基于Star-RTMPose的双目视觉定位与测量

张梦权,许四祥*,杨 玉,吴端正
(安徽工业大学机械工程学院,安徽马鞍山 243032)

摘 要: 针对传统双目视觉特征点检测算法效率低、匹配精度不足、对光照变化敏感以及参数调优复杂,导致双目视觉定位与测量精度受限的问题,本文提出一种基于Star-RTMPose(Star-enhanced Real-Time Multi-person Pose estimation)的双目视觉定位与测量方法. 本文以钢铁冶金行业的连铸坯为研究对象,聚焦其火焰切割后毛刺切除所需的精准定位与尺寸测量需求,给出了对应的技术实现路径. 首先,通过标定后的双目相机采集连铸坯图像,并采用LabelMe工具完成目标区域与关键点标注,将标注结果统一转换为MSCOCO(MicroSoft Common Objects in COntext)格式以适配模型训练. 随后,采用"目标检测-关键点提取"的双阶段框架实现精准检测,即先基于RTMDet(Real-Time Models for object Detection)算法快速定位连铸坯的主体区域,进而采用基于RTMPose(Real-Time Multi-person Pose estimation)的改进模型Star-RTMPose提取关键点坐标. 改进包括:在RTMPose主干引入StarTriBlock(Star Triple Block)模块,通过多支路动态融合机制增强网络对目标高层语义特征的表征能力,充分利用该阶段最大感受野与全局空间关联信息;使用基于深度可分离卷积的MaxDSC2(Maximum Depthwise Separable Convolution 2)模块替代网络头部的7×7大核卷积,并将该模块的中间通道数设定为输入通道数的0.45倍,在提升语义信息敏感度的同时降低参数量;用无参SimAM(Simple parameter-free Attention Module)注意力模块替代传统通道注意力模块,通过能量函数闭式解生成通道-空间三维联合权重,强化网络对空间特征的捕获性能,避免参数冗余. 最终,结合双目相机标定参数与三角测量原理,完成关键点三维重建与连铸坯尺寸测量. 实验结果表明:在关键点检测任务中,改进后的Star-RTMPose模型对单张图像的推理时间仅为9.86 ms,相较于基准模型RTMPose-T,其AP(Average Precision)提升1.09个百分点,PCK(Percentage of Correct Keypoints)提升0.40个百分点,NME(Normalized Mean Error)降低42.86%;改进后的模型在参数量更为精简的前提下,综合性能显著优于HRNet-W32、SwinTransformer-T等主流模型;在三维测量精度方面,本文方法对1型连铸坯长边尺寸的测量相对误差相较于传统ORB(Oriented FAST and Rotated BRIEF)算法以及改进后的FAST(Features from Accelerated Segment Test)算法分别降低了1.715个百分点和0.365个百分点. 本文方法有效解决了传统算法鲁棒性欠佳的问题,实现了检测精度与测量精度的双重提升,切实满足工业场景对高精度检测的需求.

关键词: 双目视觉;RTMPose;注意力模块;三维重建;尺寸测量;关键点检测
基金项目: 国家自然科学基金(No.51374007);安徽高校自然科学研究重点项目(No.KJ2020A0259)
中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112(2025)12-4317-13
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250422

Binocular Vision Localization and Measurement Based on Star-RTMPose

ZHANG Meng-quan, XU Si-xiang*, YANG Yu, WU Duan-zheng
(*College of Mechanical Engineering, Anhui University of Technology, Ma'anshan, Anhui* 243032, *China*)

Abstract: A binocular vision localization and measurement method based on star-enhanced real-time multi-person pose estimation (Star-RTMPose) is proposed to address the problems of low efficiency, insufficient matching accuracy, sensitivity to illumination changes, and complex parameter tuning of traditional binocular vision feature point detection algorithms, which limit the accuracy of binocular vision localization and measurement. Taking continuous casting billets in the iron and steel metallurgy industry as the research object, this method focuses on the precise positioning and dimension measurement requirements for burr removal after flame cutting, and proposes a corresponding technical implementation approach. Firstly, images of continuous casting billets are collected using calibrated binocular cameras. The LabelMe tool is

收稿日期:2025-05-26;录用日期:2025-12-01;责任编辑:覃怀银
*通讯作者:许四祥

then used to annotate target regions and keypoints, which are uniformly converted to the microsoft common objects in context (MSCOCO) format to adapt to model training. Subsequently, a two-stage framework of "target detection-keypoint extraction" is adopted to achieve precise detection: the real-time models for object detection (RTMDet) algorithm is first used to quickly locate the main area of the continuous casting billet, and then the improved real-time multi-person pose estimation (RTMPose) model, Star-RTMPose, is used to extract keypoint coordinates. The improvements include: introducing the star triple block (StarTriBlock) module into the RTMPose backbone to enhance the network's ability to characterize high-level semantic features of the target through a multi-branch dynamic fusion mechanism, making full use of the maximum receptive field and global spatial correlation information of this stage; replacing the 7×7 large kernel convolution at the network head with the maximum depthwise separable convolution 2 (MaxDSC2) module based on depth-separable convolution, setting the intermediate channel number of this module to 0.45 times the input channel number to improve the sensitivity to semantic information while reducing the number of parameters; substituting the traditional channel attention module with the parameter-free simple parameter-free attention module (SimAM) attention module, which generates channel-spatial three-dimensional joint weights through the closed-form solution of the energy function, strengthens the network's ability to capture spatial features, and avoids parameter redundancy. Finally, by combining the calibration parameters of the binocular camera with the triangulation principle, the three-dimensional reconstruction of keypoints and the dimensional measurement of continuous casting billets are completed. The experimental results show that: in the keypoint detection task, the inference time of the improved Star-RTMPose model for a single image is only 9.86 ms; compared with the baseline model RTMPose-T, its average precision (AP) is improved by 1.09 percentage points, percentage of correct keypoints (PCK) by 0.40 percentage points, and normalized mean error (NME) is reduced by 42.86%; on the premise of more streamlined parameters, the comprehensive performance of the improved model is significantly superior to that of mainstream models such as HRNet-W32 and SwinTransformer-T. In terms of three-dimensional measurement accuracy, the relative error of the proposed method for measuring the long side dimension of Type 1 continuous casting billet is reduced by 1.715 and 0.365 percentage points compared to the traditional oriented fast and rotated brief (ORB) algorithm and the improved features from accelerated segment test (FAST) algorithm, respectively. This method effectively addresses the issue of poor robustness in traditional algorithms, achieving dual improvements in detection accuracy and measurement accuracy, and thereby meeting the demand for high-precision detection in industrial scenarios.

Key words: binocular vision; RTMPose; attention module; three-dimensional reconstruction; dimensional measurement; keypoint detection

Foundation Item(s): National Natural Science Foundation of China (No. 51374007); Key Program of Natural Science Research in Anhui Provincial Universities (No.KJ2020A0259)

1 引言

现代钢铁冶金行业为方便连铸坯的运输、储存及后续轧制加工,多采用火焰切割机对大断面连铸坯进行定尺切割. 然而,在火焰切割过程中,液态熔渣有时无法完全从切缝中排出,导致其在切口的下边缘冷却并凝结成不规则的硬质毛刺. 这些毛刺的形成不仅严重损害了连铸坯的表面质量,还可能对后续加工流程造成显著的负面影响. 因此,在火焰切割连铸坯后,去除这些毛刺成为了一项至关重要的工序. 许四祥等[1]提出一种机器人携带等离子枪去除毛刺的方案,然而该方案缺乏对连铸坯的精确定位能力,且无法动态调整切割轨迹,致使等离子枪切割轨迹与实际切口轮廓出现失配,进而造成漏切与错切现象. 为此,采用双目视觉方法对连铸坯进行定位与测量,以辅助机器人精准切除毛刺.

