

摘要

随着图像处理技术的发展，分化出如图像识别、分割、分类等分支，因其能够提供真实可用的特征信息而成为研究热点。在研究图像过程中，人们往往只对突出的前景信息感兴趣，不需要过多关注背景信息。为了更高效地利用图像信息，本文从图像分割与分类角度出发，提出基于双分支融合注意力机制的图像分割算法对图像中前景物体进行分割，以及基于残差网络融合多层感知机制图像分类算法对分割出的前景物体进行分类，主要研究工作和成果如下：

(1) 针对 DeepLabV3+ 中存在多个物体影响目标物体的重要程度，从而导致目标物体细节信息损失分割不细致的问题，本文提出一种基于双分支融合注意力机制的图像分割算法。首先引入多维注意力机制捕获图像细节信息，其次设计双分支特征提取机制扩大特征提取能力以细化图像边缘信息，最后联合采用两种损失函数聚焦目标物体弱化背景特征，提高算法分割精度。在 PASCAL VOC 2012 和 Cityscapes 数据集上进行实验，本文算法的 MIoU 值相较于 DeepLabV3+ 分别提高了 1.14% 和 0.65%，MPA 值分别提高了 0.73% 和 1.51%。

(2) 针对 EfficientNet 中深度可分离卷积执行缓慢以及输入高分辨率图像占用大量内存从而影响分类效果的问题，本文提出一种基于残差网络融合多层感知机制的图像分类算法。首先将前期阶段中的深度可分离卷积替换为常规卷积，以少量参数量的开销提高分类精度；其次控制输入图像分辨率，并对 SE 通道注意模块进行优化，增强对图像通道特征的提取；最后设计多层感知机制和残差网络标记、投射卷积特征，进一步提高算法分类准确度。在 CIFAR-100 和城市街景数据集上进行实验，本文算法的 TOP-1 准确率相较于 EfficientNet 分别提高了 0.54% 和 0.69%。

(3) 基于本文所提出的两种算法，开发并完成了交互式图像分割与分类系统，该系统不但能够通过对图像进行分割获取目标物体信息，还能够对目标物体进行定位分类，为后续图像处理技术奠定了合理与精确的依据，满足人们对更精确图像分割与分类的要求，也实现了文中所提方法的实际应用意义。

关键词：图像分割，图像分类，残差网络，注意力机制，多层感知机制

目 录

摘 要.....	I
ABSTRACT.....	III
1 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 图像分割算法.....	2
1.2.2 图像分类算法.....	4
1.3 本文贡献.....	5
1.4 本文组织架构.....	6
1.5 本章小结.....	7
2 图像分割与分类概述及相关理论.....	9
2.1 图像分割与分类概述.....	9
2.2 卷积神经网络 CNN.....	10
2.3 注意力机制.....	12
2.4 多层感知机制.....	15
2.4.1 多层感知机制运算过程.....	15
2.4.2 常用的激活函数.....	17
2.5 本章小结.....	17
3 基于双分支融合注意力机制的图像分割算法.....	19
3.1 DeepLabV3+图像分割算法概述.....	19
3.2 基于 DeepLabV3+模型的改进算法.....	21
3.2.1 优化多维注意力机制.....	21
3.2.2 双分支特征提取机制.....	22
3.2.3 优化损失函数.....	23
3.2.4 算法描述.....	24
3.3 实验分析.....	25
3.3.1 实验配置及评价指标.....	25

3.3.2 PASCAL VOC 2012 数据集.....	26
3.3.3 Cityscapes 数据集.....	29
3.3.4 消融实验.....	30
3.4 本章小结.....	31
4 基于残差网络融合多层感知机制的图像分类算法.....	33
4.1 EfficientNet 分类算法概述.....	33
4.2 基于 EfficientNet 模型的改进算法.....	35
4.2.1 优化 MBConv 结构.....	36
4.2.2 优化通道注意模块.....	37
4.2.3 改进滑动多层感知机制.....	38
4.2.4 算法描述.....	39
4.3 实验分析.....	40
4.3.1 实验配置及评价指标.....	40
4.3.2 CIFAR-100 数据集.....	41
4.3.3 城市街景数据集.....	43
4.3.4 消融实验.....	45
4.4 本章小结.....	46
5 图像分割与分类系统的设计与实现.....	47
5.1 开发环境.....	47
5.2 应用场景.....	47
5.3 需求分析.....	47
5.3.1 功能需求.....	48
5.3.2 非功能需求.....	48
5.4 系统的设计与实现.....	49
5.4.1 系统设计.....	49
5.4.2 系统实现.....	50
5.5 本章小结.....	53
6 总结与展望.....	55
6.1 总结.....	55
6.2 展望.....	55

