

分类号： TP391  
论文编号： 2020022284

密 级： 公开

贵 州 大 学  
2023 届硕士研究生学位论文

# 基于深度学习的蝴蝶种类识别研究

学科专业： 电子信息

研究方向： 图像处理

导 师： 谢本亮 副教授

研 究 生： 李飞

中国 · 贵州 · 贵阳

2023 年 5 月

# 硕士学位论文答辩委员会名单

## 基于深度学习的蝴蝶种类识别研究

答辩人： 李飞

答辩委员会委员：

贵州大学：\_\_\_\_\_ 马奎 教授 \_\_\_\_\_（主席）

贵州师范大学：\_\_\_\_\_ 雷涛 副教授 \_\_\_\_\_

贵州大学：\_\_\_\_\_ 郭祥 副教授 \_\_\_\_\_

答辩秘书：张正丽 博士

答辩时间：2023 年 05 月 20 日

答辩地点：博学楼 626

# 基于深度学习的蝴蝶种类识别研究

## 摘要

蝴蝶种类丰富,在维持生态系统稳定和保护生物多样性等方面发挥着重要作用。同时,它们翅膀上的图案颜色丰富多彩、形状各异,观赏价值颇高。但是,相差较远的蝴蝶种类间的颜色、纹理和形状复杂多样,而同一种特别是同一属下的蝴蝶在形状、颜色和纹理方面却具有较大的相似性。此外,部分蝴蝶还具有拟态性。上述因素的存在,严重影响了自然环境中蝴蝶种类识别效果。因此,从蝴蝶本身特征出发,探讨出高效准确的识别方法,是目前急需解决的问题。早期对蝴蝶种类识别总是依赖昆虫学家,但这种方法耗时耗力。随着计算机技术的发展,使用深度学习技术来解决图像处理任务已经取得显著效果。基于现有研究中出现的问题,本文将深度学习技术应用到蝴蝶种类识别中,分别从蝴蝶图像的分割和蝴蝶图像的识别两方面进行研究,通过各主流网络间的对比选出适合蝴蝶种类识别的基准模型,并结合蝴蝶图像特征对基准模型进行了改进和优化。本文主要完成的工作如下所示:

1. 自然环境下的蝴蝶图像数据集制作。针对现有研究中自然环境下蝴蝶图像数据集的数量和种类较少的情况,本文采用了多种数据获取方式,包括实地拍摄蝴蝶图像、从 Kaggle 官方数据网站下载和编程 Python 爬虫代码的方式等。通过将当地蝴蝶种类和官方数据集中蝴蝶种类相结合,构建了自然环境状态下的 32 种共计 7825 张蝴蝶图像数据集。为了增强模型的泛化能力,本文采用高斯噪声、椒盐噪声、左右翻转、增强亮度等数据增强方式对蝴蝶图像数据集进行数据扩充。

2. 针对自然环境中蝴蝶种类识别率低等问题,本文以上述 32 种自然环境下的蝴蝶图像数据集为研究对象,提出了一种改进的 DeepLabv3+蝴蝶图像分割模型,以去除蝴蝶图像中的复杂背景,改进后的模型融合了置换注意力机制与密集空洞空间金字塔池化。通过使用 EfficientNetV2-S 网络替换原始 DeepLabv3+网络中的主干特征提取网络 Xception,提高了主干网络的特征提取能力。虽然原始的空洞空间金字塔池化网络已经在一定程度上增大了网络的感受野,但是针对大型蝴蝶,还需要进一步增加网络感受野才能捕捉到较为完整的蝴蝶图像特征,同时原始空洞空间金字塔池化网络的信息利用率较低的问题也需要解决,因而将密集

级联的思想引入到了空洞空间金字塔池化网络中,这不仅可以增大网络感受野还能增加采样点的密集性。使用置换注意力网络对蝴蝶的高维语义信息做进一步提取,以获得更加丰富的蝴蝶图像信息。实验结果表明,改进后的蝴蝶图像分割网络能够有效分割自然环境下的蝴蝶图像,平均分割精度达到了 96.19%,同时模型的像素准确率达到了 98.60%。与原始 DeepLabv3+网络相比,改进后的网络分割能力更强,可以更有效地分割自然环境下的蝴蝶图像与背景。

3. 针对蝴蝶图像识别效率低的问题,本文先将蝴蝶与其复杂背景进行分割,得到了去除背景噪声的蝴蝶图像,再以这些图像为研究对象,提出了一种基于残差网络的轻量级可变形蝴蝶种类识别模型。该模型以 ResNet-18 网络为基础,从蝴蝶种类数量分布情况、形状大小多样和模型轻量化等方面入手,对基础模型进行了改进。针对蝴蝶种类中存在的数量分布不均衡现象,引入 PolyLoss-Focal Loss 损失函数进行训练,有效解决了蝴蝶因种类分布不均带来的识别效果较差的问题;此外,为了应对蝴蝶的大小不一、形状多样的特点,引入可变形卷积替换普通卷积,使网络的感受野增大,针对较大的蝴蝶,网络能够捕捉到它们比较完整的信息,以此来提高网络的识别率。从模型参数量方面考虑,借鉴深度可分离卷积的思想,对原始的深度可分离卷积进行了改进,改进后的模型能更大程度的减小模型参数量与运算量。实验结果表明,改进后的蝴蝶图像识别模型取得了更好的识别效果,模型的平均识别准确率达到 96.23%,而模型参数量仅 1.04MB。该模型能够有效识别 32 类蝴蝶图像。

**关键词:** 深度学习; 蝴蝶种类识别; 图像分割; 注意力机制;  
深度可分离卷积

# Research on butterfly species identification based on deep learning

## ABSTRACT

Butterflies are diverse in species and play an important role in maintaining ecosystem stability and preserving biodiversity. Additionally, their wings display a wide range of colorful patterns and unique shapes, making them highly valuable for observation. However, the colors, patterns, and shapes of butterfly species that are distantly related can be complex and diverse, while those of the same species, especially within the same genus, tend to exhibit greater similarity in shape, color, and pattern. Furthermore, some butterflies possess mimicry. These factors have a significant impact on the identification of butterfly species in the wild, making it imperative to investigate reliable identification methods based on butterfly traits. While entomologists were essential for early species recognition, this process required a lot of time and effort. With the advancement of computer technology, deep learning methods have been used to solve image processing problems, and the outcomes have improved. This paper applies deep learning techniques to butterfly species recognition based on current research issues. Separate studies on butterfly picture segmentation and butterfly image recognition were conducted. By comparing mainstream networks, the benchmark model appropriate for butterfly species identification was selected. Then, the benchmark model was enhanced and optimized by combining butterfly image features. The main work accomplished in this paper is as follows:

1. Butterfly image dataset production in natural environment. Due to the limited quantity and variety of butterfly image datasets in natural environments in existing research. This paper utilized field photography techniques to capture butterfly images, downloaded images from the main Kaggle data website, and utilized web crawler programming to collect butterfly images from the internet. By combining local butterfly species with those from official databases, a total of 7,825 butterfly images of 32 species in their natural habitats were created. The butterfly image dataset can be improved with

gaussian noise, salt and pepper noise, left-right flip and enhanced luminance to increase the generalizability of the model.

2. Aiming at the problems such as low recognition rate of butterfly species in natural environments, this paper utilizes the 32 butterfly image datasets in natural environments mentioned above as the research object. A DeepLabv3+ butterfly image segmentation model incorporating the shuffle attention mechanism and dense atrous spatial pyramidal pooling is proposed to remove complex backgrounds from butterfly images. To improve segmentation accuracy of the model, the Xception backbone feature extraction network in the DeepLabv3+ network is replaced with the EfficientNetV2-S network. Although the original atrous spatial pyramidal pooling network has increased the network's field of perception to some extent, it is necessary to further increase the field of perception for large butterflies to capture more complete butterfly image features. Additionally, the low information utilization problem of the original atrous spatial pyramidal pooling network must also be addressed, and the idea of dense cascade is introduced into the atrous spatial pyramidal pooling network. To extract more detailed information about butterfly images, the shuffle attention network is used to extract high-dimensional semantic information. The experimental results demonstrate that the enhanced butterfly image segmentation network can successfully segment butterfly images in natural environments. Compared to the DeepLabv3+ network, the mean intersection over union of the improved model reached 96.19%. Meanwhile the pixel accuracy reached 98.60%. The method presented in this article can successfully separate butterflies from backgrounds in natural environments.

3. To address the problem of low efficiency in butterfly image recognition, this paper first segmented butterflies with their complex backgrounds to obtain butterfly images with background noise removed. Based on these images, a lightweight deformable butterfly species recognition model was proposed using residual networks. The model is based on the ResNet-18 network, which was improved by accounting for the distribution of butterfly species, their variety in sizes, shapes and their lightness. To address the uneven distribution of butterfly species, the PolyLoss-Focal Loss function

is introduced for training, which effectively improves butterfly recognition. Deformable convolution was used instead of regular convolution to accommodate butterflies of various sizes and shapes. The network's sensory field was expanded to fully perceive larger butterflies with larger sensory fields, resulting in increased recognition rates. Taking inspiration from the depthwise separable convolution, an enhanced version has been proposed to significantly reduce the number of model parameters and operations. The experimental results show that the improved model achieves superior performance in butterfly recognition, with an average recognition accuracy of 96.23%, and the number of model parameters of just 1.04MB. It is capable of accurately identifying 32 distinct types of butterfly images.

**Key words:** Deep learning, Butterfly species identification, Image segmentation, Attentional mechanisms, Depthwise separable convolution.

## 目 录

第一章 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	3
1.2.1 深度学习研究现状.....	3
1.2.2 图像分割研究现状.....	5
1.2.3 蝴蝶种类识别的国内外研究现状.....	6
1.3 主要研究内容.....	8
1.4 论文章节安排.....	9
第二章 深度学习的相关理论基础 .....	11
2.1 卷积神经网络.....	11
2.1.1 卷积层.....	11
2.1.2 池化层.....	12
2.1.3 全连接层.....	13
2.2 主流图像分割网络概述.....	13
2.2.1 SegNet 网络 .....	13
2.2.2 PSPNet 网络 .....	14
2.2.3 HRNet 网络 .....	15
2.2.4 DeepLabv3+网络 .....	16
2.3 主流图像分类网络概述.....	17
2.3.1 VGG 网络 .....	17
2.3.2 GoogLeNet 网络 .....	18
2.3.3 ResNet 网络 .....	18
2.3.4 MobileNet 网络 .....	20
2.3.5 ShuffleNet-V2 网络 .....	22
2.3.6 EfficientNet 网络 .....	23
2.4 本章小结.....	23
第三章 基于改进的 DeepLabv3+蝴蝶图像分割模型 .....	25
3.1 蝴蝶图像数据集构建.....	25
3.1.1 蝴蝶图像数据集预处理.....	27
3.1.2 数据标注.....	28
3.1.3 数据增强.....	29
3.2 蝴蝶图像分割模型构建.....	30
3.2.1 蝴蝶图像分割模型总体结构.....	31
3.2.2 EfficientNetV2-S 主干网络 .....	32



3.2.3 改进空洞空间金字塔池化.....	33
3.2.4 引入置换注意力模块.....	35
3.3 实验结果与分析.....	37
3.3.1 实验环境.....	37
3.3.2 参数设置.....	37
3.3.3 评价指标.....	37
3.3.4 结果分析.....	38
3.4 本章小结.....	42
第四章 基于残差网络和可变形卷积的蝴蝶种类识别模型 .....	43
4.1 蝴蝶图像识别数据集构建 .....	43
4.1.1 数据集制作.....	43
4.1.2 数据集划分.....	44
4.1.3 数据增强.....	44
4.2 蝴蝶种类识别模型构建.....	44
4.2.1 蝴蝶种类识别总体模型结构.....	45
4.2.2 引入 PolyLoss-Focal Loss 损失函数.....	45
4.2.3 引入可变形卷积模块.....	46
4.2.4 改进深度可分离卷积.....	48
4.3 实验结果与分析.....	50
4.3.1 实验环境.....	50
4.3.2 参数设置.....	50
4.3.3 评价指标.....	50
4.3.4 结果分析.....	51
4.4 本章小结.....	56
第五章 总结与展望 .....	57
5.1 论文工作总结.....	57
5.2 未来展望.....	58
参考文献.....	60

# 第一章 绪论

## 1.1 研究背景与意义

昆虫物种繁多、形态多样，是世界上数量最多的动物。其中，人类已经发现的昆虫数量就达到了 150 多万种，它们的物种多样性在生态系统功能方面发挥了重要作用<sup>[1]</sup>。蝴蝶便是最常见的一种昆虫。

蝴蝶属于昆虫纲第二大目——鳞翅目<sup>[2]</sup>，因为它们的身体部位有鳞片<sup>[3]</sup>，故此得名。蝴蝶种类丰富，目前全世界已知的种类约有 18000 种<sup>[4]</sup>，其中，中国的蝴蝶种类约占全世界的九分之一<sup>[5]</sup>。蝴蝶翅膀上的鳞片色彩斑斓，飞行时翅膀展开非常美丽，具有很高的观赏价值<sup>[6]</sup>。同时，蝴蝶对其生存环境和气候的要求颇高，不同区域的温度和湿度不同，其种类分布也会不同，因此蝴蝶也被确定为环境和气候变化的指示性昆虫。而蝴蝶的这种特性，又导致了它们的迁徙，这也就促进了植物之间花粉的传播，尤其是远亲植物间的繁殖，蝴蝶这种帮助远亲植物授粉的行为<sup>[8]</sup>，促进了植物的多样性。由此可见，蝴蝶在生物多样性的保护和生态系统稳定性的维持中发挥了重要作用<sup>[9]</sup>。综上所述，对蝴蝶种类的自动识别研究在气候变化、环境检测和保护生物多样性等方面具有重要意义。

为了躲避天敌的捕杀，蝴蝶自身颜色会随着环境的变化而发生变化，呈现出与所处环境相类似的保护色，如图 1-1 所示，合上翅膀的枯叶蛱蝶完美地融入到枯叶中，成功地隐藏了自己。同时，不同种类的蝴蝶在颜色、形状和纹理等方面具有差异性，而相同种属间的蝴蝶在颜色、纹理或者形状等特征中又存在相似性，如图 1-2 所示，绿带燕凤蝶与燕凤蝶非常相似。此外，同一种类的蝴蝶雌性和雄性之间也存在着差异，如图 1-3 所示，玉带凤蝶的雌性和雄性具有明显的差异。不仅如此，蝴蝶图像还受到成像拍摄角度多样和拍摄亮度多变等因素的影响，使得同一种类蝴蝶图像具有复杂多样的呈现形式。综合分析可知，对自然环境中蝴蝶种类进行自动识别研究存在着一定的难度。



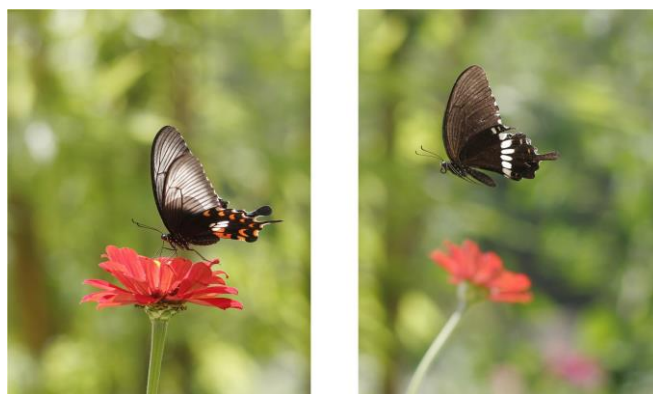
图 1-1 枯叶蛱蝶



(a) 绿带燕凤蝶

(b) 燕凤蝶

图 1-2 绿带燕凤蝶与燕凤蝶



(a) 雌性玉带凤蝶

(b) 雄性玉带凤蝶

图 1-3 雌性和雄性的玉带凤蝶

传统的蝴蝶种类鉴别主要依赖于昆虫学家，因为对蝴蝶种类进行鉴别时需要根据其翅膀正反面的纹理和颜色等特征进行识别，这不仅需要有足够经验的专业人员，还需要专业设备，这种鉴别方式会消耗巨大的人力和物力。如果能探索出一种新的、准确高效的鉴定方法来对蝴蝶种类进行自动识别，进而取代传统的人工鉴定，那将是非常有价值的。

随着科学技术的不断发展，为了提高蝴蝶种类识别效率，科研人员开始使用计算机技术对其种类识别进行研究。在早期的研究中，主要使用蝴蝶的标本图像进行识别，这些图像基本都是昆虫学家和蝴蝶爱好者收集的标本，涉及到的样本

数量和种类较少。同时,这种图像背景简单,图像中蝴蝶的翅膀和纹理信息清晰明了<sup>[10]</sup>。使用传统机器学习方法对蝴蝶种类自动识别的步骤为:蝴蝶图像采集与预处理、特征提取和优化以及设计合适的分类器<sup>[11]</sup>。其中,特征提取和优化用于提取蝴蝶的颜色、纹理和形状等特征。

近年来,深度学习技术不断发展,使用该技术解决昆虫种类识别问题已经取得了较好效果<sup>[12]</sup>。不仅实现了种类地自动识别,还提高了识别效率,减少了人力和物力的消耗。然而,蝴蝶种类繁多,其所处的自然环境十分复杂;同时,蝴蝶本身颜色、纹理和形状等特征信息多样;除此之外,现有自然环境下的蝴蝶图像数据量较少。所以,对蝴蝶种类进行自动识别研究仍然是一个难题。因此,本文对蝴蝶种类自动识别进行研究具有非常重要的意义。

## 1.2 国内外研究现状

### 1.2.1 深度学习研究现状

深度学习是机器学习的一个分支,随着该技术的迅猛发展,其在医疗健康诊断、工业生产自动化、自动驾驶、物体检测和自然语言处理等方面取得了巨大成果<sup>[13]</sup>。深度学习技术是以数据样本为依据来学习数据之间的内在规律,使用深度学习技术解决实际问题的最终目的是让计算机能够像人的大脑一样,具备学习和分析问题的能力,从而解决很多复杂场景下的难题,例如无人驾驶。目前,深度学习技术的飞速发展使得人工智能技术取得了巨大进步,为人类社会的发展带来了许多新的机遇和挑战。

1943 年,Warren McCulloch 和 Walter Pitts<sup>[14]</sup>通过用计算机模拟生物神经细胞单元结构和工作原理的过程,提出了神经网络这个概念,同时构建了人工神经元模型,这是深度学习技术的开篇。1958 年,Rosenblatt 等人<sup>[15]</sup>提出了感知机模型(Perceptron),这是首次把神经网络用到了实践上。然而,在接下来的一段时间里,人工神经网络都处于低谷期;但随着计算机的快速发展,到了 1986 年,Rumelhart 和 Hinton 等人<sup>[16]</sup>提出了反向传播网络(Back Propagation Network),解决了单层感知器无法解决的问题,但是该网络仅限于几个神经网络层。随着网络层数的增加和计算成本的限制,神经网络变得很难训练,反向传播算法出现了梯度弥散问题。1998 年,LeCun 等人<sup>[17]</sup>提出了基于梯度学习方法来训练卷积神经

网络的模型，并在 MNIST 数据集上进行实验，成功地对字母表中的手写字母进行了准确识别，但是对于较大数据集图像的解决效果并不友好。直到 2006 年 Hinton 等人<sup>[18]</sup>提出了深度置信网络(Deep Belief Network)，这是一种新的解决方案，即通过预训练来微调参数，并用标记数据的监督学习对网络进行微调，深度置信网络的这种特点有效地训练了深度神经网络，解决了深层网络优化的难题。该工作代表着神经网络迎来了一个新的时代，此时神经网络才最终更名为深度学习。随后，深度学习技术开始迅猛发展。2012 年，Alex Krizhevsky 等人<sup>[19]</sup>提出了深度卷积神经网络 AlexNet，该网络的提出掀起了卷积神经网络的研究热潮，至此，使用深度学习技术代替传统的机器学习方法，被广泛应用于各大领域。AlexNet 网络构成简单，卷积层只有 5 层、全连接层也只有 3 层，该网络成功地对 ImageNet 数据集中 1000 个类别图像进行了识别。2014 年，Simonyan 等人<sup>[20]</sup>提出了 VGG 网络，该网络通过增加网络深度的方式，达到了提升网络性能的目的，网络最深达到了 19 层，但是引入的模型参数量很大，无法在移动设备上部署。2015 年，He 等人<sup>[21]</sup>创造性的提出了 ResNet 网络，通过对一系列的残差块进行堆叠使用，解决了因为网络层数增加带来的问题，即模型会出现梯度消失和梯度爆炸，网络层数最深达到了 152 层。之前所做工作是为了提升网络性能，期望能真正将这些技术运用到实际中，来解决实际问题，但是随着网络深度的加深，网络的参数量会变得很大，无法在移动设备上部署。所以，大量研究都在尽量保证网络性能的前提下来减小模型参数量，以此获得轻量化的模型。在图像识别领域，轻量级的网络主要有 MobileNet 系列网络<sup>[22][23][24]</sup>、ShuffleNet 系列网络<sup>[25][26]</sup>和 EfficientNet 系列网络<sup>[27][28]</sup>。同时，深度学习技术也广泛应用到了图像分割和目标检测任务上，常见的图像分割算法有 FCN<sup>[29]</sup>、U-Net<sup>[30]</sup>、SegNet<sup>[31]</sup>、PSPNet<sup>[32]</sup>和 DeepLab 系列网络<sup>[33][34]</sup>。在目标检测任务中，常用的算法有 RCNN 网络<sup>[35]</sup>、Faster RCNN 网络<sup>[36]</sup>和 YOLO 系列的经典网络<sup>[37][38]</sup>。

近年来，深度学习技术已经成为解决医疗诊断、工业生产及检测和农业自动化等实际问题的重要工具，这是因为深度学习技术在图像处理任务上取得了显著的效果。故此，越来越多的研究者都开始使用深度学习技术来解决昆虫图像识别等相关问题。2021 年方伟政<sup>[39]</sup>以 ResNet50 网络为基础模型，通过在基础模型中添加注意力机制、双线性池化和修改损失函数等方法，在自建的细粒度昆虫图像