在传统双目视觉特征点检测方法中,特征匹配是至关重要的一个环节. 常用的特征点提取和描述算法有 SIFT(Scale-Invariant Feature Transform)[2]、SURF(Speeded Up Robust Features)[3]和 ORB(Oriented FAST and Rotated BRIEF)[4]等. 宋超群等[5]针对传统 FAST(Features from Accelerated Segment Test)角点检测算法存在的阈值需人工设定、角点冗余等问题,提出基于自适应阈值和邻域灰度均值的改进方法,提高了匹配准确率和测量精度. 宋祥等[6]通过图像差分增强边缘对比度以提升特征点提取精度,采用环形邻域策略优化(Modified-SURF, M-SURF)描述符,并基于垂直方向 Haar 小波响应符号特征进行多区间特征划分生成高维描述符,有效提升了特征点匹配精度. Xu 等[7]通过局部区域采样与子区域划分剔除低偏移量子区域,并基于灰度差异阈值生成描述符,有效抑制了光滑区域噪声位模式的干扰,降低了特征点的误匹配率. 虽然传统特征点提取与特征匹配算法已得到很大改善,但是依然存在对光照变化敏感、参数调优复杂和实时性差等不足.

深度学习对计算机视觉领域产生了革命性的影

响,极大地推动了图像处理和分析技术的发展. 利用深度学习进行关键点检测是当前研究的热点方向,其核心是通过神经网络自动学习图像中具有判别性的特征表示,从而准确定位目标对象的关键点. 关键点检测早期主要应用在人体姿态估计中. Cao等[8]提出了应用于人体姿态估计的OpenPose算法,通过热力图和部件亲和场(Part Affinity Fields,PAFs)的结合,OpenPose能够精确地定位关键点并关联肢体. Sun等[9]提出了能在整个网络过程中维持高分辨率特征的算法HRNet(High-Resolution Network). 江佳鸿等[10]以HRNet为基础,设计了一种多分辨率网络,具有广泛的适用范围. Yuan等[11]提出高分辨率HRFormer(High-Resolution transFormer)算法,延续HRNet的多分辨率设计,通过非重叠窗口自注意力降低计算复杂度,并在前馈网络中嵌入3×3深度可分离卷积促进跨窗口信息交互,显著提升了人体姿态估计的效率与精度. 然而,这些神经网络结构复杂、参数量大,训练和推理需占用大量计算资源. Jiang等[12]探讨了姿势估计中的关键因素,提出高性能实时多人姿势估计框架RTMPose(Real-Time Multi-person Pose estimation). 该框架采用轻量化骨干网络CSPNeXt(Cross Stage Partial NeXt),结合SimCC(Simple Coordinate Classification)[13]关键点定位方法,在保持网络低复杂度的同时实现较高精度. 然而,RTMPose在光照条件变化、拍摄距离变化等真实场景下性能会出现下降.

基于此,提出一种基于改进RTMPose的双目视觉定位与测量方法. 基于RTMPose的原始结构,对其整体架构进行优化:首先,在主干网络中嵌入StarTriBlock(Star Triple Block)模块,通过增强网络对高级图像特征的提取能力,有效提升模型的特征学习性能. 其次,采用SimAM(Simple parameter-free Attention Module)模块替换CSPLayer(Cross Stage Partial Layer)中的通道注意力模块,弥补了原网络空间学习能力的不足. 最后,以MaxDSC2(Maximum Depthwise Separable Convolution 2)模块替换头部网络的7×7大核卷积,在改善模型语义信息敏感度的同时,实现网络精度提升与参数量优化的双重目标. 通过对RTMPose网络的改进,模型对连铸坯关键点的检测精度显著提高,最终根据三角测量原理完成关键点三维重建,实现连铸坯的精准定位与尺寸测量.

本文所提出的Star-RTMPose(Star-enhanced Real-Time Multi-person Pose estimation)模型及其与RTMDet(Real-Time Models for object Detection)构成的双阶段检测框架,其核心思想具备良好的通用性与迁移潜力. 尽管本研究以连铸坯为具体研究对象,但该方法并未依赖于连铸坯特有的先验知识. 其技术路径是通过目标检测网络定位物体主体,再使用关键点检测网络精确定位其预定义关键点,最后结合双目视觉进行三维重建,这是计算机视觉中一种通用的物体定位与测量范式. 该框架可迁移至其他需要对规则或不规则工业部件进行高精度定位与尺寸测量的场景,例如钣金件的孔位定位、焊接机器人的焊缝跟踪与测量等. 模型中的核心改进模块,如StarTriBlock、SimAM、MaxDSC2,旨在增强网络的特征提取能力、空间注意力机制与计算效率,这些改进对于光照变化、尺度变化等复杂工业环境下的各类视觉任务均具有积极的促进作用.

2　双目视觉定位与测量方法

基于改进RTMPose的双目视觉定位与测量方法主要由数据集制作、网络训练、定位与测量三部分构成,具体流程如图1所示.

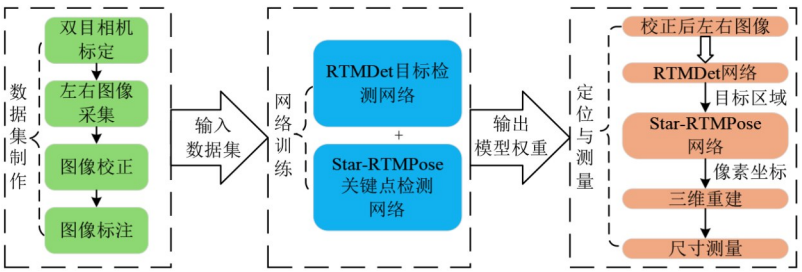


图1　连铸坯的定位与尺寸测量流程

对于网络训练部分,使用相同的数据集对RTMDet目标检测网络和RTMPose关键点检测网络进行训练. 其中定位与测量部分的核心流程是一个两阶段网络结构. 首先,将经过立体校正后的左右图像分别输入RTMDet目标检测网络,该网络负责快速、准确地定位图像中连铸坯的主体区域,并输出其边界框. 接着,将RTMDet检测到的连铸坯边界框区域从原图中裁剪出来,作为Star-RTMPose关键点检测网络的输入;Star-RTMPose网络在该目标区域内进行精细的关键点检测,精确输出左右图像中连铸坯关键点的二维像素坐标. 最后,结合双目相机的标定参数,根据三角测量原理完成关键点的三维重建,最终实现连铸坯的精准定位与尺寸测量.

2.1　数据集制作

连铸坯经火焰切割后,切缝处未排出的液态熔渣

会在切口下边缘凝结成不规则毛刺，如图2所示．通过获得连铸坯A、B、C、D四个关键点的空间坐标，即可实现连铸坯的空间定位与尺寸测量．为了提高模型的鲁棒性和泛化能力，实验使用标定后的双目深度相机在多视角、多旋转角度、多距离及不同光照等条件下对4种连铸坯进行图像采集．根据双目相机的标定结果对采集到的图像进行立体校正处理，共获得1 255幅分辨率为1 600×1 200像素的数据集图像．随后使用LabelMe标注工具逐幅标注图像，在框选连铸坯主体区域后按顺时针顺序完成关键点标注，部分标注图像如图3所示．标注完成后，将生成的JSON（JavaScript Object Notation）文件统一转换为MSCOCO（MicroSoft Common Objects in COntext）格式．最终按8∶2的比例随机划分数据集，得到包含1 004幅图像的训练集和251幅图像的测试集．

