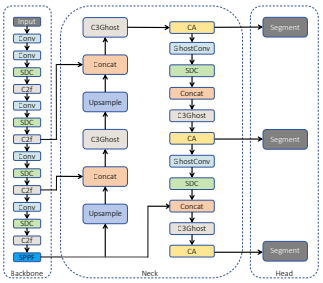
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 5 月, 第 53 卷, 第 5 期

基于空间深度转换的露天矿道路可行驶区域分割模型



赵瑞祥[1,3], 顾清华[2,3*], 李少博[3], 李滨洋[2]

[1]西安建筑科技大学信息与控制工程学院, 陕西 西安 710055;

[2]西安建筑科技大学资源工程学院, 陕西 西安 710055;

[3]西安建筑科技大学, 西安市智慧工业感知计算与决策重点实验室, 陕西 西安 710055

摘要: 露天矿区非结构化路面环境复杂多变, 特别是在中远距离视角下, 车辙、水坑等小尺度区域在图像中呈现为低分辨率细节目标, 其纹理与边界信息在特征提取过程中严重丢失, 极易导致小目标漏检与分割边缘模糊, 对无人驾驶车辆的路径规划安全构成严峻挑战。针对该问题, 提出一种基于空间深度转换的露天矿卡行车可行驶区域实例分割模型(SCG-YOLO)。首先构建空间到深度变换与可变形卷积(space-to-depth and deformable convolution, SDC)模块, 通过空间深度解耦重组保留车辙纹理、水坑边缘等细粒度信息, 再结合可变形卷积的动态采样特性, 适配非结构化路面中目标的非刚性形变, 加强模型对路面纹理细节和地形边缘的感知能力。其次引入坐标注意力(coordinate attention, CA)机制, 通过抑制无关信息并增强车辙、水坑等关键区域的特征响应, 进一步改善分割掩膜边缘清晰度与区域完整性。最后采用 Ghost 模块对 Neck 层进行轻量化重构, 作为工程约束以确保模型在维持高精度的前提下满足实时性需求。基于现场采集的露天矿区道路数据集进行实验验证, 结果表明: 模型 mAP50 达到了 82.6%, 较基准模型提高了 11.7%, 且分割边缘清晰度得到改善; 同时保持 425.31 f/s 实时性, 为露天矿复杂环境下无人驾驶提供了可行的环境感知方案。

关键词: 露天矿; 可行驶区域; 空间深度转换; 可变形卷积

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250216 | CSTR: 32245.14.oee.2026.250216 | 中图分类号: TP391 | 文献标识码: A

赵瑞祥, 顾清华, 李少博, 等. 基于空间深度转换的露天矿道路可行驶区域分割模型 [J]. 光电工程, 2026, **53**(5): 250216

Zhao R X, Gu Q H, Li S B, et al. Drivable area segmentation model for open-pit mine roads based on spatial depth conversion[J]. *Opto-Electron Eng*, 2026, **53**(5): 250216

1 引 言

随着露天矿区自动驾驶技术[1]的快速发展, 准确而高效地分割矿卡可行驶区域成为提升矿区作业安全性和效率的关键。但露天矿区非结构化道路和复杂环境如粉尘、光照变化及多样障碍物[2]对现有图像分割方法构成挑战, 因此亟需对露天矿区非结构化道路可行驶区域的精准、高效分割方法进行研究。

在结构化道路场景中, 基于卷积神经网络(convolutional neural network, CNN)的方法已取得显著进展。近年来, 众多学者已在特征鲜明、结构简化、表面平滑且边缘清晰的城市结构化道路领域内, 成功实现了道路可行驶区域的精确分割, 如徐卓栋[3]通过优化 U-Net 架构实现了在晴、雨、夜天气条件下对城市道路的精准语义分割。王云艳等人[4]提出一种改进型 DeepLab v3 的城市道路图像语义分割算法, 改善了语义分割过程

收稿日期: 2025-07-22; 修回日期: 2026-02-08; 录用日期: 2026-02-11; 网络出版日期: 2026-05-25

基金项目: 国家自然科学基金 (52074205, 52374135); 陕西省金属矿智能开采理论及技术创新团队项目 (2023-CX-TD-12)

*通信作者: 顾清华, qinghuagu@126.com

版权所有©2026 中国科学院光电技术研究所

中的误分割问题以及分割边界的不连续性状况。徐博文等人[5]提出一种基于改进 SOLO 网络的城市道路场景实例分割方法，可提升无人驾驶技术在城市道路交通场景下对实例分割的精确度。

然而，上述方法依赖清晰道路标识与稳定环境特征，难以适应矿区非结构化场景的复杂多变。这一难题主要源于非结构化道路所独有的几方面特性：1) 此类道路缺乏清晰可辨的车道线标识，道路边界模糊，形状多变且不规则，从而极大地增加了道路感知与分割的难度；2) 道路路面的质量普遍较低，周边自然环境原始且多变，呈现出显著的差异性，进一步加剧了道路场景理解与分割的复杂性；3) 路面上常见的车辙、水坑等障碍物不仅影响车辆的行驶稳定性，还可能对行车安全构成严重威胁，这些复杂因素均对道路分割算法的性能提出了更高要求。

部分学者在非结构化道路场景下的分割任务中已取得了一定的研究成果与成效。宋亮等人[6]提出一种改进 SOLOv2 的非结构化道路图像实例分割方法，通过优化 SOLOv2 架构提升了非结构化道路场景下目标实例分割的精确度。孟庆宽等人[7]将混合扩张卷积融入特征提取网络，增加感受野，增强了对复杂道路场景的分割准确性。Yang 等人[8]提出一种名为 MRNet 的深度学习框架，通过多尺度特征融合与注意力机制，实现了多场景非结构化道路的可行驶区域精准检测。

以上方法在不同程度上提升了非结构化道路的整体分割精度，但在处理露天矿区特定场景时仍面临严峻挑战。核心问题在于，矿区车载摄像头采集的图像中，车辙、水坑等关键路面要素往往位于中远距离，呈现出低分辨率、小尺度且纹理特征微弱的特点。现有的主流分割网络多采用跨步卷积或池化操作进行下采样，降低计算量的同时导致细粒度空间信息严重丢失。在此过程中，车辙的浅层纹理和水坑的不规则边缘极易被抑制，造成识别漏检或分割边缘与实际地形严重不符，难以满足矿卡对厘米级边缘感知的安全需求。

语义分割与实例分割是计算机视觉的核心任务，前者在于对图像中每个像素进行类别标注，后者旨在精确区分并勾勒出图像中各个独立对象的轮廓。在无人驾驶场景中，非结构化环境下的障碍物往往呈现出显著的个体差异性 (如形态、尺度、姿态的多样化)，而自动驾驶系统的决策与控制过程高度依赖个体级的空间特征信息。鉴于此，相较于仅能实现类别划分的语义分割，实例分割更能满足无人驾驶任务对环境感知的精细化需求，故本文以实例分割为核心技术手段，实现露天矿道路可行驶区域的识别。

实例分割分为两阶段与单阶段方法。两阶段实例分割[9]在处理实时性高的露天矿卡可行驶区域分割时，因速度慢、精度低且计算成本高而受限。而单阶段方法通过单次前向传递同步完成分类、检测与掩膜生成，具有效率高、模型小、易部署等优势。Daniel 等人[10]提出的 YOLACT 作为首个实现单阶段实例分割的模型，通过分解任务提升了推理效率与性能。后续 YOLO 系列[11-14]亦取得显著进展[15]。但在露天矿区等复杂场景中，YOLO 模型[16]存在小目标边缘感知与掩膜精度不足的问题。

基于上述分析，本文在 YOLOv8-seg 框架上进行针对性改进，围绕中远距离低分辨率和小尺度区域的精细实例分割这一核心任务，提出一种基于空间深度转换的露天矿卡可行驶区域实例分割模型 (SCG-YOLO)。首先设计并嵌入 SDC 模块，通过分离特征图的空间维度与通道维度，结合可变形卷积，实现了细粒度特征提取，在降低分辨率损失的同时，强化了对低质量图像中复杂区域的掩膜生成能力。其次引入 CA 机制，增强分割模型的特征提取与表达能力，使其能够更有效地捕捉矿区复杂场景中的关键信息。最后将原网络 Neck 层中的 Conv 和 C2f 模块替换为 Ghost 模块从而降低模型的复杂度，提升运行效率。为了更好地适应露天矿区的复杂恶劣环境，本文将露天矿分割目标划分为行车可行驶区域、车辙和水坑 3 类，以便更精细地刻画矿区场景。

