

基于轻量化高分辨率网络的双目视觉定位与测量

任加琪 许四祥* 董宾卉 汤 澳 宋昱宸

安徽工业大学机械工程学院, 马鞍山, 243032

摘要:针对基于特征点检测的双目视觉测量效率低、神经网络计算复杂度高等问题,提出了基于轻量化高分辨率网络(HRNet)的双目视觉定位与测量方法。轻量化HRNet以HRNet为基准,先替换卷积模块、缩减参数量,再引入Transformer提取全局图像特征,最后使用多级上采样融合策略捕获多尺度特征信息。与原HRNet模型相比,轻量化HRNet模型参数减少95.40%,计算量、归一化平均误差分别减小94.27%和6.25%;三维测量上,轻量化HRNet与双目视觉结合方法的相对误差达到0.256%,能在低算力硬件上实现高精度检测。

关键词:双目视觉;高分辨率网络;轻量化;关键点检测;尺寸测量

中图分类号: TP391.41

DOI:10.3969/j.issn.1004-132X.2026.01.021

开放科学(资源服务)标识码(OSID):



Binocular Vision Localization and Measurement Based on Lightweight HRNet

REN Jiaqi XU Sixiang* DONG Binhui TANG Ao SONG Yuchen

School of Mechanical Engineering, Anhui University of Technology, Ma'anshan, Anhui, 243032

Abstract: Aiming at the problems of low efficiency in binocular vision measurements based on feature point detection and high computational complexity of neural networks, a binocular vision localization and measurement method was proposed based on a lightweight HRNet. The lightweight HRNet was built upon the original HRNet by replacing the convolutional modules to reduce the number of parameters, introducing Transformer to extract global image features, and employing a multi-level upsampling fusion strategy to capture the multi-scale feature information. Compared with the original HRNet model, the lightweight HRNet reduces model parameters by 95.40%, while computational loads and normalized mean errors are decreased by 94.27% and 6.25% respectively. In terms of 3D measurement, the relative errors of the method combining lightweight HRNet with binocular vision reach 0.256%, enabling high-precision detection on hardware with low computational power.

Key words: binocular vision; high resolution net (HRNet); lightweight; landmark detection; measurement

0 引言

冶金行业中,使用火焰切割连铸坯时,液态金属会在连铸坯切面底部生成一条不规则且硬度大的毛刺,严重影响连铸坯质量,因此去毛刺是切割连铸坯过程中不可或缺的工序。等离子弧去连铸坯毛刺方案^[1]利用机器人末端执行器切除毛刺,但存在连铸坯偏移导致无法定位的问题,故需使用双目视觉进行定位与测量,协助机器人去除毛刺。

双目视觉通过算法处理采集的图像来代替人

工测量,是无接触定位与测量的一种理想方法。传统的视觉检测方法有尺度不变特征变换(scale invariant feature transform, SIFT)、定向FAST和旋转BRIEF(oriented FAST and rotated BRIEF, ORB)、方向梯度直方图(histogram of oriented gradient, HOG)。XU等^[2]使用低偏移值区域去除法分配特征点主方向,利用差异阈值得到描述符,有效减小了光滑区域因噪声引起的描述符不稳定的影响。周书华等^[3]筛选高熵区域并结合局部二值模式(local binary patterns, LBP)与旋转二进制鲁棒独立基本特征(rotated binary robust independent elementary features, rBRIEF)进行特征匹配,提高了检测效率和匹配正确率。宋祥等^[4]通过改变邻域形状的方式增强弱边缘的描

收稿日期:2024-04-24

基金项目:国家自然科学基金(51374007);安徽高校自然科学基金研究重点项目(KJ2020A0259)

述,提高了加速非线性扩散(accelerated nonlinear diffusion, KAZE)特征检测算法的匹配精度。上述方法整体上对参数比较敏感,实际应用具有局限性。

深度学习方法为计算机视觉和机器学习带来革命性的进步^[5]。卷积神经网络(convolutional neural networks, CNN)^[6-7]是图像识别中最具代表性的深度学习模型之一,但存在存储开销大、计算效率低、无法获取全局信息等问题。全卷积网络(fully convolutional networks, FCN)^[8]和U-Net网络^[9]虽在不同程度上弥补了CNN的不足,但其结构中的下采样操作容易造成特征信息的丢失。高分辨率网络(high resolution net, HRNet)^[10]通过并行和交互多个分辨率全程保持高分辨率特征,在依赖精细空间结构的位置感知任务(如人体姿态估计)中表现出色,但高分辨率分支会引入较高的计算复杂度与较多参数,同时其低分辨率分支因感受野有限且深度较浅,因此建模的语义表征不够充分。在具有最强语义表示的特征图后增加更多阶段的U-HRNet^[11]通过增强上下文信息的提取,有效提高了模型的全局语义表征能力。针对网络参数量大的问题,MobileNets^[12]、ShuffleNet^[13]能有效减少参数,降低计算量。基于多尺度池化金字塔的结构化自蒸馏学习模型^[14]可提高网络的特征表示能力。

