DOI：10.13718/j.cnki.xdzk.2025.07.016
邵晓艳，董文永，赵雪专，等. 基于改进 DeepLabv3＋的安全帽佩戴分割算法［J］. 西南大学学报(自然科学版)，2025，47(7)：185-195.

基于改进 DeepLabv3＋的安全帽佩戴分割算法

邵晓艳[1]，　董文永[2]，　赵雪专[1]，　李玲玲[1]，　薄树奎[1]

1. 郑州航空工业管理学院 计算机学院，郑州 450046；2. 武汉大学 计算机学院，武汉 430072

摘要：针对物流园区空间跨度大、作业设备繁多导致安全帽佩戴检测分割难度增加的问题，提出一种基于改进 DeepLabv3＋的安全帽佩戴分割算法。该算法采用 ResNet-101 膨胀残差网络进行特征提取；在编码阶段引入卷积注意力机制融合模块，有效增强特征区域表征能力；在特征提取阶段引入图像特征网格化模块，将低分辨率图像进行平均切分，有助于获得局部图像的小目标特征。将该算法在 SHWD(Safety Helmet Wearing Detect)数据集中训练测试，结果表明：算法的像素准确率达到 89.23%，相比 DeepLabv3＋提升了 2.21 个百分点，有效提高了复杂场景下物流园区安全帽佩戴分割精度。

关　键　词：神经网络；注意力机制；膨胀卷积；语义分割
中图分类号：TP391　　**文献标识码**：A
文 章 编 号：1673－9868(2025)07－0185－11　　开放科学(资源服务)标识码(OSID)：

Segmentation Algorithm of Helmet Wear Based on the Improved DeepLabv3＋

SHAO Xiaoyan[1]，　DONG Wenyong[2]，　ZHAO Xuezhuan[1]，
LI Lingling[1]，　BO Shukui[1]

1. School of Computer Science，Zhengzhou University of Aeronautics，Zhengzhou 450046，China；
2. School of Computer Science，Wuhan University，Wuhan 430072，China

Abstract：To address the challenges of increased difficulty in safety helmet wearing detection and segmentation caused by large spatial spans and numerous operational equipment in logistics parks，a safety helmet wearing segmentation algorithm based on improved DeepLabv3＋ was proposed. The algorithm used ResNet-101 expansion residual network to extract features. In the coding stage，the convolutional attention

收稿日期：2024－07－12
基金项目：国家自然科学基金项目(U1904119)；河南省科技攻关计划项目(252102210034)；河南省重点研发专项(231111212000)；航空科学基金项目(20230001055002)。
作者简介：邵晓艳，硕士，副教授，主要从事机器学习、模式识别、计算机视觉等研究。
通信作者：赵雪专，博士，副教授。

mechanism fusion module was introduced to effectively enhance the ability of feature region representation. In the feature extraction stage, the image feature grid module was introduced to average segmentation of low-resolution images, which was helpful to obtain small target features of local images. The algorithm was trained and tested with the Safety Helmet Wearing Detect (SHWD) Dataset, and the results showed that the pixel accuracy of algorithm reached 89.23%, which was 2.21 percentage points higher than DeepLabv3+, effectively improving the segmentation accuracy of safety helmet wearing in the logistics parks in complex scenes.

Key words: neural network; attention mechanism; dilated convolution; semantic segmentation

近年来，在电子商务和供应链升级的推动下，我国物流基础设施规模持续扩大，各类物流园区数量呈现爆发式增长。根据中华人民共和国应急管理部最新发布的行业安全白皮书数据显示，在物流作业场所发生的安全事故中，因未规范佩戴安全帽防护装备而导致的伤害事件占比较高。为应对这一安全隐患，新修订的《中华人民共和国安全生产法》特别强化了物流企业在个人防护装备管理方面的主体责任。在此背景下，融合计算机视觉、深度学习等新一代信息技术的智能安全监测解决方案，正逐步成为提升物流园区安全管理水平的重要技术手段，并在物流企业中取得显著应用成效[1-2]。

随着新一代信息技术与物流产业的深度融合，智能感知和工业互联技术为物流园区安全管理体系创新注入了新动能。一般情况下安全帽检测[3-5]分为3种方式：① 将感应设备嵌入到安全帽内；② 根据形状、颜色等多个特征来进行检测；③ 结合深度学习算法进行智能检测。前2种检测方式操作过程比较复杂，检测速度比较缓慢。以深度学习为基础的神经网络算法，对复杂环境适应性强，且检测效率更高[6]，因此基于深度神经网络的识别技术成为安全帽佩戴检测与监管的一种重要方法[7-8]。

基于神经网络算法对安全帽图像进行检测和分割，实际上是采用计算机视觉研究中的语义分割算法对图像进行像素级分类[9-10]。Ren等[11]提出的Faster R-CNN算法相较于单阶段检测算法更加精准，可以解决多尺度、小目标问题，但速度相较于单阶段检测算法更慢[12]。刘雅洁等[13]在主干网络添加坐标注意力机制，提高模型对关键特征的注意力，更聚焦于训练安全帽相关目标特征并提高准确率，但对于密集小目标的检测效果不够理想。Wu等[14]为进一步提高检测精度，将单阶段检测算法和注意力机制进行融合，但该算法同样存在对密集目标检测效果不够理想的问题[15]。田乔鑫等[16]改进了传统的分类算法LeNet-5[17]，将支持向量机和神经网络进行结合，该算法因为存在后处理操作，能有效提高目标识别精度，但是在实时性上有待提高。戴天虹等[18]引入EfficientDet算法，将预设边框用于目标检测架构，能进一步提高对小目标的检测精度。

