

基于分组卷积与双注意力机制的河流水面污染图像分类^①



宋一格^{1,2}, 王宁^{1,2}, 李宏昌³, 武 曠⁴

¹(中国科学院大学, 北京 100049)

²(中国科学院 沈阳计算技术研究所, 沈阳 110168)

³(阜新市生态环境保护服务中心, 阜新 123008)

⁴(辽宁省生态环境监测中心, 沈阳 110161)

通信作者: 王宁, E-mail: wangning@sict.ac.cn

摘 要: 河流水面污染物是危害河流资源的重要污染物, 及时发现并处理水面污染物可以有效地保护河流环境以及水资源, 能进一步实现减污降碳, 提升生态系统碳汇能力. 随着智能化的大范围推广, 传统的河流水面污染物的监测处理方法已经不能满足当今的需求. 针对辽河流域水面污染问题, 本文将计算机视觉技术应用到了河流水面污染分类上, 提出了基于分组卷积与双注意力机制的河流水面污染图像分类算法模块 (grouped convolution dual attention, GCDA), 在分组卷积的基础上引入简化的双注意力机制, 使用较少的参数量增强了网络对图像的特征提取能力, 进一步提升图像分类效果. 通过固定截取图像的方式对辽河流域中的温泉城水站取水口、王营河入细河、高台子断面、津源污水排口和清源污水处理厂溢流口 5 个河流监控摄像图像做了预处理工作并建立了一个河流水面污染物数据集, 图像分为污染和未污染两类, 通过实验证明在此数据集上, 添加使用 GCDA 模块的网络相较于原网络以及分别添加空间、通道注意力机制的网络在河流水面污染图像的二分类任务中效果有明显提升.

关键词: 分组卷积; 注意力机制; 特征融合; 河流污染; 图像分类; 深度学习; 卷积神经网络

引用格式: 宋一格, 王宁, 李宏昌, 武曠. 基于分组卷积与双注意力机制的河流水面污染图像分类. 计算机系统应用, 2022, 31(9): 250-256. <http://www.c-s-a.org.cn/1003-3254/8688.html>

Image Classification of River Water Surface Pollution Based on Grouped Convolution and Dual Attention Mechanism

SONG Yi-Ge^{1,2}, WANG Ning^{1,2}, LI Hong-Chang³, WU Jian⁴

¹(University of Chinese Academy of Sciences, Beijing 100049, China)

²(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

³(Fuxin Ecological and Environmental Protection Service Center, Fuxin 123008, China)

⁴(Liaoning Ecological and Environmental Monitoring Center, Shenyang 110161, China)

Abstract: Water surface pollutants of rivers are the main pollutants that endanger river resources. Timely detection and treatment of water surface pollutants can effectively protect the river environment and water resources and further boost the pollution and carbon reduction and the carbon sink capacity of the ecosystem. With the widespread promotion of intelligence, traditional monitoring and processing methods for water surface pollutants of rivers can no longer meet the current needs. To address water surface pollution in the Liaohe River basin, this study applies computer vision technology to the classification of water surface pollution and proposes a classification algorithm module based on grouped convolution and the dual attention (GCDA) mechanism for images of water surface pollution. Specifically, a simplified dual attention mechanism is introduced into the network on the basis of grouped convolution, which uses fewer parameters to enhance the network's ability to extract features of images and further enhances the effect of image

① 基金项目: 辽宁省中央引导地方科技发展专项 (2021010211-JH6/105); 辽宁省科学技术计划 (2020JH2/10300113); 沈阳市科技计划 (20-203-5-50)

收稿时间: 2021-12-03; 修改时间: 2021-12-31; 采用时间: 2022-01-28; csa 在线出版时间: 2022-06-16

classification. The method of capturing images at a fixed position is performed on images from five river monitoring cameras in the Liaohe River basin for preprocessing. The five cameras refer to the ones in the hot spring intake of Chengshui Station, the confluence of Wangyinghe River and Xihe River, Gaotaizi Section, Jinyuan Sewage Outlet, and the overflow port of Qingyuan Sewage Treatment Plant. In addition, a dataset for water surface pollutants of rivers is established, and these images are categorized as polluted and unpolluted ones. Experiments indicate that compared with the original network and the network that adds space and channel attention mechanisms separately, the network with the GCDA module demonstrates better performance on this dataset in the dichotomous classification of images of water surface pollutants.

