

施工现场的人机距离检测方法综述

郝旭¹ 武文红¹ 牛恒茂² 石宝¹ 乌尼尔¹ 王嘉敏¹ 褚宏坤¹

¹ 内蒙古工业大学信息工程学院 呼和浩特 010080

² 内蒙古建筑职业技术学院建筑工程与测绘学院 呼和浩特 010020

(3459106065@qq.com)

摘要 随着建筑行业的发展,施工机械的使用日益频繁,由此带来的安全问题也愈发严峻。近年来,全国范围内发生的生产安全事故中,建筑起重机械类事故占比显著。因此,如何有效监测并预防施工现场工人与施工机械之间的潜在风险,成为当前研究的热点。首先,系统归纳了基于定位技术和深度学习方法的工人与施工机械距离检测技术,重点介绍深度学习的方法并阐述其关键技术;其次,根据距离检测方法总结国内外的研究现状,并对各方法的优势及局限性进行对比分析;然后,通过目前研究面临的挑战,提出相应的改进策略;最后,给出未来发展趋势,为相关领域的研究者提供有价值的参考。

关键词: 深度学习;施工现场;距离检测;深度估计

中图分类号 TP391

Survey of Man-Machine Distance Detection Method in Construction Site

HAO Xu¹, WU Wenhong¹, NIU Hengmao², SHI Bao¹, WU Nier¹, WANG Jiamin¹ and CHU Hongkun¹

¹ College of Information Engineering, Inner Mongolia University of Technology, Hohhot 010080, China

² College of Construction Engineering and Surveying and Mapping, Inner Mongolia Technical College of Construction, Hohhot 010020, China

Abstract With the development of the construction industry, the use of construction machinery is becoming more and more frequent, and the resulting safety problems are becoming more and more serious. In recent years, among the production safety accidents that have occurred nationwide, construction lifting machinery accidents account for a significant proportion. Therefore, how to effectively monitor and prevent the potential risks between construction site workers and construction machinery become a hot topic of current research. Firstly, this paper systematically summarizes the distance detection technologies between workers and construction machinery based on positioning technology and deep learning method, focusing on the method of deep learning and expounding its key technologies. Secondly, according to the distance detection method, the research status at home and abroad is summarized, and the advantages and limitations of each method are compared and analyzed. Then, through the challenges faced by the current research, the corresponding improvement strategies are proposed. Finally, the future development trend is given for the follow-up research, which provides valuable reference for researchers in related fields.

Keywords Deep learning, Construction site, Distance detection, Depth estimation

随着建筑业的发展,施工机械在施工过程中得到了广泛的应用。然而,由其引发的安全问题也日益凸显,2017—2021年,全国共发生房屋市政工程生产安全事故3622起,死亡4198人,其中建筑起重机械类事故占总数的35%以上^[1];2022年,全国共发生特种设备事故和相关事故108起,死亡101人,其中起重机械事故25起,死亡23人^[2]。因此,对施工现场的工人和机械进行防范,并及时采取预防措施,有助于避免或减少施工现场生产安全事故的发生。

如今,工人与施工机械的安全检测研究的重点还是面向工人的安全检测,如检测工人是否佩戴安全防护装备^[3-6]、工人的不安全行为^[7-9]等;但工人与施工机械的活动交织在一起,

形成了一个复杂多变的空间环境。在这个环境中,工人与施工机械之间的距离、位置关系以及相对位置的动态变化都是影响施工安全的关键因素。因此,实时、准确地恢复和监测空间信息,对于预防事故和提高施工效率具有至关重要的意义。

在测量工人与设备距离的过程中,定位技术被广泛应用,其主要通过部署各类设备^[10-12],直接恢复施工现场工人与设备的空间位置信息。同时,随着深度学习技术的发展,卷积神经网络(Convolutional Neural Networks, CNNs)^[13]也在此领域得到广泛应用,单目深度估计^[14]则因其非接触、实时性强、成本低廉等优点备受关注。该方法通过摄像头捕捉施工现场的图像信息,利用图像处理和分析技术提取目标的位置信息。

基金项目:国家自然科学基金(62066035);内蒙古自治区高等学校科学技术研究项目(NJZY22374);内蒙古自治区自然科学基金(2024QN06021)

This work was supported by the National Natural Science Foundation of China(62066035), Scientific and Technological Research Project of Universities in Inner Mongolia Autonomous Region (NJZY22374) and Natural Science Foundation of Inner Mongolia Autonomous Region (2024QN06021).

通信作者:武文红(wwh801225@163.com)

施工现场工人与施工机械的距离检测,需要判断两者之间的空间关系,恢复目标的深度信息则是距离测量的关键。

Cai 等^[15]的综述侧重于无监督单目深度估计方法的总结以及无监督学习方式下的创新与挑战;而 Jiang 等^[14]梳理和总结了单目深度估计的方法,并分别从监督、半监督和无监督 3 类学习方式进行总结归纳。不同于此,本文首先将介绍基于定位技术以及基于深度学习的测距方法,详细阐述其原理;其次,通过对上述方法在施工现场应用的研究现状进行深入分析,重点介绍基于深度学习的应用研究并揭示当前面临的挑战与难题;然后,结合单目深度估计在实际应用中存在的问题,总结相应的改进策略,最后,基于当前研究的局限性和未解决的挑战提出未来可能的发展趋势,从而为相关领域的研究者提供有价值的参考。

1 人机距离检测方法概述

建筑领域中工人与施工机械的距离检测^[16-19]在不断发展,它涉及到从定位技术的方法到最新的深度学习技术的广泛应用。早期在建筑施工现场用于确定工人与机械位置的方法是定位技术,包括射频识别(Radio Frequency Identification, RFID)、全球定位系统(Global Positioning System, GPS)和超宽带(Ultra-Wideband, UWB)等,这些方法通过部署特定设备来直接恢复出工人与施工机械的空间位置信息。然而,随着深度学习技术的兴起,测距方法也得到了极大的创新。其中,根据相机数量可以划分为单目深度估计和双目深度估计两种方法。单目深度估计通过单个相机捕捉图像,并利用深度学习中的卷积神经网络等技术,从图像中提取目标的深度信息。而双目深度估计则通过两个相机同时拍摄同一场景,利用视差原理计算目标距离。图 1 给出了目前工人与施工机械间测距的方法。

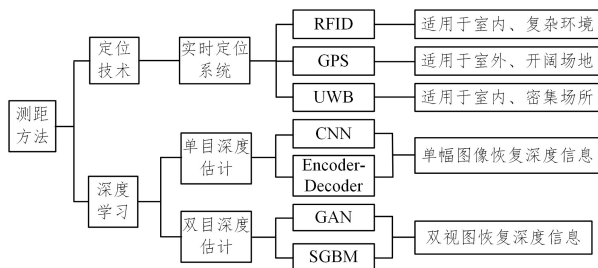


图 1 测距方法

Fig. 1 Ranging method

1.1 基于定位技术的方法

传统的测距方法主要是采用各种距离传感器进行非接触式测距。随着技术的不断发展,在施工现场应用实时定位系统(Real-Time Location System, RTLS)也可以帮助安全管理。RTLS 被定义为能够在其覆盖范围内实时自动检测实体的位置^[20],例如无线射频识别(RFID)、全球定位系统(GPS)、超宽带(UWB)等。