基于工程实施场景下检测任务的高实时性要求，目前以YOLO算法为代表的单阶段目标检测算法应用较为广泛[19-20]。其中针对安全帽佩戴检测任务的代表性研究成果有：徐守坤等[21]基于YOLOv3系列算法进行目标检测和分割，对于安全帽给出了相应的语义描述，但在语义描述的复杂度和多样性上有待提高[22]；王玲敏等[23]在YOLOv5算法中引入坐标注意力机制，同时引入加权双向特征金字塔结构，可以有效预测安全帽的位置；吕宗喆等[24]基于YOLOv5算法优化边界框回归损失函数和置信度预测损失函数的计算方式，引入切片辅助微调和切片辅助推理对输入网络的图像进行切片处理，使小目标对象产生更大的像素区域，进而改善网络推理与微调的效果从而提高安全帽佩戴检测的精确性；张锦等[25]针对安全帽尺寸不同的问题，在YOLOv5算法中采用K-Means++算法重新聚类，再引入多光谱通道注意力机制增强信息传播并提高检测效果；杨大为等[26]在YOLOv7算法中引入卷积块注意力机制，同时增加1个小目标层，将浅层网络特征与深层网络特征融合，提高安全帽佩戴检测精度。

针对现有安全帽检测算法参数量过多、复杂度高、对不同场景泛化能力差[27]，以及原始DeepLabv3+网络收敛效率低、分割准确率受限的问题，提出一种端到端的基于改进DeepLabv3+的安全帽语义分割算

法模型 Grid-DeepLabv3＋，以期提高复杂场景下物流园区安全帽佩戴的分割精度。

1　基于改进 DeepLabv3＋的语义分割架构

相较于原始 DeepLabv3＋网络，Grid-DeepLabv3＋网络主要进行了以下 3 方面改进：

1）DeepLabv3＋主干网络采用的是 Xception 网络模型，该模型存在网络结构层数较深的问题。采用膨胀残差特征提取网络 ResNet-101 替换 Xception 网络，以减小计算量，提高了网络模型的收敛速度，同时也有助于提取图像的高维特征。

2）为有效关注物流园区图像的显著位置，忽略无关背景信息，将通道注意力机制和空间注意力机制模块放置在编码阶段，有助于提升特征区域表征能力，有效提高图像语义分割的精度。

3）为有效捕捉物流园区图像小目标特征信息，引入图像特征网格化模块，将低分辨率图像进行平均切分，更好地捕获局部区域的小目标特征信息。

基于 Grid-DeepLabv3＋网络的语义分割架构如图 1 所示。

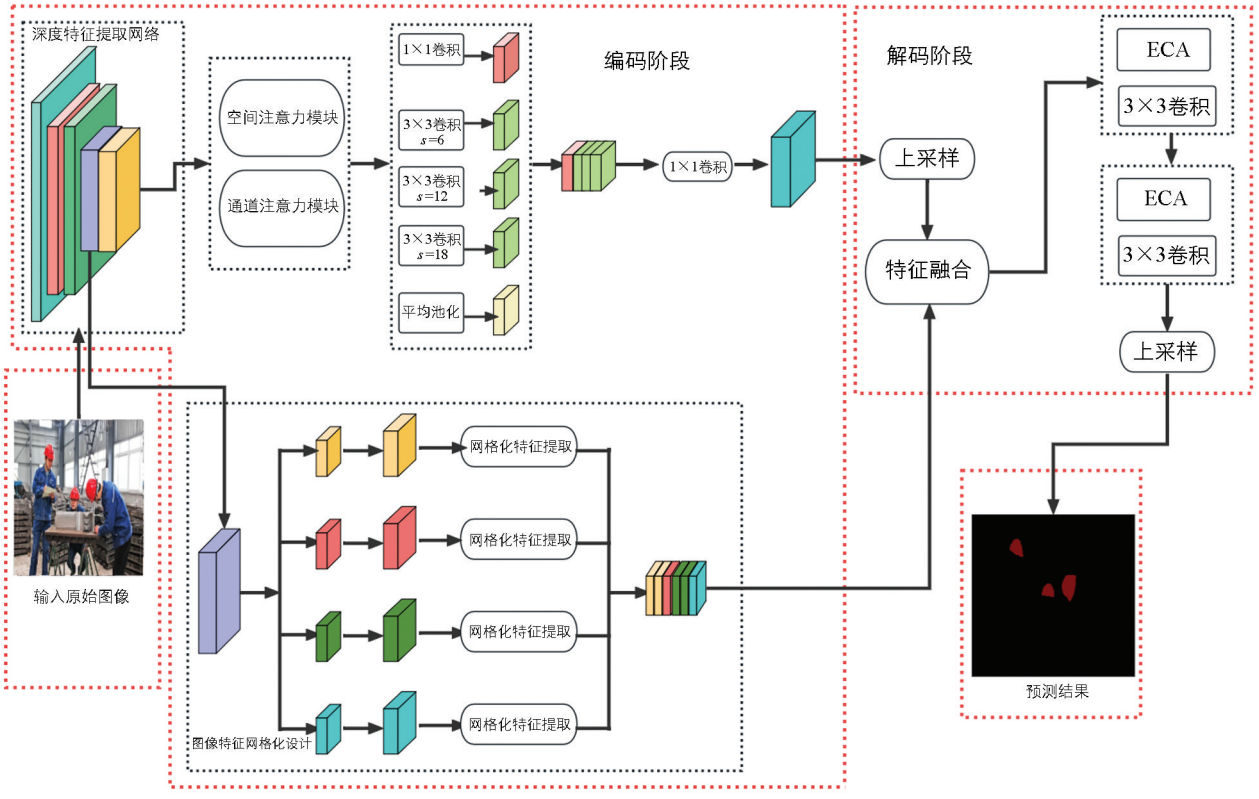


图 1　基于 Grid-DeepLabv3＋网络的语义分割架构

2　编码层网络构建

工业物流园区场景图像语义信息比较复杂，分割模型提取图像特征信息的核心部分在编码阶段。本文提出的 Grid-DeepLabv3＋网络架构编码阶段在 3 个部分进行了改进，分别为：特征提取网络、卷积注意力机制融合模块、图像特征网格化模块。具体实现方式为：首先将图像数据输入到主干网络 ResNet-101 中进行特征提取，产生不同尺度的特征图（20×20、40×40、80×80 等），对于合并得到的特征图进行不同膨胀率的膨胀卷积学习，获取其特征的同时增加感受野；然后将膨胀卷积学习完成的特征图进行拼接，经过 1×1 卷积调整通道进行上采样操作以便与接下来处理的特征图融合；最后将特征图进行网格化设计，对不同尺寸的网格化像素上采样后，进行多尺度特征融合，再将经过处理的所有特征合并，通过卷积操作实现像素级分类。具体步骤见算法 1。

