

队伍编号	MCB2302817
赛道	A

基于 Advanced-Inception_v3 模型的道路坑洼检测

摘 要

本文根据深度学习理论，在预训练模型 VGG16 模型与 Inception_v3 模型的基础上对卷积神经网络的结构进行调整，并通过模型训练与模型优化建立一个识别率高、速度快、分类准确的模型，使得该模型能够自动地检测一张新的道路图像，判断道路表面是否存在坑洼现象。

针对问题一，首先对题目所给的数据集进行数据预处理，分别对图像中特征不鲜明的图片进行清洗。由于数据集的样本数量不足，本文采用图像随机缩放变换、剪切变换、旋转、垂直、水平翻转、亮度调整、加入椒盐噪声的方法对数据集进行增广，以丰富数据集的多样性。其次，在 VGG16 深度卷积神经网络模型基础上对网络结构中的全连接层进行改进，形成 Advanced-VGG16 模型。同时，在 Inception_v3 模型的基础上将激活函数改为 sigmoid 函数形成通过 Advanced-Inception_v3 模型，使模型更好地捕获图像中的坑洼特征，提高道路坑洼的识别性能。初步通过较少迭代次数的模型训练，得出 Advanced-Inception_v3 模型的准确率提升速度更快、相同迭代次数下准确率更加高、训练开销更小。因此选取该模型作为识别坑洼道路的分类模型。

针对问题二，对问题一所选择的 Advanced-Inception_v3 分类模型进行多次训练，从损失函数的选择、优化器的选择、是否添加全局平均池化层、是否添加正则化、是否添加早停回调函数五个维度对模型进行优化，最终分类模型的分类准确率达 93%。本文选取分类准确率、召回率、查准率、F1 分数、Kappa 系数对模型进行评估效果均较好，且坑洼图像的查准率为 96%比正常图像的查准率 91%高，进一步说明本文所建立的分类模型对道路坑洼检测的效果更好，具有较强的鲁棒性与泛化能力。

针对问题三，我们对给定的测试集进行预测，我们将使用在训练过程中获得的优秀模型参数。在 Colab 平台上加载这些模型参数后，我们将进行一系列预测，并将结果保存在名为 test_result.csv 的文件中。

关键词：深度学习 VGG16 模型 Inception_v3 模型 卷积神经网络 坑洼检测

目录

一、 问题重述.....	3
1.1 问题背景	3
1.2 问题重述	3
二、 问题分析.....	3
2.1 问题一的分析	3
2.2 问题二的分析	4
2.3 问题三的分析	4
三、 模型假设.....	5
四、 符号说明.....	5
五、 模型的建立与求解.....	5
5.1 问题一模型的建立与求解	5
5.1.1 数据预处理	5
5.1.2 Advanced-VGG16 模型的建立	10
5.1.3 Advanced-Inception_v3 卷积神经网络模型.....	13
5.1.4 模型对比与模型选择	15
5.2 问题二模型的建立与求解	16
5.2.1 Advanced-Inception_v3 模型训练与模型优化.....	16
5.2.2 模型评估	19
5.3 问题三模型的建立与求解	22
六、 模型的评价、改进与推广	22
6.1 模型的优点	22
6.2 模型的缺点	23
6.3 模型的改进	23
七、 参考文献.....	24

一、问题重述

1.1 问题背景

道路建设是我国的基础设施建设，道路作为人们日常通勤、出行和旅行的主要方式应该要平坦无障碍。但由于天气、气候、高强度车辆交通等各种自然因素与社会因素的影响，导致某些道路上会出现一定深浅的坑洼，这对道路交通安全造成极大的危害，因此需要定期检查道路坑洼与及时维修。

近年来，计算机视觉的目标检测应用到道路坑洼检测问题可以提高坑洼检测效率、准确性和覆盖范围，从而帮助道路管理部门及时发现并修复道路坑洼，提升道路网络的安全性、舒适性和效率。

1.2 问题重述

本题希望结合深度学习理论从本题所给的正常道路图像与坑洼道路图像中提取重要特征，建立合适的分类模型，使得该模型能够自动地检测一张新的道路图像，判断道路表面是否存在坑洼现象。具体任务如下：

问题 1:结合给出的图像文件，提取图像特征，建立一个识别率高、速度快、分类准确的模型，用于识别图像中的道路是正常或者坑洼。

问题 2:对问题 1 中构建的模型进行训练, 并从不同维度进行模型评估。

问题 3:利用已训练的模型识别测试集中的坑洼图像，并将识别结果放在“test result.csv”中。



正常道路示例



坑洼道路示例

二、问题分析

2.1 问题一的分析

本问题的目标是结合深度学习理论创建一个高性能的分类模型，该可以通过提取图像特征来快速而准确地识别道路状况，将道路图像分为正常道路与坑洼道路两类。

首先，结合本题所给的图像文件，对图像进行数据清洗与数据增广，根据每张图像的像素将图像转换为数据，从而为后续的模型训练创建一个特征鲜明的数据集。本题基于 VGG16 深度卷积神经网络模型对其网络结构进行改进后提出 Advanced VGG16 模型，同时基于 Inception_v3 卷积神经网络模型深度卷积神经网络模型对其分类器中的激活函数进行改进后提出 Advanced Inception_v3 模型。

其次，为从上述两个分类模型中选出一个识别率高、速度快、分类准确的模型，在本题中将对上述两个模型进行较少迭代次数的模型训练，通过对比模型的训练开销、准确率的提升速度、相同迭代次数下的准确率这三个维度初步选择分类准确的模型。

后续问题二中将通过模型训练继续从多个方面对该分类模型进行优化。问题一的分析思路如下图 1 所示。

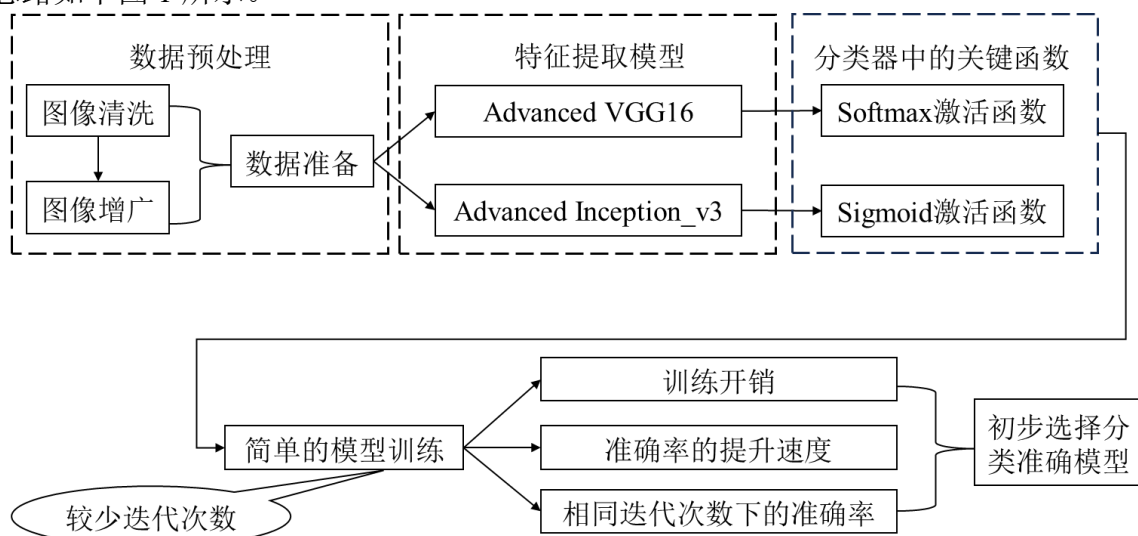


图 1 问题一分析思路图

2.2 问题二的分析

本问题需要对问题一所构建的道路图像分类模型进行训练，并从多个维度对分类模型的性能进行评估。

首先，对问题一中初步选择的分类模型进行模型训练，分别从是否在原有神经网络的全连接层后增加全局平均池化层，是否在编译模型阶段增加 L2 正则化、丢弃层，选择何种损失函数、选择何种优化器这五个方面对模型进行优化，最终建立一个识别率高、速度快、分类准确的模型。其次，本问题将根据分类模型的训练结果得出混淆矩阵，并从分类准确率、召回率、查准率、F1 分数、ROC 曲线五个维度对分类模型进行评估。问题二的分析思路如下图 2 所示。

