处。因此，我们相较于原生模型，除了增加 SENet 模块以外，只对 InceptionResNetV2 添加了一个全局平均池化层和一个全连接层，最后添加了一个二分类的输出层。最后，我们通过实验证明，模型与输出尺寸之间的关系。

4.4.1 特征融合

SENet (Squeeze-and-Excitation Networks) 模块引入了一种有效的通道注意力机制，旨在增强卷积神经网络 (CNN) 对通道特征重要性的适应性和判别力。SENet模块通过显式地建模通道间的依赖关系，能够动态地调整通道特征的响应强度。这一机制包含两个主要操作：压缩 (Squeeze) 和激励 (Excitation)，具体步骤如下：(1) 压缩：通过全局平均池化 (Global Average Pooling, GAP) 操作来实现，它将每个通道的特征图 (Feature Map) 压缩为一个单一的标量值。这个标量值是通过对通道上所有特征的平均值来获得的，从而实现对通道信息的全局汇总。(2) 激励：在压缩后，通过一个门控机制（通常是一个具有ReLU和Sigmoid激活函数的两层全连接网络）来学习每个通道的重要性权重。这个过程生成了一个用于调制每个通道的特征图的权重向量。这些增强的权重可以应用于原始特征图的每个通道，以增强或抑制通道内的信息，从而使网络更有针对性地关注重要通道，提高模型的性能 [7]。SENet的通道注意力机制可以嵌入到CNN的不同层中，以提高网络对特定任务的特征学习和表示能力。这对于提高图像分类、目标检测和语义分割等任务的性能具有帮助。通过这种方式，SENet模块能够强化有用的特征并抑制不相关的特征，从而提升CNN在图像分类、目标检测、语义分割等视觉任务中的性能 [8]。SENet模块的设计允许它被轻松地集成到现有的CNN架构中，而不需要对网络的基本结构进行大的修改。

4.5 集成学习模型分类器

为了学习到鲁棒的特征，同时充分利用预处理提取到的特征，我们采用到特征级融合的集成学习方法，即对每种特征提取技术（如RFC、LBP、HOG）应用了不同的分类器（包括KNN、SVM、RF、MLP），并选择了每个特征的最佳分类模型。然后，我们将每个特征的测试集分类结果进行投票，以综合生成最终的分类结果。硬投票（Hard Voting）和软投票（Soft Voting）是集成学习中常用的两种投票方式，用于组合多个基本模型的预测结果，以获得最终的集成预测。在本问题中，道路坑洼的检测与识别任务作为一个二元分类问题。为了达成最终的分类决策，我们采纳了硬投票策略。硬投票策略的优势在于其决策的明确性：分类结果直接由获得最多投票的类别确定。在我们的集成模型中，每个子模型对于最终投票的贡献是等权重的，即多数票决定最终预测类别。这种方法在处理明确分类任务时尤为有效，因为它简化了决策过程并减少了模型预测的不确定性 [9]。集成学习方法能够显著增强分类任务的性能，有效地减少了模型的偏差，增强了对新数据的适应性，并确保了分类结果的鲁棒性 [10]。

5 模型的求解及优化

5.1 深度卷积神经网络分类模型训练及优化

5.1.1 模型训练

训练环境见表1

Environment	Device or Version
CPU	12th Gen Intel(R) Core(TM) i7-12700F
GPU	NVIDIA GeForce RTX 4090
MEM	32GB
CUDAtoolkit	12.3
Tensorflow	2.7

表 1: 配置环境

我们首先对数据进行不同尺寸下的预处理，把原始数据分别压缩到5种尺寸下，并分别获取它们的特征图，这样做是希望能够捕获模型在不同尺度数据集上的差异。图8展示了数据处理所使用到的代码文件，以及对应生成的相关目录下的特征图。

在训练之前，如代码中‘train_merge_sz640.py’中展示的那样，我们对首先对不平衡数据集中的 pothole 类增加权重，以防止其在 pothole 类表现过差，然后对整体数据进行数据增强以及，不然过少或有序的数据可能会导致模型效果不佳。随后我们对四类特征图进行数据集划分之后，再进行 Concat 操作。不同尺寸下数据集，训练的模型表现如表2所示，从中发现，删除SENet模块会显著造成模型识别准确率下降。在五种不同的分类中， 640×640 能够使得模型取得不错的效果。但红色数据表明模型可能出现了过拟合现象。