算法1 编码层网络构建算法

输入：输入特征图 $\boldsymbol{X}_{\text{input}} \in \mathbf{R}^{B\times C\times H\times W}$

输出：输出特征图 $\boldsymbol{Y}_{\text{output}} \in \mathbf{R}^{B\times C\times H\times W}$

1. 通过3×3卷积将输入特征图拆分为4部分：$[\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4] \leftarrow \text{Conv}_{3\times 3}(\boldsymbol{X}_{\text{input}})$
2. 合并拆分结果：$\boldsymbol{X}_{\text{all}} \leftarrow [\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4]$
3. 将 $\boldsymbol{X}_{\text{all}}$ 拆分为4个子块：$\{\boldsymbol{X}_{00}, \boldsymbol{X}_{10}, \boldsymbol{X}_{01}, \boldsymbol{X}_{11}\} \leftarrow \text{Split}(\boldsymbol{X}_{\text{all}})$
4. 拼接子块：$\boldsymbol{X} \leftarrow \text{Cat}(\{\boldsymbol{X}_{00}, \boldsymbol{X}_{10}, \boldsymbol{X}_{01}, \boldsymbol{X}_{11}\})$
5. 对拼接结果进行1×1卷积：$\boldsymbol{X} \leftarrow \text{Conv}_{1\times 1}(\boldsymbol{X})$
6. 网格重组：$\boldsymbol{X}_{\text{all}}[index] \leftarrow \text{Backsplit}(\boldsymbol{X})$
7. 合并所有处理后的特征图：$\boldsymbol{X} \leftarrow \text{Cat}(\boldsymbol{X}_{\text{all}})$
8. 通过卷积生成最终输出 $\boldsymbol{Y}_{\text{output}}$

2.1 膨胀卷积

在工业物流园区的图像分割场景中，为了进一步捕捉更广阔的上下文信息，提取丰富语义特征，采用膨胀卷积(Dialted Convolution)进行主干网络特征提取。膨胀卷积是一种通过引入膨胀率来扩大感受野的特殊卷积操作。与标准卷积相比，其核心特点是在卷积核元素之间插入特定间隔，从而在不增加参数量的前提下，显著扩大感受野范围。这种设计具有3大关键优势：① 完全避免了传统下采样操作带来的信息损失，完整保留了输入特征图的空间分辨率和细节信息；② 通过灵活调节膨胀率，可以动态控制特征提取的感知范围，从局部细节到全局上下文实现多尺度特征捕获；③ 计算效率更高，不会明显增加模型的计算负担。

标准卷积计算公式为：

$$P(m, n) \cdot Q(m, n) = \sum_{i=0}^{w} \sum_{j=0}^{h} Q(i, j) \cdot P(m-i, n-j) \tag{1}$$

式中：$P(m, n)$为图像上点(m, n)处的像素值；$Q(m, n)$为卷积核。

膨胀卷积相比标准卷积而言，增加了膨胀率，其计算公式为：

$$P(m, n) \cdot Q(m, n) = \sum_{i=0}^{w} \sum_{j=0}^{h} Q(i, j) \cdot P(m-s\times i, n-s\times j) \tag{2}$$

式中：s 为膨胀卷积的膨胀率。

膨胀卷积的效果如图2所示。

a. 膨胀率s=1　　b. 膨胀率s=2　　c. 膨胀率s=3

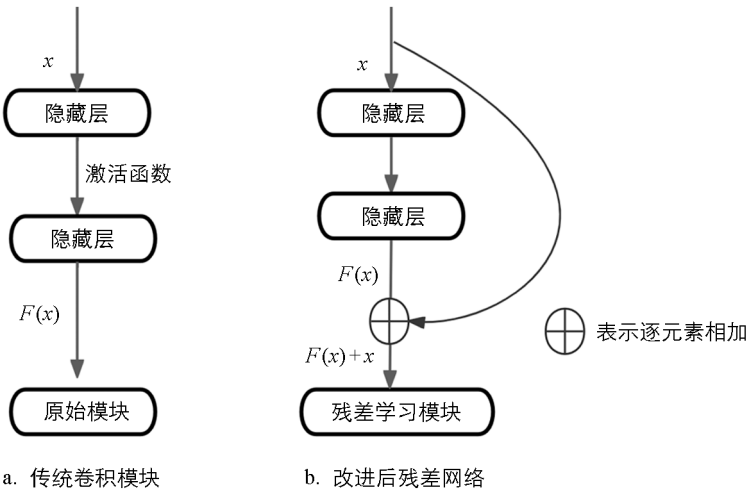
图2 膨胀卷积效果

2.2 膨胀残差网络

针对训练网络出现过拟合以及准确率不高的现象，He等[28]提出深度卷积残差网络结构，并得到广泛应用[29]。传统的卷积模块如图3a所示，改进后的残差网络如图3b所示。若输入的特征为 x，则残差函数表示为 $F(x)$，那么构建的映射函数 $H(x)$为：