基于卷积的神经网络在捕捉长距离依赖关系和全局上下文信息方面存在局限性。受到自然语言处理(natural language processing, NLP)的启发,研究者将Transformer模型^[15]引入计算机视觉领域并且取得较好效果。CHEN等^[16]结合Transformer与U-Net提出的Transunet在分割任务中取得较好效果。结合卷积与Transformer的混合模型使网络既有卷积的归纳偏置特性,又有Transformer的全局归纳建模能力。

基于此,笔者提出一种基于轻量化HRNet的双目视觉定位与测量方法,使用关键点检测网络检测双目相机采集图片的关键点。引入Transformer增强网络捕获长距离依赖关系的能力,使用多级上采样融合策略进一步强化特征的上下文信息表征能力。使用重建模块减少通道和空间中的参数冗余,增强特征学习能力,提高关键点检测精度。最后根据三角测量原理获得板坯的三维坐标信息,完成测距任务。

1 双目视觉定位与测量方法

基于轻量化HRNet的双目视觉定位与测量

方法步骤主要包括数据集制作、关键点检测网络模型训练、定位与测量。总体框架如图1所示。



图1 定位与测量总体框架

Fig.1 General framework for positioning and measurement

1.1 数据集制作及预处理

为准确去除毛刺,需对连铸坯进行定位(定位连铸坯关键点 A 、 B 、 C 、 D)与测量,如图2所示,左图中的 A_1 、 B_1 、 C_1 、 D_1 和右图中的 A_2 、 B_2 、 C_2 、 D_2 分别对应关键点 A 、 B 、 C 、 D 。

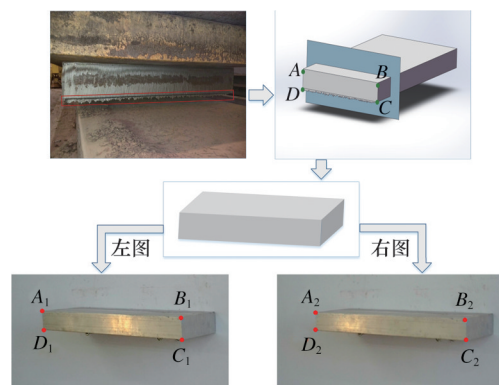


图2 连铸坯模型示意图

Fig.2 Schematic diagram of continuous casting billet model

将已标定的双目深度相机采集的连铸坯左右图片作为数据集来分析算法的可行性。首先通过标定法^[17]与OpenCV视觉库标定相机。然后在不同视角、不同亮度、不同旋转角度等条件下采集4种连铸坯的图像,根据标定结果对采集的连铸坯图片进行畸变校正和极线校正。最后使用LabelMe标注工具分别对校正后的连铸坯左右图的4个关键点进行顺序标注。数据集初始采集图片的分辨率为1600像素 \times 1200像素。立体校正(图像预处理)后,图片分辨率调整为1664像素 \times 1152像素。1220张采集图片中,训练集有1150张图片,测试集有70张图片。

为提高网络模型的泛化能力,对训练集进行离线数据增强,增强后的数据集部分图像见图3,图中的绿点为数据标注中点的位置。

数据增强方式主要包含图片平移、添加噪声、调整亮度、调整对比度等操作。增强后的数据集共有2326张图片,其中,训练集图片2256张,测试集图片70张。

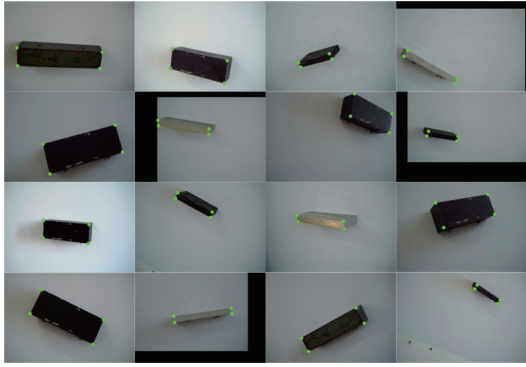


图3 部分数据集图像

Fig.3 Partial dataset image

1.2 网络模型

轻量化 HRNet 网络整体结构如图 4 所示。以 HRNet-18 为基准模型,整体结构分为预处理(Stem)、主干网络(Backbone)、颈部(Neck)、和输出头(Head)。

网络输入图片 X 的尺寸为 $H \times W$ 、通道为 3,

即 $X \in \mathbb{R}^{H \times W \times 3}$ 。在预处理阶段,输入图片 4 倍下采样得到通道数为 C 的特征图 $X_1 \in \mathbb{R}^{H/4 \times W/4 \times C}$ 。主干网络有 4 个阶段,在每个阶段依次增加 1 个平行分支。对平行分支特征图进行 2 倍下采样,通道数增加一倍。每个分支由 2 个 Shuffle block 和 1 个多分辨率融合单元组成,各单元在其指定的分辨率上进行特征提取与跨分支交互。主干网络最终输出 4 个尺度的特征图,它们宽度的分辨率分别为原图的 $1/4, 1/8, 1/16, 1/32$,通道数分别为 $C, 2C, 4C, 8C$ 。在分辨率最低的特征流上加入 Transformer 层,以获取空间的相互关系。在网络颈部采用多级上采样融合策略,逐步融合低分辨率特征图与高分辨率特征图。低分辨率特征图上采样后,与高分辨率特征图进行通道连接。之后,通过重建模块减少空间和通道的冗余,得到融合特征图。最后,通过输出头模块将特征图上采样至与输入图像相同的分辨率,并生成热图。