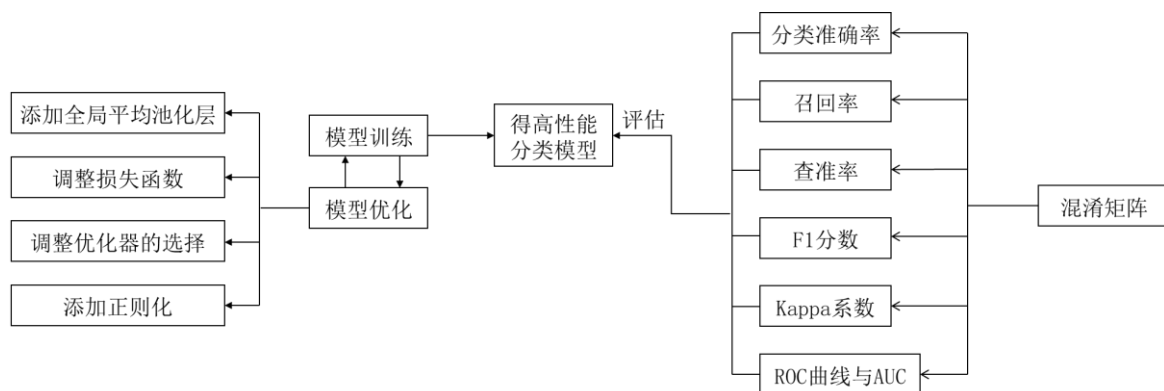


图 2 问题二分析思路图

2.3 问题三的分析

我们对给定的测试集进行预测，我们将使用在训练过程中获得的优秀模型参数。在 Colab 平台上加载这些模型参数后，我们将进行一系列预测，并将结果保存在名为 test_result.csv 的文件中。

三、模型假设

- 1、假设模型可以适应不同类型和尺寸的坑洼。
- 2、假设数据集中的道路特征足够准确划分正常道路与坑洼道路。
- 3、假设图像的天气状况、强弱光照等各类场景对道路坑洼识别的影响较小。
- 4、假设在神经网络分类预测中，局部图像特征具有代表性，可以有效地推断出整体图像的特征。

四、符号说明

符号	说明
α	图像旋转角度
x	原像素值
x_0	标准化处理后的像素值
n	神经元的个数
y_k	第 k 个神经元的输出
a_k	第 k 个输入信号
TP	将正类预测为正类的数量
FN	将正类预测为负类的数量
FP	将负类预测为正类的数量
TN	将负类预测为负类的数量

五、模型的建立与求解

5.1 问题一模型的建立与求解

5.1.1 数据预处理

为本题能训练出分类性能更好的模型，确保模型的准确性与高效性，本文将通过数据清洗与数据增广的方式准备一个高质量、有清晰定义的数据集。

数据清洗

（一）格式统一

通过检验数据集中的图像格式发现原数据集中的图像格式存在混乱不一的问题，虽图像的后缀均为 jpg，但实质上存在少量的 png 格式与 webp 格式，如下图 3 所示其中有 0.3% 的图像为 png 格式，3.0% 的图像为 webp 格式。

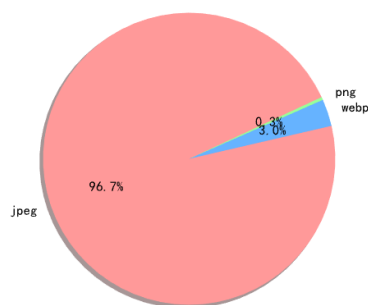


图3 图像格式占比分析图

png、webp 这两种格式的通道数为 4，jpeg 的通道数为 3，如果训练集数据包含不同通道数的图像，则需要建立足够复杂的模型来适应数据的多样性，否则模型无法充分学习图像的特征。同时考虑到本模型中所用到的卷积神经网络中的卷积操作需要根据输入图像的通道数进行相应的调整，且在训练和推断时需要跟多的计算资源，本文将统一图像格式为 png 以确保图像具有一致的通道数。

(二)质量筛选

通过对原数据集中的图像进行筛选发现，normal 类别中的样本存在 3 张特征并不鲜明的图像，如下图 4 所示。从图像可以看出 normal 173 实质上是坑洼图像。normal 298 中仅有左半边为正常道路，右半边实质上为塌方图像，该图像无效干扰信息太多，容易出现特征提取错误的情况。normal 346 两边出现大面积的纯黑边框，属于无效信息，并且图像中的道路实质上为泥泞道路而非正常道路。

若将这些特征不鲜明的图像放在训练集中的 normal 类别，可能会出现不同类别的样本在特征空间中重叠，使得所建立的分类模型难以准确区分正常道路与坑洼道路，降低模型的分类准确度与泛化能力。



图4 特征不鲜明图像

数据增广

由于题目所给数据集的图像较少，且正常图像与坑洼图像的数量差距较大，如下图 5 所示。其中正常图像有 266 张，通过数据清洗删减掉特征不鲜明的图像后剩 263 张，坑洼图像仅有 35 张。通过查阅相关文献^[1]，样本数据不平衡问题导致所建立的分类模型过多关注多数类，即本题中的正常图像，使得少数类样本即坑洼图像的分类性能下降。

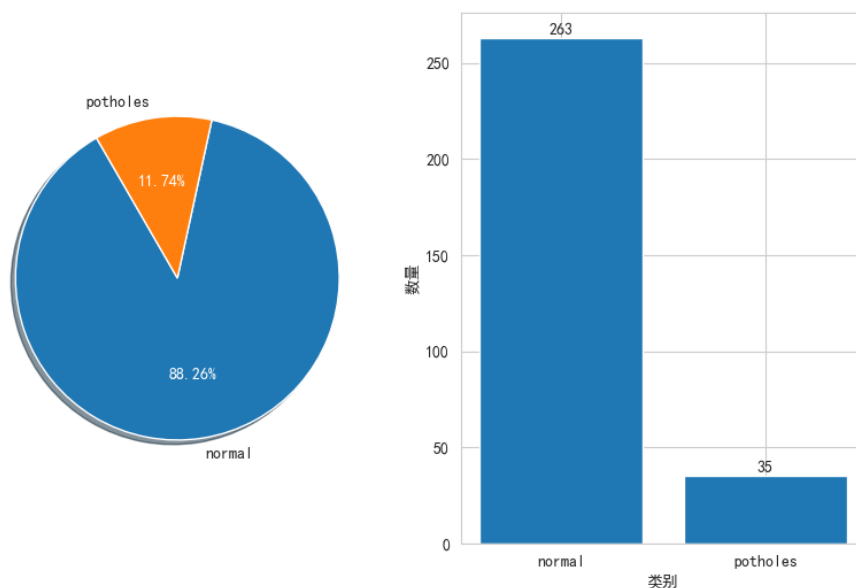


图 5 正常图像和坑洼图像的数据对比图

据研究表明，平衡的数据集能够使分类器具有良好的分类性能。本文通过过采样方法图 6，从网络公开数据集中寻找新的坑洼图像，如下所示，并添加进训练集中，使正、反类的样本数目接近。

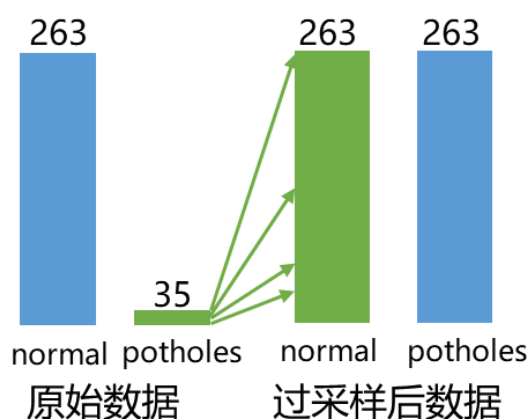


图 6 过采样示意图

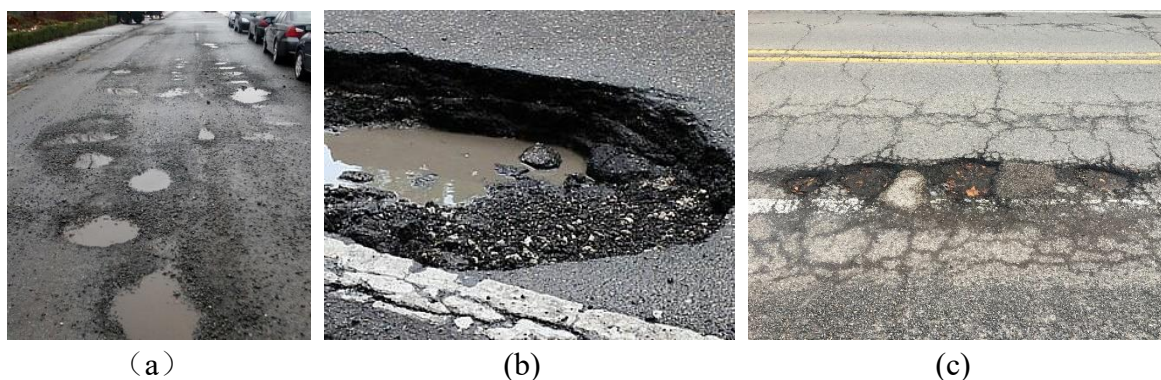


图 7 新添加的坑洼图像示例

为了填补数据集的样本不足问题，本题采用图像随机缩放变换、剪切变换、旋转、垂直、水平翻转、亮度调整、加入椒盐噪声的方法对数据集进行增广，以丰富数据集的多样性，提高分类模型对两个类别的学习能力，使本文模型更具有鲁棒性与分类性