识别任务上取得了良好的效果。2022 年曾伟辉<sup>[40]</sup>等人针对害虫图像识别率低的问题，对数据集进行数据增强，并在 ResNet50 网络中引入自校准卷积，通过多次堆叠 ResNet 网络中的卷积块来提升网络性能，最终在水稻害虫数据集上取得了 91.2% 的识别准确率。上述使用深度学习技术解决图像处理问题已经取得了较好效果，但在解决图像识别问题时，需要根据图像中物体的实际情况和特征进行具体分析。

### 1.2.2 图像分割研究现状

随着计算机的飞速发展，深度学习技术不仅在图像识别任务上应用广泛，而且在图像分割任务上也得到了普遍应用。图像分割是将图像中的信息在像素级上进行分类的过程，属于细粒度图像分类任务。图像分割在自动驾驶、生物医学、工业自动化、卫星遥感和航空影像等方面发挥着重要作用<sup>[41]</sup>，图像分割的目的是去除掉图像中的干扰信息，使研究的目标物体易于分析。

2012 年，全卷积神经网络(Fully Convolutional Networks, FCN)的诞生，在图像分割领域具有非常重要的意义。与图像分类网络不同的是，该网络使用卷积层替换了原始卷积神经网络中的全连接层，网络可以接受任意大小的输入图像，然后对输入的图像进行分割。但是它对上下文信息的处理能力不是那么有效；在全卷积神经网络的启发下，2015 年 U-Net 网络应运而生，它是在卷积神经网络上进行的改进，U-Net 网络的组成结构是编码器—解码器的结构，是非常对称的<sup>[42]</sup>。U-Net 网络处理全局上下文信息的能力优于全卷积神经网络，主要用来解决医学图像领域的分割问题；同年 Badrinarayanan 等人提出了 SegNet 网络结构，以解决自动驾驶和机器人领域的图像分割问题。SegNet 的创新之处在于它的解码器部分对于低分辨率的输入特征图的采样方式，具体来说，解码器中的上采样操作使用的是编码器中最大池化所用的像素索引来进行的非线性上采样，该操作减少了内存的占用，并提高了运行速率。但 SegNet 网络没能实现高低层语义信息的融合<sup>[43]</sup>；金字塔场景解析网络(Pyramid Scene Parseing Network, PSPNet)在 2017 年被提出，它的核心模块是金字塔池化模块，该模块可以学习到目标物体的不同尺度信息，并将它们进行整合，以提高网络获取多尺度特征的能力；到了 2018 年，DeepLab 系列迎来了它的最后一个版本 DeepLabv3+。该网络也是由编码模

块和解码模块构成,其中,编码模块通过使用不同大小的膨胀率实现了目标物体的多尺度信息提取,解码模块则用于恢复目标物体的边界信息<sup>[44]</sup>。

随着深度学习技术的不断发展,越来越多的研究者倾向于使用图像分割技术来对昆虫图像进行分析和研究。陈冬梅等人<sup>[45]</sup>以田间作物上的昆虫为研究对象,为了对昆虫图像进行准确而高效地分割,以 ResNet101 为特征提取主干网络,使用端到端的分割算法对复杂背景下的昆虫图像与背景进行了准确实时分割。Wang 等人<sup>[46]</sup>将害虫图像首先分成若干块,然后为了提高分割性能,使用 K-means 聚类法对玉米叶、番茄叶和辣椒叶片上的粉虱进行初步分割,最后,使用椭圆的三个数字形态学特征来去除叶脉,实现了害虫与农作物叶片的分割,所提方法对害虫图像的分割准确而稳定。虽然深度学习技术已经在昆虫分割方向取得了一些成果,但是上述研究中的分割图像背景较为单一,所得到的分割效果还有待进一步提升。

目前,使用深度学习方法对自然环境中蝴蝶图像分割的研究还很少,Tang 等人<sup>[47]</sup>通过在 U-Net 网络中引入扩展编码器 HCDC 模块,以提取蝴蝶高维语义特征,实现了 Leeds Butterfly 数据集中 10 类 832 张蝴蝶图像的自动分割;同样地,黄世国等人<sup>[48]</sup>针对复杂背景下的蝴蝶图像分割困难的问题,以 F<sup>3</sup>Net 网络为基础,在迁移学习方法的基础上,通过对模型参数的重新训练和优化,得到了蝴蝶图像自动分割模型;Filali 等人<sup>[49]</sup>也是以 Leeds Butterfly 数据集中蝴蝶图像为研究对象,通过使用 Graph cut 算法来对蝴蝶图像进行分割。上述三位研究者均是对 Leeds Butterfly 数据集中蝴蝶图像进行的分割,该数据集中的蝴蝶种类和数量较少,无法满足实际需求。

### 1.2.3 蝴蝶种类识别的国内外研究现状

早期对蝴蝶种类进行识别都要依赖专业人员,这是因为普通大众缺乏相应的专业知识,很难根据蝴蝶的颜色和纹理等特征来区分它们的种类。但是,一味依赖昆虫学家来进行蝴蝶识别会耗费极大的人力和物力;而且蝴蝶种类繁多,不同种类的蝴蝶,它们的颜色、纹理和形状等方面具有很大差异,而同一属下的蝴蝶又较为相似;在不同的生长阶段,有些蝴蝶甚至会呈现出不同的颜色,还有一些蝴蝶雌雄间的差异也是比较大的;此外,蝴蝶会模仿其生存环境,所以很难将蝴

蝶和背景区分开来。上述这些问题,会给蝴蝶种类识别带来很大困难。也正是因为这些因素,越来越多的研究人员开始借助深度学习技术来对蝴蝶种类进行识别研究。

在早期使用计算机技术对蝴蝶种类进行识别时,研究者首先会对构建的蝴蝶图像进行预处理操作,然后对蝴蝶特征进行提取和优化,最后设计分类器来对蝴蝶种类进行分类。例如,Kaya 等人<sup>[50]</sup>使用局部二值模式(Local Binary Pattern, LBP)和人工神经网络(Artificial Neural Network, ANN)相结合的方法,对 5 个种类的 50 张蝴蝶图像进行了识别。其中, LBP 用于提取蝴蝶的纹理信息,人工神经网络用于对种类进行分类;李凡等人<sup>[51]</sup>使用多尺度曲率直方图(Histograms of Multi-Scale Curvature, HoMSC)提取蝴蝶翅膀的形状信息,并使用图像块的灰度共生矩阵(Gray-Level Co-occurrence Matrix of Image Blocks, GLCMoIB)提取蝴蝶翅膀的纹理信息,并使用 KNN 分类器对 50 个种类的 750 张图像进行识别;Kartika 等人<sup>[52]</sup>在对蝴蝶特征进行提取之前,先对蝴蝶数据集进行了归一化操作,再通过 RGB 和 HSV 方法提取蝴蝶的颜色特征,并使用 LBP 提取蝴蝶的纹理特征,同时还提取了蝴蝶的形状特征。通过提取到的 3 种蝴蝶特征,对 10 类 890 张蝴蝶进行识别,准确率达到 75%;Chen 等人<sup>[53]</sup>考虑到蝴蝶纹理和形状差异较大的情况,提出了一种新的解决方法,即高斯卷积角(Gaussian Convolution Angle, GCA)。该方法由从高斯卷积脉络角度和高斯卷积纹理角度分别提取蝴蝶图像的脉络特征和纹理特征,对 Leeds Butterfly 数据集的 10 类 832 张蝴蝶图像进行了有效识别,识别准确率达到 92.03%。由于自然环境下蝴蝶所处的背景十分复杂,而蝴蝶本身的颜色、纹理和形状信息又纷繁复杂,因此,通过传统的机器学习方法很难提取到蝴蝶图像中的更多有效特征。周爱明等人<sup>[54]</sup>为了对比深度学习技术和传统的模式识别技术,在同一数据集上,分别基于 CaffeNet 和支持向量机(Support Vector Machine, SVM)进行试验,结果表明,使用深度学习的蝴蝶种类识别模型的识别准确率更高,模型的泛化能力更强。所以,在目前的研究中,几乎都是使用深度学习技术来解决蝴蝶种类自动识别难题。

Almryad 等人<sup>[55]</sup>基于迁移学习,以采集到的 10 种土耳其蝴蝶生态照为研究对象,分别使用 VGG16、VGG19 和 ResNet50 这 3 种网络进行实验。结果表明,在 VGG16 模型上取得的效果最好,在测试集上的准确率达到 79.5%;谢娟英



等人<sup>[56]</sup>以 RetinaNet 网络为基础模型,通过引入 DSEA(Direct Squeeze-and-Excitation with global Average pooling)和 DSEM(Direct Squeeze-and-Excitation with global Max pooling)两种注意力机制来对基准模型进行改进,同时使用可变形卷积来增强网络对不同大小蝴蝶的特征提取能力。Rajeena 等人<sup>[57]</sup>首先对 15 种蝴蝶图像进行数据增强,然后基于迁移学习的方法,分别使用 VGG16、VGG19、MobileNet、Xception、ResNet50 和 Inceptionv3 网络对蝴蝶图像进行识别,实验结果证明,使用 Inceptionv3 网络时取得的效果最好。周文进等人<sup>[58]</sup>针对自然环境中的蝴蝶种类识别率低的问题,以 YOLOv3 为基础模型,通过在主干网络中引入通道注意力 MultiSE1D 模块来增强网络的特征提取性能,同时,使用一维卷积替换全连接层,有效降低了模型的参数量。实验结果表明,改进后的模型最终取得的 mAP 达到了 83.2%,网络学习到了蝴蝶的细粒度特征。虽然以上方法使用了深度学习技术来解决蝴蝶种类识别问题,但研究中还存在着识别率低或者所研究的蝴蝶种类少的问题,同时所使用的网络模型都存在着结构复杂和参数大的问题。

### 1.3 主要研究内容

本文以自然环境下的蝴蝶图像为研究对象,利用深度学习技术来研究蝴蝶图像种类自动识别的问题,旨在解决自然环境中蝴蝶所处背景干扰强、识别率低等难题。具体的研究内容包括以下几个方面:

1. 自然环境下蝴蝶图像数据集的构建。因为使用深度学习技术对样本进行训练时,需要足量的图像才能防止出现过拟合等现象。本文以常见的凤蝶科蝴蝶和粉蝶科蝴蝶为研究对象,通过数码相机拍摄、从数据网站获取和使用 Python 语言编写爬虫代码的方式,构建了 32 种 7825 张自然环境下的蝴蝶图像数据集。本文所用蝴蝶生态图像均经过昆虫学家识别确认。为了防止过拟合,本文使用高斯噪声、椒盐噪声、左右翻转、增强亮度等数据增强方式对原始蝴蝶图像数据集进行扩充,得到了最终的蝴蝶图像数据集。

2. 提出了一种改进的 DeepLabv3+蝴蝶图像分割模型,该模型融合了置换注意力机制与密集空洞空间金字塔池化。具体而言,首先以 DeepLabv3+网络为基准模型,使用 EfficientNetV2-S 替换 DeepLabv3+的主干提取网络 Xception,提升

主干网络特征提取能力。其次对加强特征提取网络进行改进，将密集连接的思想引入到空洞空间金字塔池化网络，以提取蝴蝶多尺度特征，同时增加采样点的密集性，使获得的蝴蝶特征更加丰富。最后为了增强蝴蝶特征的信息权重，削弱冗余信息对蝴蝶本身特征的干扰，本文在改进的空洞空间金字塔池化网络后引入了轻量级置换注意力机制模块。与其它主流分割网络进行对比，改进后的蝴蝶图像分割网络效果更优，实验证明了该方法的有效性。

3. 构建了一种以残差网络为基础的蝴蝶种类自动识别模型。首先将去除了背景噪声干扰的蝴蝶图像与原图进行叠加，制作识别网络的蝴蝶图像数据集，并使用数据增强方法对蝴蝶图像进行扩充。其次考虑到本文蝴蝶图像种类在数量上的分布不均衡问题，实验中使用 PolyLoss-Focal Loss 损失函数代替交叉熵损失函数进行训练。根据蝴蝶形状和大小的不同，引入可变形卷积增强网络对不同形状、不同大小蝴蝶图像的特征提取能力。最后借鉴深度可分离卷积将普通卷积分解为深度卷积和逐点卷积的思想，提出了改进的深度可分离卷积。与深度可分离卷积相比，改进的深度可分离卷积能从更大程度上降低模型的参数量与运算量。改进后的蝴蝶图像识别网络的识别效果更好、模型更轻量化，与它主流分类网络相比，体现出了该网络的有效性。

## 1.4 论文章节安排

本文使用深度学习技术来对自然环境下的蝴蝶种类识别进行研究。分别从图像分割和图像识别两方面入手，针对现有研究中蝴蝶种类识别存在的问题进行改进。论文具体章节安排如下：

第一章：绪论。首先介绍了对蝴蝶种类识别研究的背景和意义，对深度学习的研究现状进行了分析。其次叙述了图像分割的研究现状，并对目前蝴蝶图像的分割进行了总结，同时从国内外研究分析了蝴蝶种类识别的研究现状。最后介绍了本文的主要研究内容和章节安排。

第二章：深度学习的相关理论基础。首先概述了卷积神经网络的结构和主要组成成分。其次介绍了本文用到的主流图像分割网络的构成和特点。最后对本文使用的主流图像识别网络做了简要描述。

第三章：基于改进的 DeepLabv3+蝴蝶图像分割模型。首先制作了自然环境

下的蝴蝶图像数据集，并对其进行了数据标注和数据增强。其次介绍了蝴蝶图像分割的总体模型结构，并从网络分割性能的角度进行考虑，对原始 DeepLabv3+ 网络的主干网络进行替换。对原始空洞空间金字塔池化网络进行了改进，同时引入轻量级置换注意力模块，增强蝴蝶所在区域和所在通道的权重，以此增强蝴蝶高维语义信息的特征提取能力。最后对实验环境、参数设置和图像分割领域的常用评价指标进行了介绍，并对分割网络的结果进行了分析。

第四章：基于残差网络和可变形卷积的蝴蝶种类识别模型。首先制作了蝴蝶图像识别所需要的数据集，并对识别数据集进行划分和数据增强。其次，从蝴蝶种类识别的整体模型结构，以及对引入的 PolyLoss-Focal Loss 损失函数、可变形卷积和提出的改进的深度可分离卷积四个部分分别进行了介绍。最后对用于蝴蝶图像识别的实验环境、参数设置和评价指标进行了介绍，并对实验结果进行了分析。

第五章：总结与展望。对本文的主要研究成果和研究中存在的不足进行了总结，并对未来的蝴蝶种类识别研究进行了展望。

## 第二章 深度学习的相关理论基础

最近几年以来,深度学习技术飞速发展,该技术已经被用来解决了一些实际问题,例如在自动驾驶、医学诊断和工业瑕疵检测等领域。深度学习技术的广泛使用,不仅提高了人们的生活质量,还降低了人力和物力的成本。由于本文工作涉及到了一些深度学习的基本理论知识,本章节首先介绍了卷积神经网络的概念和基本组成结构,然后对文中用到的一些图像分割算法进行了简要概括,最后对常用的图像识别网络进行了介绍。

### 2.1 卷积神经网络

卷积神经网络(Convolutional Neural Networks, CNN)是一种前馈神经网络,它在图像处理领域具有很大的优势,并取得了很好的应用效果。如运用图像识别、目标检测和图像分割等算法,解决了科研和人类生活生产中的诸多实际问题。

卷积神经网络主要是由输入层、卷积层、池化层、全连接层和输出层构成<sup>[59]</sup>。其中卷积层在卷积神经网络中发挥着重要作用,因为输入图像的特征主要依靠该层进行提取。接着将经过卷积层后得到的特征图送到池化层中,池化层对特征图的大小进行处理,起到降低维度的作用。全连接层起到了桥梁作用,将池化层与输出层进行连接,全连接层能够对前边层所提取到的特征进行处理,最后根据类别个数实现图像分类的目的。卷积神经网络的基本结构如图 2-1 所示。

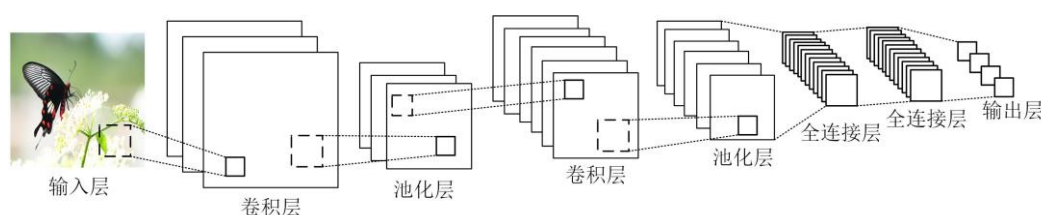


图 2-1 卷积神经网络结构

#### 2.1.1 卷积层

卷积层是卷积神经网络结构中最重要的一部分,因为它的每一个卷积核都能对输入图像的数据进行卷积操作,从而提取出输入图像的特征。不同的卷积核进行卷积操作后会形成不同的特征向量,所有的特征向量构成了一个卷积层。网络对图像数据的特征提取能力越强,表现出来的效果就越好。

卷积层中的第一次卷积运算操作是从输入特征图的最左上角开始，以卷积核的大小为依据进行卷积运算。在图 2-2 中，对卷积运算的具体操作进行了详细的解释。从图 2-2 可知，第一次卷积操作时，卷积核的大小为  $3 \times 3$ ，它与输入特征图最左上角的  $3 \times 3$  元素，在对应位置的对应点上先进行相乘操作，然后进行相加操作，最后得到第一次卷积后的结果。第二次及以后的卷积操作，除了考虑卷积核大小，还要考虑步长大小，并根据步长进行移动，直到最后一次卷积完成，得到该通道的输出特征图，然后将所有通道的输出特征堆叠在一起，得到完整的输出特征图。

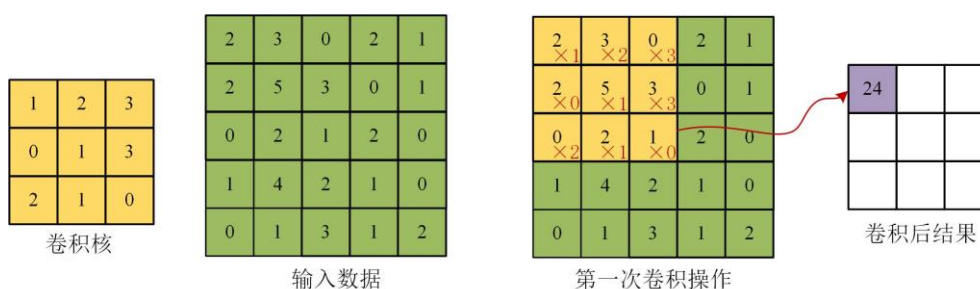


图 2-2 卷积运算操作

### 2.1.2 池化层

池化层中的池化过程往往紧随卷积运算过程之后。池化操作的主要作用是对特征图进行采样，而采样的目的是为了对数据进行降低维度处理，以此来提高运算的效率。通常情况下，池化操作会使用最大池化(Max Pooling)或者平均池化(Average Pooling)这两种池化方式。

池化层接收到卷积层输出的特征图后，会将其划分成若干个小块，然后对每个小块进行处理，处理后的每个小块仅输出一个数字。若输出的是最大值，那便是最大池化，如图 2-3 所示。如果输出的是小块中所有值的平均值，那就是进行的平均池化操作，如图 2-4 所示。

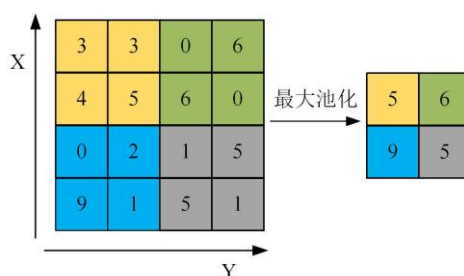


图 2-3 最大池化操作

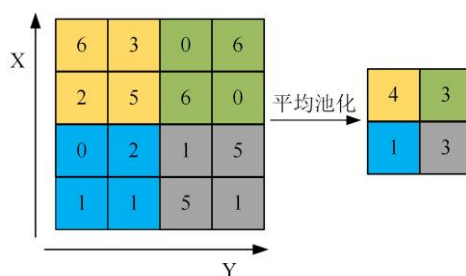


图 2-4 平均池化操作

### 2.1.3 全连接层

输入图像的特征会通过卷积层和池化层进行提取，然后将提取到的信息传递给全连接层。在卷积神经网络中，全连接层是输出层之前的那一层，它的每个神经元会将它前一层和后一层的所有神经元一一连接。该连接方式可将通过卷积层和池化层后得到的特征图中所有信息在全连接层中进行整合与传递。最后，全连接层与输出层相连接，输出层再根据所有需要分类的类别数目进行分类操作，以达到图像分类的目的。

## 2.2 主流图像分割网络概述

图像分割通常分为基于传统方法的图像分割和基于深度学习的图像分割，然而，基于传统方法的图像分割对背景复杂的图像处理效果欠佳，所以目前对图像进行分割使用的基本都是基于深度学习的图像分割方法。基于深度学习的图像分割算法已经在医学图像、遥感和农业等领域广泛应用。本节将对 SegNet、PSPNet、HRNet 和 DeepLabv3+这四种常用的分割网络做简单介绍。

### 2.2.1 SegNet 网络

2017 年，Badrinarayanan 等人提出了 SegNet 分割网络，该网络主要由卷积层堆叠而成，网络结构如图 2-5 所示。由图 2-5 可知，SegNet 网络由编码器和解码器两个部分组成<sup>[60]</sup>，这两个部分是十分对称的。

SegNet 分割网络的编码模块用到的是 VGG16 网络中去除了全连接层的 13 层结构，以提高模型的高维特征提取能力。相比于直接使用 VGG16 网络，去掉 3 层全连接层后，模型不会引入过大的参数量，达到提高分割效果和降低参数量