Key words: grouped convolution; attention mechanism; feature fusion; river pollution; image classification; deep learning; convolutional neural network (CNN)

1 引言

河流水面污染是水体污染的一种, 污染物通过直接或间接的方式被排入河流环境, 造成损害河流资源和人类健康的影响. 河流资源是淡水资源重要的组成部分, 地球上人类可利用的淡水资源占比不足 1%, 保护河流资源是非常迫切和必要的. 随着经济和社会的快速发展, 我国城镇化建设步伐日益加快, 大量的人口聚集到城市生活, 给城市周边的环境和生态造成了严重的影响, 许多河流的水面上出现大量的污染物, 这些污染物由各种自然和人工废弃物组成, 在水面堆积腐烂严重影响水环境和水质, 在城市和景观水域影响市容和旅游环境^[1]. 如果不能得到及时清理, 势必会破坏生态环境, 威胁人类的生存和发展.

本文主要的研究对象是河流水面的污染物, 例如垃圾塑料、自然脱落的树枝、富营养化藻类等污染物. 当前国内针对此类污染物处理主要是人工清理和遥感测试的两种方法. 人工清理的方式不能及时的清理污染物, 造成对河流环境的污染和河流生态的破坏, 不能有效的保护河流环境, 同时需要大量的人力物力. 遥感方法使用成本较高, 针对细长的河流会造成一定的技术图片和时间成本的浪费, 同时需要专业人员对遥感图像进行分析. 两种方法在时效性和经济成本上都不能很好的解决上述污染物的监测问题.

为解决上述问题, 本文结合实际情况和计算机视觉技术在相关领域的应用, 结合现有图像制作了一个河流水面污染图像数据集, 提出了一种基于分组卷积与双注意力机制的河流污染图像分类算法. 目前国内的主要河道都部署了河流监控摄像头, 通过拍摄图片的方式对河流资源进行监控, 本文提出的方法利用这些图像资源对模型进行训练测试并构建了一种基于分

组卷积和双注意力机制的算法模块, 将其添加到标准网络中可以有效的对河流水面污染物的图像进行分类, 判别河流水面污染的是否存在污染. 实验表明, 本文提出的 GCDA 模块添加到 ResNet50 算法上, 相较于原始网络以及单独添加注意力机制模块的网络在河流水面污染二分类实验中效果明显提升. 适用于实际工程, 能有效解决河流水面污染物监测问题, 有一定的实际意义.

2 相关工作

计算视觉技术是用计算机来处理图像, 从数字图像中提取特征, 使计算机具有人一样的视觉能力. 2012年, Krizhevsky 等^[2]创造的大型的深度卷积神经网络 AlexNet 在图像分类任务上, 以准确率提升 10 多个百分点的成绩取得了 ImageNet 竞赛的冠军. 这标志着卷积神经网络逐渐成为了计算机视觉中的图像分类和目标检测领域的主要方法. 计算机视觉技术在当今已经得到广泛应用, 例如车辆检测^[3], 人脸识别^[4], 自动驾驶^[5]等. 解决河流污染问题, 众多的企业和学者提出了自己的解决方法. 腾讯云提出了河道漂浮物识别方案. 华为云提出了智能水体 (WaterGo) 方案. 左建军等^[6]提出了水面漂浮物智能监控技术. 唐小敏等^[7]提出了基于 SSD 深度网络的河道漂浮物检测技术. 许静波^[8]提出了水面漂浮物监测及估算系统. 雷李义等^[9]提出基于深度学习的水面漂浮物目标检测方法.