能。

（一）图像随机缩放变换

下图 8 的图像采用了随机缩放变换方法，其中缩放变换允许的最大缩放因子为 0.2，代表图像的宽度和高度都可以在其原始尺寸的基础上扩大或缩小不超过 20%。



图 8 缩放变换前后对比图

（二）图像随机水平错切变换

下图 9 的图像采用了水平错切变换方法，其中错切变换范围为 0.2，表示图像的上下部分将相对于垂直轴进行平行位移，导致图像在水平方向上变得略微倾斜，上半部分向左移动，下半部分向右移动，从而改变了图像的形状。



图 9 水平错切变换前后对比图

（三）图像随机旋转

数据增广的一种常见技术为图像旋转，这种方法将图像围绕其特定的点（旋转中心）进行不同角度的旋转，通过这个旋转中心创建出多个新的训练样本。使所训练出的分类模型能够识别和理解不同角度或方向呈现的图像，更好的适应现实中的多样性条件与视角，提高本文分类模型的泛化能力。

下述用矩阵的形式具体表示图像的旋转。设点 $A_0(u_0, v_0)$ ，通过逆时针旋转 α 角度后所对应的点为 $A_1(u_1, v_1)$ ，图像变换过程的公式如下

$$\begin{cases} u_0 = r \cos \beta \\ v_0 = r \sin \beta \end{cases} \quad (1)$$

$$\begin{cases} u_1 = r \cos(\beta + \alpha) = r \cos \alpha \cos \beta - r \sin \alpha \sin \beta = u_0 \cos \alpha - v_0 \sin \alpha \\ v_1 = r \sin(\beta + \alpha) = r \sin \alpha \cos \beta + r \cos \alpha \sin \beta = u_0 \sin \alpha + v_0 \cos \alpha \end{cases} \quad (2)$$

用矩阵表示如下：

$$\begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_0 \\ v_0 \\ 1 \end{pmatrix} \quad (3)$$

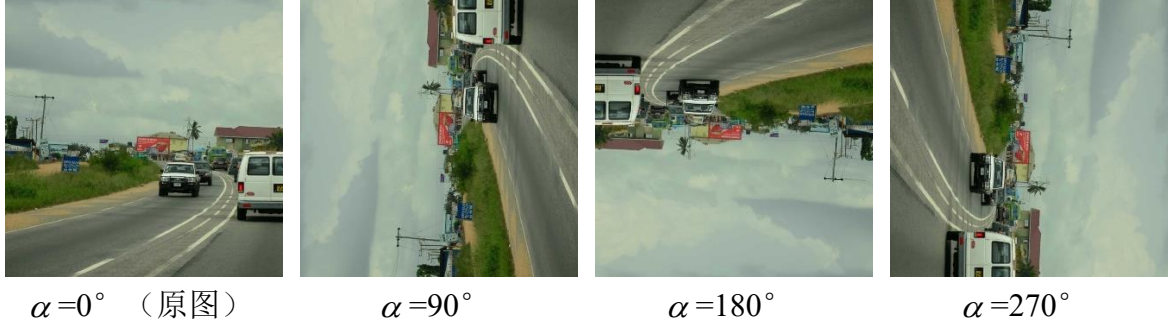


图 10 以 normal40 为例的图像旋转

(四) 图像垂直、水平翻转

下图 11 分别对原数据集中的图像 normal19 采用了水平翻转与垂直翻转的方式。图像的水平翻转是将图像从左到右进行镜像对称，但上下位置不变。图像的垂直翻转是将图像从上到下进行镜像对称，但左右位置不变。



图 11 以 normal19 为例的垂直、水平翻转

(五) 图像随机亮度调整

本题将对原数据集中的图像进行亮度调整如图 12 所示，将亮度变换的范围设置在 0.5-1.8 之间，通过调整图像的像素值使得图像变得更亮或更暗，使所训练的分类模型更好地应对不同光照条件下的图像。



图 12 亮度调整前后对比图

(六) 椒盐噪声

本文通过将椒盐噪声添加到图像作为数据增广的一部分，如下图 13 所示。在原始图像的基础上，随机替换 1%到 5%的像素点，使它们成为黑点或白点，从而模拟噪声的效果。这样做有助于训练分类模型更好地适应真实世界中可能存在的噪声，使其

能够在处理具有噪声的图像时表现更好。



图 13 加椒盐噪声前后对比图

数据准备

本文选用 rescale 标准化的方法对图像的像素值进行缩小，将像素值范围缩放至 $[0, 1]$ ，以减少后续训练过程中计算的复杂性，提高数值的稳定性，使本文所建立的分类模型更好地捕捉图像特征。具体转换公式如下：

$$x_0 = \frac{x}{225} \quad (4)$$

其中 x 为原像素值， x_0 为标准化处理后的像素值。这种简单标准化的方法在像素值的范围缩放上不涉及平均值和标准差的计算，使数据集更加高效缩放到一个适合模型训练的范围。

5. 1. 2 Advanced-VGG16 模型的建立

卷积神经网络（CNN）是专门设计用于处理图像及相似数据的神经网络结构。其独特之处在于利用卷积层对图像进行局部特征提取，使网络能够有效捕捉图像中的模式和结构，其结构图 14 如下。

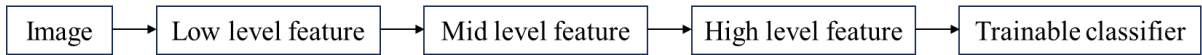


图 14 卷积神经网络结构图

VGG16^[2]（Visual Geometry Group 16）是一个经典的深度卷积神经网络模型，其中 VGG16 的主要特点在于其深度卷积结构和前向传播体系。这意味着它通过一系列卷积核计算来提取图像特征，然后通过前向传播的方式将这些特征传递给全连接层，以最终实现图像分类和其他计算机视觉任务。这种结构强调了对图像的逐层抽象表示和特征学习，使其在大规模图像分类任务中表现出色。

VGG16 网络结构非常一致且简单，具有较强的可理解性。其采用小尺寸的 3×3 卷积核，而不是更大的卷积核，这有助于该神经网络更好地捕捉图像的局部特征，同时减少了参数数量，提高计算效率。因此，本文采用 VGG16 卷积神经网络模型进行图像的特征提取与图像分类。

一、VGG16 模型的网络结构

VGG16 的网络结构包括输入层、卷积层、池化层、全连接层、输出层。其中 VGG16 包含了五个卷积层组，每个卷积层组都由一系列卷积层和池化层组成。这些卷积层组逐渐增加网络的深度，从而使该模型学习不同层次的特征。原始 VGG16 网络结构如下图所示 15 所示。

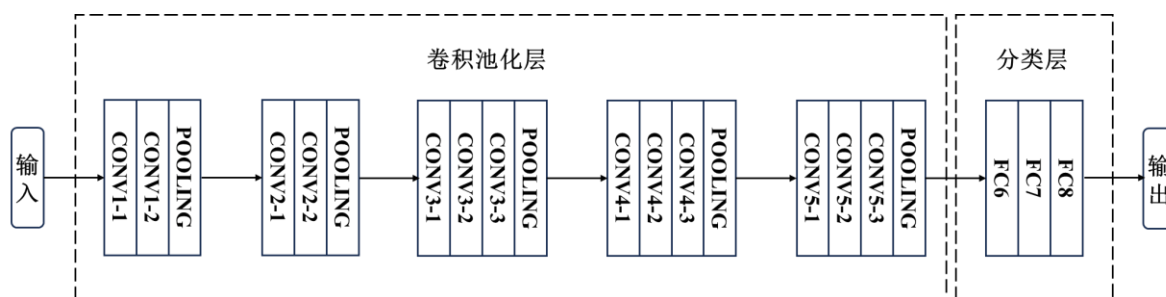


图 15 原始 VGG16 网络结构图

1.输入层

VGG16 模型的输入层被设计为接受具有特定尺寸和通道数的图像，即 224x224 像素的彩色图像，其中包括红色、绿色和蓝色通道（RGB）。

2.卷积层

VGG16 模型包含 13 个卷积层，其中卷积核大小为 3x3，步幅为 1，填充为 1 以提取图像的局部特征。这些卷积层逐渐增加网络的深度，在不同层次上学会不同的特征，从边缘和纹理到更高级别的特征，并且每层都使用 ReLU 激活函数来引入非线性。简化的卷积过程图如下图 16。

