◀油气人工智能▶

基于定位姿态估计的修井机系统轨迹识别与工况分析*

王利君[1] 张德君[1] 李 赓[2] 张 汇[1] 普宏槟[2] 传 平[1]
（1. 中国石油新疆油田分公司数智技术公司 2. 中国石油新疆油田分公司工程技术部）

王利君，张德君，李赓，等. 基于定位姿态估计的修井机系统轨迹识别与工况分析［J］. 石油机械，2026，54（10）：1-11.

Wang Lijun，Zhang Dejun，Li Geng，et al. Trajectory identification and working condition analysis of workover rig system based on pose estimation［J］. China Petroleum Machinery，2026，54（10）：1-11.

摘要：在油气增储上产及高效开发需求驱动下，油气田修井作业现场亟须提升智能化监测水平。传统人工与半自动化作业模式下应用的计算机视觉监测技术难以实现修井机关键构件高精度实时定位，轨迹跟踪与工况分析相互独立，大钩摆动偏移、载荷异常等安全风险难以自动辨识。针对该工程难题，提出了一套集成快速姿态估计与精细轨迹识别和工况分析的系统框架，利用基于姿态引导的非局部注意力机制改进姿态估计YOLO-Pose方法，设计全局粗定位与局部特征细化相结合的分层局部化姿态估计策略，该方法面向游车大钩结构定制5个专用关键点，并配套井口单点标注方案，同步完成井架、大钩双目标联合检测，依托关键点像素距离建立图像与真实物理空间尺度换算关系；引入姿态引导轻量化通道注意力机制，以设备实时姿态信息驱动特征权重动态分配，抑制井场遮挡、光影带来的无效噪声，实现轨迹跟踪与工况分析深度耦合；基于连续帧关键点坐标建立空间坐标转换模型，结合牛顿动力学模型求解大钩位移、速度、加速度等时序运动参数，构建大钩-井口垂向夹角量化判定准则，完成动态受力解析与摆动危险工况自动预警。研究结果表明：模型经300轮迭代训练后趋于稳定收敛，目标与关键点检测精度达99.91%、召回率100%，边界框与关键点$P_{A@0.5}$均为99.5%，关键点平均精度均值$P_{mA@0.5:0.95}$可达98.21%；现场实测可完整复现大钩起下管运动轨迹，运动参数曲线可清晰区分启停、匀速、静置等作业阶段，设定5°安全阈值能够精准识别大钩异常摆动状态。该方法可在复杂井场环境下实现修井机关键部件高精度定位、连续轨迹稳定跟踪与工况量化诊断，提前预判作业安全隐患。研究结果可为复杂井场作业提供智能化解决方案。

关键词：井场工况；YOLO-Pose；轨迹识别；修井机系统；注意力机制；姿态估计

中图分类号：TE928/319 **文献标识码：**A DOI：10.12473/CPM.202507026

Trajectory Identification and Working Condition Analysis of Workover Rig System Based on Pose Estimation

Wang Lijun[1] Zhang Dejun[1] Li Geng[2] Zhang Hui[1] Pu Hongbin[2] Chuan Ping[1]
（1. *Digital and Intelligent Technology Company*, *PetroChina Xinjiang Oilfield Company*; 2. *Engineering Technology Department*, *PetroChina Xinjiang Oilfield Company*）

Abstract: Driven by the demands for increased oil and gas reserves/production and efficient development,

* 基金项目：新疆维吾尔自治区“天山英才”优秀工程师项目“新疆油田生产指挥中心系统建设”（2022TSYCGC0002）。

© Editorial Office of *China Petroleum Machinery*. OA under CC BY-NC-ND 4.0

there is an urgent need to enhance the intelligent monitoring level at field workover sites. Traditional computer-vision-based monitoring techniques, mainly applied in manual and semi-automated operation modes, are unable to achieve high-precision real-time positioning of key components of workover rigs. Trajectory tracking and working condition analysis are performed independently, and safety risks such as hook swing deviation and load anomalies are difficult to identify automatically. To address this engineering challenge, a systematic framework integrating rapid pose estimation, fine trajectory identification, and working condition analysis was proposed. The YOLO-Pose method for pose estimation was improved using a pose-guided non-local attention mechanism. A hierarchical localized pose estimation strategy combining global rough positioning and local feature refinement was designed, with five keypoints customized for the block-hook assembly. In conjunction with the wellhead single-point annotation scheme, the derrick and hook were detected synchronously. Based on the pixel distance of keypoints, the conversion relationship between image scale and real physical space scale was established. A pose-guided lightweight channel attention mechanism was introduced to dynamically allocate feature weights driven by real-time equipment pose information, suppressing invalid noise caused by wellsite occlusion and lighting variations, and achieving deep coupling of trajectory tracking and working condition analysis. A spatial coordinate conversion model was constructed based on continuous-frame keypoint coordinates, and combined with the Newtonian dynamics model, the time-series kinematic parameters of hook displacement, velocity, and acceleration were solved. A quantification criterion for the vertical angle between the hook and the wellhead was defined. Ultimately, dynamic force analysis and automatic warning of dangerous swing conditions were completed. The results show that the model tends to converge stably after 300 training epochs, achieving a detection accuracy of 99.91% and a recall of 100% for targets and keypoints, with both bounding-box and keypoint $P_{\mathrm{A@0.5}}$ reaching 99.5%, and keypoint $P_{\mathrm{mA@0.5:0.95}}$ reaching 98.21%. Field measurements can fully reproduce the hook motion trajectory during pipe running and pulling, and the kinematic parameter curves can clearly reflect operational stages such as start-stop, uniform motion, and idling. A safety threshold of 5° enables accurate identification of abnormal hook swing states. This method can achieve high-precision positioning of key components of workover rig, stable tracking of continuous trajectory, and quantitative diagnosis of working condition in complex wellsite environments, providing early warning of operational safety hazards. The research offers an intelligent solution for complex wellsite operations.