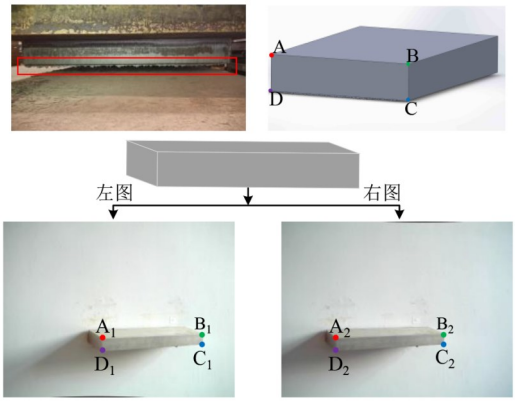


图2 连铸坯模型示意图

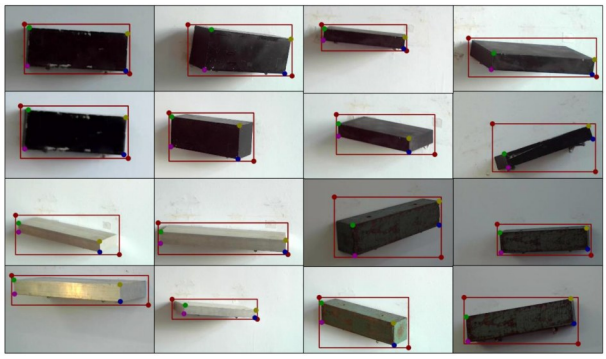


图3 部分标注图像

2.2 网络模型

2.2.1 RTMDet网络模型

RTMDet网络[14]作为YOLO（You Only Look Once）系列衍生的新一代实时目标检测模型，架构如图4所示．其采用改进的CSPDarkNet（Cross Stage Partial DarkNet）为主干，通过大核深度可分离卷积扩大感受野并优化参数量，实现高效特征提取；颈部融合主干多尺度特征，增强跨层级语义与细节的互补；检测头通过参数共享、动态标签分配设计降低复杂度并提升精度．该模型在工业场景中平衡实时性与精度，为后续关键点检测提供高质量目标区域，有效缩小关键点检测网络的搜索范围，提升整体系统的效率和鲁棒性．

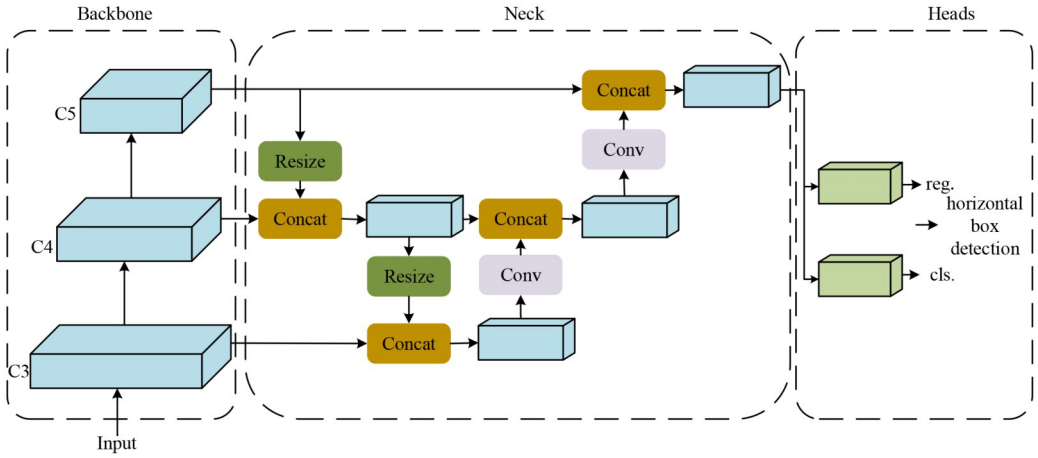


图4 RTMDet连铸坯目标检测网络架构

2.2.2 RTMPose网络模型

RTMPose-T是一个简单高效的实时人体姿态估计框架，如图5所示．相较于传统2D关键点检测网络依赖复杂上采样层生成高分辨率热图并通过后处理减小量化误差的方案，RTMPose通过CSPNeXt主干网络对特征图进行连续的下采样，并采用SimCC方法进行关键点定位．该方法通过坐标解耦设计和密集bin划分策略，将关键点横、纵坐标解耦为独立一维分类任务，无需通过网络上采样生成高分辨率热图即可基于下采样特征图完成定位，在降低量化误差的基础上实现亚像素级定位精度，既降低计算成本，又保持高精度检测性能．

CSPNeXt网络由卷积模块（Convolution Module，ConvModule）、跨阶段部分层（Cross Stage Partial Layer，

CSPLayer)和空间金字塔快速池化瓶颈模块(Spatial Pyramid Pooling Fast Bottleneck,SPPFBottleneck)三大模块组成. 其中,ConvModule作为下采样和通道调整的基本模块,包含Conv2d、BN(Batch Normalization)和SiLU(Sigmoid Linear Unit)激活函数;CSPLayer通过级联结构增强网络的特征提取能力,平衡计算效率与表征丰富度;SPPFBottleneck对不同尺度特征图进行池化,有效捕捉多尺度语义信息以提升模型性能.

RTMPose的头部网络由7×7卷积层、全连接层(Fully Connected layer,FC)、门控注意力单元(Gated Attention Unit,GAU)和坐标分类器构成. 其中,7×7卷积层将384通道的特征图压缩至k个通道,生成k个关键点的初始特征表示;每个关键点的特征表示经展平处理为一维向量后,通过全连接层扩展至256维;随后,门控注意力单元通过自注意力机制融合全局与局部空间信息;最后,坐标分类器基于k个关键点特征向量,分别在水平和垂直方向上执行分类,从而实现精准的关键点定位.

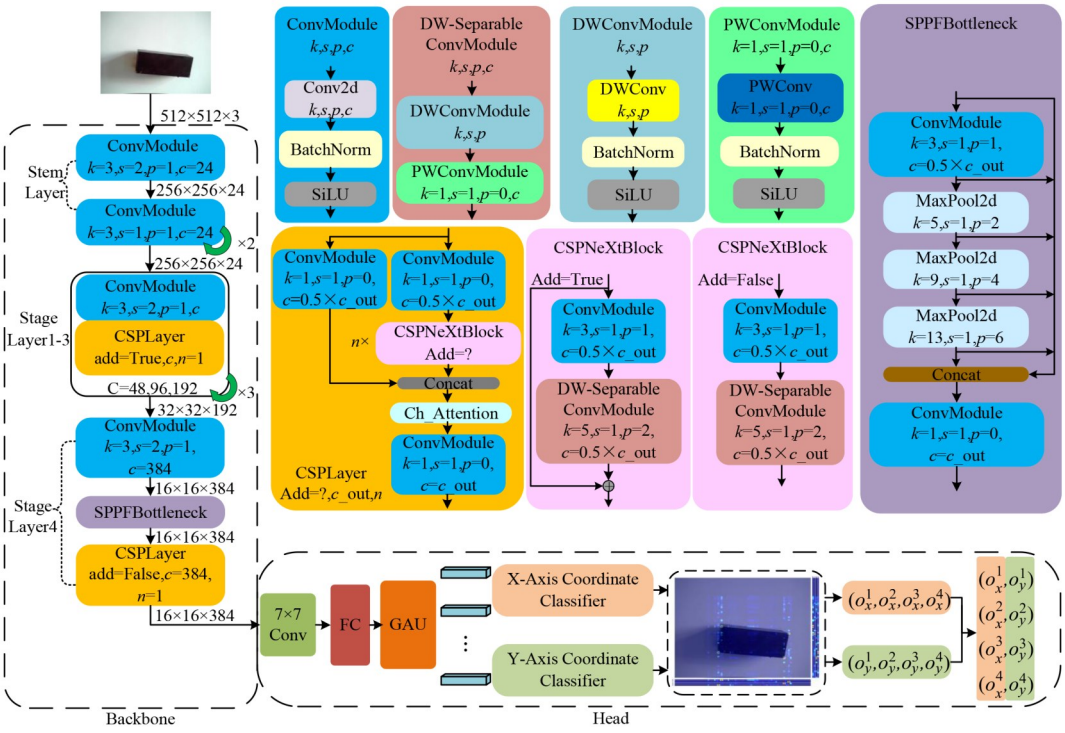


图5　RTMPose-T连铸坯关键点检测网络架构

2.2.3　Star-RTMPose网络模型

Star-RTMPose网络模型结构如图6所示. 首先,为增强模型对高级图像特征的学习能力,在RTMPose-T的主干部分引入StarTriBlock模块. 然后,以SimAM注意力模块替代CSPLayer中原有的通道注意力模块,在不增加参数量的前提下提升模型的空间学习能力. 最后,将头部的7×7卷积层替换为MaxDSC2模块,在降低复杂度的同时,进一步优化模型的推理精度.