$$H(x) = x + F(x) \tag{3}$$

采用 ResNet-101 残差网络代替网络模型中的骨干提取网络，通过引入残差结构有效缓解深层网络中的梯度消失问题，提升模型表现。残差块是其核心单元，它允许网络学习输入和输出之间的残差，而不是直接从输入到输出的映射。残差块的结构按顺序包括 1×1 卷积(用于减少通道数)、3×3 卷积(用于主要特征提取)和 1×1 卷积(恢复通道数)，并在每个卷积操作后进行批归一化和 ReLU 激活处理。ResNet-101 网络结构包括 1 个 7×7 的卷积层(64 通道，步幅为 2)和 1 个 3×3 最大池化层(步幅为 2)，随后是多个残差块组。具体细节如下：Conv2 组有 3 个残差块(64 通道)；Conv3 组有 4 个残差块(128 通道)；Conv4 组是关键部分，包含 23 个残差块(256 通道)；Conv5 组包含 3 个残差块(512 通道)。在最后一个残差块组后，使用全局平均池化层降维，然后通过全连接层输出分类结果。ResNet-101 的跳跃连接(Skip Connection)使每个残差块的输入不仅经过卷积处理，还可以直接跳跃到块的输出处，并与卷积结果相加，从而缓解梯度消失问题，增强网络的训练稳定性和深层网络的可训练性。



a. 传统卷积模块　　b. 改进后残差网络

图 3　传统卷积模块与改进后残差网络对比

残差网络可以通过深度的增加来提升特征提取的准确率，其主要特点是容易优化。残差内部通过跳跃连接结构，有效缓解过深的神经网络结构中出现梯度消失的弊端，在一定程度上防止模型过拟合发生。在 ResNet-101 网络中有多个卷积阶段，针对这些阶段均进行下采样操作从而将图像特征图进行分辨率缩减，通常每个阶段将特征图分辨率缩减为上一阶段的一半。残差膨胀网络在不增加参数量的基础上，能够扩大卷积核的感受野，有助于深层次语义信息的提取，但是对于物流园区场景图像的小目标关注度不够。鉴于以上原因，本文采用 2、2、2、3、6 的膨胀卷积代替传统卷积进行下采样操作，这样既可以确保特征图的分辨率大小不变，又可以扩大每个阶段的感受野。在 $F(x)$中经过大量实验，选取合适位置引入膨胀卷积对该模块输入的特征图进行特征编码。膨胀卷积在不增加计算复杂度的情况下，能够在保持特征图空间分辨率的同时，捕捉更广泛的上下文信息，它比增加卷积核尺寸或网络深度更加高效。详细算法如算法 2 所示。

算法 2　膨胀残差网络算法

输入：图像 $\boldsymbol{I}\in\mathbf{R}^{C\times H\times W}$

输出：特征图 $\boldsymbol{F}\in\mathbf{R}^{B\times C\times H\times W}$

1. 残差块 Conv2 包含：ResidualblockConv2[i]←[Conv(1×1，64)，Conv(3×3，64)，Conv(1×1，64)]
2. 跳跃连接：Output[i]←ResidualblockConv2[i]＋Input
3. 残差块 Conv3：ResidualblockConv3[i]←[Conv(1×1，128)，Conv(3×3，128)，Conv(1×1，128)]
4. 跳跃连接：Output[i]←ResidualblockConv3[i]＋Input
5. 残差块 Conv4：ResidualblockConv4[i]←[Conv(1×1，256)，Conv(3×3，256)，Conv(1×1，256)]
6. 跳跃连接：Output[i]←ResidualblockConv4[i]＋Input
7. 残差块 Conv5：ResidualblockConv5[i]←[Conv(1×1，512)，Conv(3×3，512)，Conv(1×1，512)]
8. 跳跃连接：Output[i]←ResidualblockConv5[i]＋Input
9. 各卷积层应用空洞卷积，膨胀率为(2，2，2，3，6)：D_{Conv}←DilatedConv(s＝2，2，2，3，6)
10. 将空洞卷积后的多尺度特征拼接：Concat_f←Concat(D_{Conv})
11. 通过 1×1 卷积输出特征图：$\boldsymbol{F}$←Conv(1×1，Concat_f)

2.3 卷积注意力机制融合模块

为提升特征提取的有效性，本文创新性地将轻量化的卷积注意力机制融合模块嵌入卷积神经网络架构，在保持模型参数量不变的前提下，通过双注意力机制优化特征学习过程。该模块能够有效抑制背景噪声干扰，动态强化关键特征表达，并自适应地生成通道与空间维度的特征权重分布。具体而言，网络编码器部分采用了如图4所示的并行注意力结构，其中通道注意力(Channel Attention，CA)模块专注于特征通道间的关系建模，而空间注意力(Spatial Attention，SA)模块则重点捕捉特征图的空间相关性。

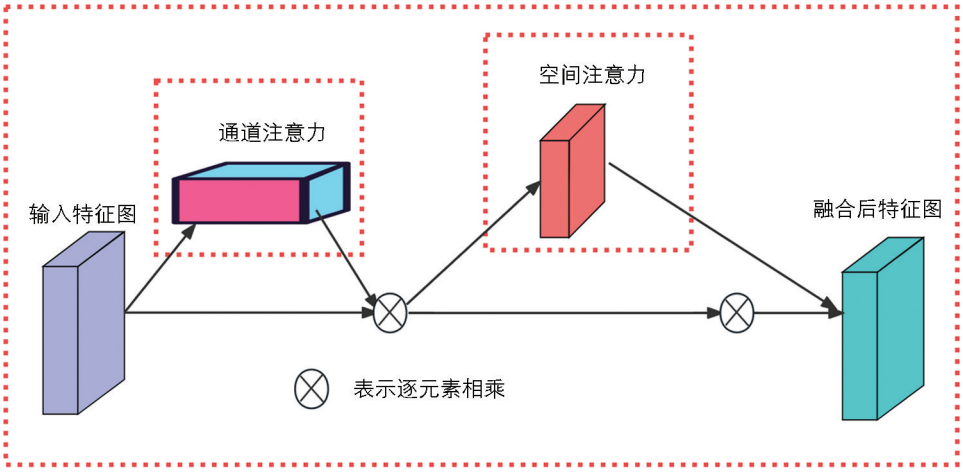


图4 卷积注意力机制融合模块

针对工业物流园区场景特征图 $\boldsymbol{A}$ 的处理流程如下：首先输入图像特征图经过通道注意力模块处理，输出维度为一维通道权重向量，记作 $\boldsymbol{A}_a \in \mathbf{R}^{C\times 1\times 1}$；随后空间注意力模块对特征图进行空间域分析，生成二维空间权重矩阵 $\boldsymbol{A}_b \in \mathbf{R}^{C\times H\times W}$。2个注意力特征的计算可形式化表示为：