参考文献.....	57
致 谢.....	63
攻读学位期间的科研成果.....	65

1 绪论

图像分割与分类作为高效的图像处理技术，在计算机视觉、模式识别等方面有着广泛的应用前景，进而成为备受关注的重要领域，也吸引了众多学者深入研究。本章从背景意义出发，研究国内外分割与分类算法模型并简述了本文主要贡献和各章节安排。

1.1 研究背景与意义

图像分割与分类^[1]问题在近二十年中受到广泛的关注和发展，国内外学者提出的方法在各应用领域也取得了一定的成果。分割与分类是对图像进行的初步处理，接下来如目标识别、检测等任务的质量，都依赖于图像分割与分类的效果如何。

图像的分割和分类被广泛应用于许多领域，包括安防中的人脸识别、智能视频分析，交通场景下的物体识别、车辆检测等。日常生活中随处可见图像分割与分类的运用，该技术代替了传统人工操作，对人类的生活方式产生了巨大的影响。

图像处理技术要解决的基本问题可归结为图像中存在什么物体，该物体处于什么位置。如图 1-1(a)所示，对于一张图片，图像分类要解决的问题是，在这张图片中主要存在哪些物体，比如公交车；图像分割要解决的问题是该物体在图像中的位置，并给出物体的轮廓信息，如图 1-1(b)所示。要实现真实复杂场景的自动分析和理解，必须先判断出哪些目标物体位于图像的哪个位置，即分割和分类问题。由于图像分割与分类技术的重要性，对图像的精确分割与分类方法进行深入的研究具有重大的理论与现实意义。



图 1-1 图像分割与分类

与此同时，随着智能系统研究进程的不断深入，像是无人驾驶、城市无人机等智能设备出现在人们的生活中，图像处理技术理所当然地成为此类应用的重要基础，而对图像进行分割和分类则直接关系到这些智能设备的安全性。与目标检测相比，图像分割与分类的结果更加精确，可以具体地分割出人们所需要的物体，比如在自动驾驶领域中，传感器采集到道路场景之后，可以利用图像分割与分类技术，将行人、车辆等目标物体轮廓分割出来，从而引导车辆避让目标，确保安全。图像中所蕴含的信息是海量的，如何从这些图像中提取有价值的信息尤为重要，因此，需要对图像分割和分类进行深入研究。

1.2 国内外研究现状

1.2.1 图像分割算法

图像分割，是根据特定准则将一张图像划分为不同区域的过程，即将前景信息和背景信息分离开来，并在此基础上进行后续的图像处理工作。

(1) 基于阈值的图像分割方法

基于阈值的图像分割方法利用目标之间不同的特征差别，设定不同的灰度阈值，对图像灰度直方图分类划分出前景物体和背景。Liu 等人^[2]基于最大类间差算法对图像进行处理，通过不断的调整阈值来加大像素点间的差距，以此来获取更明显的分割区域。基于阈值的分割方法比较适合于目标和背景有较大差别的图像，该类图像的像素灰度值差异较大。然而这种方法对于复杂的图像，分割效果不佳，对图像的语义、空间特征提取不充分。

(2) 基于边缘的图像分割方法

基于边缘的图像分割方法利用像素点差异，检测到灰度值变化大的点认定为区域边界点，之后连接起来得到几个区域以达到分割目的。常见的边缘检测方法有 Roberts^[3]、Sobel^[4]、PreWitt^[5]、Canny^[6]等。Wang 等人^[7]先利用图像灰度化处理得到差别较大的目标灰度值和背景灰度值，基于此再使用小波多尺度边缘检测法能很容易划分出目标区域并完成分割。基于边缘检测法的分割方法实现原理简单，分割效果也比较优良，但其无法平衡抗噪性和分割精度，也不能得到良好的区域结构。

