doi:10.3969/j.issn.1001-4616.2025.03.014

基于图像融合和注意力机制的图像分类

黄文秀[1],周术诚[2],陈新元[1,3],周忠眉[4],王榕国[1]

(1.福州工商学院工学院,福建 福州 350715)
(2.福建农林大学计算机与信息学院,福建 福州 350002)
(3.吉隆坡大学信息技术学院,马来西亚 吉隆坡 50250)
(4.闽南师范大学计算机学院,福建 漳州 363000)

[**摘要**] 图像分类作为计算机视觉领域的一个关键任务,在多种应用场景中具有重要意义. 针对图像分类中的准确性和鲁棒性问题,提出了一种基于图像融合和注意力机制的分类方法. 首先,选用了 ResNet-152 作为图像分类的基础模型,并对公开数据集进行了预处理. 在特征融合阶段,采用了三个并行支路,分别应用不同大小的卷积核进行特征提取. 随后,在残差网络结构后引入了注意力机制,综合了构建的格拉姆矩阵、平均池化和最大池化,以突出模型对分类有益的区域. 在实验阶段,在公共图像数据集上进行大量实验,结果表明所提出的方法在实际应用中表现出很好的效果,分类准确度从最初的 96.68%提高到 98.87%. 同时,相较于传统方法具有更好的鲁棒性. 因此,本研究为图像分类领域提供了一种有效的改进方法,具有广泛的应用前景.

[**关键词**] 图像融合,注意力机制,深度学习,图像分类,卷积网络
[**中图分类号**]TP301　[**文献标志码**]A　[**文章编号**]1001-4616(2025)03-0120-09

Image Classification Based on Image Fusion and Attention Mechanism

Huang Wenxiu[1],Zhou Shucheng[2],Chen Xinyuan[1,3],Zhou Zhongmei[4],Wang Rongguo[1]

(1.School of Technology,Fuzhou Technology and Business College,Fuzhou 350715,China)
(2.College of Computer and Information Sciences,Fujian Agriculture and Forestry University,Fuzhou 350002,China)
(3.College of Information Technology,Universiti Kuala Lumpur,Kuala Lumpur 50250,Malaysia)
(4.School of Computer Science,Minnan Normal University,Zhangzhou 363000,China)

Abstract:As a key task in the field of computer vision,image classification is of great significance in many application scenarios. Aiming at the accuracy and robustness of image classification,a classification method based on image fusion and attention mechanism is proposed. Firstly,ResNet-152 is selected as the basic model of image classification,and the public data set is preprocessed. In the feature fusion stage,three parallel branches are used,and convolution kernels with different sizes are used to extract features. Then,the attention mechanism is introduced after the residual network structure,and the Gram matrix,average pooling and maximum pooling are integrated to highlight the areas where the model is beneficial to classification. In the experimental stage,through a large number of experiments on public image data sets,the results show that the proposed method has a good effect in practical application,and the classification accuracy has increased from the initial 96.68% to 98.87%. In addition,compared with traditional methods,it has better robustness. Therefore,this study provides an effective improvement method for the field of image classification,which has a wide application prospect.

Key words:image fusion,attention mechanisms,deep learning,image classification,convolutional networks

在数字图像获取和存储技术飞速发展的大背景下,随着数字图像的普及,图像处理和计算机视觉领域正在迅速演进. 在这个发展趋势的带动下,图像分类作为一个核心问题引起了广泛的研究和关注. 图像分类的目标是将输入的图像准确划分为预定义的类别,为自动化识别系统提供基础支持,如人脸识别、医学影像分

收稿日期:2025-01-25.
基金项目:国家自然科学基金项目(61672159)、福建省自然科学基金项目(2022J01398)、福建省中青年教师科技类教育科研项目(JAT211024)、福建省中青年教师科技类教育科研项目(JAT211001)、福建省终身教育提质培优项目(ZS22026).
通讯作者:周忠眉,博士,教授,研究方向:数据挖掘、人工智能. E-mail:251128555@qq.com

析、智能交通等应用. 然而,图像分类面临着一系列挑战,包括图像的复杂性、多样性和噪声等因素,这些因素使得传统的图像分类方法在处理复杂场景和变化较大的图像时表现不佳. 为了迎接这一系列挑战,研究者们持续地探索并提出新的方法和技术,以不断优化图像分类的性能,使其更准确、更具鲁棒性.

近年来,随着深度学习技术的迅猛发展,计算机视觉领域取得了令人瞩目的成就,其中包括对象识别和目标检测等方面得到了广泛的应用[1]. 2022 年,Wang 等[2]创新性地引入了一种注重注意力机制的多标签图像分类模型,将注意力机制和 GCN 相结合,以端到端的方式同时捕获局部和全局标签相关性,为多标签图像分类任务提供了一种全面而有效的解决方案. 姜昊等[3]2023 年创新性的引入了特征融合金字塔模块(feature fusion pyramid module,FFPN),该模块通过提取不同层次的输出并实施特征对齐和融合,以有效地综合多尺度特征. 相对于传统的 FPN 模块[4],FFPN 在上采样过程中改变了融合方式,更为出色地突显了图像的细粒度信息. 然而,该方法只针对细粒度的图像,并没有捕获局部和全局的特征,因此实现图像的精确分类仍然是一项极具挑战性的任务.