分组卷积, 就是在对特征图进行卷积的时候, 首先对特征图分成小组再卷积. 用 g 表示分组大小, 与普通卷积比, 分组卷积的参数个数和计算代价都减少 g 倍. 这个 g 被称为基数, 它与深度和宽度一起描述了模型. 分组卷积首先在 AlexNet 被采用, 目的是将模型分到更多的 GPU 资源上, 后来被 Deep Roots^[10] 研究指出

卷积组可以更好地学习特征表示, ResNexts^[11] 提出即使在复杂情况下, 基数 g 的增加也能够提高分类精度。

注意力机制源于对人类视觉的研究, 人类视觉通过快速扫描全局图像, 获得需要重点关注的目标区域, 可以将有限的处理资源分配至重要的部分, 从而获得更多关注区域的相关信息. 在计算机视觉研究中, 主要有两种广泛使用的注意力机制: 空间注意力机制和通道注意力机制, 它们分别用于捕获像素级成对关系和通道依赖性. 注意力机制的意义在近年来得到了广泛的研究. 注意力机制偏向于信息更丰富的特征表达式, 从自然语言处理的机器翻译^[12] 开始使用, 到图像理解的图像字幕^[13] 都有广泛应用. 在计算机视觉研究中, 主要有两种广泛使用的注意力机制: 通道注意力机制和空间注意力机制, SENet^[14] 使用两个全连接层建立了一个轻量的门控机制, 通过通道关系来校正特征图. 在空间注意力机制方面, CBAM^[15] 通过计算特征图中每个空间点之间的相关矩阵引入了空间注意力机制并与通道注意力做了序列组合. 而 DANet^[16] 则是通过对来自不同分支的两个注意模块进行叠加, 实现了两种注意力机制的结合。

3 算法设计

本节是对 GCDA 模块结构的介绍. 该模块使用一种新的融合方法提升特征提取能力. 首先对输入特征分组, 然后对每组的子特征使用简化的双注意力机制提取特征, 对两种注意力机制进行特征融合后聚合为模块的输出. 使用较少参数增强模块的特征提取能力. 本节首先介绍了分组卷积的过程和优点, 然后介绍了双注意力机制的处理和融合, 最后是对模块在其他网络上添加使用的介绍。

3.1 分组卷积的参数量和计算量

分组卷积的合理使用能减少网络的参数量和计算量, 下面我们将介绍分组卷积的实现原理. 如图 1 所示。

图 1(a) 为普通卷积示意图, 此时的输入的特征图大小为 $W \times H \times C_1$, 分为对应的是特征图的宽、高和通道数. 单个卷积核大小是 $w_1 \times h_1 \times C_1$, 分为对应单个卷积核的宽、高和通道数. 输出特征图的大小是 $W \times H \times C_2$, 分别对应特征图的宽、高和通道数. 参数量是 $w_1 \times h_1 \times C_1 \times C_1 \times C_2$, 计算量 $w_1 \times h_1 \times C_1 \times C_1 \times C_2 \times W \times H$.

分组卷积是将普通卷积的输入特征图分成 g 组, 每个卷积核也相应地分成 g 组, 在对应的组内做卷积,

如图 1(b), 输入的特征图每组大小为 $W \times H \times C_1/g$, 分为对应的是特征图的宽、高和通道数, 共 g 组. 单个卷积核大小是 $w_1 \times h_1 \times C_1/g$, 对应单个卷积核的宽、高和通道数. 输出特征图的大小是 $W \times H \times C_2/g$, 分别对应的是特征图的宽、高和通道数. 参数量和计算量分别为: $\frac{w_1 \times h_1 \times C_1 \times C_1 \times C_2}{g}, \frac{w_1 \times h_1 \times C_1 \times C_1 \times C_2 \times W \times H}{g}$. 可以看出分组卷积的参数量和计算量是普通卷积的 $\frac{1}{g}$.

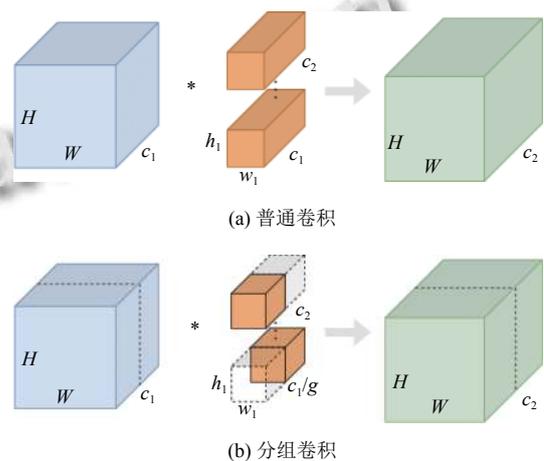


图 1 普通卷积和分组卷积示意图

3.2 分组卷积双注意力机制模块

注意力机制的使用能是有效的增加网络的提取特征能力, 通道注意力机制和空间注意力机制的有效结合能进一步提高网络效率, 同时也会带来参数量和计算量的提升. 为解决这一问题, 本文在使用分组卷积的基础上, 根据注意力机制的本质使用简化版的双注意力融合结构, 提出了 GCDA 模块, 在尽量少增加网络参数的情况下提升网络性能. GCDA 结构如图 2.