Keywords: wellsite working condition; YOLO-Pose; trajectory identification; workover rig system; attention mechanism; pose estimation

0　引　言

近年来，传统井场作业模式逐渐显现出效率低下、精度欠缺及存在安全隐患等诸多局限性，基于机器学习的智能化技术逐渐成为油气田开发领域的研究热点[1]。基于卷积神经网络（convolutional neural networks，CNN）的模型在井场视频监控数据处理方面展现出卓越的性能，为修井机智能化系统的研发提供了技术基础[2]。如张菲菲等[3]通过强化数据集成，构建融合模型和开发自适应方法，系统分析了关键技术与挑战。深度神经网络在特征提取、非线性建模及大规模数据处理方面优势显著，为井场作业挑战提供可行方案。姿态估计技术在多个工业领域均展现出了广泛的应用前景与价值[4]。在油气田作业场景下，姿态估计技术可为关键设备的精准定位与运动轨迹分析提供有力支持[5]。但如何利用视频监控系统提供的现场作业数据，通过计算机视觉技术进行视频数据分析，实现对监控视频的智能化高效利用，需要进一步深入地研究。此外，将姿态估计、轨迹识别和工况分析技术相结合，可提升作业精度与效率，增强系统对复杂工况的适应性和响应能力。

为此，针对油气田开发中的修井机智能化运行提出以下技术与方法：①高效局部化姿态估计技术，实现修井机在复杂井场环境中的高效定位；②姿态引导下的轨迹识别与工况深度耦合分析，实现动态实时分析与精准预测；③基于姿态估计的修井机大钩工况分析，通过实时监测大钩的空间位置、姿态和受力情况，实现大钩摆动的自动判定和潜在故障风险预测。这些研究旨在为复杂

井场作业提供智能化解决方案，推动行业效率提升和安全发展。

1　快速精细定位的姿态估计技术

1.1　自下而上的姿态估计

姿态估计技术从图像或视频中识别目标关键点位置，获取运动信息。在油气田开发智能监测中，该技术可通过捕捉修井机大钩等部件关键位置和运动轨迹获取实时工况信息。

在计算机视觉领域中，姿态估计可分为自上而下[6-8]和自下而上的检测方法[9-10]。自上而下的方法先进行目标检测，分离图像中的个体，再在各目标区域执行姿态估计。其目标分割步骤能精确定位个体，可以有效解决拥挤场景中的姿态估计问题。例如，使用 Faster-RCNN（region-based convolutional neural networks）目标检测方法定位目标位置和数量[11]，然后在每个目标的检测框中单独应用姿态估计策略，基于上下文信息实现姿态估计。但这种方法对精度敏感性要求高，且算法训练所需要的时间随着图像的复杂度增大逐渐延长[12]。自下而上的方法先检测所有关键点，再根据空间和几何关系聚类为不同个体。该方法计算效率高，但在复杂背景或个体密集场景中容易误匹配。因此，自下而上的方法通常使用高斯热图计算和匹配不同个体，只是这种做法增加了额外的后处理，导致模型复杂度更高[13]。综合考虑自上而下与自下而上方法优点及图像特征，选择无热图的自下而上 YOLO-Pose 算法来实现修井机关键点检测与姿态信息获取。

1.2　修井机整体关键点定位

YOLO-Pose 姿态估计算法基于 YOLO（you only look once）目标检测框架建立。YOLO 作为目前应用程度最为广泛的目标检测框架，在准确性和复杂性方面均处于领先水平[14]，笔者基于此框架联合目标检测与姿态估计构建神经网络。在通用领域中，YOLO 在进行检测时其目标检测头将每个目标使用特定维度的 Anchor（先验框）进行匹配。Anchor 是目标检测中一种预先定义的、具有不同尺寸和长宽比的边界框，用于在特征图上生成可能包含目标对象的候选区域，从而定位图像中的目标对象。经典的通用领域数据集 COCO（common objects in context）包含 80 个类别[15]，每一个先验框预测 85 个元素，分别对应 80 个类的边界框、对象分数及置信度分数，而每个网格的位置中又包含 3 种不同形状的先验框[16]。YOLO-Pose 算法通过新增关键点检测头并将关键点信息一并添加进先验框中，以扩大先验框的检测目标，实现一次性预测出目标框及关键点。

修井作业现场的姿态估计不需要多个目标的检测，可被归结为单一目标的修井机检测问题。根据修井机的井架大钩结构，每个大钩目标包含 5 个关键点（见图 1），每个关键点都有对应的位置及置信度分数。每一个先验框预测 5 个关键点的 15 个元素。为了综合判定大钩的运动轨迹及其运动过程中与井架是否存在偏移，利用目标检测头同时检测大钩-游车系统（后简称大钩系统）与井架 2 个目标。每个目标框的识别需要 6 个元素，则整体联合目标检测头预测需要 12 个元素。综上，对于具有 6 个关键点的先验框，其总体预测向量（prediction vector）$\boldsymbol{P}_{\mathrm{v}}$ 可以定义为：

$$\boldsymbol{P}_{\mathrm{v}}=\left\{\boldsymbol{C}_x^1,\boldsymbol{C}_y^1,\boldsymbol{W}^1,\boldsymbol{H}^1,\boldsymbol{B}_{\mathrm{conf}}^1,\boldsymbol{C}_{\mathrm{conf}}^1,\boldsymbol{C}_x^2,\boldsymbol{C}_y^2,\boldsymbol{W}^2,\boldsymbol{H}^2,\boldsymbol{B}_{\mathrm{conf}}^2,\boldsymbol{C}_{\mathrm{conf}}^2,K_x^1,K_y^1K_{\mathrm{conf}}^1,\ldots,K_x^6,K_y^6,K_{\mathrm{conf}}^6\right\} \tag{1}$$

式中：$\boldsymbol{C}_x$、$\boldsymbol{C}_y$、$\boldsymbol{W}$、$\boldsymbol{H}$、$\boldsymbol{B}_{\mathrm{conf}}$ 和 $\boldsymbol{C}_{\mathrm{conf}}$ 由目标检测头负责预测，分别为预测框的中心点横坐标、纵坐标、框的宽度、框的高度、框置信度分数和类置信度分数，角标 1 和 2 分别为修井机的大钩系统及井架；剩余部分由关键点检测头负责预测，分别为大钩及井口的每个关键点的横坐标、纵坐标及置信度分数。关键点 K^1~K^5 对应大钩系统的 5 个关键点，K^6 对应井架（井口）的关键点。

在进行关键点检测任务时，第一步是对训练数据进行标注，对于通用领域中的人体姿态估计任务中，所制定的关键点标注规则通常包括 17 个特征点，其序号对应关系分别为 0-鼻尖、1-左眼、2-右眼、3-左耳、4-右耳、5-左肩关节、6-右肩关节、7-左肘关节、8-右肘关节、9-左腕关节、10-右腕关节、11-左髋关节、12 -右髋关节、13-左膝关节、14-右膝关节、15-左踝关节、16-右踝关节[15]，其分布位置如图 2 所示。

在进行大钩的关键点检测时，根据大钩系统的运动部件结构特征设置 5 个关键点分布，其对应关系分别为 0-游车中心、1-游车左端、2-游车右端、3-大钩链接点、4-大钩低端，分布位置如图 1 所示。对于井口仅需要确定井口位置信息，因此设定 1 个关键点定位井口。根据设计的关键点信息，定位实际运行时的大钩关键位置，为后续的计