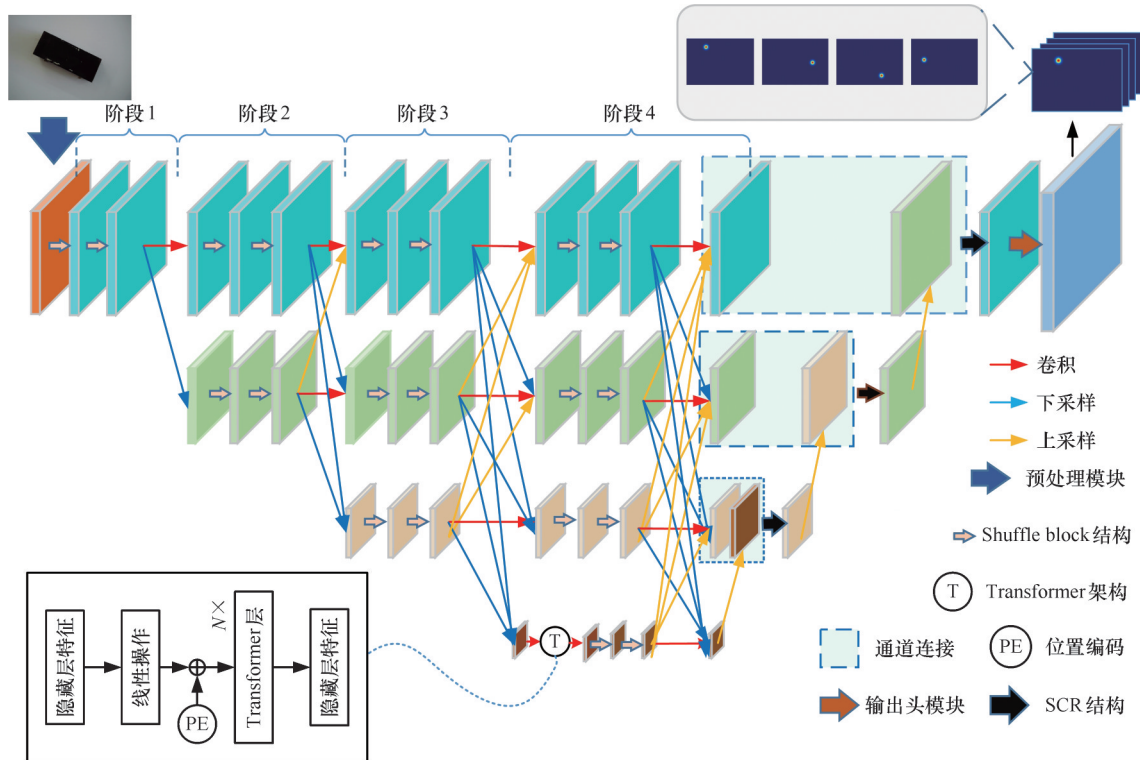


图4 轻量化 HRNet 连铸坯关键点检测网络架构

Fig.4 Lightweight HRNet continuous casting billet key point detection network architecture

1.2.1 主干网络轻量化

为减少参数,将预处理模块中的第二个卷积和所有残差单元替换为图 5 所示的 Shuffle block,其中,Conv 表示卷积操作,DWConv 表示深度卷积,Channel Split 表示通道分离,Concat 表示通道连接,Channel Shuffle 表示通道混洗。先对输入的特征图进行通道分组卷积(分 2 组依次进行卷积),提高模型的表达能力。通道连接后,进行通

道混洗,交换不同通道的信息,并将特征融合过程中所有标准卷积替换为深度可分离卷积。

卷积部分使用的深度可分离卷积主要包含深度卷积(减少模型参数并提取空间特征)和逐点卷积(扩展通道数并提取通道特征)。

假设输入特征图的尺寸为 $H \times W \times C$,输出通道数为 M ,其中, C 为输入通道数。当进行深度可分离卷积时,深度卷积的核大小为 $D \times D \times 1$,

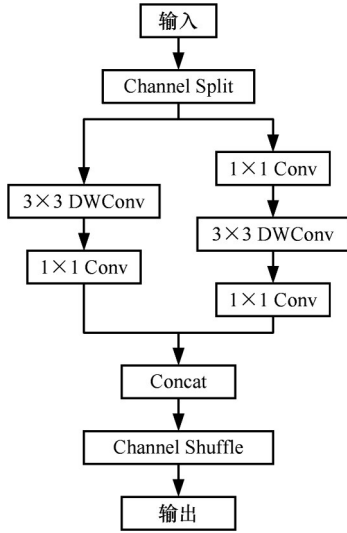


图5 Shuffle block 结构

Fig.5 Shuffle block structure

逐点卷积的核大小为 $1 \times 1 \times C$ 。由于输出通道数为 M ，所以逐点卷积核个数为 M ，输出特征图尺寸为 $H_{out} \times W_{out} \times M$ 。深度可分离卷积的原理如图6所示。

