面向复杂环境的人机协作装配意图识别方法

何家威[1]，张朝阳[1,2]，叶子健[1]，史天佑[1]，郑坤明[3]
（1. 江南大学机械工程学院，214122，江苏无锡；2. 江苏省食品先进制造装备技术重点实验室，214122，江苏无锡；3. 江南大学智能制造学院，214122，江苏无锡）

摘要：针对复杂环境中人机协作装配系统存在的识别装配动作准确率较低、动作相似度高时无法准确感知装配者意图，以及由此导致的机器人配合操作者的装配效率较低的问题，从装配动作信息和装配零件信息之间的联系出发，建立了面向复杂环境的人机协作装配意图识别融合模型。基于骨架特征，采用双流自适应图卷积神经网络（2S-AGCN）模型针对装配进行动作识别；提出改进的YOLOV8模型，以提高装配零件在无序和遮挡环境下的识别准确率；结合当前装配任务，设计了包含装配动作和装配零件信息的装配推理规则，规划装配顺序。在装配动作数据集以及装配泵体零件数据集上对所提方法进行了验证。实验结果表明：交并比为0.50～0.95时，YOLOV8模型对零件识别的平均准确率达到88.361%；所提出的融合模型对于操作者装配意图的识别准确率达到96.66%，验证了人机协作系统识别操作者装配意图的可行性和有效性。将装配动作信息与装配零件信息相结合，有助于在复杂的人机协作环境中准确地识别出操作者的装配意图。

关键词：人机协作；动作识别；双流自适应图卷积神经网络；YOLOV8模型；意图识别融合模型
中图分类号：TH166 **文献标志码**：A
DOI：10.7652/xjtuxb202512004 **文章编号**：0253-987X(2025)12-0044-14

Human-Robot Collaborative Assembly Intention Recognition Method for Complex Environments

HE Jiawei[1], ZHANG Chaoyang[1,2], YE Zijian[1], SHI Tianyou[1], ZHENG Kunming[3]
(1. School of Mechanical Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China; 2. Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment and Technology, Wuxi, Jiangsu 214122, China; 3. School of Intelligent Manufacturing, Jiangnan University, Wuxi, Jiangsu 214122, China)

Abstract: To address the issues of low accuracy in recognizing assembly actions and the inability to accurately perceive human operators' intention under high action similarity in human-robot collaborative assembly systems within complex environments, as well as the resulting inefficiency in robot-operator collaboration, this study establishes an intention recognition fusion model for human-robot collaborative assembly in complex environments by integrating assembly action information and assembly part information. Based on skeletal features, a two-stream adaptive graph convolutional neural network (2S-AGCN) model is employed for action recognition in assembly tasks. An improved YOLOV8 model is proposed to enhance the recognition accuracy of assembly parts in disordered and occluded environments. Considering with the current assembly

收稿日期：2025-04-28。 作者简介：何家威(2001—)，男，硕士生；张朝阳(通信作者)，男，副教授，硕士生导师。 基金项目：国家自然科学基金资助项目(51805213,52205015)。
网络出版时间：2025-08-28 网络出版地址：https://link.cnki.net/urlid/61.1069.T.20250828.1449.006

task, assembly reasoning rules incorporating both assembly actions and part information are designed to plan the assembly sequence. The proposed method is validated on an assembly action dataset and an assembly pump part dataset. The experimental results show that when the intersection over union (I_{oU}) threshold ranges from 0.50 to 0.95, the improved YOLOV8 model achieves a mean average precision of 88.361% for part recognition. The proposed fusion model achieves an intention recognition accuracy of 96.66%, demonstrating the feasibility and effectiveness of the human-robot collaborative system in recognizing operator assembly intention. Integrating assembly action information with assembly part information contributes to accurately identifying operator assembly intention in complex human-robot collaborative environments.

Keywords: human-robot collaboration; action recognition; two-stream adaptive graph convolutional networks; YOLOV8 model; intention recognition fusion model

随着工业互联网、人工智能、大数据等技术在制造服务业中的发展，新一代信息通信技术与智能加工技术深度的融合，制造业正在由数字化、网络化等向智能化转型。在智能制造生产中，人-机-物的系统性联系越来越紧密，装配是制造生产中的关键工序，装配工艺质量是决定产品最终性能与可靠性的关键性因素。在自动化生产车间中，通常采用人机分离的模式，机器人通常被运用于执行相对简单、重复和繁重的装配任务，但不能处理复杂和高度灵活的装配任务。此外，在复杂的产品装配任务中，由于机器人几乎没有认知能力和灵活性，所以人工操作仍然必不可少。因此，人机协作（HRC）装配系统作为一种能够结合人和机器人优势的新型装配模式逐渐成为了热门的研究方向。目前，HRC 不应仅限于结构化传统工业环境中的工作单元，还应扩展到完整的开放和共享的制造场景。人类和机器人可以在非稳定环境中协作，参与个性化的小批量和多品种的制造活动，例如产品组装、拆卸等较为复杂的活动。

与传统制造系统相比，协作机器人的行为管理不是基于传统的预编程指令，而是通过视觉感知或触觉感知等方式识别操作者的意图，从而更好地配合操作者完成 HRC 装配工作。因此，在装配工作中，机器人是否可以准确地识别操作者的装配意图至关重要，只有准确识别操作者的装配意图才能使机器人配合操作者完成相关装配。

基于上述目标，国内外研究人员围绕人机协作状态感知和机器人路径规划等主题开展了大量研究，取得了丰富的理论成果和方法积累。戴汝为[1-2]提出了以人为中心的系统构建思想，系统的核心应该是人而不是机器设备，要突出强调人员在系统中的作用。

Matheson 等[3]指出人机协作在制造领域的应用主要包括装配、拆卸、质量控制和机器送料等方面。李梓响[4]研究了人机协同的装配线配平问题，提出了更为高效的编码解码方法，并通过群体智能算法实现生产规模最小化节拍优化。Xiao 等[5]利用 Kinect 深度传感器获取深度图像，结合工作空间其他信息对人体姿态进行识别。王政伟等[6]利用深度传感器获取人的实时骨骼特征与机器人自身的三维模型，并在此模型之上计算人机最小距离所对应的空间坐标。Vieira 等[7]提出了名为时空占用模式（STOP）的特征描述符，该特征描述符保留了空间和时间的上下文信息，从而灵活地适应动作变化。Xiao 等[8]提出了一种高效动态剩余注意机制的残差卷积网络的运动动作识别方法，解决了传统 3D 卷积神经网络在运动动作识别上识别率低的问题。姚冬安等[9]在深度卷积神经网络的基础上，提出了快速区域卷积神经网络（Faster R-CNN）模型对装配场景进行感知，通过该方法可以准确识别装配环境下人的作业意图。在 HRC 装配过程中，操作者的装配动作信息是获取装配意图的重要组成部分。Zhu 等[10]使用光流模型提取局部光流特征，并结合全局轮廓特征来识别人体动作。Urgo 等[11]基于人体姿态估计库（OpenPose）识别操作者的位置，然后基于隐马尔可夫模型构建监控方法来识别缺失的操作或不安全的行为。Berg 等[12]提出了一种基于隐马尔可夫模型的人机协作装配任务动作识别方法，通过动作识别使机器人意识到人当前正在执行的任务，以便能够适应人类的工作方式柔性调整。Al-amin等[13]提出了一种基于骨架数据的 CNN 分类器的个性化系统，用于识别操作者的装配动作，提高了异构工人的动作识别精度。Yan 等[14]首先提出了将图神经网络应用于基于行为的动作识别任