RFID 是一种非接触式的双向数据通信技术,依赖于无线射频信号实现其功能,广泛应用于复杂的室内环境中^[21]。该技术具备对目标对象进行快速识别并实现数据交换的能力。

GPS 系统由多个卫星组成,每个卫星都会不断地向地面发送包含自己位置和时间信息的信号。通过测量一组卫星到 GPS 接收机的距离、GPS 信号从卫星到接收机的持续时间以及光速来计算位置^[22]。

UWB 利用纳秒级的非正弦波窄脉冲传输数据,因此其所占的频谱范围很宽,适用于室内等密集多径场所的高速无线接入。它具有较短的脉冲,使反射信号能够从原始信号中过滤,以帮助克服室内环境中的多径失真,并提供更准确的结果^[23]。

综上所述,RFID、GPS 和 UWB 技术各自展现出独特的优势与适用场景。具体而言,RFID 与 UWB 技术因其技术特性,更适用于室内环境及复杂条件下的精确定位需求;而 GPS 技术则因其依赖于卫星信号,更适用于室外及开阔区域的定位应用。在实际应用过程中,可以根据具体需求及环境条件选择合适的技术。

1.2 基于深度学习的方法

在计算机视觉领域,深度估计是理解和分析三维场景的关键。随着深度学习技术的飞速发展,基于深度学习的测距方法得到了广泛关注。根据相机数量的不同,这些方法可以划分为单目深度估计与双目深度估计。

1.2.1 评价指标

深度估计的评价指标主要有均方根误差(RMSE)、对数均方根误差(RMSE log)、相对误差(Abs Rel)、平方相对误差(Sq Rel)和准确率(Accuracy)这 5 种评价指标^[24],详细的计算公式及评价准则如表 1 所列。其中, d_i 表示像素 i 点处的预测灰度值; d_i^* 表示真实灰度值; N 表示真实灰度值像素点数量; thr 表示阈值,常取 $1.25, 1.25^2, 1.25^3$ 。

表 1 深度估计评价指标

Table 1 Evaluation index of depth estimation

评价指标	计算公式	评价准则
均方根误差	$RMSE = \sqrt{\frac{1}{N} \sum_{i \in N} \ d_i - d_i^*\ ^2}$	表示预测深度和真实深度之间的平均差值的平方根。RMSE 越小,表示误差越小,性能越好
对数均方根误差	$RMSE \log = \sqrt{\frac{1}{N} \sum_{i \in N} \ \log(d_i) - \log(d_i^*)\ ^2}$	表示预测深度和真实深度之间的绝对差值与真实深度的比值的平均值
相对误差	$Abs \text{ Rel} = \frac{1}{N} \sum_{i \in N} \frac{ d_i - d_i^* }{d_i^*}$	反映预测深度的可信度,相对误差越小,表示预测深度越接近真实深度
平方相对误差	$Sq \text{ Rel} = \frac{1}{N} \sum_{i \in N} \frac{\ d_i - d_i^*\ ^2}{d_i^{*2}}$	与相对误差相比,平方相对误差对较大的误差更为敏感,因此可以更准确地反映预测深度与真实深度之间的差异
准确率/%	$Accuracy \text{ of } d_i \text{ s. t. } \max\left(\frac{d_i}{d_i^*}, \frac{d_i^*}{d_i}\right) = \delta < thr$	反映分类器对整个样本的判定能力,即将正的判定为正,负的判定为负的能力

1.2.2 单目深度估计

CNN^[13]在图像处理和计算机视觉任务中展现了卓越性能,尤其是在特征提取方面。CNN通过多层卷积、池化和全连接层的学习过程,能够自动地从原始图像中提取有效的特征表示,并逐步抽象出更高级别的特征。

在单目深度估计领域,CNN同样发挥了关键作用。单目深度估计是指仅使用一个摄像头捕获的单一视角图像来估计深度信息。相较于双目或多目深度估计,单目深度估计具有设备简单、成本低廉的优势,且可以方便地集成到各种设备上。

单目深度估计^[25]的基本原理是通过分析图像数据来推断场景中物体的相对或绝对距离。主要通过CNN提取图像特征,然后采用深度估计输出深度图。具体流程如图2所示。

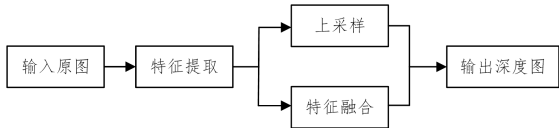


图2 单目深度估计流程图

Fig. 2 Flow chart of monocular depth estimation

在计算机视觉中,通常用深度图来表示这种距离信息,其中每个像素的值对应于场景中某一点的深度值。为了实现这一目标,深度学习模型需要学习从二维图像到三维深度图的映射关系。

传统的单目深度估计方法主要基于图像本身的特征和线索来计算深度值,如边缘、纹理、阴影^[26-28]等。然而,这些方法往往受到光照、遮挡等因素的影响,且难以处理复杂的场景。

随着深度学习技术的发展,基于CNN的单目深度估计方法逐渐成为主流。这些方法通过大量训练数据来学习图像到深度图的映射关系,具有更高的精度和鲁棒性。

在基于深度学习的单目深度估计方法中,Encoder-Decoder结构得到了广泛的应用。Encoder部分负责从输入图像中提取特征信息,通过多层卷积和池化操作,逐步将图像转换为紧凑的特征表示。Decoder部分则负责将编码器提取的特征信息解码为深度图,通过上采样和卷积操作,逐步恢复图像的空间分辨率,并生成与输入图像对应的深度图。

在训练过程中,基于深度学习的单目深度估计方法通常采用监督学习^[29]的方式。模型需要输入大量的带有真实深度标签的图像数据。这些数据可以通过激光雷达、RGB-D相机等设备获取。通过最小化预测深度图与真实深度图之间的误差,模型可以逐渐学习到从图像到深度图的映射关系。

在模型架构方面,基于深度学习的单目深度估计方法通常采用编码器-解码器结构。编码器部分负责提取图像中的特征信息,解码器部分则负责将特征信息转换为深度图。图3是深度估计网络的结构图^[30]。首先, $n \times n$ convolution模块表示采用 $n \times n$ 大小的卷积核进行线性变换,这一过程有效地提取了图像中的局部特征,为后续处理提供了基础。其次, Batch normalization层作为网络优化的一部分,负责对激活值进行归一化处理,通过稳定每一层的输入分布,加速训练过程并提高模型的泛化能力。最后,Up-Convolution模块,即上采样卷积层,在解码器部分,它通过增加特征图的空间维度,逐步恢复出原始图像的空间分辨率,是生成深度图的关键步骤。