针对上述问题,通过对相关文献的调研,深入研究了当前图像分类领域的相关算法和各种卷积神经网络的结构与特征. 目前针对将图像特征融合和注意力机制相结合用于图形分类的研究还比较少见. 因此,本文提出了一种基于图像融合和注意力机制的图像分类方法. 首先综述了深度学习在图像分类领域的应用现状,分析了现有方法的优势和局限性. 在此基础上,本文明确了研究思路:通过引入特征融合和注意力机制,优化 ResNet-152 模型,以提高图像分类的准确性和鲁棒性,具体方法包括构建并行特征提取分支,利用不同尺寸的卷积核捕捉多尺度特征;在残差网络后引入注意力机制,强化模型对关键区域的识别. 在公共图像数据集上的实验验证显示,本文所提方法在分类准确度和鲁棒性方面均取得了显著提升.

1 基于 ResNet-152 的深度残差网络模型架构

1.1 残差神经网络

按照理论推断,深度神经网络能够使权重空间上的损失函数更加平滑. 这样,在经过适当训练后,网络应当更容易找到优质的局部最小值,从而获得更高的精度. 然而,尽管经过长时间的研究和探索,深度神经网络在最大深度方面仍然受到一定限制. 试图通过强行叠加多层非线性变换[5](如卷积层和全连接层)来提高网络表达能力的尝试通常导致深度网络难以收敛,出现梯度问题,从而直接影响网络的训练效果. 在 ResNet[6]提出之前,神经网络的层数通常最多只有两位数,而 ResNet 的网络层数可以达到 152 层. 值得注意的是,尽管深度残差神经网络层数更多,但其计算复杂度相对于 VGG 网络[7]并不高,表明深度残差网络具有较低的复杂度.

2015 年,张婷婷等[8]首次提出了一种深度残差网络结构,如图 1 所示. 该图采用捷径连接的方式,将输入值 N 直接传递到后续网络层,而不仅仅作为第 $X-1$ 层的输入. 通过这种使用恒等映射的方法,既不增加大量参数以及计算的复杂度,还解决了随着网络层次加深导致的精度下降问题.

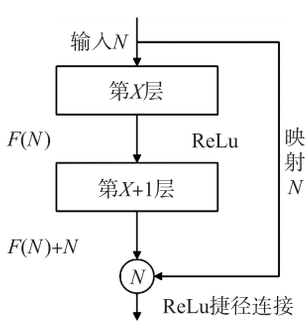


图 1 深度残差结构图

Fig. 1 Structure diagram of depth residual

1.2 图像分类基础模型选择

随着人工智能和深度学习技术的飞速发展,自 2012 年 AlexNet[9]提出后,国内外研究人员将目光聚焦在深度神经网络上,积极参与深度学习领域的研究. 为了应对图像分类的挑战,迫切需要比较和选择现有的算法模型. 目前,一些备受关注的算法包括 SVM[10]、KNN[11]、CNN[12]、AlexNet、VGGNet[13]、ResNet、GoogleNet[14]、Inception[15]等,其中,AlexNet、VGGNet、ResNet、GoogleNet 和 Inception 都被视为深度学习领域的经典.

为了寻找适用的基础分类模型,运用公开的 ImageNet 数据集进行了涉及五种模型的图像分类实验. 实验设置进行 1 000 次的迭代,通过对比各模型在分类准确度和损失值方面的表现(表 1),旨在选定最为合适的深度模型结构,作为进一步改进的基础模型.

经过对表 1 的观察,可以明显看出 Inception 模型和 ResNet 模型(包括五个不同结构的模型)在训练集和测试集上都取得了显著的精度优势,并且损失值相对较低. 针对选择基础模型的问题,考虑了

Inception 结构和 ResNet 结构. 虽然 Inception 结构在复杂性和参数数量上相对较高,但需要大量计算资源,可能造成计算机资源的浪费. 可见增加深度卷积神经网络的层数通常对于处理多分类和非线性问题有着显著的积极效果,因此最终选用 ResNet-152 模型作为改进的基础模型.

表 1 ImageNet 数据集上各网络模型的性能对比

Table 1 Performance comparison of each network model on ImageNet dataset

网络模型	训练集		测试集		网络模型	训练集		测试集	
	准确度	损失值	准确度	损失值		准确度	损失值	准确度	损失值
LeNet	0.732	0.760	0.743	0.751	Inception V3	0.958	0.259	0.914	0.482
AlexNet	0.842	0.288	0.767	1.711	ResNet-18	0.962	0.199	0.908	0.469
VGG-16	0.952	0.132	0.873	0.961	ResNet-34	0.971	0.183	0.919	0.458
VGG-19	0.966	0.180	0.891	0.831	ResNet-50	0.970	0.184	0.921	0.432
GoogleNet	0.969	0.181	0.921	0.587	ResNet-101	0.973	0.185	0.922	0.429
Inception V2	0.948	0.195	0.908	0.491	ResNet-152	0.975	0.169	0.925	0.398

ResNet-152 是深度残差神经网络系列中的一种模型. 它是 ResNet 系列中一个较深的模型,具有 152 层的深度,相对于之前的 ResNet-34,提供了更高的网络容量和表示能力. 深度网络通常能够更好地捕捉图像的复杂特征和语义信息,有助于提高模型的表示能力. ResNet-152 在图像分类领域的先进性能证明了其在复杂任务上的有效性. 因此,选用 ResNet-152 作为基础网络模型.

