

基于 HOI 集合预测的带电作业场景元素交互关系解析方法

马富齐^{1,2}, 刘永文¹, 王 波², 贾 嵘¹, 董旭柱², 马恒瑞²

(1. 西安理工大学电气工程学院, 西安 710054; 2. 武汉大学电气与自动化学院, 武汉 430072)

摘 要: 带电作业风险辨识与防护对保障人员生命安全及电网安全运行意义重大。现有风险辨识研究主要集中在安全防护工具的规范佩戴、人员动态不安全行为辨识及电力设备故障隐患检测等单场景元素风险辨识, 对场景元素间的交互关系考虑不足。为此, 该文提出了一种基于人物交互(human object interactions, HOI)集合预测的带电作业场景元素交互关系解析方法。该方法首先基于残差网络的视觉特征提取模块感知带电作业影像的降维扁平化特征, 并利用位置编码模块获取图像像素特征的位置信息; 然后基于多头注意力机制编码器捕获全局交互关系的增强型特征向量, 利用解码器得到固定数量的 HOI 交互特征查询向量, 并通过对增强型特征向量交叉感知的方式实现对查询向量的迭代解码; 其次以基于二分图匹配方式的多层感知机集合预测模块进行最优交互关系的判别, 从而实现了带电作业场景元素交互关系的精细解析。最后以登高作业场景为例, 结果表明所提方法的平均准确率可达 94.40%, 且在不同环境下均具有较好的识别效果。所提模型可为带电场景关系检测提供参考。

关键词: 带电作业; 安全影像解译; 场景元素关系解析; 集合预测; 人物交互

Analysis Method for Interaction Relationships of Elements in Electrified Operation Scenarios Based on HOI Set Prediction

MA Fuqi^{1,2}, LIU Yongwen¹, WANG Bo², JIA Rong¹, DONG Xuzhu², MA Hengrui²

(1. School of Electrical Engineering, Xi'an University of Technology, Xi'an 710054, China;

2. School of Electrical and Automation, Wuhan University, Wuhan 430072, China)

Abstract: Risk identification and protection of electrified operation is of great significance to ensure the safety of personnel and the safe operation of power grid. Existing researches on risk identification mainly focus on the standardized wearing of safety protective tools, identification of unsafe dynamic behaviors of personnel, and detection of potential faults in electrical equipment, all of which do not sufficiently consider the interaction relationships between scene elements but address the risks in single scene elements. To this end, the paper proposes an interaction relationship analysis method for energized work scenario elements based on the prediction of human object interactions (HOI) sets. The method firstly perceives the downsampled flattened features of the electrified operation image based on the visual feature extraction module of the residual network, and obtains the position information of the image pixel features by using the position encoding module. Then, based on the multi-head attention mechanism encoder, enhanced feature vectors capturing global interaction relationships are obtained, and the decoder is used to generate a fixed number of HOI interaction feature query vectors. The iterative decoding of the query vectors is achieved through cross-perception of the enhanced feature vectors. Subsequently, an optimal interaction relationship is determined by using a multi-layer perceptron ensemble prediction module based on a bipartite graph matching approach, thereby achieving a detailed analysis of the interaction relationships of elements in live-line operation scenarios. Finally, the high-altitude operation scenario is taken as an example, and the results show that the proposed method achieves an average accuracy of 94.40% and demonstrates good recognition performance in various environments. The proposed model can serve as a reference for detecting interaction relationships in live-line scenarios.

Key words: electrified operation; safety image interpretation; scene element relationship analysis; ensemble prediction; human object interactions

0 引言

随着电力系统规模的日益扩大, 电力生产安全面临着巨大挑战, 各类电力安全风险辨识方法被广泛运用在带电作业场景中, 极大地提高了电力系统的稳定性。然而, 实际带电作业场景是涉及人员、工具、设备机械等多元目标的复杂交互过程, 安全风险通常蕴藏在复杂的交互关系中。仅 2021 年第二季度, 全国就发生电力人身伤亡事故 13 起、死亡 13 人^[1]。其中, 由于工作人员与工具、设备间的错误交互动作导致的安全事故占比 38.46%。由此可见, 研究带电作业场景元素交互关系检测方法, 对提高违章行为风险辨识准确率、保障带电作业本质安全水平具有重要意义。

当前, 国内外对于违章风险检测的研究主要集中在安全防护工具的规范佩戴、工作人员的行为辨识等单场景元素上, 且已广泛应用在电力生产、施工建筑、煤矿作业等工业领域。如文献[2]以掩膜区域卷积神经网络为基础, 提出了一种用于监控安全帽佩戴规范性的智能图像处理方法, 实现了在多目标场景下对安全帽佩戴规范度的精确辨识。文献[3]通过改进坐标后处理的单人关键点检测和区域定位的方式, 提出了一种基于关键点检测和注意力机制的违规着装识别算法, 大幅度提升了违规着装的识别准确率。文献[4]融合了分割网络与跟踪网络的特点, 在孪生网络框架的基础上提出了一种将跟踪与分割相融合的算法, 实现了在复杂情况下对人体高速移动及姿态变化的高精度追踪, 进而减少了带电作业场景中因工作人员操作不当引起的电力事故。文献[5-6]为解决建筑施工与煤矿作业场景下防护工具的规范佩戴问题, 分别提出基于 S3-YOLOv5s 与 YOLOv5-ShuffleNetv2 模型的防护设备检测算法, 实现了防护设备检测精度的大幅提升。

然而, 在实际带电作业场景中, 安全风险复杂多样, 仅靠对单场景元素的检测远远不够^[7]。近年来, 随着人工智能及计算机技术的不断发展, 元素交互关系解析方法已被陆续应用在带电作业场景中。如文献[8]通过语义分割和边缘检测相协同的方式, 将多模态高级特征融合在全局和语义视图中, 并通过图内与图间元素的信息交互, 有效提取出生产作业影像的上下文信息, 进而实现了电力场景的智能理解。文献[9]将多对象交互性考虑到风险识别中, 提出将语义特征、空间位置特征及视觉特征相