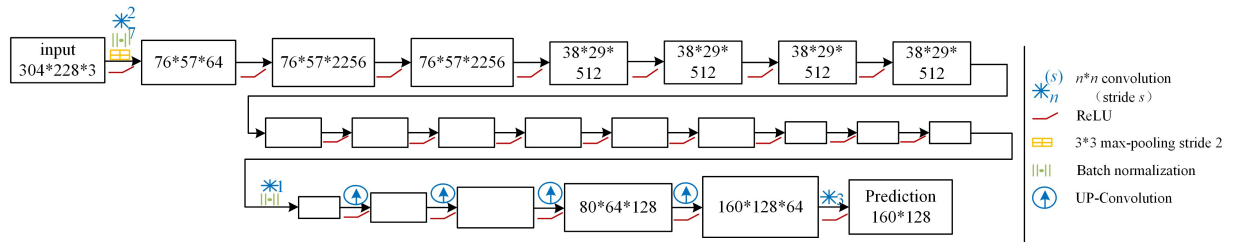


图3 深度估计网络

Fig. 3 Deep estimation network

1.2.3 双目深度估计

双目深度估计^[31]是利用两台摄像机,位置相对固定,同时拍摄被测物体的图像,得到两张或多张图像的左右视图并进行畸变矫正;然后,通过找出这些图像的对对应关系,即进行特征点提取和匹配,建立被测物的几何模型;最后,通过“视差”测量原理,将图像二维坐标转换到空间三维坐标,得到被测物体的三维信息。图4为双目深度估计流程图。

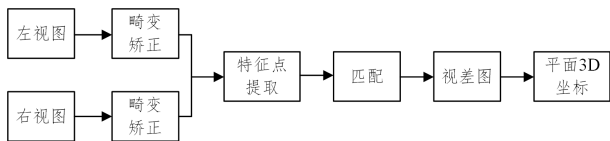


图4 双目深度估计流程图

Fig. 4 Flow chart of binocular depth estimation

随着深度学习和计算机视觉技术的不断发展,生成对抗网络(Generative Adversarial Nets, GAN)和半全局块匹配(Semi-Global Block Matching, SGBM)算法为这一领域提供了新的解决方案。

生成对抗网络(GAN)通过其独特的学习机制,能够模拟并生成与真实数据极其相似的新数据。在建筑工地测距技术的应用中,GAN可以学习施工现场的真实环境数据分布,并生成模拟的施工现场数据^[32]。这些模拟数据可以涵盖各种可能的施工情况和工人与机械的位置关系,从而为测距提供丰富的数据源。

另一方面,SGBM算法作为一种高效的立体匹配方法,在双目深度估计中发挥着重要作用。SGBM算法^[33]通过计算两幅图像之间对应像素块的相似度来估计场景的深度。相较于单目深度估计,双目深度估计可以获取更准确的深度信息,但设备成本较高,且计算复杂度较大。

综上所述,基于深度学习的测距方法在计算机视觉领域具有广泛的应用前景。其中,单目深度估计以其设备简单、成本低廉的优势受到广泛关注。单目深度估计方法作为一种新兴的技术手段,在提升测距精度和鲁棒性方面展现出了巨大的潜力。未来,随着技术的不断进步和应用场景的不断拓展,基于深度学习的测距技术将在更多领域发挥重要作用。

2 施工现场人机距离检测应用研究

施工现场工人与施工机械间的距离检测主要是判定工人是否处于危险区域,而危险区域是指在建筑施工现场存在潜在危险源的区域,这些危险源可能包括高处坠落、坍塌、机械伤害、触电等。危险区域的识别是预防事故发生的前提和基础。根据实际场景并结合 Guo 等^[34]的研究,可将机械伤人的危险区域划分为落物区和碰撞区,详细划分规则如表 2 所列。

表 2 危险区域的划分范围

Table 2 Division range of dangerous areas

危险区域	施工机械	危险范围
落物区	塔吊及吊车 起重机	旋转状态时,机身为中心、悬臂长为半径的圆形范围
碰撞区	挖掘机 压路机 运输车	工作区域外扩 T 时间可达范围

然而在图像或视频处理中,实现人机距离检测的核心在于准确获取目标的三维信息。这一目标对于确保人机交互的安全性、有效性以及智能化至关重要。下文将深入探讨并阐述各种测距技术在人机检测领域的具体应用和研究进展。

2.1 基于定位技术的人机距离检测

在建筑安全领域的研究受益于各种定位设备。全球导航卫星系统(GNSS)等工具已经变得突出,而实时定位系统(RTLS)也越来越受欢迎。这些 RTLS 包含一系列技术,包括 RFID、GPS 和 UWB 等^[35]。

Duan^[36]和 Fan^[37]采用 GPS 技术搭建施工现场工人与施工机械的轨迹跟踪系统;Zhang^[38]和 Dai^[39]利用 RFID 对人员、材料和机械进行识别和定位;Teizer^[40]和 Zhang 等^[41]采用 UWB 的定位方法获取工人的空间位置信息。此外,激光雷达点云方法对三维空间关系识别^[42-43]是有效的,也用于施工安全^[44],通过深度预测重建伪激光雷达点云。然而,该方法在估计深度信息时存在标量恢复和表示可转移性的问题。

随着技术的发展,移动技术、云计算和可穿戴设备^[45]的发明为推进该领域的研究前沿创造了更好的可能性。Kong^[46]进行了基于低功耗蓝牙(BLE)的位置跟踪与建筑物信息融合的研究,通过定义不安全区域监测工人的位置,并根据工人与不安全区域的接近程度进行统计分析,最后利用建模(BIM)和基于云的通信来自动监测建筑工地的安全状况。

综上所述,尽管多种定位技术为实时监测施工现场安全提供了可能性,但由于其高昂的初始投入成本,这些技术在实际应用中难以实现大规模批量部署。此外,诸如 RFID 和 GPS 的 RTLS 方法受限于工人需要携带单独的设备以便定位的要求。这一操作不仅增加了施工的复杂度,还可能对工人的正常施工活动造成一定程度的干扰。因此,利用定位技术监测工人与设备位置信息还存在一定的局限性。表 3 列出了各定位技术在施工现场应用中的优势与局限性。

表 3 定位技术的优势及局限性

Table 3 Advantages and limitations of positioning technologies

方法	优势	局限性
RFID	成本低,适用于白天与黑夜	精度差
UWB	易于安装,数据采集准确	成本相对较高,容易耗电
GPS	远距离检测差	造价高

2.2 基于深度学习的人机距离检测

当前,单目深度估计主要依赖于小孔成像模型和三角测量原理来实现人机距离的精确检测。Shi 等^[47]运用相机成像的几何原理,成功地计算出了工人与吊车危险区域之间的距离。Kim 等^[48]则提出了一种基于计算机视觉的创新方法,该方法利用无人机获取数据,并结合目标检测实现人机距离检测,从而对建筑设备和工人的位置进行高效估计。此外,还有一种通过二维图像来测量位置和距离的技术被提出。Shan^[49]提出了一种在俯视图角下进行的单目测距方案,其通过对相机标定获得相机的内外参数,并结合小孔成像原理实现卡车盲区目标的距离检测。

虽然这些方法在理论上无需引入额外的传感器即可实现距离测量,但在实际应用中,复杂的施工现场环境对它们提出了严峻的挑战。具体而言,由于摄像头的部署角度并非始终固定且单一,这些方法在处理多角度或倾斜视角的图像时可能会受到限制。因此,对于非标准视角下的图像,其测距或深度估计的精度和适用性可能会受到显著影响。