1.3 注意力机制

近来,注意力机制模型[16]在计算机视觉领域崭露头角,广泛应用于各种任务,其中如图像分类,已成为该模型的重要应用领域. 在计算机视觉领域,引入注意力机制的方式有多种变化. 一种常见的区分是硬注意力和软注意力[17]. 硬注意力将注意力权重二值化为 0 或 1,而软注意力采用概率分布的形式,将注意力权重调整为在 0 到 1 之间的连续值. 相对而言,软注意力机制更加灵活,能够实现对权重的微调,因此在图像分类领域中更为常见且被广泛应用.

对于给定的输入图像 $X \in R^{h\times w\times 3}$,其标签为 $Y \in R^{C}$,采用 one-hot 编码. 这里,h 表示输入图像的高度,w 表示输入图像的宽度,C 为图像标签的类别数量. 首先,将图像输入某个基线卷积神经网络(convolutional neural network,CNN)的主干部分 ϕ,通过该基线 CNN 的多个卷积层、激活函数和池化操作,得到特征图 $F \in R^{h\times w\times 3}$,其中 h 为特征图的高度,w 为特征图的宽度. 通常情况下,基线 CNN 的特征图 F 经过池化层的处理[18],其定义如下:

$$F_{\text{avgpool}} = \frac{\sum_{i,j}^{w,h} F_{i,j}}{w \times h}. \tag{1}$$

在基线 CNN 中,经过池化操作后的特征图 F 通常被送入全连接层,如下所示:

$$A_{\text{avgpool}} = F_{\text{avgpool}} \times W_{fc}, \tag{2}$$

式中,W_{fc}代表全连接层的权重. 为了简化表示,这里省略了偏差 b.

在经过池化操作后,针对细粒度图像分类任务,由于标签往往涉及多个类别,基线 CNN 通常采用 softmax 函数进行激活,其数学表达如下:

$$y = \frac{e^{A_{\text{avgpool}}}}{\sum_{i}^{c} e^{A^{i}_{\text{avgpool}}}}, \tag{3}$$

式中,y 表示网络输出的预测标签,网络通过对损失函数的优化来进行训练. 在这里,使用了交叉熵损失函数如下所示:

$$L(y,\hat{y}) = -\sum \log(y) \times \hat{y}, \tag{4}$$

式中,$\hat{y}$ 表示实际标签的真实值.

在细粒度图像分类问题中,为了更好地学习精细特征,通常将基线 CNN 用作原始图像的特征提取器. 在全卷积网络(full convolution network,FCN)[19]中,研究者通过使用 1×1 的卷积层替代全连接层,生成了高维特征表示图 heatmap,并在这一表示的基础上进行了细致的图像语义分割. 这一操作如下所示:

$$A = F \times W_{1\times 1Conv}, \tag{5}$$

式中,F 代表基线 CNN 在平均池化层之前提取的特征,$W_{1\times 1Conv}$则表示 1×1 卷积层的权重.

经过 1×1 卷积层处理和全局池化后得到的特征 A^{*}_{avgpool},即通过基线 CNN 和 FCN 特征提取器获得的类激活映射(class activation mapping,CAM)[20]. 这个特征是后续多重注意力机制将要处理的关键. CAM 有助于网络聚焦在特定的局部区域.

$$A^{*}_{\text{avgpool}} = \frac{\sum_{i,j}^{w,h} A_{i,j}}{w \times h}. \tag{6}$$

2 基于图像融合和注意力机制的图像分类方法

2.1 图像特征融合机制

多尺度卷积核的特征提取已被多项研究证明能够有效提升模型性能. 例如,Neshat 等[21]提出的 Inception 架构通过并行使用不同尺寸的卷积核(如 1×1、3×3、5×5),显著增强了模型对图像局部和全局特征的捕捉能力. 类似地,Luo 等[22]在 ResNeXt 中引入分组卷积与多分支结构,通过组合不同感受野的卷积核,进一步提高了特征的多样性和分类鲁棒性. 上述研究表明,并行使用不同大小的卷积核可通过捕捉多粒度特征来增强模型的表达能力,为本文的多支路设计提供了理论支持.

图 2 展示了本文的优化策略,通过对输入部分的改进,引入并行特征提取的机制,并在支路末端将提取的特征融合,实现了多特征的有效整合. 具体实现过程中,采用了三条平行支路,每一支路使用不同大小的卷积核. 第一支路使用 1×1 的卷积核进行卷积操作,用于降维和增强特征提取能力;第二支路和第三支路采用 3×3 大小的卷积核,将多尺度的特征合并,使得下一层网络能够捕捉不同尺寸的特征. 所有支路中卷积操作的步长均设为 1,以保持输入图像的大小. 最后,在两条支路的末端进行特征融合,既提升了模型的特征提取能力,又增强了模型的非线性表达能力.