由图 2 可知, 分组卷积双注意力机制模块由分组、双注意力和聚合 3 部分组成. 首先, 特征提取的特征图先被分成 g 个小组. 然后, 每个小的特征图从通道注意力机制和空间注意力机制处理后进行特征融合. 最后, 得到的每个小的特征图在聚合为模块的输出. 网络通过嵌入分组卷积双注意力机制模块, 可以用较少的参数量使网络更关注有意义的特征. 具体的过程可以表示为:

$$out = \sum_{i=1}^g FS(X_1, X_2) \quad (1)$$

$$X_1 = \sigma(W_1 \cdot F_{gp}(F_i) + b_1) \quad (2)$$

$$X_2 = \sigma(W_2 \cdot GN(F_i) + b_2) \cdot F_i \quad (3)$$

其中, F_i 是输入特征 F 的第 i 个特征组, F_{gp} 是全局池化操作, W_1, b_1 是通道注意力特征权重和参数, W_2, b_2 是空间注意力特征权重和参数, σ 为 Sigmoid 函数^[17],

特征融合操作即先求和再做一次卷积, out 为 GCDA 模块的输出特征. 可以看出 GCDA 模块的输入输出维度都相同的, 这一特点提高了模块对其他网络适用性.

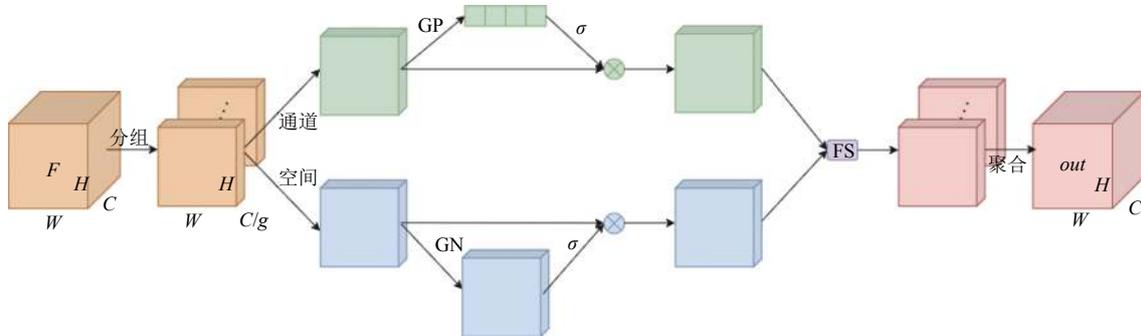


图2 GCDA 模块结构图

3.3 双注意力机制

下面我们将详细介绍 GCDA 模块中的双注意力机制的具体结构和流程.

通道注意力, 对于输入是图像的神经网络来说, 一个维度描述的是图像的尺度空间, 另一个维度就是通道. 经过特征提取器提取特征后, 每个通道描述图像的部分特征, 而对于各种类别, 不同特征通道的重要程度也不一样, 所以要给予其不同的权重来表示, 让分类器能够更加关注的正向的特征, 帮助分类器关注对网络有意义的通道.

SENet 使用通道注意力机制来实现了效果明显的图像分类, SENet 使用全局池化对空间维度进行压缩, 经过两次的全连接层和激活函数层实现了通道注意力机制. SENet 的目的是最大限度的提升准确性, 两次全连接层使得网路参数量增重. 本文使用了一次全连接层的方式实现了通道注意力机制. 结构如图3所示.

