



DOI:10.12404/j.issn.1671-1815.2406113

引用格式: 卢昱杰, 陈宇凡, 魏伟. 施工场景遮蔽图像特征推理与语义补全方法: 以施工围栏为例[J]. 科学技术与工程, 2025, 25(16): 6898-6912.

Lu Yujie, Chen Yufan, Wei Wei. Image feature inference and semantic completion in occluded construction scenes: a case study of construction fences[J]. Science Technology and Engineering, 2025, 25(16): 6898-6912.

施工场景遮蔽图像特征推理与语义补全方法: 以施工围栏为例

卢昱杰^{1,2,3}, 陈宇凡¹, 魏伟^{1*}

(1. 同济大学土木工程学院, 上海 200092; 2. 同济大学工程结构性能演化与控制教育部重点实验室, 上海 200092;

3. 同济大学上海智能科学与技术研究院, 上海 200092)

摘要 施工安全巡检易因视野遮挡导致特征缺失, 出现危险误判情况。为提升施工风险识别效率, 针对特征遮挡问题, 以施工围栏检测为例, 提出一种基于 amodal 补全技术的遮挡特征推理方法。首先, 利用图像合成技术建立存在视野遮挡的围栏检测数据集。然后, 结合 YOLOv8 实例分割和 ASBU (amodal segmenter based on boundary uncertainty estimation) 特征补全网络, 推测围栏被遮挡部分的视觉特征。遮挡特征补全后的施工围栏可应用于闭环检测等多种施工安全监管任务。研究选择多个施工场地的围栏图像进行案例验证, 可实现精准的围栏缺失特征补全 (平均交并比 mIoU > 95.5%)。研究成果为施工场景的遮挡特征推理提供研究范式, 有效提升了智能化施工安全监管效率。

关键词 施工安全监管; 施工特征遮挡; amodal 特征补全; 实例分割; 计算机视觉

中图分类号 TU714;

文献标志码 A

Image Feature Inference and Semantic Completion in Occluded Construction Scenes: A Case Study of Construction Fences

LU Yu-jie^{1,2,3}, CHEN Yu-fan¹, WEI Wei^{1*}

(1. College of Civil Engineering, Tongji University, Shanghai 200092, China; 2. Key Laboratory of Performance Evolution and

Control for Engineering Structures of the Ministry of Education, Tongji University, Shanghai 200092, China;

3. Shanghai Research Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China)

[Abstract] In construction safety inspections, visual obstructions often lead to missing features, resulting in dangerous misjudgments. To improve the efficiency of risk identification in construction, a method for occluded feature inference based on amodal completion technology was proposed, using construction fence detection as a case study. First, a dataset of fence detection images with visual obstructions was created using image synthesis techniques. Then, a combination of YOLOv8 instance segmentation and the amodal segmenter based on boundary uncertainty estimation (ASBU) feature completion network was used to infer the visual features of the occluded parts of the fence. The completed features of the occluded construction fences can be applied to various construction safety monitoring tasks, such as closed-loop detection. The approach was validated using fence images from multiple construction sites, achieving precise feature completion for occluded fences (average intersection ratio mIoU > 95.5%). The research results provide a framework for feature inference in occluded construction scenes, which enhances the efficiency of intelligent construction safety supervision.

[Keywords] construction safety supervision; occluded construction features; amodal feature completion; instance segmentation; computer vision

规范的施工安全管理将有效避免工程危险事故发生, 是保障工程顺利进行和项目成功交付的关键因素。近年来, 建筑施工事故层出不穷, 2020 年,

全国共发生房屋市政工程生产安全事故 689 起, 其中起重机械伤害事故占事故总数的 30.43%, 严重威胁到工作人员的生命安全^[1]。

收稿日期: 2024-08-15; 修订日期: 2025-03-17

基金项目: 国家重点研发计划(2022YFC3801700); 中央高校基本科研业务费专项资金(2024-1-ZD-02, 22120250350); 国家自然科学基金(52078374)

第一作者: 卢昱杰(1985—), 男, 汉族, 贵州贵阳人, 博士, 教授。研究方向: 智能建造、低碳建造与工程管理。E-mail: lu6@tongji.edu.cn。

* 通信作者: 魏伟(1997—), 男, 汉族, 山东济南人, 博士研究生。研究方向: 智能建造。E-mail: 2410045@tongji.edu.cn。

投稿网址: www.stae.com.cn

为了提升施工安全监管效率,计算机视觉技术已逐步应用于智能化施工监控。基于目标检测、目标分类、实例分割、目标跟踪等计算机视觉技术,可用于结构健康监测^[2-4]、安全装置佩戴监测^[5-6]、人机安全距离监控^[7-8]、脚手架安全检查^[9]、围栏缺口检测^[10]等。然而,由于建筑工地的施工主体繁多,施工监控画面中的关注区域易被其他机械、工人等多种变化且不规则的前景遮挡,且遮挡发生的比例和位置分布随机性强,易造成关键识别特征缺失,降低风险识别精度,致使错误报警频发。近年来,已有学者针对特征补全问题展开了相关研究。

在计算机视觉领域,图像修复(Inpainting)是一种实用的解决方法,该方法旨在根据图像的上下文信息,为图像中缺失部分生成视觉上合理且语义一致的内容。苏婷婷等^[11]使用深度卷积对抗生成网络(deep convolutional generative adversarial network, DCGAN)^[12]对人脸补全修复,并建立了针对人脸的相关性损失函数以促进 DCGAN 的优化过程,该方法在 50% 加噪比例下实现了较完整的图像补全。Angah 等^[13]提出了一种基于 U-Net^[14]的上下文编码器模型来去除工人遮挡,并结合条件生成对抗网络(conditional generative adversarial nets, cGAN)^[15]完成图像修复。这种方法能够在去除前景遮挡后实现背景的高质量填补,从而获得更清晰的工作现场视图。然而,这类方法的目的是使补全部分与图片整体保持视觉观感一致性,缺乏对遮挡部分特征的推理能力。

除图像修复外,遮挡特征提取与推理也是一种常用的特征补全方法。Yan 等^[16]利用遮挡部分与未遮挡部分之间的全局结构约束来恢复遮挡特征,以提升汽车或行人在部分遮挡情况下的检测性能。但该方法仅适用于目标处于图像边缘或受拍摄视野限制导致部分特征不可见的情况。还有学者采用深度学习提取特征的方法,Li 等^[17]针对施工工人遮挡场景,提出了一种基于改进生成对抗插补网络(generative adversarial imputation nets, GAIN)^[18]的遮挡修复方法,该方法通过卷积神经网络提取工人可见骨架信息构建特征矩阵,并预测因遮挡产生的特征缺失值。但该方法依赖视频帧之间的上下文信息,且工人骨架特征相对简单,难以应对更复杂场景下的遮挡问题。

上述方法虽然在一定程度上解决了遮挡问题,但它们的特征提取过程主要集中在如图像颜色、纹理等显式特征上,这些特征通常依赖预定义的规则,缺乏对物体整体结构的隐式理解,即这些方法无法结合对象的形状、结构等先验知识来推理,容