Epoch=20	Validation_Accuracy_epoch=20	Validation_Accuracy_epoch=40	5
240 × 240	85.25%	85.25%	
360 × 360	77.92%	78.52%	
480 × 480	80.32%	83.61%	
560 × 560	67.21%	13.21%	
640 × 640	88.52%	85.10%	
640 × 640 without SENet module	78.62%	77.92%	

表 2: Epoch = 20 下的验证数据集准确率

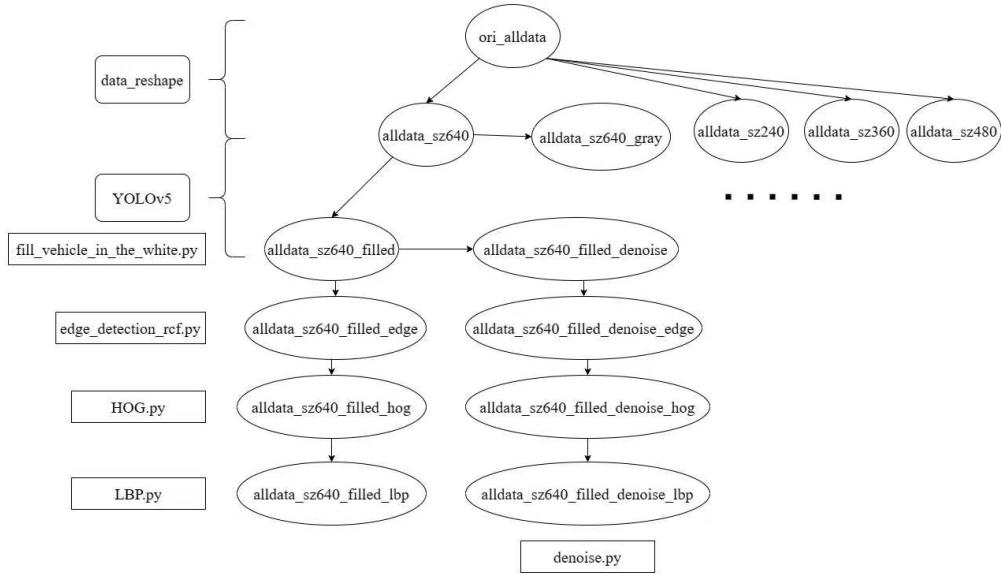


图 8：我们对原始数据预处理后的数据目录

5.1.2 优化

我们的训练目标是获取理想的分类模型，但在实际的过程中，我们得到的模型往往是不平衡和小样本的，这样在数据增强的过程中，很有可能会存在“用力过猛”，导致模型在训练的过程中，好比“难度加大”，这也会导致，每隔一个或几个 Epoch，验证数据集准确率波动过大，如图9b和图9d所示。同样模型在训练的过程中过小的‘batch_size’会导致加载的批次陷入某个固定的分类中，使得验证数据准确率突然降低。解决方案是采用遍历方法尝试不同的L1/L2正则值搭配方案，采用解决过拟合问题的办法，如图9d的“陷入”频率明显低于图9c和图9b。但同时要注意，过高的正则化可能会导致模型欠拟合，而过低则可能无法解决过拟合问题。通常需要通过实验来找到最佳的正则化系数。此外，正则化可能会影响训练速度和模型最终的性能 [11]。

5.2 集成学习模型

我们对每个模型和每个组合进行了网格搜索CV，找出了每个模型和组合的最佳参数，然后计算了不同的模型性能矩阵，用于模型和特征组合的比较分析。然后，我们对每个组合进行了比较，然后选择了最优化的模型和特征组合。