2 YOLOv8 实例分割模型

YOLOv8 通过整合前沿算法模块并进行深度优化，显著提升模型性能与泛化能力。该模型在边缘计算与工业场景中具有优异表现[17]，其架构包含 n、s、m、l、x 五种规格，参数量与检测精度呈正相关，支持用户根据应用需求进行模型选型。模型采用 Backbone-Neck-Head 模块化架构 (图 1)，通过层次化特征融合机制构建高效的多尺度检测系统，其中 YOLOv8-seg 的单阶段实例分割能力可兼顾检测效率与轮廓分割精度。因此，本研究以 YOLOv8-seg 为基准框架，提出 SCG-YOLO 模型。

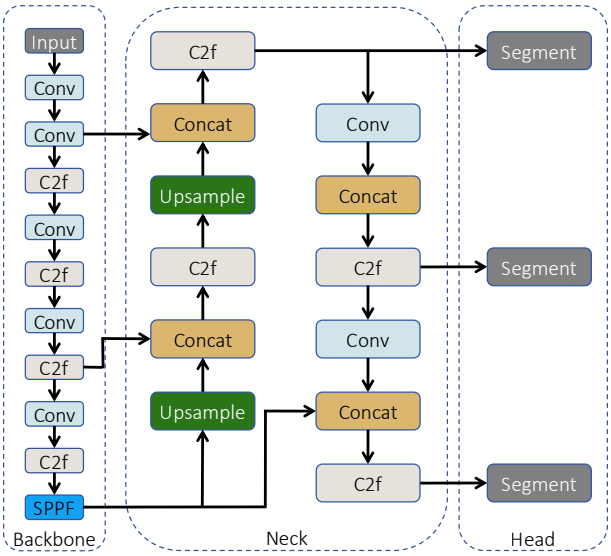


图 1 YOLOv8-seg 结构图

Fig. 1 YOLOv8-seg structural diagram

3 SCG-YOLO 模型结构

为解决露天矿区非结构化道路场景中，因中远距离成像导致的低分辨率图像下车辙、水坑等小尺度目标细粒度特征丢失及分割边缘模糊的问题，本研究在 YOLOv8-seg 实例分割模型的基础上进行改进，提出 SCG-YOLO 模型。

如图 2 所示，模型整体沿用 YOLO 系列经典的 Backbone-Neck-Head 架构，并在 Backbone 和 Neck 部分针对性集成了三个模块：SDC 细粒度特征提取模块、CA 特征增强模块以及 Ghost 轻量化模块。这些模块协同工作，旨在提升模型对复杂矿区场景下可行驶区域和小尺度目标的实例分割精度，同时保障算法的实时性，为露天矿区无人驾驶提供环境感知支撑。

3.1 SDC 细粒度特征提取模块

在露天矿区这类非结构化道路且复杂的环境中，由于作业区域广阔，所采集的图像数据通常涵盖了位于中远距离的小尺寸目标对象，这些目标因距离因素而呈现出低分辨率、小尺度的特点，且其形态常因视角、地形或遮挡而发生局部几何变形。精准分割这些对象对于自动驾驶系统的决策与路径规划至关重要，因此需要提高模型在处理低分辨率图像和小目标时的性能。

CNN[18] 架构在处理这类任务时存在局限性，一方面，步长卷积和池化操作导致的细粒度信息丢失严重影响了对低分辨率图像和小目标的分割能力；另一方面，标准卷积固定的几何结构难以有效适应复杂场景中目标可能出现的形变。为此，本文通过将空间到深度变换与非步长卷积 (space-to-depth+non-strided convolution, SPD-Conv)[19] 模块中的非步长卷积改进为可变性卷积，构建

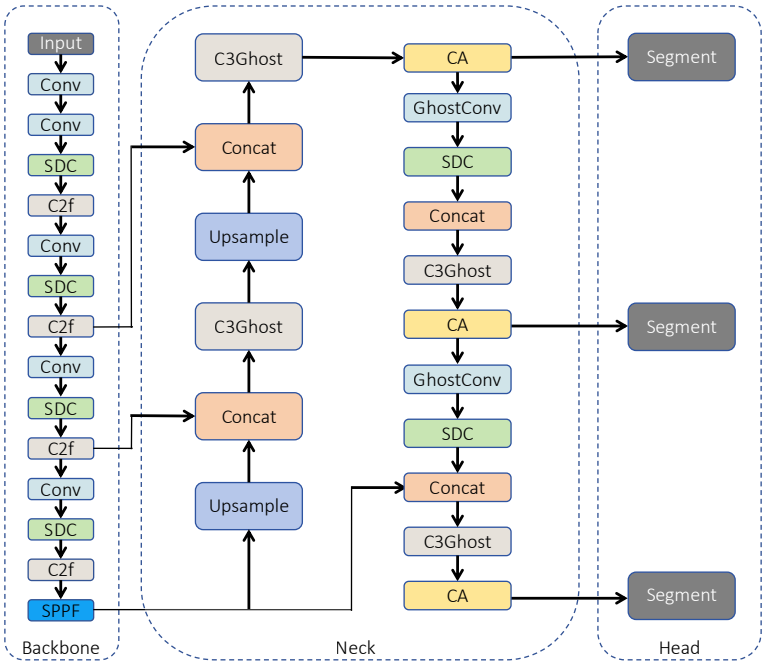


图 2 SCG-YOLO 模型结构图

Fig. 2 SCG-YOLO model structure diagram

了 SDC 模块。该模块利用空间维度与通道维度的解耦重组，并结合可变形卷积强大的几何建模能力，从而实现了露天矿区道路特征的高效层次化深度表达，提升了对低分辨率图像中小目标 (尤其是非规则形态目标) 的分割精度。

空间到深度 (Space-to-depth，SPD) 层：空间到深度转换层的作用是将输入特征图的空间维度 (高和宽) 分解重组为深度维度 (通道数)，从而增加特征图的深度。具体来说，假设输入特征图 x 的尺寸为 $S\times S\times C_1$，其中 S 是空间尺寸，C_1 是通道数。$scale$ 是尺度因子。通过空间到深度转换，特征图 x 会被拆分为 $scale^2$ 个子特征图，每个子特征图的空间尺寸减小为 $\frac{S}{scale}\times\frac{S}{scale}\times C_1$，而通道数增加，随后，这些子特征图沿通道维度进行拼接，得到空间到深度转换后的中间特征图 x'，尺寸为 $\frac{S}{scale}\times\frac{S}{scale}\times scale^2C_1$。例如，当尺度因子 $scale=2$ 时 (如图 3 所示)，输入特征图经过下采样后得到 4 个子特征图，每个子特征图的尺寸为 $\frac{S}{2}\times\frac{S}{2}\times C_1$。然后，这些子特征图沿着通道维度进行拼接，得到新的特征图，其尺寸为 $\frac{S}{2}\times\frac{S}{2}\times 4C_1$。这样，空间维度被压缩，而深度维度被扩展，从而保留了更多的特征信息。

可变形卷积 (Deformable convolution) 层：在空间到深度转换之后，SDC 模块引入可变形卷积层以取代标准的非步长卷积层。可变形卷积的核心创新在于其卷积核的采样位置是可学习的偏移量 (offset) 动态调节的 (如图 4 所示)，而非固定于规则的矩形网格。这使得卷积操作能够适应图像中目标物体的几何形变与空间分布的非规则性。具体地，可变形卷积层作用于 SPD 层的输出特征图，通过学习额外的偏移域 (offset field)，引导卷积运算在更灵活的采样点上提取特征。相比于固定卷积核，可变形卷积在维持空间分辨率的前提下，改善了模型对露天矿区非结构化道路场景中复杂目标形态 (如不规则轮廓、遮挡、视角变化) 的建模能力，并能更精确地聚焦于目标的关键部位，有效缓解了固定结构卷积在处理此类复杂场景时潜在的特征提取不充分和表征力不足的问题。

可变形卷积输出的特征公式如式 (1) 所示：

$$y(p_0)=\sum_{k=1}^{k} w_k\times x'(p_0+p_k+\Delta p_k)\Delta m_k\,, \tag{1}$$

式中：x'表示空间到深度转换后的中间特征图，同时作为可变形卷积层的输入；y 表示经过可变形卷积处理后的输出特征图；k 表示卷积核采样点数量；w_k 表示学习权重；p_0 表示卷积操作的中心点；p_k 表示预先指定偏移量；Δp_k表示像素点 k 的可学习偏移值；Δm_k表示像素点 k 的调制权重，取值范围为 [0, 1]。