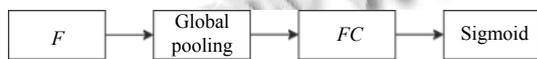


图3 本文通道注意力机制示意图

由图3可知, F 为特征向量, 通过简单地使用全局平均池 (GAP) 来嵌入全局信息, 生成信道统计信息, 它可以通过空间维度收缩来计算, 通过 Sigmoid 函数激活得到通道注意力机制的输出:

$$X_{i1} = \sigma(FC(s)) \cdot F = \sigma(W_1 \cdot s + b_1) \cdot F \quad (4)$$

$$s = F_{gp}(F) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F(i, j) \quad (5)$$

其中, $F \in \mathbb{R}^{c/g \times 1 \times 1}$, $s \in \mathbb{R}^{c/g \times 1 \times 1}$, $W_1 \in \mathbb{R}^{c/g \times 1 \times 1}$, $b_1 \in \mathbb{R}^{c/g \times 1 \times 1}$, σ 为 Sigmoid 函数.

空间注意力, 图像中并不是所有的区域对任务的贡献都是同样重要的, 只有与任务相关的区域才是注意的, 对于分类任务的主体, 空间注意力模型就是寻找网络中最重要的部位进行处理. 图像经过特征提取后形成特征图, 空间注意力机制能降低其他信息对分类器的干扰, 从而提升分类性能, 本文使用的空间注意力机制结构图如图4所示.

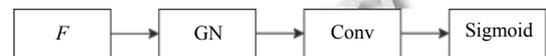


图4 本文空间注意力机制示意图

由图4可知, F 为特征向量, 使用 group norm^[18] 对 F 进行归一化处理获得空间信息, 在通过卷进操作增强, 最后经过 Sigmoid 激活函数激活获得空间注意力机制的输出.

$$X_{i2} = \sigma(W_2 \cdot GN(F) + b_2) \cdot F \quad (6)$$

其中, $F \in \mathbb{R}^{c/g \times 1 \times 1}$, $W_2 \in \mathbb{R}^{c/g \times 1 \times 1}$, $b_2 \in \mathbb{R}^{c/g \times 1 \times 1}$, σ 为 Sigmoid 函数.

最后, 是两种注意力机制的特征融合, 我们采用先求和在卷积的方式进行两种特征图的特征融合, 令模块输出的特征结合出两种注意力机制的特征.

$$X_i = W_3 \cdot (X_{i1} + X_{i2}) + b_3 \quad (7)$$

其中, $W_3 \in \mathbb{R}^{c/g \times 1 \times 1}$, $b_3 \in \mathbb{R}^{c/g \times 1 \times 1}$.

3.4 GCDA 模块部署

GCDA 模块由于输入和输出的特征图维度相同,

可以在神经网络的每个卷积后之后插入,所以 GCDA 模块可以集成到 GoogLeNet^[19] 和 ResNet^[20] 等标准的框架中. 本节最后,我们以结构比较复杂的 ResNet 为基础,介绍一下 GCDA 模块的部署使用方法. ResNet 是由 He 等^[20] 提出的网络模型,目的是解决神经网络的退化现象,即是模型随着网络结构的深度增加,会产生梯度消失问题,导致模型的训练逐渐困难且难于收敛. 这一方法的使用可是极大的加深网络深度,示意图如图 5 所示.

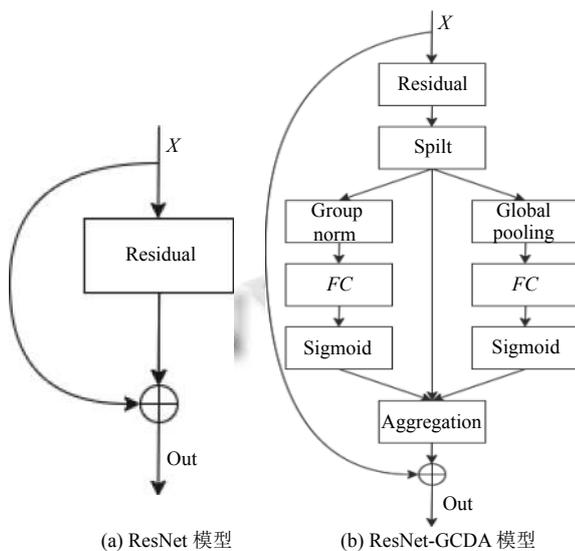


图 5 ResNet 部署 GCDA 示意图

由图 5 可以看到,我们将 GCDA 模块部署到了 ResNet 的主干分支中,使其在与分支求和之前发挥作用. 同时我们在实验部分做了消融实验,通过在网络上添加不同模块与 GCDA 模块进行对比实验,来证明 GCDA 模块的有效性.