结合的方式来构建关系检测模块, 并将该模块与 Faster R-CNN 相结合, 进而实现了元素交互关系的智能理解及电力场景的自动危险预警。文献[10]将注意力机制引入到关系交互的检测中, 并将视觉特征、空间特征与其融合, 从而实现了场景元素之间的关系推理。文献[11]通过引入人员姿态的方式, 实现了工作人员与工器具间关系的判别。综上所述, 当前方法大多主要聚焦在安全帽、绝缘手套、工作人员等单一场景元素诱因的违章风险上, 然而实际带电作业违章行为风险蕴藏在人员、电力设备、安全工器具和环境等多场景元素的交互关系中。因此, 如何感知和理解带电作业场景元素之间的交互关系, 实现带电作业场景安全风险的精细化辨识, 是降低带电作业生产安全事故发生概率、提高带电作业本质安全水平的关键。

为此, 本文聚焦于带电作业场景中场景元素交互关系的精确辨别问题, 利用固定视频监控、安全布控球^[12-13]等感知的视觉影像为数据来源, 在分析场景元素间交互关系复杂性的基础上, 提出了一种基于人物交互(human object interactions, HOI)集合预测的元素交互关系智能解析方法, 从而实现了带电作业场景中人员与工器具间复杂交互关系的精细化辨识。

1 带电作业场景视觉特征分析

带电作业场景元素交互关系解析的目的在于从现场影像中识别出各场景元素的目标类型及相应位置, 并推理出各元素间的交互关系, 为带电作业场景的智能理解及违章行为风险的准确辨识提供基础, 然而当前检测方法大多以单一场景元素检测为主, 对场景元素间的交互关系考虑不足。如图 1 所示, 以登高作业场景为例, 该场景主要涉及工作人员(person)、登高梯(ladder)两种单场景元素, 及登梯(ascend)、扶梯(hold)、抬梯(lift)、在梯旁(beside)4 类交互动作。然而在实际的带电作业现场中通常会出现场景元素相互遮挡、交互关系相互干扰的情况, 仅对单一场景元素检测并不能解决上述问题。如图 1(a)、图 1(b)与图 1(d)所示, 图中均出现了登高梯被大量遮挡的情况, 且图 1(a)、图 1(b)的背景元素较复杂, 干扰较大。如图 1(c)所示, 当工作人员与登高梯距离较近时, 在梯旁状态与扶梯状态的交互特征会较为相似, 同时当实际带电作业场景出现多类交互关系时, 交互关系之间也会相互干扰。因此,

如何充分挖掘单一场景元素间的交互关系, 准确快速地识别出元素间的交互类型, 是实现带电作业场景智能理解的关键。

2 基于 HOI 集合关系预测的带电作业场景元素交互关系解析方法

带电作业场景元素交互关系解析的难点在于: 在实际带电作业场景中, 电力设备、工具具与工作



图1 带电作业场景元素交互关系示意图

Fig.1 Schematic diagram of the interaction relationship between the elements of the electrified operation scene

人员之间的交互关系复杂多样, 且不同作业环节的交互关系相互影响、相互交叉。2018 年, Georgia Gkioxari 等人首次将深度学习引入人物交互(HOI)检测算法中, 不同于传统对单一元素的检测方法, HOI 的核心思想在于检测到单场景元素的同时, 推断出以人为中心的元素交互关系, 进而实现对场景的智能解析。本文借鉴 Transformer^[14]的思想, 以端对端的目标检测方法为基础, 提出了一种基于并行编解码集合关系预测的带电作业场景元素交互关系解析方法。通过将图像特征、位置编码与元素交互关系相结合的方式, 实现了工作人员与工具具关系的精准辨识。该方法的具体架构如图 2 所示(其中, W 、 H 、 d 分别代表图像特征的宽度、高度与深度; $L \times N$ 、 $M \times L$ 、 $M \times L$ 则代表不同类型运行向量的数量及维度), 该方法首先通过 Resnet50 提取电力影像特征图并对输出特征进行扁平化处理, 其次将处理后的图像特征与位置编码一同传入到基于 Transformer 的编码模块中得到全局储存并将全局储存、位置编码(本文的位置编码均采用了正余弦的编码方式)、HOI 位置嵌入模块输入到基于自注意力机制的并行解码模块中, 通过交叉注意力层(解码器中的多头注意力层)的点注意力操作, 实现解码器查询向量与编码器键值向量之间的相似度筛选及权重向量的相应更新, 最后通过基于多层 MLP(multi-layer perceptron)结构的集合预测层对更新后的查询向量进行关系三元组的集合预测, 进而实现了带电作业场景元素交互关系的智能解析。