$$\boldsymbol{A}' = \boldsymbol{A}_a(\boldsymbol{A}) \otimes \boldsymbol{A} \tag{4}$$

$$\boldsymbol{A}'' = \boldsymbol{A}_b(\boldsymbol{A}') \otimes \boldsymbol{A}' \tag{5}$$

其中：⊗为物流园区场景图像对应元素相乘。

2.3.1 通道注意力模块设计

通道注意力机制是通过MaxPool和AvePool 2种池化操作实现对图像空间信息的汇总。在物流园区场景图像语义分割任务中，该运算有助于迅速确定图像中的重点部位，有效提升特征区域的表征能力。通道注意力机制先进行MaxPool和AvgPool池化操作，然后将生成的特征图进行全连接操作输出，全连接阶段采取参数共享策略，最终生成的特征图经过Sigmoid激活函数处理，输出通道注意力权重向量，具体可表示为：

$$\begin{aligned}\boldsymbol{A}_a(\boldsymbol{A}) &= S\{F[\mathrm{MaxPool}(\boldsymbol{A})] + F[\mathrm{AvgPool}(\boldsymbol{A})]\} = \\ &S\{\omega_1[\omega_0(\boldsymbol{A}^c_{\max})] + \omega_1[\omega_0(\boldsymbol{A}^c_{\mathrm{avg}})]\}\end{aligned} \tag{6}$$

式中：$\boldsymbol{A}$ 为输入的特征图；S 为Sigmoid激活函数；F 为全连接操作；ω_0、ω_1 为共享参数；c 为特征通道数；$\boldsymbol{A}_a(\boldsymbol{A})$为通道注意力模块参数矩阵。

2.3.2 空间注意力模块设计

空间注意力模块通过AvgPool和MaxPool池化操作，分别提取通道的特征图信息，然后通过5×5的卷积操作进行合并，最后生成空间注意力特征图，具体可表示为：

$$\begin{aligned}\boldsymbol{A}_b(\boldsymbol{A}) &= S\{f^{5\times 5}[\mathrm{AvgPool}(\boldsymbol{A}), \mathrm{MaxPool}(\boldsymbol{A})]\} = \\ &S\{f^{5\times 5}[\boldsymbol{A}^c_{\mathrm{avg}}, \boldsymbol{A}^c_{\max}]\}\end{aligned} \tag{7}$$

式中：$\boldsymbol{A}$ 为输入的特征图；S 为Sigmoid激活函数；$f^{5\times 5}$ 为卷积核大小为5×5的卷积核；$\boldsymbol{A}_b(\boldsymbol{A})$为空间注意力模块参数矩阵。

2.4　图像特征网格化模块

本文提出的 Grid-DeepLabv3＋网络模型引入图像特征网格化模块。在骨干网络提取图像特征过程中，图像分辨率是一个逐渐下降的过程。当分辨率下降到原始图像的 1/32 时，将这幅图像进行网格化平均分割，有助于获取局部区域的小目标信息。当然也可以在分辨率下降到原始图像的 1/8、1/16 等其他比例时进行网格化分割，但过少的分割不足以关注到小目标，过大的分割又会迅速增加模型的参数量，经过反复实验，确定在图像分辨率下降至 1/32 时进行图像特征网格化分割。

网格化设计特征提取的具体实施方法如算法 3 所示。首先对第 3 层特征图进行网格化设计，将 4 种网格化特征像素通过上采样操作恢复到原来尺寸大小，实现将图像中的小目标进一步放大，有助于提高语义分割的精度，以更好地提取不同目标大小的关键信息，并且增加感受野大小；随后将不同尺度大小特征图上采样进行多尺度特征融合来丰富特征图的内容表示；最后，第 4 层特征图上采样完成后进行特征融合操作，同时将特征图通道数量通过 1×1 卷积减少到对应类别数量，实现对图像进行像素级分类。

算法 3　网格化设计特征提取算法

输入：第 3 层特征图 $\boldsymbol{X}_{\text{input3}} \in \mathbf{R}^{B\times C\times H\times W}$；第 4 层特征图 $\boldsymbol{X}_{\text{input4}} \in \mathbf{R}^{B\times C\times H\times W}$

输出：分类后的图像 $\boldsymbol{Y} \in \mathbf{R}^{B\times C\times H\times W}$

1. 将第 3 层特征图 $\boldsymbol{X}_{\text{input3}} \in \mathbf{R}^{B\times C\times H\times W}$ 拆分为 4 个子块：$[\boldsymbol{X}_{3,1}, \boldsymbol{X}_{3,2}, \boldsymbol{X}_{3,3}, \boldsymbol{X}_{3,4}] \leftarrow \boldsymbol{X}_{\text{input3}}$
2. 对子块进行上采样：$\boldsymbol{X}_{3,\text{all}} \leftarrow \text{Upsampling}([\boldsymbol{X}_{3,1}, \boldsymbol{X}_{3,2}, \boldsymbol{X}_{3,3}, \boldsymbol{X}_{3,4}])$
3. 应用融合函数 fusion，遍历 $\boldsymbol{X}_{3,\text{all}}$ 中的每个子块：$\boldsymbol{X}_{3,\text{all}}[index] \leftarrow \text{fusion}(\boldsymbol{X})$
4. 类似地，对第 4 层特征图 $\boldsymbol{X}_{\text{input4}}$ 进行融合：$\boldsymbol{X}_{4,\text{all}}[index] \leftarrow \text{fusion}(\boldsymbol{X})$
5. 将融合后的第 3 层和第 4 层特征进行拼接：$\boldsymbol{X} \leftarrow \text{Concat}(\boldsymbol{X}_{3,\text{all}} + \boldsymbol{X}_{4,\text{all}})$
6. 通过 1×1 卷积生成最终输出 $\boldsymbol{Y}$