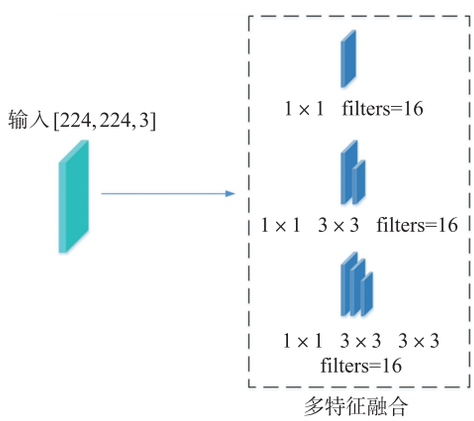


图 2 提出的图像特征融合机制

Fig. 2 The proposed image feature fusion mechanism

具体的特征融合方法表现为:假设某一种图像类别在经过三条支路进行特征提取后,得到的每条支路的具体特征向量如下:

$$\begin{aligned} &C_{1,1,1}, \quad C_{2,1,1}, \quad \cdots, \quad C_{16,1,1};\\ &C_{2,1,1}, \quad C_{2,1,2}, \quad \cdots, \quad C_{16,1,2};\\ &\vdots\\ &C_{1,1,16}, \quad C_{2,1,16}, \quad \cdots, \quad C_{16,1,16}, \end{aligned} \tag{7}$$

式中,$C_{a,b,c}$表示 a,c 大小的第 b 种类别图像的分块. 通过计算每条支路提取的特征与标准样本之间的欧式距离,形成相应的隶属度矩阵:

$$\begin{aligned} &m_{1,1,1}, \quad m_{2,1,1}, \quad \cdots, \quad m_{16,1,1};\\ &m_{2,1,1}, \quad m_{2,1,2}, \quad \cdots, \quad m_{16,1,2};\\ &\vdots\\ &m_{1,1,16}, \quad m_{2,1,16}, \quad \cdots, \quad m_{16,1,16}, \end{aligned} \tag{8}$$

式中,隶属度定义为:

$$m_{a,b,c} = \frac{1}{1+\left(\dfrac{d_{a,b,c}}{d_{b,c}}\right)^{t}}, \tag{9}$$

$$d_{a,b,c}=C_{b,c}-C_{a,b,c}, \tag{10}$$

$$d_{b,c}=\frac{1}{10}\sum_{m=1}^{10} d_{a,b,c}, \tag{11}$$

式中,$C_{b,c}$代表整体未分类的图像,而 $C_{a,b,c}$表示标准图像的特征向量. 通过计算这两个向量之间的欧氏距离($d_{a,b,c}$),引入模糊因子 t(控制模糊度,经过实验确定 $t=0.25$ 时效果最佳),计算得到隶属度矩阵.

最后,对于每个类别的图像,沿着列方向对其隶属度矩阵进行求和,形成融合结果,从而得到最终的隶属度向量$[m1,m2,m3,m4]^{\mathrm{T}}$,即特征融合的最终结果.

2.2　特定区域注意力机制

图 3 展示了本文构建的注意力机制,其关键功能是集中关注图像中特定区域,从而提高模型的分类性能. 因此,本研究的关注点是如何使模型能够精确地关注图像中的显著差异区域. 该注意力机制主要分为全局和局部两个部分.

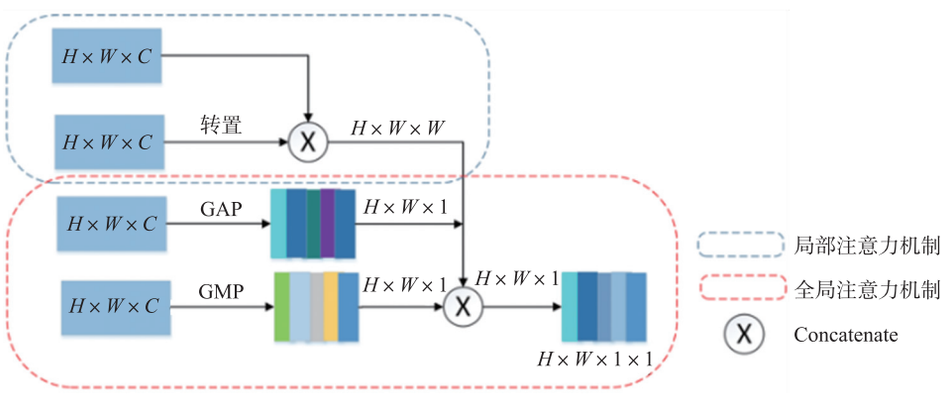


图 3　特定区域的注意力机制

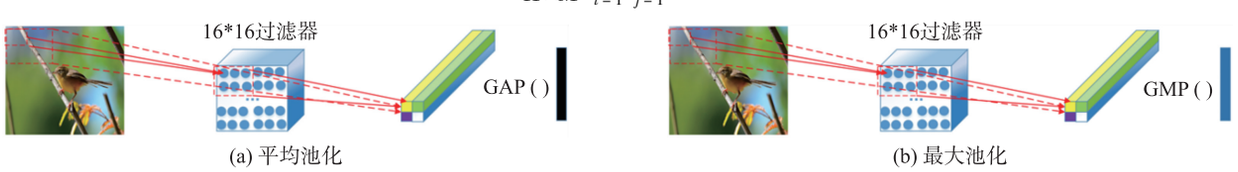
Fig. 3　Attention mechanism of specific regions

如图 4 所示,全局注意力机制在第一个残差网络模块中的神经元负责捕捉图像中一个小区域的特征. 这个小区域的大小是由 16×16 的过滤器决定的,被称为图像的神经元感受区. 16×16 的过滤器能够在第一个残差网络模块中捕捉到图像中的一个较大区域——“神经元感受区”,从而提供足够大的感受野来感知全局特征. 此外,在 ResNet 等模块化网络中,16×16 与后续模块的架构层次相匹配,便于全局和局部注意力机制的融合与衔接. 随着网络层数的增加,每个神经元对应的感受区也随之扩大,因为下一个残差网络模块与前一层中相同的 16×16 区域进行卷积操作. 在图中,黄色特征向量经过第一个残差网络模块,其感受区域由红色虚线框表示. 其他特征向量(绿色、蓝色等)同样代表图像中不同的感受野. 最终通过进行池化操作,将从图像中提取的关键特征转换为一个向量表示. 具体而言,在图中展示了两种不同的池化方法,分别是平均池化和最大池化. 这一步骤旨在提炼出图像中最显著和有代表性的信息,以便更有效地进行后续处理.