5.2.1 SVM正则化的选择

在支持向量机（SVM）模型中，正则化是用来控制模型的复杂性以及防止过拟合的重要超参数。正则化的选择在很大程度上影响模型的性能。

SVM的目标函数为：

$$J(w) = C \cdot \sum_{i=1}^n L(y_i, w^T x_i) + \frac{1}{2} \|w\|^2 \quad (5)$$

其中， C 是正则化参数，他是一个超参数，控制了损失函数和正则化项之前的权衡； $L(y_i, w^T x_i)$ 是损失函数，通常用于测量每个训练样本的误分类程度； w 是模型的权重向量； x_i 是第 i 个训练样本的特征向

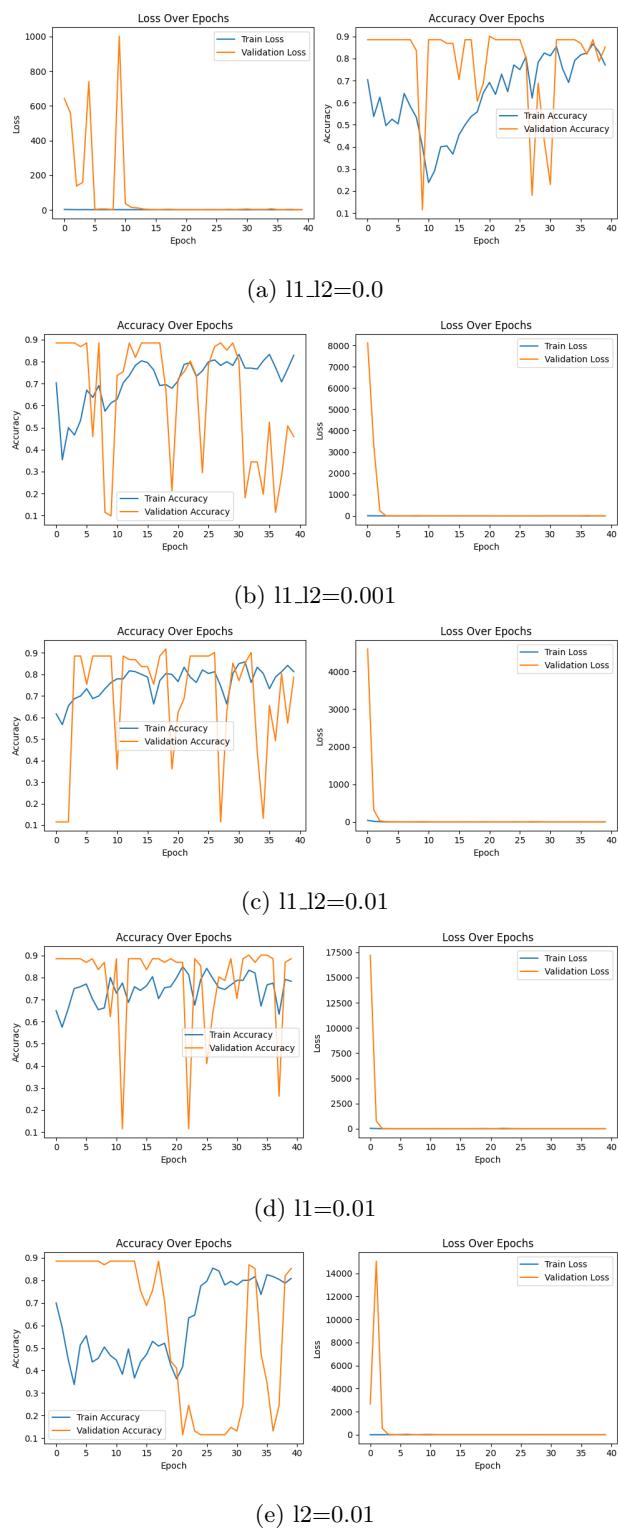


图 9: 不同的L1, L2正则值的训练曲线和准确率。

量: $\|w\|^2$ 是 \mathcal{L}_2 正则项, 用于对权重进行平方惩罚, 以控制模型的复杂性。

5.2.2 网络融合策略优化

如图10我们选择了 4 个网络用于模型集成, 让每种特征图像采用投票策略决定道路是否坑洼的最终预测值。对于投票法, 我们这里采用硬投票的方式。预处理得到的 3 个特征图分别对 4 个基模型寻求最佳分类模型, 运行后针对同一张遥感图片会生成 3 张标签图, 对这 3 张标签图逐像素进行投票来决定每一像素最终的标签。其中在对最终结果投票时, 我们对模型的输出采用更严格的判定策略, 对于每个子模型的分类结果, 为降低假负率, 我们采用少数服从多数的原则, 当至少两个模型判定其属于坑洼道路时我们才判定其为坑洼道路 (标记为 1), 否则为非坑洼道路 (标记为 0)。训练结果见图.11