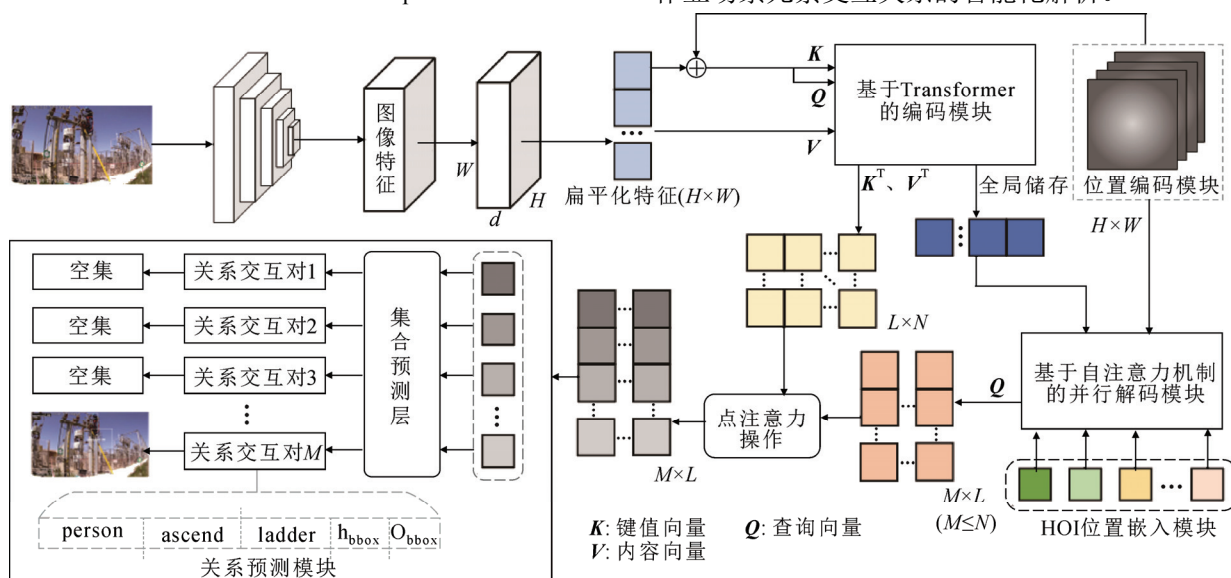


图2 基于 HOI 集合关系预测的带电作业场景元素交互关系解析模型框架图

Fig.2 Framework diagram of the interaction relationship analysis model of the elements of the electrified operation scene based on the prediction of HOI ensemble relationship

2.1 基于残差神经网络的全局视觉特征提取模块

视觉特征的准确提取是实现带电作业场景中元素交互关系智能解析的首要任务。目前存在 AlexNet、VGG、GoogleNet、DenseNet、ResNet 等多种视觉特征提取网络^[15-19]。其中, ResNet 网络通过引入残差块结构, 有效解决了因卷积神经网络层数加深导致的权重衰减、过拟合、梯度消失等问题, 极大提升了网络的特征提取能力^[20-21]。本文综合考虑网络的视觉提取能力与算力需求, 选择了 ResNet50 网络作为视觉特征提取模块的主干网络。

该网络的模型结构图如图 3 所示(其中 Conv7×7、Stride、Padding 分别代表 7×7 大小的卷积块、步长、填充操作; MaxPool2d、BottleNeck、avg、fc 分别代表二维最大池化层、瓶颈模块、全局平均池化与全连接层; ×3、×3、×6 代表各模块的数量; “数字×数字”代表各阶段的特征尺寸; “数字, 数字”代表各阶段通道数的前后变化)。由于编-解码模块的输入为一维序列, 因此在利用深度残差网络提取视觉特征后, 需要对视觉特征进行扁平化处理, 并将空间和深度维度折叠为一维。本文首先利用 1×1 卷积对提取出的视觉特征($H \times W \times d$)进行降维处理, 使其维数由 d 变为 c 。最后通过扁平化操作(flatten 层)将处理后的视觉特征拉伸为 d 个长度为 $H \times W$ 的一维特征序列。

2.2 基于自注意力的并行编-解码器关系检测模块

本文在传统编-解码结构的基础上, 构建了基于自注意并行编-解码器的关系检测模块(见图 4), 以此来捕获元素交互关系的差异性特征, 该模块主要包括基于 Transformer 的编码模块与基于自注意力机制的并行解码模块两部分, 其中编码器模块由多头自注意力层、前馈神经网络层、残差连接与归一化层几部分组成, 其具体架构遵循文献^[22]。

该模块以扁平化的视觉特征与位置编码作为输入, 通过堆叠多个自注意力层和前馈神经网络层, 并结合残差连接和归一化操作, 使编码器能够逐层学习到输入图像特征序列中不同位置编码之间的关联程度, 进而增强了不同关系三元组间的特征差异性。同时由于多头机制的原因, 模型可以并行学习不同关注点的全局特征表示, 进而提高模型的全局理解能力。