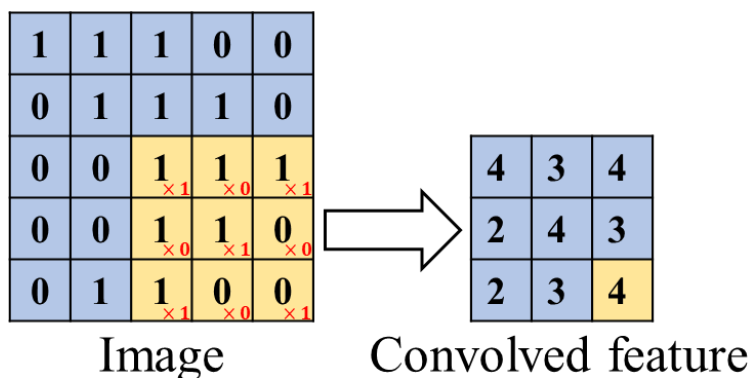


图 16 卷积过程图

3.池化层

VGG16 的池化层采用最大池化操作，使用 2x2 的池化窗口，步幅为 2，以减小特征图的空间维度并保留重要特征。池化层的主要目的是减小特征图的空间分辨率，从而降低计算复杂性和防止过拟合。池化过程图如下图 17。

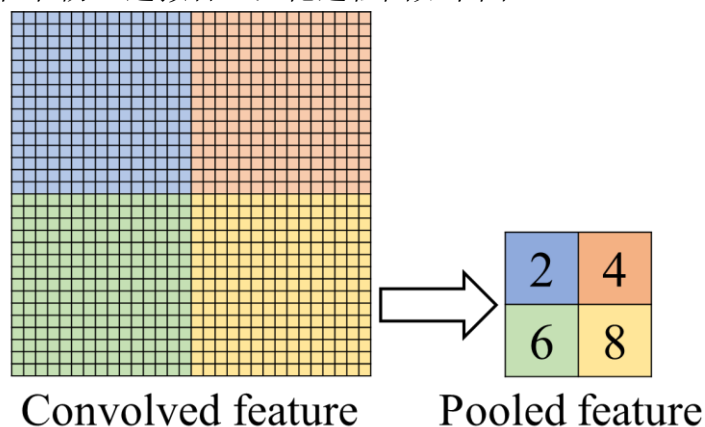


图 17 2x2 池化过程图

4.全连接层

VGG16 的全连接层如下图 18 所示，其由三个密集连接的层组成，每个层包含 4096 个神经元，用于将前面的卷积和池化层提取的特征映射到图像分类的输出概率，最终实现对不同类别的图像分类。

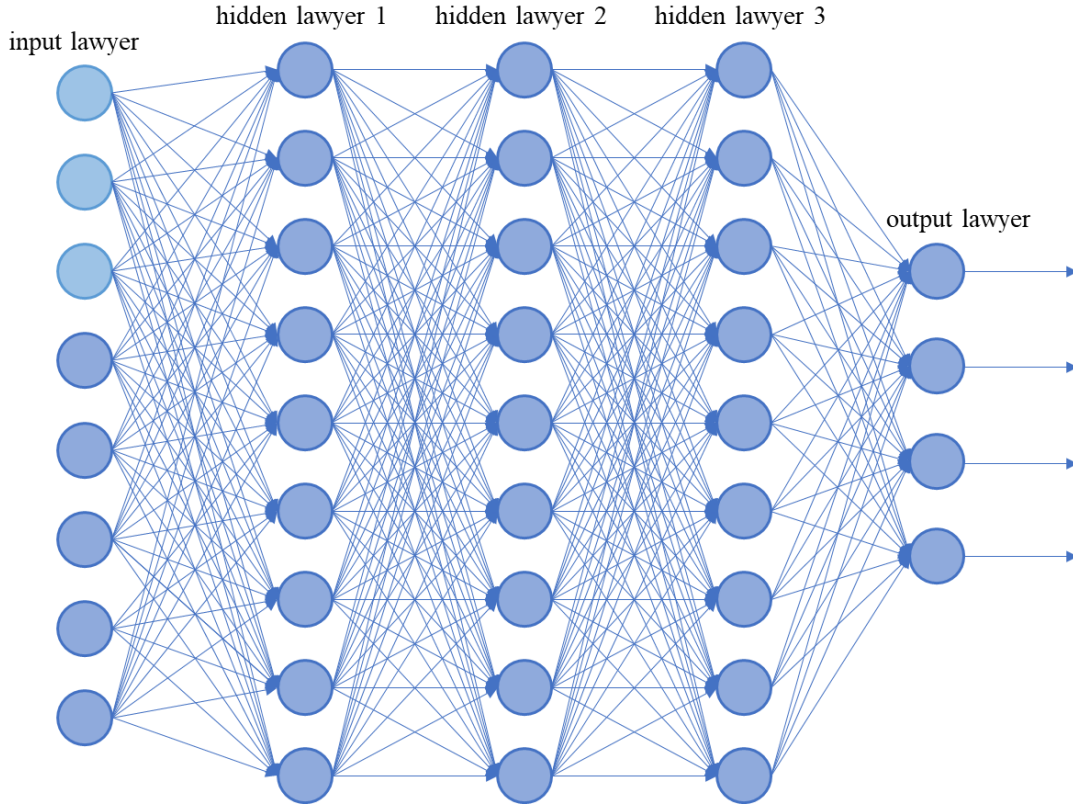


图 18 全连接层示意图

5.输出层（分类）

VGG16 模型的输出层是一个全连接层，在输出层后的 Softmax 激活函数将网络的原始输出转换为概率分布，从而进行对输入的图像进行分类。Softmax 激活函数的公式如下：

$$y_k = \frac{e^{a_k}}{\sum_{i=1}^n e^{a_i}} \quad (5)$$

其中，上式假设输出层共有 n 个神经元， y_k 表示第 k 个神经元的输出， a_k 表示输入信号。

三、网络迁移 Advanced-VGG16 模型

根据迁移学习理论，本文在初始 VGG16 深度卷积神经网络模型的基础上进行改进形成 Advanced-VGG16 作为特征提取模型，如下图 19 所示。

Advanced-VGG16 模型主要是在原模型的基础上对网络结构中的全连接层进行改进，在每个全连接层后添加丢弃层，使全连接层和丢弃层交替连接，这种交替连接的设计能够有效地提高神经网络的鲁棒性和泛化能力，通过在全连接层输出和输入丢弃层中的一部分神经元，减少过拟合风险，同时促进特征的更好学习。并且，额外添加了网络结构的层数以增加模型的深度和特征表示能力，这有助于更好地捕获坑洼图像中的细节和纹理特征。使得本文所建立的 Advanced-VGG16 模型更好地捕获图像中的坑洼特征，提高道路坑洼的识别性能。

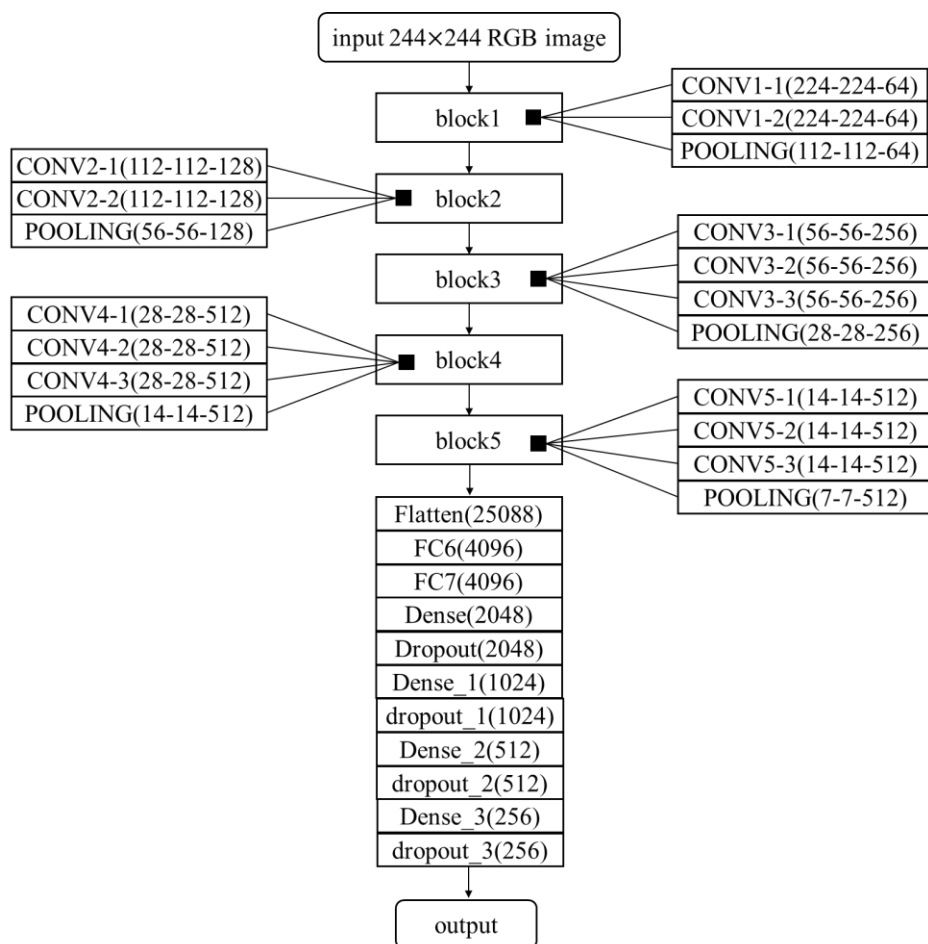


图 19 Advanced-VGG16 深度卷积神经网络模型

5.1.3 Advanced-Inception_v3 卷积神经网络模型

一、Inception 模块化设计

Inception_v3 模型是由 Google 公司开发的用于图像分类和识别任务的卷积神经网络模型。Inception_v3 模型独特之处在于其采用了模块化设计，Inception 模块包含了多个不同尺寸的卷积核和池化层，这种多尺度的设计使得网络能够同时捕捉到图像中不同大小物体的特征。通过 Inception 模块的设计，Inception-v3 网络增加了深度，同时使用了更小的卷积核，实现了相对较少的参数和更高的效率。其模块结构如下图 20 所示：