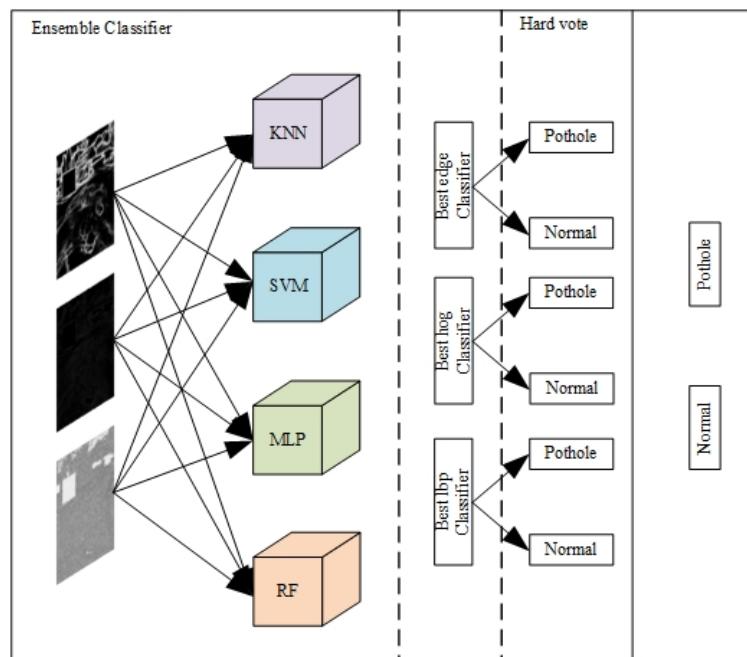


图 10: 集成学习模型

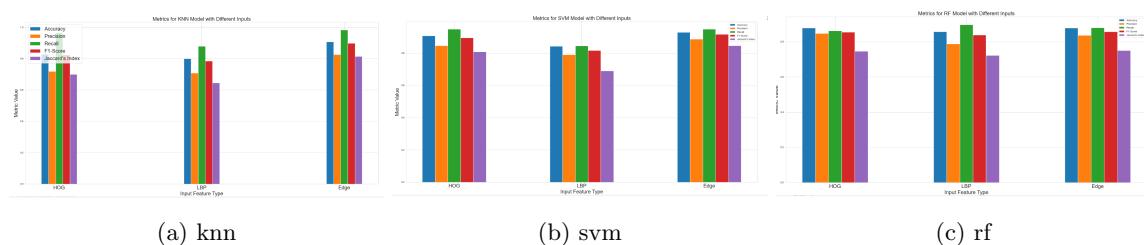


图 11: KNN, SVM, RF 模型的评分

6 模型的评价

6.1 模型的优点

(1) 在本研究中，我们采取了一种创新的预处理方法，将目标检测模块集成到分类模型的预处理步骤中，对干扰项物体识别并进行将像素值改变为255的“背景化”操作。这种方法的提出是基于在此二分类问题中，模型把正常的道路识别为道路坑洼的概率更大，即假负率较高，真正率较低。在道路坑洼的识别任务中，降低假负率对于提升模型的实际应用价值至关重要。因此，为了减少高假负率的问题，我们提出了在分类前排除干扰项的策略。

(2) 在本研究中，我们采用了集成学习策略，通过将不同特征提取技术与不同分类器结合起来，生成了一个综合的预测结果。这种方法的效率高且易于实施，显著减少了预测误差，并减轻了对单一模型的依赖，从而提高整体模型的鲁棒性，即便个别模型性能不佳，整体模型依然能保持稳定的性能。此外，它还提升了模型对新数据的泛化能力。在本次道路坑洼检测的应用中，该方法可以优化并精确目标点位置并减少背景噪声对道路检测结果的干扰。