解码器模块与编码器模块相比增加了多头注意力层结构, 该模块以编码器的全局特征、位置编码模块与 HOI 位置嵌入模块作为输入, 为了更好地

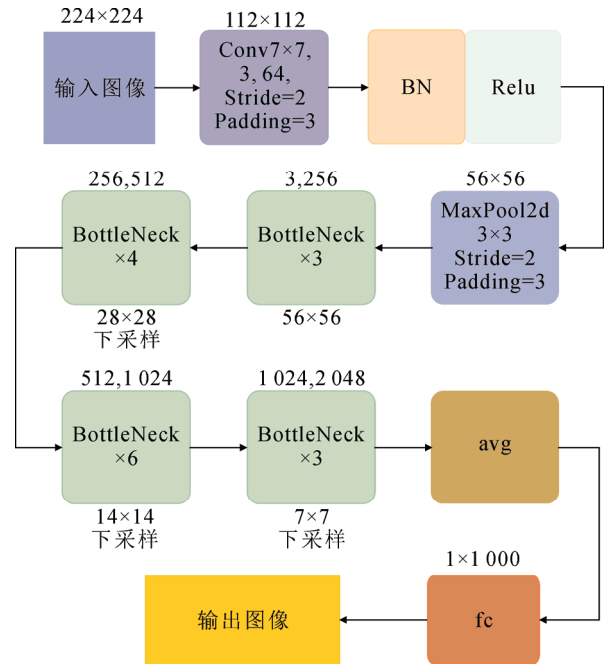


图3 Resnet50 模型结构图

Fig.3 Resnet50 model structure diagram

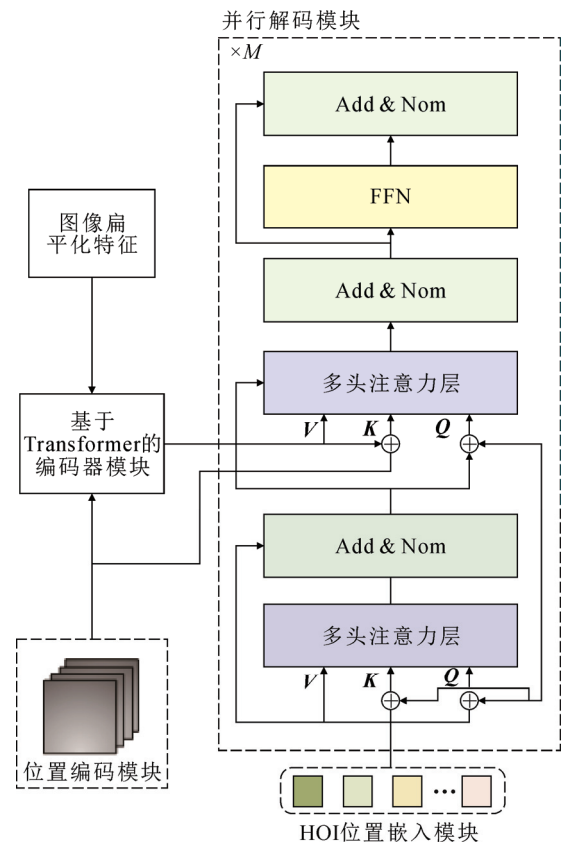


图4 基于自注意力机制的并行解码器结构图

Fig.4 Architecture of parallel decoder based on self-attention mechanism

提高解码器模块对图像全局关联性的理解能力与建模能力, 本文将原 Transformer 结构中的掩码多头注

注意力层用多头自注意力层代替,其具体网络架构图见图4(其中 V 、 K 、 Q 分别代表内容向量、键值向量、查询向量; $\times M$ 代表由 M 个解码模块堆叠而成; FFN、Add、Norm 分别代表前馈神经网络、残差连接与层归一化)。本文在解码器的输入端引入 HOI 位置嵌入模块,该模块由固定数量的 HOI 查询向量组成(其初始化向量通过采用“0 加位置编码”方式来实现查询向量间的有效区分),其数量决定了图像上关系三元组的最大预测数量。本文首先通过多头自注意力操作来增强不同 HOI 查询向量间的任务理解能力,该操作可减少不同 HOI 查询向量对同一区域或同一目标查询的情况,进而实现对冗余目标(相同关系三元组的特征)的有效筛选。其次,通过多头注意力多层的点注意力操作,使 HOI 查询向量对键值向量全局特征中的不同关系三元组进行特征学习并更新其权重,其具体关系式如下:

$$\text{IA}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V = WV \quad (1)$$

式中: $\text{IA}(Q, K, V)$ 为自注意力函数; Q 代表并行多头自注意力操作转换后的 HOI 查询向量; d_k 代表 K 的维度; W 代表注意力头权重向量; Softmax 代表归一化注意力权重操作。最后, HOI 查询向量通过在多个解码器的不断迭代,进而实现对不同三元关系组图像特征的有效查询及更新。

2.3 基于集合预测的场景元素关系解析模块

该模块主要由多层感知机组成,其中多层感知机也称深度神经网络,由多个感知机堆叠而成^[23],具有出色的非线性匹配和泛化能力^[24]。该模块以基于自注意力机制的并行编-解码器模块的输出(即每个 HOI 查询向量对应输出的 embedding)作为输入,并通过多个 MLP 结构将输出的 embedding 转换为关系三元组。具体来说,在集合预测层中共设置了 5 个 MLP 层,前 3 个 MLP 层为单层结构用以分别预测人员、交互动作及工器具的置信度,后 2 个 MLP 层为三层结构用以预测人员的边界框及工器具的边界框。最后,本文将场景元素交互关系检测转换成真实值和预测值两个集合之间的匹配问题,并以二分图匹配的方式优化该模型的损失。

在匹配过程中,本文遵循两个基本的匹配策略。1) 在所有预测值均正确预测出目标类别与交互动作的情况下,模型将会优先把场景元素交互实例分配给动作类别、目标定位均识别较准确的关系三元组。2) 在部分预测值未正确预测出目标类别或交

互动作时,模型优先选择动作类别与目标定位均识别准确率较高的关系三元组,而当交互动作类别准确率与目标定位准确度相左时,模型则会优先选择交互动作较准确的关系三元组,如图5所示(图中 P 、 L 分别代表人员边界框、工器具边界框),模型将优先选择虚线交互对(图5中用单实线的交互对来表示真实值),其具体原因可见本文消融实验。

同时,考虑到真实值与预测值数量不匹配问题,本文将添加空集到场景元素交互关系集合中进行集合匹配,匹配的成本函数如下:

$$L_{\text{cost}} = \sum_i^N L_{\text{match}}(T, P) = \sum_i^N L_{\text{match}}(t^i, p^{f(i)}) \quad (2)$$

式中: T 、 P 分别代表真实值与预测值; i 、 N 代表序列号; $f(i)$ 代表分配给第 i 个场景元素交互关系实例的预测关系组的索引; $L_{\text{match}}(t^i, p^{f(i)})$ 代表第 i 个场景元素交互实例 t^i 与预测关系组 $p^{f(i)}$ 之间的匹配成本。

在训练过程中,本文则是首先在两个集合间找到最优的一对一匹配。具体匹配计算式如下:

$$L_{\text{match}}(t_j^i, p^{f(i)}) = \beta_1 \sum_{j \in p, r, l} \alpha_j L_{\text{cls}}^j + \beta_2 \sum_{k \in p, l} L_{\text{box}}^k \quad (3)$$