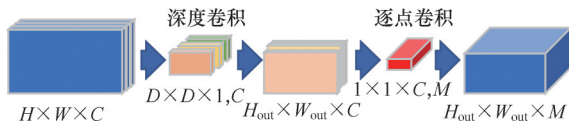


图6 深度可分离卷积原理图

Fig.6 Depth separable convolution schematic

那么深度可分离卷积的参数量与标准卷积的参数量之比为

$$\eta = \frac{D \times D \times 1 \times C + 1 \times 1 \times C \times M}{D \times D \times C \times M} = \frac{1}{M} + \frac{1}{D^2} \quad (1)$$

计算量之比为

$$\lambda = \frac{H \times W \times D \times D \times 1 \times C + 1 \times 1 \times C \times M \times H \times W}{H \times W \times D \times D \times C \times M} = \frac{1}{M} + \frac{1}{D^2} \quad (2)$$

由式(1)、式(2)可知，深度可分离卷积与标准卷积的参数量之比 η 和计算量之比 λ 相同，都是 $1/M + 1/D^2$ 。输出通道数为 128、256 时， $\eta \approx 1/D^2$ 。使用 3×3 卷积时，深度可分离卷积的参数量约是标准卷积的 $1/9$ 。

1.2.2 注意力机制

Transformer 模型输入为序列，先将特征图映射到指定维度，再展平得到 n 个列向量组成的序列 $\{t_i \in \mathbb{R}^h | i=1, 2, \dots, n\}$ ，其中， h 为隐藏层的特征图数量，文中， $h=256, n=117$ 。为构建序列的空间位置信息，加入可学习的位置编码 $E_{pos} \in \mathbb{R}^{h \times n}$ ，得到二维序列：

$$S_0 = [t_1 \quad t_2 \quad \dots \quad t_n] + E_{pos} \quad (3)$$

随后将二维序列 S_0 输入 Transformer 层，Transformer 层结构如图7所示，其中 Layer Normer 表示层归一化，MLP(multi-layer perceptron)代表多层感知机，MatMul代表矩阵乘法运算， q, k 以及 v 分别表示查询向量、键向量以及值向量。

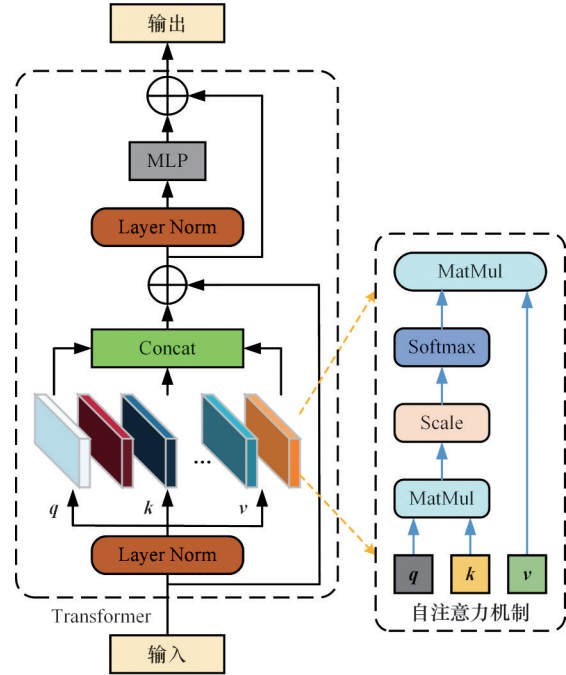


图7 Transformer 层结构

Fig.7 Transformer layer architecture

S_0 进入 Transformer 层后，通过多头自注意力机制得到注意力矩阵

$$A = \sum_{j=1}^n \text{softmax} \left(\frac{q_i \cdot k_j}{\sqrt{d_k}} \right) v_j \quad (4)$$

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_i \exp(x_i)} \quad (5)$$

式中： q_i 为第 i 个序列的查询向量； k_j, v_j 分别为第 j 个序列的键向量和值向量； d_k 为多头注意力的头部大小，即键向量 k 的维度，防止 $\text{softmax}(\cdot)$ 的梯度过小； x_i 为 $q_i \cdot k_j$ 的未归一化注意力权重中的第 i 个元素。

然后使用多层感知机进行加工处理。

1.2.3 多级上采样融合

针对低分辨率分支特征信息利用不足的问题，在颈部使用多级上采样融合的方式自底向上逐步融合多尺度特征。低分辨率特征图向高分辨率特征图的融合过程如图8所示。首先，将低分辨率特征图上采样至与到高分辨率特征图相同的分辨率。然后，将上采样后的特征图与对应的高分辨率特征图进行通道拼接。融合后的特征图包含的特征信息较多，故接入特征建模模块(SCR)，以抑制冗余特征，得到低分辨率与高分辨率的融

合特征图。SCR由2个 1×1 卷积和1个空间和通道重构卷积(spatial and channel reconstruction convolution, SCConv)^[18]组成,空间和通道重构卷积可自适应抑制特征冗余并促进代表性特征的学习,有效降低模型的计算复杂度和参数量。