(3) 本文模型探索了图像的不同特征提取方法在分类模型上的准确率以及不同图像大小的输入对模型准确率的影响，探索迁移效果。

6.2 模型的缺点

虽然通过优化，我们的模型在性能上有了明显的提高，能够识别出与道路坑洼具有相似灰度和灰度变化的区域，但它在预测某些如湖泊等特定色块时仍然面临挑战。为了解决这一问题，我们需要增强模型对强化原始图像信息，扩充训练数据集，并对模型进行进一步的细化和改进。

6.3 模型的改进

(1) 通过网络剪枝和参数共享的策略改进本文深度学习模型。以此来追求模型的轻量化和高效性，以减小模型的参数量和计算负担，同时确保模型对道路坑洼检测和识别的准确率。(2) 扩大训练集。在数据增强操作中，我们仅对原始数据进行180度旋转、对角与镜像旋转。我们构想采用更多数据增强操作，如增加对比度、图像加噪等以扩充数据量。(3) 尝试优化预处理中目标检测模块对象的选择。在训练集中我们通过Roboflow平台手动对坑洼以外的干扰项物体进行绘制边界框(Bounding Box)，不排除一些潜在的干扰项物品并未在我们的目标检测对象范围以内进行处理，从而使我们的模型有待提高。

6.4 模型的推广

本项目目标为道路坑洼图像的二分类，在实际应用过程中，可将本模型加以扩展适用于道路坑洼图像的多分类任务中，对坑洼的具体形状和深浅大小划分成不同的类别，例如小坑、中坑、大坑等。通常来说，对坑洼进行细分和分类可以更好地帮助城市规划和维护部门进行道路维护和修复工作。

7 附录

所有数据代码，模型代码，训练代码均见：<https://github.com/npukuju11/potholes-detection>

参考文献

- [1] Xudong Dong, Shuai Yan, and Chaoqun Duan. A lightweight vehicles detection network model based on yolov5. *Engineering Applications of Artificial Intelligence*, 113:104914, 2022.
- [2] S. N. Shivappriya, M. Jasmine Pemeena Priyadarsini, Andrzej Stateczny, C. Puttamadappa, and B. D. Parameshachari. Cascade object detection and remote sensing object detection method based on trainable activation function. *Remote Sensing*, 13(2), 2021.
- [3] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Jia-Wang Bian, Le Zhang, Xiang Bai, and Jinhui Tang. Richer convolutional features for edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1939–1946, 2019.
- [4] Tingting Lu, Dongping Ming, Xiangguo Lin, Zhaoli Hong, Xueding Bai, and Ju Fang. Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network. *Remote Sensing*, 10(9), 2018.
- [5] Dirk Valkenborg, Axel-Jan Rousseau, Melvin Geubbelsmans, and Tomasz Burzykowski. Support vector machines. *American Journal of Orthodontics and Dentofacial Orthopedics*, 164(5):754–757, 2023.
- [6] Peng Tao, Jie Cheng, and Luonan Chen. Brain-inspired chaotic backpropagation for mlp. *Neural Networks*, 155:1–13, 2022.
- [7] Zhenfeng Shao and Jiajun Cai. Remote sensing image fusion with deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1656–1669, 2018.
- [8] Junsuo Qu, Zongbing Tang, Le Zhang, Yanghai Zhang, and Zhenguang Zhang. Remote sensing small object detection network based on attention mechanism and multi-scale feature fusion. *Remote Sensing*, 15(11), 2023.
- [9] Giang Ngo, Rodney Beard, and Rohitash Chandra. Evolutionary bagging for ensemble learning. *Neurocomputing*, 510:1–14, 2022.
- [10] Wenbo Qiu, Jianghan Zhu, Guohua Wu, Huangke Chen, Witold Pedrycz, and Ponmuthurai Nagarathnam Suganthan. Ensemble many-objective optimization algorithm based on voting mechanism. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(3):1716–1730, 2022.
- [11] Rafael C. Gonzalez. Deep convolutional neural networks [lecture notes]. *IEEE Signal Processing Magazine*, 35(6):79–87, 2018.