SDC 模块被嵌入到主干网络与 Neck 层，用于重构多尺度特征传递路径。浅层通过空间深度转换侧重保留车辙纹理、水坑边缘等细节信息；深层通过可变形卷积的大感受野捕捉目标全局形态。通过空间深度转换减少跨尺度融合时的细节损失，同时利用可变形卷积对齐不同层级的小目标特征分布，将深层语义信息与浅层边缘信息精准匹配，重点强化车辙、水坑等小尺度危险区域的轮廓表征能力，SDC 模块缓解传统池化和步长卷积导致的小目标边缘模糊问题，成为模型解决中远距离低分

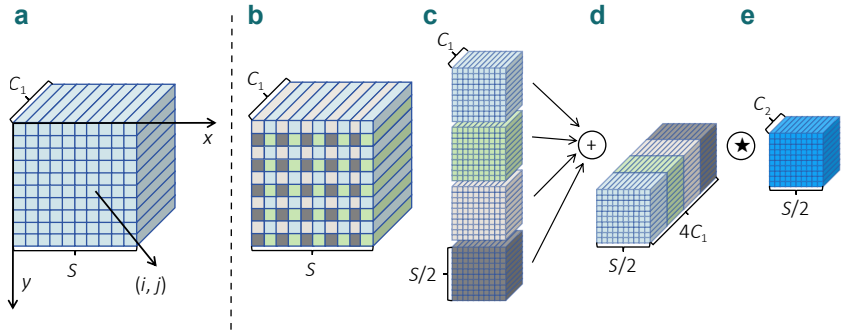


图 3 尺度因子为 2 时的 SPD-Conv 示意图。(a) 标准特征图；(b) 空间到深度操作过程；(c) 空间到深度操作后的输出特征图；(d) 空间到深度操作后采用的非步长卷积层；(e) 经步长为 1 卷积处理后的输出特征图

Fig. 3 Schematic diagram of SPD-Conv with a scale factor of 2. (a) Standard feature map; (b) Process of the space-to-depth operation; (c) Output feature map after the space-to-depth operation; (d) Non-strided convolution layer after the space-to-depth operation; (e) Output feature map after convolution with a stride of 1

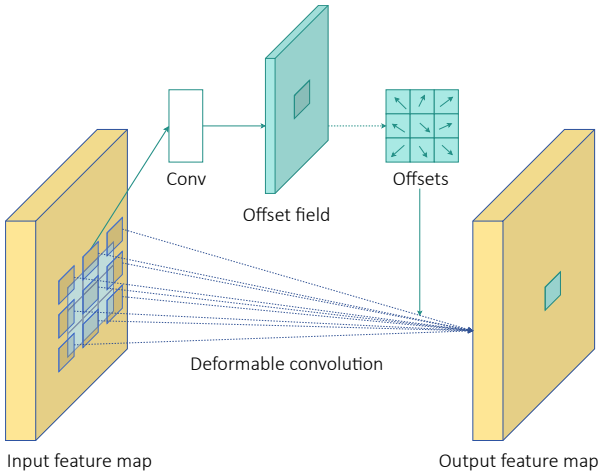


图 4 可变形卷积工作原理

Fig. 4 Working principle of deformable convolution

辨率和小尺度区域细粒度特征丢失核心问题的关键技术支撑。

3.2 CA 坐标注意力特征增强模块

针对露天矿背景复杂导致的水坑、车辙边缘分割模糊问题，注意力机制通过动态提升可行驶区域 (如路面纹理、边界和安全路径) 的特征权重，抑制背景干扰 (如阴影、矿石堆积和无关设备)，提升模型在复杂露天矿环境中的判别能力。主流的注意力机制方法有轻量级注意力机制 (shuffle attention，SA)[20]、卷积块注意力模块 (convolutional block attention module，CBAM)[21]、高效多尺度注意力机制 EMA[22] 及多头自注意力机制 (multi-head self-attention, MHSA)[23] 等。综合考虑露天矿区特殊环境、算法复杂度与特征增强效果，本文在分割模型的 Neck 层引入 CA[24] 机制，其核心目标是在 SDC 模块捕捉到的细粒度特征基础上，进一步强化关键目标的特征响应，帮助模型更鲁棒地应对光照变化、地形不平整等复杂工况，减少小目标的误检与漏检，保障无人车路径规划的安全性与可靠性。

CA 作为一种聚焦于空间坐标的注意力机制，其核心目标在于强化特征表达效能。它能够基于输入特征的空间位置信息，以动态化的方式对权重予以适应性调整，实现对图像中目标区域的精确关注。通过并行地沿水平和垂直方向编码特征图，有效整合位置信息至通道注意力中，既保留跨通道交互，又增强空间敏感性，促进图像关键特征的精准定位与分割。与传统机制相比，CA 机制呈现出更为卓越的计算效率与自适应特性。其结构示意图如图 5 所示。

CA 机制的运算流程主要涵盖坐标信息嵌入以及坐标注意力生成两个关键步骤，具体过程如下所示。

坐标信息嵌入：传统方法用全局池化处理数据，会使位置信息大量流失难以保存。CA 注意力机制以一对一特征编码操作取代全局池化，沿宽度与高度方向聚合特征挖掘整合空间特征，获取含空间信息的特征图。具体是对输入特征图，在宽度方向用池化核 (1、W) 平均池化，高度方向用池化核 (H、1) 执行相同操作，以高度为 H 的第 C 通道为例，输出公式如式 (2) 所示。

$$Z_C^H(H)=\frac{1}{W}\sum\nolimits_{0\leqslant i<W}X_C(H,i), \tag{2}$$

式中：H,W,C 分别为输入特征图的高度、宽度、通道数；$X_C(H,i)$为第 C 个通道在高度位置 H、宽度位置 i 处的特征值。

第 C 通道在宽度为 W 处的输出如式 (3) 所示：

$$Z_C^W(w)=\frac{1}{H}\sum\nolimits_{0\leqslant j<H}X_C(j,W), \tag{3}$$

式中：$X_C(j,W)$为第 C 个通道在高度位置 j、宽度位置 W 处的特征值。

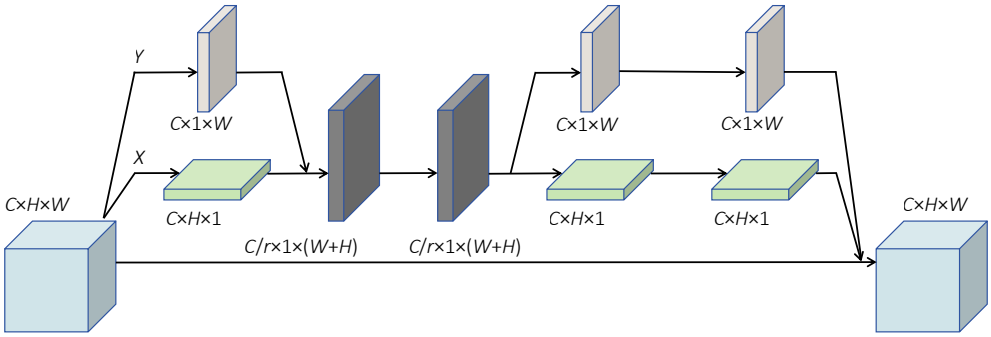


图 5 CA 注意力机制结构

Fig. 5 CA attention mechanism structure

坐标注意力生成：对所生成的特征图执行拼接操作，即融合式 (2) 与式 (3) 的输出结果，随后通过 1×1 卷积运算生成包含水平与垂直方向空间信息的特征图，为进一步提升该特征图的表达能力，对其进行批量归一化处理，并应用非线性激活函数 (δ) 进行激活，如式 (4) 所示：

$$f=\delta\left(F_1\left[z^{\mathrm{h}},z^{\mathrm{w}}\right]\right), \tag{4}$$

式中：f 为中间特征图，$f\in\mathbb{R}^{C/r\times(H+W)}$；$F_1$ 为 1×1 卷积变换函数；$[\cdot,\cdot]$ 为沿空间维度的 concatenate 操作；$z^{\mathrm{h}},z^{\mathrm{w}}$ 分别为水平与竖直方向特征图。

为捕获空间关键信息，将融合了水平与竖直方向空间信息的特征图 f 依其空间维度重新拆解成两个独立的特征张量。随后，对每个张量分别应用 1×1 卷积与 Sigmoid 激活函数 (σ) 进行处理。经过此独立处理流程，计算所得的输出即为沿水平方向的空间注意力权重 g^{h} 以及沿竖直方向的空间注意力权重 g^{w}，其表达式如下所示：

$$g^{\mathrm{h}}=\sigma\left(F_{\mathrm{h}}\left(f^{\mathrm{h}}\right)\right), \tag{5}$$

$$g^{\mathrm{w}}=\sigma\left(F_{\mathrm{w}}\left(f^{\mathrm{w}}\right)\right), \tag{6}$$