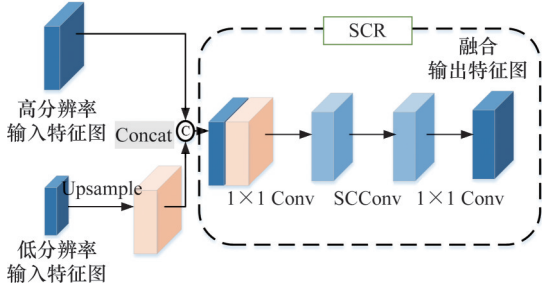


图8 不同分辨率特征图融合

Fig.8 Fusion of feature maps at different resolutions

SCConv由空间重建单元(SRU)和通道重建单元(CRU)构成,整体结构如图9所示。

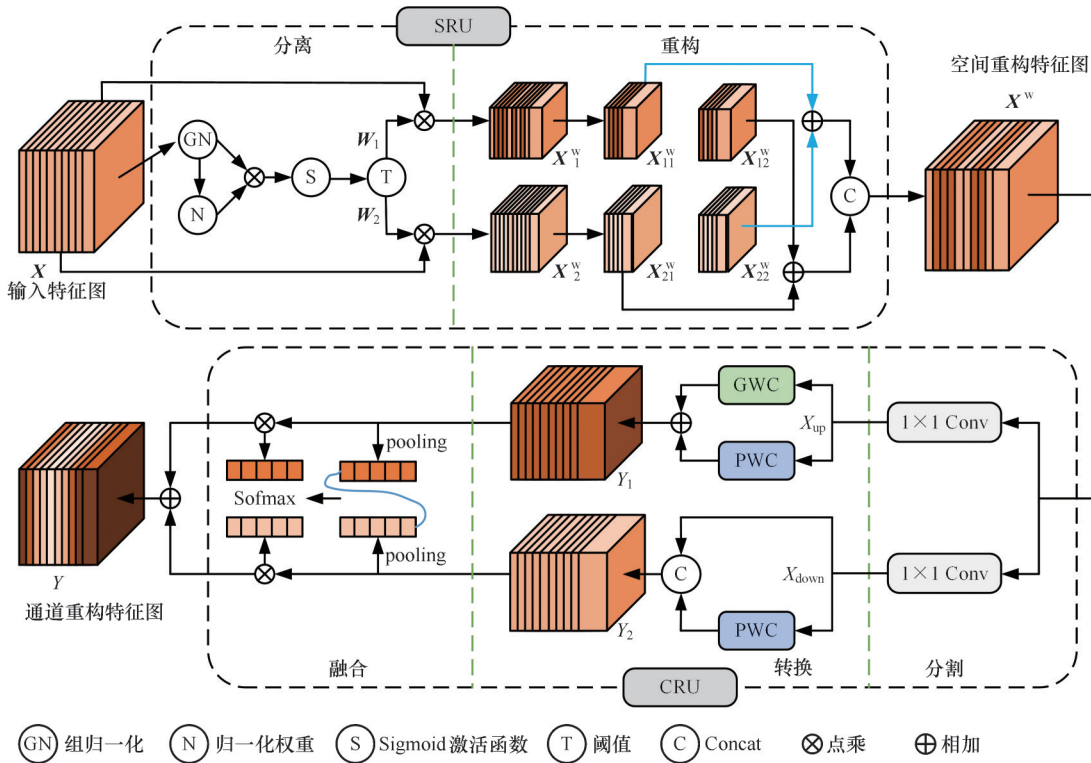


图9 SCConv模块结构图

Fig.9 Structure diagram of the SCConv module

将输入特征图 X 与掩码 W_1 、 W_2 相乘,得到信息量大的特征 X_1^w 和信息量小的冗余特征 X_2^w ,然后将 X_1^w 分通道拆分成 X_{11}^w 、 X_{12}^w ,将 X_2^w 分通道拆分成 X_{21}^w 、 X_{22}^w ,并使用交叉重建方式将信息量大和信息量小的特征融合,获得空间重构特征图:

$$X^w = X_1^w \cup X_2^w \quad (8)$$

$$X_1^w = X_{11}^w \oplus X_{22}^w \quad X_2^w = X_{21}^w \oplus X_{12}^w$$

$$X_1^w = W_1 \otimes X \quad X_2^w = W_2 \otimes X$$

式中: \cup 表示通道连接操作; \oplus 表示逐元素相加; \otimes 表示逐

元素相乘。

空间重构单元通过分离和重建抑制空间的冗余。首先将输入特征图 $X \in \mathbb{R}^{H \times W \times C}$ 进行组归一化:

$$\text{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta^2$$

$$\gamma = (\gamma_1, \gamma_2, \dots, \gamma_C)$$

式中: μ 、 σ 分别为均值和方差; ϵ 为极小常数; γ 、 β 分别为可训练的缩放和平移参数向量。

对组归一化后的特征进行进一步处理,得到归一化权重:

$$W_\gamma = (\gamma_1, \gamma_2, \dots, \gamma_C) / \sum_{j=1}^C \gamma_j \quad (6)$$