在经过第一个残差网络模块后,对特征图 U 执行池化操作,分别得到了用 $\boldsymbol{F}_A$ 和 $\boldsymbol{F}_M$ 表示的两个大小为 $H\times W\times 1$ 的特征图(公式(12)和(13)). 同时采用平均池化和最大池化的并行操作,相较于仅进行平均池化,能够更好地维持信息,弥补了池化操作可能带来的信息丢失.

$$\boldsymbol{F}_A=\frac{1}{H\times M}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j). \tag{12}$$

$$\boldsymbol{F}_M=\frac{1}{H\times M}\sum_{i=1}^{H}\sum_{j=1}^{W} u_l(i,j). \tag{13}$$



(a) 平均池化　　(b) 最大池化

图 4　残差网络模块后的两种不同池化方法

Fig. 4　Two different pooling methods after residual network module

局部注意力机制针对每个残差网络结构的输出(用 $\boldsymbol{F}_i$ 表示,大小为 $H\times W\times C$),借助格拉姆矩阵[23]构建了一个注意力机制模块. 通过将 $\boldsymbol{F}_i$ 与其转置 $\boldsymbol{F}_i^{\mathrm{T}}$ 相乘,得到 $\boldsymbol{F}_L$,其大小 $H\times W\times W$ 为:

$$\boldsymbol{F}_L=\boldsymbol{F}_i\times\boldsymbol{F}_i^{\mathrm{T}}. \tag{14}$$

通过将输出值与其矩阵相乘,巧妙地使特征图中原本数值较大的像素点值增大,而数值较小的像素点值减小. 这种处理方式有效地放大了图像中需要重点关注的区域,有助于在进行类别判断时排除无关紧要的特征信息.

最终,将经过不同池化操作的 $\boldsymbol{F}_A$ 和 $\boldsymbol{F}_M$ 以及 $\boldsymbol{F}_L$ 相乘,即综合考虑了一个全局特征和两个局部特征,得到了图像的总体特征 $\boldsymbol{F}$. 这种综合利用全局和局部信息的方法有助于更全面地捕捉图像的关键特征,提高了图像分类的准确性.

$$\boldsymbol{F}=\boldsymbol{F}_L\times\boldsymbol{F}_A\times\boldsymbol{F}_M. \tag{15}$$

2.3 图像分类方法总体流程

在应用 ResNet-152 深度残差网络的图像识别中,常面临两个关键问题:首先,由于图像的表面特征相似,例如厕纸和纸张,在图像特征提取阶段易发生混淆;其次,不同种类图像在多个方面存在相似性,导致传统的 Softmax loss 函数难以准确分类,容易造成误判. 为解决这些挑战,本文提出了一种经过优化的残差网络结构 All-ResNet-152,该结构能够有效聚焦图像中显著差异区域,提高图像分类的准确性. 这一深度残差网络结构包含输入层、卷积层、池化层、全连接层和输出层核心部分.

本文所提 All-ResNet-152 结构如图 5 所示,经过了三项关键的优化措施. 首先,引入了多特征融合,旨在充分整合输入图像的多方面信息. 其次,在残差网络模块之后,巧妙地加入了注意力机制,使模型能够有针对性地聚焦于对分类至关重要的区域,从而实现更为准确的特征提取. 最终,联合使用 Softmax loss 和 Center loss 函数进行网络参数的优化,增强了网络的性能和鲁棒性. 这些改进共同构筑了一个更为高效和精密的模型.