的目的。SegNet 分割网络在结构上是对称的，解码部分与编码部分一一对应，编码部分使用到了最大池化，以此来降低输入特征图的维度，该操作还会记录下最大池化所在的位置，然后解码的时候根据编码时记录的位置索引号进行恢复，SegNet 网络中保存最大池化索引的操作会减小模型的运算量与参数量。最后，使用 Softmax 函数得到每个类别中像素点的最大预测值，实现最终的图像分割。

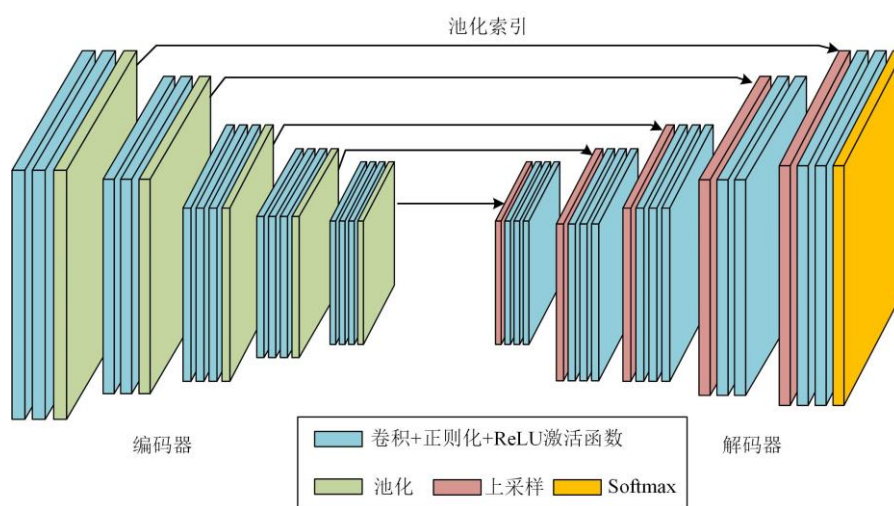


图 2-5 SegNet 网络结构<sup>[60]</sup>

### 2.2.2 PSPNet 网络

PSPNet 于 2017 年被提出，该网络的提出是为了对场景问题进行解析，最大的特点是实现了多尺度特征提取。PSPNet 是在 FCN 网络的基础上进行的改进，提出了金字塔池模块，该结构可以提取到图像中需要分割物体的多尺度特征，从而捕捉到更多的全局信息，降低分割错误率。PSPNet 网络的具体结构图如图 2-6 所示。

输入图像首先经过 ResNet 网络进行特征提取，得到的特征图接着就会进行池化操作。PSPNet 网络的池化操作主要由金字塔池模块实现，而金字塔池模块又是 PSPNet 网络结构中最突出的部分。该模块会将输入进来的特征层进行细化得到 4 个子区域，使用  $1 \times 1$ 、 $2 \times 2$ 、 $3 \times 3$  和  $6 \times 6$  的卷积核对每个子区域进行平均池化操作，以提取不同尺度的特征。然后在每个金字塔层级后使用  $1 \times 1$  的卷积核来降低维度，使其与全局信息权重一致。使用双线性插值法将四个不同尺度的特征结果进行上采样，得到与输入特征图大小相同的输出特征图，然后将它们与输入特征图拼接在一起，最后进行卷积操作，得到预测结果。



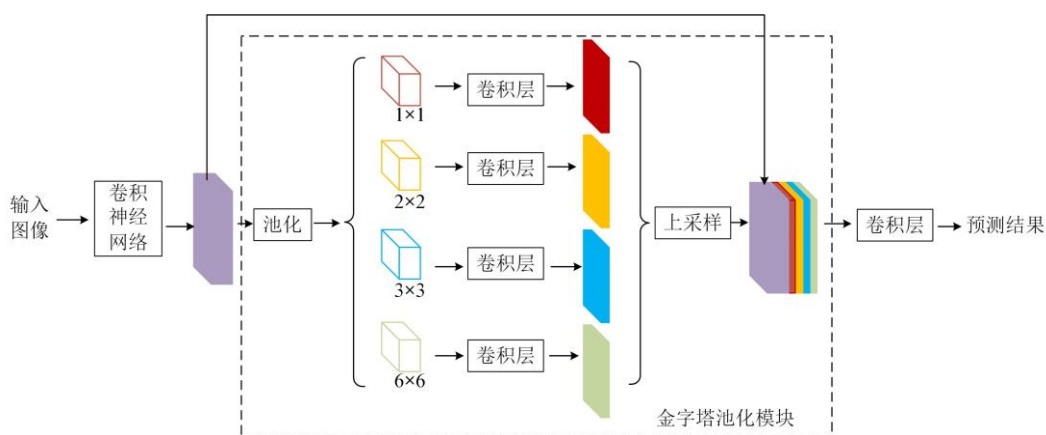


图 2-6 PSPNet 网络结构<sup>[32]</sup>

### 2.2.3 HRNet 网络

2019 年，微软亚洲研究院提出了 HRNet 网络<sup>[61]</sup>，该网络的特点在于当网络层数较深时，网络的高分辨率仍然可以得到保持。HRNet 网络最开始提出来是用于人类姿势估计，其与以往的网络不同之处在于，它采用高分辨率特征图与低分辨率特征图并行连接的方式，使网络的高分辨率特征得以保持。

HRNet 网络的主干网络结构如图 2-7 所示。由图 2-7 可知，不同分辨率的卷积是并联的，这使得整个过程中的高分辨率特征得到了保持。区别于传统分类网络，HRNet 网络首先构建的是高分辨率子网络，然后再逐步引入低分辨率子网络，最后将高低分辨率的各个子网络并行连接。HRNet 网络的低分辨率特征与高分辨率特征并行连接，实现了信息交互。这种信息交互使得多尺度语义信息融合在了一起。HRNet 中高分辨率的保持，使得图像中的目标物体的空间位置信息得以保持，这对图像分割任务来说是非常重要的。

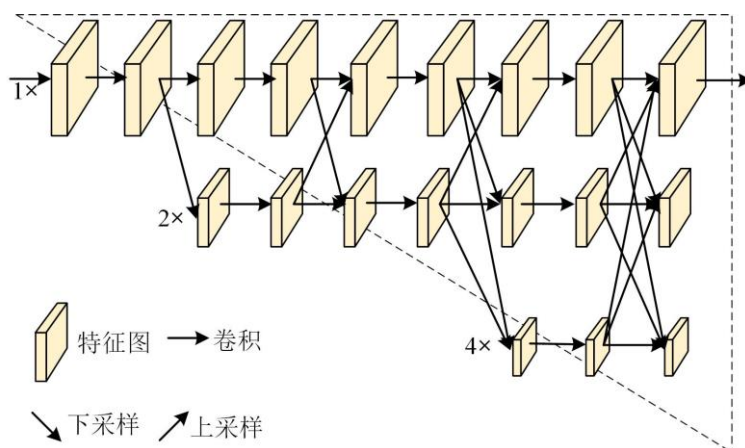


图 2-7 HRNet 主干网络结构<sup>[61]</sup>



## 2.2.4 DeepLabv3+网络

DeepLabv3+网络由编码模块和解码模块组成<sup>[62]</sup>。其中，编码模块的主要作用是对输入的图像进行特征提取，它由主干特征提取网络 Xception<sup>[63]</sup>和空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)<sup>[64]</sup>构成。具体操作为：首先输入图像通过 Xception 网络获得浅层和深层两个特征层。然后深层特征会通过 1 个  $1\times 1$  卷积层、3 个膨胀率为 6, 12, 18 的  $3\times 3$  空洞卷积层以及 1 个池化层进行特征提取，接着将得到的特征进行堆叠。最后，将堆叠后的结果经过  $1\times 1$  卷积进行通道数调整。利用不同膨胀率的空洞卷积可以提高网络感受野，使网络能够提取到图像中目标物体的多尺度特征，起到加强特征提取的作用，进而提高网络分割精度。

解码模块起到了恢复边界信息的作用。其主要是对浅层特征层和编码模块的输出特征进行操作，从而获得输入图像中每个像素点所对应种类的预测结果。具体操作为：首先将通过 Xception 网络获得的浅层信息经过  $1\times 1$  卷积处理；然后再与编码模块获得的较高语义信息进行 4 倍上采样后的结果进行特征融合；接着利用  $3\times 3$  卷积处理拼接后的特征图，最后再采取 4 倍上采样操作，将分割图像大小恢复成与输入图像相同的大小。DeepLabv3+网络的结构图如图 2-8 所示。DeepLabv3+网络的编码模块和解码模块实现了低层语义信息和高层语义信息的融合，提升了图像分割精度。

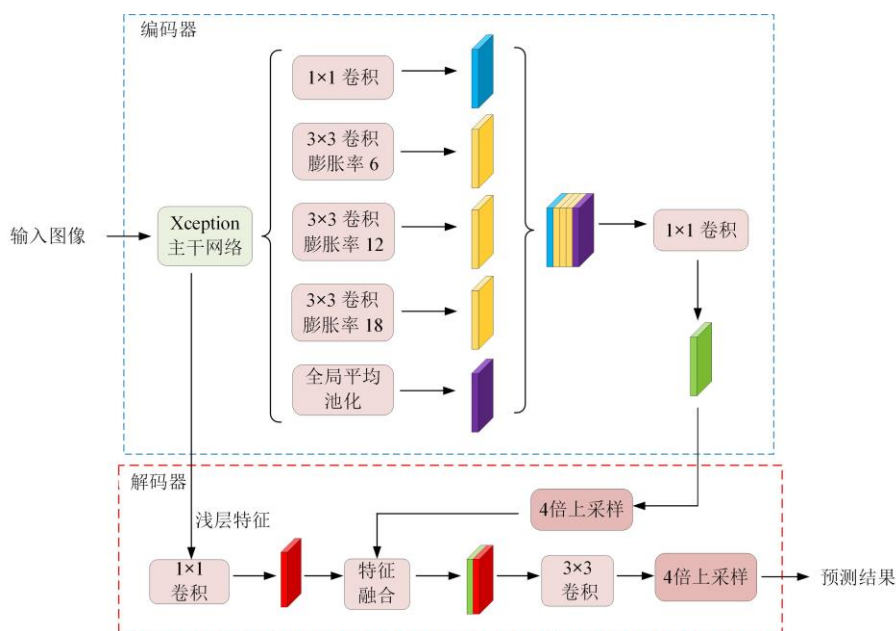


图 2-8 DeepLabv3+网络结构<sup>[62]</sup>

## 2.3 主流图像分类网络概述

图像分类任务的核心是对图像中的物体进行准确分类。通常，网络会根据标签类别对输入图像中的物体进行分类，分类误差越小，分类效果越好。深度学习技术在图像分类任务中表现出了卓越的能力，不仅可以对不同物体之间的类别进行划分，还可以实现具体物体类别之间的更细粒度的分类。本节将介绍目前常用的四种主流分类网络。

### 2.3.1 VGG 网络

VGG 网络从网络深度对网络性能的影响出发，证明了增加网络深度能在一定程度上提升网络的性能<sup>[65]</sup>。常见的 VGG 网络结构如表 2-1 所示。

表 2-1 常见的 VGG 网络结构<sup>[66]</sup>

VGG11	VGG11-LRN	VGG13	VGG16	VGG19
input ( $224 \times 224$ RGB image)				
conv3-64	conv3-64 LRN	[conv3-64] $\times$ 2	[conv3-64] $\times$ 2	[conv3-64] $\times$ 2
		maxpool		
conv3-128	conv3-128	[conv3-128] $\times$ 2	[conv3-128] $\times$ 2	[conv3-128] $\times$ 2
		maxpool		
[conv3-256] $\times$ 2	[conv3-256] $\times$ 2	[conv3-256] $\times$ 2	[conv3-256] $\times$ 3	[conv3-256] $\times$ 4
		maxpool		
[conv3-512] $\times$ 2	[conv3-512] $\times$ 2	[conv3-512] $\times$ 2	[conv3-512] $\times$ 3	[conv3-512] $\times$ 4
		maxpool		
[conv3-512] $\times$ 2	[conv3-512] $\times$ 2	[conv3-512] $\times$ 2	[conv3-512] $\times$ 3	[conv3-512] $\times$ 4
		maxpool		
		FC-4096		
		FC-4096		
		FC-4096		
		soft-max		

VGG 中常用的是 VGG16 和 VGG19, 该网络是对 AlexNet 网络进行的改进, 其核心思想是网络中全部使用  $3 \times 3$  的卷积核来对 AlexNet 中较大的  $5 \times 5$  和  $7 \times 7$  卷积核进行替换。VGG 网络用 2 个  $3 \times 3$  卷积核来代替 1 个  $5 \times 5$  的卷积核, 当然 1 个  $7 \times 7$  的卷积核是需要用 3 个  $3 \times 3$  卷积核来代替的。通过这种处理方式, 网络的深度得到了加深, 从而网络的性能得到了有效提升, 同时还能保证处理前后网络具有相同的感受野。

### 2.3.2 GoogLeNet 网络

GoogLeNet 网络<sup>[67]</sup>是一种基于 Inception 模块的深度学习神经网络。由 2.3.1 分析可知, VGG 网络是通过增加网络深度来提升性能的, 但是当层数过深时, 网络可能会出现准确率饱和甚至下降的问题, 这导致网络变得很难优化。同时过深的网络会导致网络模型参数量过大, 无法将其应用到实际模型部署中。对此, 谷歌提出了 Inception 模块。

Inception 模块是 GoogLeNet 网络的核心。该结构是一种稀疏性结构<sup>[68]</sup>, 可以将特征高效地表达出来。Inception 模块由卷积和池化组成, 其结构如图 2-9 所示。其中,  $1 \times 1$  卷积,  $3 \times 3$  卷积和  $5 \times 5$  卷积可以提取到目标物体的不同尺度信息, 同时, 最左边支路上的  $1 \times 1$  卷积还具有降低维度的作用, 起到了减小模型的运算量的作用。

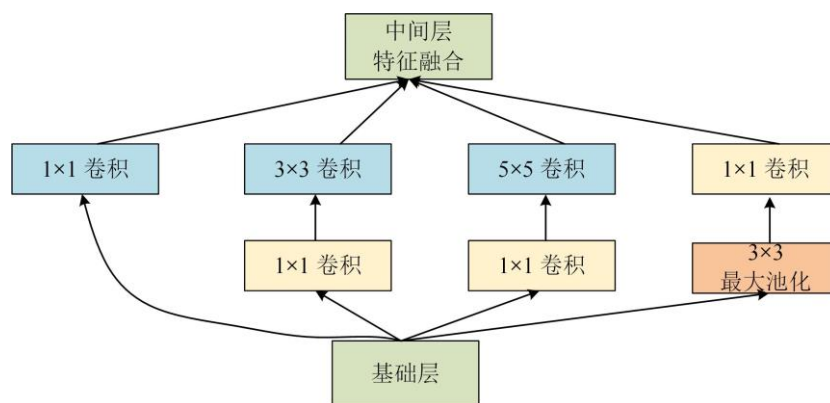


图 2-9 Inception 模块结构<sup>[67]</sup>

### 2.3.3 ResNet 网络

深度残差网络 ResNet 的提出是具有里程碑意义的。虽然加深网络深度可以在一定程度上提升网络的性能, 但是网络深度并不是越深越好。层数太深的网络

不仅无法取得较好的网络性能，还会陷入梯度消失或者梯度爆炸的难题。ResNet 网络的提出恰好解决了这个问题，它的网络结构如表 2-2 所示。

表 2-2 ResNet 网络结构<sup>[21]</sup>

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

ResNet 网络由一系列的残差块堆叠而成，残差块的单元结构图如图 2-10 所示。该结构中使用了残差连接，当输入为 $x$ 时，经过权重层后得到的残差映射为 $F(x)$ ，再将其与快捷连接支路相加，得到的融合特征输出为 $H(x) = F(x) + x$ ，则 $F(x) = H(x) - x$ 。ResNet 网络通过层与层之间的残差连接，实现了浅层特征的引入，有效解决了网络过深引发的问题。

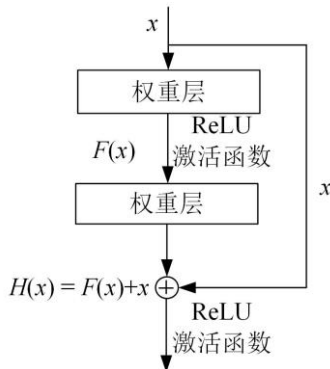


图 2-10 残差结构示意图<sup>[21]</sup>

由表 2-2 可知，ResNet 网络最小的网络层数有 18 层，最大可达 152 层。针对网络层数的不同，ResNet 提出了两种残差结构，结构如图 2-11 所示。其中左边的残差结构适用于 ResNet18 和 ResNet34，使用的是两个  $3 \times 3$  卷积。右边的残差结构适用于 ResNet50、ResNet101 和 ResNet152，这些网络结构都比较深，为了对通道维度进行处理，通过在残差结构中使用两次  $1 \times 1$  卷积实现，使得网络性能得到保证的同时，减小网络的运算量和参数量。

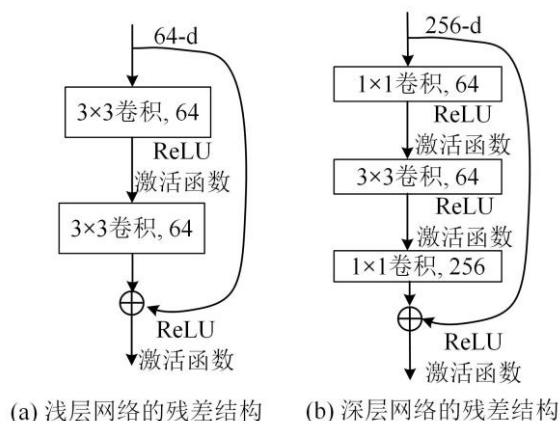


图 2-11 ResNet 网络中的两种残差结构<sup>[21]</sup>

### 2.3.4 MobileNet 网络

2017 年，谷歌提出了 MobileNet 网络，它在轻量级卷积神经网络中具有重要意义。MobileNet 网络的轻量化得益于深度可分离卷积<sup>[69]</sup>，与普通卷积相比，深度可分离卷积能够大幅度减小模型的参数量与运算量。这是因为深度可分离卷积由深度卷积与逐点卷积组成，这种网络构成实现了模型轻量化。图 2-12 为深度可分离卷积的网络结构。

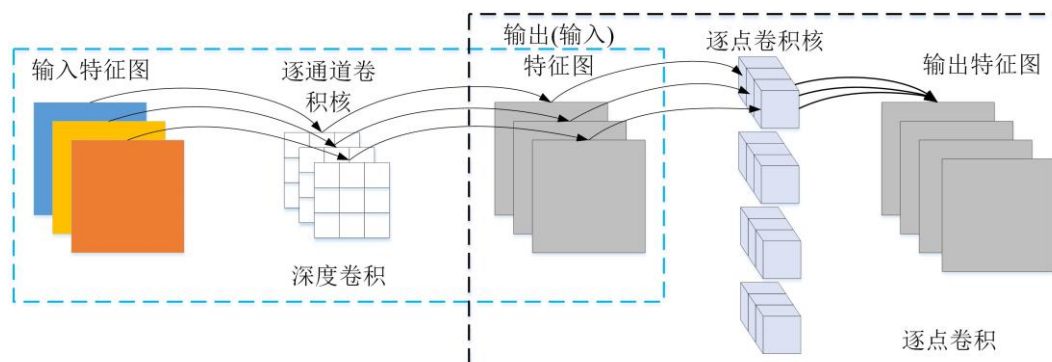


图 2-12 深度可分离卷积网络结构<sup>[70]</sup>

MobileNet 网络的结构如表 2-3 所示，其中 dw 表示深度卷积。

表 2-3 MobileNet 网络结构<sup>[22]</sup>

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5×5 Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
5×5 Conv s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool $7 \times 7$	$7 \times 7 \times 1024$
FC / s1	$1024 \times 1000$	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

深度可分离卷积中的一个卷积核对应一个输入通道,而普通卷积中的卷积核对所有输入通道都是起作用的。同时,逐点卷积是指卷积核大小为  $1 \times 1$  的普通卷积,它能够将深度卷积获得的通道信息整合在一起,图 2-13 为普通卷积和深度可分离卷积的基本结构。因为 MobileNet 网络中提出了深度可分离卷积,而该卷积又能大幅度减小模型参数量与运算量,所以该网络成为了轻量级网络的经典。

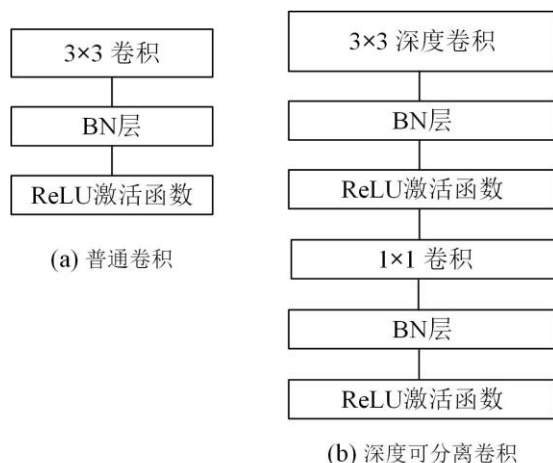


图 2-13 普通卷积与深度可分离卷积的基本结构<sup>[22]</sup>

### 2.3.5 ShuffleNet-V2 网络

继 MobileNet 网络之后，轻量级卷积神经网络飞速发展。2018 年旷世提出了更高效、更轻量级的 ShuffleNet-V2 卷积神经网络，该网络是在 ShuffleNet 网络的基础上进行的改进。ShuffleNet-V2 主要由基本单元和下采样单元组成<sup>[71]</sup>，基本单元和下采样单元的具体结构组成如图 2-14 所示。其中基本单元通过通道分离和通道混洗实现了信息融合；下采样单元是将特征图输入到两个分支路，两支路在经过卷积操作后，会在通道维度进行拼接，得到的特征图通道数为原始特征图通道数的两倍，最后将特征图进行通道混洗。下采样单元的处理方式进一步增强了网络的特征提取能力。