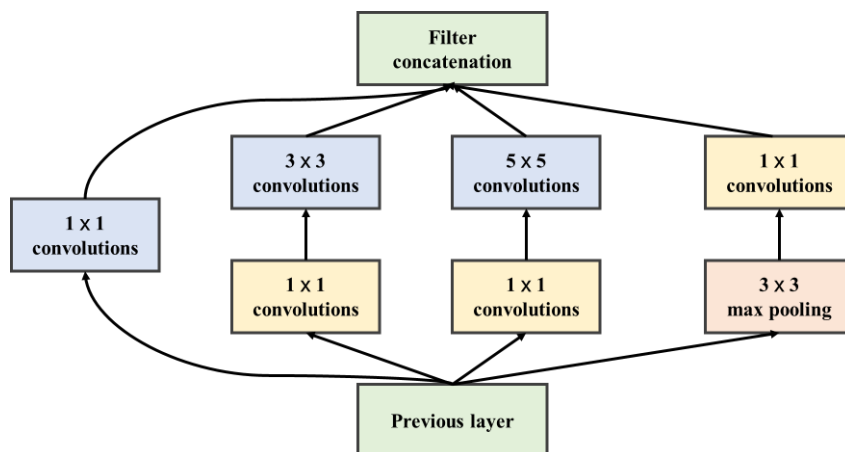


图 20 Inception 模块结构图

二、Inception v3 模型的网络结构

与 VGG16 深度卷积神经网络模型类似, Inception_v3 卷积神经网络模型的网络结构也包括输入层、卷积层、池化层、全连接层、输出层, 在此基础上 Inception_v3 还包括了 Inception 模块, 并且在最后一个 Inception 模块之后使用全局平均池化层来汇总特征图, 将最终的特征图变换成一个固定大小的向量。这种汇总技术有助于减少过拟合, 并使网络更好地理解整个图像的内容。Inception_v3 卷积神经网络模型的网络结构如下图 21 所示:

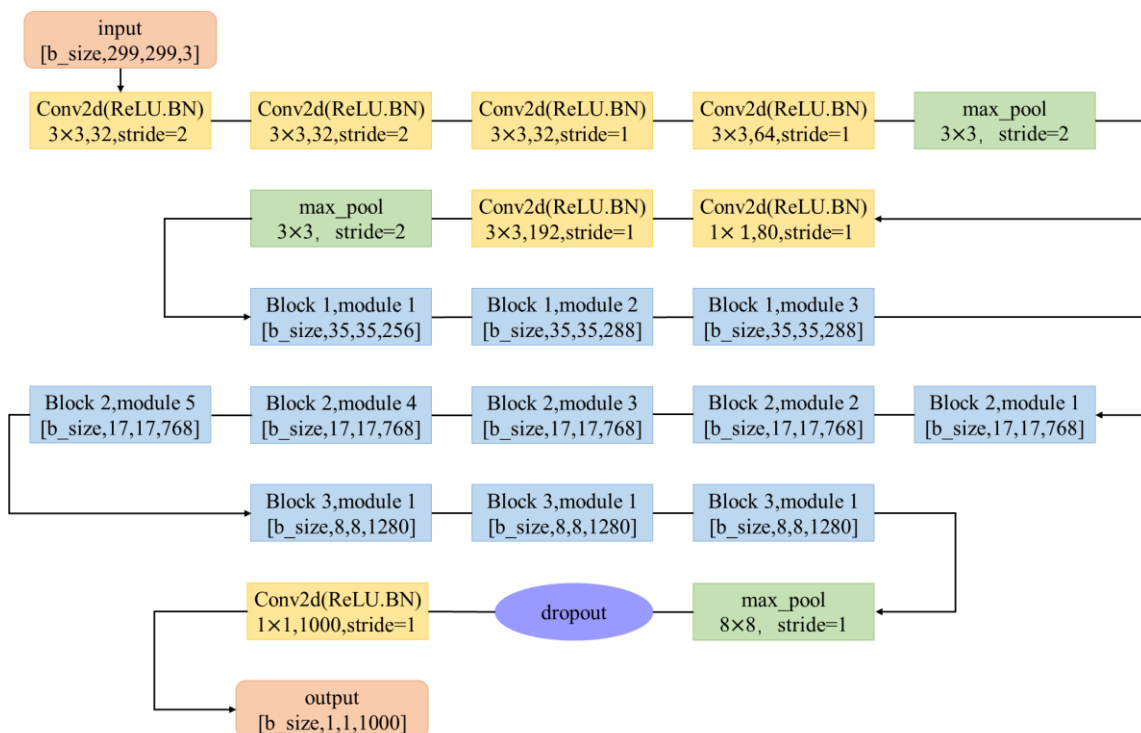
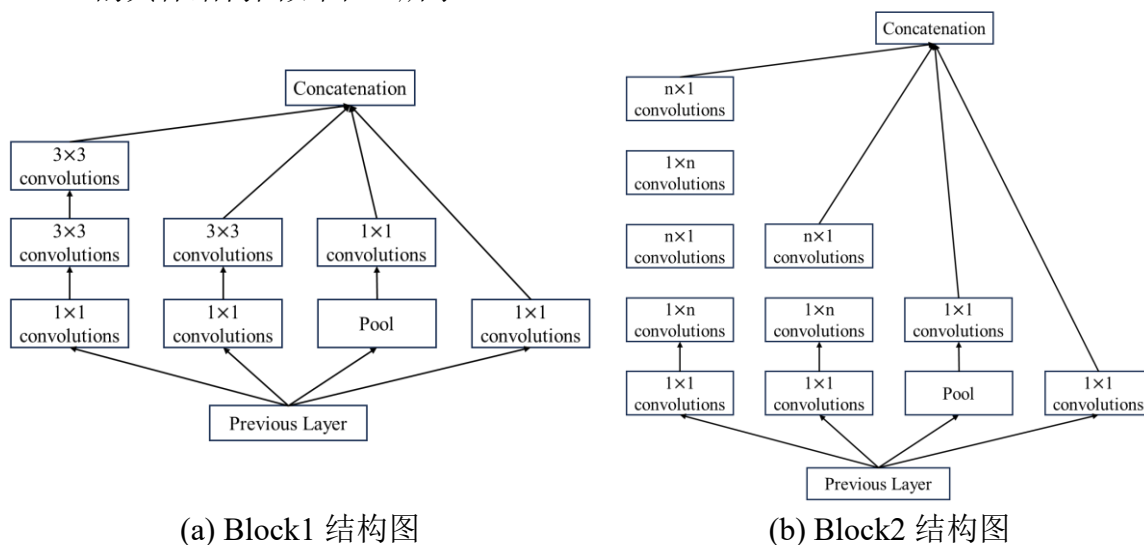
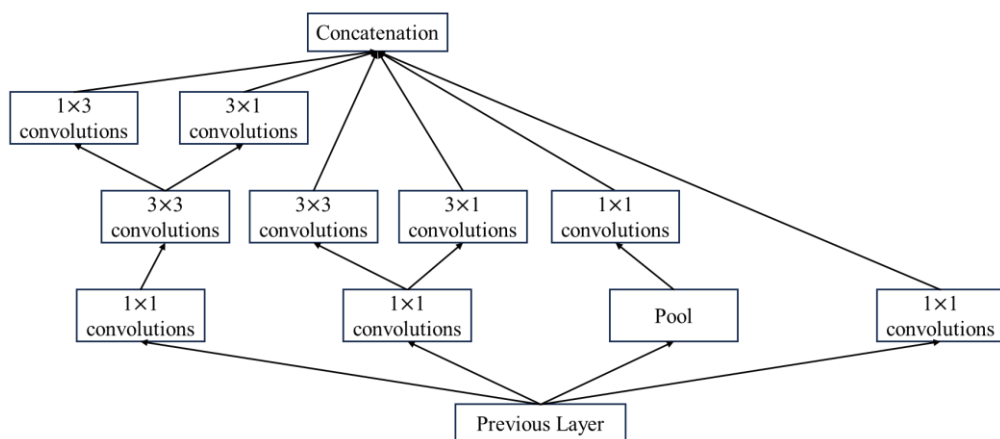