式中: L_{cls}^j 代表人员、工器具与交互动作的分类损失; t_j^i 代表在场景元素交互关系实例的集合中第 j 个实例的类别标签(本文采用标准的 Softmax 交叉熵损失计算方法); L_{box}^k 代表人员边界框与工器具边界框的回归损失(该回归损失由 L_1 损失与 giou 损失的加权和组成); α 与 β 为超参数的损失权重。其次本文采用匈牙利算法来解决二分图最大匹配问题,进而得出最优匹配解,具体计算式如下:

$$\hat{\sigma} = \arg \min_{\sigma \in \theta} L_{\text{cost}} \quad (4)$$

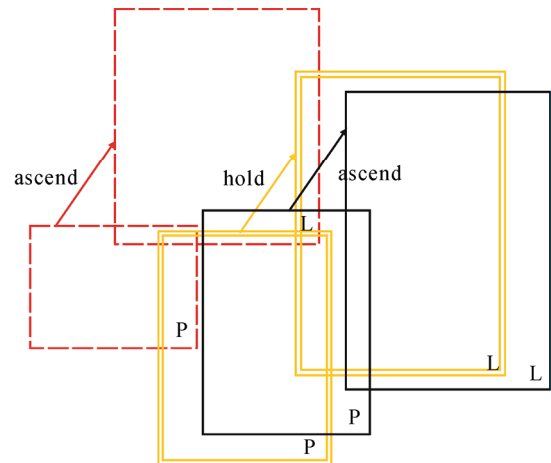


图5 匹配策略示意图

Fig.5 Schematic of matching strategy

式中: θ 代表一对一匹配的解空间; L_{cost} 代表损失成本。本文在完成场景元素交互关系实例与预测场景元素交互关系组的一对一最优匹配后, 也利用式(3)对匹配对之间的网络损失进行了计算。虽然两个过程用的损失函数相同, 但理论上它们的超参数有所区分, 可能会出现不同的最优解。在本文的实际操作中, 由于超参数的搜索空间较庞大, 会带来较高的计算成本, 本文使用了与目标检测算法 DETR 相同的方法。

3 实验结果与分析

3.1 数据描述

为了验证所提方法对带电作业场景元素交互关系精细化辨识的可行性, 本文以登高作业为例, 选取登梯、扶梯、抬梯、在梯旁 4 种交互关系作为检测对象, 并以某变电站的现场影像作为自建数据集的数据源。

为提高关系检测模型的泛化性, 本文选取的带电作业影像包含了不同作业地点及作业时间。考虑到带电作业影像前后帧的连贯性, 对带电作业影像每隔 30 s 进行抽帧, 并对相似度较高、遮挡严重的负样本进行适当剔除, 为了提高特征提取网络的识别精度, 本文使用大型公开数据集 COCO 对其中的视觉特征提取模块进行预训练并对正样本图像进行亮度增强等预处理操作, 最终得到 1 155 张样本图像(每张样本图片中包含多种场景元素交互关系)。同时, 本文将随机选取样本图像的 80%作为训练集, 20%作为验证集, 即 924 张图像作为训练集, 231 张图像作为测试集, 自建数据集中各交互关系的图片数量及目标数如表 1 所示(由于 1 张图片中存在多种不同关系目标类型, 因此不同关系目标类型图片会有所重叠, 故各目标类型图片的数值总和远大于总计图片数)。

3.2 实验环境描述

为了测试所提基于 HOI 集合预测的带电作业场景元素交互关系解析方法的性能, 本文将所提方法在服务器中进行了模型的训练及相应的性能测试。服务器的具体配置如表 2。

3.3 模型参数设置及训练

本文在对自建数据集进行模型训练时, 目标类别分为 3 类(即工作人员、梯子及背景物), 关系类别分为 5 类(即登梯、扶梯、爬梯、在梯旁及背景关系)。因此, 本模型的类别数量 n_{class} 及动作数量 n_{action}

表 1 登高作业数据集具体说明

Table 1 Aerial work data set specifications						
目标 类型	训练集		测试集		总计	
	图片数	目标数	图片数	目标数	图片数	目标数
登梯	575	611	117	131	692	742
扶梯	276	327	69	74	345	401
抬梯	253	354	61	69	314	423
在梯旁	312	393	78	93	390	486
总计	924	1 685	231	367	1 155	2 052

表 2 实验环境配置说明

Table 2 Instructions for configuring the experimental environment	
配置名称	具体信息
服务器类型	DELL Precision T5820 GPU 服务器
CPU	i9-10980XE, 18 核, 3.0 GHz
GPU	2*RTX3090, 48 GB
内存	128 GB
硬盘	10 GB 固态硬盘
系统	Linux Ubuntu20.04

分别选取为 3 与 5。由于所提方法不存在非极大性抑制问题, 因此需要对预测的集合进行相应的分数排序。

根据 Detr 的经验, 本文将 top_k (表示按置信度排序后保留前 k 个最相关结果)取值为 100, 即将前 100 个分数较高的集合进行关系匹配。同时该模型将迭代次数 epoch 设置为 250, 模型预测标签的置信度设置为 0.5, 初始学习率 r_{learning} 设置为 10^{-4} , 并选择在 110 代逐步减小学习率以提高模型的精度。该模型主要的损失函数如上图所示, 其中图 6(b)与图 6(d)可用来衡量目标物的定位情况, 图 6(c)可用来衡量模型的分类情况。由上述损失图 6 可见, 该模型的损失值在 250 代的训练过程中逐步收敛, 最终趋于平稳。以此作为初始化的参数模型进行收敛性的分析, 本模型利用其损失函数来反映模型收敛性能。

3.4 消融实验

场景元素交互关系三元组各个元素的相对重要程度及匹配策略的正确选择是影响本文模型各类损失的关键因素。因此, 本节对场景元素的相对主导程度及匹配策略进行了不同消融实验(其中, NonRare 代表包含登梯动作的数据集部分, Rare 代表不包含登梯动作的数据集部分, Full 代表完整的数据集, 训练损失($\downarrow$)代表训练损失下降百分比, mAP 代表全类平均精度)。由式(3)可知, 本文的匹配成本可具体展开为:

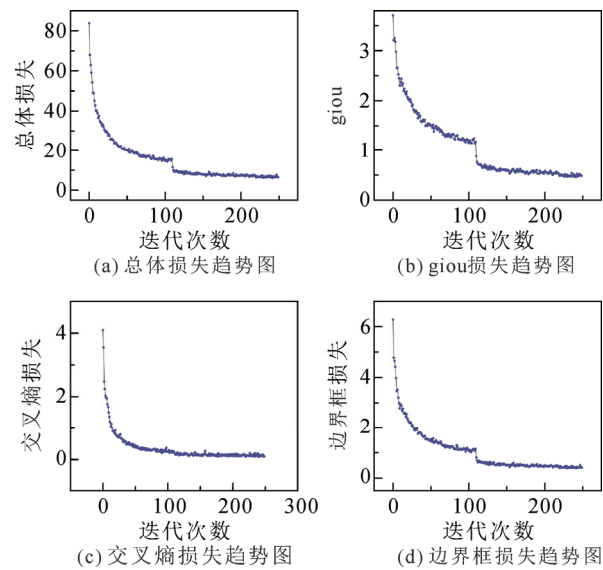


图 6 损失函数汇总图

Fig.6 Loss function summary plot

$$L_{\text{match}} = \beta_1(\alpha_p L_{\text{cls}}^p + \alpha_r L_{\text{cls}}^r + \alpha_l L_{\text{cls}}^l) + \beta_2 \sum_{k \in \{p, l\}} L_{\text{box}}^k \quad (5)$$

式中： α_p 、 α_r 、 α_l 分别代表训练过程中的人员权值、交互动作权值及交互对象权值的相对重要程度； β_1 、 β_2 则分别代表场景元素分类与场景元素定位的主导程度。本文的场景元素交互关系三元组(人员/交互动作/登高梯)可看为一个 2/4/2 的分类问题，但由于人员具有较多的训练样本且在每个关系三元组都有该元素的存在，故本文对人员的分类默认为较简单的分类(即设置 $\alpha_p=1$)。通过在 $\beta_1=\beta_2=1$ 的情况下，对 α_r 和 α_l 进行消融实验来验证交互动作与交互对象的相对重要程度。如表 3 所示，在 $\alpha_r=2$ ， $\alpha_l=1$ 时，模型实验效果达到最佳。因此，交互动作与交互对象相比对本文所提模型具有较大的影响。

同时考虑到图 5 的情况，本文在 $\alpha_r=\alpha_l=1$ 的情况下也对图 5 中的两种匹配情况进行了相应的消融实验。由式(5)可知，当 $\beta_1=1$ ， $\beta_2=1$ 时，关系三元组的分类与目标对象的定位具有相同的主导程度，即在该情况下两种匹配情况对模型具有相同的影响效果。当 $\beta_1=1$ ， $\beta_2=2$ 时，目标对象的定位具有较高占比，即此时目标对象的定位具有较高的主导程度，模型将会优先把双实线交互对分配给交互关系实例。反之，当 $\beta_1=2$ ， $\beta_2=1$ 时，关系三元组的分类具有较高的主导程度，在该情况下，模型则会优先选择虚线交互对。如表 4 所示，当 $\beta_1=2$ ， $\beta_2=1$ 时，所提方法具有较好的实验效果。因此，在本模型的匹配策略中，场景元素的正确分类要比场景元素的准

表 3 分类损失系数 α 对模型性能影响分析

Table 3 Classification loss coefficient α impact analysis on

		model performance			
α_r	α_l	训练损失 (↓)	mAP/%		
			Full	Rare	NonRare
1.0	1.0	72.98	84.24	82.07	90.74
1.0	2.0	70.42	83.41	81.74	88.43
2.0	1.0	76.10	85.51	83.60	91.22

表 4 匹配成本系数 β 对模型性能影响分析

Table 4 Matching cost coefficients β impact analysis on model performance

β_1	β_2	训练损失 (↓)	mAP/%		
			Full	Rare	NonRare
1.0	1.0	72.98	84.24	82.07	90.74
1.0	2.0	71.66	82.74	80.55	89.32
2.0	1.0	75.73	85.28	82.86	92.53

确定位更重要，模型将会优先把虚线交互对分配给交互关系实例。

3.5 模型检测效果对比分析

为了验证所提方法的有效性，本文选用 4 种具有代表性的图像分类算法与所提方法进行了不同场景下的实验对比。其中 Hog+SVM^[25-26]为经典的机器学习算法，该方法通过将方向梯度直方图与支持向量机相结合的方式实现了对图像特征的有效区分。RegNet、Swim transformer 与 Vision transformer 为典型的深度学习算法，其中 RegNet 通过引入网络设计空间的方式寻找出最佳的特征网络参数组合，进而实现了目标特征的准确辨识。Swim transformer^[27-28]与 Vision transformer 则是通过将卷积神经网络的局部感知能力与编-解码器的全局建模能力相结合的方式捕捉不同尺度图像特征，进而实现图像的精确分类。具体测试结果如图 7—图 9 所示。

如图 7 所示，在单动作场景下(该场景下仅有 1 种交互动作且背景物相对较少)，4 类对比方法均能正确识别出了图片中登高作业的交互关系，但 Hog+SVM 算法的置信度要明显低于其他几类深度学习算法，仅以 0.71 的置信度识别出图像的抬梯动作。与 4 类方法相比，本文所提方法有近 1 的置信度，且在准确判别交互关系的同时也对交互对象进行了较准确的预测，更有利于登高作业场景的智能理解。