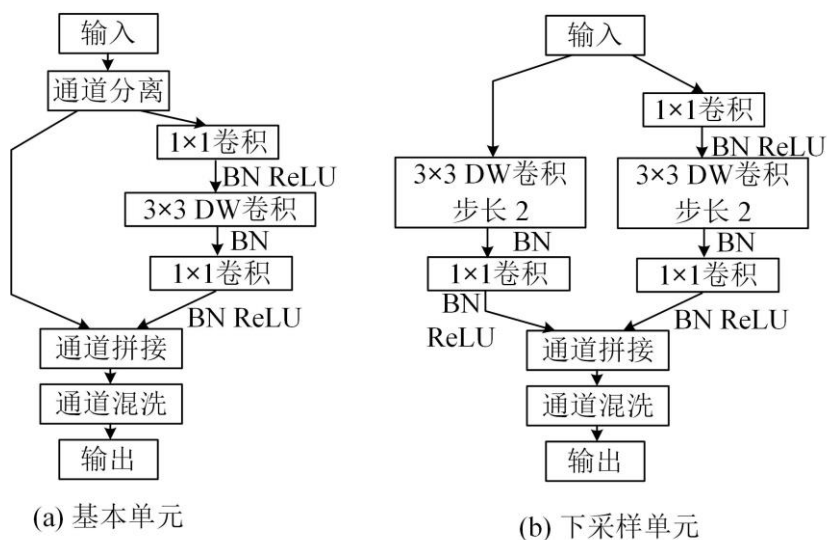


图 2-14 ShuffleNet-V2 的基本单元和下采样单元<sup>[26]</sup>

### 2.3.6 EfficientNet 网络

为了平衡网络宽度、网络深度和分辨率,谷歌团队提出了 EfficientNet 网络。EfficientNet 网络能够搜索到网络宽度、网络深度和分辨率三者的最佳值是通过使用神经网络结构搜索(Neural Architecture Search, NAS)技术实现的。在相似参数量的网络中, EfficientNet 网络的准确率更高。EfficientNet 网络的基本模块是 MBConv 模块<sup>[72]</sup>,该模块由两个  $1 \times 1$  卷积、BN 层、Swish 激活函数、深度可分离卷积、SE 注意力机制和 Dropout 层组成,具体的组成结构由图 2-15 所示。其中深度可分离卷积用于减小模型参数量,进而提升效率。SE 注意力的引入可以增强特征提取能力,得到更好的效果。EfficientNet 网络中最常见的是 EfficientNet-B0,其具体结构参数如表 2-4 所示。

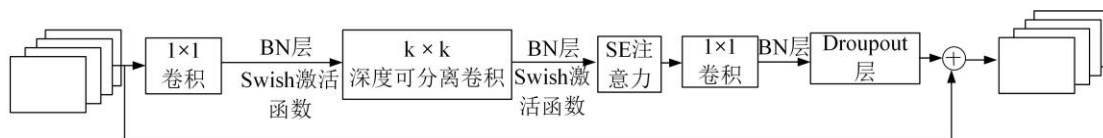


图 2-15 MBConv 模块结构<sup>[72]</sup>

表 2-4 EfficientNet-B0 网络结构参数<sup>[27]</sup>

Stage	Block name	Block number	Resolution	Channels
1	Conv3×3	1	224×224	32
2	MBConv, k3×3	1	112×112	16
3	MBConv, k3×3	2	112×112	24
4	MBConv, k5×5	2	56×56	40
5	MBConv, k3×3	3	28×28	80
6	MBConv, k5×5	3	14×14	112
7	MBConv, k5×5	4	14×14	192
8	MBConv, k3×3	1	7×7	320
9	Conv1×1 & Pooling & FC	1	7×7	1280

## 2.4 本章小结

本章主要介绍了深度学习的相关理论。首先,从卷积神经网络的概述和组成入手,对卷积神经网络中的卷积层、池化层和全连接层的相关理论做了简要概述。



其次，介绍了本文所使用的主流图像分割网络的理论。最后，从网络结构和模型特点等方面，对经典的图像分类网络进行了简要概述。这为本文后续的蝴蝶图像分割和种类识别奠定了基础。

### 第三章 基于改进的 DeepLabv3+蝴蝶图像分割模型

自然环境下的蝴蝶多喜欢飞舞在花丛中，导致采集到的图像背景十分复杂。同时，蝴蝶自身颜色也会根据生存环境的变化而变化，演变完美的甚至会完全融入到当地环境中，如枯叶蝶。此外，蝴蝶图像还受到成像拍摄角度多样等因素的影响，这些因素使得对自然环境下的蝴蝶种类识别存在一定困难。对自然环境中蝴蝶图像进行分割的目的，是为了去除图像中的复杂背景干扰，便于后期对蝴蝶图像种类识别。传统的图像分割方法主要有：基于阈值的分割方法、基于区域的分割方法和基于边缘的分割方法<sup>[73]</sup>，但这些方法对背景噪声复杂的自然图像处理能力较差，无法适用于自然环境下的蝴蝶图像分割。鉴于此，本章以 DeepLabv3+ 为蝴蝶图像分割的基础模型，并从蝴蝶图像特征和网络结构出发，对原始 DeepLabv3+ 网络进行了改进。首先，对主干网络进行替换，以加强对蝴蝶特征的提取；其次，针对空洞空间金字塔池化网络感受野较小和采样点稀疏导致的信息丢失等问题，进行了改进，并引入置换注意力网络对特征进行进一步提取，最终，构建了适用于自然环境下的蝴蝶图像分割模型。

#### 3.1 蝴蝶图像数据集构建

目前公开的蝴蝶图像数据集种类和数量较为有限，特别是自然环境下的蝴蝶生态照的种类和数量更是稀少。对此，本文以常见的凤蝶科蝴蝶和粉蝶科蝴蝶为研究对象，结合贵州省常见的蝴蝶种类和官方竞赛数据集中蝴蝶种类，构建了自然环境下的蝴蝶图像数据集，共计 32 种蝴蝶种类，包含 7825 张图像。该数据集中蝴蝶种类和数量的具体分布情况如表 3-1 所示。

表 3-1 蝴蝶图像数据集分布

蝴蝶种类(中文学名)	蝴蝶种类(拉丁学名)	数量
曙凤蝶	<i>Atrophaneura horishanus</i>	195
瓦曙凤蝶	<i>Atrophaneura varuna</i>	156
白斑麝凤蝶	<i>Byasa dasarada</i>	102
多姿麝凤蝶	<i>Byasa polyeuctes</i>	254
小黑斑凤蝶	<i>Chilasa epycides</i>	77

表 3-1 蝴蝶图像数据集分布

续表

蝴蝶种类(中文学名)	蝴蝶种类(拉丁学名)	数量
统帅青凤蝶	Graphium_agamemnon	356
宽带青凤蝶	Graphium_cloanthus	224
青凤蝶	Graphium_sarpedon	231
旖凤蝶	Iphiclides_podalirius	253
燕凤蝶	Lamproptera_curius	206
绿带燕凤蝶	Lamproptera_meges	105
锤尾凤蝶	Losaria_conon	92
钩凤蝶	Meandrusa_payeni	91
褐钩凤蝶	Meandrusa_sciron	85
红珠凤蝶	Pachliopta_aristolochiae	219
红基美凤蝶	Papilio_alcmenor	119
窄斑翠凤蝶	Papilio_arcturus	117
碧凤蝶	Papilio_bianor	187
达摩凤蝶	Papilio_demoleus	238
金凤蝶	Papilio_machaon	253
美凤蝶	Papilio_memnon	192
巴黎翠凤蝶	Papilio_paris	234
玉带凤蝶	Papilio_polytes	246
绿凤蝶	Pathysa_antiphates	175
丝带凤蝶	Sericinus_montelus	237
迁粉蝶	Catopsilia_pomona	433
斑缘豆粉蝶	Colias_erate	335
报喜斑粉蝶	Delias_pasithoe	486
钩粉蝶	Gonepteryx_rhamni	907
鹤顶粉蝶	Hebomoia_glaucippe	459
东方菜粉蝶	Pieris_canidia	266
黑纹粉蝶	Pieris_melete	295

本文蝴蝶图像数据集来源有 3 种：首先，与昆虫研究所专家合作，在贵州省不同地区使用佳能数码相机进行拍摄；其次，通过 Kaggle 官方数据网站 (<https://www.kaggle.com>) 和第三届中国数据挖掘竞赛(国际首次蝴蝶识别大赛)<sup>[74]</sup> 获取；最后，编写 Python 爬虫代码，从各个网站获取所需种类的蝴蝶图像。本文所用蝴蝶生态图像均经过昆虫学家识别确认。数据集的种类构成如图 3-1 所示，图中(1)-(32)列出了这 32 类蝴蝶种类，分别表示的是曙凤蝶、瓦曙凤蝶、白斑麝凤蝶、多姿麝凤蝶、小黑斑凤蝶、统帅青凤蝶、宽带青凤蝶、青凤蝶、旖凤蝶、燕凤蝶、绿带燕凤蝶、锤尾凤蝶、钩凤蝶、褐钩凤蝶、红珠凤蝶、红基美凤蝶、窄斑翠凤蝶、碧凤蝶、达摩凤蝶、金凤蝶、美凤蝶、巴黎翠凤蝶、玉带凤蝶、绿凤蝶、丝带凤蝶、迁粉蝶、斑缘豆粉蝶、报喜斑粉蝶、钩粉蝶、鹤顶粉蝶、东方菜粉蝶和黑纹粉蝶。

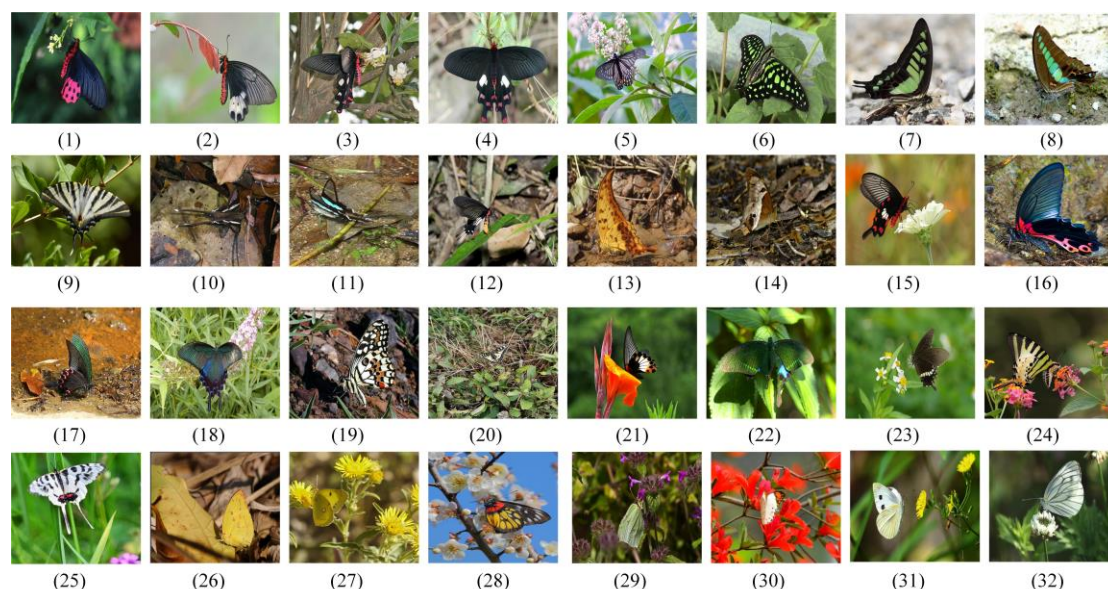


图 3-1 蝴蝶图像数据集

### 3.1.1 蝴蝶图像数据集预处理

由于本文蝴蝶图像数据集来源较广，需要先将其进行预处理操作，然后再送入网络中。本文对蝴蝶图像数据集进行的预处理操作主要有：数据筛选、图像规范、数据标注和数据增强四个部分。首先，对上述三种途径获得的蝴蝶图像进行筛选，以保证图像质量。其次，对图像尺寸进行规范化，以满足网络的输入要求。接着，对处理后的图像进行数据标注，得到蝴蝶分割的图像数据集。最后，鉴于

本文数据样本较少，且蝴蝶类别之间存在着数量分布不均衡的现象，为了防止训练过程中出现过拟合，本文对数据集进行了数据增强，以提高模型的泛化能力。

### 1. 数据筛选

由于蝴蝶图像数据集来源较广，所以需要获得的蝴蝶图像进行处理。首先，去除了蝴蝶拍摄时得到的模糊图像。其次，删除了网页所得蝴蝶的标本图像，并从公开数据集中去除了本文不需要的蝴蝶种类。最后，对经过上述步骤处理后的蝴蝶图像种类进行最后的识别和确认。

### 2. 图像规范

图像规范指的是，对经过数据筛选后的蝴蝶图像进行尺寸和种类类别的统一化处理。由于本文蝴蝶图像来源有三种途径，蝴蝶图像保存的格式和图片大小等均不统一，所以对图像进行规范化处理是必要的。首先，将所有蝴蝶图像调整为.jpg 格式。然后，使用蝴蝶的拉丁学名对所有蝴蝶图像进行命名，每一种类下的蝴蝶图像命名方式按照拉丁学名和阿拉伯数字的顺序命名。考虑到本文所用的图像识别网络和图像分割网络对输入图像尺寸的要求，将所有的蝴蝶图像尺寸调整为  $224 \times 224$ 。

#### 3.1.2 数据标注

本文使用广泛应用于图像标注的 Labelme 工具，对经过数据筛选和图像规范化处理后的蝴蝶图像进行标注。使用 Labelme 标注工具对蝴蝶图像进行标注的操作是：将需要标注的蝴蝶图像导入后，根据蝴蝶在图像中的形状轮廓进行标注。因为图像分割的目的是去除背景，所以在标注时，将类别分为两类，一类是蝴蝶，一类是背景。标注完就会生成对应的.Json 格式文件，然后通过 Python 程序将其转换为可视化的图片标签。图 3-2 展示了 Labelme 标注工具的界面，图 3-3 为根据.Json 文件生成的可视化图片标签。最后，将标注完成的蝴蝶分割数据集按照 7:3 的比例划分为训练集和测试集。

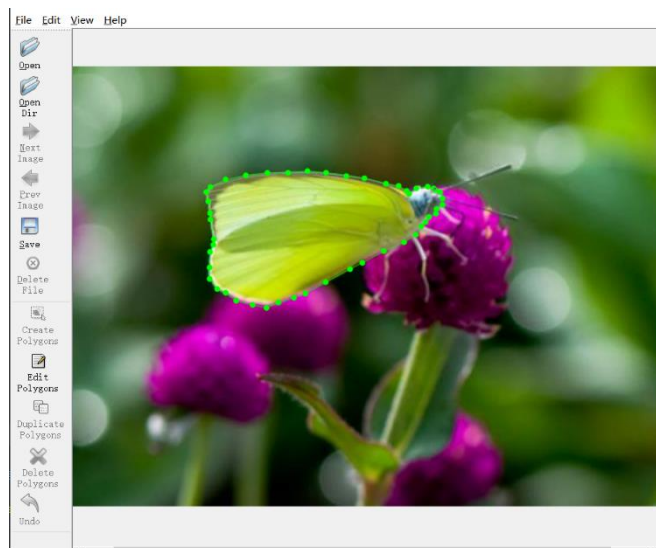


图 3-2 蝴蝶分割图像的标注界面



图 3-3 标签的可视化

### 3.1.3 数据增强

使用深度学习技术对图像进行处理时，样本数量越丰富，模型的训练效果越好，模型的泛化能力也就越强。然而，由于本文数据集的数量较少且种类间数量存在分布不均衡的情况，为避免训练过程中出现过拟合，本章使用高斯噪声、椒盐噪声、左右翻转、增强亮度等数据增强方式，对数据集进行扩充。扩充后，蝴蝶图像数据集共 39125 张。因为本章是对蝴蝶图像进行分割处理，标签图像是 RGB 图像，其中包含有蝴蝶的标注信息。所以在对原始图像做数据增强时，必须对蝴蝶原始图像对应的标签图像进行相同的处理。

以巴黎翠凤蝶为例，选择其中一张图像进行数据增强。图 3-4 为原始蝴蝶图像，经过高斯噪声、椒盐噪声、增强亮度和左右翻转等数据处理后得到的图像。



图 3-4 数据增强

### 3.2 蝴蝶图像分割模型构建

蝴蝶图像分割的效果对蝴蝶种类识别的准确性有着直接影响,这是因为自然环境下蝴蝶背景十分复杂,需要先将蝴蝶图像的复杂背景去除。本章以 DeepLabv3+网络为蝴蝶图像分割的基础模型,针对原始 DeepLabv3+主干网络特征提取能力较差、总体分割精度有待提高的问题,本章首先使用 EfficientNetV2-S 替换 DeepLabv3+网络中的 Xception 主干特征提取网络,替换后的网络可以提取到更加丰富的蝴蝶特征,从而提升蝴蝶图像的分割精度。同时,针对 DeepLabv3+网络中的空洞空间金字塔池化模块感受野较小,蝴蝶图像有效信息利用率低的问题,对空洞空间金字塔池模块进行改进,引入密集空洞空间金字塔池化(DenseNet ASPP, DASPP)作为加强特征提取网络,通过多层级联堆叠的方式,网络感受野进一步增大,获得更加丰富的蝴蝶图像多尺度信息,同时解决了原始网络采样点稀疏而造成的信息利用率的问题。最后,在密集空洞空间金字塔池化后引入置换注意力机制,通过特征分组和通道置换方式,使网络更加关注蝴蝶本身特征,增大了蝴蝶特征信息权重,降低了背景噪声的干扰,进而提升了分割精度。置换注意力机制属于轻量级网络,既增强了网络对蝴蝶特征的提取能力,又不会增加模型参数量。基于上述分析,本章对 DeepLabv3+网络进行改进,构建了融合置换注意力与密集空洞空间金字塔池化的 DeepLabv3+蝴蝶图像分割模型(DeepLabv3+ SA-Net, DS-Net)。



### 3.2.1 蝴蝶图像分割模型总体结构

本章结合蝴蝶图像特征，对 DeepLabv3+ 的主干网络和空洞空间金字塔池化模块进行了改进。提出了一种改进的 DeepLabv3+ 蝴蝶图像分割算法，该算法融合了置换注意力机制和密集空洞空间金字塔池化，其总体框架如图 3-5 所示。

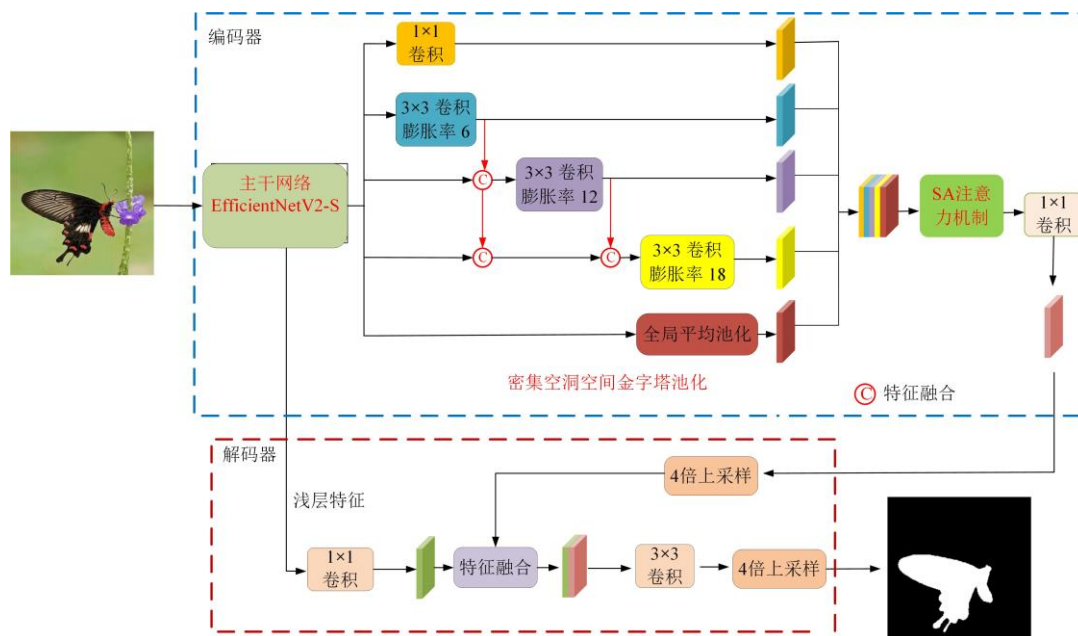


图 3-5 蝴蝶图像分割模型总体结构

与 DeepLabv3+ 一样，DS-Net 网络也由编码模块和解码模块组成。编码模块的主要任务是对输入的蝴蝶图像进行特征提取，解码模块则起到了恢复边界信息的作用。本章主要改进的是 DeepLabv3+ 的编码模块，首先，使用 EfficientNetV2-S 网络替换 DeepLabv3+ 的主干提取网络 Xception，以提升蝴蝶分割精度。其次，将 DenseNet<sup>[75]</sup> 网络中的密集连接思想引入到空洞空间金字塔池化模块中，提取蝴蝶多尺度特征的同时，增加采样点的密集性，使提取到的蝴蝶特征更加丰富。最后，为增强蝴蝶特征信息权重，削弱冗余信息对蝴蝶本身特征的干扰，在改进的空洞空间金字塔池化网络后引入轻量级置换注意力机制模块。

编码部分的具体操作是：首先，输入的蝴蝶图像经过 EfficientNetV2-S 网络进行特征提取，得到浅层和深层两个特征层；然后，深层特征会被送到密集空洞空间金字塔池化进行加强特征提取，并将得到的特征进行堆叠。堆叠后的结果会输入到置换注意力模块中，通过特征分组和通道置换，将通道和空间信息融合在一起；最后，使用 1x1 卷积对通道数进行调整，得到蝴蝶图像的高级语义信息，