式中：f^{h}、f^{w} 分别为沿水平方向和竖直方向拆分得到的特征图，$f^{\mathrm{h}}\in\mathbb{R}^{C/r\times H}$，$f^{\mathrm{w}}\in\mathbb{R}^{C/r\times W}$；$F_{\mathrm{h}}$、$F_{\mathrm{w}}$ 分别为对应方向上的 1×1 卷积变换；$g^{\mathrm{h}}\in\mathbb{R}^{C\times H\times 1}$，$g^{\mathrm{w}}\in\mathbb{R}^{C\times W\times 1}$。

最终通过归一化和加权处理，得到输出特征图，使原始特征图获得注意力权重，提升特征聚焦能力。公式如式 (7) 所示：

$$y_c(i,j)=x_c(i,j)\times g_c^{\mathrm{h}}(i)\times g_c^{\mathrm{w}}(j), \tag{7}$$

式中：$x_c(i,j)$ 和 $y_c(i,j)$ 分别表示为第 c 个通道在空间位置 (i,j) 处的输入特征值和输出特征值；i、j 为像素坐标值；c 为通道索引；$g_c^{\mathrm{h}}(i)$ 和 $g_c^{\mathrm{w}}(j)$ 分别表示第 c 个通道沿水平和竖直方向的空间注意力权重。

相较于 CBAM 模块局限于捕获局部空间相关性而难以建模全局依赖关系，SE 模块 (squeeze-and-excitation module) 聚焦通道信息编码却忽视空间位置关系，而 CA 模块通过强化关键区域特征权重，抑制矿石堆积、粉尘阴影等背景干扰，在特征信息提取的全面性上表现较好。

3.3 Ghost 轻量化模块

虽然 SDC 与 CA 模块改善了中远距离低分辨率和小尺度区域的分割性能，但也不可避免地增加了模型的计算复杂度。考虑到车载域控平台的算力约束，本节在 Neck 层引入 GhostNet 方法进行轻量化设计，通过特征冗余替代策略降低计算量与参数量，其核心是利用廉价线性变换生成“幽灵特征图”，在尽量维持路面纹理、障碍物边缘等关键识别精度的同时压缩卷积计算成本。这使得模型能高效部署于车载嵌入式设备，满足露天矿区复杂环境下的实时分割需求。

GhostNet 提出的 Ghost 卷积模块结合传统卷积与线性运算，利用了特征图冗余性。GhostConv 模块先通过标准卷积压缩通道数，生成少量固有特征图，再对固有特征图实施廉价线性变换 Φ_i (i=1,2,...,s-1)，生成 Ghost 特征图，最后将对应的固有特征图与 Ghost 特征图拼接得到最终特征图，实现高效特征提取。GhostConv 模块如图 6 所示，其中 Identity 表示恒等映射，即保留标准卷积得到的固有特征图。

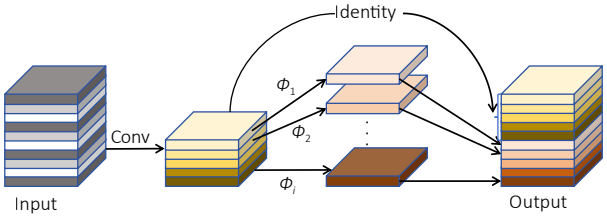


图 6 GhostConv 模块

Fig. 6 GhostConv module

假设输入特征图的尺寸大小为 $H\times W\times C$，输出特征图尺寸大小为 $h'\times w'\times n$，其中 H、W、C 分别表示输入特征图的高度、宽度、通道数，h'、w'、n 分别表示输出特征图的高度、宽度、通道数，采用的卷积核尺寸大小为 $k\times k$，线性变换卷积核尺寸大小为 $d\times d$，经过 Φ_i 操作生成的 Ghost 特征图连同原固有特征图，共得到 s 个相关特征图。则传统卷积与 GhostNet 卷积所需参数量分别为式 (8) 和式 (9)：

$$P_{\mathrm{Conv}}=n\times h'\times w'\times C\times k\times k, \tag{8}$$

$$P_{\mathrm{GhostConv}}=\frac{n}{s}\times h'\times w'\times C\times k\times k+(s-1)\times\frac{n}{s}\times h'\times w'\times d\times d, \tag{9}$$

式中：P_{Conv} 表示传统卷积所需的参数量；$P_{\mathrm{GhostConv}}$ 表示 GhostConv 所需的参数量；n/s 表示标准卷积生成的固有特征图数量；$(s-1)\times n/s$ 表示通过廉价线性变换生成的

Ghost 特征图数量。

那么传统卷积操作的参数量与 Ghost 卷积操作的参数量的比值 r_c 如式 (10)：

$$r_c=\frac{C\times k\times k}{\frac{1}{s}\times C\times k\times k+\frac{s-1}{s}\times d\times d}\approx\frac{s\times C}{s+C-1}\approx s. \tag{10}$$

从上述公式可知，Ghost 卷积操作的参数量可近似等于传统卷积操作的 1/s。

将 Bottleneck 替换成 GhostBottleneck 插入 C3 模块形成 C3Ghost 模块，如图 7 所示。

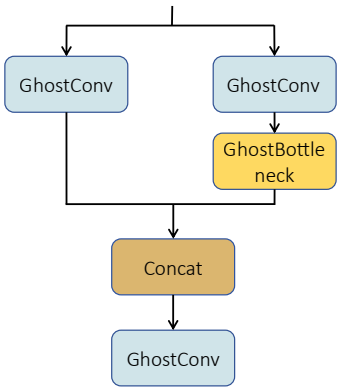


图 7 C3Ghost 模块

Fig. 7 C3Ghost module

因此将分割模型网络 Neck 层中的 Conv 模块和 C2f 模块分别替换成 GhostConv 模块和 C3Ghost 模块，模型在保持性能的同时实现了计算量和参数量的有效缩减，而且通过减少冗余计算，提高了网络的运行效率。

4 实验结果与分析

4.1 实验环境

本实验采用服务器操作系统为 Ubuntu 22.04，显卡为 NVIDIA GeForce RTX 3090(24G)，编程语言为 Python 3.11；深度学习环境为 CUDA 11.8、PyTorch 2.1.0。模型训练参数设置：输入图片尺寸为 640 pixel×640 pixel，批处理量 (batchsize) 为 16，训练周期 (epochs) 为 100 次，50 个训练周期无提升后自动停止，其余参数为默认值。

4.2 实验数据集构建及评价指标

本实验中的数据集采集于大型露天开采矿现场，使用 Basler 相机采集。分割目标包括露天矿卡行车可行驶区域、车辙和水坑 3 类。在多样化的季节变化与各异的光照环境条件下，共计收集并获取了 565 张图像数据样本。本实验采用 labelme 标注工具将其精细标注后转换成 YOLO 格式的数据集，通过 Python 代码实现随机抽样，将 90% 的样本划分为训练集用于模型训练，10% 作为验证集验证模型训练情况，图像原分辨率是 2448 pixel×2048 pixel，输入至网络前，将图像尺寸调整为 640 pixel×640 pixel。

针对本实验中实例分割任务有以下常用评价指标可对算法的结果进行评估。

精确度 (P) 和召回率 (R) 的计算公式如式 (11) 和式 (12) 所示：

$$P=\frac{T_{\rm TP}}{T_{\rm TP}+F_{\rm FP}}, \tag{11}$$

$$R=\frac{T_{\rm TP}}{T_{\rm TP}+F_{\rm FN}}, \tag{12}$$

式中：$T_{\rm TP}$(true positive, TP) 表示正确分割出图像中的待分割目标且 IoU 大于阈值；$F_{\rm FP}$(false positive, FP) 表示没有正确分割出图像中的待分割目标且 IoU 小于阈值；$F_{\rm FN}$(false negative, FN) 表示没有分割出图像中的目标。

平均精度 mAP 的计算公式如式 (13) 所示：

$$mAP=\frac{1}{N}\sum\nolimits_{n=0}^{N}\int_0^1 P_n(r)\,{\rm d}r, \tag{13}$$

式中：N 为数据集中分割目标的类别个数；P_n为某一类别的平均精度 (average precision，AP)。AP 表示每个类别的精确度-召回率 (precision-recall，P-R) 曲线所围出的面积，该面积与模型性能优劣呈正相关关系。mAP50 是将 IoU 阈值设为 0.5，计算每一类图片的 AP，最后将所有类别作平均。mAP50-95 是指 IoU 阈值从 0.5 至 0.95，以 0.05 为步长时，多个 AP 的平均值。