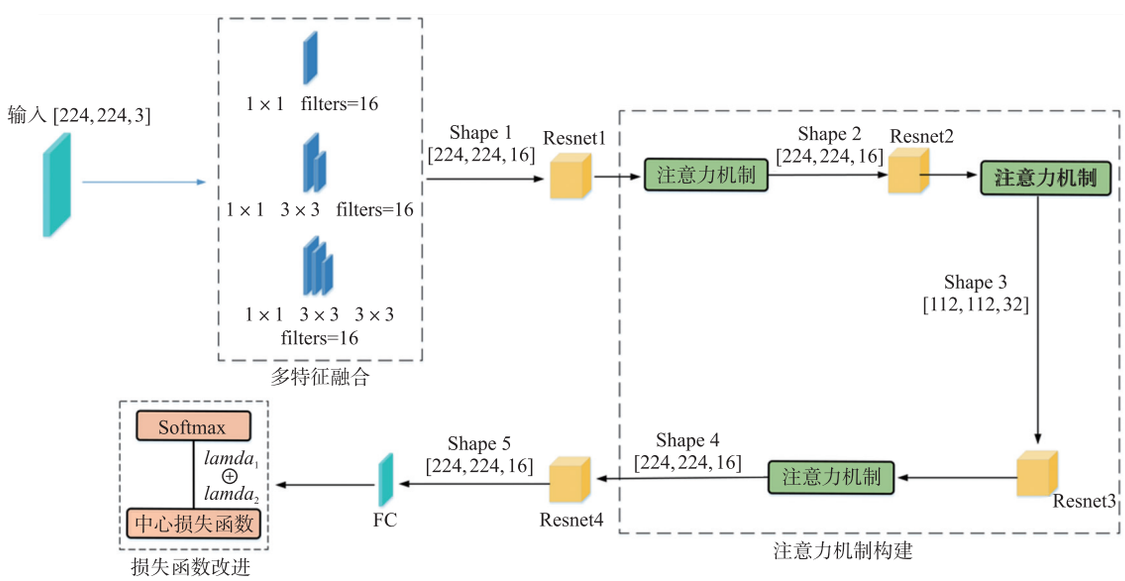


图 5 All-ResNet-152 结构的流程图

Fig. 5 The flow chart of All-ResNet-152 structure

3 实验结果与分析

3.1 实验环境与评价指标

实验阶段,本文选择在一台服务器上进行研究,主要运用了深度学习框架 PyTorch 进行实验. 具体的软硬件配置细节如表 2 所示.

在评估深度学习以及机器学习模型的性能时,常用的指标有很多. 本研究着重使用准确度、召回率和 $F1$ 分数三项指标,以全面、多角度地评估所构建模型的性能表现. 这些指标能够提供模型在图像分类任务中整体表现的量化度量.

(1)准确度是用来度量模型在图像分类任务中预测正确的能力的指标. 它以正确预测的图像数量占总图像数量的比例来衡量模型的整体性能,如下所示:

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN},\tag{16}$$

式中,TP 表示混淆矩阵中成功识别为正例的样本数量,TN 表示成功识别为负例的样本数量,FP 表示将负例错误地识别为正例的样本数量,FN 表示将正例错误地识别为负例的样本数量. 这些指标是评估深度学习模型性能的关键要素,帮助更全面地了解模型在不同情境下的表现.

表 2　实验配置参数

Table 2　Experimental configuration parameters

环境	组成	配置
软件环境	操作系统	Ubuntu
	编程语言	Python3.7
	深度学习框架	Pytorch 1.10.1
硬件环境	CPU	Intel Core i9-13900K
	显卡	GTX 4090
	显存	24 GB
	CUDA	11.4

(2)召回率是评估模型识别出真正正例的能力,计算方式为真正例的数量除以实际为真的样本总数量,如下所示:

$$Recall=\frac{TP}{TP+FN}.\tag{17}$$

(3)$F1$ 分数是综合考虑了精确度和召回率的评价指标,具有平衡性能的优势. 通过 $P-R$ 曲线,可以调整分类阈值,观察模型在不同条件下的表现,从而更全面地了解其性能. $F1$ 分数和 $P-R$ 曲线为评估模型在图像分类任务中的有效性提供了有力的工具,如下所示:

$$F1=\frac{2\times Accuracy\times Recall}{Accuracy+Recall}.\tag{18}$$

3.2　实验数据与对比实验

本文使用公开 CUB-200-2011 鸟类数据集[24],对提出的网络模型进行实验验证. CUB-200-2011 鸟类数据集是细粒度图像分类研究中的重要数据集. 总计包含 11 788 张图像,分为 200 个子类,其中训练集和测试集分别包含 5 994 和 5 794 张图像. 该数据集不仅为每张图像提供了边界框和关键点信息的详细注释,还附带了有关鸟类个体属性的额外信息,为研究者提供了更全面的数据资源.

本研究通过在输入层引入三条并行支路的多特征融合策略,有效将多尺度特征汇聚在一起,显著提升了模型对复杂特征的提取和拟合能力. 为了确定最优的特征融合情况,比较了单一支路、两条并行支路、三条并行支路以及四条并行支路这四种方案在准确度和 $F1$ 分数上的表现.

在实验中,对四种特征融合情况进行 2 400 步的迭代训练. 从图 6 和图 7 的结果来看,随着并行支路的增加,模型的准确度逐渐提高,在并行三条支路时最高达到 99.7%的峰值. 然后,准确度呈下降趋势. 类似地,$F1$ 分数在并行三条支路时达到最大值. 因此,实验证明在输入层采用三条并行支路进行多特征融合,能够显著提高图像分类模型的准确性和综合性能.

通过在 ResNet-152 模型中嵌入格拉姆矩阵构建的注意力机制,本文进行了一系列实验,旨在评估其对图像分类模型性能的影响. 实验结果如图 8 所示. 显然,在迭代次数达到 200 次时,优化后的模型明显超越了传统的 ResNet-152 模型. 随着迭代次数的增加,两者之间的性能差距趋于稳定. 值得指出的是,传统的 ResNet 模型在应对图像数据集方面显示出相对有限的性能,整体表现并不令人满意. 因此,经综合考虑,引入格拉姆矩阵注意力机制的 ResNet-152 模型在准确度方面展现出卓越的分类效果.