在双目深度估计的领域中,SGBM 算法占据了核心地位。该算法通过全局最优化的搜索策略,在图像块之间执行精确的匹配过程,以达成视差图的高精度计算和鲁棒性提升。已有研究^[49-52]采纳了双目视觉技术,文献[50]在 YOLOv5s 目标检测的基础上,结合 SGBM 算法的双目测距技术,引入空间距离作为报警触发条件,实现了更精确的人员安全监控。同时,通过引入注意力机制优化特征提取,解决了光照不足导致的图像模糊问题,进一步提高了目标识别的精度。文献[51-53]则利用双目视觉技术结合相机标定的方法,并利用相机的内外参数信息实现距离测算,从而为施工现场的安全监控提供了有力的技术支持。虽然双目相机可以获取安全施工区域^[54]以及定位^[55]物体的深度和位置,但由于其可视范围较短,价格相对较高,维护工作量大,很少部署在建筑工地上。同样地,通常在 1~8m 范围内采集深度信息的 RGB 深度相机,如 VICON 和 Microsoft Kinect,特别容易受到来自铁磁辐射和太阳辐射的各种噪声的影响,因此不适用于室外环境^[56]。另一方面,用于估计深度的立体相机是现有 3D 检测模型^[57]的替代选择。然而,立体相机通常伴随着较大的计算复杂性和带宽需求,且其在建筑环境中的部署相对受限。这些限制条件显著制约了立体相机在建筑领域的实际应用。

随着深度学习技术的进步,视觉传感领域的技术在工人安全监控研究中得到了广泛探索,包括在复杂环境中准确检测工人、设备和材料的能力^[58-59]。另外,该技术还能够有效提取工人的运动轨迹,并实时监测可能导致安全事故的异常行为。在深度学习的诸多应用中,基于 CNN 的方法尤为突出,它们被成功应用于工人的检测与跟踪^[60-61]、安全帽检测^[62]、各种姿势和变化背景下工人姿态的分析^[63],以及识别有摔倒风险的工人^[64]。

在人机距离检测的范畴内,基于深度学习的方法不仅局限于工人和设备的检测与定位,更进一步扩展到了预测事故发生概率以及评估工人发生事故风险的研究中^[65-66]。一系列研究成果表明,虽然通过计算工人和设备检测的中心坐标来预测碰撞概率是一种方法,但这种方法往往因忽视设备的实际体积而低估了真实距离^[67]。因此,有研究者建议直接使用深度学习检测结果中的包围盒大小来考量设备的体积,以

更准确地反映实际距离。例如,检测建筑实体及其关键顶点的二维包围盒,就是重建每个对象^[56]的三维包围盒。然后,该方法根据摄像机光轴在三维世界坐标系中与地平面平行的假设,在已知内参数的情况下,估计施工安全深度和范围。以上针对深度信息的模型都没有在公开的数据集上进行测试和验证,因此这些传统的学习范式很难被评估。

为了突破现有技术的局限,基于单相机的深度估计技术因其低成本、高帧率和高分辨率以及更广泛的适用性^[68],成为了研究热点,并展现出极高的可扩展性和潜在应用价值。在文献[69-72]中,研究者们利用单目深度估计恢复出目标的深度信息,为深度信息的感知和理解提供了有力的支持。其中,文献[69]针对施工现场重型设备作业区域的安全问题,通过结合 RGB 图像、Mask 区域信息及深度信息,将二维目标转

换为三维点云空间进行处理,实现了对工人与重型设备之间空间关系的判断。文献[71]通过迁移学习等方法,构建了适用于施工现场的深度估计数据集和损失函数,实现了对施工现场物体深度的有效估计。文献[72]考虑单目深度估计面临的诸多挑战,如光照变化、遮挡、场景复杂度等,提出了一种基于多尺度特征提取、编解码结构与分阶段优化策略的单目图像深度估计方法,以此提升单目深度估计的性能。总的来说,这些研究不仅展示了单目深度估计的独特优势,还进一步拓宽了其在实际应用中的可能性和潜力。然而,值得注意的是,现有的数据集主要聚焦于自动驾驶领域,对于施工场景^[73]而言,获取相关公共数据集仍面临一定的挑战。因此,针对施工场景的单目深度估计的研究和发展显得尤为重要。表 4 列举了部分单目深度估计的方法。

表 4 单目深度估计的方法
Table 4 Monocular depth estimation methods

方法	结构	优势	局限性
ACAN ^[74]	Encoder-Decoder	充分利用上下文信息,提高深度估计准确度	对数据集要求高,泛化能力弱,网络结构相对复杂
DenseDepth ^[75]	Encoder-Decoder	采用对抗性和多尺度进行训练,能够处理复杂的场景和图像	需要大量数据进行训练,计算资源要求较高
BTS ^[76]	Encoder-Decoder	引入了新型局部平面制导层,提升模型精度,保持高精度的同时具有较好的实时性	模型的性能取决于数据的质量,对数据依赖较高
Monodepth2 ^[77]	Encoder-Decoder	提高了深度估计的精度和鲁棒性;自动掩码机制有利于处理遮挡和动态物体	需要标注的深度图进行训练;复杂场景和极端光照条件精度较差
NeW CRFs ^[78]	Transformer	利用全连接 CRFs 的强大表征力,提升深度估计精度;适用于全景图深度估计	计算量较大,对硬件要求较高

综上所述,尽管单目深度估计在当前研究中受到广泛关注,但在实际应用中仍面临各种挑战。一方面,其精度相较于激光测距尚显不足,这导致在恢复目标空间信息时,估计信息与真实信息之间存在较大的误差。另一方面,当前公开用于深度估计的数据集主要集中在自动驾驶领域,这也在一定程度上限制了其在其他领域的应用和发展。因此,下文将针对人机距离检测过程中面临的问题,提出相应的改进方法。

3 存在的问题及改进策略

在建筑施工领域的实际应用中,关于单目深度估计恢复目标深度信息的研究相对较少,因此其在真实场景下的应用面临着诸多挑战与限制。首先,现有的单目深度估计方法大多基于深度学习的技术,对数据集的依赖性较强。然而,建筑施工领域的专用数据集相对匮乏,导致模型训练困难,泛化能力受限。其次,在建筑施工环境中,单目深度估计技术面临着复杂多变的光照条件、遮挡现象以及目标形态多样性等挑战。这些因素极大地影响了深度信息的准确性和可靠性,限制了其在建筑施工过程中的应用效果。

3.1 数据集问题

目前,针对真实世界深度信息的获取,已存在一系列公开数据集,如 KITTI 数据集^[79]等,它们被广泛用于自动驾驶领域的相关研究。然而,在建筑施工领域^[44,56]获取此类包含深度信息的公开数据集仍然面临极大的挑战。鉴于标记数据的稀缺性和标注过程的高成本,无标签的建筑数据集成为挑战,研究者们主要通过数据增强及下采样、无监督、半监督以及自监督等方法来解决建筑数据集无标签的问题。其中,自监督学习方法成为一种有效的替代方案。这种方法通过利用大量