参数量 (Parameter) 的值越小表示参数越少、模型越轻便，使用的计算资源越少；浮点运算次数 (floating point operations，FLOPs) 值越低表明计算时所需的计算资源越少，计算复杂度也越低。帧速 (frames per second, FPS) 表示模型每秒能处理的图片数量，单位为 f/s，用于衡量模型分割速度及实时性。

4.3 结果分析

4.3.1 规模模型对比

为了更深入地了解 YOLOv8 分割模型针对此场景在不同规模下的性能表现，本实验特别设计了针对 n、s、m、l、x 五种不同规模模型的对比实验。这五种规模模型分别代表了从轻量级到超大规模的不同配置。其不同规模性能差异如表 1 所示。

表 1 YOLOv8-Seg 不同规模对比

Table 1 Comparison of YOLOv8-Seg at different sizes

Model	P_B/%	R_B/%	mAP50/%		mAP50-95/%		Parameters/M	FLOPs/G	FPS
			Box	Mask	Box	Mask			
YOLOv8n-seg	68.4	60.0	62.7	57.4	40.3	32.8	3.3	12.0	495.72
YOLOv8s-seg	74.3	67.7	70.9	67.7	45.9	41.1	11.8	42.4	402.94
YOLOv8m-seg	77.6	71.1	76.0	72.9	49.2	45.3	27.2	110.0	221.76
YOLOv8l-seg	79.7	73.7	78.3	75.3	51.1	47.0	46.0	220.1	159.96
YOLOv8x-seg	81.1	75.2	79.8	77.2	52.0	47.9	71.7	343.7	100.86

实验结果清晰地展示了模型规模与性能之间的正相关性，尤其是在关键指标 mAP50 上的显著差异。具体而言，尽管 YOLOv8s-seg 在 mAP50 这一阈值上的表现略差于 YOLOv8m-seg，但其复杂度显著降低，且帧速率远高于后者，展现出了更为优越的计算效率。在全面权衡实时处理能力、分割精度以及模型轻量化等实际应用中的需求后，选择 YOLOv8s-seg 作为本实验的分割基准模型。

4.3.2 注意力机制实验

针对露天矿区路面分割任务中，矿石纹理、设备阴影等背景干扰与可行驶区域特征存在强耦合性，传统特征编码网络难以实现空间相关性较弱的干扰信息与有效目标特征的精准分离等问题，前文 3.2 节提出了引入 CA 机制的改进策略。为验证 CA 机制的有效性与合理性，并进一步量化其特征聚焦效果，设计 CBAM、MHSA 和 CA 机制三种注意力模块的对比实验，并分别将它们集成到分割基准模型中，在统一的自制数据集与训练环境下开展性能测试，不同注意力模块对网络性能的影响差异如图 8 所示。

实验结果显示，CA 机制在提升模型性能方面展现出了一定的优势。在 mAP50 这一关键评估指标上，该模型的表现优于在相同条件下分别应用了 MHSA 和 CBAM 的模型，相比 MHSA 提升了 5.6%，相比 CBAM 提升了 1.7%。这一结果验证了引入 CA 机制的有效性。

为直观验证 CA 机制的特征聚焦效果，通过在模型推理阶段设置 visualize=True 参数，调用 YOLOv8 框架内置的注意力可视化接口，对图 9 所示的露天矿道路原图生成对应的注意力热力图，如图 10 所示。

图 10 本质上是特征响应强度的伪彩色映射图。图中色彩的变化直观反映了模型对不同区域的关注程度。颜色越接近暖色，表示该位置的注意力权重越高；颜色越接近冷色，表示该位置的注意力权重越低。图中高响应区域主要分布在车辙纹理、水坑轮廓及道路边界等关键可行驶区域，而低响应区域则集中在矿石堆积、车辆阴影等背景区域。由此可见，CA 机制能够提升关键区域特征权重、抑制无关干扰信息，从而有助于提高分割边缘的清晰度与整体分割精度。这一可视化结果进一步

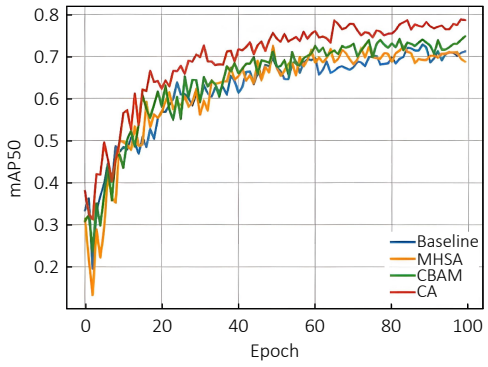


图 8 注意力机制对比

Fig. 8 Comparison of attention mechanism



图 9 露天矿道路原图

Fig. 9 Original image of open-pit mine road

验证了将 CA 作为特征增强模块的合理性。

4.3.3 主干网络对比分析

在露天矿卡行车可行驶区域实例分割任务中，输入图像常包含中远距离小目标 (如车辙、水坑边缘)，且非结构化路面纹理细节易受粉尘干扰而模糊，这对主干网络的特征提取性能提出特殊要求：既要在低分辨率条件下保留小目标轮廓信息，又需强化对复杂纹理的感知能力。分割基准模型的主干网络采用传统跨步卷积结构，在处理此类场景时存在细粒度信息丢失问题，导致分割边缘模糊与小目标漏检。对于上述问题，前文提出了 SDC 模块。本节设计对比实验来检验该结构设计的有效性，对分割基准模型分别引入 SDC 模块、EfficientNetV2[25] 和基于 Haar 小波的下采样技术 (Haar wavelet downsampling, HWD)[26] 三种不同的主干网络架构，以此通过使用相同自制数据集在相同训练环境下的对比分析，揭示它们对模型性能的具体影响。其影响差异如图 11 所示。

在本次实验中，EfficientNetV2 在分割任务上的 mAP50 指标并未达到预期。主要原因在于其设计架构以分类任务为核心导向，EfficientNetV2 通过复合缩放策略优化通道数与深度，特征提取过程更侧重全局语义信息的聚合，而露天矿可行驶区域分割需精准捕捉水坑边缘、车辙纹路等局部细粒度空间特征，二者任务目标的差异导致 EfficientNetV2 无法适配矿区分割需求。

HWD 通过引入小波变换中的 Haar 小波进行下采样，旨在保留图像的边缘和纹理信息，这对于分割任务至关重要。然而，实验结果显示，虽然 HWD 在理论上具有保持图像重要特征的潜力，但在实际应用中，其未能有效结合 YOLOv8 的框架优势，导致在 mAP50 上的提升并不显著。这表明在选择主干网络时，除了考虑其理论上的优势外，还需综合考虑其与特定任务框架的兼容性。

最终引入 SDC 模块的分割模型在 mAP50 上达到了 80.3%，这一成绩不仅较未改进任何主干网络的基准模型提升了 9.4%，也优于在相同条件下分别引入 EfficientNetV2 主干网络和 HWD 主干网络所得到的模型，相比改进了 EfficientNetV2 主干网络的模型提升了 10.7%，相比改进了 HWD 主干网络的模型提升了 8.9%。这表明基于空间到深度转换与可变形卷积相结合的 SDC 模块，更契合露天矿道路场景，能够更好地在下采样过程中保留小目标与边缘细节，有效缓解低分辨率条件下小目标细粒度信息丢失的核心问题，是模型精度提升的关键，验证了 3.1 节中 SDC 模块结构设计的合理性。

4.3.4 消融实验

为验证本文算法改进模块是否合理及其实际效果，以分割基准模型开始，逐步增加改进模块来进行消融实验，结果如表 2 所示，其中 SDC、CA 和 Ghost 分别代表相应的改进模块。

SDC 模块通过可变形卷积学习目标形变偏移量，使采样点能动态聚焦于车辙边缘、水坑轮廓等关键区域，有效解决了 SPD-Conv 因固定卷积结构导致的形变适配不足问题，尽管其参数量和 FLOPs 有所增加，但精度提升显著，mAP50 达到 80.3%，表明 SDC 模块能有效解决低分辨率下小目标细粒度信息丢失问题。在 SDC 模块基础上引入 CA 注意力机制，mAP50 进一步提升至 82.5%，表明 CA 有效帮助模型在保留下来的丰富特征中筛选出关键信息，抑制了背景干扰。