使用Sigmoid函数将重新加权的特征图权重映射到 $[0, 1]$ 内,并设置阈值门控。将大于阈值的权重设置为1,形成二进制掩码 W_1 ;将小于阈值的权重设置为0,形成二进制掩码 W_2 ,即

$$W_i = \text{Gate}(\text{sigmoid}(W_\gamma \otimes \text{GN}(X))) \quad i = 1, 2 \quad (7)$$

元素相乘。

空间重构后,特征图在通道维度上仍存在冗余。通道重建主要包含3个阶段即分割、转换和融合,可进一步减少特征在通道维度上的冗余。

1)分割阶段。先将空间重构特征图按1:1比例分为2个通道,然后利用 1×1 卷积压缩特征图的通道提高计算效率,最后将空间重构特征图 X^w 分为上部分 X_{up} 和下部分 X_{down} 。

2)转换阶段。 X_{up} 使用逐分组卷积(GWC)和

逐点卷积(PWC)提取特征 Y_1 , X_{down} 使用逐点卷积(PWC)提取特征 Y_2 。

3)融合阶段。通过全局平均池化 $pooling(*)$ 生成全局空间信息并堆叠,使用 softmax 函数将堆叠的全局空间信息映射为特征向量 β_1 和 β_2 ,并对 2 组特征进行通道连接,得到拥有通道细节的特征图:

$$Y = \beta_1 Y_1 + \beta_2 Y_2 \quad (9)$$

$$\beta_i = \frac{\exp(\theta_i)}{\exp(\theta_1) + \exp(\theta_2)} \quad \theta_i = pooling(Y_i) \quad i = 1, 2$$

2 实验结果与分析

2.1 实验环境与模型训练

实验采用 64 位 Windows 11 操作系统,处理器为 5800H,内存 16 GB,显卡为 RTX3060,框架是 PyTorch1.13。模型输入尺寸为 288×416 (高 \times 宽)。训练过程中采用均方误差损失(mean squared error loss, MSE Loss)函数来优化算法,调整模型参数,减小预测值和真实值的差。模型训练采用适应性矩估计 Adam 优化器,初始学习率设为 2×10^{-3} ,单卡训练批次样本大小(per-GPU batch size)为 4,迭代训练 100 轮。前 50 轮中,每 2 轮更新学习率,学习率衰减为原来的 $3/4$ 。

2.2 实验评价指标

实验采用均方根误差 E_{mse} 、归一化平均误差 E_{nme} 、网络的参数量 N_p 、浮点运算次数(计算量) N_F 、推理时间 T 评估关键点检测网络的性能。均方根误差

$$E_{rmse} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (10)$$

式中: N 为关键点数量; y_i, \hat{y}_i 分别为第 i 个关键点的真实值和预测值。

直接反映预测点与真实点的坐标差。使用归一化平均误差

$$E_{nme} = \frac{1}{Nd} \sum_{i=1}^N \|y_i - \hat{y}_i\|_2 \quad (11)$$

式中: d 为连铸坯长边的欧氏距离。

消除模型尺寸不同带来的误差不合理变化。

2.3 关键点检测精度对比实验

2.3.1 不同模型关键点检测精度对比

为评估提出模型的有效性,在连铸坯数据集上对比不同网络的训练结果,如表 1 所示。相较于 U-Net, Transunet 的均方根误差 E_{mse} 减小 39.60%,归一化平均误差 E_{nme} 减小 34.21%,虽然关键点检测精度提高较多,但带来了较大的参数量(原来的十几倍)和计算量。相较于 HRNet,本文算法的参数量减小 95.40%,计算量减小 94.27%,均方根误差减小 3.36%,归一化平均误

差减小 6.25%,在参数大幅减少的情况下,推理最快,归一化平均误差最小。

表 1 数据集上各算法的性能评价与对比

Tab.1 Performance evaluation and comparison of algorithms on datasets

算法	U-Net	Transunet	HRNet	本文算法
$N_p/10^6$	5.70	93.20	22.65	1.04
$N_F/(10^9 s^{-1})$	25.73	59.01	18.34	1.05
E_{mse}/mm	3.56	2.15	2.08	2.01
$E_{nme}/\%$	0.76	0.50	0.48	0.45
T/ms	86.21	102.21	116.27	62.50

为进一步验证轻量化 HRNet 的关键点检测性能,对不同模型进行关键点检测,实验的可视化结果如图 10 所示,其中,红点为预测点,绿点为真实点。图 10 中,有些红点被绿点覆盖了,有些红点露出来一些,最终可以判断本文算法的预测点最接近实点。