将变量 r 作为分割比率，r 的取值可以为 1、2、3、4 等。由图 5 可知，分割后小格子内的图像蕴含各个位置的空间特征信息。对于每个小格子图像，通过上采样操作恢复到原来尺寸大小，从而将图像中的小目标进一步放大，这样有助于提高语义分割精度。本文通过引入图像特征网格化模块的设计，能进一步捕获到小目标信息，有效提升图像分割能力。

a. r=2　　b. r=3　　c. r=4

图 5　图像特征网格化设计

在图像特征网格化模块中对于原始分辨率 1/32 的图像，经过上采样还原成分割前的大小，这对有效提取小目标物体提出了挑战。本文在骨干特征提取网络中引入膨胀卷积来分别获取多尺度特征信息，其结构如图 6 所示。特征提取网络结构分 2 步进行特征提取：① 采用卷积核大小为 1×1 的滤波器，通过卷积操作降低特征图的通道数；② 将 2 个膨胀卷积进行串联操作，进一步扩大网络模型的感受野，同时还能保证图像分辨率的大小。最后将 2 步操作合并成一个新的特征图，进行 1 次全局池化操作。通过该操作，既可以有效提取中间层信息，又可以对图形中小目标特征进行有效提取，整体上提升了图像语义的分割精度。

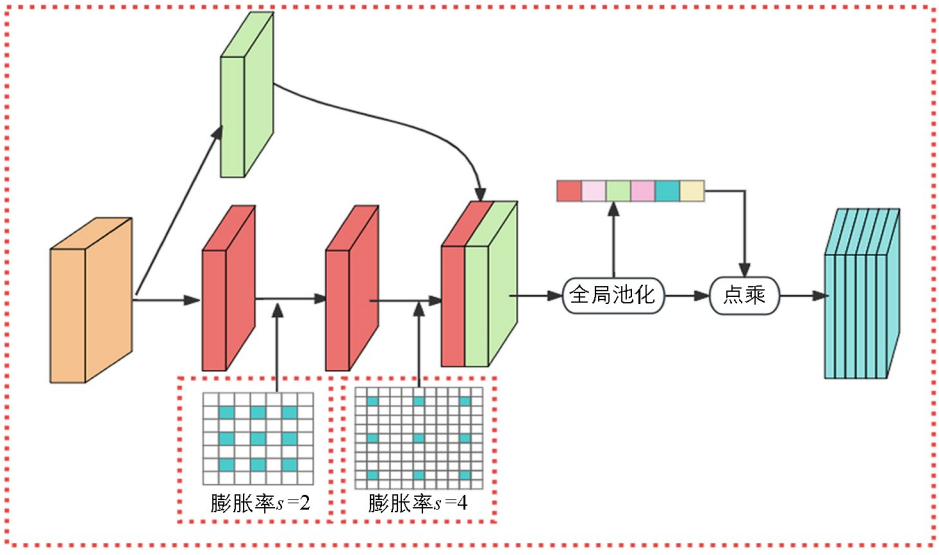


图6 图像特征网格化模块

3 实验结果与分析

3.1 数据集与实验环境设置

为验证语义分割模型Grid-DeepLabv3+适应物流园区安全帽语义分割任务的有效性，采用开源数据集SHWD(Safety Helmet Wearing Detect)进行实验分析。SHWD数据集作为专门用于安全帽佩戴检测的数据集，其设计初衷是涵盖多种复杂场景下的安全帽佩戴情况。该数据集存在多种干扰因素，如光照变化大、人员密集、遮挡情况多、背景复杂多变等。因此从一定程度上讲，SHWD数据集能够反映复杂场景下的安全帽佩戴检测问题。该数据集包含7 581张标注图像，在数据预处理阶段，使用Labelme工具对所有样本进行精细标注，并将数据集划分为6 000张训练样本、581张验证样本和1 000张测试样本，确保模型训练和评估的科学性。

实验硬件和软件环境参数配置如表1所示。

表1 实验环境参数配置

环境类别	设备名称	参数
硬件环境	CPU	i7 13代
	GPU	GeForce GTX 4090
	内存	32G
软件环境	操作系统	Windows 10
	编程语言	Python 3.6
	深度学习框架	PyTorch 1.10.0
		Torchvision 0.6.1
	CUDA	10.2

为提高算法的泛化性能，数据预处理使用MOSAIC数据增强方法。该方法是基于多种图像增强技术设计的一种综合性策略，旨在通过多种手段提升模型的泛化能力。该策略结合了旋转、翻转、颜色调节、高斯噪声添加等操作，从而生成更加多样化和更具挑战性的训练样本。考虑到硬件资源的限制，本文采用640×640的图像分块处理策略。

3.2 定量评估指标

实验选用召回率R、平均交并比$mIoU$、像素准确率P_a共3个评估指标来衡量改进模型的性能，其中

分类任务 4 个指标及其意义如下：TP 表示预测正确的真正例，即模型预测为正例，实际为正例；FP 表示预测错误的假正例，即模型预测为正例，实际为反例；FN 表示预测错误的假反例，即模型预测为反例，实际为正例；TN 表示预测正确的真反例，即模型预测为反例，实际为反例。