务，即时空图卷积神经网络（ST-GCN）。将空间和时间卷积结合到骨架行为识别任务中显示出出色的鲁棒性。Shi 等[15]设计了双流自适应图卷积网络（2S-AGCN），通过骨骼数据的二阶信息建模，并构建双流框架，可以同时对一阶和二阶信息进行建模，从而显著提高识别精度。Andrianakos 等[16]采用单激发多框探测器（SSD）算法识别装配部件和操作者的手，自动监控装配操作的执行情况。宋栓军等[17]将改进的 YOLOV3 算法应用于装配零件的判断，从而提高零件识别的成功率。Gao 等[18]基于 ST-GCN 混合一维卷积神经网路（1DCNN）提高识别装配动作准确率。Zhang 等[19]运用 ST-GCN 和 YOLOX 基于姿态识别和目标检测识别装配意图。Fan 等[20]基于视觉的自适应建立了人机协作数字孪生模型，使数字孪生模型运用于人机协作系统。Li 等[21]在基于深度强化学习的机器人运动生成中释放混合现实能力，实现了安全的人机协作。Zheng 等[22]基于视觉语言引导和深度强化学习提出了非结构化人机协同制造任务实现方法。Zhang 等[23]在人机协同装配中基于 RGB 的高度相似人体动作预测。Wang 等[24]面向以人为中心的智能制造提出了基于大型语言模型（LLM）的视觉和语言协作机器人导航的方法。黄思翰等[25]基于大语言模型和机器视觉的智能制造系统提出了人机自主协同作业方法，从而达到人机协作自动进行。

现有研究主要集中于在简单人机协作环境中识别单一的基本动作，这些方法在复杂动作识别上稳定性不足，且缺乏对操作者装配意图的深层感知。如何让机器人真正理解生产任务和人类意图，以实现自主协同配合，相关探索仍较为匮乏。为此，本文提出了一种面向复杂环境的人机协作系统装配意图识别方法。创新性地构建了一个融合框架，将基于骨架的装配动作识别与装配部件识别相结合，以期精准地感知操作者的协作意图。

1 HRC 中操作者装配意图识别框架

在人机协作装配过程中，机器人能够识别操作者的装配动作并且理解操作者的装配意图是前提。为了更好地理解操作者的装配动作和意图，使操作者和机器人协同完成装配任务，本文提出了 HRC 中操作者装配意图识别框架，如图 1 所示。该框架由 3 个模块组成。第 1 个模块是基于计算机视觉和深度感知技术，在装配过程中实现对物理车间中操作者的装配动作信息和装配零件信息的获取；第 2 个模块是基于 2S-AGCN 模型，提取骨骼特征识别操作者的装配动作；第 3 个模块是基于 YOLOV8 模型提出一种改进方法引入注意力机制卷积注意力模块（CBAM）进行装配零件的识别。最后，结合装配动作信息和装配零件信息，准确识别出操作者的装配意图。

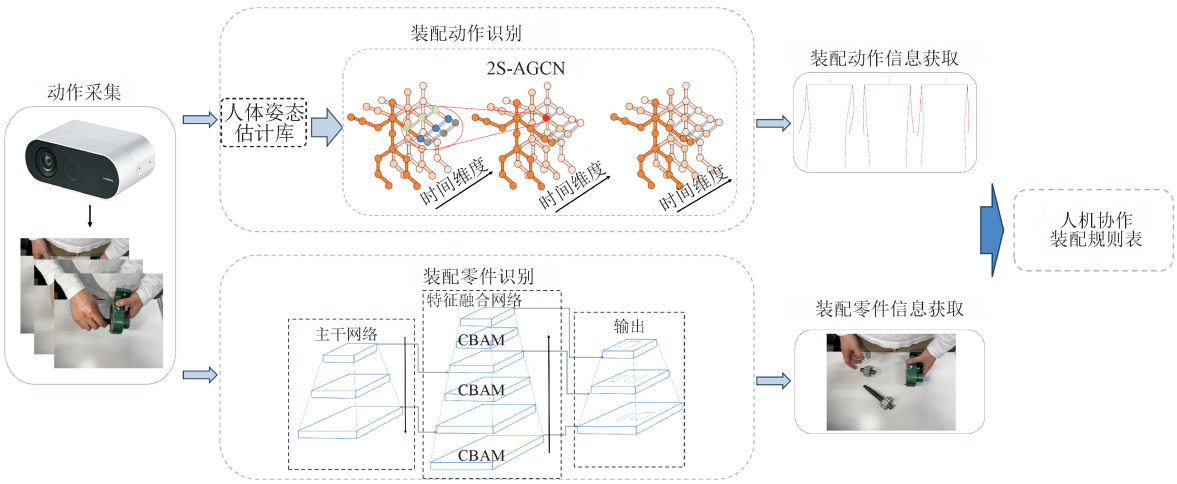


图 1 人机协作装配操作者意图识别框架

Fig. 1 Human machine collaborative assembly operator intention recognition framework

在该框架中利用 Orbbec Femto Bolt 深度相机拍摄装配动作信息以及装配零件信息，首先利用 OpenPose 提取装配动作中的 18 个骨骼点数据，并使用 2S-AGCN 模型对操作者的装配动作进行识别。其次，采用改进的 YOLOV8 模型对装配零件进行识别。最后，针对装配序列的灵活性，本文设计了一种基于装配步骤的推理方法，通过结合装配动作信息和装配零件信息识别操作者的装配意图。

2　装配动作识别 2S-AGCN 模型

2S-AGCN 双流自适应图卷积神经网络将骨骼数据的二阶信息（骨骼的长度和方向）利用起来，骨骼的二阶数据更具有信息量和判别性，使用了双流框架，可以同时对于一阶和二阶信息进行建模，从而显著提高识别精度。

2.1　双流自适应图卷积神经网络 2S-AGCN 模型

在人机协作装配中，操作者的动作识别需快速且准确。常见的人类行为识别模态包括 RGB 图像、深度图像、光流和身体骨架等。其中，基于骨架的方法因轻量化和速度快而更具优势。本文引入 2S-AGCN 模型识别装配动作。

在该模型中，图结构与卷积参数共同训练更新：全局图表征数据的共性模式，个体图则刻画样本的独特模式，并在不同层分别优化，从而增强了模型的灵活性和泛化性。一阶信息为关节的绝对坐标，二阶信息为关节间的相对关系向量，包含骨骼的长度与方向。模型采用双流框架融合两类信息，通过学习骨骼在时间维度上的运动轨迹，实现高效动作识别。2S-AGCN 模型的总体架构如图 2 所示。给定样本后，首先根据关节数据计算骨骼数据，随后将关节数据和骨骼数据分别输入 J 流（关节流）和 B 流（骨骼流），最后将两个流的 softmax 分类器分数相加得到融合结果，并预测动作标签。