易在识别或分割任务中产生不连贯结果,且在应对复杂场景时存在局限。

为解决上述问题,提出一种基于 amodal 补全的遮挡补全方法。在计算机视觉领域,amodal 补全指的是在遮挡特征缺失的情况下感知目标整体特征的能力^[19],该补全方法涉及 modal 和 amodal 两种掩码。前者是指目标可见部分的掩码(未被遮挡部分),而后者是指物体可见和不可见部分掩码(被遮挡部分)的并集。所提出的遮挡补全方法可广泛应用于施工安全管理任务。其中,人机作业是施工危险系数最高,最需进行安全监管的施工任务之一^[20]。为防止人机碰撞事故,规范要求开挖深度超过 2 m 及以上的基坑周边应安装防护围栏且确保围栏封闭^[21]。为避免施工特征遮挡对围栏闭合判断造成影响,以施工围栏闭合检测任务为例,验证了遮挡补全方法在施工安全监管任务中的实际应用效果。首先,基于虚拟数据合成方法生成视野遮挡的围栏数据集。其次,将合成数据集应用于 ASBU^[22](amodal segmenter based on boundary uncertainty estimation)网络,提取目标可见部分的上下文信息,推理遮挡部分的不可见特征。最终,基于完整的施工目标特征应用于围栏闭合检测等施工安全监管任务。研究方法为施工目标遮挡特征推理研究提供理论借鉴,为复杂建造场景智能化施工安全动态管控提供实践参考。

1 基于 amodal 补全的遮挡补全方法

研究目标是施工场景下物体被遮挡部分的特征补全,并以围栏为例,基于补全结果对围栏闭合状态进行判断,整体的研究内容分为:①基于前背景组合的遮挡数据集合成。将挖掘机作为前景对象,合成到围栏背景图像中,形成一套有真值的挖掘机遮挡围栏数据集;②基于 YOLOv8 的实例分割。识别场景中围栏及挖掘机的像素尺度信息,得到围栏的可见掩码;③基于 ASBU 的 amodal 补全。基于 YOLOv8 分割的可见部分掩码,输入 ASBU 预测被遮挡部分,补全形成围栏的完整掩码;④基于几何轮廓提取的围栏闭合判断。通过一系列图像层级的后处理操作,以围栏的 amodal 掩码为输入获取围栏中心线,并基于中心线连续情况判断实际围栏是否闭合。研究框架如图 1 所示。

1.1 基于前背景组合的围栏遮挡数据合成

ASBU 网络需要输入包含目标可见及被遮挡区域的完整图像以进行特征推理,而现实实拍数据多为单帧静态图像,难以获取被遮挡区域的图像信息。

为了解决上述问题,提出一种图像合成的方法

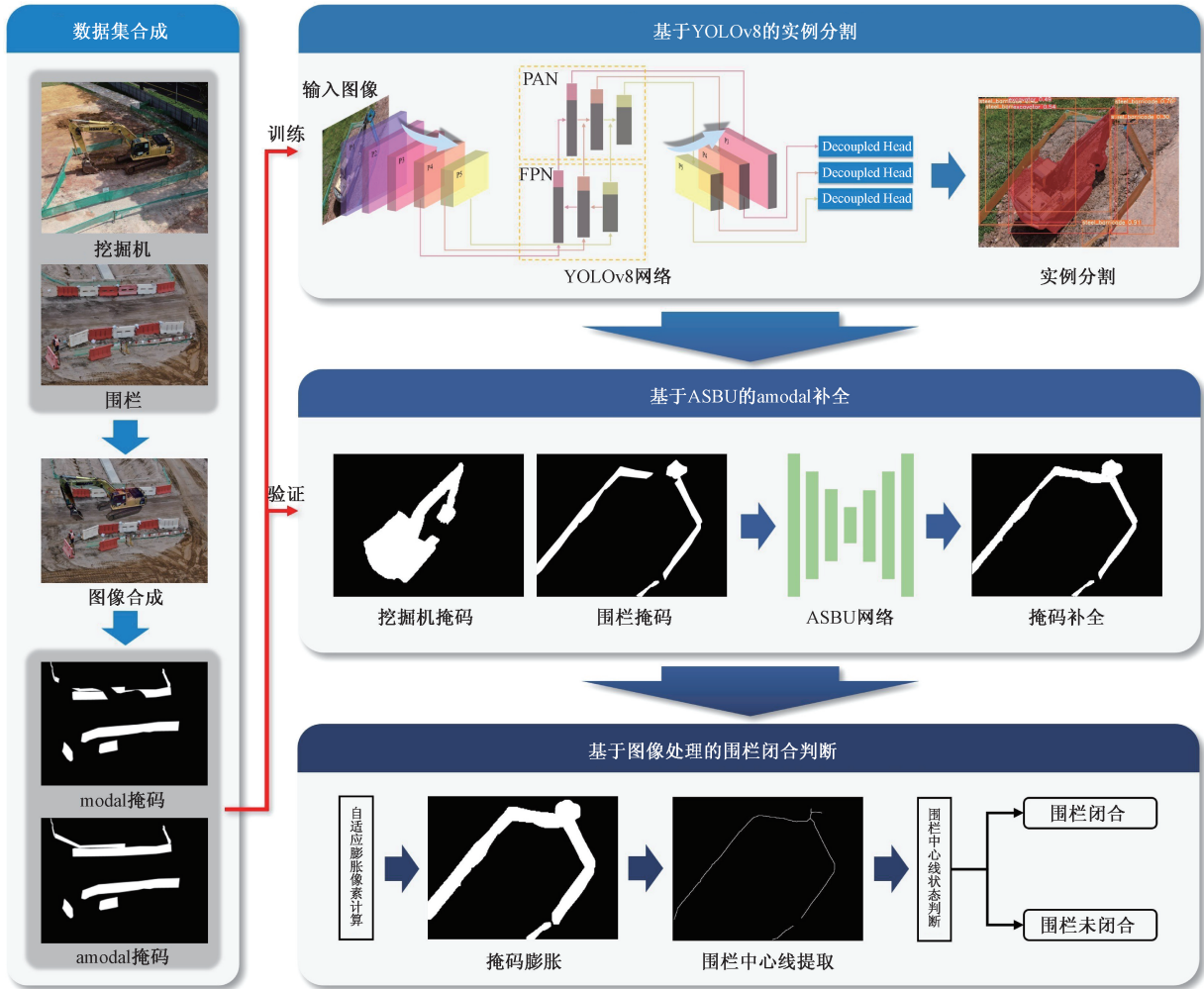


图1 研究框架

Fig. 1 Research framework

来获取包含遮挡区域真值的虚拟数据集,具体的方法框架如图2所示。

首先对挖掘机图像与围栏图像进行标注,以生成挖掘机前景掩码与未被遮挡的围栏掩码。然后,将挖掘机与围栏图像按照观测视角划分为平视(俯仰角:0°~30°)、轻微俯视(俯仰角:30°~50°)、俯视(俯仰角:50°~80°)3个类别。

进一步地,从图像中分离前景掩码并将其放置于同类观测视角的围栏背景中,使得挖掘机前景遮挡围栏的部分区域。通过以上步骤,既能生成围栏遮挡图像,又可获取未被遮挡的围栏掩码(即围栏amodal掩码)。

根据前景对象在背景图像中的位置,可得到合成图像中前景物体的标注信息。基于背景图像中标注的围栏掩码,在挖掘机合成到图像中后,通过图像位运算,可计算得到围栏的可见部分掩码。经过数据处理,可形成如图2右侧所示的各类掩码数据:挖掘机(前景)掩码、围栏 amodal 掩码及围栏

modal 掩码。经过处理的数据可根据具体需求生成 amodal 及 modal 数据集。

在图像合成过程中,首先调整前景对象的图像尺寸使其与背景图像在同一个像素尺度,接着将原图像中的前景对象插入到背景图像的随机位置。由于图像合成的前景背景来自不同的图像,因此合成图像多出现像素伪影^[23],即前景背景交界处存在明显的分辨率与颜色突变。采用泊松融合图像处理技术^[24],通过缩小图像亮度和颜色的梯度差异,使前景背景边界的外观过度更加自然,提升合成图像的外观真实性。泊松图像处理方法整体流程如图3所示。