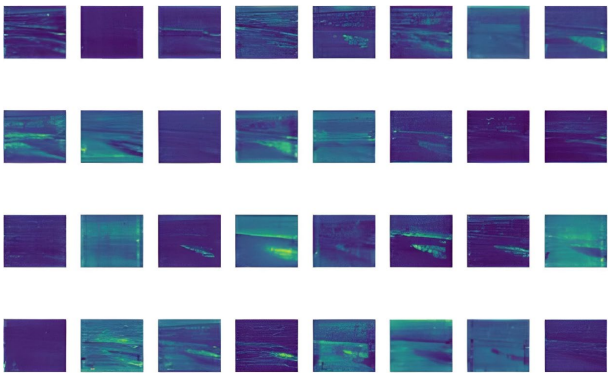


图 10 CA 注意力机制特征权重热力图

Fig. 10 Feature Weight Heatmap of CA attention mechanism

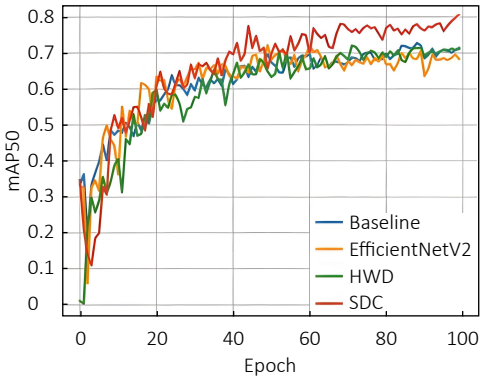


图 11 主干网络对比

Fig. 11 Comparison of backbone network

表 2 消融实验对比

Table 2 Comparison of ablation experiments

Experiments	SPD-Conv	SDC	CA	Ghost	mAP50/%		mAP50-95/%		Parameters/M	FLOPs/G	FPS
					Box	Mask	Box	Mask			
1					70.9	67.7	45.9	41.1	11.8	42.4	402.94
2	√				76.2	72.3	49.2	44.6	12.2	49.8	398.21
3		√			80.3	75.9	51.8	47.3	13.6	79.5	379.31
4		√	√		82.5	77.8	53.2	48.8	13.7	81.9	375.99
5		√	√	√	82.6	78.1	53.5	49.0	11.2	42.7	425.31

最终引入 Ghost 模块，在保持 mAP50 为 82.6% 高精度的同时，参数量降至 11.2 M，FPS 提升至 425.31 f/s，证明 Ghost 轻量化设计成功抵消了 SDC 和 CA 模块带来的计算负担，在保持精度的前提下满足矿区实时性需求。

4.3.5 泛化性实验

为了进一步验证 SCG-YOLO 模型在不同非结构化场景下的鲁棒性与泛化能力，并排除模型对特定自制矿区数据集产生过拟合的可能性，本研究选取了 RUGD (Robotic Urban Ground Dataset) 公开数据集进行实验。RUGD 数据集包含大量复杂的野外环境图像，涵盖土路、碎石路、水域和草地等多种非结构化地形，与露天矿区路面在视觉特性与分割难点上具有较高相似性，同样存在路面不平整、边界模糊及纹理复杂等挑战，因此可用于检验模型在数据分布变化条件下的鲁棒性。

本实验选取 RUGD 数据集中与矿区场景最为相关的土路、碎石及水域类别进行训练与测试，保持与前述实验相同的环境配置。SCG-YOLO 模型与基准模型 YOLOv8s-seg 在 RUGD 数据集上的性能对比结果如表 3 所示。

由表 3 可知，SCG-YOLO 模型在 RUGD 数据集上的各项指标均优于基准模型 YOLOv8s-seg，其中 SCG-YOLO 的 mAP50 达到 71.1%，较基准模型提升了 5.2%。虽然受场景差异、类别定义与标注规范不一致等因素影响，模型在 RUGD 上的分割精度较自建数据集有所下降，但其相较于基准模型的性能优势，表明模型能够有效捕捉非结构化道路环境的共性特征，也证明了模型并非仅对自制矿区数据集有效，在跨场景、跨数据分布条件下仍具备稳定的性能增益，具有一定的泛化性能。

4.3.6 不同算法性能对比实验

为验证本文算法的有效性和优势，对比评估本文算法与多种现有实例分割算法在相同自制数据集和实验环境下的性能。结果如表 4 所示。

其中分割基准模型、YOLOv11s-seg 分割模型与改进后的 SCG-YOLO 实例分割模型的分割预测效果可视化对比如图 12 所示。

由表 4 的实验结果可知，在自制露天矿区数据集中，与经典实例分割模型 Mask R-CNN 和 YOLACT 相比，本文所提出的算法不仅在模型轻量化程度与计算效率层面表现出一定的优势，同时在 mAP50 (Box) 和 mAP50 (Mask) 上也实现了对传统模型的超越。在与 YOLO 系列早期版本，诸如 YOLOv5s-seg 和 YOLOv6s-seg 的比较中，尽管 YOLOv5s-seg 的参数量与浮点运算次数 (FLOPs) 略低于本文算法，但本文算法在 mAP50 指标上相较于 YOLOv5s-seg 有着显著提升。此外，本文算法在综合性能上全面优于 YOLOv6s-seg。而在与 YOLO 系列较新版本 YOLOv11s-seg 对于特定的露天矿非结构化道路场景和自建数据集的对比中，虽然 YOLOv11s-seg 在 mAP50(Mask) 指标上取得了一定成绩，达到了 71.5%，但本文所提出的算法在分割精度上仍更高。同时，本文算法在参数量和 FLOPs 方面与 YOLOv11s-seg 相比，虽略有差异，但综合分割精度、资源消耗与

表 3 SCG-YOLO 与 YOLOv8s-seg 在 RUGD 数据集上的性能对比结果

Table 3 Performance comparison results of SCG-YOLO and YOLOv8s-seg on the RUGD dataset

Model	mAP50/%		mAP50-95/%	
	Box	Mask	Box	Mask
YOLOv8s-seg	65.9	62.8	42.6	38.1
SCG-YOLO (Ours)	71.1	67.3	46.1	42.2

表 4 不同算法性能对比

Table 4 Comparison of performance of different algorithms

Model	mAP50/%		Parameters/M	FLOPs/G
	Box	Mask		
YOLOv8s-seg (Baseline)	70.9	67.7	11.8	42.4
Mask R-CNN	68.5	65.4	46.3	239.3
YOLACT	63.4	58.6	34.8	61.8
YOLOv5s-seg	73.6	64.6	9.7	37.8
YOLOv6s-seg	71.1	66.3	16.6	56.7
YOLOv11s-seg	68.7	71.5	10.1	35.4
FastInst-R50	75.5	73.4	28.5	56.6
Mask2Former-S	79.8	77.2	52.6	286.5
SCG-YOLO (Ours)	82.6	78.1	11.2	42.7