经过  $1 \times 1$  卷积处理后的结果会输出到解码模块进行再处理。

解码模块主要是对经过主干特征提取模块输出的浅层特征和来自编码模块的高维语义信息进行解码操作,从而获得预测结果。该结果为输入的蝴蝶图像中每个像素点所对应的种类。具体操作为:首先,将通过 EfficientNetV2-S 网络获得的浅层信息经过  $1 \times 1$  卷积处理;然后,将经过 4 倍上采样后的编码模块的高维语义信息与浅层信息进行特征融合,并利用  $3 \times 3$  卷积提取更细化的特征;最后,通过 4 倍上采样将图像大小恢复成与输入图像相同大小的图像。DS-Net 网络的编码模块和解码模块实现了低层语义信息和高层语义信息的融合,提升了复杂背景下蝴蝶分割的精度。

### 3.2.2 EfficientNetV2-S 主干网络

原始 DeepLabv3+使用 Xception 作为主干特征提取网络,但蝴蝶图像背景噪声复杂,蝴蝶自身颜色、形状和纹理多样,使用 Xception 作为主干特征提取网络会导致模型分割精度较低。为了更好地适应蝴蝶数据集的实际情况,本研究使用 EfficientNetV2-S 作为主干特征提取网络,因为相比于 Xception, EfficientNetV2-S 网络拥有更高的识别准确率。EfficientNetV2-S 网络主要由融合移动翻转瓶颈卷积结构(Fused-MBConv)模块和移动翻转瓶颈卷积结构(MBConv)模块组成<sup>[76]</sup>,其网络结构参数如表 3-2 所示。

表 3-2 EfficientNetV2-S 网络结构参数<sup>[28]</sup>

Stage	Operator	Stride	#Channels	#Layers
0	Conv $3 \times 3$	2	24	1
1	Fused-MBConv1, $k3 \times 3$	1	24	2
2	Fused-MBConv4, $k3 \times 3$	2	48	4
3	Fused-MBConv4, $k3 \times 3$	2	64	4
4	MBConv4, $k3 \times 3$ , SE0.25	2	128	6
5	MBConv6, $k3 \times 3$ , SE0.25	1	160	9
6	MBConv6, $k3 \times 3$ , SE0.25	2	256	15
7	Conv $1 \times 1$ & Pooling & FC	—	1280	1

由表 3-2 可知, Fused-MBConv 模块应用于 1-3 阶段。该模块由 1 个  $3 \times 3$  卷

积、SE 注意力和 1 个  $1 \times 1$  卷积组成，其结构图如图 3-6(a)所示。在网络浅层中使用 Fused-MBConv 模块不仅增强了特征提取能力，还提高了网络的训练速度。MBConv 模块应用于 4-6 阶段，它由两个  $1 \times 1$  卷积、深度可分离卷积和 SE 注意力组成，其结构图如图 3-6(b)所示，其中深度可分离卷积用来减小模型参数量。MBConv 模块和 Fused-MBConv 模块的结合使用，使得 EfficientNetV2-S 网络在模型精度、参数量和推理速度上达到了平衡。因此，本章选择 EfficientNetV2-S 作为 DeepLabv3+新的主干特征提取网络，以此达到提升蝴蝶图像分割精度和像素点准确率的目的。

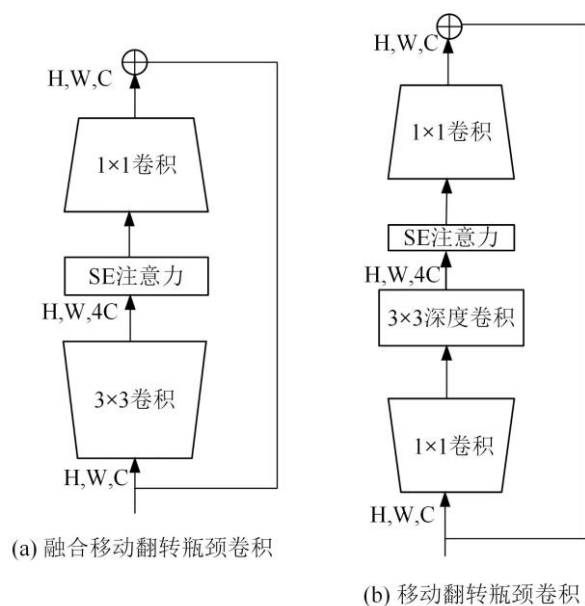


图 3-6 Fused-MBConv 与 MBConv 网络结构<sup>[28]</sup>

### 3.2.3 改进空洞空间金字塔池化

DeepLabv3+模型使用空洞空间金字塔池化作为加强特征提取的网络，其网络结构图如图 3-7(a)所示。由图 3-7(a)可知，ASPP 由 1 个  $1 \times 1$  卷积层、3 个膨胀率为 6, 12, 18 的  $3 \times 3$  空洞卷积层和 1 个池化层组成，能够提取到部分蝴蝶的不同尺度特征。然而，由于 ASPP 中各层之间是相互独立的，后层无法利用前层提取到的特征，所以网络的最大感受野，就是膨胀率最大的那层空洞卷积的感受野，导致网络感受野受到限制。ASPP 增大感受野是通过不同膨胀率空洞卷积实现的，但当膨胀率增大时，需要补的零会越来越多，导致采样点越来越稀疏，使得采样信息大量丢失，从而影响蝴蝶图像分割精度。

为了解决上述问题，受到 DenseNet 网络密集连接思想的启发，本章使用密集空洞空间金字塔池化替换 ASPP，DASPP 的网络结构图如图 3-7(b)所示。由图 3-7(b)可知，DASPP 网络采用了密集连接方式，可将前边层的输出输入到下一层卷积层，使网络获得更大的感受野和更密集的采样点，从而提高网络对蝴蝶特征的提取能力。

对于膨胀率为  $r$ ，卷积核尺寸为  $k$  的空洞卷积，它的感受野大小如式(3-1)所示：

$$R = (r - 1) * (k - 1) + k \quad (3-1)$$

因 ASPP 各层间相互独立，所以 ASPP 的最大感受野应为  $r=18$  的  $3 \times 3$  的空洞卷积，其最大感受野如式(3-2)所示：

$$R_{max} = (r - 1) * (k - 1) + k = 37 \quad (3-2)$$

DASPP 模块能够实现密集连接，即将前边层的输出结果输入到下一层的卷积层，从而使前边层的采样结果在下一层继续采样，以此获得更大的感受野。两层空洞卷积进行密集连接后，最大感受野的计算公式如式(3-3)所示：

$$R_D = R_1 + R_2 - 1 \quad (3-3)$$

与 ASPP 保持相同膨胀率的 DASPP 的最大感受野如式(3-4)所示：

$$R_{Dmax} = R_1 + R_2 + R_3 - 2 = 73 \quad (3-4)$$

式中  $R_1$ 、 $R_2$ 、 $R_3$  分别表示膨胀率  $r$  为 6、12、18，卷积核大小为 3 时，所得密集空洞卷积的感受野大小。

由(3-2)式和(3-4)式对比可以知道，DASPP 具有更大的感受野，能够更全面地捕获蝴蝶图像特征信息，因此更有利于实现蝴蝶图像的特征提取。

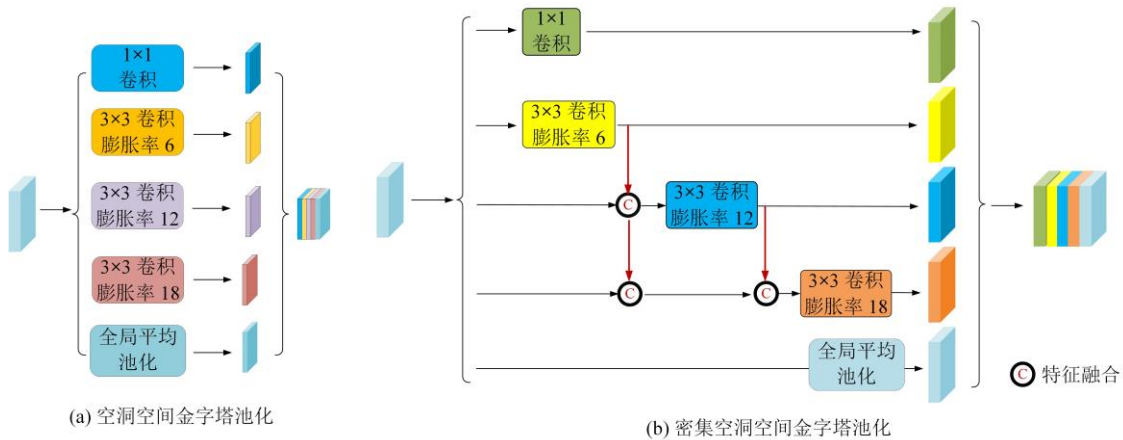


图 3-7 ASPP 与 DASPP 网络结构对比

图 3-8 所示为 ASPP 和 DASPP 在一维和二维下的采样对比图，此处是以膨

膨胀率  $r$  为 12 时进行的分析。图 3-8(a)和图 3-8(b)为一维情况下的采样情况，图 3-8(c)为二维的采样情况。图 3-8(a)为膨胀率为 12，卷积核大小为 3 的空洞卷积，其感受野为 25，但是采样到的像素点仅有 3 个，即参与运算的像素点数只有 3 个。图 3-8(b)为膨胀率为 12，卷积核大小为 3 的密集空洞卷积，该层密集空洞卷积有来自上一层，即膨胀率为 6 的那一层的采样输出，所以膨胀率为 12 的密集空洞卷积层中参与运算的采样点个数增加到了 7 个。此外，图 3-8(c)为 ASPP 与 DASPP 在二维下的采样图，可以看出，DASPP 中参与运算的点数明显比原始 ASPP 多，由 9 个变成了 49 个。由此参与计算的采样点变得更加密集，解决了因 ASPP 采样稀疏带来的信息丢失问题。

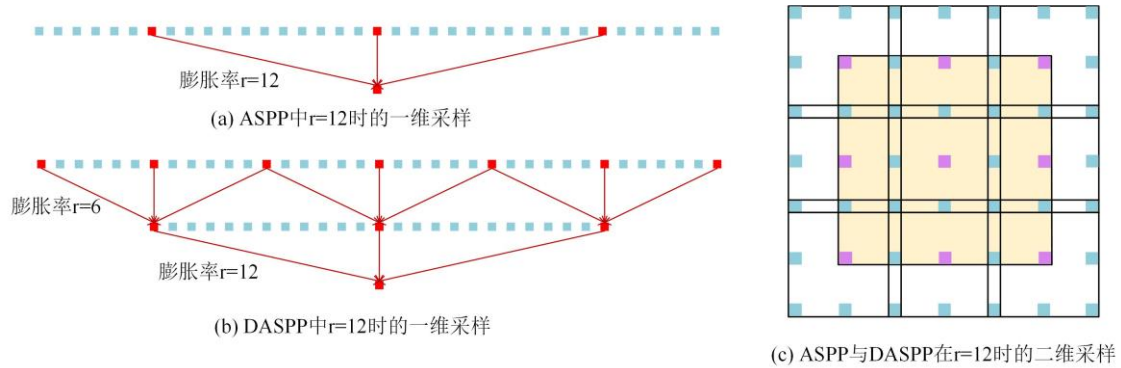


图 3-8  $r=12$  时 ASPP 与 DASPP 采样点密集性对比

### 3.2.4 引入置换注意力模块

2021 年杨育彬等人提出了置换注意力(Shuffle Attention, SA)<sup>[77]</sup>，该注意力网络通过特征分组和通道置换，实现了通道和空间注意力的融合，从而在不牺牲模型参数量与计算量的前提下，提升了网络性能。首先置换注意力将输入特征  $X \in R^{C \times H \times W}$  拆分成多组( $X = [X_1, \dots, X_G]$ )，其中  $X_k \in R^{C/G \times H \times W}$ ；然后对每组特征进行并行处理，其中每组信息又从通道和空间两个维度分别进行处理；最后将各组特征进行融合。对于通道注意力，输入为  $X_{k1} \in R^{C/2G \times H \times W}$ ，首先通过全局平均池化得到通道维度信息  $s$ ，对  $H \times W$  维度得到的信息  $s$  表达式如式(3-5)所示：

$$s = F_{gp}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{k1}(i, j) \quad (3-5)^{[77]}$$

然后经过  $F_c(s)$  增强特征，最后经过 sigmoid 处理后得到的结果与输入特征相乘，从而得到通道维度上的输出结果  $X'_{k1}$ ，该操作能够增强蝴蝶所在通道的特征表示能力。其计算表达式如式(3-6)所示：

$$X'_{k1} = \sigma(F_c(s)) \cdot X_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \quad (3-6)^{[77]}$$

式中  $W_1 \in R^{C/2G \times 1 \times 1}$ ;  $b_1 \in R^{C/2G \times 1 \times 1}$ 。

对于空间注意力，输入为  $X_{k2} \in R^{C/2G \times H \times W}$ ，首先使用组归一化(Group Normalization, GN)<sup>[78]</sup>获得空间维度信息，然后经过  $F_c(\cdot)$ 对空间特征进行增强。接着使用 sigmoid 处理，sigmoid 激活函数的图像如图 3-9 所示，计算公式如式 3-7 所示：

$$f(x) = \frac{1}{1+e^{-x}} \quad (3-7)$$

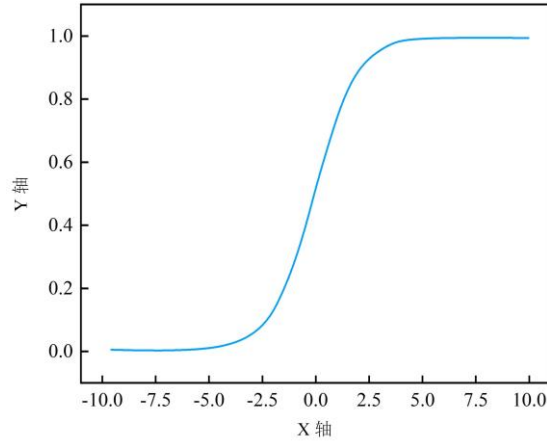


图 3-9 Sigmoid 激活函数

最后将经过 sigmoid 函数归一化后的结果与输入特征相乘，得到空间维度上的输出结果  $X'_{k2}$ 。其计算表达式如式(3-8)所示：

$$X'_{k2} = \sigma(W_2 \cdot GN(X_{k2}) + b_2) \cdot X_{k2} \quad (3-8)^{[77]}$$

式中  $W_2 \in R^{C/2G \times 1 \times 1}$ ;  $b_2 \in R^{C/2G \times 1 \times 1}$ 。经过空间注意力操作后，网络会增加蝴蝶所在区域的特征权重。

将通道维度和空间维度的输出结果  $X'_{k1}$  和  $X'_{k2}$  进行拼接(concat)操作，得到  $X'_k \in R^{C/G \times H \times W}$ ，最后将所有组的结果聚合到一起，采用通道置换操作，以获得各通道间的跨组信息。置换注意力的网络结构如图 3-10 所示。

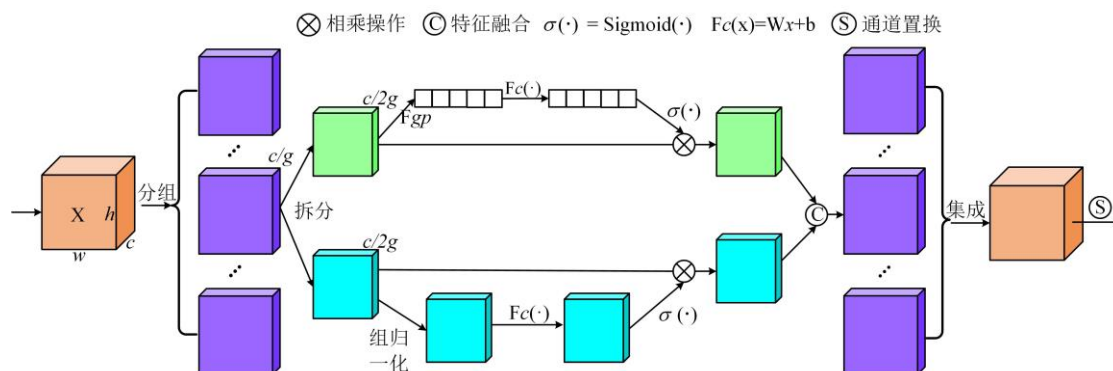


图 3-10 置换注意力网络结构<sup>[77]</sup>

### 3.3 实验结果与分析

#### 3.3.1 实验环境

本章基于PyTorch深度学习框架，使用Python语言实现，实验训练与验证均在同一平台上进行。实验环境配置为：CPU为Intel(R) Core(TM) i7-11700KF @ 3.60GHZ，内存为32GB；GPU为NVIDIA GeForce GTX 3080 12GB。所采用的软件工具为：PyTorch版本为1.7.1；Python版本为3.8。

#### 3.3.2 参数设置

根据实验设备性能，将批量大小(Batch size)设置为16，优化器选用的是随机梯度下降算法(Stochastic Gradient Descent, SGD)<sup>[79]</sup>，使用的损失函数是交叉熵损失函数(Cross Entropy Loss)<sup>[80]</sup>，初始学习率设置为0.01，其后每10个轮次学习率减半一次，总轮次(Epoch)设置为100，动量(Momentum)为0.9，为了防止出现过拟合现象，增加了权重衰减(Weight decay)因子，并将其值设置为0.0005。

#### 3.3.3 评价指标

在图像分割领域，分割精度、像素准确率(Pixel Accuracy, PA)、精确率(Precision)、召回率(Recall)和 F1 值(F1-score)是常见的评价指标。

其中，平均分割精度是图像分割领域最常用的评价指标，本文旨在将蝴蝶图像中的蝴蝶与背景分开，所以图像分割中只有背景和蝴蝶两个类别。平均分割精度反映的是预测结果与真实结果在像素级别上的匹配程度，具体地，蝴蝶图像平

均分割精度指的是蝴蝶分割精度与背景分割精度的平均值,其计算公式如式(3-9)所示:

$$mIoU = \frac{IoU_{\text{正例}} + IoU_{\text{反例}}}{2} = \frac{\frac{TP}{TP+FP+FN} + \frac{TN}{TN+FN+FP}}{2} \quad (3-9)$$

像素准确率是指模型预测正确的像素点数目占全部像素点数的比例,其计算公式如式(3-10)所示:

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \quad (3-10)$$

精确率是指模型预测为蝴蝶的像素中,正确预测为蝴蝶的像素点数目占全部预测为蝴蝶的像素点的比例。召回率是实际为蝴蝶的像素中,正确预测为蝴蝶的像素点数目占实际总蝴蝶像素点的比例。F1 值是精确率和召回率的调和平均数,所以选取 F1 值作为本章的另外一个评价指标。精确率、召回率和 F1 值的计算公式如式(3-11)、(3-12)和(3-13)所示。

$$Precision = \frac{TP}{TP+FP} \quad (3-11)$$

$$Recall = \frac{TP}{TP+FN} \quad (3-12)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3-13)$$

式中  $TP$ 、 $TN$ 、 $FP$  和  $FN$  分别表示实际为蝴蝶预测也为蝴蝶、实际为背景预测也为背景、实际为背景但预测成蝴蝶和实际为蝴蝶但是预测成背景的像素点数量。

本章选取平均分割精度、像素准确率和 F1 值作为蝴蝶图像分割任务的评价指标,从多个方面反应模型的分割效果。

### 3.3.4 结果分析

#### 1. 不同主干特征提取网络对比分析

为了验证所提方法的有效性,本章以 DeepLabv3+为基础网络,通过替换主干网络、改进空洞空间金字塔池化模块以及引入置换注意力模块进行了实验。

首先使用 EfficientNetV2-S 替换 Xception 主干特征提取网络,以提升模型分割精度。为了验证该方法的有效性,在实验参数设置和实验环境均保持一致的情况下进行了如下验证实验,其结果如表 3-3 所示。

表 3-3 不同主干特征提取网络对比

Model	Pixel Accuracy/%	mIoU/%	F1-score/%
Xception	98.16	95.25	96.27
EfficientNetV2-S	98.35	95.58	96.54

由表 3-3 可知, 使用 EfficientNetV2-S 网络替换原始 DeepLabv3+网络中的 Xception 后, 模型中像素点的准确率达到 98.35%, 平均分割精度 mIoU 值达到 95.58%, F1-score 达到 96.54%。与原始 DeepLabv3+网络相比, 替换主干网络后的模型具有更好的平均分割精度和像素准确率, 实验结果表明, 使用 EfficientNetV2-S 网络作为 DeepLabv3+网络的主干特征提取网络后, 蝴蝶图像分割效果更好, 不仅能提升蝴蝶图像的平均分割精度, 还可以提升像素点的准确率和 F1 值。故本章选取 EfficientNetV2-S 作为蝴蝶图像分割模型的主干特征提取网络。

## 2. 消融实验

将密集空洞空间金字塔池化和置换注意力模块分别引入 DeepLabv3+网络, 以探讨他们对网络性能的影响, 实验结果如表 3-4 所示。其中, 密集空洞空间金字塔池化作为加强特征提取网络, 提取蝴蝶高维语义信息; 置换注意力则对经过密集空洞空间金字塔池化模块后的特征, 进行更进一步的提取, 而且还会增加模型参数量。此处 DeepLabv3+网络的主干特征提取网络是 EfficientNetV2-S, 其中 DASPP 表示密集空洞空间金字塔池化, SA 表示置换注意力模块。

表 3-4 消融实验

Model	EfficientNetV2-S	DASPP	SA	Pixel Accuracy/%	mIoU/%	F1-score/%
1	√	—	—	98.35	95.58	96.54
2	√	√	—	98.43	95.94	96.86
3	√	√	√	98.60	96.19	97.05

√表示使用该方法, —表示没有使用相应方法

由表 3-4 可知, 在空洞空间金字塔池化网络中引入密集连接后, 模型分割精度较原始空洞空间金字塔池化提升了 0.36%, 这表明密集空洞空间金字塔池化使网络的感受野进一步增大, 参与计算的采样点更加密集, 提取到的蝴蝶特征更加丰富, 改进后的网络能够捕获到更多的蝴蝶本身信息, 网络提取蝴蝶特征能力更