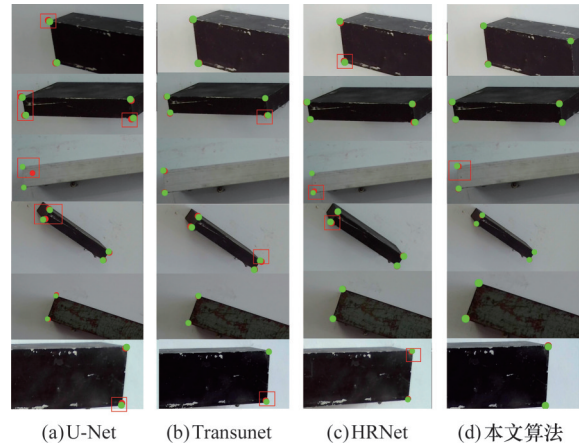


图 10 实验可视化结果

Fig.10 Experimental visual results

2.3.2 三维测量精度对比

首先,轻量化 HRNet 网络输出连铸坯左右图关键点的二维坐标。然后通过标定确定的双目相机的内外参数实现坐标的转化并测量连铸坯,同时将同一连铸坯不同拍摄角度的测量均值作为测量结果。由表 2 可知,不同尺寸的 4 个连铸坯长边测量绝对误差在 3mm 之内,相对误差均小于 1%,测量精度均符合测量要求。但短边的测量误差较大,2 号、3 号连铸坯的短边相对误差分别为 6.062% 和 5.784%,影响短边测量误差原因如下:相机的标定结果存在误差;目标物体之间的尺度差异导致小尺寸物体的检测精度低。

由表 3 可知,本文算法的测量精度最高,运行时间最短。1 号连铸坯长边测量的相对误差达 -0.256% ,比改进 KAZE 算法^[4]的精度高。推理运行时间大幅减少,运行时间是 ORB 算法的 12.475%。1 号连铸坯长边的测量精度略低于

表2 本文算法长方体连铸坯尺寸测量结果

Tab.2 Measurement results of cuboid continuous casting billet sizes based on the proposed algorithm

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	300.340	299.570	301.012	0.770	0.670	-0.256	0.223
	宽	68.300	70.161	68.181	1.861	0.120	2.725	-0.175
2	长	301.900	304.095	298.969	2.195	1.929	0.727	-0.639
	宽	31.200	33.091	31.146	1.891	0.054	6.062	-0.172
3	长	204.120	204.477	203.991	0.357	0.128	0.175	-0.063
	宽	30.580	32.345	30.891	1.769	0.312	5.784	1.019
4	长	262.680	261.626	263.008	1.054	0.328	-0.401	0.125
	宽	100.960	100.673	101.368	0.287	0.408	-0.285	0.404

表3 各算法的测量结果

Tab.3 Measurement results of each algorithm

算法	1号连铸坯长边			3号连铸坯短边			运行时间/s
	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	305.680	300.340	1.778	27.091	30.580	-11.409	19.918
ORB	295.340		-1.663	25.836		-15.513	0.505
改进ORB ^[3]	298.773		-0.520				0.640
改进KAZE ^[4]	301.431		0.363	28.305		7.440	3.556
Transunet*深度学习 ^[19]	299.844		-0.165	32.807		7.284	0.105
本文算法	301.012		-0.256	32.345		5.784	0.063

由表4可知,模型结构简化后,参数量、计算量、推理时间分别减小73%、60%和52.4%,但精度未显著降低,测试的归一化平均误差从0.48%增大到0.56%。结构简化减少了参数冗余和计算量,加快了推理。

表4 简化模型结构消融实验结果

Tab.4 Ablation experiment results of simplified model structural

方法	$N_p/10^6$	$N_f/(10^9s^{-1})$	$E_{nme}/\%$	T/ms
HRNet	22.64	18.34	0.48	116.27
HRNet-1	12.90	12.20	0.50	80.00
HRNet-2	10.08	9.93	0.52	60.61
HRNet-3	5.98	7.42	0.56	55.25

2.4.2 各模块的消融实验

虽然模型结构简化显著减少了参数冗余和计算,降低了模型复杂度,但导致模型检测精度降低。为在保持轻量化优势的基础上进一步优化计算复杂度和提高检测精度,改进了模型。

为验证改进模型的有效性,对主干网络轻量化、注意力机制、颈部多级上采样融合进行了消融实验。实验结果见表5,基准模型为HRNet-3,LT代表主干网络轻量化(包含残差单元替换成Shuffle block,融合部分的卷积改为深度可分离卷积),TF代表Transformer模型,SC代表多级上采

Transunet*深度学习方法^[19],但是运行时间大幅缩短。本文算法有效平衡了测量精度和运行时间,性价比最高。

2.4 消融实验

2.4.1 简化模型结构消融实验

HRNet模型主干网络部分包含4个阶段,4个阶段的重复次数分别为1、2、4、2。每个阶段包含多个并行的卷积流,每个卷积流重复4次堆叠残差单元。为降低模型复杂度,将残差单元的堆叠次数改为2(记为A改变),4个阶段设置为不重复(记为B改变)。HRNet-1为A改变、B不改变,HRNet-2为A不改变、B改变,HRNet-3为A改变、B改变。

表5 各模块消融实验结果

Tab.5 Ablation experiment results of each module

LT	TF	SC	$N_p/10^6$	$N_f/(10^9s^{-1})$	$E_{nme}/\%$
			5.98	7.42	0.56
✓			0.67	1.78	0.62
✓	✓		0.94	1.78	0.54
✓		✓	0.77	1.05	0.56
✓	✓	✓	1.04	1.05	0.45

样融合。

由表5可知,主干网络轻量化大幅度减小参数量和计算量,但归一化平均误差增大10.71%。在主干网络轻量化基础上增加Transformer导致网络参数增加270 000,归一化平均误差减小12.90%。颈部使用多级上采样融合后,网络参数增加90 000,归一化平均误差减小9.68%。加入Transformer和多级上采样融合的网络模型减小了归一化平均误差,提高了检测精度,在性能和效率之间做到很好的平衡。相较于基准模型,改进模型参数减少82.63%,归一化平均误差减小19.64%,验证了各个模块的有效性。