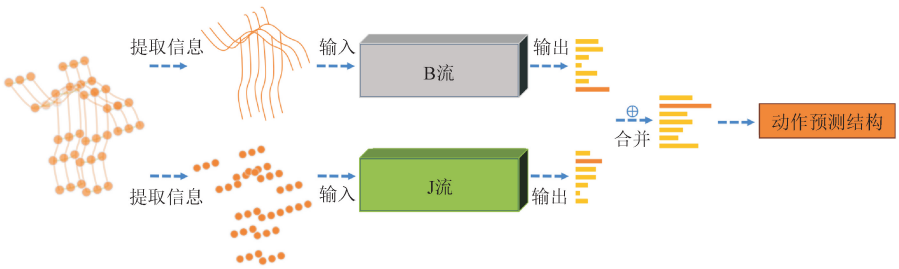


图 2　2S-AGCN 模型的总体架构示意图

Fig. 2　Schematic diagram of the overall architecture of 2S-AGCN

2.2　2S-AGCN 图卷积神经网络模型

2.2.1　双流自适应图卷积网络

自适应的图卷积层如图 3 所示，以端到端学习的方式，自适应的图卷积层将图的拓扑与网络的其他参数一起进行优化。对于不同的层和样本，图是唯一的，这大大增加了模型的灵活性。同时，它被设计成残差分支，保证了原始模型的稳定性。图卷积在空间维度上的网络特征图实际上是一个 $C\times T\times N$ 的张量，N 为顶点数，T 为时间长度，C 为通道数，$\boldsymbol{W}_k$ 为权重向量。具体来说，图卷积每层共有 3 种类型的图，即 $\boldsymbol{A}_k$、$\boldsymbol{B}_k$ 和 $\boldsymbol{C}_k$，下标 k 表示图卷积里不同邻接矩阵划分的索引，设 k 为 1、2、3 分别表示把邻接矩阵划分为自身、同侧、对侧。$\boldsymbol{A}_k$ 是归一化的 $N\times N$ 邻接矩阵，它代表了人体的物理结构；$\boldsymbol{B}_k$ 也是一个 $N\times N$ 的邻接矩阵，与 $\boldsymbol{A}_k$ 相反，$\boldsymbol{B}_k$ 的元素在训练过程中与其他参数一起被参数化和优化。$\boldsymbol{B}_k$ 的值没有任何约束，这意味着完全根据训练数据学习图。通过这种数据驱动的方法，模型不仅能够准确识别任务，还能针对不同层次所包含的信息更新个性化的图。然而，它无法生成原始物理图中不存在的新连接。从这个角度来看，$\boldsymbol{B}_k$ 比 $\boldsymbol{A}_k$ 更具灵活性。$\boldsymbol{C}_k$ 是一种数据依赖图，它为每个样本单独学习一个唯一的图。θ_k 和 φ_k 分别是 $\boldsymbol{C}_k$ 的参数。不是直接用 $\boldsymbol{B}_k$ 或 $\boldsymbol{C}_k$ 替换原来的 $\boldsymbol{A}_k$，而是将它们添加到 $\boldsymbol{A}_k$ 中。$\boldsymbol{B}_k$ 的值以及 θ_k 和 φ_k 的参数被初始化为 0，K_v 为空间维度上的卷积核的大小，K_v 设为 3。图 3 中橙色方框表示该参数是可

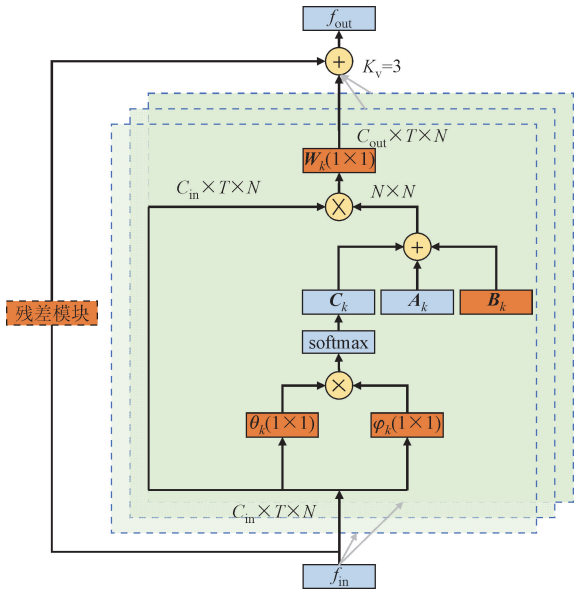


图 3　自适应图卷积层示意图

Fig. 3　Schematic diagram of adaptive convolution

学习的，1×1 表示卷积核的大小，$\oplus$ 表示逐元的求和，$\otimes$ 表示逐元素相乘操作。只有当入口通道数 C_{in} 与出口通道数 C_{out} 不同时，才需要使用残差模块。图的拓扑结构实际上是由邻接矩阵 $\boldsymbol{A}_k$ 和掩码 $\boldsymbol{M}_k$ 决定的。$\boldsymbol{A}_k$ 决定了两个顶点之间是否存在连接，$\boldsymbol{M}_k$ 决定了连接的强度。为了使图结构自适应，输出 f_{out} 的表达式如下

$$f_{out}=\sum_{k=1}^{K_v}\boldsymbol{W}_k f_{in}(\boldsymbol{A}_k+\boldsymbol{B}_k+\boldsymbol{C}_k) \tag{1}$$

这种设计在不降低原有性能的前提下，增强了模型的灵活性。为了判断两个顶点之间是否存在连接及两个顶点间的强度，采用归一化的高斯嵌入函数来计算顶点间的相似度。

2.2.2 自适应图卷积块 AGCN

在时间维度上，采用 $K_t\times1$ 的卷积核对 $C\times T\times N$ 的输入特征进行卷积，以捕捉动作在时间上的动态变化；在空间维度上，则基于骨架邻接矩阵进行卷积，卷积核大小为 $K_v\times3$。空间卷积和时间卷积之后依次接批归一化(BN)层与 ReLU 激活函数。自适应图卷积块如图 4 所示，一个基本块由空间卷积、时间卷积和一个丢弃率为 0.5 的 Dropout 层组成，并通过残差连接来稳定训练过程。

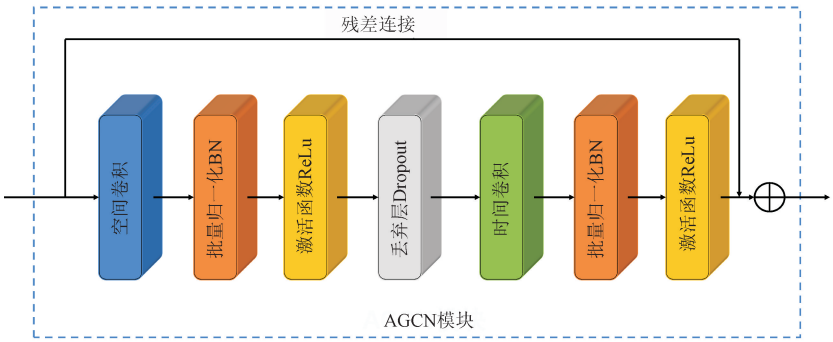


图 4 自适应图卷积块示意图

Fig. 4 Schematic diagram of adaptive plot volume block

双流自适应图卷积网络 2S-AGCN 的自适应图卷积模块(AGCN) 如图 5 所示，2S-AGCN 总共有 9 个块，每个区块的输出通道数分别为 64、64、64、128、128、128、256、256、256。开始时添加一个数据批标准化 BN 层来对输入数据进行规范化，然后输入到一个全局平均池化层，将不同样本的特征映射池化到相同大小。最后的输出被发送给 softmax 分类器函数来获得预测结果。

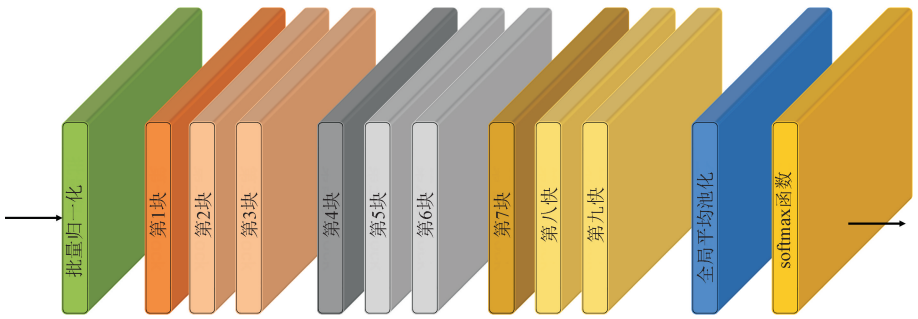


图 5 自适应图卷积网络 AGCN 模块

Fig. 5 Adaptive convolutional network AGCN block

3 改进的识别装配件 YOLOV8 模型

在 HRC 装配中，操作者经常与不同的装配部件进行交互，以完成复杂的产品装配任务。由于操作者的动作具有较高的相似性，同一动作可能涉及不同装配序列对应的不同装配部件，如果仅使用基于骨架的操作者装配动作识别结果，则 HRC 装配中操作者意图识别困难且精度低。

为提高 HRC 装配中操作者意图识别的准确性和有效性，应将操作者的装配动作识别与装配部件识别相结合。本文设计了嵌入 CBAM 的改进的 YOLOV8 模型(YOLOV8_CBAM)来识别装配部件。

注意力机制是指人类视觉对局部信息的选择性注意，可以关注关键信息，提高 YOLOV8 网络模型的计算性能。在加入 CBAM 之前，YOLOV8 模型对于装配零件的目标检测存在特征冗余、小目标检测能力有限等缺陷，因为 CBAM 是一种轻量级的空间注意力模块，通用性强，所以将 CBAM 模块嵌入到 YOLOV8 模型中以减少背景干扰，使模型能够更好地关注装配部件。