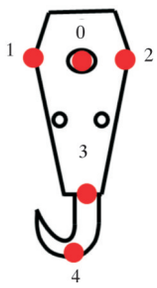


图 1　大钩系统的关键点标注方案示意图

Fig. 1　Keypoint annotation scheme for the hook system

算与分析提供依据。同时，根据现场使用的大钩规模获得实际大钩尺寸与图像中关键点 1、2 之间的距离之比，便能够计算图像与现实的比例尺 λ。

1.3　整体关键点局部化

整体关键点局部化是一种姿态估计策略，将任务分为粗略定位和局部精细化 2 个子任务。首先在全图或检测框内快速定位关键点，获取整体空间结构；再在粗定位区域优化，提高预测准确性。这种方法通过分阶段定位关键点提升大钩姿态估计精度和效率。

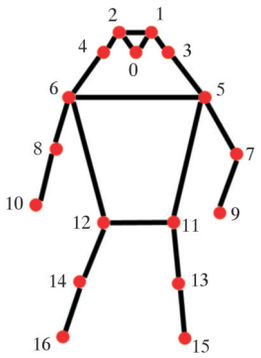


图 2　通用任务的关键点标注方案示意图

Fig. 2　Keypoint annotation scheme for general tasks

这里先用 YOLO 网络获取大钩区域候选框，再在框内粗略定位关键点，初步估计其相对位置，锁定大钩位置和形状后，生成粗略姿态估计。

在粗略定位的基础上，增加模型卷积层数，利用级联金字塔网络[17]逐层细化图像局部特征，使用更高精度的局部化模型在每个关键点的初步位置周围进行细化，以间接提升对大钩关键点定位的精度。模型的整体流程如图 3 所示。

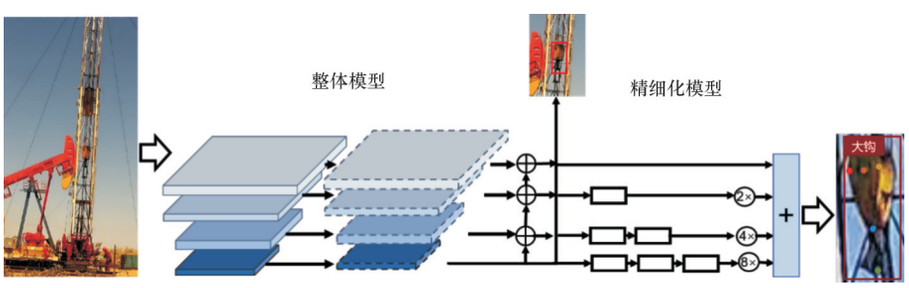


图 3　整体关键点局部化流程示意图

Fig. 3　Overall keypoint localization process

2　系统轨迹识别与工况分析

2.1　姿态引导的注意力机制

在修井机系统智能化中，轨迹识别与工况分析对提升效率至关重要。笔者用姿态引导注意力机制识别轨迹、整合姿态信息，动态调整关注点，捕捉关键特征。复杂作业环境下，轨迹数据常含冗余和噪声，该机制可引导模型聚焦关键特征，过滤干扰信息，显著提高识别精度，使系统更准确地捕捉运动规律。

注意力机制效仿人类视觉系统中的选择性注意机制[18]，通过构建一种专注于图像关键区域的模型架构，降低图像处理的计算复杂度，并提升模型效能。该机制通过加权计算输入数据，权重体现信息相关程度。具体实现时，输入数据先经特征提取模块转为特征向量，再由与任务状态关联的查询向量筛选重要特征向量，确保模型聚焦相关信息。

注意力机制的核心包括 3 个组成部分：查询（Query）、键（Key）和值（Value）。查询与各个键之间的相似度通过得分函数计算，常用的得分函数包括加性得分函数和点积得分函数。得分函数输出一个分数向量，表示每个向量的重要性，表达式如下：

$$\boldsymbol{e}_l = \mathrm{score}(\boldsymbol{q}, \boldsymbol{k}_l) \tag{2}$$

式中：$\boldsymbol{q}$ 为查询向量，$\boldsymbol{k}_l$ 为第 l 个向量，$\boldsymbol{e}_l$ 为对应的注意力分数。

注意力分数通过对齐函数转换成注意力权重。常用的对齐函数是 Softmax 函数，它将注意力分数转换成一个概率分布，确保所有权重的和为 1：

$$\boldsymbol{a}_l = \frac{\exp(e_l)}{\sum_{j=1}^{n}\exp(e_j)} \tag{3}$$

式中：$\boldsymbol{a}_l$ 为第 l 个注意力权重，n 为特征向量的总数。

最后，模型使用这些注意力权重来计算上下文向量，即所有值向量的加权平均。其中每个值向量的权重由对应的注意力权重决定：

$$\boldsymbol{c}=\sum_{l=1}^{n}\boldsymbol{a}_l\boldsymbol{v}_l \tag{4}$$

式中：$\boldsymbol{v}_l$为第l个值向量，$\boldsymbol{c}$为上下文向量，被用于生成模型的输出。

注意力权重的原理如图4所示。

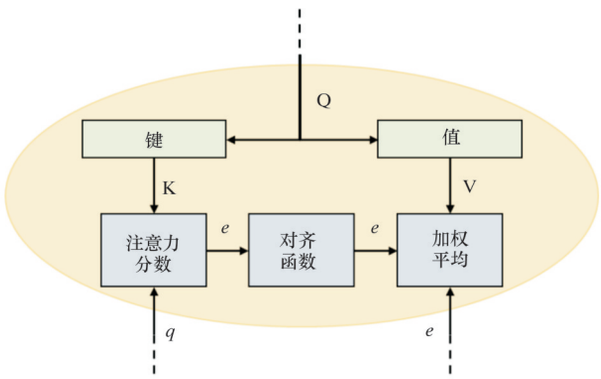


图4　注意力机制的原理图

Fig. 4　Principle of the attention mechanism

通道注意力机制调控图像特征通道，强化关键信息，削弱次要信息，动态地调整特征图通道权重，使网络聚焦重要特征。在深度卷积神经网络中，各卷积层输出的特征图表征不同维度特征。该机制突出相关特征通道，抑制低贡献或噪声通道，优化模型性能。Wang等[19]提出通道注意力机制来增强深度卷积神经网络的性能，解决传统SENet[20]（squeeze-and-excitation networks）通道注意力机制在计算效率和参数数量上的不足，实现特征图中通道间依赖关系的轻量级捕获。

2.2　修井机轨迹跟踪及识别

按设计的数据标注方案处理井场作业场景图像数据，获得模型输入数据集，包含修井机姿态信息。将修井机图像输入YOLO-Pose网络，经主干生成多分辨率特征图，PANet（路径聚合网络）进行特征融合后送入检测头，联合目标检测与关键点检测，输出修井机井架和大钩的检测框、关键点坐标，并连成骨架图。模型框架如图5所示。姿态估计阶段，使用姿态非最大抑制策略消除冗余姿态估计，保留最可能姿态[21-22]，对姿态估计网络进一步调整关键点，实现精确姿态估计，通过姿态引导对准技术微调修井机姿态，提高准确性。