未标记数据来训练深度网络,进而学习数据中的内在结构和规律,展现出在某些场景下甚至优于传统监督学习方法的潜力^[80]。在自监督学习方法的推动下,自监督单目深度估计(SSMDE)技术在近年来取得了显著的进步,并在多个基准测试中取得了最先进的性能^[81-82]。其主要通过设计合理的自监督任务(如视图合成、光流估计等),使模型在无需显式深度标签的情况下,学习从单目图像中预测深度信息的能力。这种能力不仅降低了对大规模标注数据集的依赖,还促进了深度估计技术在更多实际场景中的应用。Shen 等^[44]针对建筑场景分析,深入研究了自监督表示学习方法,并将相关模型成功迁移至该领域。他们通过设计适用于建筑环境的自监督学习任务,如利用建筑物结构的几何一致性或时间序列图像中的变化等,训练深度网络,以学习建筑场景中的深度信息。这一研究为建筑场景下的深度估计提供了新的解决思路。在三维场景重建与深度估计领域,多个数据集的发布极大地推动了计算机视觉与机器学习研究的进展。Make3D^[83]作为早期探索,通过提供 534 对高分辨率室外彩色图像及其对应的激光扫描深度图,为从单目图像中估计深度奠定了基础。随后,NYU Depth v2^[84]数据集扩展了室内场景的理解范畴,其包含的 1449 对 RGB-D 图像,由 Microsoft Kinect 相机捕获,不仅促进了三维重建、物体识别与姿态估计等任务的发展,还成为室内环境感知研究的重要资源。

为了进一步推动深度估计技术的泛化能力,Depth in the Wild^[85]数据集应运而生。该数据集通过收集近 50 万张来自互联网的多样化图像,并借助众包方式标注深度信息,构建了一个大规模且覆盖广泛场景与对象类别的深度图像库。在自动驾驶领域,ApolloScape^[86]数据集则代表了自动驾驶场景理解的新高度。该数据集不仅涵盖了来自中国多个城市的复杂

交通场景,还集成了高分辨率 RGB 图像、激光雷达点云及详尽的 3D 物体标注,为深度估计、语义分割、实例分割及三维物体检测等关键任务提供了全面的数据支持。

此外,Cityscapes^[87]数据集通过车载相机采集的来自 50 个不同城市的图像数据,结合像素级及实例级的语义分割标签与深度图。综上所述,这些数据集不仅丰富了计算机视觉与机器学习领域的研究内容,也为技术的实际应用奠定了坚实的基础。表 5 归纳总结了各类公开数据集。

表 5 公开数据集
Table 5 Public datasets

数据集	图片数量	适用场景	采集方式
Make3D ^[83]	534	室外	激光扫描仪
KITTI ^[79]	93000+	室外	激光扫描仪
NYU depth v2 ^[84]	1449	室外	RGB-D 相机
Depth in the wild ^[85]	495000	综合	互联网+人工标注
ApolloScap ^[86]	147000+	室外	激光雷达
Cityscapes ^[87]	222973	室外	相机+人工标注

3.2 深度信息缺失问题

首先,单目深度估计依赖于从单一视角的图像中推断深度信息,这导致了其在处理复杂施工场景时的局限性。施工现场通常包含多种不同的物体和遮挡情况,如建筑材料、设备、工人等,它们之间的相互作用和遮挡可能导致某些区域的深度信息无法准确获取。

为应对运动物体遮挡的难题,Godard 等^[88]提出了一种创新的损失函数设计,优化深度估计模型在动态场景中的表现。同年,Casser 等^[89]则提出了一种策略,通过减少动态场景中运动物体的深度信息重投影误差,来增强模型的鲁棒性。Guizilini 等^[90]进一步提出了一种名为 PackNet 的新型网络架构,该架构专注于学习视觉信息以获取更优的深度预测性能。此外,Guiziliniu 等^[81]采用了预训练网络来生成具备语义感知能力的特征。这一方法不仅提升了深度估计的精度,同时也为处理动态场景问题提供了有力的支持。然而,引入语义分割等额外约束可能增加模型的计算负担。为此,Peng 等^[91]提出了一种新颖的方法,通过自蒸馏损失和数据嫁接技术来降低模型的计算成本。Bian 等^[92]则针对相邻图像间深度估计不一致的问题,引入了一种新的几何一致性损失。该损失函数有效地提升了深度估计模型在不同图像帧间的连贯性和稳定性,为动态场景下的深度估计提供了更为可靠的解决方案。

此外,施工现场的光照条件和环境变化也是导致深度信

息缺失的重要因素。光照不均匀、阴影遮挡、天气变化等都可能影响图像的视觉效果,进而影响深度估计的准确性。特别是在光线不足或恶劣天气条件下,图像的清晰度和对比度可能大幅下降,使得深度信息的提取变得更加困难。

在深度估计领域,弱光环境是一个公认的挑战。虽然已存在受 GAN^[93]启发的方法用于图像增强,但这些方法并未专门针对深度估计任务中的光照条件进行优化。弱光环境不仅影响图像的亮度、对比度和噪声水平,还可能导致纹理细节丢失和边缘模糊,这些因素都会直接影响深度估计的准确性。

为了提升深度估计在弱光环境下的性能,研究者们开始探索注意力机制的应用。Hu 等^[94]提出的通道注意力机制,能够根据不同通道在弱光条件下的表现差异,为每个通道分配不同的权重,从而实现对重要特征的有效提取。Woo 等^[95]则进一步研究了如何通过级联和并行的方式,将不同类型的注意力机制进行融合,以捕捉图像中更为复杂的上下文信息。Yang 等^[96]则提出了一种基于非局部均值和注意力机制的模块,该模块能够捕获特征图中像素之间的长程依赖关系,从而增强模型对弱光环境中光照变化的适应能力。

为了充分利用图像中的多尺度信息,Wu 等^[97]提出了一种多层上下文和多模态融合网络 MCMFNet。该网络能够集成不同尺度的上下文特征映射,并通过学习深度图中的目标边缘来增强深度估计的精细度。同时,针对低分辨率深度图缺乏场景中的空间和物体信息的问题,研究者们也在探索如何通过高分辨率深度图来补充这些信息,以提高深度估计的语义准确性。

在弱光环境下,深度估计任务往往面临更大的挑战。为了提升模型的性能,研究者们开始尝试利用语义信息作为辅助指导。Jiao 等^[98]提出了一种具有代表性的深度估计方法,该方法充分利用了语义信息来指导深度估计任务。辅助信息大致可以分为两类:一类是单独使用某一方面的感知信息,如光照、纹理等;另一类则是使用多个场景感知信息的组合^[99]。在利用这些辅助信息时,关键在于如何增强模型对信息的整合能力,以充分利用语义辅助信息来指导深度估计任务。

在深度估计领域,特别是在处理复杂施工场景时,面临的挑战尤为显著。施工场景中的多种物体和遮挡情况,以及弱光环境下光照条件的变化,均对深度信息的准确获取造成了困难。为解决这些问题,研究者们提出了一系列创新的方法。表 6 列出了改进深度估计的其他方法。