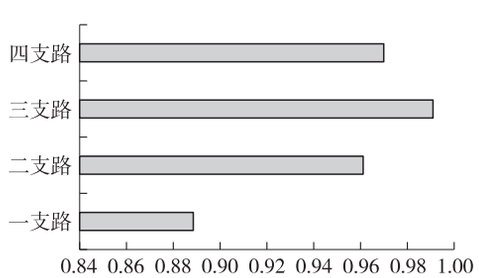


图 6　特征融合四种情况的准确度比较

Fig. 6　Accuracy comparison of four cases of feature fusion

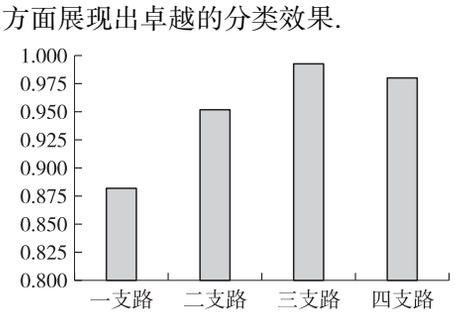


图 7　特征融合四种情况的 $F1$ 分数比较

Fig. 7　Comparison of $F1$ scores in four cases of feature fusion

在鸟类数据集上,比较了 ResNet-152 基础模型和经过注意力机制优化后的网络模型 All-ResNet-152 的损失曲线,如图 9 所示. 查看图中曲线的走势,可以发现两个模型的初始损失值大致在 2 左右,并且都呈现出急剧减小的趋势. 随着迭代次数的增加,损失值逐渐接近于零,并在约 1 200 次迭代后趋于稳定. 值得注意的是,添加了注意力机制的网络模型始终保持较低的损失值,呈现出更为乐观的整体下降趋势.

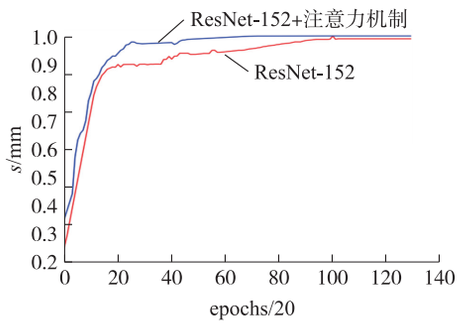


图 8　添加注意力机制准确率对比图

Fig. 8　Comparison chart of accuracy of adding attention mechanism

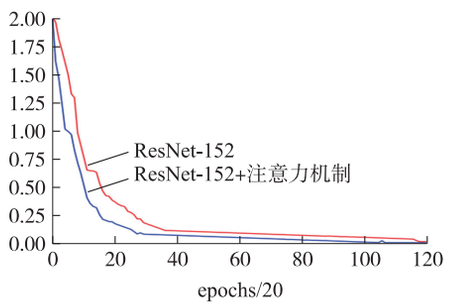


图 9　添加注意力机制损失曲线对比图

Fig. 9　Comparison diagram of loss curve of adding attention mechanism

在 CUB-200-2011 鸟类公共数据集上,进行了模型对比实验,包括 GCNet、CaffeNet 和提出的 All-ResNet-152 模型. 通过进行 500 步的迭代训练,对这些模型的分类时间进行了比较,结果如表 3 所示. 从结果可以明显看出,All-ResNet-152 模型在评价分类时间上表现最优,同时成功将图像准确分类到其对应类别. 这表明所引入的优化能够有效提取图像的分类特征,使得分类时间显著变快.

表 3　不同模型的分类时间对比

Table 3　Comparison of the classification time of different models

模型	分类时间/min
All-ResNet-152	0.98
GCNet[25]	1.02
CaffeNet[26]	1.20

在三种经典分类数据集(CUB-200-2011、Stanford-cars 和 Stanford-dogs)上与多个不同图像分类算法进行比较的表现如表 4 所示. 结果显示,新提出的 All-ResNet-152 模型在图像分类性能方面取得了显著的改进. 相较于最新的细粒度图像分类算法 CATNet,在实验结果上,提出的 All-ResNet-152 模型表现更为出色. 这表明,All-ResNet-152 模型所引入的优化能够有效提取图像的分类特征,显著增强模型的鲁棒性,使得分类结果更为精确.

表 4　多个图像分类算法的对比

Table 4　Comparison of multiple image classification algorithms

模型	准确度/%		
	CUB-200-2011	Stanford-cars	Stanford-dogs
ResNet-50	76.3	87.4	73.8
MAMC[27]	86.1	92.7	84.7
MACNN[28]	88.4	92.1	84.6
CATNet[29]	90.7	94.2	90.4
All-ResNet-152	95.6	97.3	94.5

4　结论

本研究创新性地将经典的卷积神经网络模型与新方法相结合,提出了一种基于优化后的 ResNet-152 的算法,为解决研究问题提供了重要参考和借鉴. 优化主要集中在对 ResNet-152 网络模型的改进,包括在输入部分引入特征融合和构建新型注意力机制. 通过在公共图像数据集上的一系列实验,证明该方法相较于传统方法在图像分类任务中具有更高的性能. 实验结果显示,经过优化后的模型在分类准确度方面取得了显著提升,从最初的 96.68% 提高到 98.87%. 期望该方法在未来的研究和应用中得到进一步验证和推广.

[参考文献]

[1] 李少凡,高尚兵,张莹莹,等. 基于轻量化网络与嵌入式的分心行为协同检测系统[J]. 南京师范大学学报(工程技术版),2023,23(1):25-32.