在图像领域要处理的泊松融合函数如式(1)所示,该公式的目标是最小化融合图像中融合部分的梯度与原图像中前景对象梯度之间的差值。

$$\min_f \iint_{\Omega} |\nabla f - \nu|^2 \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (1)$$

式(1)中: f 为待求解的结果图像; f^* 为已知的目标

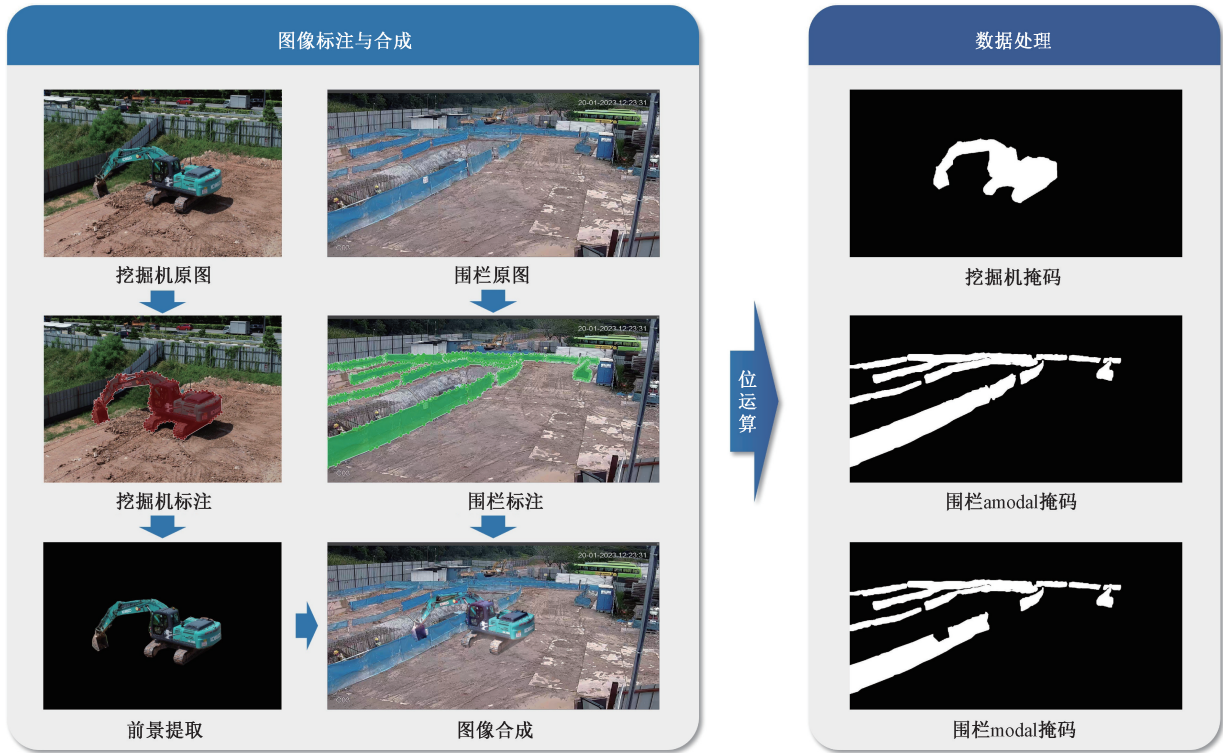


图2 图像合成与结果处理框架

Fig. 2 Image synthesis and results processing framework



图3 图像合成整体流程

Fig. 3 Overall process of image synthesis

图像; Ω 为待融合区域; ∇ 为梯度算子; v 为梯度场, 即原图像的梯度。

在实际计算中, 图像被表示为一个二维离散网络, 对于每一个像素点 (i, j) , 离散化的拉普拉斯算子如式(2)、式(3)所示。

$$\Delta f_{i,j} = f_{i+1,j} + f_{i-1,j} + f_{i,j+1} + f_{i,j-1} - 4f_{i,j} \quad (2)$$

$$\Delta f_{i,j} = \nabla \cdot v_{i,j} \quad (3)$$

式中: Δ 为拉普拉斯算子; $f_{i,j}$ 为像素点 (i, j) 对应的像素值; ∇ 为散度运算符; $v_{i,j}$ 为原图像在 (i, j) 位置的梯度。

为了方便计算机求解, 将公式转换为矩阵形式有式(4), 由于未知量巨大, 采用 Jacobi 迭代法来求解, 迭代次数设置为 20 000 次, 每次迭代都会更新图像中每个像素的值, 并逐渐逼近最终解, 如式(5)所示。

$$Ax = b \quad (4)$$

$$x_i^{(k+1)} = \frac{1}{A_{i,i}} \left(b_i - \sum_{j \neq i} A_{i,j} x_j^{(k)} \right) \quad (5)$$

式中: x 为待求解目标图像的像素值; b 为包含梯度信息的向量; A 为稀疏矩阵, 对应像素的邻域关系; $x_i^{(k)}$ 为第 i 个未知量在第 k 次迭代时的值; $A_{i,j}$ 为 A 矩阵中第 i 行、第 j 列的元素; b_i 表示向量 b 的第 i 个元素。

为了合成图像中的挖掘机与围栏背景的整体色调一致, 采用颜色直方图匹配算法调整挖掘机的颜色分布, 使其更接近背景图像的颜色分布。

1.2 基于 YOLOv8 的实例分割

通过 1.1 节方法获得数据集, 其中形成的 modal 数据集可用于训练 YOLOv8 的实例分割。将所输入的图片调整至统一尺寸输入卷积神经网络提取特征, 再通过检测头完成对目标的分类及掩码预测。

基于 YOLOv8 实例分割网络的结构如图 4 所示,主要由骨干网络(Backbone)、颈部(Neck)、头部(Head)构成,其中,骨干网络从图像中提取图像特征,颈部作为中间层,对骨干网络提取的特征进行多尺度融合,头部作为模型的最后一层,采用 Softmax 激活函数,输出目标所在的边界框、类别概率和像素级分割掩码。

1.3 基于 ASBU 的 amodal 补全

1.3.1 ASBU 网络架构

基于 YOLOv8 实例分割获取的挖掘机和围栏的 modal 掩码,输入 ASBU^[22] 网络进行围栏的 amodal 补全,其网络架构如图 5 所示。

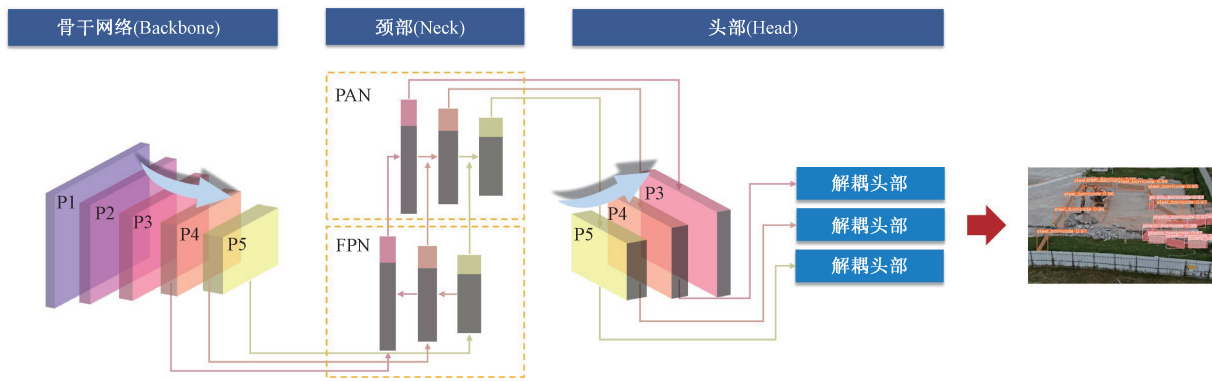
ASBU 接受以下数据作为训练的输入集:

- ①RGB图像;
- ②被遮挡物的 modal 掩码;
- ③遮挡物掩码及其与被遮挡物的边界。

在输入的数据集中,ASBU 随机选取遮挡物和被遮挡物及它们的 modal 掩码,并计算被遮挡物被前景遮挡后的掩码。同时,根据遮挡物被遮挡的情况计算遮挡发生的边界掩码,该掩码被估计为遮挡物和被遮挡物的形态学放大掩码的交集区域。

将数据输入 UNet 后,将输出一个 $H \times W \times 2$ 的特征图,该特征图的两个通道分别用于 amodal 分割预测和不确定性预测,其中 H 和 W 分别为输入图像的高度和宽度。不确定性预测旨在隐式地捕获训练对象实例的“先验形状”,通过最小化预测的不确定性,网络能够捕捉到物体的合理形状^[22]。

amodal 分割预测值由 Sigmoid 函数输出,不确定性预测值由 Softplus 函数输出,分别如式(6)、式(7)所示。ASBU 采用的损失函数结合了数据误差和不确定性两方面,具体如式(8)、式(9)所示。



P_n 为第 n 层特征图;FPN(feature pyramid networks)为特征金字塔网络; PAN(path aggregation network)为路径聚合网络

图 4 YOLOv8 网络架构

Fig. 4 YOLOv8 network architecture

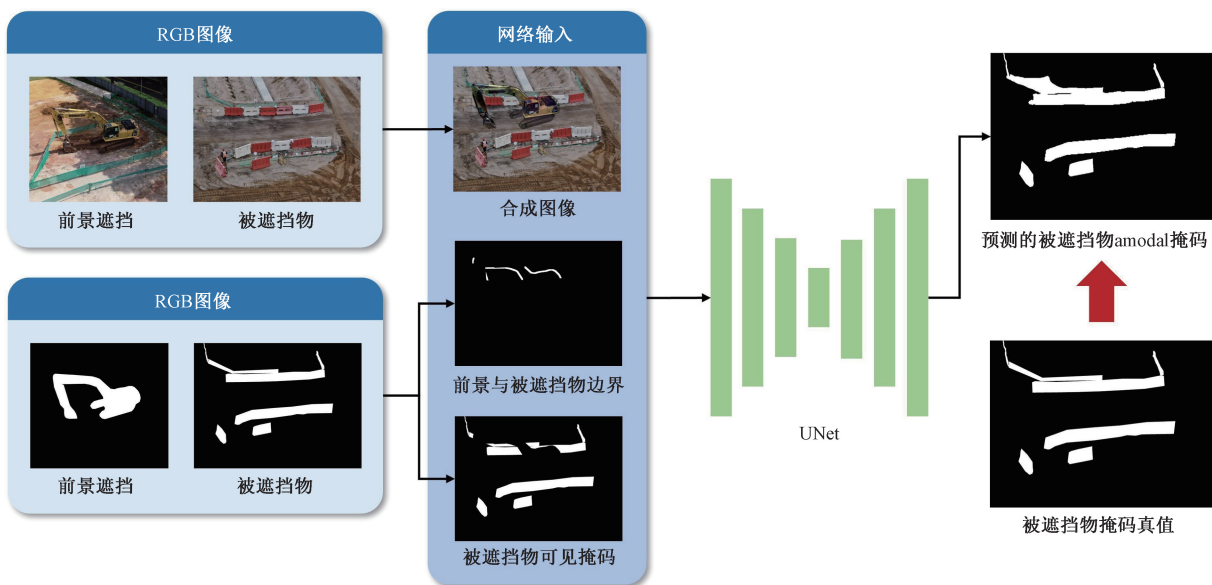


图 5 ASBU 网络架构

Fig. 5 ASBU network architecture

通过引入该不确定性加权的损失函数,综合考虑了给定的真值 modal 掩码(伪 amodal 掩码)和输出的预测 amodal 掩码的差异及预测的不确定性。

$$m_i = \text{sigmoid}(z) = \frac{1}{1 + e^{-z}} \quad (6)$$

$$u_i = \text{softplus}(z) = \ln(1 + e^z) \quad (7)$$

式中: $m_i \in [0, 1]$ 为 amodal 掩码预测值; i 为像素的索引; sigmoid、softplus 为神经网络的激活函数; z 为函数的输入; $u_i \in \mathbf{R}^+$ 为不确定性预测值。

$$L = \frac{1}{N} \sum_{i=1}^N J(m_i^c = 0)L_i + \lambda J(m_i^c = 1)L_i \quad (8)$$

$$L_i = \frac{1}{2} \left[\left(\frac{m_i^* - m_i}{u_i} \right)^2 + u_i^2 \right] \quad (9)$$

式中: m_i^c 为遮挡物掩码值; $J(\cdot)$ 为推断函数; m_i^* 为真实掩码值; λ 为用于平衡不同区域的损失权重的常数, 设置为 5, 即遮挡物内部被赋予更多的权重; L_i 为第 i 个像素的损失函数。

1.3.2 特征推理精度评价指标

ASBU 算法采用平均交并比 (mIoU) 和像素准确率 (pixel accuracy, PA) 进行精度评价。平均交并比 [式(10)] 用于衡量预测 amodal 掩码与真值掩码之间的重叠程度; 像素准确率 [式(11)] 的数值结果为预测的 amodal 掩码中正确预测的像素与真值掩码像素的数量比。

$$\text{mIoU} = \frac{|m_p \cap \text{GT}|}{|m_p \cup \text{GT}|} \quad (10)$$

$$\text{PA} = \frac{|m_p \cap \text{GT}|}{|\text{GT}|} \quad (11)$$

式中: m_p 为预测掩码; GT 为真实掩码; $|\cdot|$ 为集合中元素的数量, 即像素数。

1.4 基于图像处理的围栏闭合判断

1.4.1 围栏掩码自适应膨胀方法

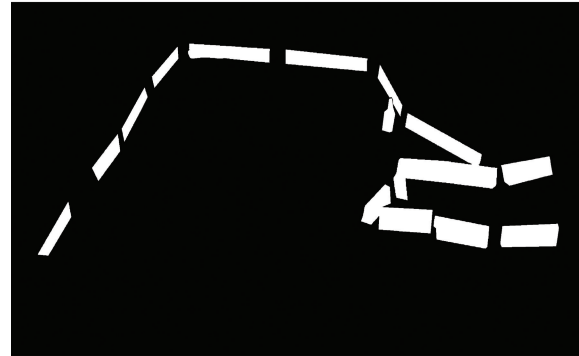
由于围栏间的微小间隙 [图 6(a)] 也会影响其闭合状态判断。提出一套基于围栏掩码特征的形态学像素膨胀方法来解决上述问题, 该方法综合考虑图像的俯仰角和围栏的规范间距, 将规范间距设定为掩码膨胀尺寸, 以实际与像素距离的等比关系为尺度提出像素膨胀估算公式, 见式(12)。

为了计算条状掩码的近似宽度, 以假想长方形作为条状围栏掩码形状, 掩码的周长和面积作为已知量, 掩码的宽度和高度作为未知量, 得到掩码近似宽度的计算公式, 见式(13)。考虑到拍摄俯仰角对高度的透视影响, 掩码的近似高度计算公式见式(14)。

$$D_i = \lambda' d \frac{W_i}{H_i} = \frac{\lambda' d W_i}{h_i \cos \theta} \quad (12)$$