在动态识别方面，构建运转动态识别模型，利用曲线提取和运动分析识别修井机动态行为，理解其运动模式。最终，用轨迹识别技术准确识别修井机大钩轨迹，为井场作业监控和分析提供重要数据支持。

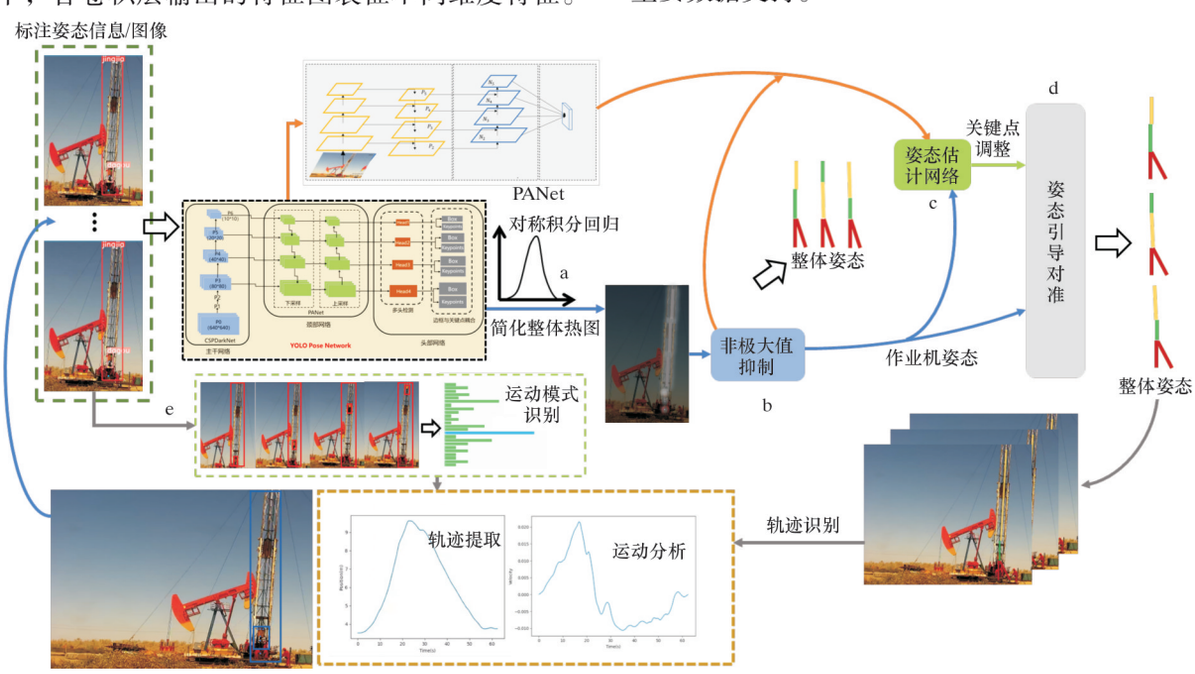


图5　模型框架图

Fig. 5　Model framework

2.3　修井机系统工况分析

通过现场监控视频获取大钩运动信息，构建位移-时间曲线，计算速度和加速度曲线，并计算大钩受力情况。具体流程如图6所示。

将大钩和井口采油树在视频图像中的实时坐标通过坐标转换算法转换为全局坐标系下的坐标信息。计算两者相对距离、角度及方位等几何参数，据此判定对齐情况，如图7所示。

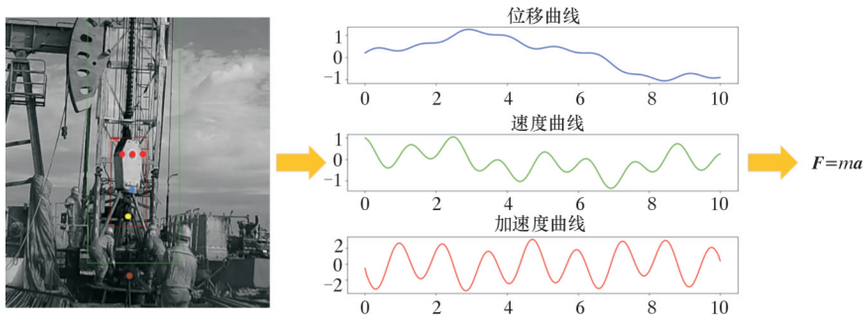


图 6　模型框架图

Fig. 6　Model framework

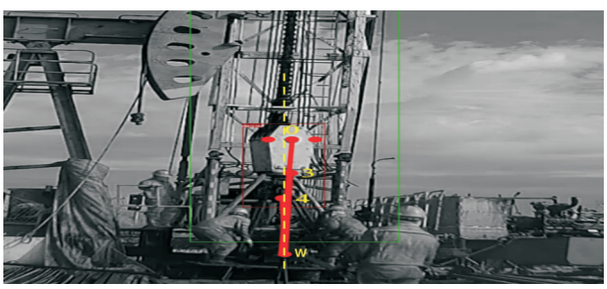


图 7　大钩摆动姿态估计示意图(图片来源于测试集)

Fig. 7　Hook swing pose estimation (Photo is sourced from the test set)

利用大钩本身的空间信息进行图像方位矫正，将图中大钩点位 0、3、4 的位置信息记为（x_{h0}，y_{h0}）、（x_{h3}，y_{h3}）、（x_{h4}，y_{h4}），井口位置记为（x_w，y_w），由此可以得到 3 段连线与垂直线的夹角：

$$\begin{cases} \theta_1 = \arctan\left|\dfrac{x_{h0}-x_w}{y_{h0}-y_w}\right| \\ \theta_2 = \arctan\left|\dfrac{x_{h3}-x_w}{y_{h3}-y_w}\right| \\ \theta_3 = \arctan\left|\dfrac{x_{h4}-x_w}{y_{h4}-y_w}\right| \end{cases} \tag{5}$$

式中：θ_1、θ_2、θ_3为 3 个位置夹角，rad。

根据 3 段线段的夹角，设定垂线判定准则：

$$\bar{\theta} = \frac{\theta_1+\theta_2+\theta_3}{3} < \varepsilon \tag{6}$$

若得到结果$\bar{\theta}$小于设定阈值ε，则判定大钩与井口位置处于一条垂线上，即大钩本身没有发生大的摆动，反之认定当前作业处于危险状态。

基于姿态估计技术的修井机系统工况分析为智能化管理提供了更加精确的数据支撑和分析工具。

3　轨迹识别与工况判断效果分析

3.1　数据集构成

本方法用于识别修井作业现场的修井机及大钩运动轨迹。利用布控球和监控摄像机获取作业视频和图像数据，经标注和转换处理后用于模型训练。试验数据来自某油田 10 个小修作业现场的监控视频，抽帧处理后共收集 8 914 张修井机图像，按 YOLO 格式标准化处理形成数据集。数据集按 8∶2 比例划分为训练集和验证集，7 313 张图像用于训练，1 783 张用于验证。