图 21 原始 Inception v3 网络结构图

上述 Inception_v3 网络结构主要包括三种 Block, 见图 21 蓝色层 Block1、Block2、Block3 的具体结构图如图 22 所示:





(c) Block3 结构图

图 22 Block 具体结构图

三、Advanced-Inception_v3 模型

原始 Inception_v3 模型的分类器中采用 softmax 函数作为激活函数，根据迁移学习理论，本文在 Inception_v3 模型的基础上将激活函数改为 sigmoid 函数以适应新的分类任务。Sigmoid 函数通过一种曲线映射，将输入值变换为介于 0 和 1 之间的输出值，其公式如下：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

sigmoid 函数与 softmax 函数相比，Softmax 输出是一个概率分布，所有类别的概率总和为 1。但 Sigmoid 输出是单独的概率值，对于每个输出神经元都是独立的，并不要求总和为 1。本文采用 sigmoid 激活函数，希望使得所建立的 Advanced-Inception_v3 模型更准确地探测图像中的道路坑洼特征，以提高道路坑洼的检测性能。

但改变激活函数从 Softmax 函数到 Sigmoid 函数可能会导致模型对道路分类任务的适应性产生重大影响。后续在问题二将通过模型训练进一步对 Advanced-Inception_v3 模型的网络结构进行调整，可能会从添加全局平均池化层、添加正则化、调整损失函数与优化器等方面考虑，以适应道路图像分类任务。

5.1.4 模型对比与模型选择

为从 Advanced-VGG16 模型与 Advanced-Inception_v3 模型中初步选择高性能的道路坑洼分类模型，本小问将进行较少迭代次数的简单模型训练，分别从训练开销、准确率的提升速度、相同迭代次数下的准确率三个指标选择合适的模型。下图 23 是 Advanced-VGG16 模型与 Advanced-Inception_v3 模型在 SGD 优化器、二进制交叉熵损失函数且没有添加正则化的条件下经过 20 轮迭代训练的模型准确率与损失率的对比图。

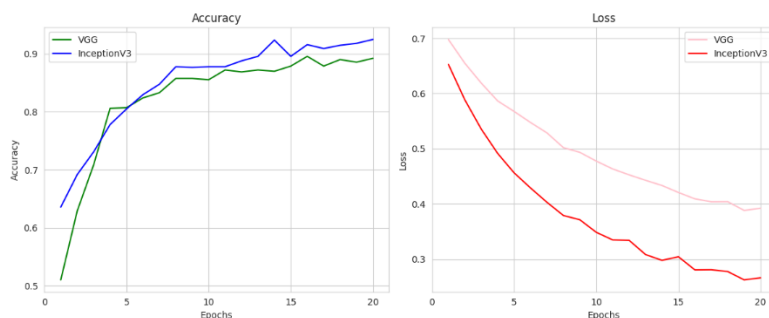


图 23 Advanced- VGG16 模型与 Advanced- Inception_v3 模型 accuracy 和 loss 对比图

从上图 23 可以看出，在相同迭代次数的情况下，Advanced-Inception_v3 模型的准确率除了在第 4 轮略微比 Advanced-VGG16 模型逊色外，其他轮次的准确率都高于 Advanced-VGG16 模型。同时，从总体上看，Advanced-Inception_v3 模型准确率的提升速度也比 Advanced-VGG16 模型的提升速度更快。

为比较两个模型的训练开销，本问将从"Total params"、"Trainable params" 和 "Non-trainable params" 三个参数评估模型复杂性与可训练性。其中，Total params 表示模型的总参数数量，Trainable params 表示需要通过训练来优化的参数，Non-trainable params 表示包含在模型中保持不变的参数。如下

表 1 所示，Advanced-Inception_v3 模型与 Advanced-VGG16 模型相比，其在这三个关键维度上都有显著的优势。

表 1 Advanced- VGG16 模型与 Advanced- Inception_v3 模型训练开销对比图

	Advanced-VGG16	Advanced-Inception_v3
Total params	134268738(512.19MB)	21806882(83.19MB)
Trainable params	8194(32.01KB)	4098(16.01KB)
Non-trainable	134260544(512.16MB)	21802784(83.17MB)

首先，考虑"Total params"，Advanced-Inception_v3 模型的总参数数量明显较少，而较小的模型规模意味着更低的存储需求，更适合部署在资源有限的环境中，如移动设备和边缘计算。此外，模型加载和传输速度更快，有利于实时应用。

其次，关注"Trainable params"，Advanced-Inception_v3 模型具有更少的可训练参数，由于需要更新的参数更少，使得模型的训练速度更快。同时，这降低了过拟合的风险，有助于模型更好地泛化到新数据。

最后，在"Non-trainable params"方面，Advanced-Inception_v3 模型同样领先。非可训练参数更少，使模型更简单，计算速度更快。这在需要快速推断的实时应用中非常有利，也提高了模型的可解释性，因为更简单的模型更容易理解和分析。

总之，Advanced-Inception_v3 模型在总参数数量、可训练参数数量以及非可训练参数数量上均明显优于 Advanced-VGG16 模型。这使得它成为适用于资源受限环境、需要快速训练和推断、以及对模型可解释性有要求的任务的理想选择。

综合 Advanced-Inception_v3 模型的准确率提升速度更快、相同迭代次数下准确率更加高、训练开销更小，本题将选取 Advanced-Inception_v3 模型作为识别坑洼道路的分类模型。

5.2 问题二模型的建立与求解

5.2.1 Advanced-Inception_v3 模型训练与模型优化

本问对上述所选择的 Advanced-Inception_v3 分类模型进行多次训练，从损失函数的选择、优化器的选择、是否添加全局平均池化层、是否添加正则化、是否添加早停函数五个维度对模型进行优化，从而建立一个识别率高、速度快、分类准确的模型用于自动检测图像的道路状况，识别图像是正常道路还是坑洼道路。

一、损失函数的选择

本问将在 Advanced-Inception_v3 模型训练中分别对比使用二元交叉熵损失函数与稀疏交叉熵损失函数，训练结果如下图 24 所示。从图中可以看出二元交叉熵损失函数的准确率明显高于稀疏交叉熵损失函数，同时训练损失也明显低于稀疏交叉熵损失函数。说明使用二元交叉熵函数作为损失函数可以使模型的分类性能更好。后续训练将在使用二元交叉熵损失函数基础上从其他方面进行模型改进。

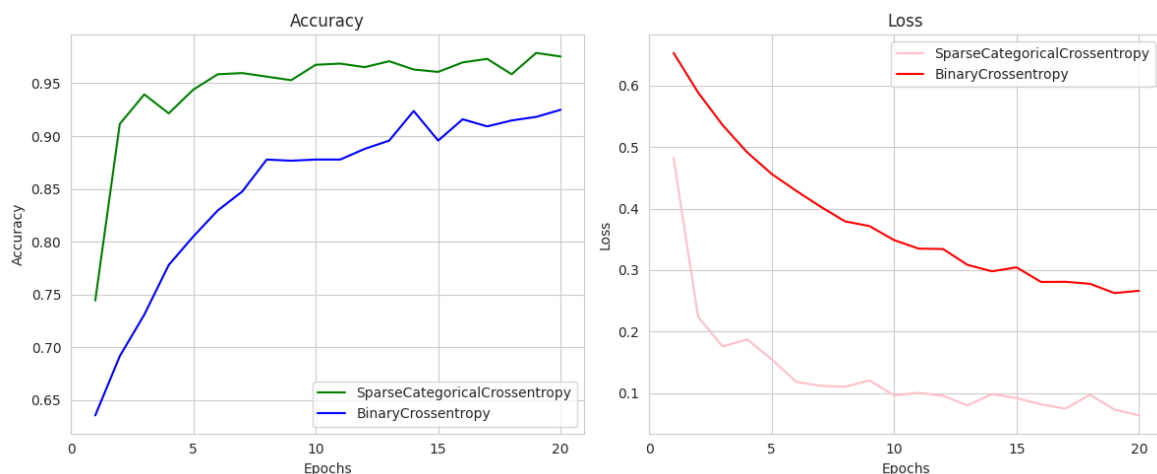


图 24 不同损失函数的 accuracy 和 loss 对比图

二、优化器的选择

本问在模型训练将使用 SGD 优化器与 Adam 优化器分类准确率与训练损失进行对比,如下图 25 所示。从图中可以看出 Adam 优化器的准确率明显高于 SGD 优化器,同时训练损失也明显低于 SGD 优化器。说明使用 Adam 优化器优化器可以使模型的分类性能更好。