3.1　CBAM 注意力机制模块

注意力机制最初应用在机器翻译上取得了理想的效果，并逐渐应用在计算机视觉领域，CBAM 的主要思想是通过关注重要的特征并抑制不必要的特征来增强网络的表示能力。模块首先应用通道注意力关注重要的特征，然后应用空间注意力关注这些特征的重要位置。通过这种方式，CBAM 模块增强了模型对图像关键信息的关注，提高了特征的表示力度。CBAM 结合了通道注意力模块（CAM）和空间注意力模块（SAM），如图 6 所示。

CAM 和 SAM 的计算式如下

$$F' = M_c(F) \otimes F \tag{2}$$

$$F'' = M_s(F) \otimes F' \tag{3}$$

式中：F 为输入特征图；$M_c(F)$、$M_s(F)$ 分别为通道注意力与空间注意力权重；F' 为输入 F 的输出特征矩阵；F'' 为输入 F' 的输出特征矩阵。

CAM 中采用并行的平均池化层和最大池化层，每个池化层分别经过两个卷积模块，然后相加经过 sigmoid 激活函数得到注意力概率图。

对于 SAM 就是判断特征图中每个部分中像素的重要程度，需要综合每个部分所有通道的特征。

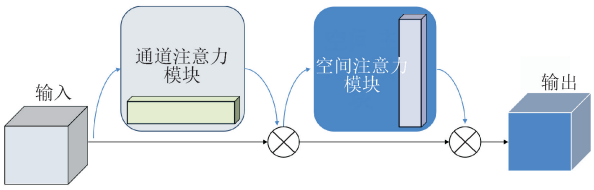


图 6　注意力机制 CBAM 模块的原理图

Fig. 6　Schematic diagram of attention mechanism CBAM module

3.2　改进的 YOLOV8 模型

嵌入 CBAM 的改进 YOLOV8 网络结构如图 7 所示，C2f 为跨阶段部分瓶颈卷积模块，C2f1 为第 1 个模块，SPPF 块为多尺度特征信息提取器模块。本文在特征融合网络引入 3 个 CBAM 模块。C2f6 输出后为 CBAM1，特征图有 256 维通道，C2f7 输出后为 CBAM2，特征图有 512 维通道，C2f8 输出后为 CBAM3，特征图有 1 024 维通道。一方面，嵌入 CBAM 的 YOLOV8 模型可以提高其对装配件的聚焦能力，另一方面，本文引入的 CBAM 根据 C2f 的输出通道数设置输入通道数，因此对推理速度影响不大。

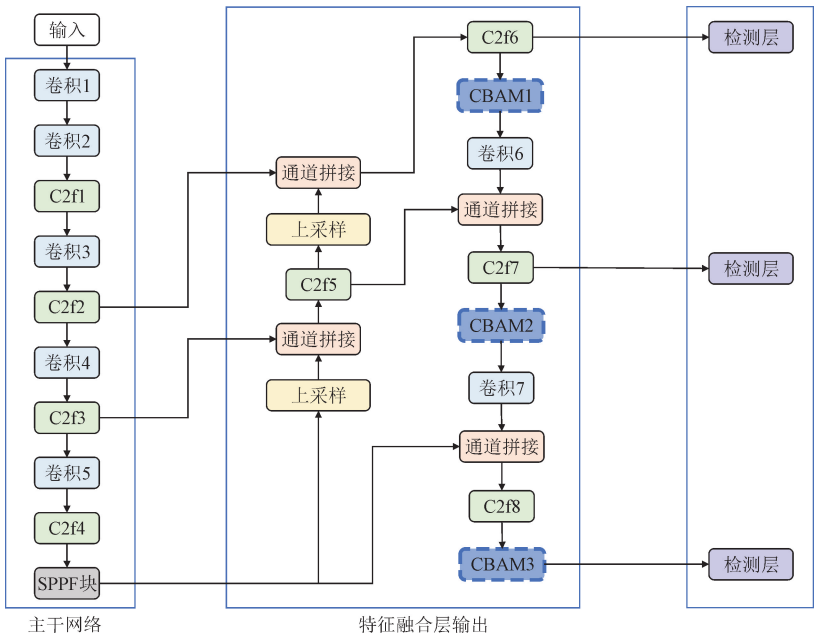


图 7　嵌入 CBAM 的 YOLOV8 模型

Fig. 7　YOLOV8 model embedded in CBAM

4 意图识别融合模型

由于单识别操作者的装配动作和装配零件无法准确识别操作者装配意图，但是动作识别和零件识别模型是不同模态的模型，因此在装配意图识别中，构建 2S-AGCN 的骨架时序特征与 YOLOV8_CBAM 的空间视觉特征的协同机制模型，如图 8 所示。该融合模型通过骨架时序特征与图像目标检测特征的互补信息，准确识别操作者的装配意图。

该协同识别模型的输入为操作者在装配场景中动作全过程的视频序列，提取输入视频的特征帧，I_f 为第 f 帧图像，f 为视频总帧数，得到特征图集合 Q，Q 的表达式如下

$$Q = \{I_1, I_2, \cdots, I_f\} \tag{4}$$

将视频帧输入人体姿态估计库进行骨架特征提取，得到人体动作特征 P_{pose}，表达式如下

$$P_{\text{pose}} = \{(x_l, y_l) \mid l = 1, \cdots, N\} \tag{5}$$

式中：l 表示人体的第 l 个关键点的编号。将骨架姿态数据 P_{pose} 输入 2S-AGCN 模型后，可以得到操作者的动作特征。将视频中所有帧的骨架特征按时间顺序拼接成时间序列张量，并输入 2S-AGCN 模型，最终得到动作特征序列 S_f，S_f 表示第 f 帧对应的动作特征，该特征序列作为协同识别模型的输入。

每隔 5 帧抽取 1 帧图像作 YOLOV8_CBAM 模型的输入，同时增加注意力机制关注操作者的手部零件，输入 YOLOV8_CBAM 模型得到装配零件的特征 Q_f，检测当前装配零件，再将结果编码 Q_f 时序作为协同识别模型的输入。

由于 2S-AGCN、YOLOV8 特征维度和时间尺度可能不完全匹配，利用对齐权重参数 W 进行尺度对齐。对齐后骨架动作特征 $\widetilde{S}_f$ 和零件视觉特征 $\widetilde{Q}_f$ 时间维度一致，表达式如下

$$\begin{cases} \widetilde{S}_f = WS_f \\ \widetilde{Q}_f = WQ_f \end{cases} \tag{6}$$

将 $\widetilde{S}_f$ 和 $\widetilde{Q}_f$ 拼接融合输入线形层后进行非线性变化，得到融合后的时序特征 F_f，表达式如下

$$F_f = \text{ReLU}(W[\widetilde{S}_f, \widetilde{Q}_f]) \tag{7}$$

将融合的时序特征输入 LSTM，进行时序上下文信息编码后得到输出特征 h_f，最后输入 softmax 分类器函数进行意图识别的分类，得到分类结果 $\hat{y}$，表达式如下

$$h_f = \text{LSTM}(F_f) \tag{8}$$

$$\hat{y} = \text{softmax}(Wh_f) \tag{9}$$

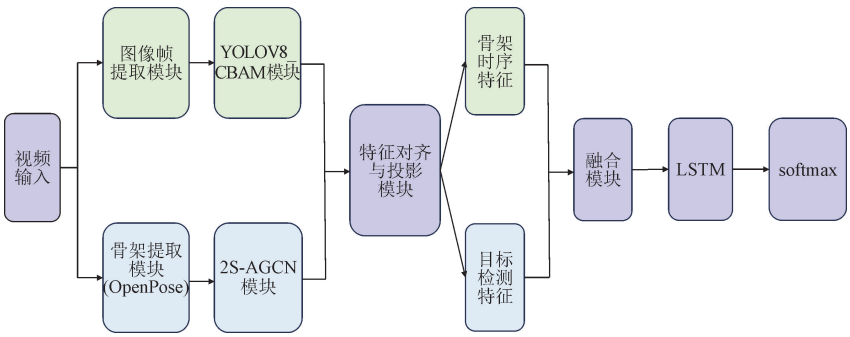


图 8 意图识别融合模型

Fig. 8 Intent recognition fusion model

5 案 例

本节以基于人机协作的泵体装配任务为例，建立使用 Femto Bolt 深度相机拍摄的装配动作和装配零件数据集，验证本文所提出的方法。实验均在 Win-windows11(64 位)系统下进行。CPU 为 inter i5-13500H，显卡为 NVIDIA RTX 4060 (8 GB)，2S-AGCN 模型和改进 YOLOV8 模型均基于 python3.8 语言和 PyTorch 2.4.1 深度学习框架构建。