强。在引入置换注意力模块后，网络的平均分割精度较单独引入密集空洞空间金字塔池化网络更高，这说明置换注意力机制通过特征分组和通道置换的方式，增强了蝴蝶所在通道和所在区域的特征表示能力，进一步加强了网络对蝴蝶特征的提取能力。

为了体现消融实验效果，图 3-11 以 4 类蝴蝶生态照为研究对象，列出了消融实验的可视化结果。其中组 1 表示的是：使用 EfficientNetV2-S 替换原始 DeepLabv3+主干网络后的实验可视化结果。组 2 是在组 1 的基础上，在 ASPP 网络中引入密集连接后的可视化结果。与组 1 相比，引入密集连接思想后，网络的加强特征提取能力更强，网络感受野更大，采样点利用率更丰富，因此整体来看，组 2 的分割效果优于组 1。组 3 是在组 2 的基础上，引入了置换注意力机制，通过特征分组和通道置换，蝴蝶所在区域和所在空间的权重信息会增加，由此获得的分割图像更加精准。

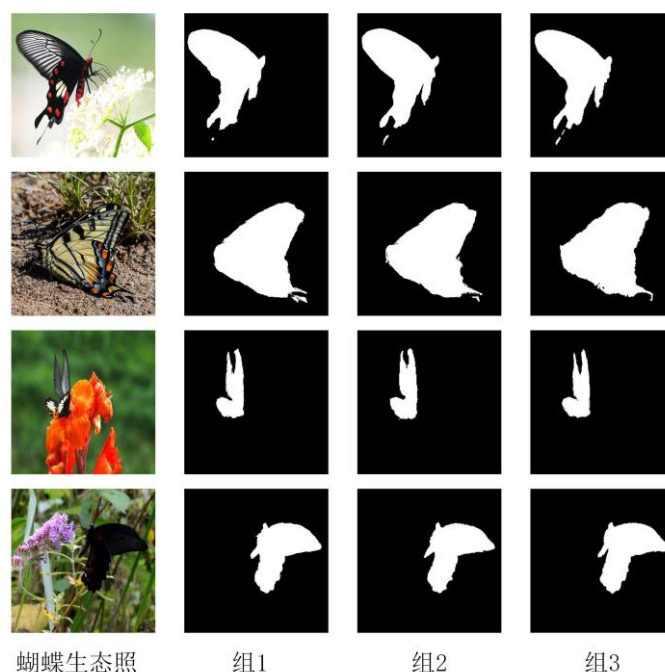


图 3-11 消融实验可视化结果

### 3. 模型对比实验

为了验证本文提出的 DS-Net 模型的有效性，本章选取主流的分割网络 PSPNet、SegNet、HRNet 和 DeepLabv3+模型与本文模型进行对比实验。实验在相同环境和参数设置下进行，结果如表 3-5 所示。

由表 3-5 可知, DS-Net 网络的平均分割精度 mIoU 最高, 达到了 96.19%, 分别比 PSPNet、SegNet、HRNet 和 DeepLabv3+高出 2.9%、1.91%、1.57%和 0.94%。这表明, DS-Net 网络对自然环境下蝴蝶图像中提取到的蝴蝶特征最丰富, 分割效果最好; 此外, 在像素点准确率和 F1-score 两个评价指标上, DS-Net 网络也取得了最好的效果。综上所述, 在对比网络中, DS-Net 网络得到了最佳的平均分割精度和最好的像素准确率, 即 DS-Net 网络对蝴蝶图像的分割效果最优, 与标签图像的匹配程度最高。

表 3-5 分割模型对比实验

Model	Pixel Accuracy/%	F1-score/%	mIoU/%
PSPNet	97.43	94.67	93.29
SegNet	97.81	95.51	94.28
HRNet	97.95	95.78	94.62
DeepLabv3+	98.16	96.27	95.25
DS-Net	<b>98.60</b>	<b>97.05</b>	<b>96.19</b>

图 3-12 为 PSPNet、SegNet、HRNet、DeepLabv3+和 DS-Net 五种分割网络的分割效果对比图。从左到右依次为原图图像, 标签图像, 经过 PSPNet 网络、SegNet 网络、HRNet 网络、DeepLabv3+网络和 DS-Net 网络后的分割图像。从图 3-12 可知, PSPNet 网络的分割效果最差, 因为在蝴蝶颜色和背景相似的图像中, 背景对 PSPNet 网络的分割效果产生了很大干扰, 所以无法得到较为完整的蝴蝶分割图像; SegNet 网络和 HRNet 网络的分割效果略优于 PSPNet 网络, 但是它们对边缘信息的处理能力仍不够; DeepLabv3+网络的分割效果优于前三个网络, 但得到的分割图像与真实标签图像之间仍存在一定差距; DS-Net 网络在所有对比网络中取得了最佳的分割效果, 能够更好地处理蝴蝶边缘信息, 生成的分割图像也几乎能够接近标签图像。这是因为 DS-Net 的主干特征提取网络能力得到了加强, 同时, 在加强特征提取网络中引入了密集连接思想, 实现了加强特征提取网络间前边层信息与当前层信息的叠加, 这样采样点更密集, 采样到的蝴蝶信息更全, 同时网络感受野增大, 网络能够观察到大型蝴蝶的全局信息; 此外, 本章还引入了置换注意力网络, 以进一步提取蝴蝶所在通道和区域的特征, 而且该操

作不会增加模型的参数量。上述结果表明，DS-Net 网络能够更有效地分割复杂背景下的蝴蝶图像进行。

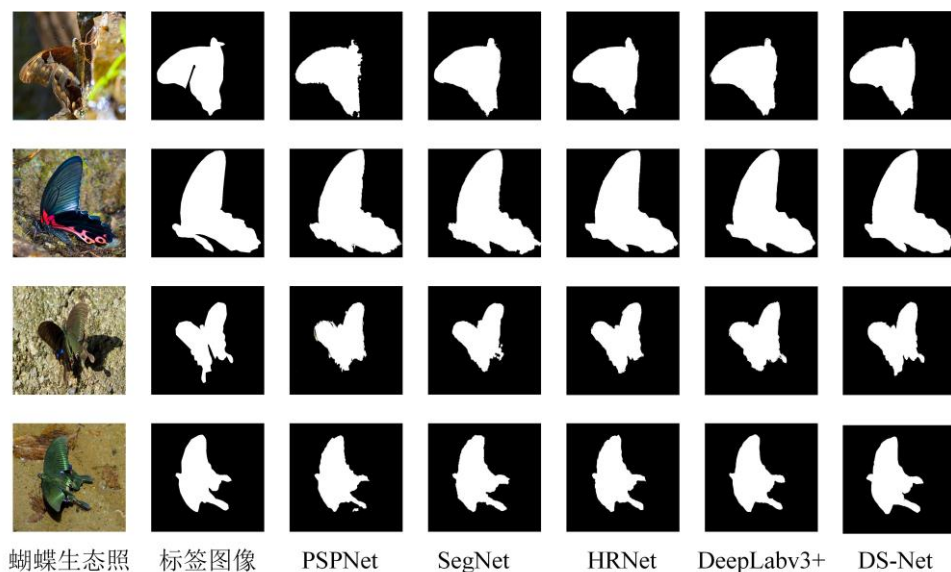


图 3-12 不同分割网络效果图比较

### 3.4 本章小结

本章对原始 DeepLabv3+网络结构进行了分析，同时从自然环境下蝴蝶图像特征入手，对 DeepLabv3+网络进行了改进，提出了一种融合置换注意力机制与密集空洞空间金字塔池化的 DeepLabv3+蝴蝶图像分割模型，达到了去除蝴蝶图像复杂背景的目的。实验结果表明，所提出的 DS-Net 网络可以提取到更加丰富的蝴蝶图像特征，该网络既有高分割精度又有高分割准确率，能够对复杂背景下的蝴蝶图像进行准确的自动分割。具体来说，首先使用 EfficientNetV2-S 替换原始 DeepLabv3+的主干特征提取网络 Xception，以增强主干网络特征提取能力。接着对加强特征提取网络进行改进，将多层级联的思想引入到空洞空间金字塔池化，得到密集空洞空间金字塔池化模块，该模块不仅可以使网络感受野进一步增大，还能使参与运算的采样点变得更加密集，从而提取到更加丰富的蝴蝶特征。最后为了提高蝴蝶所在通道和区域的特征表达能力，在密集空洞空间金字塔池化模块后引入轻量级置换注意力机制，以此来提高网络性能。在与其他主流分割网络对比时，DS-Net 方法具有更高的平均分割精度和更好的分割准确率。能够对 32 类蝴蝶图像进行自动分割，图像中像素准确率达到了 98.60%，平均分割精度达到了 96.19%。

## 第四章 基于残差网络和可变形卷积的蝴蝶种类识别模型

蝴蝶赖以生存的自然环境非常复杂,该因素对蝴蝶种类识别效果的影响较大。通过第三章的分析与实验,得到了去除背景噪声的蝴蝶分割图像,本章将对这些处理过的蝴蝶图像种类进行识别。为了提升蝴蝶种类识别效果,本章从蝴蝶的自身特征出发,提出了一种基于残差网络的轻量级可变形卷积的蝴蝶种类自动识别方法。该方法以 ResNet-18 网络为基础,引入可变形卷积改变原始普通卷积的固定采样位置,增大网络感受野,网络可以学习到更加丰富的蝴蝶图像特征,进而提升识别效果;然后从降低模型参数量的角度出发,对深度可分离卷积进行了改进;最后考虑到蝴蝶种类数量之间的分布不均衡现象,使用 PolyLoss-Focal Loss<sup>[81]</sup>进行训练,最终构建了轻量级可变形残差网络(Lightweight Deformable ResNet, LD-ResNet)。实验结果表明,该网络在蝴蝶种类识别方面表现最佳,模型参数量最小,可为蝴蝶种类自动识别研究提供一种高效且准确的解决方案。此外,LD-ResNet 网络采用轻量化的网络结构和有效的特征提取方法,具有较强的通用性和可扩展性,可为其他昆虫种类的自动识别研究提供参考。

### 4.1 蝴蝶图像识别数据集构建

本章所用的蝴蝶图像数据集与第三章所用数据集图像保持一致,只是本章数据集是第三章去除了蝴蝶图像中背景噪声的图像。蝴蝶图像识别部分的数据集构建主要有:数据集制作、数据集划分和数据图像增强。首先将蝴蝶的分割图像进行处理得到本章数据集;其次根据图像分类任务中数据集要求对蝴蝶图像进行数据集划分;最后为了增强模型的鲁棒性,对数据集进行数据增强操作。

#### 4.1.1 数据集制作

为了得到去除背景噪声的蝴蝶彩色图像,需要使用 Python 的 OpenCV,将经过分割网络后的图像叠加在原图上进行处理。考虑到蝴蝶种类繁多,部分蝴蝶颜色较深,所以本文将去除了噪声的所有蝴蝶图像背景统一转换成白色,处理后的蝴蝶图像识别部分数据集的示例如图 4-1 所示。蝴蝶图像识别部分的图像尺寸与

分割部分一致，仍然为  $224 \times 224$ ，且图像后缀以 .jpg 格式储存。按照蝴蝶类别的拉丁文名创建文件夹，然后将对应的蝴蝶图像放入其中，每个文件夹中的图像命名按照阿拉伯数字顺序命名。



图 4-1 蝴蝶图像识别数据集示例

#### 4.1.2 数据集划分

在 4.1.1 中，已经得到了处理后的 32 类蝴蝶种类识别数据集。为了得到训练和验证的蝴蝶图像，首先创建了 `train` 和 `val` 两个文件夹，然后在这两个文件夹下新建了以 32 类蝴蝶种类拉丁文命名的文件夹，最后按照 7:3 的比例将原始的 7825 张蝴蝶识别图像数据集划分到训练集和测试集中。

#### 4.1.3 数据增强

为了增强蝴蝶图像识别模型的泛化能力，使用 3.1.3 节中的数据增强技术，对蝴蝶图像识别数据集的训练集进行数据增强。

### 4.2 蝴蝶种类识别模型构建

本章构建了一种轻量级的可变形残差网络(LD-ResNet)，用于蝴蝶图像的自动识别。首先从蝴蝶特征入手，引入可变形卷积增大网络感受野；其次借鉴深度可分离卷积的思想，提出了改进的深度可分离卷积，与原始深度可分离卷积相比，改进后的深度可分离卷积可以更大程度地减小模型参数量与运算量；最后考虑到蝴蝶种类数量之间分布不均的情况，引入 PolyLoss-Focal Loss 损失函数进行训练。所用方法是从蝴蝶自身特征和模型轻量化角度出发，通过理论分析和实验验证，证明了所提方法的有效性。

### 4.2.1 蝴蝶种类识别总体模型结构

本章提出的蝴蝶图像识别网络以 ResNet-18 为基础，根据蝴蝶形状和大小的不同，使用可变形卷积替换 ResNet-18 网络中的倒数第 1 个和倒数第 2 个卷积块中的普通卷积，以提取到蝴蝶更加完整的信息；并且为了更大程度降低模型的参数量，提出了改进的深度可分离卷积，改进后的蝴蝶图像识别模型总体结构如图 4-2 所示。由图 4-2 知，模型的总体结构由特征提取网络(Feature Extraction Network)和全局平均池化(Globe Average Pool, GAP)组成，使用改进的深度可分离卷积替换原始 ResNet-18 网络中卷积块 2 和卷积块 3 的普通卷积，并在卷积块 4 和卷积块 5 中用可变形卷积和改进的深度可分离卷积替换原始 ResNet-18 网络中的普通卷积；最后将输出特征图送入全局平均池化层处理。

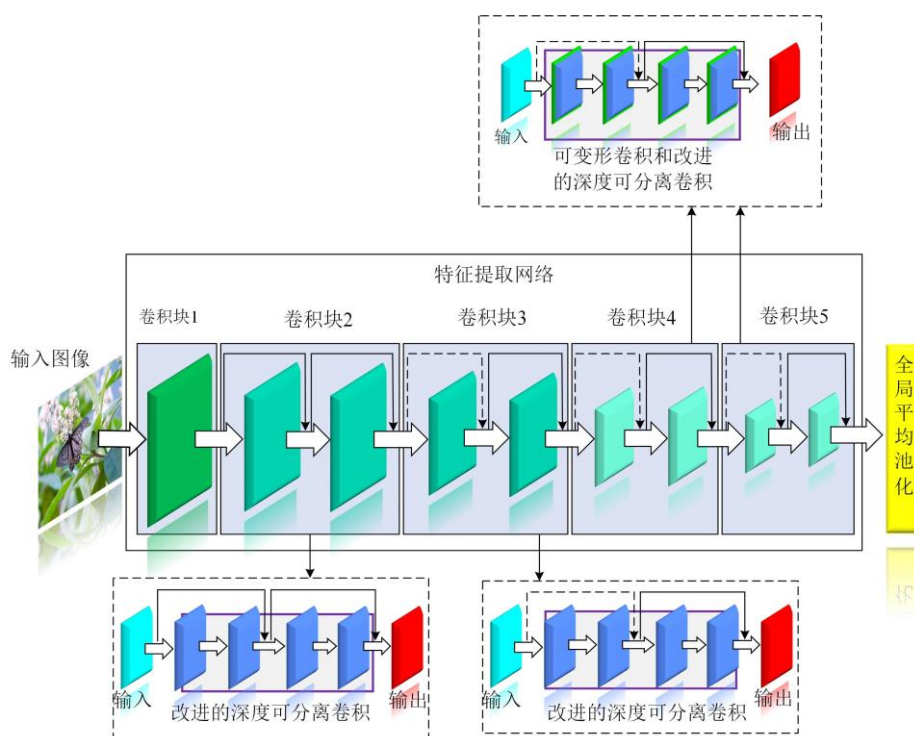


图 4-2 LD-ResNet 模型总体结构

### 4.2.2 引入 PolyLoss-Focal Loss 损失函数

在蝴蝶图像识别数据集中，蝴蝶各个种类之间的数量分布非常不平衡。其中钩粉蝶的数量最多，高达 907 张，而小黑斑凤蝶数量最少，仅有 77 张，数量最多的种类竟然是最少种类的 12 倍。这种数量分布上的极不平衡现象，对模型的

整体识别率产生较大的影响。对于数量较多的种类，网络能够对它们的特征进行充分地提取，这样提取到的特征就会非常丰富，那么网络对这种数据量丰富的蝴蝶就会有很好的识别效果；而对数量较少的种类，因为数据量较小，网络对这种数据量小的蝴蝶种类特征的学习能力有限，提取到的特征也就不那么充分，这使得模型对这种数据量小的蝴蝶种类的识别率低，进而影响整体识别率。为了解决这个问题，本章使用 PolyLoss 损失函数来替代交叉熵损失函数，该方法是 2022 年提出来的，发表在深度学习领域的顶级会议上。该函数有 PolyLoss-Cross Entropy 和 PolyLoss-Focal Loss 两种形式，结合数据集分布情况，本文选取 PolyLoss-Focal Loss 作为本章的损失函数进行实验。PolyLoss-Focal Loss 损失函数是基于焦点损失函数 Focal Loss<sup>[82]</sup>提出的。虽然 Focal Loss 损失函数可以解决数据样本中的分布不均衡问题，但是对于不同数据集，效果会有所不同，而且使用 Focal Loss 损失函数进行训练时，为了得到较好的训练效果，需要不断地调整参数因子。鉴于此，本章使用 PolyLoss-Focal Loss 损失函数进行实验。

PolyLoss-Focal Loss 损失函数是将 Focal Loss 损失函数通过泰勒公式展开的，展开后会得到一系列的多项式函数，展开的形式如式(4-1)所示。使用焦点损失函数进行实验时，需要对两个可调因子进行多次实验才能得出与蝴蝶图像数据集匹配的可调参数，而对于 PolyLoss-Focal Loss 损失函数，仅需要调整第 1 个多项式的系数即可。使用 PolyLoss-Focal Loss 损失函数进行训练时，它基于多项式来对样本的权重进行拟合，使得模型更加关注蝴蝶种类较少的样本，进而从整体上提升蝴蝶种类识别的效果。这种方法能够有效地解决数据样本分布不均衡问题，而且更加稳定，不需要频繁调整参数因子。

$$\sum_{j=1}^{+\infty} \alpha_j (1 - P_t)^j \quad (4-1)^{[81]}$$

式中  $\alpha_j$  表示多项式损失权重； $P_t$  表示目标标签的预测概率。

#### 4.2.3 引入可变形卷积模块

本文的蝴蝶图像数据集都是自然环境下的蝴蝶图像，所以它们都是拍摄所得。在对蝴蝶进行拍摄时，容易受到成像视角、拍摄距离、当地天气变化和蝴蝶状态等因素的影响，这些因素的存在，会使蝴蝶在图像中的占比不一样，同时蝴蝶呈现出来的形状也会不一样。当拍摄距离近时，图像中蝴蝶占据的比例就大；当拍



摄距离远时，图像中蝴蝶的占比就小；同时蝴蝶飞行时会展开翅膀，而休息时则会合上，所以拍摄得到的形状也会不一样；另外，阴天拍摄的蝴蝶图像和晴天拍摄的蝴蝶图像的亮度也不同。ResNet-18 网络中除了第一个卷积层，其余部分均使用的是  $3 \times 3$  的普通卷积，普通卷积中的采样方式是对输入特征图的固定位置进行采样，导致网络对大小不一的蝴蝶处理能力有限，对于较大的蝴蝶，很难采样到其完整特征，影响蝴蝶种类的识别效果。鉴于此，本章将原始 ResNet-18 网络的倒数第 1 个和倒数第 2 个残差结构中的普通卷积替换为可变形卷积 (Deformable Convolution, DC)<sup>[83]</sup>。与标准卷积的固定几何结构采样方式不同，可变形卷积通过在每个卷积核的采样点上添加一个偏移量，使得网络能够根据可变偏移量，以灵活的几何结构对输入特征图进行采样，使采样点排列由原来的固定位置变得不规则，由此感受野的范围得到改变，网络拥有的自适应能力由输入自适应偏移量控制，可以根据蝴蝶图像数据集中蝴蝶的形状和大小来调整感受野，进而获取到更多的有效采样点信息。这样网络对大小不一样和形状不同的蝴蝶都有很好的特征提取能力。可变形卷积的结构如图 4-3 所示。

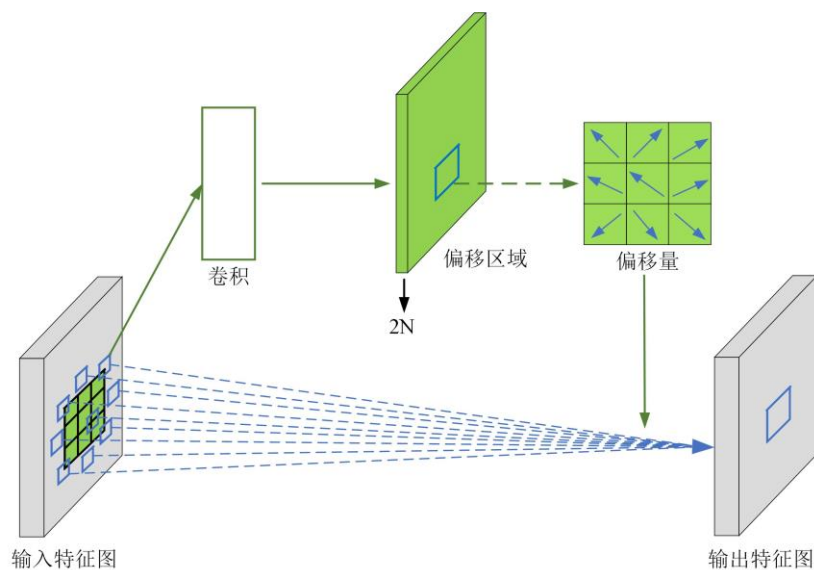


图 4-3 可变形卷积结构<sup>[84]</sup>

以卷积核大小为  $3 \times 3$  的卷积为例，对于普通卷积，其输出特征图中的任意一点  $P_0$ ，用卷积核  $w$  对网格  $R(R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\})$  中的采样点进行加权计算，公式表示如式(4-2)所示：