表 6 改进策略

Table 6 Improvement strategies

改进策略	关键技术	优势	局限性
改进网络结构的方法 ^[100-103]	ResNet,FCN、多尺度信息提取以及结合循环神经网络和生成式对抗网络	提升模型的特征提取能力和全局信息的感知能力	增加了模型的复杂性和计算量,局部信息易丢失
改进损失函数的方法 ^[104-105]	分层次嵌入损失、自动均衡损失权重	简化深度估计的复杂性,提升模型的准确性和鲁棒性	针对特定场景,泛化能力有限
分类的方法 ^[106-108]	有序回归框架、特征融合	降低了计算复杂度,保留了深度信息的连续性	损失精度,增加计算成本
部分深度信息的方法 ^[109-110]	条件变分自编码器、最大先验估计	更好地捕获物体的边缘和细节,多任务并行提高模型的效率	信息量较少,预测位置区域深度时增加不确定性

综上所述,单目深度估计在施工现场工人与设备距离检测中面临深度信息缺失的问题,这主要体现在复杂场景处理、光照和环境变化以及相对位置变化等方面。为了解

决这些问题,需要深入研究单目深度估计的原理和算法,并结合施工现场的实际情况,探索更加准确和可靠的深度估计方法。

4 未来发展趋势

在建筑施工领域,采用深度估计的方法检测工人与施工机械的距离,未来发展趋势可以归纳为以下几点。

1) 标记数据的稀缺性和标注过程的高成本,使得自监督与弱监督学习方法将在未来深度估计任务中占据更加重要的地位。这些方法能够利用大量未标记或仅部分标记的数据进行模型训练,有效减少对数据标注的依赖,同时提高模型的泛化能力。

2) 针对建筑施工领域的特定场景(如弱光环境、动态场景等),未来研究将更加注重模型的优化和定制。注意力机制和上下文信息的利用将成为提升深度估计性能的关键。通过引入注意力机制,模型能够更加关注图像中的重要区域,提高特征提取的准确性和效率。同时,上下文信息能够帮助模型更好地理解图像中的场景和物体关系,提高深度估计的语义准确性;语义信息能够提供关于场景和物体的额外信息,有助于指导深度估计过程。未来研究可以注重语义信息的辅助与整合,通过结合语义分割、目标检测等任务的结果,提高深度估计的准确性和鲁棒性。

3) 建筑施工现场对深度估计任务的实时性要求较高,因此模型轻量化将是未来研究的重要方向。通过采用轻量级网络结构、剪枝、量化等技术手段,降低模型的计算复杂度和存储需求,提高模型的实时性和部署便捷性。

4) 随着传感器技术的不断发展,建筑施工现场将能够获取到包括图像、视频、激光雷达等在内的多模态数据。未来的深度估计模型将更加注重多模态数据的融合,以充分利用各种数据源中的互补信息,提高深度估计的准确性和鲁棒性。

结束语 采用深度学习的方法检测施工现场下人机的距离,节省人力成本的同时也能够降低机械伤人事故发生的概率。虽然测距任务能够达到实际需求,但仍存在数据集匮乏以及深度信息丢失的问题,未来还需对深度学习的方法继续学习并加以改进,从而更准确地检测施工现场的人机安全。因此,基于深度学习的施工现场人机距离检测有很大的研究前景。

参考文献

- [1] Analysis of building safety situation in China[DB/OL]. <https://zhuanlan.zhihu.com/p/562479748>.
- [2] 2022 National Special Equipment Safety Report[DB/OL]. https://www.samr.gov.cn/tzsbj/qktb/tb/art/2023/art_67857a52cc3b41bd9cff0560e6b9b503.html.
- [3] WU L F. Safety belt recognition and detection for aerial work based on convolutional neural network [J]. *Modern Computer*, 2021, 27(24): 124-130.
- [4] LI H, WU L Z, XUE X C, et al. Safety inspection of high-altitude operations based on computer vision [J]. *Chinese Journal of Safety Science*, 2023, 33(9): 69-75.
- [5] YANG D W, ZHANG C C. Detection algorithm of helmet wearing based on improved YOLOv7 [J]. *Journal of Shenyang University of Technology*, 2024, 43(1): 16-21.
- [6] FENG P Y, QIAN Y R, FAN Y Y, et al. Safety helmet detection algorithm based on improved Cascade R-CNN [J]. *Microelectronics and Computer*, 2024(1): 63-73.
- [7] FAN B Q, DONG B Y, WANG B, et al. Identification and application of unsafe behaviors of subway construction workers based on deep learning [J]. *Chinese Journal of Safety Science*, 2023, 33(1): 41-47.
- [8] CHENG J, CHENG Z, SU Z, et al. Unsafe behavior management technology based on deep learning [J]. *Construction Technology (Chinese and English)*, 2023, 52(20): 118-121.
- [9] LI W J, LIU X. Research on identification and early warning system of unsafe behavior of underground personnel based on deep learning [J]. *Metal Mine*, 2023(3): 177-184.
- [10] JIANG W, ZHANG L, ZHANG Y, et al. Infrastructure construction personnel management system supported by mobile positioning technology [J]. *VW Standardization*, 2022(23): 70-72.
- [11] SONG Z Y. Research on early warning of building construction object strike accident [J]. *Engineering Technology Research*, 2023, 8(8): 214-216.
- [12] DER Y W, CHANG H K, HSIE Y C, et al. An Image Recognition-based Distance Measurement Technique for Real-time Locating of Construction Workers[C]// 2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability(ECBIO). IEEE, 2021: 148-151.
- [13] LINDSAY G W. Convolutional neural networks as a model of the visual system: Past, present, and future [J]. *Journal of Cognitive Neuroscience*, 2021, 33(10): 2017-2031.
- [14] JIANG J J, LI Z Y, LIU X M. Review of monocular depth estimation methods based on deep learning [J]. *Journal of Computer Science*, 2022, 45(6): 1276-1307.
- [15] CAI J C, DONG F M, SUN S F, et al. A review of unsupervised monocular depth estimation research [J]. *Computer Science*, 2024, 51(2): 117-134.
- [16] DING J F, XIAO W T, LI M Y, et al. Substation cross-border violation detection combining visual perception and RTK positioning [J]. *Computer Technology and Development*, 2023, 33(8): 206-213.
- [17] SUN Y H, XUE Z G, WANG C B, et al. Research on subway construction personnel positioning technology based on video image analysis and monocular vision technology [J]. *Electronic Design Engineering*, 2023, 31(15): 187-190, 195.
- [18] ZHANG X H, CHEN X, YANG W J, et al. Research on visual positioning method of anchoring equipment based on single laser beam information [J]. *Coal Science and Technology*, 2024, 52(1): 311-322.
- [19] LI K W. Research and application of monocular depth estimation for transmission line scenarios [D]. Jinan: Shandong University, 2023.
- [20] LI H, CHAN G, WONG J K W, et al. Real-time locating systems applications in construction [J]. *Automation in Construction*, 2016, 63: 37-47.
- [21] PAUL A, PULANI S, MAHESWARI J U. Digital Twin Framework for Worker Safety using RFID Technology[C]// ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. IAARC Publications, 2023: 254-261.
- [22] BRUSSELAERS N, HUANG H, MACHARIS C, et al. A GPS-based approach to measure the environmental impact of construction-related HGV traffic on city level [J]. *Environmental*