5.1 装配零件识别

5.1.1 构建装配零件数据集

使用 Femto Bolt 深度相机，拍摄 5～6 s 左右的装配过程视频，然后利用计算机视觉库(opencv)进行抽帧处理，得到 745 幅(20%)P1 底座、740 幅(19.9%)P2 短齿轮轴、742 幅(19.9%)P3 长齿轮轴、730 幅(19.6%)P4 泵盖、769 幅(20.6%)P5 螺栓，总共 3 726 幅图片的数据集，泵体由这 5 个零件装配构成，如图 9(a)所示。使用目标检测标注工具 labelimg 对数据集进行标签化处理，对主要的零部件打标签，将标签的位置等信息做归一化处理并统一成文本格式，形成标签文件，标签的分布如图 9(b)所示。

采用精度(A_P)和平均精度(M_{AP})作为性能评价指标。本文采用 A_P 评价某类型装配件的识别效

果，表达式如下

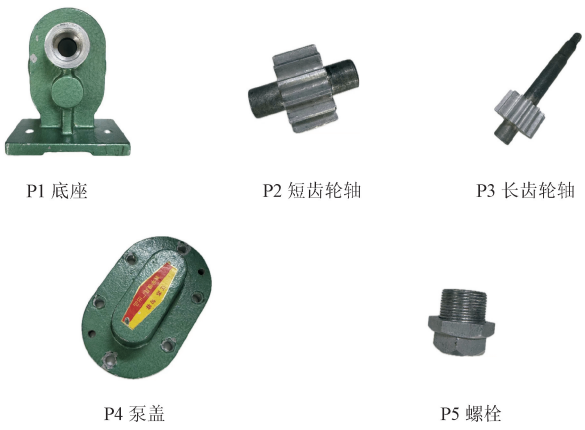
$$A_{\mathrm{P}}=\int_0^1 P(R)\mathrm{d}R \tag{10}$$

式中：P 为准确率；R 为召回率。M_{AP} 是 5 类装配件的 A_{P} 的平均值。召回率 R 以及准确率 P 的计算式如下

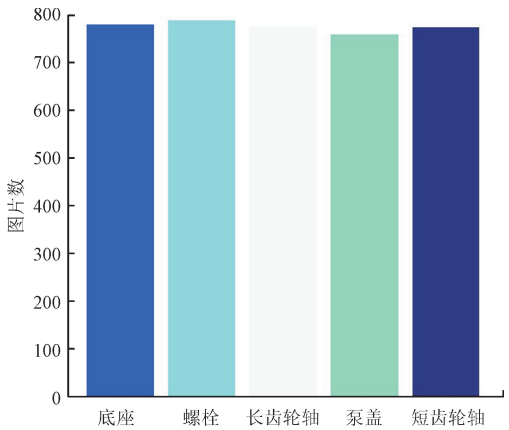
$$R=T_{\mathrm{P}}/(T_{\mathrm{P}}+F_{\mathrm{N}}) \tag{11}$$

$$P=T_{\mathrm{P}}/(T_{\mathrm{P}}+F_{\mathrm{P}}) \tag{12}$$

式中：T_{P} 代表模型将实际为正类别的样本正确预测为正类别；F_{N} 代表模型将实际为正类别的样本错误预测为负类别；F_{P} 代表模型将实际为负类别的样本错误预测为正类别。召回率 R 表示模型正确识别为正类的样本的数量占总的正类样本数的比例，值越高说明有更多的正类样本被模型预测正确，模型的效果越好。准确率 P 表示在模型识别为正类样本中，真正为正类的样本所占比例，值越高意味着模型预测为正类的可信度较高。



(a)5 个零件



(b)标签的分布

图 9　泵体零件数据集

Fig. 9　Pump body part data set

5.1.2　装配零件实验结果

本研究共采集 3 726 幅图像，其中 80%(2 980 幅)用于训练，10%(373 幅)用于验证，10%(373 幅)用于测试。为提升模型的泛化能力，对训练集进行了多种数据增强处理，以 33.3%的概率进行视角变化、遮挡(遮挡面积不超过图像的 30%)以及光线变化，亮度调整范围为 0.5～1.5，色温范围为 2 500～7 500 K，对图像施加 0.8～1.2 倍随机缩放，随机旋转±15°，角点偏移±10%，随机噪声强度为 0～0.3，模拟不同拍摄角度、光照条件、遮挡及传感器噪声等实际情况。经过上述增强处理，训练集最终扩展至 5 960 幅图像。

装配零件数据集原始 YOLOV8 模型和改进模型召回率、准确率的对比见表 1。由表 1 可知，YOLOV8_CBAM 模型召回率 R 为 98.719%，比原模型提高 3.444%，说明新模型能有效减少漏检，其准确率 P 为 98.303%，比原始模型提高了3.001%，意味着模型预测为正类的可信度较高，避免误判带来的负面影响。

表 1　装配零件数据集原始 YOLOV8 模型和改进模型召回率、准确率的对比

Table 1　Comparison of assembly parts dataset between original YOLOV8 and improved model recall/precision

模型	零件	召回率/%	准确率/%
YOLOV8	底座	95.369	94.671
	短齿轮轴	94.313	95.712
	长齿轮轴	97.425	98.116
	泵盖	96.162	96.135
	螺栓	93.106	91.876
	平均值	95.275	95.302
YOLOV8_CBAM	底座	98.632	98.406
	短齿轮轴	98.831	98.712
	长齿轮轴	99.621	99.317
	泵盖	99.217	98.519
	螺栓	99.312	96.561
	平均值	98.719	98.303

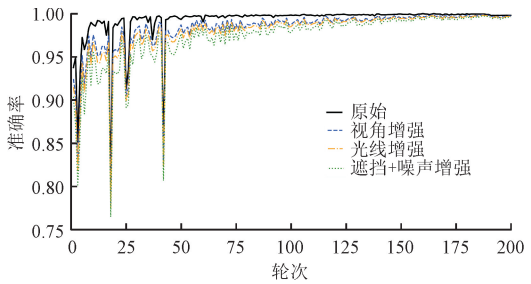
装配零件数据集原始 YOLOV8 模型和改进 YOLOV8 模型的精度、平均精度的对比见表 2。由表 2 可知，YOLOV8_CBAM 模型装配零件识别算法在交并比 I_{oU} 为 0.50 时的 M_{AP} 达到 97.38%，比原 YOLOV8 装配零件识别算法提高了 1.01%。当 I_{oU} 为 0.50～0.95 时，改进 YOLOV8 算法的 M_{AP} 达到

88.361%，比原 YOLOV8 算法提高了 4.83%。除了长齿轮轴零件的 A_P 为 100.00%，其他装配零件的 A_P 均有不同程度的提高。利用数据增广后的训练集训练得到的结果如图 10 所示，YOLOV8_CBAM 模型在不同数据增强策略的情况下，I_{oU} 为 0.50 时，M_{AP} 达到 96.34%，I_{oU} 为 0.50 ～ 0.95 时，M_{AP} 达到 87.735%，同时 R 为 97.435%，P 为 97.541%，说明该模型具有较好的跨场景鲁棒性。

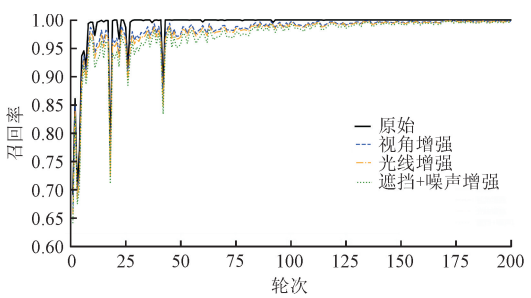
表 2 装配零件数据集原始 YOLOV8 模型和改进 YOLOV8 模型的精度、平均精度的对比

Table 2 Comparison between original and improved YOLOV8 of assembly part data set