4 实验结果及分析

4.1 实验环境及参数设置

本文实验在 Ubuntu 16.04.6 LTS 系统下利用 Python 语言,在百度飞浆平台完成,训练框架为 PaddlePaddle 2.1.2. 硬件信息为 CPU 是 Intel(R) Xeon(R) Gold 6271C,内存 32 GB,显卡为 Tesla v100,显存 16 GB,硬盘是 100 GB. 实验共设置 50 个迭代周期, batch size 设置为 64,选用 SGD 优化算法^[21] 训练模型,动量系数 0.9,学习率设置为 0.01 每间隔 10 个迭代周期学习率衰减 0.1 倍,使用交叉熵损失,定义如下:

$$L = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (8)$$

其中, M 为类别数, y_{ic} 为符号函数, p_{ic} 指观测样本 i 属于类别 c 的预测概率.

实验中采用 Accuracy, Precision 和 F1-score 作为评价指标,具体定义如式 (9)–式 (12) 所示:

$$F1\text{-score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (9)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

其中, TP , FP , FN , TN 分别为真正例,假正例,假反例和真反例.

4.2 数据预处理

本文实验数据源于辽河流域河流水面拍摄图像. 我们采用了辽河流域温泉城水站取水口、王营河入细河、高台子断面、津源污水排口和清源污水处理厂溢流口 5 个河流的监控摄像头连续 2 个月的拍摄图像,摄像头是固定机位,每 10 分钟拍一次图片. 由于晚上没有光源,我们选取早上 8 点到下午 6 点的拍摄图像作为原始数据,图片分为污染类和未污染类两个大类. 总的样本数为 8 987 张,污染类 3 812 张,未污染类 5 170 张. 其中训练集 6 822 张,验证集 1 098 张,测试集 1 067 张.

结合实际图片和实验结果我们发现直接使用事物原始图像由于拍摄视野较大包含了河道周围的草木行人等干扰因素,不利于网络对图片的分类. 如图 6 所示,图 6(a) 为温泉城水站取水口拍摄图像,除去河流水面部分还有大面积的草木土地等. 为解决这一问题,我们决定截取主要河流部分作为输入数据,如图 6(b) 所示. 由于摄像头机位固定,拍摄图片的角度和范围是不会变的,这种方法能保证截取图像始终是同一位置 有利于图像的批量处理,并且去除了其他干扰因素.

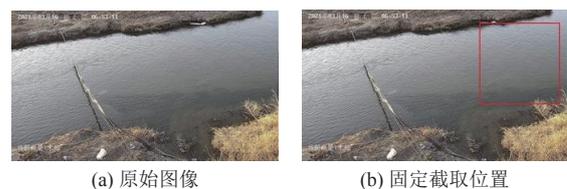


图 6 温泉城水站取水口图像截取示意图

为提升数据集的质量,我们在 VGGNet16^[22]、GoogLeNet 和 ResNet50 网络上分别对原始图像,随机

截取, 中心截取和固定位截取 4 种不同处理方法得到的数据集上做二分类对比实验, 结果如表 1 所示.

表 1 不同算法在 4 种数据集上的 $F1$ -score

处理方法	VGGNet16	GoogLeNet	ResNet50
原始数据	0.7324	0.7365	0.7368
随机截取	0.7017	0.7168	0.7115
中心截取	0.7431	0.7291	0.7457
固定位截取	0.7785	0.8035	0.8478

由表 1 可以看出采用固定位截取原始图片的方式在 3 种网络中都能有效的提高网络的准确率, 证明了本文数据预处理的必要性和有效性.

4.3 消融实验

为证明 GCDA 模块的融合双注意机制的有效性, 我们在实验参数设置相同的情况下, 以 ResNet50 为基础分别添加空间注意力机制模块、通道注意力机制模块和 GCDA 模块进行对比实验. 实验数据采用的上文描述的固定位截取方式获取的数据集做二分类任务. 训练过程的损失收敛曲线和准确度收敛曲线如图 7.