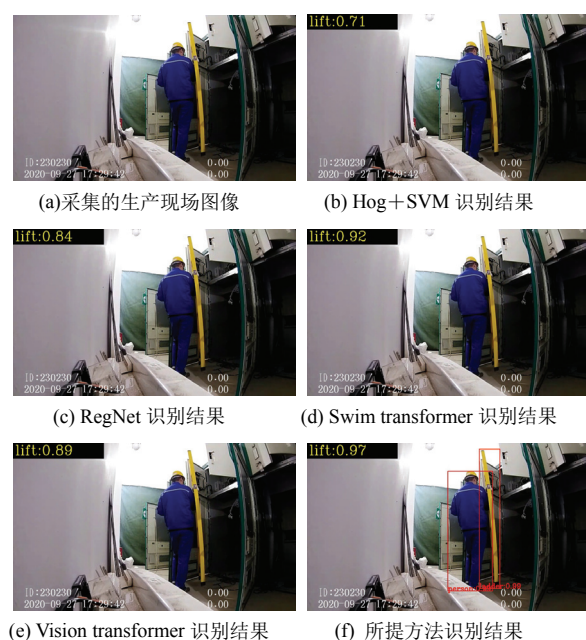


图7 单动作场景测试效果图

Fig.7 Single-motion scene test results



图9 暗光场景测试效果图

Fig.9 Dark light scene test results



图8 多动作场景测试效果图

Fig.8 Multi-motion scene test results

如图8所示,在多动作场景下(该场景具有多种交互动作且背景物相对复杂),由于人员遮挡与环境复杂等原因,Hog+SVM与RegNet方法仅正确识别出了图像中的一类交互动作,且置信度均低于0.9。Swim transformer与Vision transformer则正确识别出了图像中的所有交互动作,但两种方法并不能判断出同一交互动作的具体数量,且相对于单动作场景,其置信度均有所下降。与4种方法相比,所提

方法具有较好的测试效果,在正确识别出交互动作的同时,也对交互数量进行了有效辨识。

如图9所示,在暗光环境下,由于图像中的交互关系特征较为模糊,Hog+SVM、RegNet未能正确识别出图像中的交互动作。Vision Transformer混淆了图像中的前景目标与背景目标,导致模型识别出无效的背景动作。Swim transformer则是以较低的置信度识别出了图像中的交互关系。与之相比,本文所提的方法在正确区分出前景目标与背景目标的同时,仍能准确判断出图像中的交互关系。

综上所述,相比与其他4类算法,本文所提方法在不同的带电作业登高场景上均有较高的置信度与较好的识别效果。为了更好地验证所提方法的有效性,本文采用自建数据集(数据集具体描述见表1)对上述5种方法进行了相应的性能对比分析。由表5可知,Hog+SVM检测方法性能最差,对于交互关系类型的平均准确率仅65.90%,对于各个交互关系的检测准确率也比较低,其中登梯、扶梯、在梯旁3类动作的检测准确率均低于70%。RegNet检测方法与Hog+SVM相比检测效果有所提升,其平均准确率达到74.14%,但仍有较大不足。Vision transformer与Swim transformer的检测性能相当,相比于上述两种方法均有较大提升,其平均准确率也都达到85%以上,但在模型性能上Swim transformer要略高于Vision transformer。与以上4类检

表 5 登高作业关系判别结果对比

Table 5 Comparison of the results of the relationship between work at heights

所用方法	登高作业场景下不同交互动作的识别准确性/%				准确率/%
	Ascend	Hold	Lift lad-	Beside	
	ladder	ladder	der	ladder	
Hog+SVM	67.64	61.56	70.92	63.49	65.90
RegNet	76.92	71.23	78.85	69.54	74.14
Swim trans-former	92.08	86.92	84.54	81.46	86.25
Vision trans-former	90.77	84.62	87.08	78.92	85.35
本文所提方法	97.65	94.94	95.42	89.60	94.40

测方法相比, 本文所提方法对于 4 类不同交互动作的准确率均超过 85%, 且平均准确率达 94.40%, 对于数据量较多的登梯动作的识别准确率达 97.65%, 远远高于其他 4 种方法。由此可见, 本文所提出方法能够显著提高对登高作业场景元素交互关系的检测准确率, 极大地提高了模型对交互关系的智能理解能力。

3.6 实际场景测试效果

本文在识别出具体场景元素交互关系的前提下, 也进行了场景元素交互关系安全风险的研究。以登高作业场景为例, 根据交互关系的相互依存性可大致分为两类交互关系安全风险, 即: 1) 当检测到登梯动作时, 应伴随扶梯动作; 2) 当检测到抬梯动作时, 抬梯人数应不少于 2 人。

为此, 本文在不同的带电作业现场对场景元素交互关系安全风险进行了相应测试, 具体测试结果如图 10—图 12 所示。从测试图中可以看出, 本文所提出的基于 HOI 集合预测的带电作业场景元素交互关系解析方法在不同的登高作业现场下, 仍能准确的识别出具体的交互对象、交互动作及相应的安全风险, 进而验证了本文所提方法的有效性和优越性。

4 结论

1) 本文提出了一种基于 HOI 集合预测的带电作业场景元素交互关系解析方法, 将交互关系特征学习转换成对嵌入层关联特征向量的重新编码问题, 进而实现了对带电作业场景中场景元素交互关系的精准识别及安全风险的正确辨识。