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (4-2)^{[83]}$$



对于可变形卷积, 因为有偏移量 $\Delta p_n$ 的引入, 所以采样点的排列就由原来的固定位置变得不规则, 能够对特征图进行灵活采样, 公式也就变成了(4-3):

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (4-3)^{[83]}$$

式中 $\Delta p_n$ 为采样点的偏移量。

#### 4.2.4 改进深度可分离卷积

深度可分离卷积(Depthwise Separable Convolution, DSC)模块由两个部分组成, 一个部分是深度卷积(Depthwise Convolution), 另一个部分是点卷积(Pointwise Convolution)。深度卷积也叫逐通道卷积, 该部分是对通道维度进行操作, 深度卷积中一个卷积核会处理一个通道的信息, 当深度卷积操作完成后, 得到的特征图数量与输入的通道数量保持一致。因为该操作只是对通道进行处理, 得到的特征图没有包含跨通道的信息, 所以还需要将深度卷积后的特征图进行点卷积处理。逐点卷积和普通卷积一样, 只是逐点卷积的卷积核大小为  $1 \times 1$ , 它对不同通道之间的信息进行卷积操作, 实现了跨通道信息的融合。与普通卷积相比, 深度可分离卷积能较大幅度减小模型的参数量与运算量。

本章借鉴该想法对原始的深度可分离卷积进行改进, 提出了改进的深度可分离卷积(Improved Depthwise Separable Convolution, IDSC), 相较于原始的深度可分离卷积, 改进后的深度可分离卷积能够更大程度地减小模型的运算量与参数量。通过 2.3.4 节中对深度可分离卷积的分析可知, 深度可分离卷积中的参数量与运算量主要来自  $1 \times 1$  卷积。对此, 本章将深度可分离卷积的原始逐点卷积进行了改进, 仅使用一半的通道进行逐点卷积, 另外一半使用深度卷积进行处理, 最后在通道维度进行拼接, 得到改进后的深度可分离卷积。改进后的深度可分离卷积由两次深度卷积和一次逐点卷积组成, 具体过程是: 与深度可分离卷积一样, 输入特征图  $F_1$  首先处理通道维度上的信息, 即先进行第一次深度卷积得到特征图  $F_2$ ; 与原始深度可分离卷积不同的是, 改进的深度可分离卷积将原始的  $C_{out}$  分成两个  $\frac{C_{out}}{2}$  进行两次卷积。使用  $1 \times 1 \times \frac{C_{out}}{2}$  进行点卷积操作, 得到了特征图  $F_3$ ; 然后特征图  $F_3$  会进行第二次深度卷积得到  $F_4$ , 只是第二次的深度卷积中的通道数为  $\frac{C_{out}}{2}$ 。最后将特征图  $F_3$  与特征图  $F_4$  进行拼接操作, 得到最后的输出特征图  $F$ 。

原始的深度可分离卷积与改进后的深度可分离卷积的结构对比图如图 4-4 所示。

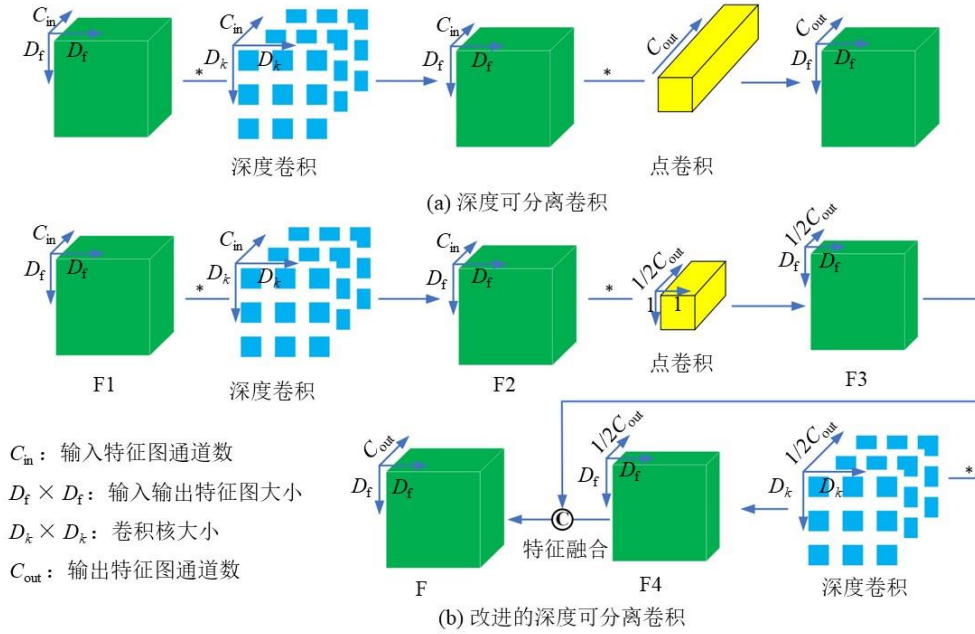


图 4-4 深度可分离卷积与改进后的深度可分离卷积结构对比

假设输入特征图  $F_1$  的大小为  $D_f \times D_f \times C_{in}$ ，卷积核大小为  $D_k \times D_k$ ，输出特征图  $F$  的大小为  $D_f \times D_f \times C_{out}$ ，对于深度可分离卷积，它的参数量计算公式如式(4-4)所示：

$$x_1 = D_k \times D_k \times C_{in} + C_{in} \times C_{out} \quad (4-4)$$

而对于改进后的深度可分离卷积，它的参数量计算公式如式(4-5)所示：

$$x_2 = D_k \times D_k \times C_{in} + C_{in} \times \frac{C_{out}}{2} + D_k \times D_k \times \frac{C_{out}}{2} \quad (4-5)$$

深度可分离卷积与改进后的深度可分离卷积在参数量上的差值计算公式如式(4-6)所示：

$$x_1 - x_2 = \frac{C_{out}}{2} \times (C_{in} - D_k^2) \quad (4-6)$$

因为  $C_{out}$  肯定是大于零的数字，由式(4-5)可知，只要  $C_{in} > D_k^2$ ，也就是只要输入的通道数大于卷积核的平方时，改进后的深度可分离卷积的参数量就会比原始的深度可分离卷积要小。从表 2-2 可以看出，除了 ResNet-18 网络的第一个卷积块的输入通道数小于卷积核的平方，其余四个模块均满足  $C_{in} > D_k^2$  的条件。所以，本章将改进后的深度可分离卷积嵌入第二个至第五个残差块中，以此来减小模型的参数量，从而得到轻量级网络。

## 4.3 实验结果与分析

### 4.3.1 实验环境

蝴蝶种类识别部分的实验环境设置与 3.3.1 节保持一致。

### 4.3.2 参数设置

综合考虑实验设备性能和实验训练效果后,本章将批量大小(Batch Size)设置为 16,即每次处理 16 张蝴蝶图像,迭代次数(Epoch)设置为 200。在损失函数优化器方面,使用随机梯度下降算法(Stochastic Gradient Descent, SGD)作为损失函数的优化器,初始学习率设置为 0.01,学习率随着训练轮次变化,每 20 轮次衰减 0.5,同时动量因子设置为 0.9。

### 4.3.3 评价指标

在图像分类任务中,常用的评价标准包括准确率(Accuracy)、精确率(Precision)、召回率(Recall)和F1值(F1-score)等。

#### (1) 平均识别准确率

蝴蝶图像的平均识别准确率(Average Recognition Accuracy, ARA)是指验证集中蝴蝶种类分类正确的数量与验证集中所有蝴蝶总样本数量的比值,其计算方法如式(4-7)所示:

$$ARA = \frac{1}{N} \sum_{i=1}^N \frac{X_i}{Y_i} \times 100\% \quad (4-7)$$

式中N为蝴蝶图像识别数据集中的数据类别总数,因为本文是对32种蝴蝶图像进行识别,所以这里N的值为32;  $Y_i$ 为第*i*类蝴蝶图像总数量;  $X_i$ 为第*i*类蝴蝶图像中分类正确的数量。

#### (2) 精确率

精确率(Precision)指的是预测为某一类的样本中,实际为该类的样本数量与全部预测为该类的数量的比值,这一评价指标也叫“查准率”,因为它是针对正例的预测结果。精确率的计算公式如式(4-8)所示。

$$P = \frac{TP}{TP+FP} \quad (4-8)$$

式中 $TP$ 表示预测类别为正例，实际类别也为正例的样本数； $FP$ 表示预测类别为正例，实际类别为负例的样本数。

### (3) 召回率

召回率（Recall）指的是实际属于某个类别的样本中，被正确预测为该类的样本数量与总的实际属于该类别样本数量的比值，这一评价指标也叫“查全率”，因为它是以实际正样本为判断依据的一种评价指标，用于检查所有实际正样本中有多少是真正被预测为正样本。召回率的计算公式如式(4-9)所示。

$$R = \frac{TP}{TP+FN} \quad (4-9)$$

式中 $TP$ 表示预测类别为正例，实际类别也为正例的样本数； $FN$ 表示预测类别为负例，实际类别为正例的样本数。

### (4) F1值

在一般情况下，精确率和召回率会出现一个高一个低的状况，这时需要引入F1值来进行综合考虑。F1值是两者的调和平均值。F1值的特点是，在尽量提高精确率值和召回率值的同时，减小两者间的差异。F1值的计算公式如式(4-10)所示，通常F1值越高，代表模型的分类效果越好。

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (4-10)$$

式中 $P$ 为精确率； $R$ 为召回率。

### (5) 模型参数量

模型大小可以通过模型参数量(Parameters)表示出来，因为模型的参数量一般都比较大，所以常用单位MB来衡量，模型参数量主要来自卷积层和全连接层。一个模型的大小会直接由它的参数量反应出来，模型参数量还会影响对内存的占用。所以在保证模型识别准确率的前提下，参数量越小越好。

考虑到本章主要目的是对蝴蝶图像进行识别，故以蝴蝶种类的平均识别准确率为主要评价指标，同时为了得到轻量化模型，本章也会考虑模型参数量作为另外一个评价指标。综上所述，选取平均识别准确率、F1-score和模型参数量作为本章的评价标准。

## 4.3.4 结果分析

### 1. 蝴蝶图像分割必要性分析

在蝴蝶图像识别之前，本文首先对背景复杂的蝴蝶图像数据集进行了分割。为了验证该方法的必要性和有效性，本文进行了实验并使用了两类不同的蝴蝶图像数据集进行对比。这两类数据集在数量和种类上均保持一致，分别为最原始的蝴蝶图像数据集和仅去除了背景噪声干扰的蝴蝶图像数据集。实验的环境和参数设置也完全一致，实验均是在原始 ResNet-18 网络上进行的验证，得到的实验结果如表 4-1 所示。

表 4-1 去除背景噪声前后对比

数据集	Average Recognition	F1-score/%	Parameters/MB
	Accuracy/%		
最原始数据集	88.56	88.69	11.19
仅去除背景噪声数据集	92.23	92.35	11.19

由表 4-1 可知，在相同环境和参数设置下，使用去除了背景噪声干扰后的蝴蝶图像数据集进行实验时，所得到的结果要优于最原始数据集的结果，这表明，对蝴蝶种类进行识别之前，对图像进行分割是必要的，该操作能大幅度提升模型的平均识别准确率。

## 2. 数据增强必要性分析

为了验证数据增强方法对蝴蝶种类识别模型的必要性，以仅去除蝴蝶图像背景噪声的数据集为基础，将它的训练集部分进行数据增强，并以数据增强前后的蝴蝶图像数据集为研究对象，分别进行实验，实验均是在 ResNet-18 网络上测试的，得到的实验结果如表 4-2 所示。

表 4-2 数据增强前后对比

数据集	Average Recognition	F1-score/%	Parameters/MB
	Accuracy/%		
仅去除背景噪声数据集	92.23	92.35	11.19
数据增强后数据集	94.64	94.73	11.19

由表 4-2 可知，对仅去除蝴蝶图像背景噪声的数据集进行数据增强后，在 ResNet-18 网络的各项评价指标均有明显提升，平均识别准确率从原来的 92.23% 提高到了 94.64%。这表明，使用高斯噪声、椒盐噪声、左右翻转和增强亮度等数据增强方式对蝴蝶图像数据集进行数据增强后，可以进一步提高蝴蝶识别网络的

性能，增强模型的泛化能力。

### 3. 消融实验

为了验证本章所提方法的有效性，以去除蝴蝶背景噪声并进行了数据增强的蝴蝶图像数据集为实验对象，使用ResNet-18网络为基础模型，对可变形卷积和PolyLoss-Focal Loss损失函数的有效性进行验证，实验结果如表4-3所示。其中第1组实验的损失函数使用的是交叉熵损失函数(CrossEntropyLoss)，第2组实验和第3组实验使用的都是PolyLoss-Focal Loss损失函数。

表 4-3 以 ResNet-18 网络为基础的消融实验

Groups	PolyLoss-Focal Loss	Deformable convolution	Average Recognition
			Accuracy/%
1	—	—	94.64
2	√	—	95.53
3	√	√	97.15

√表示使用该方法，—表示没有使用相应方法

由表 4-3 可知，与使用 CrossEntropyLoss 损失函数相比，使用 PolyLoss-Focal Loss 损失函数进行实验后，网络的平均识别准确率提升了 0.89%，这表明使用 PolyLoss-Focal Loss 损失函数后，网络对蝴蝶种类数量较少的种类的识别能力增强，较好解决了种类数量间分布不平衡的问题；将 ResNet-18 网络的最后两个残差块中普通卷积替换成可变形卷积后，模型的平均识别准确率较单独引入 PolyLoss-Focal Loss 损失函数提升了 1.62%。这表明引入可变形卷积后网络的采样范围改变，网络感受野增大，提取蝴蝶全局特征能力更强，提取到的蝴蝶信息更丰富。

### 4. 改进的深度可分离卷积对模型参数量的影响

为了验证改进的深度可分离卷积对模型参数量的影响，本章以 ResNet-18 为基础模型，以去除蝴蝶背景噪声并进行了数据增强的蝴蝶图像数据集为实验对象。由 4.2.3 分析可知，只有当输入通道数大于卷积核的平方时，改进的深度可分离卷积才能从更大程度上降低模型参数量。因此，分别选用深度可分离卷积和改进的深度可分离卷积替换 ResNet-18 网络中第 2 个至第 5 个卷积块的普通卷积，实验结果如表 4-4 所示。

表 4-4 深度可分离卷积和改进后的深度可分离卷积的参数量对比

Model	Average Recognition Accuracy/%	Parameters/MB
ResNet-18	94.64	11.19
ResNet-18+DSC	93.73	1.46
ResNet-18+IDSC	93.72	0.88

由表 4-4 可知,与原始 ResNet-18 网络相比,引入改进的深度可分离卷积后,虽然模型的平均识别准确率有些许下降,但是模型参数量却大大降低。将深度可分离卷积与改进的深度可分离卷积进行比较,两者的平均识别准确率几乎没有区别,但是改进的深度可分离卷积可以从更大程度上降低模型参数量。使用改进的深度可分离卷积后,模型参数量大幅度减小,与原始 ResNet-18 模型相比,减小了 92.14%;与深度可分离卷积相比,减小了 39.73%。由此可得,改进的深度可分离卷积能够更有效地减小模型参数量,使网络更加轻量化。

#### 5. 不同识别模型之间的对比分析

为了验证本章模型(LD-ResNet)的有效性,本章选取主流识别网络 MobileNet、VGG16、ResNet-18、GoogleNet、ShuffleNet-V2 和 EfficientNet 模型与 LD-ResNet 模型进行对比。以去除蝴蝶背景噪声并进行了数据增强的蝴蝶图像数据集为实验对象,实验是在相同的环境和参数设置下进行的,结果如表 4-5 所示。

表 4-5 不同识别模型的对比

Model	Average Recognition Accuracy/%	F1-score/%	Parameters/MB
MobileNet	93.24	93.36	4.12
VGG16	94.19	94.31	134.43
ResNet-18	94.64	94.73	11.19
GoogleNet	95.12	95.17	5.60
ShuffleNet-V2	95.51	95.54	1.26
EfficientNet	96.47	96.49	5.34
LD-ResNet	96.23	96.26	1.04

虽然上述实验是以已去除背景噪声并进行数据增强后的蝴蝶图像识别数据集为研究对象的,但是由表 4-5 可知,在所有对比网络中,MoileNet 网络的识别效果仍然不够理想,这是因为 MoileNet 网络主要解决的是模型轻量级问题,而

该模型轻量化是以牺牲识别率为代价的；使用 VGG16 网路对蝴蝶图像识别时，其平均识别准确率优于 MoileNet 网络，但是该网络的参数量过大，无法满足实际应用；使用基准的 ResNet-18 网络进行蝴蝶种类识别时，平均识别准确率达到 94.64%；与前三种网络相比，GoogleNet 网络在平均识别准确率和参数量方面都表现出了较好效果，但仍不及 LD-ResNet；ShuffleNet-V2 网络提出了四条设计高效网络准则，根据这四条准则设计出了网络的核心基本单元和下采样单元，由此 ShuffleNet-V2 网络取得了较好的平均识别准确率和较小的模型参数量，但效果仍低于 LD-ResNet 网络；EfficientNet 网络凭借神经网络结构搜索技术，平衡了网络宽度、网络深度和分辨率，在本文数据集上取得了很好的平均识别准确率，但与 LD-ResNet 网络相比，其参数量过大，不适用于模型部署。

LD-ResNet 的平均识别准确率达到 96.23%，网络参数量仅为 1.04MB，是所有对比网络中模型参数量最小的网络，其平均识别准确率仅低于 EfficientNet 网络，但两者相差并不大，另外，LD-ResNet 网络的 F1-score 也比 EfficientNet 网络小一点点。综上所述，LD-ResNet 网络在蝴蝶种类识别方面取得了良好的识别效果，同时模型参数量最小。

## 6. 可视化对比结果分析

为进一步说明 LD-ResNet 网络的有效性和可行性，本文将基准模型和 LD-ResNet 模型的实验曲线结果进行了对比。ResNet-18 模型与 LD-ResNet 模型的平均识别准确率和损失值曲线如图 4-5 所示。

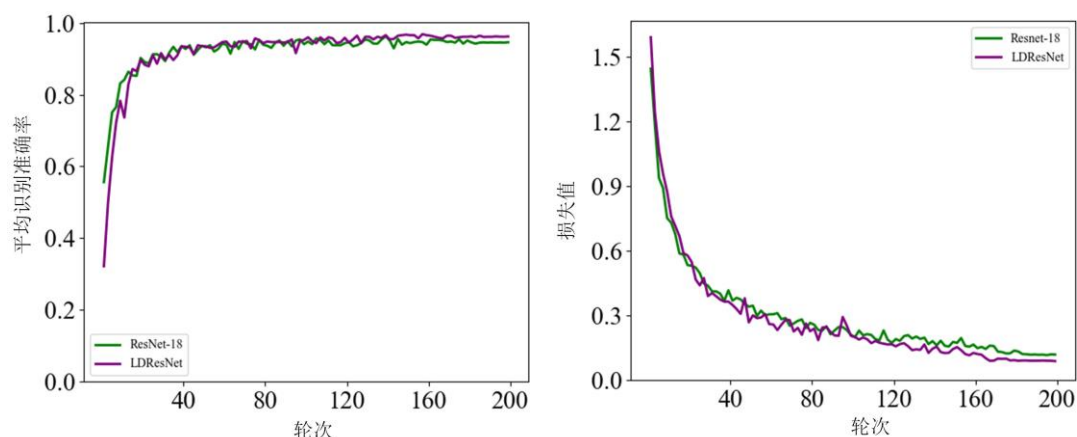


图 4-5 ResNet-18 模型与 LD-ResNet 模型的平均识别准确率和损失值



其中绿色线表示 ResNet-18 模型，紫色线表示 LD-ResNet 模型。图 4-5 所示结果是每 2 个轮次保留一次结果，并根据结果绘制曲线图。由图 4-5 可知，LD-ResNet 网络早于 ResNet-18 网络收敛，并且平均识别准确率高於基准模型，而损失值则低於基准模型。因此，与基准模型相比，LD-ResNet 模型具有更优秀的表示效果。

#### 4.4 本章小结

针对现有研究中蝴蝶种类识别存在的识别率低等问题，本章提出了一种基于残差网络的轻量级可变形蝴蝶种类自动识别模型，用来对蝴蝶种类进行识别。实验结果表明，在训练时，使用 PolyLoss-Focal Loss 损失函数，网络的平均识别准确率优于原始 ResNet-18 网络，这说明使用 PolyLoss-Focal Loss 进行训练时，数量较少的蝴蝶种类在损失函数中占据了更大的比重，从而提升了识别率；引入可变形卷积后，网络感受野增大，能更有效地提取较大蝴蝶图像的特征，从而全面保证了蝴蝶图像数据集中所有蝴蝶特征都能得到较好提取；改进的深度可分离卷积能更大幅度减小模型的参数量，使模型变得轻量化。在与其他主流识别网络进行对比时，LD-ResNet 网络具有最小的模型参数量，其平均识别准确率也仅低於 EfficientNet 网络，但两者相差并不大。该模型在对 32 类蝴蝶图像进行识别时，平均识别准确率达到 96.23%，模型的参数量仅为 1.04MB。

## 第五章 总结与展望

### 5.1 论文工作总结

蝴蝶作为昆虫的第二大目——鳞翅目，其种类丰富，数量繁多。蝴蝶的分布情况受环境、气候和温度的影响很大，所以它们在气候变化和环境检测等方面发挥着重要作用。然而，由于蝴蝶颜色、纹理和形状等特征复杂多样，如何高效准确地识别出它们的种类是一个急需解决的问题。在研究早期，对蝴蝶种类识别的研究都是依赖于昆虫学家，这不仅需要专业的设备，还需要昆虫学家具备丰富的识别经验和专业知识，这种识别方法费时费力，无法满足实际需求。近年来，随着计算机技术的发展，深度学习技术已经在图像识别任务上取得了显著成果，如医学图像分割、行人检测以及自动驾驶等领域。所以本文考虑使用深度学习技术解决自然环境中蝴蝶种类自动识别问题。鉴于自然环境中蝴蝶图像背景复杂，蝴蝶大小、颜色和纹理信息多样的问题，首先使用图像分割技术去除自然环境中蝴蝶的复杂背景，然后将去除了背景的蝴蝶图像送入分类网络进行识别，得到满足需求的蝴蝶图像识别模型。本文具体的研究内容如下所示：