(a) 原始图像



(b) 围栏掩码



(c) 膨胀后围栏掩码

图 6 围栏掩码自适应膨胀

Fig. 6 Adaptive expansion of fence mask

$$W_i = \begin{cases} \frac{p_i - \sqrt{p_i^2 - 16A_i}}{4}, & p_i \geq 4\sqrt{A_i} \\ \frac{p_i}{4}, & p_i < 4\sqrt{A_i} \end{cases} \quad (13)$$

$$H_i = h_i \cos \theta \quad (14)$$

式中: D_i 为膨胀像素; λ' 为放大系数, 取值为 1.1 ~ 1.3; W_i 为图像中第 i 个条状掩码的近似宽度; d 为规定的最大围栏间距; θ 为图像对应的俯仰角; H_i 为图像俯仰角下的换算实际高度; A_i 为掩码的面积; p_i 为掩码的周长; h_i 为现实中围栏的实际高度。

h_i 、 d 的设置参考《建筑施工现场安全防护设施技术规程》(DB42/T 535—2020) 的围栏尺寸和立杆间距, 但该规范中并没有阐明若围栏断开, 断开部

分规定的最大尺寸限制,因此对于实际围栏间距和高度未知的情况,可考虑采用标准中的围栏高度和栏杆间距代入式(12)做估算。

经上述步骤,存在微小间隙的离散掩码区域[图6(b)]可膨胀并连接为完整的掩码封闭体[图6(c)],为后续闭合逻辑判断奠定基础。

1.4.2 围栏闭合情况判断

对膨胀后的围栏掩码使用 OpenCV 库的 skeletonize 算法(也称二值图像细化),提取条状掩码的骨架线(后称中心线),获得简洁的围栏中心线信息。围栏闭合判断的整体流程如图7所示。

基于精炼后的围栏中心线,通过对其连续性的判断来验证围栏是否封闭。完整闭合的中心线应无端点(端点指线段在图像中的起点或终点,它只有1个相邻像素),因此若存在端点,则该端点所在的分支必定为非闭合。

本次研究旨在通过端点检测方法判断围栏的闭合情况。具体步骤为:首先,检索图像中的端点,并使用3×3的卷积核跟踪曲线的延伸方向。当遇到另一端点或曲线分叉点时,停止检索并删除该线

条上的像素。此过程重复进行,直至图像中不再存在端点。此时,保留在图像中的即为闭合的中心线;若不存在闭合的中心线,则表明图像中的围栏未闭合。当图像中检测到满足条件的闭合中心线时,将图像中的闭合区域与非闭合区域以颜色区分,如图8所示。

2 数据集与模型训练

通过网络检索与施工现场实拍的方式共收集挖掘机图像651张,围栏图像286张。为了保证数据多样性与泛化性,图像数据包含不同大小、远近和俯仰角的目标物体。

基于1.1节所述方法初步生成22468张合成数据。受前景放置位置随机性的影响,挖掘机可能位于语义不合理的空间位置,如漂浮于空中,或与背景中的围栏不存在遮挡关系。因此,通过人工审查的方式最终筛选出7190张图像数据集。如前所述,所有的图像附带包含modal和amodal的语义信息,因此无需额外的人工标注。数据集的归类与合成效果如表1所示。

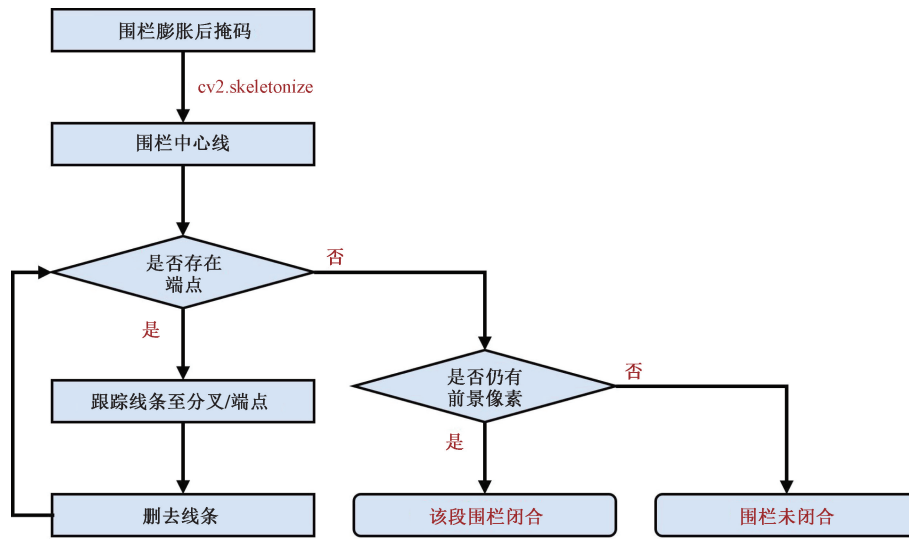
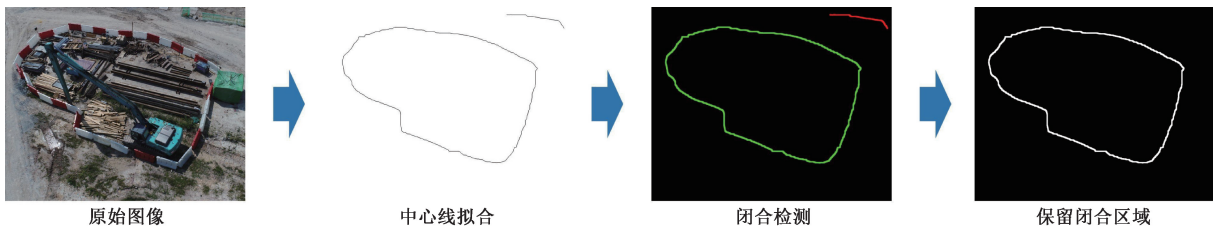


图7 围栏闭合判断流程

Fig.7 Fence closure judgment process



绿色表示检测闭合;红色表示检测未闭合

图8 围栏中心线提取

Fig.8 Fence centerline extraction

表 1 数据归类与合成
Table 1 Dataclassification and synthesis











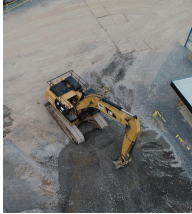




俯仰角/(°)	类别				
	挖掘机	围栏	合成效果	围栏 modal 掩码	围栏 amodal 掩码
0 ~ 30					
30 ~ 50					
50 ~ 80					

表 2 YOLOv8 模型训练环境配置

Table 2 YOLOv8 model training environment configuration

名称	设置
操作系统	Windows 11
处理器	12th GenIntel(R) Core(TM) i9-12900H 2.50 GHz
内存/GB	64.0
显卡	NVIDIA GeForce RTX 3070 Ti Laptop GPU
深度学习框架	Pytorch 1.13
运算平台	CUDA 11.7
编程语言	Python 3.9
开发环境	PyCharm Community Edition

表 3 YOLOv8 模型训练参数设置

Table 3 YOLOv8 model training parameter settings

参数	设置
预训练模型(model)	YOLOv8 m-seg
输入图像尺寸(image size)	640 × 640
训练轮次(epochs)	200
每批训练样本数(batch size)	8
优化器(optimizer)	SGD
初始学习率(lr0)	0.01
学习率因子(lrf)	0.01