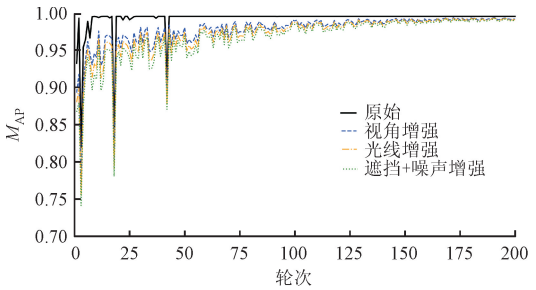
模型	零件	A_P/%
YOLOV8	底座	97.04
	短齿轮轴	95.45
	长齿轮轴	100.00
	泵盖	98.73
	螺栓	90.63
	$M_{AP}(I_{oU}=0.50)$	96.37
	$M_{AP}(I_{oU}=0.50\sim0.95)$	83.531
YOLOV8_CBAM	底座	97.24
	短齿轮轴	96.39
	长齿轮轴	100.00
	泵盖	99.50
	螺栓	93.77
	$M_{AP}(I_{oU}=0.50)$	97.38
	$M_{AP}(I_{oU}=0.50\sim0.95)$	88.361



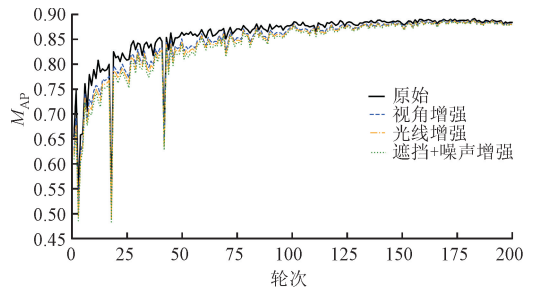
(a)准确率



(b)召回率



(c) $M_{AP}(I_{oU}=0.50)$



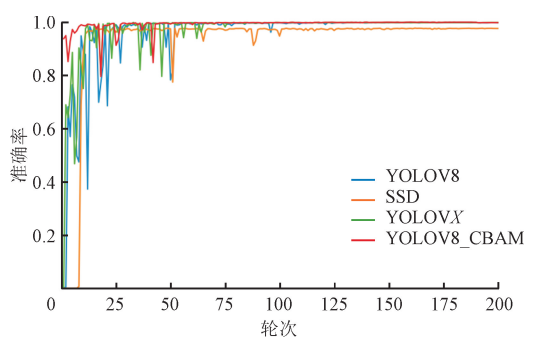
(d) $M_{AP}(I_{oU}=0.50\sim0.95)$

图 10 装配零件数据集在原始视角和数据增强情况下的比较

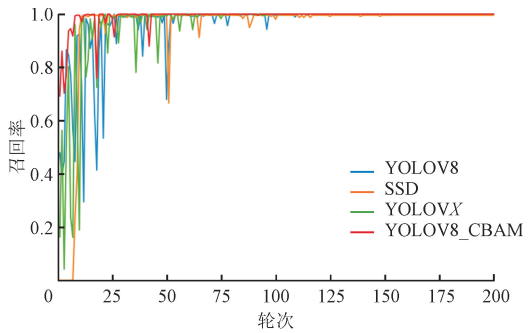
Fig. 10 Comparison curve of assembly parts dataset under original perspective and data augmentation conditions

由于采用了改进策略，YOLOV8_CBAM 模型在较小零件螺栓的识别中取得了较高的召回率 99.312%和准确率 96.561%，这表明改进的模型可以增强对相似或无序部件的识别效果。改进后的算法还在一定程度上增强了关键信息的识别性能，说明改进对于小尺寸零件识别是有效的。综上所述，引入 CBAM 模块可以使 YOLOV8 模型更加关注小尺寸装配部件，提高装配系统性能。

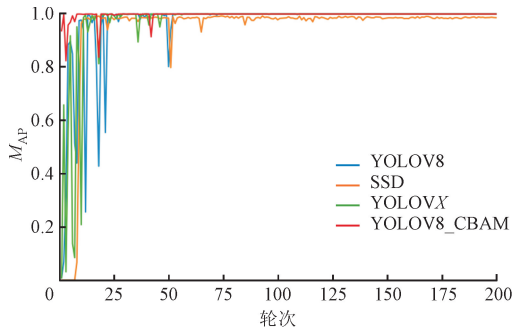
为了更好地保证 YOLOV8_CBAM 模型提升识别准确率，本文还验证了 SSD、YOLOVX 的对比性能，如图 11 所示。



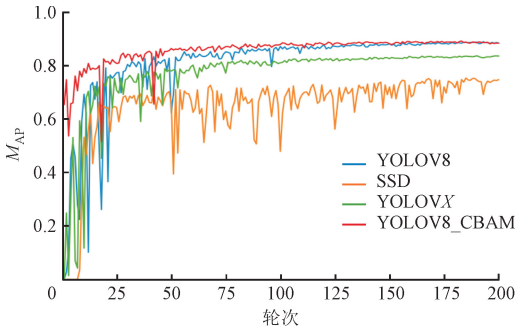
(a)准确率



(b)召回率



(c) M_{AP}(I_{oU} = 0.50)



(d) M_{AP}(I_{oU} = 0.50 ~ 0.95)

图 11　装配零件数据集在 SSD、YOLOV*X*、YOLOV8 和 YOLOV8_CBAM 模型下的比较

Fig. 11　Comparison curve of assembly part data set in SSD, YOLOV*X*, YOLOV8 and YOLOV8_CBAM

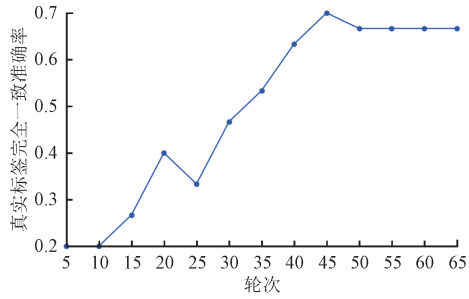
5.2　装配动作识别

5.2.1　构建装配动作数据集

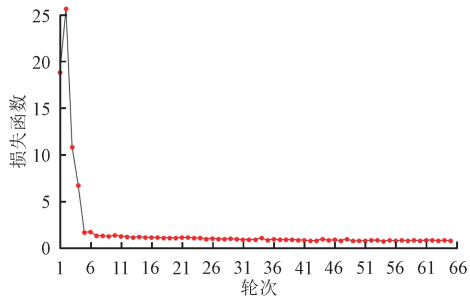
使用 Femto Bolt 深度相机，对操作者的装配动作进行拍摄。装配操作按 6 个方向收集，即左前、左上、正前方、前上、右前和右上，记录 5 个操作者的装配动作。RGB-D 视频包括深度模式(640×576 分辨率)和色彩模式(1 280×720 分辨率)，共采集 400 个视频片段，每个视频为 5～6 s，生成装配动作数据集，利用 OpenPose 进行逐帧处理，在空间对齐上，对每一帧生成一个 JSON 文件，保存检测到的人体关节点坐标。在时间对齐上，对所有视频统一处理成 30 帧/s，视频的每一帧和骨架帧是天然对齐的，并且数据集遵循 Kinetics 数据集的格式。在数据集中一共拥有 5 个类别的装配动作，包括 A1 底座装配、A2 短齿轮轴装配、A3 长齿轮轴装配、A4 泵盖装配、A5 螺栓装配对装配动作，使用 OpenPose 进行骨骼点提取后，输入 2S-AGCN 模型中进行训练。

5.2.2　装配动作实验结果

随机选择 320 个视频片段作为训练数据集，80 个视频片段作为测试数据集。图 12(a)和 12(b)展示了装配动作数据集上 2S-AGCN 模型训练的真实标签完全一致准确率和损失函数。