3.2　试验过程

（1）预训练处理。采用 MS COCO 项目，该项目源自微软于 2014 年出资标注的 Microsoft COCO 数据集[15]。训练时使用 COCO2017 数据集，该数据集训练集包括 57 000 张图像，验证集与测试集分别包含 5 000 及 20 000 张图像。预训练从 COCO 数据集提取通用特征，减轻修井机检测任务的学习负担。

（2）训练阶段。训练集的批次在每个周期中以无序方式输入，以增强所训练模型的性能。优化器选择随机梯度下降法（stochastic gradient descent，SGD），基础学习率设置为 0.001。输入图像大小标准化为 640×640。对于输出，网络的学习准则采用目标关键点相似度（object keypoint similarity，OKS）指标，最终输出预测关键点图像及训练过程的各项损失变化。

（3）测试阶段。此阶段的批次大小设置为 16，输入图像大小保持不变。在图像测试的过程中，模型会对修井机的井架及大钩预测匹配的先验框，且同时关联所有目标的边界框位置及每个关键点的坐标信息。

3.3　模型训练效果分析

试验使用 4 个 NVIDIA TESLA V100 32 GiB GPU 加速训练。批次大小是平衡训练动态和效率的关键。小批量加快模型更新，引入高梯度噪声，增强收敛性但延长训练时间；大批量处理更多样本，提供稳定梯度估计，但增加计算需求。本试验选取 32 批次大小，平衡收敛效率和计算资源。

模型循环总数设为 300，总训练时间为 1.322 h，试验训练完成后模型的实际性能如图 8 所示。在 Linux 5.15.0-60-generic 系统，NVIDIA L20 GPU 上测试平均 FPS 结果为 119.51 帧/s，该处理速度能够满足现场实时监控需求。

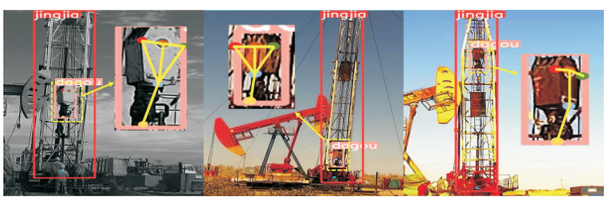


图 8　井架-大钩-井口的目标与关键点监测代表性试验结果（图片来源于测试集）

Fig. 8　Representative experimental results of target and keypoint detection for derrick-hook-wellhead (Photos are sourced from the test set)

对于模型的测试结果评估，选择精度（R_p）、召回率（R_c）、平均精度 $P_{A@0.5}$ 与平均精度均值 $P_{mA@0.5:0.95}$ 作为评价指标[23]。精度衡量的是模型进行的所有阳性检测中真正检测的比例，表明在不包括假阳性的情况下模型识别的准确性。召回率衡量的是所有实际阳性实例中真阳性检测的比例，反映了模型能够检测所有相关关键点而不遗漏任何真阳性。$P_{A@0.5}$ 表示 OKS 阈值设置为 0.5 时的平均精度，而 $P_{mA@0.5:0.95}$ 表示阈值设置为 0.5~0.95 之间的平均精度均值。

基础模型的目标边界框检测及关键点检测的各项评估指标如图 9 所示。由图 9 可以看到，模型的精度与召回率随着训练过程先快速上升，随后趋于稳定。完成 300 轮训练后，除关键点检测精度外其余各项指标仍低于 0.9。带有注意力机制模型试验结果如图 10 所示。在完成 300 轮训练后，此时模型的边框和关键点精度达到 99.91%，召回率均达到 1，这表明模型能够准确预测。完成训练后边框及关键点的 $P_{A@0.5}$ 达到 99.5%，且边框 $P_{mA@0.5:0.95}$ 达到 92.48%，关键点 $P_{mA@0.5:0.95}$ 达到 98.21%。

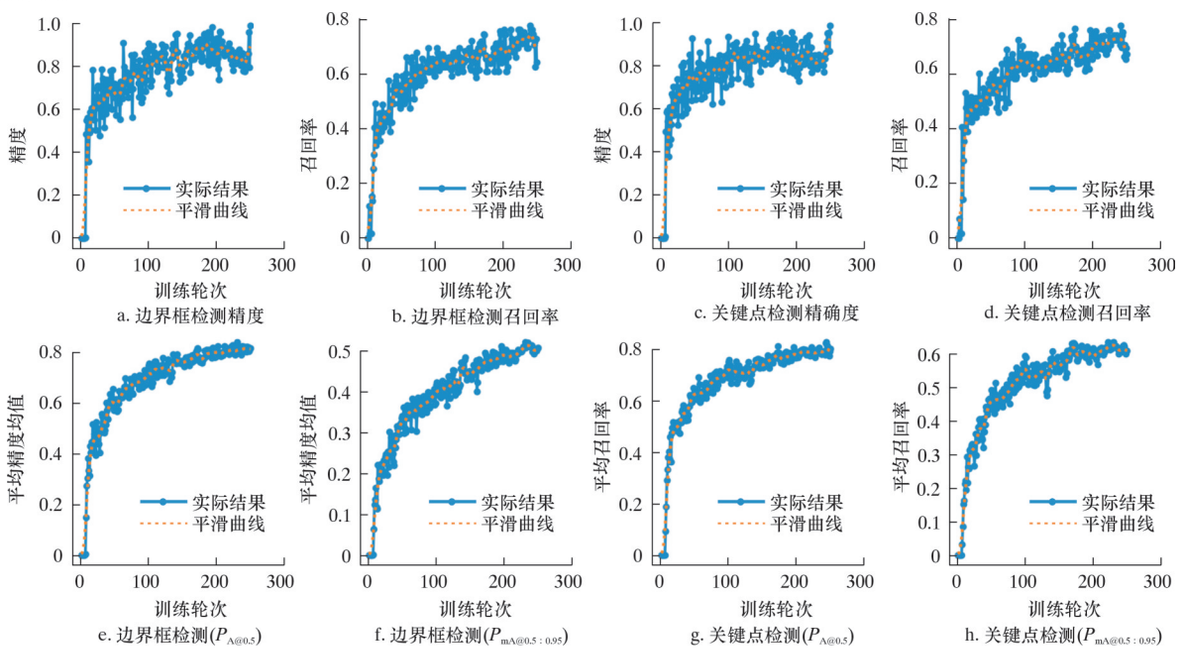


图 9　基础模型的各项评估指标曲线

Fig. 9　Evaluation index curves of the baseline model

损失函数（loss function）是衡量预测与真实标签差异的函数，目标是最小化误差，提高预测准确性。基础模型与带有注意力机制模型训练和验证阶段的各项损失曲线如图 11 与图 12 所示。试验中，各项损失随训练次数增加先迅速下降，后缓慢降低并趋于稳定，表明模型趋于收敛，神经网络学习充分，且带有注意力机制模型的曲线更平滑、损失值更低，表明模型学习更稳定、充分。