召回率 R 是指预测样本为真值，并且该样本的标签也为真值的个数与所有真值样本个数的比值，其计算公式如下：

$$R=\frac{TP}{TP+FN} \tag{8}$$

平均交并比 $mIoU$ 是指除了背景之外，计算所有分割标签对象的交并比平均值，其计算公式如下：

$$mIoU=\frac{1}{s+1}\sum_{i=0}^{s}\frac{TP}{TP+FP+FN} \tag{9}$$

式中：s 为图像分割中待区分的目标类别总数。

像素准确率 P_a 是指预测类别正确的像素数占总像素数的比例，其计算公式如下：

$$P_a=\frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

3.3　模型消融实验结果分析

为验证 Grid-DeepLabv3＋语义分割网络模型的有效性，分别对 3 个改进模块进行了消融实验，基准模型选择 DeepLabv3＋网络。消融实验以 $mIoU$ 作为评价指标，在 SHWD 数据集中消融实验的评估结果如表 2 所示。由表 2 中数据可知，采用 DeepLabv3＋基础语义分割模型，在 SHWD 数据集上 $mIoU$ 为 67.28%。本文提出的 Grid-DeepLabv3＋网络模型在只采用膨胀残差网络的情况下，$mIoU$ 提高了 0.83 个百分点，达到 68.11%；当同时采用膨胀残差网络和卷积注意力机制融合模块进行优化时，$mIoU$ 提高了 1.98 个百分点，达到 69.26%；当再加入图像特征网格化模块时，$mIoU$ 提高了 3.93 个百分点，达到 71.21%。实验结果表明 Grid-DeepLabv3＋语义分割网络模型在 SHWD 数据集上的分割效果有明显提升。

表 2　消融实验结果对比

模型	膨胀残差网络	卷积注意力机制融合模块	图像特征网格化模块	$mIoU$/%
DeepLabv3＋				67.28
Grid-DeepLabv3＋	√			68.11
	√	√		69.26
	√	√	√	71.21

注：√代表采用该方式。

3.4　与经典语义分割网络对比

将本文提出的 Grid-DeepLabv3＋语义分割网络模型与其他经典分割算法(如 PSPNet、Unet 等)进行对比，相关评估指标有明显的提升。如表 3 所示，在 SHWD 数据集上的对比实验结果表明，Grid-DeepLabv3＋以 89.23%的 P_a 和 71.21%的 $mIoU$ 显著优于其他语义分割算法。可视化结果进一步验证了该定量分析结论的一致性，如图 7 所示。

表 3　语义分割模型实验结果对比

方法名称	P_a/%	$mIoU$/%	R/%	参数量/M
Unet	84.88	59.54	79.21	77.65
PSPNet	85.16	66.76	83.95	85.53
DeepLabv3＋	87.02	67.28	84.12	77.72
Swin-Unet	86.72	70.32	84.46	108.50
Grid-DeepLabv3＋	89.23	71.21	88.15	69.89

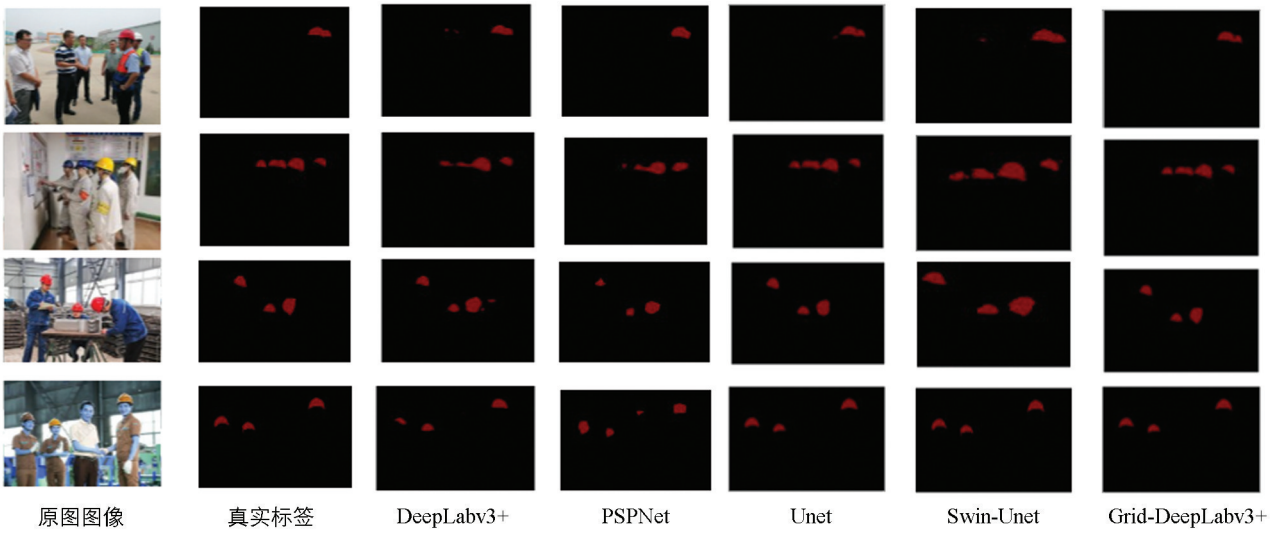


图7 各语义分割模型在SHWD数据集中的分割对比

4 结论

物流园区复杂场景下安全帽分割算法的研究，对保障物流园区的安全运行具有重要意义。本文针对目前安全帽分割存在模型识别精度低以及漏检误检等问题，提出了Grid-DeepLabv3+语义分割网络模型。该网络模型对DeepLabv3+网络模型进行了改进，在骨干网络中采用ResNet-101膨胀残差特征提取网络，相较于原来的Xception单元网络，解决了网络结构层数较深的问题，减小了计算量，提高了网络模型收敛速度；在编码阶段引入卷积注意力机制融合模块，有效增强了特征区域表征能力，进一步提升了图像语义分割的精度；在特征提取阶段构建了图像特征网格化模块，将低分辨率图像进行平均切分，更好地获取了物流园区复杂场景中局部区域的小目标特征信息。

本文提出的Grid-DeepLabv3+语义分割网络模型，相较于原始DeepLabv3+模型，在SHWD数据集上的像素准确率P_a提升了2.21个百分点，达到了89.23%。实验结果表明本算法在物流园区安全帽佩戴分割检测中效果良好，为物流园区复杂场景的图像检测分割提供了新的参考。

参考文献：

[1] 江新玲，杨乐，朱家辉，等. 面向复杂场景的基于改进YOLOX_s的安全帽检测算法[J]. 南京师大学报(自然科学版)，2023，46(2)：107-114.