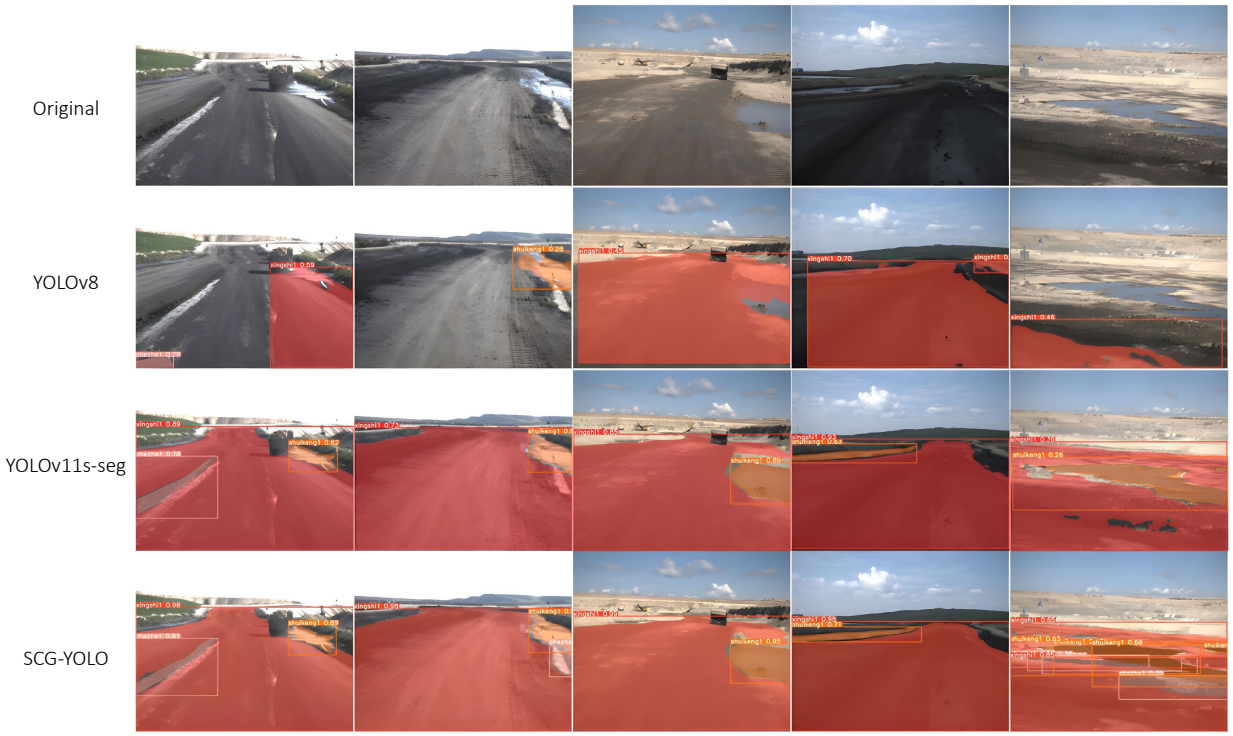


图 12 实例分割预测效果对比

Fig. 12 Performance comparison of instance segmentation prediction

实时性等多方面因素考量，本文算法在露天矿卡行车可行驶区域分割这一特定任务中，依然展现出更为优异的表现。

在与 FastInst -R50 和基于 Transformer 架构的 Mask2Former-S 分割模型的对比中，FastInst-R50 虽作为兼顾速度与精度的代表性算法，但在露天矿区这种非结构化、纹理模糊及小目标密集的特定场景下，其 mAP50 (Mask) 仍低于 SCG-YOLO 约 4.7%，且参数量约为本文模型的 2.5 倍。Mask2Former-S 依托 Transformer 的全局上下文建模能力，通常在通用数据集上表现优异，但在本实验的矿区非结构化场景中，其 mAP50 (Mask) 为 77.2%，略低于本文 SCG-YOLO 模型的 78.1%。更关键的是，Mask2Former-S 的计算代价偏高，其参数量和 FLOPs 分别达到 52.6 M 和 286.5 G，庞大的算力需求使其难以部署于资源受限的矿卡车载终端。总之，本文所提出的算法相较于其他现有模型，

不仅在分割精度上实现了提升，同时在资源消耗与实时性方面也表现出色。

图 12 第 1 列前方车辆位于道路中部并伴随扬尘干扰，导致道路表面及右侧水坑区域的纹理信息呈现出遮挡效应与边界模糊现象。在露天矿道路中远距离视角下，第 1 列、第 2 列和第 4 列场景中车辙呈现为细长条纹状、像素占比小且纹理对比度弱，以及第 3 列右侧水坑区域和第 5 列大面积积水、湿地场景中，湿滑泥面与干燥路面呈渐变过渡导致边界模糊，此类目标在图像中属于典型的小尺度目标与边界模糊目标。观察效果对比图可发现，面对复杂场景中的小目标、遮挡物或是边界模糊的情况，分割基准模型和 YOLOv11s-seg 实例分割模型的处理结果仍存在一定的局限性，如边缘不够平滑、部分细节丢失以及在某些情况下对相似目标的区分度不高，难以满足露天矿卡自动驾驶对边缘感知的高精度需求。而改进后的 SCG-YOLO 实例分割模型，有效捕获了图像中的空间位置信息与上下文依赖，通过注意力机制提升了模型对关键特征的聚焦能力，且能够对目标边界进行更精准的定位与分割。

5 结　论

针对露天矿区非结构化道路场景中，低分辨率图像下车辙、水坑等小尺度目标细粒度特征丢失及边缘模糊这一核心问题，提出了一种基于空间深度转换的实例分割模型 (SCG-YOLO)。设计并嵌入了 SDC 模块，通过空间深度解耦重组与可变形卷积的结合，有效保留了车辙纹理、水坑边缘等细粒度信息，并自适应对齐不规则、非刚性轮廓，改善了在低分辨率、中远距离条件下对小目标的边界感知和实例分割精度。引入坐标注意力机制 CA，在复杂矿区背景下增强了对关键特征的筛选与聚焦能力，进一步提升了分割边缘的清晰度。采用 Ghost 模块进行轻量化重构，在保持细粒度分割精度基本不变的前提下，有效降低模型参数量与计算复杂度，使模型在 mAP50 达到 82.6% 的高精度下，仍保持 425.31 f/s 的推理速度，实现了对露天矿卡行车可行驶区域的实时高精度分割。

此外，本研究受限于数据集规模小、多样性欠缺，图片数量及地域覆盖不广，影响了对露天矿区复杂环境的全面理解。鉴于此，未来的研究亟需致力于构建更大规模且更具多样性的数据集，以充分反映露天矿区的多样性和复杂性。

作者贡献

赵瑞祥：研究资料的收集、分析，实验的设计与完成，论文写作；顾清华、李少博：指导论文选题、论文修改；李滨洋：整理数据。

利益冲突

所有作者声明无利益冲突

参考文献

1. Chen Z F, Yu G Z, Zhang C Y, et al. Driving risk assessment and prevention strategies for autonomous vehicle in open-pits[J]. *J China Coal Soc*, 2023, **48**(4): 1782−1797.
   陈志发, 余贵珍, 张传莹, 等. 露天矿区无人驾驶行车风险评估及防控策略仿真研究[J]. 煤炭学报, 2023, **48**(4): 1782−1797.
2. Ruan S L, Li S B, Gu Q H, et al. Road obstacle detection in open-pit mines based on bidirectional feature fusion[J]. *J China Coal Soc*, 2023, **48**(3): 1425−1438.
   阮顺领, 李少博, 顾清华, 等. 基于双向特征融合的露天矿区道路障碍检测[J]. 煤炭学报, 2023, **48**(3): 1425−1438.
3. Xu Z D. Road scene recognition technology based on U-Net network[D]. Beijing: China University of Geosciences (Beijing), 2021. https://doi.org/10.27493/d.cnki.gzdzy.2021.000304.
   徐卓栋. 基于 U-Net 网络的道路场景识别技术[D]. 北京: 中国地质大学 (北京), 2021. https://doi.org/10.27493/d.cnki.gzdzy.2021.000304.
4. Wang Y Y, Wang C Y, Wu H X, et al. Image semantic segmentation based on improved Deeplabv3[J]. *Comput Simul*, 2022, **39**(10): 148−152,158.
   王云艳, 王重阳, 武华轩, 等. 基于改进型 Deeplabv3 的城市道路图像语义分割[J]. 计算机仿真, 2022, **39**(10): 148−152,158.
5. Xu B W, Lu Y N. Urban road scene instance segmentation method based on improved SOLO network[J]. *J Jilin Univ (Sci Ed)*, 2022, **60**(6): 1356−1362.
   徐博文, 卢奕南. 基于改进 SOLO 网络的城市道路场景实例分割方法[J]. 吉林大学学报 (理学版), 2022, **60**(6): 1356−1362.
6. Song L, Gu Y H, Huang J W. Improved segmentation of unstructured road image instances in SOLOv2[J]. *Laser J*, 2024, **45**(3): 133−139.
   宋亮, 谷玉海, 黄佳伟. 改进 SOLOv2 的非结构化道路图像实例分割[J]. 激光杂志, 2024, **45**(3): 133−139.
7. Meng Q K, Yang X X, Zhang M, et al. Recognition of unstructured field road scene based on semantic segmentation model[J]. *Trans Chin Soc Agric Eng*, 2021, **37**(22): 152−160.
   孟庆宽, 杨晓霞, 张漫, 等. 基于语义分割的非结构化田间道路场景识别[J]. 农业工程学报, 2021, **37**(22): 152−160.
8. Yang J, Chen J Y, Wang Y, et al. MRNet: a deep learning framework for drivable area detection in multi-scenario unstructured roads[J]. *Electronics*, 2025, **14**(11): 2242.
9. He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[J]. *IEEE Trans Pattern Anal Mach Intell*, 2020, **42**(2): 386−397.
10. Bolya D, Zhou C, Xiao F Y, et al. YOLACT: real-time instance segmentation[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019: 9157–9166. https://doi.org/10.1109/ICCV.2019.00925.
11. Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. arXiv: 1804.02767, 2018. https://doi.org/10.48550/arXiv.1804.02767.
12. Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed

and accuracy of object detection[EB/OL]. arXiv:2004.10934, 2020. https://doi.org/10.48550/arXiv.2004.10934.

13. Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.
14. Zhao Y, Shi Y B, Wang Z L. The improved YOLOV5 algorithm and its application in small target detection[C]//*15th International Conference on Intelligent Robotics and Applications (ICIRA)*, 2022: 679–688. https://doi.org/10.1007/978-3-031-13841-6_61.
15. Diwan T, Anirudh G, Tembhurne J V. Object detection using YOLO: challenges, architectural successors, datasets and applications[J]. *Multimed Tools Appl*, 2023, **82**(6): 9243−9275.
16. Chen Y Y, Wang H, Cai Y F, et al. Efficient automatic driving instance segmentation method based on detection[J]. *Automot Eng*, 2023, **45**(4): 541−550.
陈妍妍, 王海, 蔡英凤, 等. 基于检测的高效自动驾驶实例分割方法[J]. 汽车工程, 2023, **45**(4): 541−550.
17. Gu Q H, Zhou Q, Wang D. Traveling obstacle detection in open-pit mine area based on improved YOLOv8[J]. *Gold Sci Technol*, 2024, **32**(2): 345−355.
顾清华, 周琼, 王丹. 基于改进 YOLOv8 的露天矿区行车障碍物检测[J]. 黄金科学技术, 2024, **32**(2): 345−355.
18. LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proc IEEE*, 1998, **86**(11): 2278−2324.
19. Sunkara R, Luo T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//*Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2023: 443–459. https://doi.org/10.1007/978-3-031-26409-2_27.
20. Zhang Q L, Yang Y B. SA-net: shuffle attention for deep convolutional neural networks[C]//*IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021: 2235–2239. https://doi.org/10.1109/ICASSP39728.2021.9414568.
21. Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
22. Ouyang D L, He S, Zhang G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//*IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023: 1–5. https://doi.org/10.1109/ICASSP49357.2023.10096516.
23. Xiao X, Zhang D Y, Hu G W, et al. CNN-MHSA: a convolutional neural network and multi-head self-attention combined approach for detecting phishing websites[J]. *Neural Netw*, 2020, **125**: 303−312.
24. Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021: 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350.
25. Tan M X, Le Q V. EfficientNetV2: smaller models and faster training[C]//*Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021: 10096–10106.
26. Xu G P, Liao W T, Zhang X, et al. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation[J]. *Pattern Recognit*, 2023, **143**: 109819.

作者简介



赵瑞祥 (2000-)，男，硕士研究生，研究方向为无人驾驶感知。

E-mail：zhaoruiixang@163.com



【通信作者】顾清华 (1981-)，男，博士，教授，CCF 会员，研究方向为多目标优化、智能采矿。

E-mail：qinghuagu@126.com

Drivable area segmentation model for open-pit mine roads based on spatial depth conversion

Zhao Ruixiang[1,3], Gu Qinghua[2,3*], Li Shaobo[3], Li Binyang[2]

[1] School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China;
[2] School of Resource Engineering, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China;
[3] Xi'an Key Laboratory of Smart Industry Perception Computing and Decision Making, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China

Abstract

Objective The unstructured road environment in open-pit mining areas is characterized by complex topography, variable lighting conditions, and irregular boundaries. In autonomous driving scenarios for mining trucks, particularly under medium-to-long-distance viewing perspectives, critical obstacles such as ruts and puddles often appear as low-resolution, small-scale targets. Conventional convolutional neural networks typically employ strided convolutions or pooling operations for downsampling, which inevitably leads to the loss of fine-grained texture and boundary information. Consequently, existing segmentation methods frequently suffer from missed detections of small targets and blurred segmentation edges, posing severe safety risks to the path planning of unmanned vehicles. To address these challenges, this study aimed to develop a high-precision, real-time instance segmentation model, named SCG-YOLO, specifically optimized for the drivable areas of open-pit mine roads. The primary goal was to enhance the perception of low-resolution details and irregular boundaries while strictly adhering to the computational constraints of onboard embedded devices.

Methods The proposed SCG-YOLO model was constructed based on the YOLOv8-seg architecture, integrating three targeted improvement modules to handle the unique constraints of mining environments.

First, a space-to-depth and deformable convolution (SDC) module was designed and embedded into the feature extraction network. To mitigate the information loss associated with traditional downsampling, the Space-to-Depth layer decoupled and reorganized the spatial dimensions of input feature maps into depth dimensions. This operation effectively preserved fine-grained information, such as the shallow textures of ruts and the irregular edges of puddles, which are typically discarded by standard pooling layers. Furthermore, Deformable Convolution layers were incorporated within the SDC module. Unlike standard convolutions with fixed geometric structures, deformable convolutions learned offset fields to dynamically adjust the sampling grid. This capability allowed the network to adaptively model the non-rigid deformations of targets on unstructured road surfaces, ensuring accurate feature alignment for distorted shapes.

Second, a coordinate attention (CA) mechanism was introduced into the network neck. Distinct from global average pooling which compresses spatial information, the CA mechanism performed feature encoding along two spatial directions (horizontal and vertical) respectively. This process captured long-range dependencies while preserving precise positional information. By re-weighting the feature maps, the mechanism enhanced the network's focus on key drivable regions and suppressed interference from complex background factors common in mining areas, such as ore piles, dust, and equipment shadows.

Third, to address the increased computational overhead introduced by the SDC and CA modules, a Ghost module was adopted for the lightweight reconstruction of the neck layer. Standard Convolution and C2f modules were replaced with GhostConv and C3Ghost modules. The Ghost module generated a portion of the feature maps using standard convolutions and then produced redundant features via computationally inexpensive linear operations. This strategy significantly reduced the number of parameters and floating point operations (FLOPs) without compromising the representational power of the model, thereby serving as an engineering constraint to ensure real-time performance.

Results and Discussions Experimental validation was conducted using a self-collected dataset of open-pit mine roads, comprising 565 images covering various seasons and lighting conditions. The dataset included three annotated categories: drivable areas, ruts, and puddles. The proposed SCG-YOLO model demonstrated superior performance across multiple evaluation metrics.

Foundation item: National Natural Science Foundation of China (52074205, 52374135), Project of Shaanxi Province Metal Mine Intelligent Mining Theory and Technology Innovation Team (2023-CX-TD-12)

*Correspondence: Gu Q H, E-mail: qinghuagu@126.com

Quantitative analysis revealed that the model achieved a mean average precision (mAP50) of 82.6%, representing a substantial increase of 11.7% compared to the baseline YOLOv8-seg model. In terms of inference speed, the model maintained a frame rate of 425.31 f/s, satisfying the stringent real-time requirements of autonomous mining operations.

Ablation studies provided deep insights into the contribution of each component. The integration of the SDC module alone yielded a 9.4% improvement in mAP50, validating its efficacy in recovering fine-grained details lost during downsampling. The addition of the CA mechanism contributed a further 2.2% increase, confirming its ability to filter background noise and sharpen focus on relevant features. The implementation of the Ghost module successfully reduced the parameter count to 11.2 M, offsetting the computational cost of the accuracy-enhancing modules.

Comparative experiments conducted with various instance segmentation algorithms, including Mask R-CNN, YOLACT, FastInst, Mask2Former, and various YOLO iterations (v5, v6, v11), highlight the advantages of SCG-YOLO. While heavy models like Mask2Former offered competitive accuracy, their excessive computational demands (286.5 G FLOPs) rendered them unsuitable for edge deployment. Conversely, lighter models like YOLACT failed to achieve sufficient segmentation precision for small targets. SCG-YOLO achieved the optimal balance between accuracy and efficiency.

Visual qualitative analysis showed that the model produced significantly sharper segmentation masks. In scenarios with dusty backgrounds or water-filled potholes, SCG-YOLO accurately delineated boundaries that appeared blurred in baseline predictions. Generalization tests performed on the public RUGD dataset further demonstrated the model's robustness, where it outperformed the baseline by 5.2% in mAP50, proving its adaptability to diverse unstructured terrains beyond the specific training site.

Conclusions The study presents a robust solution for environmental perception in open-pit mining based on spatial depth conversion. The investigation confirms that the proposed SCG-YOLO model effectively resolves the core issues of fine-grained feature loss and edge blurring associated with low-resolution imaging of small targets in unstructured road environments. The synergistic combination of the SDC module and CA mechanism significantly enhances the model's ability to perceive rut textures and puddle boundaries under complex conditions. Simultaneously, the lightweight design utilizing Ghost modules ensures that the high computational efficiency required for autonomous driving systems is maintained. The model achieves a high segmentation precision of 82.6% mAP50 while running at speeds exceeding 400 f/s, providing a viable and safe technical foundation for the deployment of unmanned mining trucks. Future research directions include expanding the dataset to encompass extreme weather conditions and multi-vehicle interaction scenarios to further improve the model's generalization capabilities in dynamic mining environments.

Keywords: open-pit mine; drivable area; spatial depth conversion; deformable convolution

Zhao R X, Gu Q H, Li S B, et al. Drivable area segmentation model for open-pit mine roads based on spatial depth conversion[J]. *Opto-Electron Eng*, 2026, **53**(5): 250216; DOI: 10.12086/oee.2026.250216

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文