在随机抽取的 20 张图片上，模型的 O_{KS} 与 $K_{PC@0.2}$（以躯干尺寸的 0.1 倍作为阈值的正确关键点百分比）计算结果如图 13 和图 14 所示。其中，O_{KS} 均值为 0.924 6，说明预测关键点与人工标注的平均重合度已达 92.46%。$K_{PC@0.2}$ 达到 1.000，则表明当阈值放宽至躯干尺寸的 20% 时，模型对全部测试样本均实现了 100% 的正确检测。模型在定位准确性和鲁棒性 2 方面均能达到预期目标，验证了所提网络结构及训练策略的有效性。

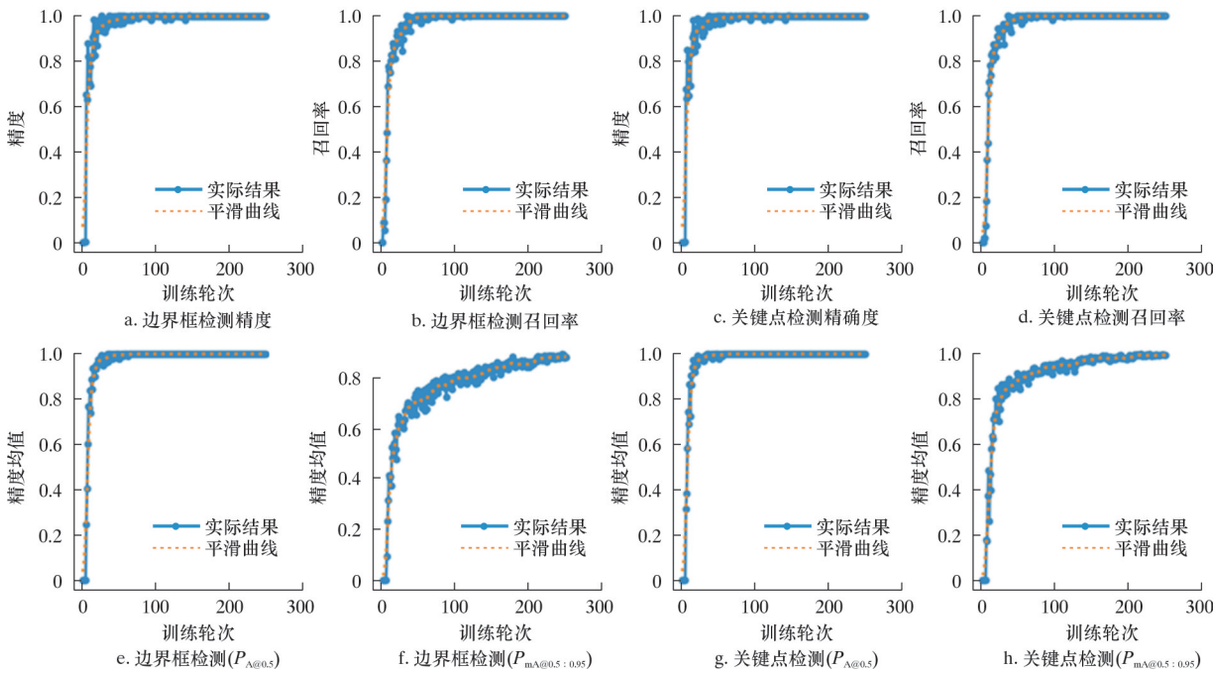


图 10　带有注意力机制模型的各项评估指标曲线

Fig. 10　Evaluation index curves of the attention-based model

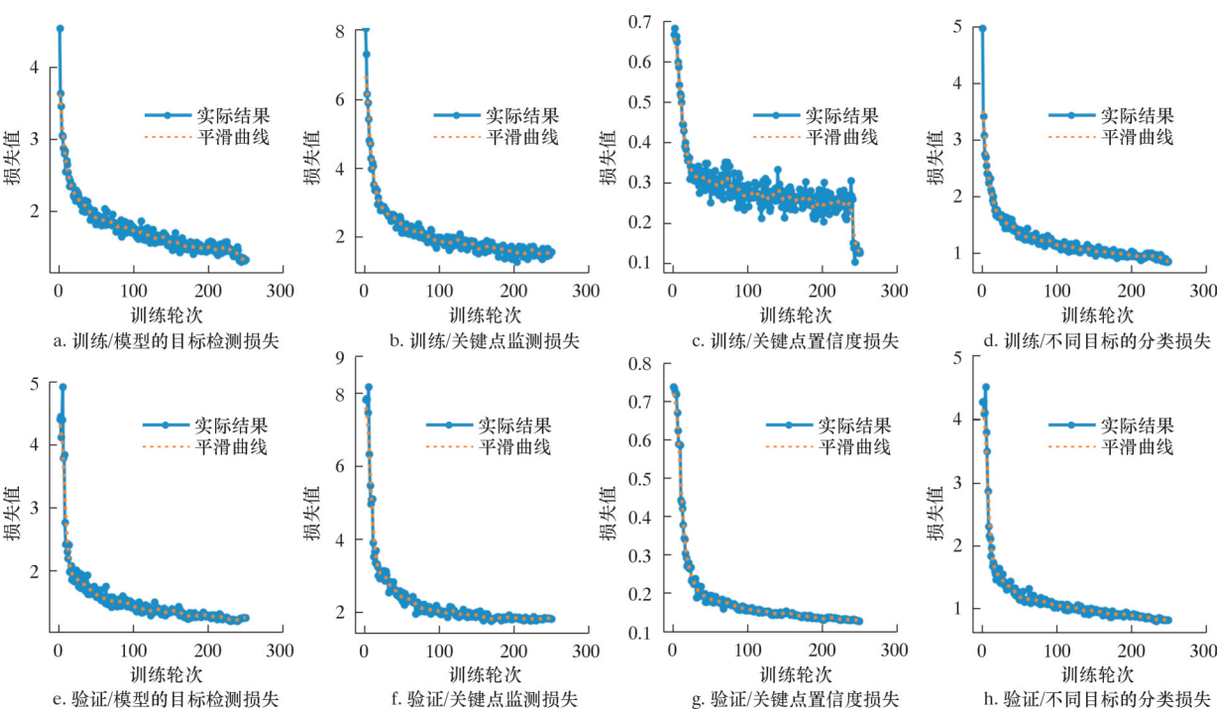


图 11　基础模型的各项损失曲线

Fig. 11　Loss curves of the baseline model

3.4　轨迹识别与工况效果分析

3.4.1　大钩运动轨迹跟踪与参数计算

输入现场作业图像后，模型输出大钩姿态信息。将某一口井的修井作业现场视频输入模型，计算得实际运动曲线（见图 15）。曲线显示大钩运行周期约 73 s，起油管后人工操作使大钩在顶端停留，与曲线一致。