(3) 基于区域的图像分割方法

基于区域的分割方法主要包括区域生长和分裂合并两种方式。前者会先选取一个像素点，持续合并相邻的类似像素点，以此标示出一个分割区域，从而对整个图像进行分割。后者基于整个图像，出现不同特征时分裂成相等的小区域，直到无法分裂时再检测相邻区域，满足一定条件则合并。Zhang^[8]将色彩融入到区域合并中，设计自适应权重因子完成分级合并分割方法，在 RGB 图像分割中取得了良好的分割效果。相比阈值方法，区域方法在相似图像中更具有优势，拥有更好的鲁棒性，排除了图像中噪点干扰，但也容易引起过度分割的问题。

(4) 基于聚类的图像分割方法

基于聚类的图像分割方法聚集有相同特征的像素点，通过反复迭代的方式对输出结果进行收敛，直至将图像划分成不同区域，而此时相同特征像素点已被聚集到同一区域块中。为了降低噪声对传统模糊聚类算法的影响，Cai 等人^[9]提出了一种核空间模糊聚类分割算法，对分类像素点进行模糊后，使用数学计算方式对目标进行模糊聚类，从而获得该方法的迭代表达式。

(5) 基于卷积神经网络的图像分割方法

近年来，卷积神经网络 (Convolutional Neural Networks, CNN) 成为图像处理应用最广泛的手段。为了应对多样的处理需求，一系列基于 CNN 的图像分割算法被广大科研学者提出。

FCN^[10]将传统的全连接操作替换为卷积操作，利用卷积和反卷积处理特征图。SegNet^[11]由 Badrinarayanan 等人提出，采用对称编解码结构，编码器以最大池化的方式降低输入图像维度，解码器保存最大池化操作的最大值位置。PSPNet^[12]金字塔模块融合了 4 种尺度特征，并且分别进行平均池化操作，卷积对高维特征图进行降维，同时对低维特征图上采样。Mask R-CNN^[17]是 He 等提出的一种在进行目标检测的同时实现高质量分割的算法模型，通过在特征图中获取多个 ROI 候选框并进行 RoIAlign 处理提高分割精度。

Oh Y 等人^[18]提出一种新型池化方法 BAP 并融入到图像分割算法，利用边界框区分出背景和前景，用前景的特征进行分割并丢弃背景特征。Lee 等人^[19]提出边界框属性映射 BBAM 结构，设置两条输出路径，一条用来图像分类，另一条利用目标检测器在边界框内定位目标对象进行分割。Zhu 等人^[22]研究了图像的低级纹理特征，设计一种新型

描述纹理信息的方式 QCO，并在此基础上引入纹理信息的增强模块 TEM 和纹理特征提取模块 PTFEM。

1.2.2 图像分类算法

图像分类是通过分析图像中所包含的特征信息，将不同种类的物体加以区分的一种图像处理技术。

(1) 基于色彩特征的分类方法

基于色彩特征的分类方法主要以物体所具有的色彩特征为依据，通常人们认为同一类物体具有相似的色彩特征，从而对物体进行区别实现分类。Swain 等人^[29]提出的色彩直方图分类方法，根据其相似性来度量分类，但也损失了部分色彩空间信息。Pass 等人^[30]提出一种新的分类标准色彩聚合矢量 (CCV)，主要通过对目标和分类图像的 CCV 进行匹配实现，该方法保留了部分色彩空间信息。Smith 等人^[31]采用色彩分割方法划分色彩集，在图像分类匹配过程中比较图像色彩集和色彩区域信息。近年来人们从不同角度研究了色彩分类方法，但由于色彩空间分为 RGB、HIS、HSV 等，在其选择上无法统一，也影响了图像分类的效果。