1. 本文首先从自然环境下蝴蝶种类识别的研究背景与研究意义入手，对深度学习研究现状、图像分割研究现状和蝴蝶种类识别的国内外研究现状进行了分析；然后，详细介绍了卷积神经网络的概述和主要组成部分；此外，还简要概述了本文使用的图像分割算法，例如 SegNet、PSPNet、HRNet 和 DeepLabv3+；最后，介绍了图像识别算法，例如 MobileNet、VGG、GoogLeNet 和 ResNet。

2. 针对现有数据集中自然环境下蝴蝶种类和数量较少以及蝴蝶种类识别率低等问题，本文构建了 32 种共计 7825 张蝴蝶图像数据集。该数据集全部为自然环境下的蝴蝶图像，蝴蝶种类以常见的凤蝶科蝴蝶和粉蝶科蝴蝶为研究对象，具体种类是根据当地环境和官方数据竞赛确定的，其中大多数种类都是本地蝴蝶类别。

3. 针对自然环境中蝴蝶图像分割问题。首先制作了适用于蝴蝶图像分割模型的数据集，并进行了数据标注和扩充；然后根据自然环境中蝴蝶图像特征，构建了蝴蝶图像分割模型；最后从实验角度对结果进行了分析。自然环境中蝴蝶图像分割模型以 DeepLabv3+网络为基础模型，首先对主干网络进行了替换；其次

对空洞空间金字塔池化模块进行了改进；最后引入了置换注意力模块，构建了蝴蝶图像分割模型 DS-Net。为了提升蝴蝶分割精度，选取 EfficientNetV2-S 网络替换原始 DeepLabv3+ 中的 Xception 网络；为了使模型能够更好地提取到蝴蝶的本身像素信息，需要增大网络感受野，增加有效信息的采样点数，所以将密集连接思想引入到空洞空间金字塔池化模块中；最后为了增大蝴蝶信息权重，降低背景噪声干扰，将置换注意力模块引入到密集空洞空间金字塔池化模块后，并通过实验证实了该方法的有效性，且置换注意力机制不会增加网络的参数量。最后得到了自然环境下蝴蝶图像分割模型，该模型的平均分割精度 mIoU 达到了 96.19%，像素点的准确率达到 98.60%，实现了自然环境中蝴蝶图像与背景的有效分割，实验证明了本方法的有效性。

4. 针对蝴蝶图像识别问题，本文采用了多种方法进行优化。首先，将分割后的图像与原图进行叠加，得到了蝴蝶识别图像数据集；其次，构建了蝴蝶图像识别模型，用于蝴蝶种类识别；最后设计实验并对结果进行了分析。蝴蝶图像识别模型以 ResNet-18 网络为基础，首先针对数据集中蝴蝶种类数量上的分布不均衡状况，使用 PolyLoss-Focal Loss 损失函数进行训练，解决了因蝴蝶种类数量间差异大带来的问题；此外，针对蝴蝶图像中存在的大小不一和形状多样，以及原始网络中的普通卷积无法提取到蝴蝶图像中的全局信息的情况，引入可变形卷积对普通卷积进行替换，以此增大网络的感受野，增强网络的特征提取能力；最后，为了减小模型参数量，以深度可分离卷积的思想为依据，提出了改进的深度可分离卷积，实验结果表明，使用改进后的深度可分离卷积，模型的参数量能更大幅度减小。通过与其他识别网络的对比，本文方法取得了 96.23% 的平均识别准确率，模型的参数量仅 1.04MB，验证了该方法的有效性。得到的 LD-ResNet 网络能够准确识别出蝴蝶种类，可为蝴蝶图像识别提供一定的技术基础。

## 5.2 未来展望

本文主要使用深度学习技术来对自然环境下的蝴蝶种类进行了自动识别研究，从蝴蝶图像分割和图像识别两个部分入手，分别对分割模型和识别模型进行了优化和分析，并通过实验证明了本文方法的有效性。然而，由于个人能力和时间等因素的影响，本文还有需要改进的地方，主要包括：

1. 使用深度学习技术来解决图像处理任务时，数据量的充足与否对网络的识别能力有较大的影响，虽然本文蝴蝶图像数据集已经有 7825 张，但若想取得更好的识别效果，还需要更丰富的蝴蝶种类和数量。在后续研究中，应该多跟随昆虫研究所专家前往省内各地拍摄蝴蝶照片，以替换现有数据集中的非当地蝴蝶。只有全部以当地蝴蝶种类为研究对象时，识别出来的蝴蝶种类才能对当地的环境变化和气候监测提供更准确的参考。

2. 因为制作蝴蝶图像分割数据集时，对每张蝴蝶图像均进行了标注，这耗费了大量时间，以至于对蝴蝶图像分割算法的研究还不够充分。虽然本文图像分割方法在背景噪声较小的蝴蝶图像上的分割效果显著，但对于背景异常复杂的蝴蝶图像的处理能力还未达到理想效果。因此，在后续的研究中，将重点研究这些蝴蝶图像背景异常复杂的蝴蝶，并探索新的方法来提升分割效果。

3. 在蝴蝶图像种类识别部分，本文以残差网络为基准模型进行了改进，取得了较好的识别效果。但是，随着深度学习技术的不断发展，将会涌现出越来越多的优秀算法。因此，在后续的研究中，将引入更多优秀的新算法来进一步提升蝴蝶种类的识别效果。

## 参考文献

- [1] 刘宇琛, 肖能文, 何帅洁, 李雪梅, 郑晓旭, 吴刚. 武陵山区来凤县不同生境下昆虫多样性分析[J]. 植物保护学报, 2021, 48(03): 697-698.
- [2] 张超, 李娟, 程海云, 段家充, 潘昭. 秦岭西段地区蝴蝶群落多样性与环境因子相关性[J]. 生物多样性, 2023, 31(01): 194-206.
- [3] 曹嘉文. 基于深度学习的蝴蝶种类识别研究[D]. 硕士学位论文 陕西师范大学, 2019.
- [4] 谢娟英, 侯琦, 史颖欢, 吕鹏, 景丽萍, 庄福振, 张军平, 谭晓阳, 许升全. 蝴蝶种类自动识别研究[J]. 计算机研究与发展, 2018, 55(08): 1609-1618.
- [5] 周爱明, 马鹏鹏, 席天宇, 王江宁, 冯晋, 邵泽中, 陶玉磊, 姚青. 基于深度学习的蝴蝶科级标本图像自动识别[J]. 昆虫学报, 2017, 60(11): 1339-1348.
- [6] 魏兰君, 李凡凡, 秦毅, 孙洪武, 廖怀建. 蝴蝶生态景观功能植物选择及评价[J]. 江苏农业科学, 2021, 49(21): 147-152.
- [7] 田宇, 张金发, 田太安, 邢济春, 张兰, 蒙仕涛. 梵净山国家级自然保护区蝴蝶物种多样性调查[J]. 山地农业生物学报, 2022, 41(06): 6-13.
- [8] 韩丹, 王成, 殷鲁秦. 北京城市蝴蝶蜜源植物网络特征及重要蜜源植物识别[J]. 生态学报, 2021, 41(22): 8892-8905.
- [9] 袁改霞. 丰林自然保护区蝴蝶群落多样性[J]. 林业资源管理, 2021(05): 87-96.
- [10] 詹琦梁. 基于深度学习的蝴蝶图像分割和识别研究[D]. 硕士学位论文 浙江工业大学, 2020.
- [11] 李凡. 基于数字图像的蝴蝶种类自动识别研究[D]. 硕士学位论文 北京林业大学, 2015.
- [12] Verma S, Tripathi S, Singh A, Ojha M, Saxena R R. Insect detection and identification using YOLO algorithms on soybean crop[C]. TENCON 2021-2021 IEEE Region 10 Conference (TENCON). IEEE, 2021: 272-277.
- [13] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521: 436-444.
- [14] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.

- [15] Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386-408.
- [16] Rumelhart D E, Hinton G E, Williams R J. Learning internal representations by back-propagating errors[J]. Nature, 1986, 323(99): 533-536.
- [17] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [18] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [19] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [21] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:770-778.
- [22] Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
- [23] Sandler M, Howard A, Zhu M L, Zhmoginov A, Chen L C. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [24] Howard A, Sandler M, Chu G, Chen L C, Chen B, Tan M X, Wang W J, Zhu Y K, Pang R M, Vasudevan V, Le Q V, Adam H. Searching for mobilenetv3[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 1314-1324.
- [25] Zhang X Y, Zhou X Y, Lin M X, Sun J. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. 2018 IEEE/CVF Conference on Computer Vision (ICCV), 2018: 6848-6856.

- [26] Ma N N, Zhang X Y, Zheng H T, Sun J. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. Proceedings of the European conference on computer vision (ECCV), 2018: 116-131.
- [27] Tan M X, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]. Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019, 97: 6105-6114.
- [28] Tan M X, Le Q. Efficientnetv2: Smaller models and faster training[C]. Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021, 139: 10096-10106.
- [29] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3431-3440.
- [30] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [31] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [32] Zhao H S, Shi J P, Qi X J, Wang, X G, Jia, J Y. Pyramid scene parsing network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2881-2890.
- [33] Chen L C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.
- [34] Chen L C, Zhu Y K, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [35] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 580-587.

- [36] Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [37] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [38] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [39] 方伟政. 基于弱监督的昆虫图像细粒度分类技术研究[D]. 硕士学位论文 电子科技大学, 2021.
- [40] 曾伟辉, 张文凤, 陈鹏, 胡根生, 梁栋. 基于 SCResNeSt 的低分辨率水稻害虫图像识别方法[J]. 农业机械学报, 2022, 53(09): 277-285.
- [41] Kim S, Nguyen L T, Shim K, Kim J, Shim B. Pseudo-label-free weakly supervised semantic segmentation using image masking[J]. IEEE Access, 2022, 10: 19401-19411.
- [42] Hou Y W, Liu Z Y, Zhang T, Li Y J. C-UNet: Complement UNet for remote sensing road extraction[J]. Sensors, 2021, 21(6): 2153.
- [43] Weng L G, Xu Y M, Xia M, Zhang Y H, Liu J, Xu Y Q. Water areas segmentation from remote sensing images using a separable residual segnet network[J]. ISPRS international journal of geo-information, 2020, 9(4): 256.
- [44] Xie J Y, Jing T W, Chen B H, Peng J J, Zhang X W, He P H, Yin H L, Sun D Z, Wang W X, Xiao, A, Lyu S L, Li J. Method for Segmentation of Litchi Branches Based on the Improved DeepLabv3+[J]. Agronomy, 2022, 12(11): 2812.
- [45] 陈冬梅, 张赫, 魏凯华, 袁琳, 金晶, 吴开华. 复杂背景下昆虫图像的快速分割与识别[J]. 江苏农业科学, 2021, 49(24): 195-204.
- [46] Wang Z B, Wang K Y, Liu Z Q, Wang X F, Pan S H. A cognitive vision method for insect pest image segmentation[J]. IFAC-PapersOnLine, 2018, 51(17): 85-89.
- [47] Tang H, Wang B, Chen X. Deep learning techniques for automatic butterfly segmentation in ecological images[J]. Computers and Electronics in Agriculture, 2020, 178: 105739.



- [48] 黄世国, 洪铭淋, 张飞萍, 何海洋, 陈亿强, 李小林. 基于 F<sup>3</sup>Net 显著性目标检测的蝴蝶图像前背景自动分割[J]. 昆虫学报, 2021, 64(05): 611-617.
- [49] Filali I, Achour B, Belkadi M, Lalam M. Graph ranking based butterfly segmentation in ecological images[J]. Ecological Informatics, 2022, 68: 101553.
- [50] Kaya Y, Kayci L, Uyar M. Automatic identification of butterfly species based on local binary patterns and artificial neural network[J]. Applied Soft Computing, 2015, 28: 132-137.
- [51] Li F, Xiong Y. Automatic identification of butterfly species based on HoMSC and GLCMoIB[J]. The Visual Computer, 2018, 34(11): 1525-1533.
- [52] Yudha Kartika D S, Herumurti D, Rahmat B, Yuniari A, Maulana H, Anggraeny F T. Combining of extraction butterfly image using color, texture and form features[C]. 2020 6th Information Technology International Seminar (ITIS). IEEE, 2020: 98-102.
- [53] Chen X, Wang B, Gao Y S. Gaussian convolution angles: Invariant vein and texture descriptors for butterfly species identification[C]. 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 5798-5803.
- [54] 周爱明, 马鹏鹏, 席天宇, 王江宁, 冯晋, 邵泽中, 陶玉磊, 姚青. 基于深度学习的蝴蝶科级标本图像自动识别[J]. 昆虫学报, 2017, 60(11): 1339-1348.
- [55] Almryad A S, Kutucu H. Automatic identification for field butterflies by convolutional neural networks[J]. Engineering Science and Technology, an International Journal, 2020, 23(1): 189-195.
- [56] 谢娟英, 鲁银圆, 孔维轩, 许升全. 基于改进 RetinaNet 的自然环境中蝴蝶种类识别[J]. 计算机研究与发展, 2021, 58(08): 1686-1704.
- [57] Rajeeva PP F, Orban R, Vadivel K S, Subramanian M, Muthusamy S, Abd Elminaam D S, Nabil A, Abulaigh L, Ahmadi M, Mona AS A. A novel method for the classification of butterfly species using pre-trained CNN models[J]. Electronics, 2022, 11(13): 2016.
- [58] 周文进, 李凡, 薛峰. 基于 YOLOv3 和注意力机制的野外蝴蝶种类识别[J]. 郑州大学学报(工学版), 2022, 43(01): 34-40.

- [59] 张港. 基于深度学习轻量化卷积神经网络的遥感图像场景分类研究[D]. 硕士学位论文 南京邮电大学, 2022.
- [60] Li G, Han W T, Huang S J, Ma W T, Ma Q, Cui X. Extraction of sunflower lodging information based on UAV multi-spectral remote sensing and deep learning[J]. Remote sensing, 2021, 13(14): 2721.
- [61] Sun K, Xiao B, Liu D, Wang J D. Deep high-resolution representation learning for human pose estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 5693-5703.
- [62] 赵祥, 王涛, 张艳, 郑迎辉, 张昆, 王龙辉. 基于改进 DeepLabv3+孪生网络的遥感影像变化检测方法[J]. 地球信息科学学报, 2022, 24(08): 1604-1616.
- [63] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1251-1258.
- [64] 孙鹏炎. 基于深度学习的行人检测与重识别算法研究与应用[D]. 硕士学位论文 中国矿业大学, 2022.
- [65] 周博超, 韩雨男, 桂志国, 李郁峰, 张权. 基于 VGG 网络与深层字典的低剂量 CT 图像去噪算法[J]. 计算机工程, 2022, 48(04): 191-196+205.
- [66] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [67] Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.
- [68] 孟月波, 穆思蓉, 刘光辉, 徐胜军, 韩九强. 基于向量注意力机制 GoogLeNet-GMP 的行人重识别方法[J]. 计算机科学, 2022, 49(07): 142-147.
- [69] 彭明杰, 唐万梅, 皮家甜, 于昕, 吴至友, 黄房婷. 改进的基于端到端学习的蜻蜓目昆虫识别算法[J]. 重庆师范大学学报(自然科学版), 2021, 38(03): 94-99+142.
- [70] 单传辉, 叶绍华, 姚万琪, 张欣. 基于深度可分离残差网络的遥感影像路网检测[J]. 计算机技术与发展, 2023, 33(04): 75-81.

- [71] 彭红星, 何慧君, 高宗梅, 田兴国, 邓倩婷, 戚春龙. 基于改进 ShuffleNetV2 模型的荔枝病虫害识别方法[J]. 农业机械学报, 2022, 53(12): 290-300.
- [72] 杨炜, 黄立红, 屈晓磊. 融合双注意力机制 EfficientNet 的沥青路面状态分类方法[J]. 机械工程学报, 2022, 58(24): 211-222.
- [73] Sridevi M. Image segmentation based on multilevel thresholding using firefly algorithm[C]. 2017 International Conference on Inventive Computing and Informatics (ICICI). IEEE, 2017: 750-753.
- [74] 谢娟英, 曹嘉文, 马丽滨, 甄文全, 陈振宁, 李晓东, 李后魂, 许升全. 蝴蝶物种自动识别研究的生态照片数据集[J]. 中国科学数据(中英文网络版), 2019, 4(03): 193-198.
- [75] Zhang K, Guo Y R, Wang X S, Yuan J S, Ding Q L. Multiple feature reweight densenet for image classification[J]. IEEE Access, 2019, 7: 9872-9880.
- [76] Qu H Y, Yang J H, Shen M G, Chen H T, Zhou D J. Fault diagnosis of rolling bearing under time-varying speed conditions based on EfficientNetv2[J]. Measurement Science and Technology, 2022, 33(6): 065023.
- [77] Zhang Q L, Yang Y B. SA-NET: Shuffle attention for deep convolutional neural networks[C]. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 2235-2239.
- [78] Wu Y X, He K M. Group normalization[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [79] 史加荣, 王丹, 尚凡华, 张鹤于. 随机梯度下降算法研究进展[J]. 自动化学报, 2021, 47(09): 2103-2119.
- [80] 李翠锦, 瞿中. 基于卷积神经网络的跨层融合边缘检测算法[J]. 计算机应用研究, 2021, 38(07): 2183-2187.
- [81] Leng Z Q, Tan M X, Liu C X, Cubuk E D, Shi X J, Cheng S Y, Anguelov D. Polyloss: A polynomial expansion perspective of classification loss functions[J]. arXiv preprint arXiv:2204.12511, 2022.
- [82] 朱琳. 基于图注意力网络的类不平衡和类重叠数据二分类算法研究[D]. 硕士学位论文 山东大学, 2022.

- [83]唐婷, 潘新. 一种基于可变形卷积的高光谱图像分类算法[J]. 光电子·激光, 2022, 33(05): 488-494.
- [84]邓中港, 代刚, 吴湘宁, 邓玉娇, 王稳, 陈苗, 涂雨, 张锋, 方恒. 基于注意力机制和可变形卷积的金属表面细微不规则损伤的图像识别模型[J]. 计算机工程与科学, 2023, 45(01): 127-135.

## 图版

图 1-1 枯叶蛱蝶 .....	2
图 1-2 绿带燕凤蝶与燕凤蝶 .....	2
图 1-3 雌性和雄性的玉带凤蝶 .....	2
图 2-1 卷积神经网络结构 .....	11
图 2-2 卷积运算操作 .....	12
图 2-3 最大池化操作 .....	12
图 2-4 平均池化操作 .....	13
图 2-5 SegNet 网络结构 <sup>[60]</sup> .....	14
图 2-6 PSPNet 网络结构 <sup>[32]</sup> .....	15
图 2-7 HRNet 主干网络结构 <sup>[61]</sup> .....	15
图 2-8 DeepLabv3+网络结构 <sup>[62]</sup> .....	16
图 2-9 Inception 模块结构 <sup>[67]</sup> .....	18
图 2-10 残差结构示意图 <sup>[21]</sup> .....	19
图 2-11 ResNet 网络中的两种残差结构 <sup>[21]</sup> .....	20
图 2-12 深度可分离卷积网络结构 <sup>[70]</sup> .....	20
图 2-13 普通卷积与深度可分离卷积的基本结构 <sup>[22]</sup> .....	22
图 2-14 ShuffleNet-V2 的基本单元和下采样单元 <sup>[26]</sup> .....	22
图 2-15 MBConv 模块结构 <sup>[72]</sup> .....	23
图 3-1 蝴蝶图像数据集 .....	27
图 3-2 蝴蝶分割图像的标注界面 .....	29
图 3-3 标签的可视化 .....	29
图 3-4 数据增强 .....	30
图 3-5 蝴蝶图像分割模型总体结构 .....	31
图 3-6 Fused-MBConv 与 MBConv 网络结构 <sup>[28]</sup> .....	33
图 3-7 ASPP 与 DASPP 网络结构对比 .....	34
图 3-8 $r=12$ 时 ASPP 与 DASPP 采样点密集性对比 .....	35
图 3-9 Sigmoid 激活函数 .....	36
图 3-10 置换注意力网络结构 <sup>[77]</sup> .....	37

图 3-11 消融实验可视化结果.....	40
图 3-12 不同分割网络效果图比较 .....	42
图 4-1 蝴蝶图像识别数据集示例 .....	44
图 4-2 LD-ResNet 模型总体结构 .....	45
图 4-3 可变形卷积结构 <sup>[84]</sup> .....	47
图 4-4 深度可分离卷积与改进后的深度可分离卷积结构对比.....	49
图 4-5 ResNet-18 模型与 LD-ResNet 模型的平均识别准确率和损失值 .....	55
表 2-1 常见的 VGG 网络结构 <sup>[66]</sup> .....	17
表 2-2 ResNet 网络结构 <sup>[21]</sup> .....	19
表 2-3 MobileNet 网络结构 <sup>[22]</sup> .....	21
表 2-4 EfficientNet-B0 网络结构参数 <sup>[27]</sup> .....	23
表 3-1 蝴蝶图像数据集分布 .....	25
表 3-2 EfficientNetV2-S 网络结构参数 <sup>[28]</sup> .....	32
表 3-3 不同主干特征提取网络对比 .....	39
表 3-4 消融实验 .....	39
表 3-5 分割模型对比实验 .....	41
表 4-1 去除背景噪声前后对比 .....	52
表 4-2 数据增强前后对比 .....	52
表 4-3 以 ResNet-18 网络为基础的消融实验 .....	53
表 4-4 深度可分离卷积和改进后的深度可分离卷积的参数数量对比.....	54
表 4-5 不同识别模型的对比 .....	54