利用提出的计算方法获取该大钩的速度与加速度，反映大钩运行的稳定性，结果如图 16 和图 17 所示。

3.4.2　修井作业大钩的稳定性分析

本方法通过大钩空间信息进行图像方位矫正，计算大钩点位 0、3、4 与井口的夹角平均值，动态衡量作业稳定性。试验中，大钩摆动角度如图 18

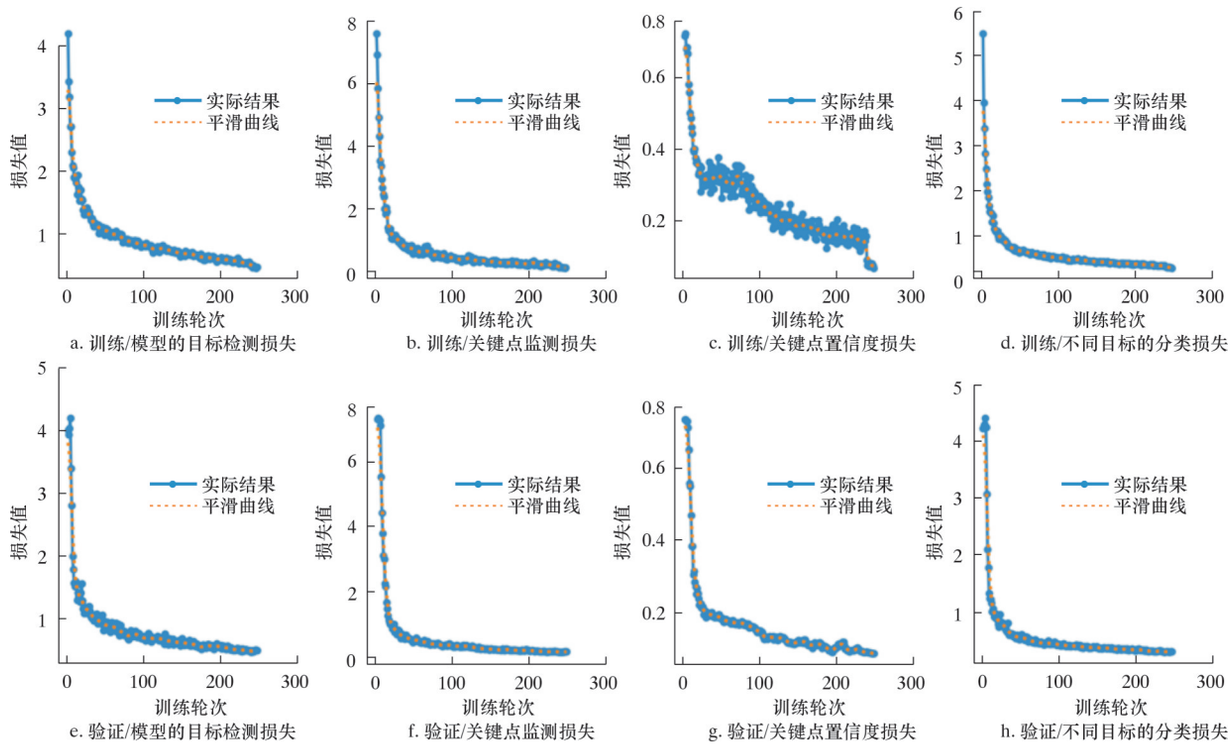


图 12　带有注意力机制模型的各项损失曲线

Fig. 12　Loss curves of the attention-based model

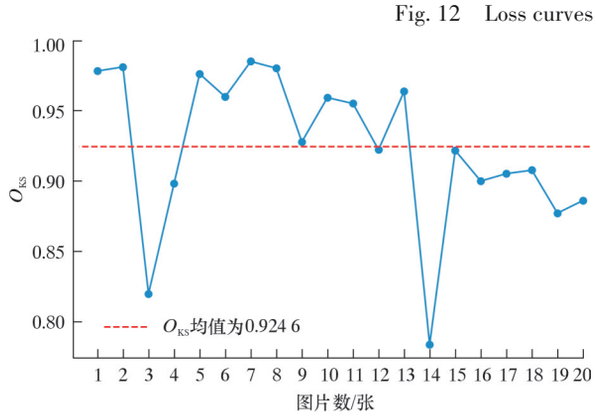


图 13　Object Keypoint Similarity（OKS）计算结果

Fig. 13　Object Keypoint Similarity (OKS) calculation results

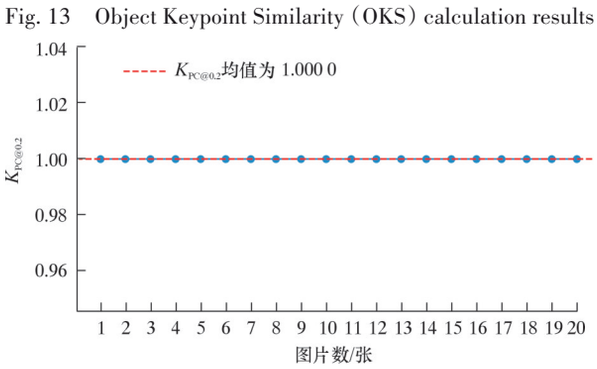


图 14　$K_{PC@0.2}$ 计算结果

Fig. 14　$K_{PC@0.2}$ calculation results

所示，曲线反映不同时刻大钩与井口夹角变化。下放油管时，大钩停顿且角度随油管撤离变大；正常作业时，角度变化小于5°。使用判定标准公式（6）整体衡量大钩运行状态，设定角度阈值为