- Impact Assessment Review, 2023, 98:106955.
- [23] WU Y F, LIU S, LIU T Y. Research on indoor motion localization algorithm based on UWB [J]. Computer Programming Skills and Maintenance, 2024(4):151-153.
- [24] EIGEN D, PUHRSCHE C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network[J]. Advances in Neural Information Processing Systems, 2014, 27.
- [25] ZHANG Y P. Vehicle detection and ranging based on convolutional neural network [D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [26] SHIBA Y, ONO S, FURUKAWA R, et al. Learning-based active 3D measurement technique using light field created by video projectors[J]. IPSJ Transactions on Computer Vision and Applications, 2019, 11:1-14.
- [27] BAO W, KOMATSU R, ASAMA H, et al. Applying Albedo Estimation and Implicit Neural Representations to Well-Posed Shape From Shading[J]. IEEE Access, 2023, 11:40038-40048.
- [28] LUO H, GAO B, XU J, et al. An approach for structured light system calibration[C]//2013 IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems. IEEE, 2013:428-433.
- [29] LI Z, CHEN Z, LIU X, et al. Depthformer: Exploiting long-range correlation and local information for accurate monocular depth estimation[J]. Machine Intelligence Research, 2023, 20(6):837-854.
- [30] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks [C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016:239-248.
- [31] WANG D D, HOU J H, WANG F Q, et al. Research on target detection and ranging based on binocular vision [J]. Electronic Production, 2022, 30(21):58-61.
- [32] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11):139-144.
- [33] ZHAO C X, ZHANG X L, YANG Y. Three-dimensional reconstruction based on SGBM semi-global stereo matching algorithm [J]. Laser Magazine, 2021, 42(4):139-143.
- [34] GUO H L, LIU W P, ZHANG W S. Research on early warning system of workers' unsafe behavior integrating BIM and PT [J]. Chinese Journal of Safety Science, 2014, 24(4):104-109.
- [35] LI H, CHAN G, WONG J K W, et al. Real-time locating systems applications in construction[J]. Automation in Construction, 2016, 63:37-47.
- [36] DUAN L J, LI J K, ZHANG B. The application of engineering personnel and vehicle positioning in the construction of hydropower stations—Taking the Changlongshan Hydropower Station as an example [J]. People's Yangtze River, 2019, 50(S1):354-357.
- [37] FAN L P. Research on safety situation assessment of construction workers' trajectory based on real-time positioning [D]. Xuzhou: China University of Mining and Technology, 2023.
- [38] ZHANG J J, HAN Y, MA G X, et al. Research on intelligent early warning system for falling accidents of construction workers based on BIM and RFID [J]. Journal of Engineering Management, 2015, 29(6):17-21.
- [39] DAI M Q. Study on hazard identification of construction site based on RFID technology [D]. Shanghai: East China University of Technology, 2014.
- [40] TEIZER J, VENUGOPAL M, WALIA A. Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and material positioning and tracking[J]. Transportation Research Record, 2008, 2081(1):56-64.
- [41] ZHANG Z T, WANG Y Y, LUO Z B, et al. Automatic detection of dangerous scenes under the interaction of tower crane and worker space [J]. Journal of Tsinghua University(Natural Science Edition), 2024, 64(2):198-204.
- [42] LIU H, LIAO K, LIN C, et al. Pseudo-lidar point cloud interpolation based on 3d motion representation and spatial supervision [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7):6379-6389.
- [43] QIAN R, GARG D, WANG Y, et al. End-to-end pseudo-lidar for image-based 3d object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:5881-5890.
- [44] SHEN J, YAN W, LI P, et al. Deep learning-based object identification with instance segmentation and pseudo-LiDAR point cloud for work zone safety[J]. Computer-Aided Civil and Infrastructure Engineering, 2021, 36(12):1549-1567.
- [45] LI M, LIU P, YU H L. Research on real-time safety early warning system for construction site construction personnel based on wearable technology [J]. Industrial Safety and Environmental Protection, 2018, 44(11):32-35.
- [46] KONG F S. Design and test analysis of beacon sensor early warning system in construction dangerous area [D]. Xi'an: Xi'an University of Architecture and Technology, 2021.
- [47] SHI H T, LI G, FAN L T, et al. Crane dangerous area intrusion pedestrian machine vision positioning algorithm [J]. Control Engineering, 2021, 28(4):759-765.
- [48] KIM D, LIU M, LEE S H, et al. Remote proximity monitoring between mobile construction resources using camera-mounted UAVs[J]. Automation in Construction, 2019, 99:168-182.
- [49] SHAN F. Study on the method of front target detection and ranging of mining trucks based on deep learning [D]. Shenyang: Northeastern University, 2020.
- [50] SHEN J, HUANG X H. Dangerous area intrusion alarm based on improved YOLOv5s algorithm [J]. Computer System Application, 2023, 32(3):157-162.
- [51] GAO Y F. Research on target detection and ranging based on binocular vision [D]. Taiyuan: Taiyuan University of Science and Technology, 2020.
- [52] WANG K Q. Research on intrusion detection of construction dangerous areas based on binocular stereo vision technology [D]. Beijing: Beijing Jianzhu University, 2023.
- [53] ZHANG R. Research on dynamic distance estimation of substation foreign body intrusion based on binocular stereo vision and deep learning [D]. Xi'an: Xi'an University of Technology, 2023.
- [54] ZHU Z, PARK M W, KOCH C, et al. Predicting movements of onsite workers and mobile equipment for enhancing construction site safety[J]. Automation in Construction, 2016, 68:95-101.
- [55] LEE Y J, PARK M W. 3D tracking of multiple onsite workers