(2) 基于纹理的图像分类方法

基于纹理的图像分类方法是根据图像中各像素的领域灰度不同进行分类。Haralick 等人^[32]提出了以纹理特征为基础的灰度共生矩阵表示法，可以提取出灰度空间相关性。Ma 等人^[33]基于研究小波变换种类，获得最符合人类视觉的 Gbaor 小波变换的结果。Matthews 等人^[34]使用相对属性对同一种纹理的相似度大小进行比较并标记，再根据相对属性，学习出纹理属性的绝对相似度度量。Cimpoi 等人^[35]定义了一组具有普适性的纹理属性集，能够描述大部分的纹理特征，并鲁棒地从纹理图像中进行估算。

(3) 基于形状的图像分类方法

对于图像中的物体来说，形状也是重要的区分特征，它通常由一条闭合的曲线包围区域来描述，基于形状的图像分类即基于物体轮廓特征进行分类。Jain 等人^[36]用封闭的直线描述形状轮廓，按角度斜率的差别来区分形状。Gunsel 等人^[37]提出基于特征形状的图像分类方法，其基本思路是通过图形中两点间高斯加权距离所构成的相邻矩阵的特征进行求解，再归一化获得形状特征。庄越挺等人^[38]提出了一种形状内角直方图的概念，用来计算邻接边缘的内角。韦东兴等人^[39]采用轮廓追踪与角点检测算法来获取一张图像的轮廓与角点，并以时间序列的形式来描述质心距离作为该图像的一个特征。

(4) 基于空间关系的图像分类方法

在图像中,人们认为物体与物体间的空间位置关系也是用来分类的一个重要评价标准,根据物体空间位置关系进行区分,更容易被人们所接受。Tanimoto 等人^[40]提出了用具有空间和波谱特征的数据元方法来表示图像中的物体,并用该数据元作为索引。Chang 等人^[41]提出用 2D-String 表示方法,实现了对图像的空间关系的归类和对某些图像符号图的重建。Cilallg 和 Jungert 等人^[42]又提出了 2DG-String 的方法,将目标分割成若干个较小的子目标,以反映目标之间的空间关系。

(5) 基于神经网络的图像分类方法

以上四种方法都是以图像特征为基础对图像进行分类,随着卷积神经网络(CNN)概念的深入,LeNet-5^[43]出现在人们视野里,主要用于手写数字识别,作为 CNN 的开山之作,LeNet-5 的出现也极大地推动了 CNN 发展。AlexNet^[44]在 LeNet 的基础上对深度改进,增加卷积层数量提取更丰富的特征,并加入 Dropout 层,抑制过拟合。VGG^[45]与 AlexNet 相比,采用了更深的网络结构,经过研究表明,增大网络深度可以对网络的性能产生一定的影响,而将大卷积核替换为小卷积核也可以显著降低参数量。但通过加深网络提高性能,也会带来梯度消失、梯度爆炸等问题。GoogLeNet^[46]从另一种角度来提升训练结果,引入 Inception 结构对不同尺度的特征信息进行融合。ResNet^[47]先对数据进行预处理,采用 BN 层解决梯度消失,并增加残差结构来缓解退化问题。MobileNet^[48]是一种嵌入深度可分离卷积降低网络参数的轻量级网络。

通常对于卷积神经网络的改进,主要是从网络深度、宽度和分辨率三个方面入手。EfficientNet^{[49][50]}很好地平衡了深度、宽度和分辨率这三个维度以提高分类精度。ViT^[51]提出了一种应用 Transformer 的分类算法,其性能在经过大量的预训练后,甚至超越了 CNN。PVT^[52]在 Transformer 中引入金字塔结构,解决了 ViT 显存可能溢出的问题。

1.3 本文贡献

本文首先对图像分割与分类概念和相关理论进行介绍,提出了两种改进图像分割与分类算法,并用大量的实验数据对本文提出的两种算法进行验证,结果表明本文提出的两种算法比其他模型在图像分割与分类任务中的表现效果更优,主要内容如下:

(1) 针对图像分割问题,本文基于 DeepLabV3+提出一种融合双分支特征提取和注意力机制的改进算法 ImDeepLabV3+。在深度卷积结构中加入注意力机制捕获被

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/185022332201012010>