(a)真实标签安全一致准确率



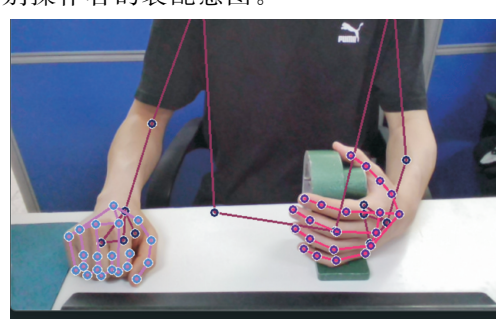
(b)损失函数

图 12　2S-AGCN 模型上真实标签完全一致准确率和损失函数

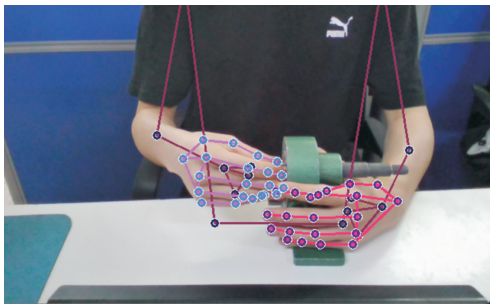
Fig. 12　True label consistency accuracy and loss function curve on 2S-AGCN model

实验结果表明，所建立的 2S-AGCN 模型能够较准确地识别部分泵体的装配动作。然而，对于一些相似的装配动作，在推断关节位置时存在较大误差，从而导致装配动作识别错误。

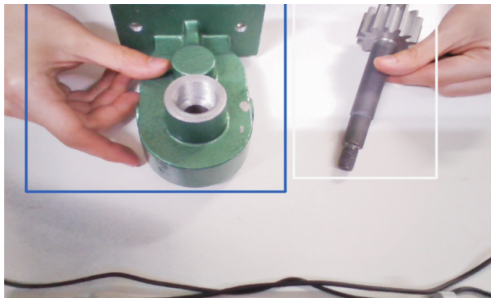
图 13 展示了泵体装配动作识别的部分捕获画面。在识别泵体装配动作时，结合装配零件的识别信息，可以对装配动作进行更准确地划分，从而更好地识别操作者的装配意图。



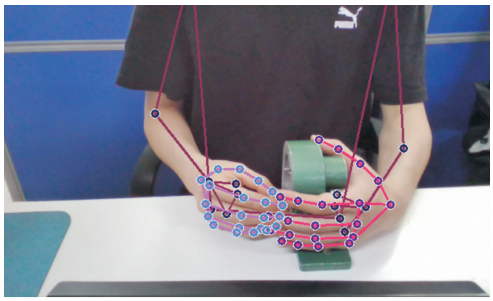
(a)底座装配动作识别



(b)长齿轮轴装配动作识别



(c)底座装配零件识别



(d)短齿轮轴装配动作识别

图 13 装配动作和零件识别的部分图片

Fig. 13 Partial pictures of assembly action and parts identification

5.3 操作者的装配意图识别

5.3.1 消融实验

在案例研究中,共设定 5 种装配动作和 5 个装配零件。考虑到装配过程的灵活性以及部分装配动作的相似性,装配动作具有一定的可选择性。例如,在泵体装配过程中,底座装配完成后可以选择装配长齿轮轴或短齿轮轴,而这两类动作在操作上较为相似,难以识别。因此,本文设计了融合模型与单一 2S-AGCN 模型的消融实验,验证融合模型的准确性。利用 2S-AGCN 模型识别装配者操作意图,显示了测试数据集中 5 个装配动作识别任务的样本数以及使用样本计算出的准确率,如表 3 所示,本研究中,每个装配动作包括 10 个测试样本。5 个装配动作中,关键装配动作的识别准确率较高,底座装配动作的识别准确率为 95.55%,螺栓装配动作的识别准确率为 95.53%。短齿轮轴装配和长齿轮轴装配动作仅在底座上移动距离不同,整体相似度高,导致短齿轮轴和长齿轮轴装配动作的识别准确率均较低,为 43.33%。在泵盖和底座装配的过程中,右手的高度不同,左手在结束时达到不同的高度和水平位置。除手腕关节外,其他关节节点的位置基本不变,导致泵盖装配动作识别率为 53.56%,被认定为长或短齿轮轴。由此可以知道,单一 2S-AGCN 模型识别装配动作的精度不高,需要结合装配部件的识别。

表 3 2S-AGCN 模型对 5 种装配意图识别的准确率

Table 3 Accuracy of 2S-AGCN model in recognizing 5 types of assembly intentions

装配意图	准确率/%
底座装配	95.55
短齿轮轴装配	43.33
长齿轮轴装配	43.33
泵盖装配	55.56
螺栓装配	95.53
平均值	66.66

本文利用意图识别融合模型对装配意图进行识别,准确率如表 4 所示,之前由于相似度太高,2S-AGCN 识别不出的动作,结合装配零件信息后融合模型可以准确识别操作者装配时手中的零件,同时加上对装配动作的大致推测,融合模型准确推测操作者装配意图,使机器人可以合理的配合操作者,两种模型准确率的对比如图 14 所示。由图 14 中可以看出,对比单一 2S-AGCN 模型,融合模型的每一个迭代都有提升,融合模型的准确率有所提升。

表 4 融合模型对 5 种装配意图识别的准确率

Table 4 Accuracy of fusion model in recognizing five types of assembly intentions

装配意图	准确率/%
底座装配	97.73
短齿轮轴装配	96.69
长齿轮轴装配	93.33
泵盖装配	96.69
螺栓装配	98.86
平均值	96.66

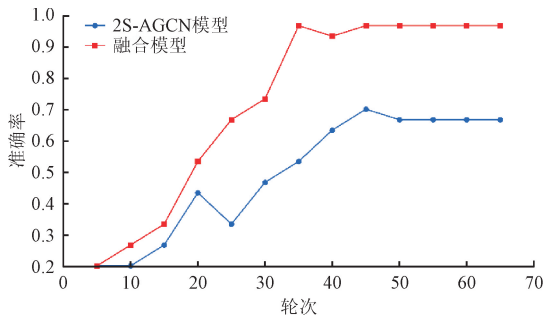


图 14　两种模型准确率对比

Fig. 14　Comparison of model accuracy

5.3.2　规则推理方法

结合当前装配任务的装配动作和装配零件，设计了一种包含装配动作和装配零件信息的规则推理方法，对于操作者的装配意图进行识别。上述装配动作数据集和装配零件数据集已经对其数据进行了编号，例如 A1 为底座装配、P1 为底座。基于装配动作和零件数据集设计出机器人的 5 种抓取动作：G1 为抓取底座；G2 为抓取短齿轮轴；G3 为抓取长齿轮轴；G4 为抓取泵盖；G5 为抓取螺栓。定义了 5 个装配动作及其对应的 5 个零件：A1 为底座装配；A2 为短齿轮轴装配；A3 为长齿轮轴装配；A4 为泵盖装配；A5 为螺栓装配；P1 为底座；P2 为短齿轮轴；P3 为长齿轮轴；P4 为泵盖；P5 为螺栓。

本文中将装配动作和装配零件的识别结合起来，准确识别操作者的装配行为意图，并推断出已经完成的装配任务和接下来要执行的装配任务。基于逻辑的操作者装配意图识别规则如图 15 所示，当摄像头捕捉后，机器人识别到手中没有零件判定为空手，此时机器人执行 G1 抓取，这时手中有 P1 并且装配者执行第一个装配动作 A1，摄像头进行捕捉后，机器人识别为“A1&P1”后进行 G2 和 G3 抓取，装配者可以进行选择后执行第 2 个装配动作，以此类推，机器人识别装配者的装配动作，当机器人识别到装配者手中有 5 个零件之后执行结束指令。与传统的完全由操作者执行装配的模式相比，人机协作装配系统在运行过程中显著减轻了操作者的认知负荷，同时提高了装配效率。