2.1 YOLOv8 模型

合成的数据集中的 modal 掩码用于训练 YOLO 模型,如前所述共包含 7 190 张图像及对应的标注数据,并按照 7:3 的比例划分为训练集和验证集。

考虑到对围栏掩码识别的精准性和完整性需求,本实验采用 YOLOv8 m-seg 作为训练模型,该模型相比于 YOLOv8s-seg 能够达到更高的分割准确率。实验的软硬件配置信息如表 2 所示,网络的超参数设置如表 3 所示。

表 4 ASBU 模型训练环境配置

Table 4 ASBU model training environment configuration

名称	内容
操作系统	Ubuntu 20.04.1 LTS
处理器	Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10 GHz
内存/GB	128.0
显卡	NVIDIA Corporation Device 2204 (rev a1)
深度学习框架	Pytorch 1.12
运算平台	CUDA 11.4
编程语言	Python 3.7
开发环境	Visual Studio Code

表 5 ASBU 模型训练参数设置

Table 5 ASBU model training parameter settings

参数	设置
骨干网络架构(backbone arch)	UNet
迭代次数(iteration)	28 000
输入图像尺寸(image size)	256 × 256
每批训练样本数(batch size)	32
优化器(optimizer)	SGD
初始学习率(lr0)	0.001
学习率因子(lrf)	0.1

注:模型迭代到 16 000 代和 24 000 代时,将学习率调整为原有的 10%。

2.2 ASBU 模型

由于 ASBU 算法逻辑包含获取挖掘机对象作为围栏前景遮挡进行训练的逻辑,因此将收集后标注的原始数据直接输入 ASBU 网络进行训练。模型训练的软硬件配置如表 4 所示,训练超参数设置如表 5 所示。此外,ASBU 的验证集使用 1.1 节的合成数据,该验证集包含合成 RGB 图像、每个物体的可见及不可见部分的掩码信息。

3 结果与分析

3.1 案例概况

研究选取中国多个施工现场的围栏图像进行闭合检测方法验证(图9)。通过手机拍照、监控视频截图等方式共收集测试数据104张,数据收集过程保证拍摄位置与拍摄对象之间的俯仰角度位于 $20^{\circ} \sim 40^{\circ}$ 。测试图像中包含红白相间的塑料围栏[图9(a)]、以及附带安全防护网的围栏[图9(b)]。部分图像存在挖掘机前景遮挡的情况[图9(c)],以测试遮挡情况下的围栏恢复与闭合检测性能。



(a) 红白相间的塑料围栏



(b) 附带安全防护网的围栏



(c) 存在挖掘机前景遮挡

图9 研究案例

Fig. 9 Case study

3.2 YOLOv8 实例分割结果

3.2.1 模型训练结果

模型在训练集和验证集上的 loss 曲线及平均精度(mean average precision, mAP)随训练轮次的变化情况如图10(a)、图10(b)所示,结果显示,训练集和验证集的边界框回归、目标回归、分类回归损失均在第150轮次后趋于平稳,未出现过拟合。模型的准确率 P 与召回率 R 分别为97.8%和94.4%,围栏分割精度 $mAP_{0.5}$ 与 $mAP_{0.5-0.95}$ ($mAP_{0.5}$ 为以0.5为IoU时算法取得的平均精度; $mAP_{0.5-0.95}$ 为IoU以0.05步长在0.5~0.95区间上的平均mAP)高达96.3%和92.5%,如图10(c)所示,满足工程现场施工围栏检测的精度要求。

3.2.2 实例分割结果

施工围栏实例分割的可视化结果如表6所示,选取的图像覆盖了不同视角的多种围栏,且包含挖掘机前景遮挡的情况。结果显示,YOLOv8算法可准确地预测施工现场图像中的挖掘机及围栏的可见掩码。

3.3 ASBU 补全结果

3.3.1 模型训练结果

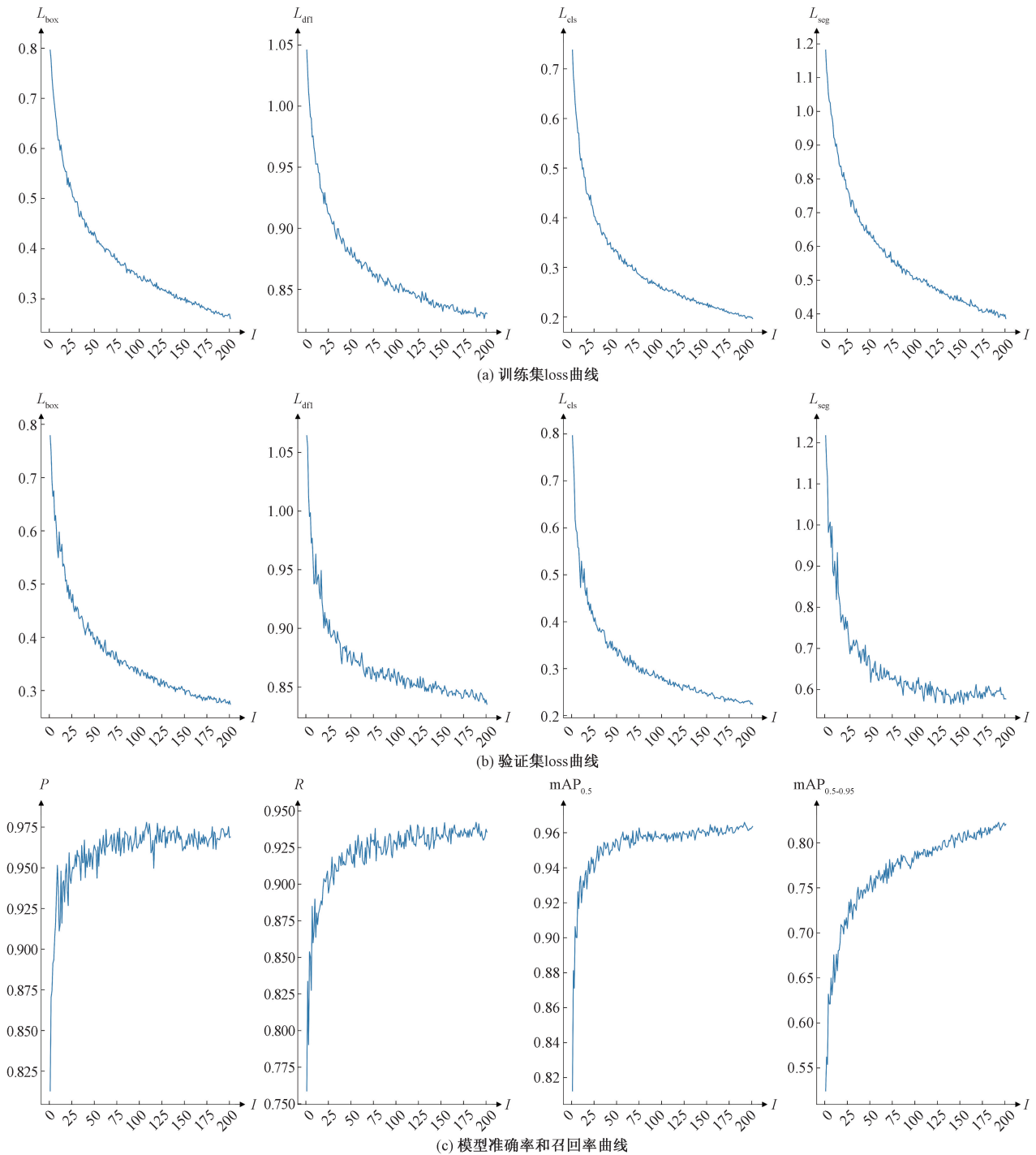
模型的损失随迭代次数的变化曲线如图11所示,训练损失在2000次迭代前急速下降,而后呈现震荡波动的缓慢下降趋势,最终损失值稳定在约0.05。