2.2.4　StarTriBlock模块

Ma等[15]提出的StarBlock模块结构如图7(a)所示. 输入特征图先经深度卷积提取空间特征,随后在两条支路上通过1×1卷积扩展通道数,一条支路通过ReLU6(Rectified Linear Unit 6)激活函数进行非线性变换,另一条支路保留原始输出,两者通过星形操作动态融合,实现特征的动态调制,增强特征的非线性表达能力. 接着,通过1×1卷积将通道数还原至原始维度,再经第二个深度卷积提取高阶空间特征,增强局部特征的表达能力. 最后,通过残差连接将输入特征与处理后的特征相加,以此增强梯度流动和特征复用. 该模块通过轻量化的深度卷积、通道扩展-压缩策略及动态特征调制,有效提升了模型的特征表达能力.

在RTMPose主干网络的第四阶段,特征具备最高语义抽象层级与最大的感受野,能够捕捉全局空间关联信息,这对关键点的精确定位至关重要. 为高效利用此类高层级特征,基于StarBlock模块设计了如图7(b)所示的StarTriBlock模块. 该模块在继承StarBlock动态调制能力的基础上,通过多支路架构提升模型在不同尺度上对高级特征信息的利用效率. 将其引入RTMPose主干网络的第四阶段,以强化模型对高层语义特征的学习能力.

2.2.5　SimAM注意力机制

SimAM[16]是一种基于空间抑制现象的轻量级三维

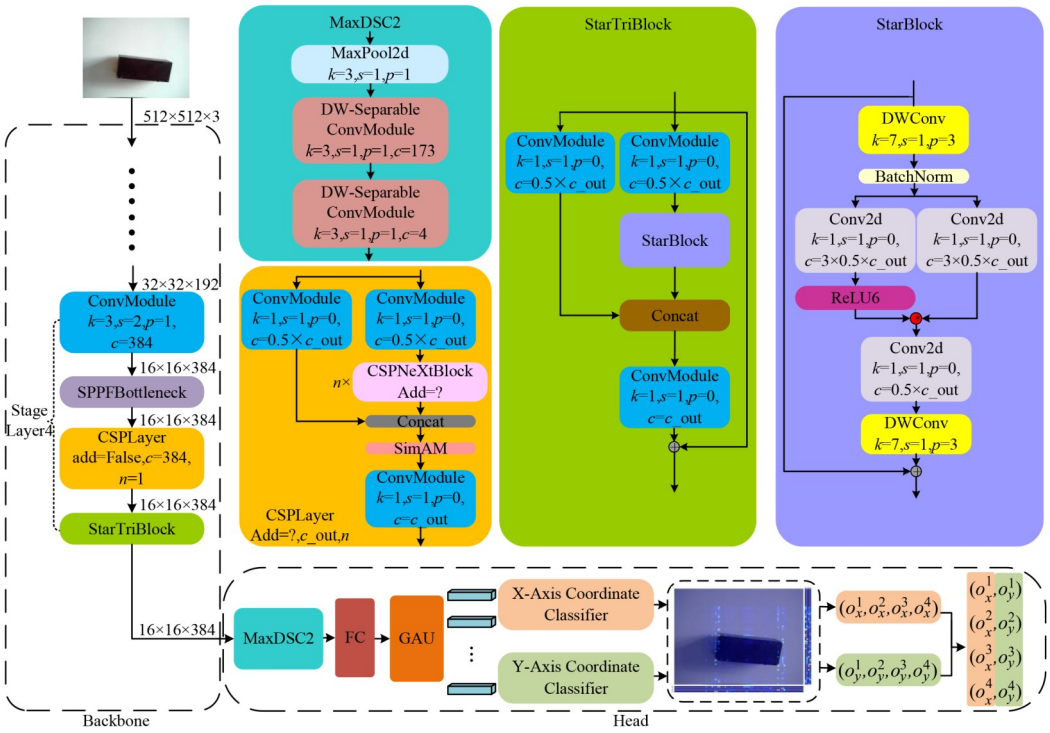
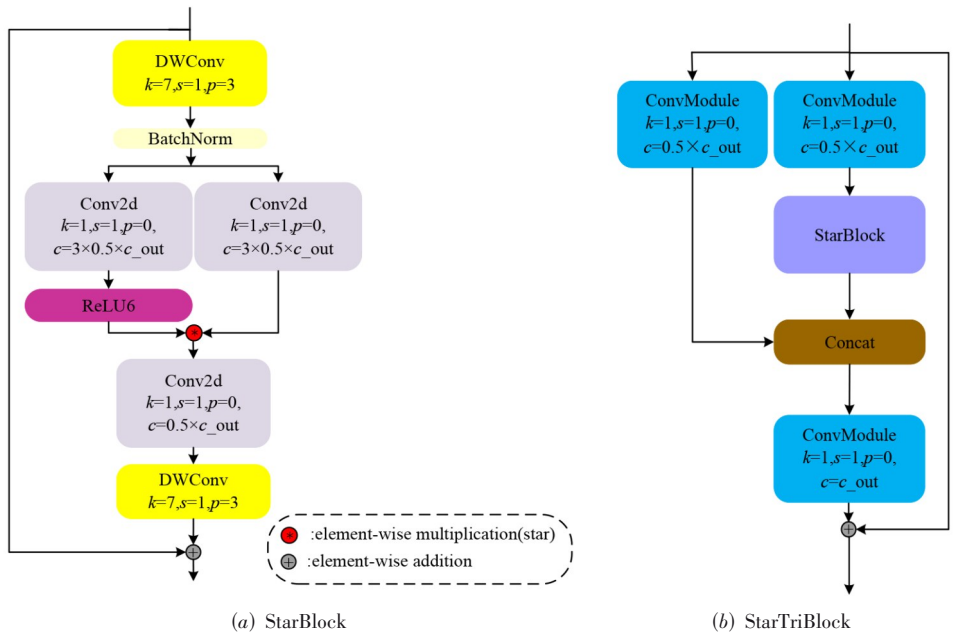


图 6 Star-RTMPose 连铸坯关键点检测网络架构



(a) StarBlock (b) StarTriBlock

图 7 StarBlock 和 StarTriBlock 结构

注意力机制，通过能量函数闭式解直接生成通道与空间维度的联合权重，在无需引入额外参数的条件下实现高效的特征选择. 该机制以神经元最小能量 e_t^* 来衡量目标神经元 t 的重要性，其中 e_t^* 定义如下：

$$e_t^* = \frac{4\left(\hat{\sigma}^2+\lambda\right)}{\left(t-\hat{\mu}\right)^2+2\hat{\sigma}^2+2\lambda} \tag{1}$$

其中，λ 是正则化参数，防止分母为零；t 是输入特征图 $\boldsymbol{X}\in R^{C\times H\times W}$ 在单个通道上的目标神经元；$\hat{\mu}$ 和 $\hat{\sigma}^2$ 是单通道上所有神经元的平均值和方差. $\hat{\mu}$ 和 $\hat{\sigma}^2$ 的表达式分别为

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{2}$$

$$\hat{\sigma}^2=\frac{1}{M}\sum_{i=1}^{M}\left(x_i-\hat{\mu}\right)^2 \tag{3}$$

其中，M表示目标神经元所在维度上神经元的个数，$M=H\times W$；x_i表示目标神经元所在维度上所有的神经元.