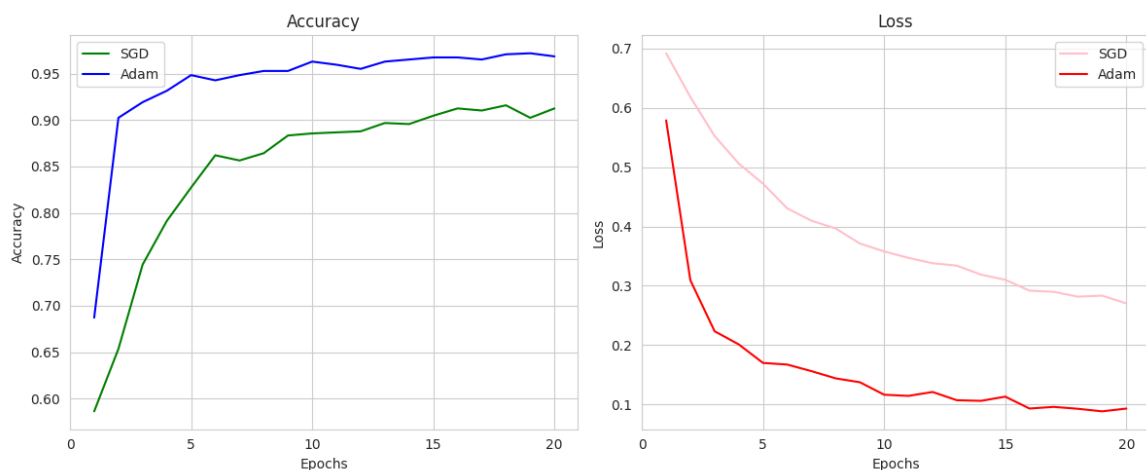


图 25 不同优化器的 accuracy 和 loss 对比图

后续训练将采用 Adam 优化器,其中 Adam 优化器被设计成结合了两个关键元素:动量 (Momentum) 和自适应学习率。这个算法的目标是提高训练速度和性能,同时克服传统优化算法的一些限制,其具体特点如下:

1.Adam 包含了动量项,有助于平滑梯度的变化,减少训练过程中的震荡,从而更快地更新模型参数。

2.Adam 能够自适应地为每个模型参数分配不同的学习率,根据每个参数的历史梯度信息来进行调整。

3.Adam 引入了一个偏差修正机制,以纠正梯度估计的偏差,有助于更好地适应学习率。

4.Adam 通常具有较好的数值稳定性,它能够处理不同尺度的梯度,因此不需要像传统优化算法那样频繁手动调整学习率和其他超参数,这减轻了训练模型的负担。

三、添加全局平均池化层

全局平均池化层在深度学习卷积神经网络中是一种重要的层类型。其主要作用是

将每个通道的图像中的所有像素值相加并求平均，最终生成一个新的 1×1 通道图。这有助于减少模型参数数量，减轻计算负担，降低过拟合风险，提高模型的泛化性能。图示如下：

为了防止过拟合，我们在 **Advanced-Inception_v3** 模型末端，即在全连接层之前，巧妙地引入了全局平均池化层。目的在于提高模型的泛化能力。

通过对整个特征图进行全局平均池化，我们可以将高维特征图压缩为一个单一值，从而减少模型的参数数量，降低计算复杂性。这有助于模型更好地适应不同数据集和避免过度拟合原训练数据。全局平均池化层的引入还有助于提高模型对输入的平移不变性，因为它不仅考虑最显著的特征，而且包含了整个特征图的信息，确保模型学习更全面的特征表示。这一策略在加强模型性能的同时，保持了网络结构的相对简洁性。

四、添加正则化

L2 正则化是一种在机器学习和深度学习中用于减少过拟合的技术。它的作用是通过向模型的损失函数添加一个正则化项，这个项基于模型权重的平方和，对权重进行惩罚，使其不过大。有助于防止模型过多依赖特定特征，提高泛化能力，减少过拟合，并改善模型性能。本文根据训练选择 `kernel_regularizer=l2(0.001)` 后模型的性能更佳。

"Dropout" 是一种深度学习正则化技术，目的是减少神经网络模型的过拟合。在训练时，Dropout 随机丢弃一部分神经元，即将它们的输出置为零，以一定概率。这导致每次迭代中不同的神经元被丢弃，迫使网络学习更鲁棒和泛化的特征。核心思想是降低神经元间的依赖性，防止某些神经元过于依赖某些特征，减少过拟合。

根据训练结果，本文在每一个全连接后面都有一个 Dropout 层用于防止过拟合，Dropout 的取值从上至下分别为 0.35、0.3、0.25、0.2；在此种情况下，最终呈现效果较优。

五、早停回调函数

为了缓解所建立的分类模型可能会在训练数据上表现得非常好，但对新数据的泛化能力较差的情况，本文在训练 **Advanced-Inception_v3** 模型的过程中加入早停回调函数，该函数会在性能停止提升时停止训练，而不会继续适应噪声或过多训练数据的特定特征，使模型的性能在训练期间变得更加稳定，以获得最佳的道路坑洼分类模型性能而不会出现过度拟合的现象。

根据上述的模型训练与模型优化本问在问题一所建立的 **Advanced-Inception_v3** 模型基础上对其神经网络结构做出了如下图 26 调整，在原模型已经具有全局平均池化层的基础上再添加一个全局平均池化层，且对网络结构中的全连接层进行改进，在每个全连接层后添加丢弃层，使全连接层和丢弃层交替连接，这种交替连接的设计能够有效地提高神经网络的鲁棒性和泛化能力，所得出的分类模型效果最佳。

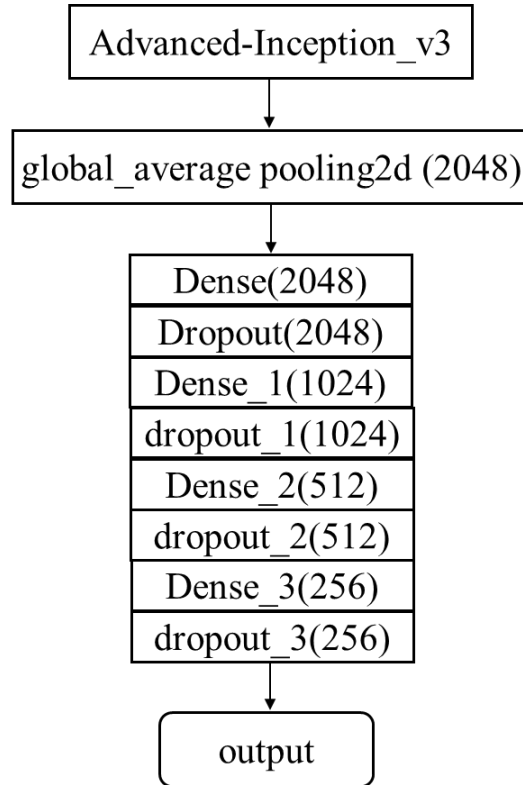


图 26 Advanced-Inception_v3 模型优化

对 Advanced-Inception_v3 模型进行优化后所得的道路坑洼分类模型在训练集与验证集上的分类准确率均一致收敛到 97%，并且是在较低的训练损失下达到较高的分类准确率，如下图 27 所示。

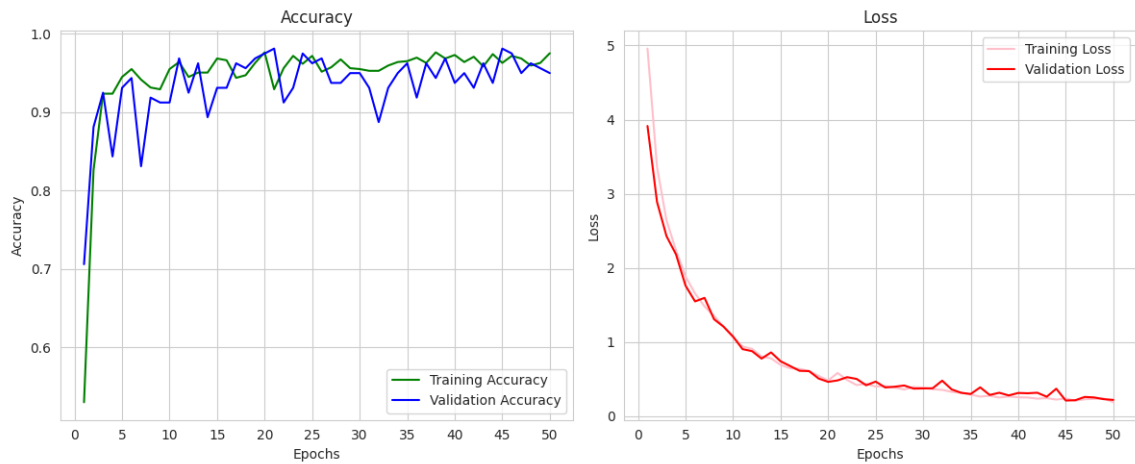
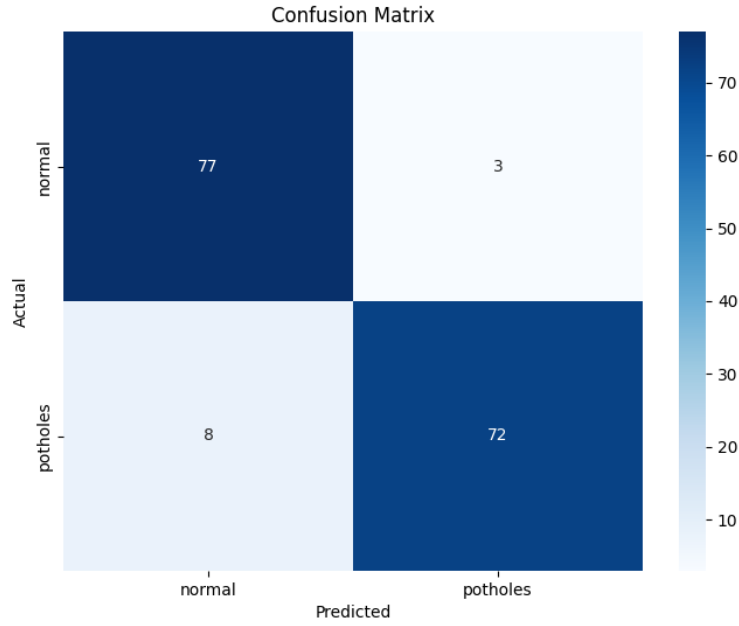