[2] WANG Y,XIE Y,ZENG J,et al. Cross-modal fusion for multi-label image classification with attention mechanism[J]. Computers and electrical engineering,2022,101:108002.

[3] 姜昊,凌萍,陈寸生保. 一种新的基于通道-空间融合注意力及 SwinT 的细粒度图像分类算法[J]. 南京师范大学学报

(工程技术版),2023,23(3):36-42.
[4] NAM E J,NAM C H,JANG K S. Face detection method based fusion RetinaNet using RGB-D image[J]. Journal of the Korea Institute of Information and Communication Engineering,2022,26(4):519-525.
[5] PANDA M K,SUBUDHI B N,VEERAKUMAR T,et al. Modified ResNet-152 network with hybrid pyramidal pooling for local change detection[J]. IEEE transactions on artificial intelligence,2023,5(4):1599-1612.
[6] SUN T,DING S,GUO L. Low-degree term first in ResNet,its variants and the whole neural network family[J]. Neural networks,2022,148:155-165.
[7] TABRIZCHI H,PARVIZPOUR S,RAZMARA J. An improved VGG model for skin cancer detection[J]. Neural processing letters,2023,55(4):3715-3732.
[8] 张婷婷,王斌,王坤,等. 基于增强层次对称点图像分析和深度残差网络的水电机组故障诊断[J]. 水利学报,2023,54(11):1380-1391.
[9] ULLAH A,ELAHI H,SUN Z,et al. Comparative analysis of AlexNet,ResNet18 and SqueezeNet with diverse modification and arduous implementation[J]. Arabian journal for science and engineering,2022,47(2):2397-2417.
[10] KURANI A,DOSHI P,VAKHARIA A,et al. A comprehensive comparative study of artificial neural network(ANN) and support vector machines(SVM) on stock forecasting[J]. Annals of DATA SCIEnce,2023,10(1):183-208.
[11] WANG Y,PAN Z,DONG J. A new two-layer nearest neighbor selection method for kNN classifier[J]. Knowledge-based systems,2022,235:107604.
[12] MOHAMMADPOUR L,LING T C,LIEW C S,et al. A survey of CNN-based network intrusion detection[J]. Applied sciences,2022,12(16):8162.
[13] LIN K,ZHOU T,GAO X,et al. Deep convolutional neural networks for construction and demolition waste classification:VGGNet structures,cyclical learning rate,and knowledge transfer[J]. Journal of environmental management,2022,318:115501.
[14] YANG L,YU X,ZHANG S,et al. GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases[J]. Computers and electronics in agriculture,2023,204:107543.
[15] STEFENON S F,YOW K C,NIED A,et al. Classification of distribution power grid structures using inception v3 deep neural network[J]. Electrical Engineering,2022,104(6):4557-4569.
[16] 陈斌,樊飞燕,陆天易. 残差混合注意力结合骨骼图卷积多人姿态识别[J]. 南京师大学报(自然科学版),2024,47(4):106-117.
[17] 张德,甄昊宇,林青宇. 结合注意力机制的生成对抗网络图像超分辨重建[J]. 福建师范大学学报(自然科学版),2023,39(3):86-93.
[18] 周湘贞,李帅,隋栋. 基于深度学习和注意力机制的微博情感分析[J]. 南京师大学报(自然科学版),2023,46(2):115-121.
[19] 王虎,尹泽泉,王雯婕,等. 基于金字塔池化网络的质子交换膜燃料电池气体扩散层组分推理方法[J]. 重庆大学学报,2024,47(1):84-92.
[20] ZHANG Q,XIAO J,TIAN C,et al. A robust deformed convolutional neural network(CNN) for image denoising[J]. CAAI transactions on intelligence technology,2023,8(2):331-342.
[21] NESHAT M,AHMED M,ASKARI H,et al. Hybrid inception architecture with residual connection:fine-tuned Inception-ResNet deep learning model for lung inflammation diagnosis from chest radiographs[J]. Procedia computer science,2024,235:1841-1850.
[22] LUO Y,XU Y,WANG C,et al. ResNeXt-CC:a novel network based on cross-layer deep-feature fusion for white blood cell classification[J]. Scientific reports,2024,14(1):18439-18451.
[23] 欧阳顺湘. 从格拉姆矩阵不等式到多维协方差不等式[J]. 高等数学研究,2023,26(1):38-40.
[24] CHEN H,PEI Y,ZHAO H,et al. Super-resolution guided knowledge distillation for low-resolution image classification[J]. Pattern recognition letters,2022,155:62-68.
[25] 李雨泽,李心安. 基于 GCNet 的毫米波波束选择算法[J]. 光通信研究,2023(6):72-76.
[26] 张广群,李英杰,汪杭军,等. 基于微调 CaffeNet 的林业图像分类[J]. 林业科学,2020,56(10):121-128.
[27] WANG J,HU C,XU H,et al. A novel multi-atlas and multi-channel(MAMC) approach for multiple sclerosis lesion segmentation in brain MRI[J]. Signal,image and video processing,2019,13:1019-1027.
[28] MUDRAKOLA S,NAGARATNA H. Multi-attention convolution neural network to detect breast cancer[J]. Computer integrated manufacturing systems,2024,30(6):75-99.
[29] ZHANG W,CHEN G,ZHUANG P,et al. CATNet:cascaded attention transformer network for marine species image classification[J]. Expert systems with applications,2024,256:124932.

[责任编辑:杜忆忱]