根据生物视觉的空间抑制理论，显著神经元通过抑制周围神经元的活性，增强自身响应，从而突出关键特征. 基于该机制，目标神经元t与同通道其他神经元的差异越显著，其重要性越高，且由式(1)可知，对应的最小能量值e_t^*越低. 因此，神经元的重要性可通过最小能量值的倒数$1/e_t^*$直接量化. 输入特征信息经SimAM注意力模块处理后，得到输出特征$\tilde{\boldsymbol{X}}$，即

$$\tilde{\boldsymbol{X}}=\mathrm{sigmoid}\left(\frac{1}{\boldsymbol{E}}\right)\odot\boldsymbol{X} \tag{4}$$

其中，$\boldsymbol{E}$表示所有神经元的三维最小能量值矩阵；⊙表示逐元素乘法. 通过sigmoid函数限制权重范围，避免极端值.

如图8所示，SimAM注意力机制直接生成覆盖通道、高度和宽度的3D注意力权重，突破了传统方法(如SE(Squeeze-and-Excitation module)[17]、CBAM(Convolutional Block Attention Module)[18]和ECA(Efficient Channel Attention module)[19])局限于单一维度或分步处理的限制，同时保持网络结构的轻量化与计算高效性. SE和ECA模块局限于通道维度，CBAM的通道-空间分步建模易导致特征关联丢失，如某通道的纹理特征与空间角点位置脱节. SimAM生成的3D注意力权重，可同时优化通道重要性与空间位置重要性，在连铸坯角点区域的边缘通道权重升高的同时，该通道内角点位置的空间权重也同步升高，实现了通道-空间特征的协同增强. 并且SE和CBAM均存在参数膨胀问题，使得网络模型的参数量和计算量有所增加，而SimAM仅仅通过能量函数闭式解直接生成权重，无需任何额外参数，有效避免了模型复杂度的上升，保障实时推理性能.

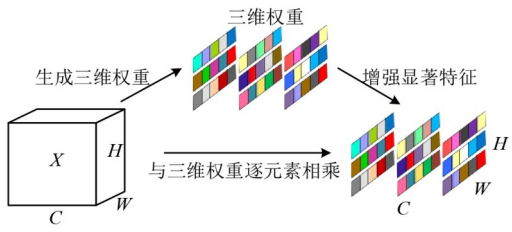


图8　SimAM工作原理图

在CSPNeXt架构中，CSPLayer作为核心组件通过跨阶段特征融合与通道注意力机制提升模型表达能力，但其传统通道注意力模块仅聚焦通道维度，存在特征建模单一化与参数冗余的问题. 替换为SimAM注意力模块后，该模块通过通道-空间联合注意力机制与无额外可学习参数的设计，在维持CSPLayer轻量化的同时，显著增强了多维度特征选择能力与计算效率. SimAM基于能量函数动态强化高区分度神经元(如关键点区域)的响应权重，并抑制同通道内冗余区域的激活，这种自适应特征聚焦机制与细粒度定位任务的需求高度契合，有效提升了模型对局部关键特征的敏感度.

2.2.6　MaxDSC2模块

深度可分离卷积[20]通过将传统标准卷积分解为深度卷积与逐点卷积两个独立操作，在保持模型性能的同时显著降低计算成本，其原理架构如图9所示.

假设输入特征图的尺寸为$H\times W\times C_{\mathrm{in}}$，深度卷积的核大小为$K\times K\times 1$且核数量为$C_{\mathrm{in}}$；逐点卷积的核大小为$1\times 1\times C_{\mathrm{in}}$且核数量为$C_{\mathrm{out}}$，输出特征图的尺寸为$H\times W\times C_{\mathrm{out}}$. 因此深度可分离卷积的参数量与标准卷积的参数量之比λ_p和计算量之比λ_c分别为

$$\begin{aligned}\lambda_p&=\frac{K\times K\times 1\times C_{\mathrm{in}}+1\times 1\times C_{\mathrm{in}}\times C_{\mathrm{out}}}{K\times K\times C_{\mathrm{in}}\times C_{\mathrm{out}}}\\&=\frac{1}{C_{\mathrm{out}}}+\frac{1}{K^2}\end{aligned} \tag{5}$$

$$\begin{aligned}\lambda_c&=\frac{H\times W\times\left(K\times K\times 1\times C_{\mathrm{in}}+1\times 1\times C_{\mathrm{in}}\times C_{\mathrm{out}}\right)}{H\times W\times K\times K\times C_{\mathrm{in}}\times C_{\mathrm{out}}}\\&=\frac{1}{C_{\mathrm{out}}}+\frac{1}{K^2}\end{aligned} \tag{6}$$

由式(5)和式(6)可知，当C_{out}较大时，λ_p和λ_c皆约等于$1/K^2$.

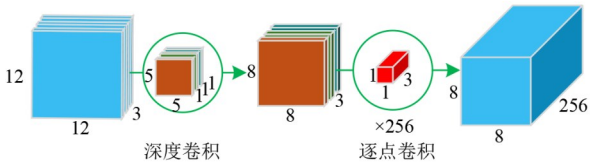


图9　深度可分离卷积原理图

RTMPose的头部网络采用7×7卷积层实现特征图通道降维. 然而，传统大核卷积因参数量随核尺寸平方增长，导致计算成本较高；同时，其大感受野在捕获全局上下文的同时，可能削弱对局部细节的建模能力，不利于细粒度语义信息的提取. 为此，本文提出了基于深度可分离卷积的MaxDSC2模块，用以替代原有的7×7卷积层.

MaxDSC2模块结构如图10所示，输入特征图首先经过最大池化层，在不降低空间分辨率的前提下减少冗余信息；随后经第一个深度可分离卷积层将通道数压缩至173. 再通过第二个深度可分离卷积层进一步将通道数降至目标关键点数量，实现对特征的层次化降维与关键信息聚焦.

MaxDSC2的输入通道数为384，中间通道数需避免

维度骤降导致的特征丢失与维度过高导致的计算冗余，中间通道数通常设计为输入通道数的0.5倍左右[21]. 经对不同中间通道数的模型性能进行比对，最终将中间通道数设定为输入通道数的0.45倍.

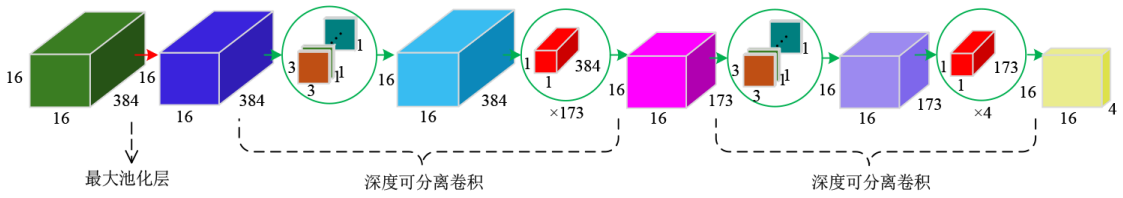


图10 MaxDSC2模块结构

3 实验结果与分析

3.1 实验环境与模型训练

实验环境配置为64位Windows10操作系统，搭载6核Xeon Gold 6142处理器，配备27.1 GB内存与RTX3080显卡，基于PyTorch1.10.1深度学习框架对目标检测模型和关键点检测模型进行训练. 为降低训练耗时与资源损耗，两种模型均在加载预训练权重的基础上开展训练，借助迁移学习加速模型收敛并提升性能.

训练目标检测模型时，输入图像统一调整为640×640分辨率. 模型训练的批量大小设置为20，共开展300轮训练. 训练过程中采用线上两阶段增强策略，即前280轮运用Mosaic拼接、随机水平翻转和随机裁剪等复合增强策略，最后20轮简化为基本增强策略以确保收敛稳定性. 优化器选用AdamW，基础学习率设定为0.004，权重衰减系数为0.05. 同时采用Linear-Cosine学习率策略，即前10轮将学习率线性预热至0.004，第150轮后通过余弦退火使其衰减至0.000 2.

训练关键点检测模型时，输入图像统一调整为512×512分辨率. 模型训练的批量大小设置为32，共开展130轮训练. 训练过程采用线上数据增强方法，通过随机水平翻转、随机旋转和随机缩放等几何变换增强技术，确保模型对不同拍摄视角和距离的鲁棒性. 优化器选用AdamW，基础学习率为0.002，权重衰减系数为0.05. 同样采用Linear-Cosine学习率策略，前10轮将学习率线性预热至0.002，第65轮后通过余弦退火使其衰减至0.000 1.