训练过程中程序每迭代2000次会进行一次训练过程中的验证,在训练结束后也可以根据对应迭代次数的权重文件在验证集上进行离线验证,在本次实验中,挑选了具有代表性的迭代权重文件进行离线验证,依次选取0、2000、16000、24000、28000次这些具有代表性迭代次数的权重文件进行离线验证,结果如表7所示。

3.3.2 amodal 补全结果

ASBU算法在验证集的 amodal 补全结果如表8所示。训练后的ASBU网络可准确预测被挖掘机遮挡部分的围栏掩码,为围栏闭合检测的可靠性提供保障。实验结果表明,不同的遮挡位置和比例将会影响特征推理的算法表现。当遮挡位于目标中部时,算法的缺失特征推理精度较高,可精准补全和连接因遮挡造成的不连续掩码;而当遮挡位置处于目标端部时,端部区域缺乏充足的上下文信息,将在一定程度上影响特征推理精度。

基于训练好的UNet权重模型,选取loss最小的第28000代权重模型用于验证集测试,试验结果表明,mIoU和像素准确率(pixel accuracy)分别高达95.454%和97.779%。预测的掩码与真实掩码几尽重合,验证了本方法在施工遮挡特征推理方面的精确度与可行性。



I 为迭代次数; L_{box} 为边界框回归损失; L_{dfl} 为分布式焦点损失; L_{cls} 为分类损失; L_{seg} 为分割损失; P 为准确率; R 为召回率; $\text{mAP}_{0.5}$ 为以 0.5 为 IoU 时算法取得的平均精度; $\text{mAP}_{0.5-0.95}$ 为 IoU 以 0.05 步长在 0.5 ~ 0.95 区间上的平均 mAP; IoU 为预测目标位置与目标真实位置的重叠率

图 10 YOLOv8 分割模型训练过程

Fig. 10 Training process of YOLOv8 segmentation model

3.3.3 遮挡补全方法对比






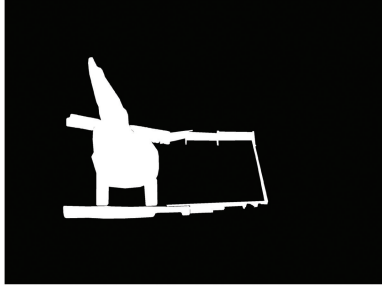


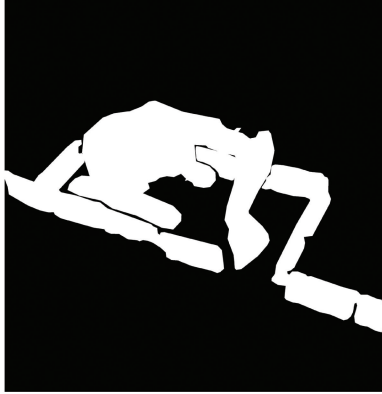


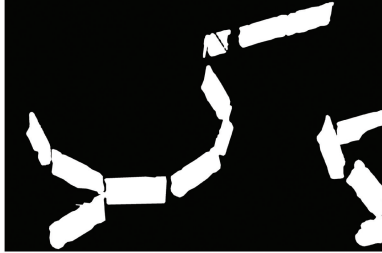


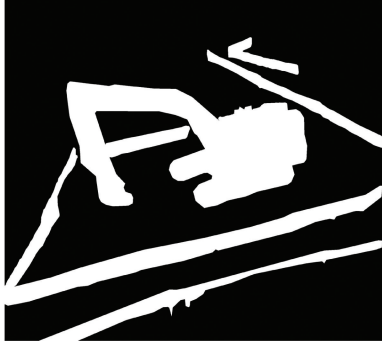
本次研究结果与其他施工场景下解决遮挡问题的方法进行对比分析, 如表 9 所示。相比于其他方法, 通过隐式理解物体的整体形态, 能够基于可见掩码预测物体整体的 amodal 掩码, 该方法能够适用于

复杂遮挡场景, 且在遮挡补全准确性上具有优势。

3.4 围栏闭合判断结果

在获取了围栏的 amodal 掩码后, 对该结果进行围栏闭合的判断验证, 共选取 100 张实际案例中的图像进行判断。表 10 为 8 个示例的判断结果, 结果

表 6 基于 YOLOv8 的实例分割结果
Table 6 Instance segmentation results based on YOLOv8

原始图像	分割结果	真值
		
		
		
		
		

注:steel_barricade 为钢制围栏;plastic_barricade 为塑料围栏;excavator 为挖掘机。

显示,以人为判断结果为基准,所提方法与未进行补全时相比,判断准确率从 68% 提升到 82%,能够准确识别画面中围栏的闭合部分。

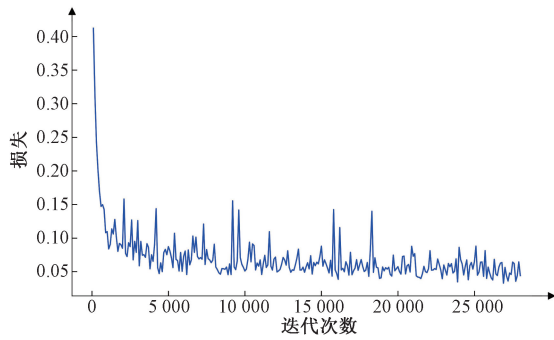


图 11 ASBU 训练 loss 曲线

Fig. 11 ASBU training loss curve

表 7 ASBU 损失验证结果

Table 7 ASBU loss validation results

迭代次数	loss 值	
	在线验证	离线验证
0	0.767 20	N.A.
2 000	0.098 38	0.090 45
16 000	0.052 97	0.048 64
24 000	0.048 69	0.045 36
28 000	0.046 39	0.040 15

注:loss 值都是训练过程中的累计 loss。

围栏判断闭合的精度受多种因素影响,如掩码识别精度、掩码膨胀尺度等。首先,因围栏闭合判断结果依赖于掩码识别的精确度,如 modal 掩码识别不完全,或预测的 amodal 掩码未将属于同一围栏的掩码正确连接,则基于该掩码的闭合判断结果的准确性也会下降。其次,不恰当的掩码膨胀尺度也可能影响判断结果,如膨胀像素过小会导致微小的围栏间隙未填补;过大则会导致邻近平行围栏融合为一个整体,造成信息丢失。

4 结论

(1) 工程建设场景的施工环境复杂,施工主体繁多,易导致监控视野遮挡,干扰危险事故判断。为了提升施工安全识别的准确性,以施工围栏识别任务为例,提出一种基于 amodal 补全的施工遮挡特征推理方法。该方法应用 ASBU 网络提取围栏可见特征,并利用全局上下文信息推断被机械遮挡的不可见信息,以恢复完整的目标特征。实验证明,该方法可在复杂施工遮挡情况下实现精准的围栏特征推理与补全,精度达到 95.5%。以围栏为案例进行研究,但该研究方法同样适用于受机械、工人、物料等多种前景目标遮挡的复杂特征推理。该方法还可预测被遮挡区域的 RGB 外观色彩特征,进而应用于基于彩色点云重建的三维场景特征还原等任务。

表 8 ASBU 补全结果

Table 8 ASBU completion results

遮挡位置	遮挡比例	原始图像	合成图像	围栏可见掩码	预测的围栏 amodal 掩码	围栏掩码真值
小幅遮挡						
中部遮挡						
	大幅遮挡					

续表

遮挡位置	遮挡比例	原始图像	合成图像	围栏可见掩码	预测的围栏 amodal 掩码	围栏掩码真值
小幅遮挡						
端部遮挡						
大幅遮挡						

注:第4列白色部分为输入 ASBU 的可见掩码;绿色表示预测和真值均为 1 的区域;红色表示预测为 1 但真值为 0 的区域,即错误预测的部分;橙色表示预测为 0 但真值为 1 的区域,即模型未能正确预测的部分。