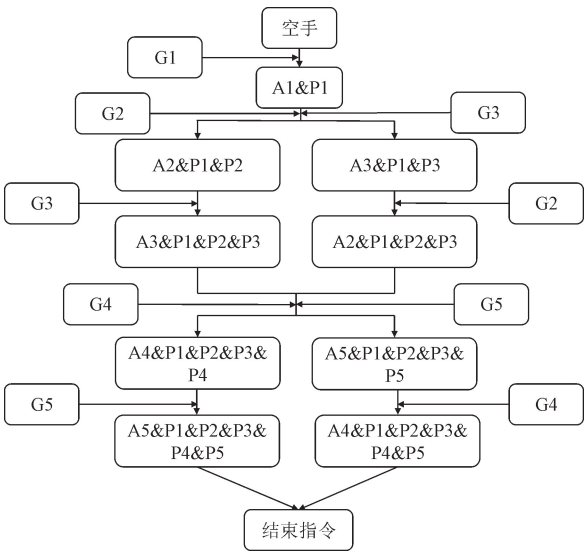


图 15　基于逻辑的操作者装配意图识别规则

Fig. 15　Rules for identifying the assembly intention of operators based on logic

6　结　论

在工业互联网和工业 4.0 的阶段背景下智能制造系统的人机协作已经成为研究热点。在人机协作装配中，如何准确识别操作者的装配意图是机器人在装配过程中面临的一个关键问题。本文提出一种面向复杂环境的人机协作装配意图识别方法，通过结合装配动作信息和装配零件信息提高了操作者装配意图的准确性。一方面，基于采集的装配动作数据集，使用 2S-AGCN 模型识别装配动作。另一方面，基于采集的装配零件数据集，设计了改进的 YOLOV8 模型，引入 CBAM 模块来提高模型对于装配零件的聚焦能力，提高识别装配零件的准确率。最后结合对装配动作和装配零件的识别结果，设计一种基于逻辑的操作者装配意图识别规则，结果表明，本文所提方法能够为人机协作系统有效地识别操作者的装配意图。

未来的研究方向包括进一步优化人机协作装配意图决策模型，探索大模型在意图识别与决策中的作用，以提升决策准确性和实时性，克服规则方法局限于当前任务的缺点。同时，可研究更高精度的人体姿态识别与目标检测算法，增强机器人对操作者动作的感知能力；结合强化学习，使机器人具备自适应决策能力。除此之外，还需突破现有人机协作系统仅适用于单人-单机模式的限制，面向车间生产场景构建适用于多人-多机器人的协作系统。

参考文献：

[1]　戴汝为. 人-机结合的智能科学和智能工程 [J]. 中国工程科学，2004(5)：24-28.
DAI Ruwei. Man-computer cooperative intelligent science and intelligent technology [J]. Strategic Study of CAE, 2004(5): 24-28.

[2]　戴汝为. 人-机结合的智能工程系统：处理开放的复杂巨系统的可操作平台 [J]. 模式识别与人工智能，2004，17(3)：257-261.

DAI Ruwei. Man-computer cooperated intelligent engineering system: a management platform for dealing with OCGS [J]. Pattern Recognition and Artificial Intelligence, 2004, 17(3): 257-261.

[3] MATHESON E, MINTO R, ZAMPIERII E G G, et al. Human-robot collaboration in manufacturing applications: a review [J]. Robotics, 2019, 8(4): 100-124.

[4] 李梓响. 人机协同双边装配线平衡建模及智能算法研究 [D]. 武汉: 武汉科技大学, 2018: 1-10.

[5] ZHENG Xiao, FU Mengyin, YANG Yi, et al. 3D human postures recognition using kinect [C]//2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics. Piscataway, NJ, USA: IEEE, 2012: 344-347.

[6] 王政伟, 甘亚辉, 戴先中. 协作环境下的人机距离建模与最小距离迭代算法 [J]. 机器人, 2018, 40(4): 413-422.
WANG Zhengwei, GAN Yahui, DAI Xianzhong. Human-robot distance modeling and minimum distance iterative algorithm in collaboration environment [J]. Robot, 2018, 40(4): 413-422.

[7] VIEIRA A W, NASCIMENTO E R, OLIVEIRA G L, et al. Stop: space-time occupancy patterns for 3d action recognition from depth map sequences [C]// Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Berlin, Germany: Springer, 2012: 252-259.

[8] XIAO Jiongen, TIAN Wenchun, DING Liping. Basketball action recognition method of deep neural network based on dynamic residual attention mechanism [J]. Information, 2022, 14(1): 13-27.

[9] 姚冬安, 徐文君, 姚碧涛,等. 基于上下文感知与图注意力网络的人机协作装配人员作业意图识别方法 [J]. 计算机集成制造系统, 2024, 30(6): 2005-2013.
YAO Dongan, XU Wenjun, YAO Bitao, et al. Human intention recognition method based on context awareness and graph attention network for human-robot collaborative assembly [J]. Computer Integrated Manufacturing Systems, 2024, 30(6): 2005-2013.

[10] ZHU Shaoping, XIA Limin. Human action recognition based on fusion features extraction of adaptive background subtraction and optical flow model [J]. Mathematical Problems in Engineering, 2015, 2015(1): 387464.

[11] URGO M, TARABINI M, TOLIO T. A human modelling and monitoring approach to support the execution of manufacturing operations [J]. CIRP Annals, 2019, 68(1): 5-8.

[12] BERG J, RECKORDT T, RICHTER C, et al. Action recognition in assembly for human-robot-cooperation using hidden Markov models [J]. Procedia CIRP, 2018, 76: 205-210.

[13] Al-AMIN M, QIN R, MONIRUZZAMAN M, et al. An individualized system of skeletal data-based CNN classifiers for action recognition in manufacturing assembly [J]. Journal of Intelligent Manufacturing, 2023, 34(2): 633-649.

[14] YAN Sijie, XIONG Yuanjun, LIN Dahua. Spatial temporal graph convolutional networks for skeleton-based action recognition [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto, C A, USA: AAAI, 2018: 7444-7452.

[15] SHI Lei, ZHANG Yifan, CHENG Jian, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2019: 12018-12027.

[16] ANDRIANAKOS G, DIMITROPOULOS N, MICHALOS G, et al. An approach for monitoring the execution of human based assembly operations using machine learning [J]. Procedia Cirp, 2019, 86: 198-203.

[17] 宋栓军, 侯中原, 王启宇, 等. 改进 YOLOV3 算法在零件识别中的应用 [J]. 机械科学与技术, 2022, 41(10): 1608-1614.
SONG Shuanjun, HOU Zhongyuan, WANG Qiyu, et al. Application of improved YOLOV3 algorithm in part recognition [J]. Mechanical Science and Technology for Aerospace Engineering, 2022, 41(10): 1608-1614.

[18] GAO Zenggui, YANG Ruining, ZHAO Kai, et al. Hybrid convolutional neural network approaches for recognizing collaborative actions in human-robot assembly tasks [J]. Sustainability, 2024, 16(1): 139-146.

[19] ZHANG Yaqian, DING Kai, HUI Jizhuang, et al. Human-object integrated assembly intention recognition for context-aware human-robot collaborative assembly [J]. Advanced Engineering Informatics, 2022, 54: 101792.

[20] FAN Junming, ZHENG Pai, LEE C K M. A vision-based human digital twin modeling approach for adaptive human-robot collaboration [J]. Journal of Manufacturing Science and Engineering, 2023, 145(12): 121002.

[21] LI Chengxi, ZHENG Pai, ZHOU Peng, et al. Unleashing mixed-reality capability in deep reinforcement

learning-based robot motion generation towards safe human-robot collaboration [J]. Journal of Manufacturing Systems, 2024, 74: 411-421.

[22] ZHENG Pai, LI Chengxi, FAN Junming, et al. A vision-language-guided and deep reinforcement learning-enabled approach for unstructured human-robot collaborative manufacturing task fulfilment [J]. CIRP Annals, 2024, 73(1): 341-344.

[23] ZHANG Yaqian, DING Kai, HUI Jizhuang, et al. Skeleton-RGB integrated highly similar human action prediction in human-robot collaborative assembly [J]. Robotics and Computer-Integrated Manufacturing, 2024, 86: 102659.

[24] WANG Tian, FAN Junming, ZHENG Pai. An LLM-based vision and language cobot navigation approach for human-centric smart manufacturing [J]. Journal of Manufacturing Systems, 2024, 75: 299-305.

[25] 黄思翰，陈建鹏，徐哲，等. 基于大语言模型和机器视觉的智能制造系统人机自主协同作业方法 [J]. 机械工程学报，2025，61(3)：130-141.
HUANG Sihan, CHEN Jianpeng, XU Zhe, et al. Human-robot autonomous collaboration method of smart manufacturing system based on large language model and machine vision [J]. Journal of Mechanical Engineering, 2025, 61(3): 130-141.

（编辑　武红江）