3.2 性能评价指标

对于网络输出关键点坐标，采用以下评价指标来评测网络性能，包括网络参数量(Parameters)、浮点运算数(Floating Point Operations，FLOPs)、基于目标关键点相似度(Object Keypoint Similarity，OKS)的平均精度(Average Precision，AP)、平均召回率(Average Recall，AR)、正确关键点百分比(Percentage of Correct Keypoints，PCK)、归一化平均误差(Normalized Mean Error，NME)以及单张图像推理时间(T). 其中，AP定义为OKS阈值从0.50到0.95(步长0.05)区间内平均精度的均值；AR则为OKS阈值从0.50到0.95(步长0.05)区间内平均召回率的均值.

OKS计算公式为

$$\mathrm{OKS}=\frac{\sum_i \exp\left\{-d_i^2/2S^2\sigma_i^2\right\}\delta\left(v_i>0\right)}{\sum_i \delta\left(v_i>0\right)} \tag{7}$$

其中，d_i表示第i个关键点的预测位置与真实位置的欧氏距离；S表示目标的尺度因子；σ_i表示第i个关键点的归一化常数；v_i为第i个关键点的可见性表示，$v_i=0$表示未标注，$v_i=1$表示不可见但标注，$v_i=2$表示标注且可见；$\delta(v_i>0)$为指示函数，当$v_i>0$时为1，否则为0.

PCK的核心是判断归一化后的关键点预测误差是否小于设定阈值，其计算公式为

$$\mathrm{PCK}=\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\left(\frac{\left\| p_i^{\mathrm{pred}}-p_i^{\mathrm{gt}}\right\|_2}{L}<\tau\right) \tag{8}$$

其中，p_i^{pred}表示第i个关键点的预测坐标；p_i^{gt}表示第i个关键点的真实坐标；L为PCK归一化因子，取值为长方体连铸坯长边的欧氏距离，用以消除目标尺寸不同所带来的影响；τ为阈值，取值为0.05；N为关键点总数；$\mathbb{I}(\cdot)$为指示函数，括号内条件成立时取1，否则取0.

NME表示归一化后的关键点预测误差的平均值，其计算公式为

$$\mathrm{NME}=\frac{1}{N}\sum_{i=1}^{N}\frac{\left\| p_i^{\mathrm{pred}}-p_i^{\mathrm{gt}}\right\|_2}{L} \tag{9}$$

其中，p_i^{pred}表示第i个关键点的预测坐标；p_i^{gt}表示第i个关键点的真实坐标；L为NME归一化因子，取值为长方体连铸坯长边的欧氏距离，用以消除目标尺寸不同所带来的影响；N为关键点总数.

3.3 对比实验

3.3.1 CSPLayer不同注意力机制的性能对比

为验证SimAM相较于其他主流注意力机制的优势，在基础模型Star-RTMPose上，仅替换CSPLay中的注意力模块(SE/CBAM/ECA/SimAM)，保持其他参数一致，实验结果如表1所示.

表1　CSPLayer不同注意力机制性能对比

注意力机制	Input Size	Params/M	FLOPs/G	AP/%	AR/%	PCK/%	NME/%	*T*/ms
SE[17]	512×512	3.92	**1.94**	98.81	99.25	99.75	0.35	10.55
CBAM[18]	512×512	3.92	1.95	99.20	99.35	99.63	0.30	11.23
ECA[19]	512×512	**3.87**	**1.94**	98.80	99.10	99.75	0.52	10.02
SimAM	512×512	**3.87**	**1.94**	**99.60**	**99.75**	**99.90**	**0.28**	**9.86**

注：表格中加粗显示的内容为各指标对应的最优表现.

从表1的实验结果可清晰看出，SimAM在不引入额外参数的前提下，各项性能指标均实现提升. 在核心指标上，其AP较SE、CBAM、ECA分别提升0.79、0.40、0.80个百分点；其AR较SE、CBAM、ECA分别提升0.50、0.40、0.65个百分点；其PCK较SE、CBAM、ECA分别提升0.15、0.27、0.15个百分点；其NME则较SE降低20.0%，较CBAM降低6.67%，较ECA降低46.15%. 尤为关键的是，得益于无参特性，SimAM在推理时间上表现最优，成为所有对比方法中推理速度最快的模型.

3.3.2　MaxDSC2不同中间通道数的性能对比

为验证MaxDSC2模块中间通道数对模型性能的影响，以Star-RTMPose为基础框架，固定其他模块不变，仅改变MaxDSC2模块中第一个深度可分离卷积输出特征图的中间通道数，在相同的连铸坯数据集上分别测试154、173、192、211、230五个值（分别选取MaxDSC2输入特征图通道数的0.40、0.45、0.50、0.55、0.60倍），实验结果如表2所示. 由表2实验结果可见，当MaxDSC2模块的中间通道数设置为173时，与154通道相比，网络模型的参数量、计算量及推理时间虽略有增加，但其AP、AR、PCK分别提升0.10、0.05、0.03个百分点，NME降低9.68%；与192通道相比，173通道的AP提升0.02个百分点，AR提升0.03个百分点，PCK二者持平，且NME降低6.67%；与211、230通道相比，173通道的AP分别提升0.69、0.89个百分点，AR分别提升0.55、0.70个百分点，PCK分别提升0.27、0.35个百分点，NME分别降低31.71%、33.33%. 上述结果充分证明选择173作为MaxDSC2中间通道数的合理性.

表2　MaxDSC2不同中间通道数性能对比

中间通道数	Input Size	Params/M	FLOPs/G	AP/%	AR/%	PCK/%	NME/%	*T*/ms
154	512×512	**3.86**	**1.93**	99.50	99.70	99.87	0.31	**9.64**
173	512×512	3.87	1.94	**99.60**	**99.75**	**99.90**	**0.28**	9.86
192	512×512	3.88	1.94	99.58	99.72	**99.90**	0.30	9.94
211	512×512	3.89	1.95	98.91	99.20	99.63	0.41	10.27
230	512×512	3.90	1.95	98.71	99.05	99.55	0.42	10.63

注：表格中加粗显示的内容为各指标对应的最优表现.

3.3.3　不同模型关键点检测精度对比

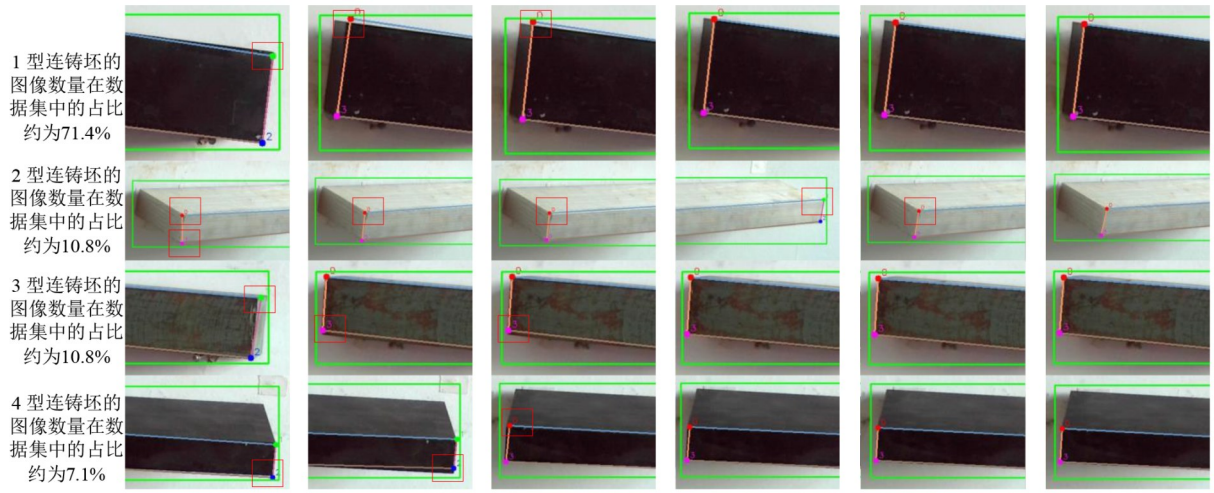
为充分验证所提模型的有效性，在相同的连铸坯数据集上与五种经典且广泛应用的主流网络开展系统的性能对比实验，包括HRNet-W32、SwinTransformer-T[22]、HRFormer-S、SimCC-MobileNetV2[23]和RTMPose-T，实验结果如表3所示.