- based on stereo vision[J]. *Automation in Construction*, 2019, 98:146-159.
- [56] YAN X, ZHANG H, LI H. Computer vision-based recognition of 3D relationship between construction entities for monitoring struck-by accidents[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2020, 35(9):1023-1038.
- [57] ZHOU Z, SI G, SUN H, et al. A robust clustering algorithm based on the identification of core points and KNN kernel density estimation[J]. *Expert Systems with Applications*, 2022, 195: 116573.
- [58] WANG K, LIU J L, HU Q Q, et al. Research on automatic detection of personal protective equipment in multi-operation scenarios [J]. *Journal of Safety and Environment*, 2023, 23(11): 3960-3967.
- [59] TIEN P W, WEI S, CALAUTIT J. A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand[J]. *Energies*, 2020, 14(1):156.
- [60] YU J Y. Research on moving target state detection in open scenes and construction areas [D]. Qingdao: Qingdao University, 2022.
- [61] WANG M, WONG P, LUO H, et al. Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques[C]// *Proceedings of the International Symposium on Automation and Robotics in Construction*. IAARC Publications, 2019:399-406.
- [62] SHI H. Research on detection algorithm of safety helmet wearing on construction site based on deep learning [D]. Wuhan: Wuhan University of Technology, 2019.
- [63] WU D. Research on safety condition monitoring technology of construction personnel based on computer vision [D]. Harbin: Harbin University of Technology, 2019.
- [64] LEE J, LEE S. Construction site safety management; a computer vision and deep learning approach [J]. *Sensors*, 2023, 23(2): 944.
- [65] GAO X R. Dynamic assessment of collision risk between construction site workers and mobile machinery [D]. Chongqing: Chongqing University, 2020.
- [66] CAO Z Y. Research on collision safety assessment of construction workers based on CV & FI [D]. Dalian: Dalian University of Technology, 2020.
- [67] ZHANG B, SONG Y B, WU B, et al. Two-stage Faster R-CNN for automatic analysis of pedestrian-vehicle collision risk at construction site [J]. *Manufacturing Automation*, 2022, 44(6): 24-27, 31.
- [68] TAO C, CAO J, WANG C, et al. Pseudo-mono for monocular 3d object detection in autonomous driving[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(8): 3962-3975.
- [69] LI P. Research and implementation of key technologies of construction site safety early warning based on target detection and depth estimation [D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [70] HU X X. Research and application of ranging method based on monocular depth estimation [D]. Beijing: North China Electric Power University, 2022.
- [71] YAN W J. Research on key technologies of 3D target recognition in construction site [D]. Chengdu: University of Electronic Science and Technology of China, 2022.
- [72] QIN S X. Research and implementation of key technologies for monocular image depth estimation [D]. Chengdu: University of Electronic Science and Technology of China, 2022.
- [73] SHEN J, YAN W, QIN S, et al. A self-supervised monocular depth estimation model with scale recovery and transfer learning for construction scene analysis[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2023, 38(9): 1142-1161.
- [74] LEE J H, HAN M K, KO D W, et al. From Big to Small; Multi-Scale Local Planar Guidance for Monocular Depth Estimation [J]. arXiv:1907.10326, 2019.
- [75] ANDRAGHETTI L, MYRIOKEFALITAKIS P, DOVESI P L, et al. Enhancing self-supervised monocular depth estimation with traditional visual odometry[C]// *2019 International Conference on 3D Vision(3DV)*. IEEE, 2019:424-433.
- [76] GUIZILINI V, AMBRUS R, PILLAI S, et al. 3d packing for self-supervised monocular depth estimation[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020:2485-2494.
- [77] LI C, YUE C, LIU Y, et al. An Improved MonoDepth2 Algorithm for Vehicle Monocular Depth Estimation[J]. *Optik*, 2024, 311:171936.
- [78] YUAN W, GU X, DAI Z, et al. New crfs: Neural window fully-connected crfs for monocular depth estimation[J]. arXiv:2203.01502, 2022.
- [79] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]// *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012:3354-3361.
- [80] ERICSSON L, GOUK H, HOSPEDALES T M. How well do self-supervised models transfer? [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021:5414-5423.
- [81] GUIZILINI V, HOU R, LI J, et al. Semantically-guided representation learning for self-supervised monocular depth [J]. arXiv:2002.12319, 2020.
- [82] BELLO J L G, KIM M. Self-supervised deep monocular depth estimation with ambiguity boosting[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(12): 9131-9149.
- [83] SAXENA A, SUN M, NG A Y. Make3d: Learning 3d scene structure from a single still image[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 31(5): 824-840.
- [84] SILBERMAN N, FERGUS R. Indoor scene segmentation using a structured light sensor[C]// *ICCV*. 2011:601-608.
- [85] CHEN W, FU Z, YANG D, et al. Single-image depth perception in the wild[J]. *Advances in Neural Information Processing Systems*, 2016, 29.
- [86] HUANG X, WANG P, CHENG X, et al. The apollo scene open dataset for autonomous driving and its application [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 42(10): 2702-2719.
- [87] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]// *Proceedings of*

- the IEEE Conference on Computer Vision and Pattern Recognition. 2016:3213-3223.
- [88] GODARD C, MAC AODHA O, FIRMAN M, et al. Digging into self-supervised monocular depth estimation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:3828-3838.
- [89] CASSER V, PIRK S, MAHJOURIAN R, et al. Unsupervised monocular depth and ego-motion learning with structure and semantics[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.
- [90] GUIZILINI V, AMBRUS R, PILLAI S, et al. 3d packing for self-supervised monocular depth estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:2485-2494.
- [91] PENG R, WANG R, LAI Y, et al. Excavating the potential capacity of self-supervised monocular depth estimation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:15560-15569.
- [92] BIAN J, LI Z, WANG N, et al. Unsupervised scale-consistent depth and ego-motion learning from monocular video[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [93] JIANG Y, GONG X, LIU D, et al. Enlightengan: Deep light enhancement without paired supervision[J]. IEEE Transactions on Image Processing, 2021, 30: 2340-2349.
- [94] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:7132-7141.
- [95] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV). 2018:3-19.
- [96] YANG Z, WANG P, XU W, et al. Unsupervised learning of geometry from videos with edge-aware depth-normal consistency [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018.
- [97] WU J, ZHOU W, LUO T, et al. Multiscale multilevel context and multimodal fusion for RGB-D salient object detection[J]. Signal Processing, 2021, 178: 107766.
- [98] JIAO J, CAO Y, SONG Y, et al. Look deeper into depth: Monocular depth estimation with semantic booster and attention-driven loss[C]//Proceedings of the European Conference on Computer Vision(ECCV). 2018:53-69.
- [99] RANJAN A, JAMPANI V, BALLE S L, et al. Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:12240-12249.
- [100] RANFTL R, BOCHKOVSKIY A, KOLTUN V. Vision transformers for dense prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 12179-12188.
- [101] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks [C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 239-248.
- [102] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
- [103] BHAT S F, ALHASHIM I, WONKA P. Adabins: Depth estimation using adaptive bins[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 4009-4018.
- [104] XIAN K, ZHANG J, WANG O, et al. Structure-guided ranking loss for single image depth prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:611-620.
- [105] WANG L, ZHANG J, WANG Y, et al. Cliffnet for monocular depth estimation with hierarchical embedding loss[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 316-331.
- [106] SAHA S, OBUKHOV A, PAUDEL D P, et al. Learning to relate depth and semantics for unsupervised domain adaptation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8197-8207.
- [107] HUYNH L, NGUYEN-HA P, MATAS J, et al. Guiding monocular depth estimation using depth-attention volume[C]//Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXVI 16. Springer International Publishing, 2020: 581-597.
- [108] LEE B U, LEE K, KWEON I S. Depth completion using plane-residual representation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13916-13925.
- [109] GUIZILINI V, AMBRUS R, BURGARD W, et al. Sparse auxiliary networks for unified monocular depth prediction and completion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11078-11088.
- [110] IMRAN S, LIU X, MORRIS D. Depth completion with twin surface extrapolation at occlusion boundaries[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2583-2592.



HAO Xu, born in 2000, postgraduate. Her main research interest is deep learning.



WU Wenhong, born in 1980, master. Her main research interests include deep learning and image processing.