表 9 遮挡补全方法对比与分析

Table 9 Comparison and analysis of occlusion completion methods

方法	算法	精度	原理	特点
基于 ASBU 的遮挡特征 amodal 补全(本文)	ASBU; amodal 补全	mIoU 为 95.454%; 像素准确率为 97.779%	基于 modal 掩码补全被遮挡物的 amodal 掩码	可基于单张静态图像推理并还原遮挡区域的真实特征
基于 U-Net 的施工现场遮挡物移除 ^[13]	U-Net; cGAN	图像平均 L1 损失为 7.39%; 标准差为 5.88%	根据图像上下文信息修复被遮挡的背景区域	可实现图像补全,但无法还原遮挡区域的真实特征
遮挡条件下的工人姿态识别 ^[17]	GAN	下游任务(姿态识别)准确率为 86.62%	基于前后帧上下文预测缺失的骨架特征	适用于视频帧输入,而难以处理单张静态图像



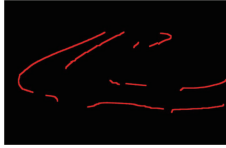






注:L1 损失为最小绝对偏差损失。

表 10 围栏闭合判断结果

Table 10 Fence closure detection results

原图	围栏补全掩码(膨胀后)	围栏闭合判断	人为判断	是否匹配	正确率/%
			闭合	是	82
			闭合	是	
			不闭合	是	

续表

原图	围栏补全掩码(膨胀后)	围栏闭合判断	人为判断	是否匹配	正确率/%
			不闭合	是	
			不闭合	是	82
			闭合	否	

注:绿色代表检测闭合,红色表示检测未闭合。

(2)本次研究仍存在一定的局限性,现有特征推理方法为基于单张图片的静态特征推理,特征推理精度与遮挡区域的位置分布有关。位于目标端部的遮挡区域易因上下文信息不充分,导致边缘遮挡特征难以推理与界定。未来研究将结合多视角图像特征匹配技术,通过融合与匹配多角度的目标特征,补充有限的单视角静态特征,提升目标边缘区域的遮挡特征推理精度。

参 考 文 献

[1] 中华人民共和国住房和城乡建设部. 住房和城乡建设部办公厅关于2020年房屋市政工程生产安全事故情况的通报[EB/OL]. (2021-05-21) [2024-06-09]. https://www.mohurd.gov.cn/gongkai/zc/wjk/art/2022/art_17339_768565.html. Ministry of Housing and Urban Rural Development of the People's Republic of China. Notice from the General Office of the Ministry of Housing and Urban Rural Development on production safety accidents in housing and municipal engineering in 2020 [EB/OL]. (2021-05-21) [2024-06-09]. https://www.mohurd.gov.cn/gongkai/zc/wjk/art/2022/art_17339_768565.html.

[2] 柏云飞. 智能传感技术在桥梁结构健康监测中的应用研究[J]. 石河子科技, 2024(3): 55-57.
Bai Yunfei. Research on the application of intelligent sensing technology in bridge structure health monitoring[J]. Shihezi Science and Technology, 2024(3): 55-57.

[3] Agyemang I O, Zhang X, Adjei-Mensah I, et al. Automated vision-based structural health inspection and assessment for post-construction civil infrastructure [J]. Automation in Construction, 2023, 156: 105153.

[4] Dong C Z, Catbas F N. A review of computer vision-based structural health monitoring at local and global levels[J]. Structural Health Monitoring, 2021, 20(2): 692-743.

[5] Fang Q, Li H, Luo X, et al. Detecting non-hardhat-use by a deep

learning method from far-field surveillance videos[J]. Automation in Construction, 2018, 85: 1-9.

[6] Fang W, Ding L, Luo H, et al. Falls from heights: a computer vision-based approach for safety harness detection[J]. Automation in Construction, 2018, 91: 53-61.

[7] Dagan D, Isaac S. Planning safe distances between workers on construction sites[J]. Automation in Construction, 2015, 50: 64-71.

[8] 曾拓程, 王佳俊, 钟登华, 等. 基于大场景视频监控的坝面施工机械潜在碰撞风险预警方法[J]. 水利学报, 2024, 55(7): 780-790, 801.
Zeng Tuocheng, Wang Jiajun, Zhong Denghua, et al. Collision risk warning method for dam surface construction machinery based on large-scale video surveillance [J]. Journal of Hydraulic Engineering, 2024, 55(7): 780-790, 801.

[9] Kim J, Kim J, Koo N, et al. Automating scaffold safety inspections using semantic analysis of 3D point clouds[J]. Automation in Construction, 2024, 166: 105603.

[10] 余萍, 宋祥宇. 基于 Open CV 围栏合规性摆放检测方法研究[J]. 科学技术与工程, 2022, 22(28): 12449-12459.
Yu Ping, Song Xiangyu. Detection method of fence compliance based on Open CV [J]. Science Technology and Engineering, 2022, 22(28): 12449-12459.

[11] 苏婷婷, 王娜. 基于图像块相似性和补全生成的人脸复原算法[J]. 科学技术与工程, 2019, 19(13): 171-176.
Su Tingting, Wang Na. Human face restoration based on natural patch likelihood and generative image inpainting [J]. Science Technology and Engineering, 2019, 19(13): 171-176.

[12] Radford A. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. arXiv Preprint, 2015: arXiv: 1511.06434.

[13] Angah O, Chen A Y. Removal of occluding construction workers in job site image data using U-Net based context encoders [J]. Automation in Construction, 2020, 119: 103332.

[14] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing

- and Computer-assisted Intervention-MICCAI 2015: 18th International Conference. Munich: Springer International Publishing, 2015: 234-241.
- [15] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv Preprint, 2014; arXiv: 1411.1784.
- [16] Yan S, Liu Q. Inferring occluded features for fast object detection [J]. Signal Processing, 2015, 110: 188-198.
- [17] Li Z, Li D. Action recognition of construction workers under occlusion [J]. Journal of Building Engineering, 2022, 45: 103352.
- [18] Yoon J, Jordon J, Schaar M. Gain: missing data imputation using generative adversarial nets[C]//International Conference on Machine Learning. Basel: PMLR, 2018: 5689-5698.
- [19] Kanizsa G, Legrenzi P, Bozzi P. Organization in vision: essays on gestalt perception[M]. New York: Praeger, 1979.
- [20] Gu B, Guo H, Huang Y, et al. Computer vision-based human-machine collision warning system on construction site [C]//IOP Conference Series: Earth and Environmental Science. London: IOP Publishing, 2022: 032014.
- [21] 湖北省住房和城乡建设厅, 湖北省市场监督管理局. 建筑施工现场安全防护设施技术规程: DB 42/T 535—2020[S]. 武汉: 湖北省市场监督管理局, 2020.
- Hubei Provincial Department of Housing and Urban Rural Development, Hubei Provincial Market Supervision Administration. Building design standard: DB 42/T 535—2020[S]. Wuhan: Hubei Provincial Market Supervision Administration, 2020.
- [22] Nguyen K, Todorovic S. A weakly supervised amodal segmenter with boundary uncertainty estimation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 7396-7405.
- [23] Dwibedi D, Misra I, Hebert M. Cut, paste and learn: surprisingly easy synthesis for instance detection[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 1301-1310.
- [24] Pérez P, Gangnet M, Blake A. Poisson image editing[M]. Berlin: Springer International Publishing, 2023.