由表3实验结果可见，改进后的RTMPose模型在精度与定位误差指标上表现最优，实现了最高的精度值以及最低的归一化平均误差. 与SimCC-MobileNetV2网络相比，该模型在参数量、计算量和推理时间仅有小幅增加的情况下，AP提升1.87%，AR提升1.48%，PCK提升0.67%，NME降低56.25%，在推理速度与模型性能之间实现了优异平衡. 将包含四种类型连铸坯的图像输入各模型进行关键点检测，可视化结果如图11所示.

表3　数据集上各算法的性能评价对比

算法	Input Size	Params /M	FLOPs/G	AP/%	AR/%	PCK/%	NME/%	*T*/ms
HRNet-W32	512×512	28.54	41.19	93.06	95.55	98.85	1.12	62.08
SwinTransformer-T[22]	512×512	32.76	10.30	91.15	92.15	98.27	1.23	19.61
HRFormer-S	512×512	7.75	16.14	92.51	95.05	98.50	1.10	76.71
SimCC-MobileNetV2[23]	512×512	**2.76**	**1.71**	97.77	98.30	99.24	0.64	**8.24**
RTMPose-T	512×512	3.44	1.77	98.51	98.85	99.50	0.49	8.64
本文算法	512×512	3.87	1.94	**99.60**	**99.75**	**99.90**	**0.28**	9.86

注：表格中加粗显示的内容为各指标对应的最优表现.



(*a*) SwinTransformer-T (*b*) HRFormer-S (*c*) HRNet-W32 (*d*) SimCC-MobileNetV2 (*e*) RTMPose-T (*f*) Ours

图 11 实验可视化结果

图 11 中最左侧一列呈现了各类连铸坯图像在数据集中的分布占比，绿框表示目标检测结果，四个不同颜色的标记点对应模型对连铸坯四个关键点的预测位置，以红框标注可视化结果中预测误差较大的点．可见，本文算法的关键点预测位置更贴近真实位置．

3.3.4 三维测量精度对比

通过改进的 RTMPose 网络获取左右图像关键点的二维像素坐标后，根据双目相机标定得到的内外参数，将二维坐标转换为三维坐标并完成连铸坯的测量．

首先通过关键点检测网络可获得某一关键点在左右图像中的像素坐标(u_l,v_l)、(u_r,v_r)，现实中某一点在相机坐标系下的三维坐标与图像中这一点的像素坐标之间的映射关系为

$$Z\begin{pmatrix}u\\v\\1\end{pmatrix}=\begin{pmatrix}f_x&0&c_x\\0&f_y&c_y\\0&0&1\end{pmatrix}\begin{pmatrix}X\\Y\\Z\end{pmatrix} \tag{10}$$

其中，c_x、c_y为图像主点在像素坐标系下的x、y坐标；f_x、f_y为相机在其自身坐标系的x轴和y轴方向上的焦距，单位为像素；u、v为关键点在图像上的像素坐标；X、Y、Z为关键点在相机坐标系下的三维坐标．

然后根据三角测量原理，可知现实中关键点的深度值为

$$Z=\frac{f_x\times b}{d} \tag{11}$$

其中，f_x为相机在水平方向的焦距，单位为像素；b为双目相机中左右两相机之间的基线长度，单位为 mm；d为关键点在左右图像上的视差值，即$d=u_l-u_r$，单位为像素；Z为关键点到相机成像平面的深度，单位为 mm．

实验所用相机的像素为正方形像素，由式(10)和式(11)可知关键点在左侧相机坐标系下的三维坐标为

$$\begin{cases}X=\dfrac{b\times\left(u_l-c_x\right)}{d}\\[2ex]Y=\dfrac{b\times\left(v_l-c_y\right)}{d}\\[2ex]Z=\dfrac{f_x\times b}{d}\end{cases} \tag{12}$$

最终，通过计算不同关键点之间的三维欧氏距离，即可得到连铸坯的实际尺寸．

以同一连铸坯在不同拍摄角度、不同拍摄距离以及不同光照强度等条件下的图像作为网络输入，对测量尺寸取均值作为最终结果，并采用绝对误差和相对误差衡量测量精度，四类连铸坯的尺寸测量数据如表 4 所示．

表 4 本文算法长方体连铸坯尺寸测量结果

类型编号	边	真实尺寸/mm	测量尺寸/mm		绝对误差/mm		相对误差/%	
			均值	最优值	均值	最优值	均值	最优值
1	长	263.020	263.275	262.790	2.522	0.230	0.959	0.087
	宽	100.920	100.625	100.941	1.906	0.021	1.889	0.021
2	长	301.740	302.802	301.729	3.491	0.012	1.157	0.004
	宽	30.300	30.568	29.893	1.945	0.407	6.418	1.345
3	长	300.100	299.278	299.582	3.282	0.518	1.094	0.173
	宽	68.040	67.967	67.937	2.942	0.103	4.323	0.151
4	长	203.840	203.271	204.049	2.052	0.209	1.007	0.102
	宽	30.460	30.075	30.504	1.034	0.044	3.394	0.144

由表 4 可知，2 型、3 型和 4 型连铸坯的宽边相对误差均值较大，其余测量结果的相对误差均值皆不超过 2%，可能的原因如下：数据集中 2 型、3 型和 4 型连铸坯图像数量占比较小；小尺寸物体的检测精度相对较差；

相机标定所得的内外参数存在误差.

如表5所示,以1型连铸坯长边和4型连铸坯宽边的测量为例,将本文算法的测量结果分别与传统双目视觉检测算法和基于深度学习的双目视觉检测算法的测量结果进行对比分析. 传统双目视觉检测算法包括SIFT、ORB以及Improved FAST[5],基于深度学习的双目视觉检测算法包括Improved TransUNet[24]和Improved HRNet[25],由表5可知,本文算法在测量精度上显著优于其他对比算法,其相对误差最小. 在1型连铸坯长边测量中,与Improved FAST算法相比,本文算法的相对误差降低0.365个百分点;相比于Improved TransUNet算法,相对误差降低0.241个百分点. 在4型连铸坯宽边测量中,本文算法较Improved HRNet算法的相对误差降低4.924个百分点.

表5　各算法测量结果对比

算法	1型连铸坯长边			4型连铸坯宽边		
	测量尺寸/mm	真实尺寸/mm	相对误差/%	测量尺寸/mm	真实尺寸/mm	相对误差/%
SIFT	257.684	263.020	2.029	27.091	30.460	11.060
ORB	267.785		1.812	25.836		15.181
Improved FAST[5]	261.806		0.462	—		—
Improved TransUNet[24]	262.132		0.338	33.470		9.882
Improved HRNet[25]	261.626		0.530	32.345		6.188
本文算法	263.275		**0.097**	30.075		**1.264**

注:表格中加粗显示的内容为相对误差对应的最优表现.

3.4　消融实验

在连铸坯数据集上进行消融实验,以测试每个改进模块对网络模型整体性能的影响,消融实验结果如表6所示.

实验以RTMPose-T作为基准模型,将实验分为7个阶段,分别测试每个改进模块及其组合的效果,以清晰展现不同模块的具体贡献.