3 结论

1)在随机拍摄角度下使用双目相机采集并构建了连铸坯的定位与测量数据集。对数据集进行

数据增强,提高了模型的泛化性和鲁棒性。

2)在 HRNet 基础框架下,简化模型结构和主干网络轻量化有效减小网络参数数量和计算量,缩短推理运行时间;引入 Transformer 模型并在网络颈部使用多级上采样融合策略减小了归一化平均误差,提高了检测精度。

3)结合深度学习和双目视觉技术实现了连铸坯的定位与测量。轻量化 HRNet 网络不仅减小模型的参数数量和计算量,还保证了连铸坯测量精度,关键点检测的归一化平均误差达到 0.45%。在三维测距上,连铸坯测量精度高于传统特征检测算法,测量相对误差达到 0.256%,满足实际测量精度与实际部署要求。

参考文献:

- [1] 许四祥,陈富强,高培青,等.一种去除板坯毛刺的系统:CN102935547B[P].2014-10-15.
Xu Sixiang, Chen Fuqiang, Gao Peiqing, et al. System for Removing Slab Burrs: CN102935547B[P]. 2014-10-15.
- [2] XU Sixiang, DONG Chenchen, ZHOU Shuhua, 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.
- [3] 周书华,许四祥,董晨晨,等.基于局部信息熵和梯度漂移的双目视觉测量算法[J].激光与光电子学进展,2023,60(12):333-341.
ZHOU Shuhua, XU Sixiang, DONG Chenchen, et al. Algorithm for Binocular Vision Measurements Based on Local Information Entropy and Gradient Drift[J]. Laser & Optoelectronics Progress, 2023, 60(12): 333-341.
- [4] 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 Xiang, XU Sixiang, YANG Lifa, 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.
- [5] 谢阳,戴逸群,张超勇,等.融合集成模型与深度学习的机床能耗识别与预测方法[J].中国机械工程,2023,34(24):2963-2974.
XIE Yang, DAI Yiqun, ZHANG Chaoyong, et al. A Method for Identifying and Predicting Energy Consumption of Machine Tools by Combining Integrated Models and Deep Learning[J]. China Mechanical Engineering, 2023, 34(24): 2963-2974.
- [6] BORRA S R, PREMALATHA B, DIVYA G, et al. Deep Hashing with Multilayer CNN-based Biometric Authentication for Identifying Individuals in Transportation Security[J]. Journal of Transportation Security, 2024, 17(1): 4.
- [7] MASUDA Y, ISHIKAWA R, TANAKA T, et al. CNN-based Fully Automatic Mitral Valve Extraction Using CT Images and Existence Probability Maps[J]. Physics in Medicine & Biology, 2024, 69(3): 035001.
- [8] SHELHAMER E, LONG J, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [M] //Medical Image Computing and Computer-assisted Intervention—MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [10] SUN Ke, XIAO Bin, LIU Dong, et al. Deep High-resolution Representation Learning for Human Pose Estimation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 5686-5696.
- [11] WANG Jian, LONG Xiang, CHEN Guowei, et al. U-Net: Delving into Improving Semantic Representation of High Resolution Network for Dense Prediction[J]. arXiv: 2210.07140.
- [12] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. arXiv:1704.04861.
- [13] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: an Extremely Efficient Convolutional Neural Network for Mobile Devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 6848-6856.
- [14] 郑云飞,王晓兵,张雄伟,等.基于金字塔知识的自蒸馏 HRNet 目标分割方法[J].电子学报,2023,51(3):746-756.
ZHENG Yunfei, WANG Xiaobing, ZHANG Xiongwei, et al. The Self-distillation HRNet Object Segmentation Based on the Pyramid Knowledge[J]. Acta Electronica Sinica, 2023, 51(3): 746-756.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, 2017:6000-6010.
- [16] CHEN Jieneng, LU Yongyi, YU Qihang, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation[J]. arXiv:2102.04306.
- [17] ZHANG Zhengyou. A Flexible New Technique for Camera Calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11):1330-1334.
- [18] LI Jiafeng, WEN Ying, HE Lianghua. SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, 2023: 6153-6162.
- [19] 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 Tongpu, XU Sixiang, SHI Yuxiang, 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.

(编辑 张洋)

作者简介:任加琪,女,2000年生,硕士研究生。研究方向为机器视觉。E-mail: 1837433005@qq.com。**许四祥***(通信作者),男,1974年生,教授,硕士研究生导师。研究方向为机器人、机器视觉。发表论文70余篇。E-mail: xsxhust@ahut.edu.cn。

本文引用格式:

任加琪,许四祥,董宾卉,等.基于轻量化高分辨率网络的双目视觉定位与测量[J].中国机械工程,2026,37(1):201-208.
REN Jiaqi, XU Sixiang, DONG Binhui, et al. Binocular Vision Localization and Measurement Based on Lightweight HRNet[J]. China Mechanical Engineering, 2026, 37(1): 201-208.