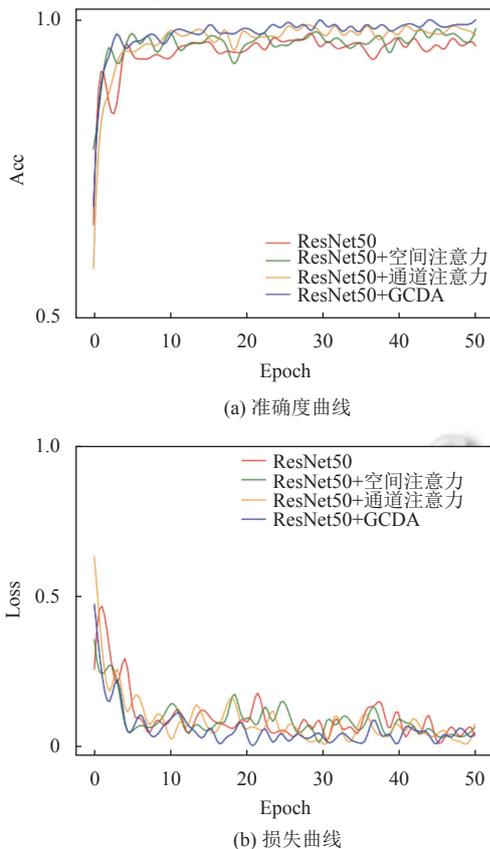


图 7 训练过程曲线图

由图 7 可以看出 4 种算法在 50 迭代后趋于收敛, 为避免过拟合现象, 我们最终将训练次数设置了 50 次.

添加 GCDA 模块的 ResNet50 相比其他 3 个算法训练效果较好. 然后在测试集进行测试, 并计算混淆矩阵:

$$\text{ResNet50} : \begin{pmatrix} 218 & 71 \\ 55 & 723 \end{pmatrix}$$

$$\text{ResNet50+通道注意力} : \begin{pmatrix} 227 & 62 \\ 56 & 722 \end{pmatrix}$$

$$\text{ResNet50+空间注意力} : \begin{pmatrix} 235 & 54 \\ 58 & 720 \end{pmatrix}$$

$$\text{ResNet50+GCDA} : \begin{pmatrix} 261 & 28 \\ 39 & 739 \end{pmatrix}$$

根据混淆矩阵采用 macro avg 方法计算 Accuracy, Precision 和 $F1$ -score 三个指标的数值, 具体结果如表 2.

由表 2 可以看出, 添加了 GCDA 模块的 ResNet50 网络在 Accuracy, Precision 和 $F1$ -score 三个指标上都有明显的提升. 为了更直观地验证 GCDA 模块的有效性, 我们使用 GradCAM^[23] 在 ResNet50 的基础上做了可视化的热力图. 如图 8 所示.

表 2 不同模块的实验结果

算法	Accuracy	Precision	$F1$ -score
ResNet50	0.8819	0.8546	0.8478
ResNet50+空间注意力	0.8894	0.8615	0.8591
ResNet50+通道注意力	0.8950	0.8661	0.8677
ResNet50+GCDA模块	0.9372	0.9167	0.9214