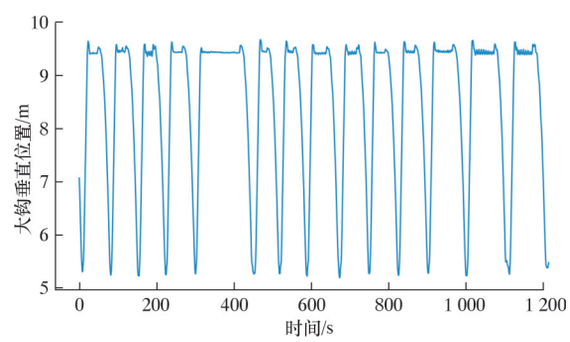


图 15　某井修井作业大钩运动曲线

Fig. 15　Hook motion curve of a workover operation of a well

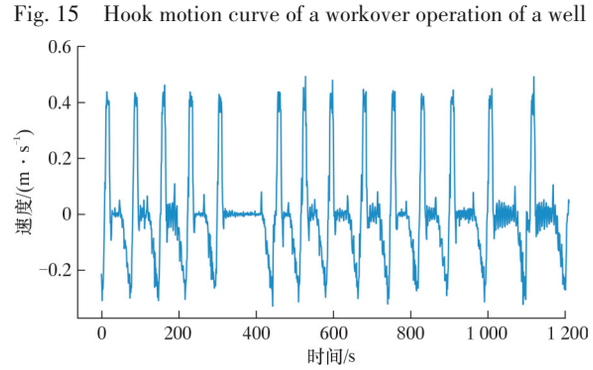


图 16　某井修井作业大钩运动速度曲线

Fig. 16　Hook velocity curve of a workover operation of a well

5°，即阈值：

$$\varepsilon = \frac{5\pi}{180} \approx 0.087 \tag{7}$$

在该井作业过程中，大钩起下作业时其值均小于阈值，表明整体运行稳定。

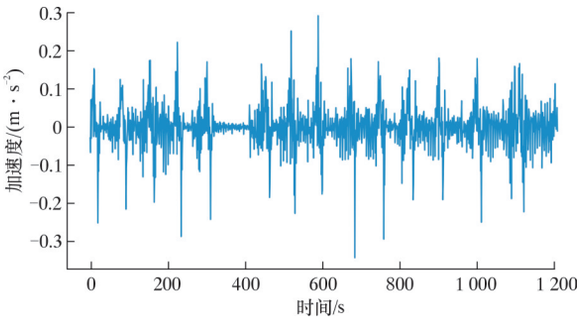


图 17　某井修井作业大钩运动加速度曲线

Fig. 17　Hook acceleration curve of a workover operation of a well

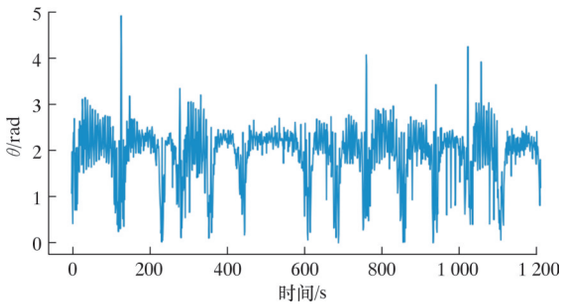


图 18　作业现场整体的角度变化曲线(图中θ是式(5)计算出的平均角度)

Fig. 18　Overall angle variation curve at the operation site (θ is the average angle from Eq. (5))

4　结　论

(1) 提出高效局部化姿态估计技术。该技术设计了基于局部特征的姿态估计方法，实现修井机及其关键部件的精确快速定位，提升姿态估计精度与效率。

(2) 开展姿态引导下的轨迹识别与动态分析。通过构建注意力机制增强的轨迹识别模型，实现修井机轨迹识别与工况分析的深度耦合，提高轨迹识别准确性及模型适应性。

(3) 基于姿态估计的工况精细分析方法。利用姿态估计技术深入剖析大钩动态受力特性，实现大钩摆动自动判定与故障风险精准预测，可提升作业安全性和效率。

参　考　文　献

[1] Sircar A, Yadav K, Rayavarapu K, et al. Application of machine learning and artificial intelligence in oil and gas industry [J]. Petroleum Research, 2021, 6(4): 379-391.

[2] Zhang Kai, Song Zewen, Xia Xiaolong, et al. Enhancing safety and workover efficiency through real-time scale-adaptive video detection of critical objects at the wellsite [J]. SPE Journal, 2024, 29(8): 3982-3999.

[3] 张菲菲,王茜,王学迎,等. 油气井工程多源多模态数据融合技术与展望[J]. 天然气工业,2024,44(9):152-166.
Zhang Feifei, Wang Qian, Wang Xueying, et al. Multi-source and multi-modal data fusion technology and its prospect in oil and gas well engineering[J]. Natural Gas Industry, 2024, 44(9): 152-166.

[4] 王晓宇,闫继宏,秦勇,等. 基于扩展卡尔曼滤波的两轮机器人姿态估计[J]. 哈尔滨工业大学学报,2007,39(12):1920-1924.
Wang Xiaoyu, Yan Jihong, Qin Yong, et al. Attitude estimation based on extended Kalman filter for a two-wheeled robot[J]. Journal of Harbin Institute of Technology, 2007, 39(12): 1920-1924.

[5] Zhang Kai, Xia Xiaolong, Song Zewen, et al. Trajectory recognition and working condition analysis of rod pumping systems based on pose estimation method with heatmap-free joint detection[J]. SPE Journal, 2024, 29(10): 5521-5553.

[6] Kocabas M, Karagoz S, Akbas E. MultiPoseNet: fast multi-person pose estimation using pose residual network [C]//Computer Vision-ECCV 2018, 2018: 437-453.

[7] He Kaiming, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2980-2988.

[8] Newell A, Yang Kaiyu, Deng Jia. Stacked hourglass networks for human pose estimation[C]//Computer Vision - ECCV 2016, 2016: 483-499.

[9] Cao Zhe, Simon T, Wei S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1302-1310.

[10] Pishchulin L, Insafutdinov E, Tang Siyu, et al. DeepCut: joint subset partition and labeling for multi person pose estimation[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 4929-4937.

[11] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[12] Toshev A, Szegedy C. DeepPose: human pose estimation via deep neural networks[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1653-1660.

[13] Tompson J, Goroshin R, Jain A, et al. Efficient object localization using Convolutional Networks [C]//2015

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 648-656.

[14] Maji D, Nagori S, Mathew M, et al. YOLO-Pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss [C] //2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022: 2636-2645.

[15] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [C] //Computer Vision - ECCV 2014, 2014: 740-755.

[16] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 778-779.

[17] Chen Yilun, Wang Zhicheng, Peng Yuxiang, et al. Cascaded pyramid network for multi-person pose estimation [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7103-7112.

[18] Larochelle H, Hinton G E. Learning to combine foveal glimpses with a third-order Boltzmann machine [C] // Proceedings of the 24th International Conference on Neural Information Processing Systems-Volume 1, 2010: 1243-1251.

[19] Wang Qilong, Wu Banggu, Zhu Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

[20] Hu Jie, Shen Li, Sun Gang. Squeeze-and-excitation networks [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[21] Liu Shu, Qi Lu, Qin Haifang, et al. Path aggregation network for instance segmentation [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[22] Yang Wei, Li Shuang, Ouyang Wanli, et al. Learning feature Pyramids for human pose estimation [C] //2017 IEEE International Conference on Computer Vision (ICCV), 2017: 1290-1299.

[23] Munea T L, Jembre Y Z, Weldegebriel H T, et al. The progress of human pose estimation: a survey and taxonomy of models applied in 2D human pose estimation [J]. IEEE Access, 2020, 8: 133330-133348.

第一作者简介：王利君，女，高级工程师，生于1987年，2013年毕业于中国石油大学（华东）控制技术与工程专业，获硕士学位，现从事油田信息化及人工智能应用研究工作。地址：（834000）新疆克拉玛依市。电话：（0990）6890627。email：wanglj_xj@petrochina.com.cn。

通信作者：张德君，高级工程师。email：zhdjun@petrochina.com.cn。

修改回稿日期：2026-07-19

（本文编辑　宋治国）