图 27 Advanced-Inception_v3 优化模型 accuracy 曲线与 loss 曲线

5.2.2 模型评估

基于混淆矩阵，计算下列准确率、召回率、查准率、F1 分数、Kappa 系数等评价指标。



（一）准确率

指被准确分类为正类别的样本数量与所有被分类为正类别的样本数量之间的比例。准确率易于理解，而且可以直观地反应模型的整体表现。其缺点是对不平衡数据和错误类型不敏感。公式如下：

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (7)$$

（二）召回率

用于评估分类模型性能的指标，它衡量了模型在所有实际正类别样本中能够正确识别的比例。它的优势是专注于检测所有真正正例，从而忽略了假负例的影响。缺点是在面对大量误报的情况下不太适用。公式如下：

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

（三）查准率

度量了模型在所有被预测为正类别的样本中，实际上是正类别的比例。简而言之，表示了模型的预测准确性。它的优点在于对假正例的敏感性较高，适用于不平衡数据。缺点是无法反应是否错过一些真正的正例。公式如下：

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

（四）F1 分数

F1 分数综合考虑精确度和召回率，用于评估分类模型的综合性能。其优点在于平衡了准确率和召回率，尤其在正负例分布差异大的情况下，适用于不平衡数据。公式如下：

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (10)$$

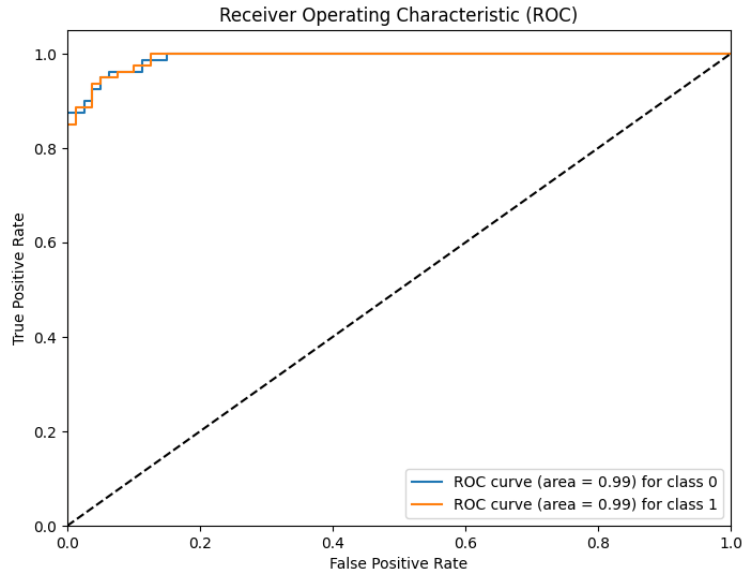
本文从坑洼道路和正常道路各随机抽取 80 张图片组成验证集。可见 Advanced-inception_v3 模型在精确度、召回率和 F1 分数方面均实现了出色的表现，均达到 0.9 以上。尤其是对坑洼道路判别的精确度高达 96% 这表明该模型在成功分类坑洼道路和正常道路方面展现了非常出色的能力。以下是具体的数据表 2：

表 2 评估指标表

	查准率	召回率	F1 分数	数量
normal	0.91	0.96	0.93	80
potholes	0.96	0.90	0.93	80

（五）ROC 曲线和 AUC

ROC 曲线是用于评估二元分类模型性能的工具。它以假阳率 FP 和真阳率 TPR 为坐标轴，绘制模型在不同阈值下的性能表现。AUC 是指 ROC 曲线下的面积，AUC 值越高，表示模型在不同阈值下的性能越好。如下图 ROC 曲线远离对角线，且向上弯曲，表明模型在不同阈值下具有很高的召回率和较低的误报率。而且 ROC 曲线下的面积为 0.99，表明模型相对与随机猜测来说性能很好。



（六）Kappa 系数

常用于影像分类的空间一致性检验的指标，也可以用于衡量分类的效果。在分类问题汇总一致性就是指模型的预测结果和实际分类结果是否一致。基于混淆矩阵的 Kappa 系数计算公式如下：

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \quad (11)$$

$$p_0 = \frac{\sum_{i=1}^c T_i}{n} \quad (12)$$

$$p_e = \frac{\sum_{i=1}^c a_i \times b_i}{n^2} \quad (13)$$

其中 p_0 是总体分类精度， C 是类别总数， T_i 是每个类别被正确分类的样本数，

a_1, \dots, a_i 是每一类的真实样本个数， b_1, \dots, b_c 是预测出来的每一类样本个数，所得为 Kappa 系数为 0.8625。

5.3 问题三模型的建立与求解

第三题要求我们对给定的测试集进行预测，我们将使用在训练过程中获得的优秀模型参数。在 Colab 平台上加载这些模型参数后，我们将进行一系列预测，并将结果保存在名为 test_result.csv 的文件中，部分结果展示如下图 28 所示：



图 28 部分分类结果示意图

可以看到，本模型预测结果不仅表现出色，即使在图像经过变形、裁切、模糊等处理后，模型也能准确识别，并且其置信度通常高达 95% 以上。这突显了该模型在精确度和抗干扰性方面的强大性能。

这一结果的展示，为我们展现了本文模型在实际应用中的强大能力，也为我们对模型性能的信心提供了坚实的基础。

六、模型的评价、改进与推广

6.1 模型的优点

1. 本文基于 VGG16，扩展至 inception_v3，选择慎重，尤其在图像分类和特征提取。通过一系列实验，微调模型以适应问题需求。深入研究后，演化 Advanced-inception_v3，经历多次改进。考虑任务需求，Advanced-inception_v3 在坑洼道路检测和分类上表现优秀。

2. 在模型改进过程中,我们采取了多种策略。通过调整损失函数以更好地适应任务需求,并采用更高效的优化器,本文实现了更快的收敛速度和提高的训练效率。此外,本文还尝试了数据增强技术,以增加数据多样性,提升模型的泛化能力。

3. 本文引入了全局平均池化层和早停函数,显著提升了模型的准确率和鲁棒性。全局平均池化有助于减少过拟合,而早停函数则有助于避免过度训练,使模型更加稳健和可靠。

4. 在评估 Advanced-inception_v3 性能时,专注验证集准确性。成功达到 X 准确率,显示其在坑洼道路分类上的可行性。实际预测效果分析确认了实际应用性能。综合而言,该模型在坑洼道路分类方面卓越,为帮助道路管理部门及时发现坑洼提供强大支持。

6.2 模型的缺点

1. 训练样本数量: 如果训练样本数量较小,可能会导致过拟合,特别是在添加了全连接层之后。

2. 可解释性: 由于模型结构较为复杂,可能难以直观理解模型在图像分类任务中的决策过程。

6.3 模型的改进

集成学习: 考虑将多个模型的预测结果结合起来,通过投票或者加权平均等方式,以获得更稳健和准确的预测。可以考虑集成学习的思想,将多个模型的预测结果结合起来,通过投票或者加权平均等方式,以获得更稳健和准确的预测,或者可以考虑分场景预测。

七、参考文献

- [1] 郑宇. 基于深度学习的样本不平衡问题研究 [D]. 重庆理工大学, 2023. DOI:10.27753/d.cnki.gcqgx.2022.000295.
- [2] 陈鹏. 基于深度学习的道路坑洼检测 [D]. 上海师范大学, 2020. DOI:10.27312/d.cnki.gshsu.2020.001756.
- [3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2021-01-01]. <https://arxiv.org/abs/1409.1556>.
- [4] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.
- [5] Simonyan, K., & Zisserman, A. (2015). Going Deeper with Convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 1-9).
- [6] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 2818-2826).
- [7] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the 31st AAAI conference on artificial intelligence (Vol. 4, p. 12).