由表6可知,在主干网络的末端引入StarTriBlock模块后,模型仅增加较少的参数量与计算量,AP提升0.69个百分点,PCK提升0.25个百分点,NME降低26.53%. 因为SimAm无需额外参数量,所以使用SimAm注意力模块替换CSPLayer中的通道注意力模块后,网络参数量降低5.81%的同时,AP提升0.29个百分点,PCK提升0.20个百分点. 将头部网络的7×7卷积层替换为MaxDSC2模块后,AP进一步提升0.39个百分点,PCK提升0.12个百分点,NME降低22.45%. 实验结果表明,模块组合优化的效果优于单一模块的效果. 改进后模型较基准模型RTMPose-T,AP提升1.09个百分点、AR提升0.9个百分点,PCK提升0.40个百分点,NME降低42.86%,且推理时间仅小幅增加,仍保持实时性,充分体现了改进方案的有效性与高效性.

表6　消融实验结果

编号	StarTriBlock	SimAm	MaxDSC2	Params/M	FLOPs/G	AP/%	AR/%	PCK/%	NME/%	*T*/ms
1	—	—	—	3.44	**1.77**	98.51	98.85	99.50	0.49	8.64
2	√	—	—	4.09	1.94	99.20	99.35	99.75	0.36	10.40
3	—	√	—	3.24	**1.77**	98.80	99.00	99.70	0.44	8.35
4	—	—	√	3.42	**1.77**	98.90	99.10	99.62	0.38	8.51
5	√	√	—	3.89	1.94	99.40	99.45	99.88	0.34	9.88
6	√	—	√	4.07	1.94	99.51	99.65	99.82	0.32	9.90
7	—	√	√	**3.22**	**1.77**	99.39	99.50	99.78	0.35	**8.31**
8	√	√	√	3.87	1.94	**99.60**	**99.75**	**99.90**	**0.28**	9.86

注:表格中加粗显示的内容为各指标对应的最优表现.

4　结论

(1)基于RTMPose-T网络架构,通过引入StarTriBlock模块、MaxDSC2模块及SimAM注意力机制,构建了高效精准的Star-RTMPose连铸坯关键点检测模型. 该模型在AP、AR、PCK和NME三项关键指标上均显著优于HRNet-W32、SwinTransformer-T和RTMPose-T等主流网络. 与推理速度最快的SimCC-MobileNetV2网络相比,AP提升1.87%、AR提升1.48%、PCK提升0.67%、NME降低56.25%,且推理耗时仅小幅增加,实现了推理速度与模型性能的优异平衡.

(2)在连铸坯测量精度上,本文算法显著优于传统双目视觉检测算法及文中对比的深度学习方法. 在

1 型连铸坯长边测量中,与 Improved FAST 算法相比,本文算法相对误差降低 0.365 个百分点;在 4 型连铸坯宽边测量中,较 Improved HRNet 算法相对误差降低 4.924 个百分点. 实验结果表明,该方法能够满足工业场景下的高精度测量需求.

参考文献

[1] 安徽工业大学. 一种去除板坯毛刺的系统: CN102935547B [P]. 2014-10-15.

[2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[3] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[4] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision. Piscataway: IEEE, 2012: 2564-2571.

[5] 宋超群, 许四祥, 杨宇, 等. 基于改进FAST和BRIEF的双目视觉测量方法[J]. 激光与光电子学进展, 2022, 59(8): 173-180.
SONG C Q, XU S X, YANG Y, et al. Binocular vision measurement method using improved FAST and BRIEF[J]. Laser & Optoelectronics Progress, 2022, 59(8): 173-180. (in Chinese)

[6] 宋祥, 许四祥, 杨利法, 等. 基于非线性扩散与高维M-SURF描述符的双目视觉测量方法[J]. 光电子·激光, 2024, 35(4): 405-413.
SONG X, XU S X, YANG L F, et al. Binocular vision measurement method based on nonlinear diffusion and high-dimensional M-SURF descriptor[J]. Journal of Optoelectronics·Laser, 2024, 35(4): 405-413. (in Chinese)

[7] XU S X, DONG C C, ZHOU S H, et al. Binocular measurement method for the continuous casting slab model based on the improved BRISK algorithm[J]. Applied Optics, 2022, 61(11): 3019-3025.

[8] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: Real-time multi-person 2D pose estimation using part affinity fields[EB/OL]. (2019-05-30)[2025-10-10]. https://arXiv.org/abs/1812.08008.

[9] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 5686-5696.

[10] 江佳鸿, 夏楠, 李长吾, 等. 基于多尺度增量学习的单人体操动作中关键点检测方法[J]. 电子学报, 2024, 52(5): 1730-1742.
JIANG J H, XIA N, LI C W, et al. Keypoint detection method for single person gymnastics actions based on multi-scale incremental learning[J]. Acta Electronica Sinica, 2024, 52(5): 1730-1742. (in Chinese)

[11] YUAN Y H, FU R, HUANG L, et al. HRFormer: High-resolution transformer for dense prediction[EB/OL]. (2021-11-07)[2025-10-10]. https://arXiv.org/abs/2110.09408.

[12] JIANG T, LU P, ZHANG L, et al. RTMPose: Real-time multi-person pose estimation based on MMPose[EB/OL]. (2023-07-03)[2025-11-11]. https://arXiv.org/abs/2303.07399.

[13] LI Y J, YANG S, LIU P D, et al. SimCC: A simple coordinate classification perspective forHuman pose estimation[C]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 89-106.

[14] LYU C Q, ZHANG W W, HUANG H A, et al. RTMDet: An empirical study of designing real-time object detectors [EB/OL]. (2022-12-16)[2025-10-10]. https://arXiv.org/abs/2212.07784.

[15] MA X, DAI X Y, BAI Y, et al. Rewrite the stars[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 5694-5703.

[16] YANG L X, ZHANG R Y, LI L D, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning. Cambridge: PMLR, 2021(139): 11863-11874.

[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.

[18] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Computer Vision - ECCV 2018. Cham: Springer, 2018: 3-19.

[19] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11531-11539.

[20] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1800-1807.

[21] LI X W, SUN K, FAN H B, et al. Real-time cattle pose estimation based on improved RTMPose[J]. Agriculture, 2023, 13(10): 1938.

[22] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierar-

chical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 9992-10002.

[23] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4510-4520.

[24] 李同谱, 许四祥, 施宇翔, 等. 基于双目视觉与Transformer的连铸坯模型定位与测量[J]. 中南大学学报(自然科学版), 2024, 55(4): 1312-1322.
LI T P, XU S X, SHI Y X, et al. Continuous casting slab model positioning and measurement based on binocular vision and Transformer[J]. Journal of Central South University (Science and Technology), 2024, 55(4): 1312-1322. (in Chinese)

[25] 任加琪, 许四祥, 董宾卉, 等. 基于轻量化HRNet的双目视觉定位与测量[J/OL]. 中国机械工程, 2024: 1-9[2025-10-10]. https://kns.cnki.net/kcms/detail/42.1294.TH.20241211.1933.008.html.
REN J Q, XU S X, DONG B H, et al. Binocular vision localization and measurement based on lightweight HRNet[J/OL]. China Mechanical Engineering, 2024: 1-9. https://kns.cnki.net/kcms/detail/42.1294.TH.20241211.1933.008.html. (in Chinese)

作者简介

张梦权　男,2001年4月出生于安徽省宿州市. 现为安徽工业大学机械工程学院硕士研究生. 主要研究方向为机器人与机器视觉.
E-mail: 2992466836@qq.com

杨　玉　男,2001年11月出生于安徽省安庆市. 现为安徽工业大学机械工程学院硕士研究生. 主要研究方向为机器人与机器视觉.
E-mail: 1308889562@qq.com

许四祥　男,1974年6月出生于湖北省汉川市. 现为安徽工业大学机械工程学院教授、硕士生导师. 主要研究方向为机器人与机器视觉.
E-mail: xsxhust@ahut.edu.cn

吴端正　男,2000年10月出生于安徽省蚌埠市. 现为安徽工业大学机械工程学院硕士研究生. 主要研究方向为机器人与机器视觉.
E-mail: 3251387273@qq.com