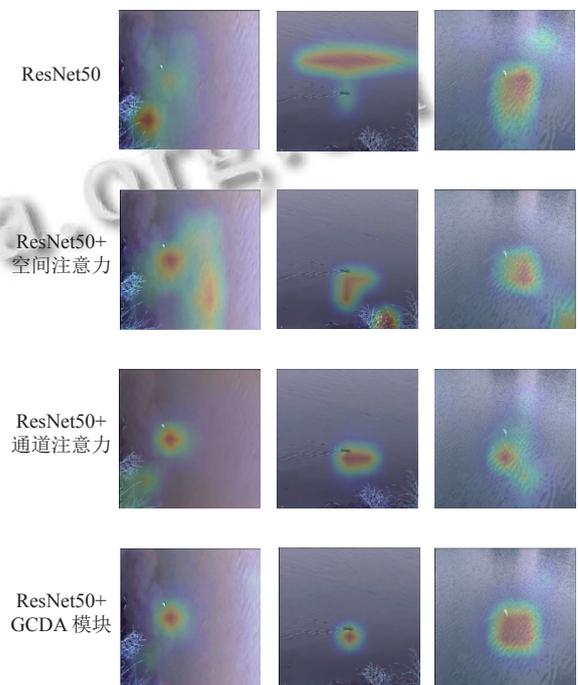


图 8 不同模块的 GradCAM 可视化图像

由图 8 可以看出, 添加了 GCDA 模块的 ResNet50 网络的热力图集中体现在水面污染物部分. 所以我们

提出的 GCDA 模块能够允许网络关注更多的、具有更多对象细节的相关区域,从而能有效的提升分类效果。

5 结论与展望

本文针对辽河流域河流水面污染分类困难问题,将计算机视觉技术中的分组卷积与双注意力机制应用到河流水面污染图像分类的任务中,结合现有图像制作了一个河流水面污染图像数据集,提出的 GCDA 模块使用分组卷积在保证准确率的情况下可以减少计算和参数量,同时引入简化的双注意力机制使模块能关注更有意义的特征信息,进一步的提升模块对图像分类的准确率。通过实验表明在河流水面污染图像二分类任务中使用 GCDA 模块的 ResNet50 相较于原始的 ResNet50 以及分别添加空间、通道注意力机制的网络有明显提升效果,满足了实际的应用需求,有良好的应用前景。

参考文献

- 1 蔡莹,唐祥甫,蒋文秀.河道漂浮物对工程影响及研究现状.长江科学院院报,2013,30(8):84-89.[doi:10.3969/j.issn.1001-5485.2013.08.019]
- 2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2012. 1097-1105.
- 3 徐子豪,黄伟泉,王胤.基于深度学习的监控视频中多类别车辆检测.计算机应用,2019,39(3):700-705.[doi:10.11772/j.issn.1001-9081.2018071587]
- 4 Sun Y, Liang D, Wang XG, et al. DeepID3: Face recognition with very deep neural networks. arXiv: 1502.00873, 2015.
- 5 张燕咏,张莎,张昱,等.基于多模态融合的自动驾驶感知及计算.计算机研究与发展,2020,57(9):1781-1799.[doi:10.7544/issn1000-1239.2020.20200255]
- 6 左建军,吴有富.水面漂浮物智能监控技术.软件导刊,2013,12(4):150-152.
- 7 唐小敏,舒远仲,刘文祥,等.基于 SSD 深度网络的河道漂浮物检测技术研究.计算机技术与发展,2020,30(9):154-158.
- 8 许静波.水面漂浮物监测及估算系统开发与建设.江苏水利,2018,(8):51-58.
- 9 雷李义,艾娇燕,彭婧,等.基于深度学习的水面漂浮物目标检测评估.环境与发展,2019,31(6):117-120,123.
- 10 Ioannou Y, Robertson D, Cipolla R, et al. Deep roots: Improving CNN efficiency with hierarchical filter groups. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5977-5986.
- 11 Xie SN, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5987-5995.
- 12 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv: 1409.0473, 2014.
- 13 You QZ, Jin HL, Wang ZW, et al. Image captioning with semantic attention. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 4651-4659.
- 14 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132-7141.
- 15 Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 2017-2025.
- 16 Fu J, Liu J, Tian HJ, et al. Dual attention network for scene segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141-3149.
- 17 Han J, Moraga C. The influence of the Sigmoid function parameters on the speed of backpropagation learning: Computational models of neurons and neural nets. International Workshop on Artificial Neural Networks. Malaga-Torremolinos: Springer, 1995. 195-201.
- 18 Wu YX, He KM. Group normalization. International Journal of Computer Vision, 2020, 128(3): 742-755. [doi: 10.1007/s11263-019-01198-w]
- 19 Szegedy C, Liu W, Jia YQ, et al. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 1-9.
- 20 He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770-778.
- 21 Bottou L. Stochastic Gradient Descent Tricks. In: Montavon G, Orr GB, Müller KR, eds. Neural Networks: Tricks of the Trade. 2nd ed. Berlin: Springer, 2012. 421-436.
- 22 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
- 23 Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 618-626.

(校对责编:孙君艳)