[2] 张震，李浩方，李孟洲，等. 改进YOLOv4的安全帽佩戴检测方法[J]. 计算机应用与软件，2023，40(2)：206-211，273.

[3] 祁泽政，徐银霞. 改进YOLOv5s算法的安全帽佩戴检测研究[J]. 计算机工程与应用，2023，59(14)：176-183.

[4] 崔海彬，蒲东兵，陆云凤，等. 基于CA-YOLO的安全帽佩戴检测[J]. 东北师大学报(自然科学版)，2023，55(3)：94-100.

[5] 王贞，邱杭，吴斌，等. 基于CCG-YOLOv8的施工场景下安全帽佩戴检测[J]. 武汉理工大学学报，2024，46(6)：73-80.

[6] 邱明明，刘超，胡正庭. 基于Openpose-CenterNet的不停电作业人员安全防护用具穿戴智能检测研究[J]. 武汉大学学报(工学版)，2024，57(6)：829-836.

[7] 徐先峰，王轲，马志雄，等. 基于改进YOLOv4颈部优化网络的安全帽佩戴检测方法[J]. 重庆大学学报，2023，46(12)：43-54.

[8] 王晓龙，江波. 基于改进YOLOX-m的安全帽佩戴检测[J]. 计算机工程，2023，49(12)：252-261.

[9] 邓珍荣，熊宇旭，杨睿，等. 面向小目标的改进YOLOv5安全帽佩戴检测算法[J]. 计算机工程与应用，2024，60(3)：78-87.

[10] FANG Q，LI H，LUO X C，et al. Detecting Non-Hardhat-Use by a Deep Learning Method from Far-Field Surveillance Videos[J]. Automation in Construction，2018，85：1-9.

[11] REN S Q，HE K M，GIRSHICK R，et al. 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.
[12] 朱旭，马溪，姬江涛，等. 基于 Faster R-CNN 的蓝莓冠层果实检测识别分析 [J]. 南方农业学报，2020，51(6)：1493-1501.
[13] 刘雅洁，伊力哈木·亚尔买买提，席凌飞，等. 改进 YOLOv5s 的安全帽佩戴检测算法研究 [J]. 计算机工程与应用，2023，59(20)：184-191.
[14] WU J X，CAI N，CHEN W J，et al. Automatic Detection of Hardhats Worn by Construction Personnel：A Deep Learning Approach and Benchmark Dataset [J]. Automation in Construction，2019，106：102894.
[15] 张炳力，王焱辉，潘泽昊，等. 基于障碍物和车位检测的单阶段多任务 YOLO-Parking 算法研究 [J]. 合肥工业大学学报(自然科学版)，2024，47(1)：1-6，61.
[16] 田乔鑫，孔韦韦，滕金保，等. 基于并行混合网络与注意力机制的文本情感分析模型 [J]. 计算机工程，2022，48(8)：266-273.
[17] 赵志宏，杨绍普，马增强. 基于卷积神经网络 LeNet-5 的车牌字符识别研究 [J]. 系统仿真学报，2010，22(3)：638-641.
[18] 戴天虹，刘超. 基于改进 EfficientDet 的雪豹红外相机图像检测方法 [J]. 哈尔滨理工大学学报，2023，28(2)：108-116.
[19] 刘歆，吴小倩. 轻量化 YOLOv4 的安全帽佩戴检测方法 [J]. 重庆邮电大学学报(自然科学版)，2023，35(4)：671-679.
[20] 李慧琴，宋赵铭，刘存祥，等. 基于 YOLOv8n 的番茄果实检测模型改进 [J/OL]. 河南农业大学学报，(2024-05-13) [2025-05-27]. https：//doi.org/10.16445/j.cnki.1000-2340.20240511.002.
[21] 徐守坤，倪楚涵，吉晨晨，等. 基于 YOLOv3 的施工场景安全帽佩戴的图像描述 [J]. 计算机科学，2020，47(8)：233-240.
[22] 徐先峰，赵万福，邹浩泉，等. 基于 MobileNet-SSD 的安全帽佩戴检测算法 [J]. 计算机工程，2021，47(10)：298-305，313.
[23] 王玲敏，段军，辛立伟. 引入注意力机制的 YOLOv5 安全帽佩戴检测方法 [J]. 计算机工程与应用，2022，58(9)：303-312.
[24] 吕宗喆，徐慧，杨骁，等. 面向小目标的 YOLOv5 安全帽检测算法 [J]. 计算机应用，2023，43(6)：1943-1949.
[25] 张锦，屈佩琪，孙程，等. 基于改进 YOLOv5 的安全帽佩戴检测算法 [J]. 计算机应用，2022，42(4)：1292-1300.
[26] 杨大为，张成超. 基于改进 YOLOv7 的安全帽佩戴检测算法 [J]. 沈阳理工大学学报，2024，43(1)：16-21.
[27] 冯勇，杨思卓，徐红艳. 基于 YOLOv8 的轻量化安全帽佩戴检测算法 [J]. 计算机应用，2024，44(S2)：251-256.
[28] HE K M，ZHANG X Y，REN S Q，et al. Deep Residual Learning for Image Recognition [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30，2016，Las Vegas，NV，USA. IEEE，2016：770-778.
[29] 汤先美，王春宇，闫顺丕，等. 基于残差和卷积神经网络的生猪图像分块压缩感知 [J]. 东北农业大学学报，2023，54(10)：70-78.

责任编辑　柳剑