2) 所提方法通过引入基于自注意力机制的并

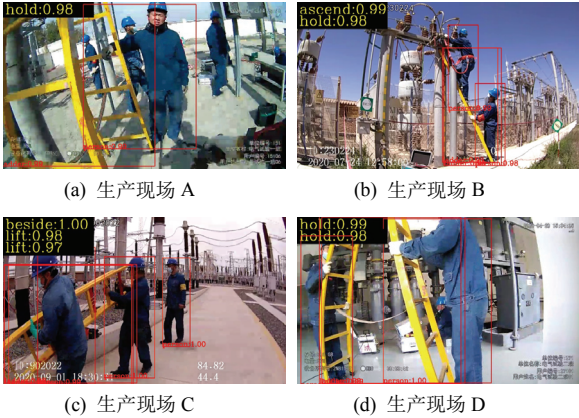


图 10 安全场景测试效果图

Fig.10 Diagram of security scenario test effect

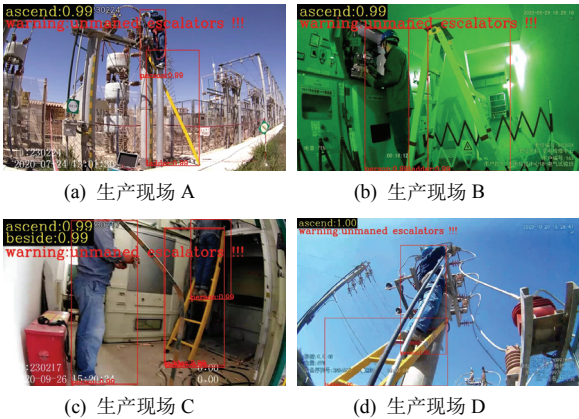


图 11 A 类安全风险测试效果图

Fig.11 Diagram of class A security risk test effect



图 12 B 类安全风险的测试效果图

Fig.12 Diagram of class B security risk test effect

行编-解码结构及集合预测模块, 将人物关系三元组合检测问题转换成了集合预测问题, 在简化模型结构的同时, 提高了模型对关系三元组全局特征的理解能力。

3) 实验结果表明, 所提方法在安全测试场景

与风险测试场景(无人扶梯、仅一人抬梯)下, 均以较高的置信度识别出了影像中的目标物与交互关系, 并有效辨识出了其中存在的风险类型, 具有较高的实用价值。

然而, 实际的带电作业生产场景复杂多样, 需综合考虑多个因素才能确保电力系统的安全、可靠和高效运行。针对不同带电作业生产场景, 同一作业类型的规章制度仍有差异, 搭建完整的电力场景规章体系, 将交互关系类型与作业场景细分化将是接下来研究的重点。

参考文献 References

- [1] 国家能源局. 2021 年二季度事故分析报告[EB/OL].[2024-06-09]. http://www.nea.gov.cn/2021-09/09/c_1310177594.htm. National Energy Agency. 2021 Second Quarter Accident Analysis Report[EB/OL].[2024-06-09]. http://www.nea.gov.cn/2021-09/09/c_1310177594.htm.
- [2] 薛明华, 艾春美, 律慧瑾, 等. 电厂安全帽佩戴安全性监控的智能图像处理[J]. 中国电机工程学报, 2022, 42(9): 3346-3353. XUE Minghua, AI Chunmei, LÜ Huijin, et al. Intelligent image processing technology for safety helmet wearing in power plant[J]. Proceedings of the CSEE, 2022, 42(9): 3346-3353.
- [3] 何国立, 齐冬莲, 闫云凤. 一种基于关键点检测和注意力机制的违规着装识别算法及其应用[J]. 中国电机工程学报, 2022, 42(5): 1826-1836. HE Guoli, QI Donglian, YAN Yunfeng. An illegal dress recognition algorithm based on key-point detection and attention mechanism and its application[J]. Proceedings of the CSEE, 2022, 42(5): 1826-1836.
- [4] 陈 汐, 韩译锋, 闫云凤, 等. 目标物智能跟踪与分割融合算法及其在变电站视频监控中的应用[J]. 中国电机工程学报, 2020, 40(23): 7578-7586. CHEN Xi, HAN Yifeng, YAN Yunfeng, et al. A unified algorithm for object tracking and segmentation and its application on intelligent video surveillance for transformer substation[J]. Proceedings of the CSEE, 2020, 40(23): 7578-7586.
- [5] 代少升, 曾 奇, 黄 炼, 等. 基于 S3-YOLOv5s 的矿井人员防护设备检测算法研究[J]. 半导体光电, 2023, 44(1): 153-160. DAI Shaosheng, ZENG Qi, HUANG Lian, et al. Research on detection algorithm of mine personnel protection equipment based on S3-YOLOv5s[J]. Semiconductor Optoelectronics, 2023, 44(1): 153-160.
- [6] 李 华, 吴立舟, 薛曦澄, 等. 基于计算机视觉的高处临边作业安全巡检[J]. 中国安全科学学报, 2023, 33(9): 69-75. LI Hua, WU Lizhou, XUE Xicheng, et al. Computer vision based safety inspection of high abutting edges[J]. China Safety Science Journal, 2023, 33(9): 69-75.
- [7] 王 波, 张迎晨, 齐冬莲, 等. 数字化安全管控视角下的全息影像: 定义、基本框架及关键技术[J]. 高电压技术, 2023, 49(8): 3335-3345. WANG Bo, ZHANG Yingchen, QI Donglian, et al. Holographic images for digital security control: definition, basic framework and key technologies[J]. High Voltage Engineering, 2023, 49(8): 3335-3345.
- [8] XU C, LI Q W, JIANG X B, et al. Dual-space graph-based interaction network for RGB-thermal semantic segmentation in electric power scene[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(4): 1577-1592.
- [9] 高 明, 左红群, 柏 帆, 等. 融合视觉关系检测的电力场景自动危险预警[J]. 中国图象图形学报, 2021, 26(7): 1583-1593. GAO Ming, ZUO Hongqun, BAI Fan, et al. Visual relationship detection-based emergency early-warning description generation in electric power industry[J]. Journal of Image and Graphics, 2021, 26(7): 1583-1593.
- [10] KUZNETSOVA A, ROM H, ALLDRIN N, et al. The open images dataset V4: unified image classification, object detection, and visual relationship detection at scale[J]. International Journal of Computer Vision, 2020, 128(7): 1956-1981.
- [11] LIU W, BAO Q, SUN Y, et al. Recent advances of monocular 2D and 3D human pose estimation: a deep learning perspective[J]. ACM Computing Surveys, 2023, 55(4): 80.
- [12] 马富齐, 王 波, 董旭柱, 等. 电力工业安全影像解译: 基本概念与技术框架[J]. 中国电机工程学报, 2022, 42(2): 458-474. MA Fuqi, WANG Bo, DONG Xuzhu, et al. Safety image interpretation of power industry: basic concepts and technical framework[J]. Proceedings of the CSEE, 2022, 42(2): 458-474.
- [13] 刘开培, 李博强, 秦 亮, 等. 深度学习目标检测算法在架空输电线路绝缘子缺陷检测中的应用研究综述[J]. 高电压技术, 2023, 49(9): 3584-3595. LIU Kaipei, LI Boqiang, QIN Liang, et al. Review of application research of deep learning object detection algorithms in insulator defect detection of overhead transmission lines[J]. High Voltage Engineering, 2023, 49(9): 3584-3595.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing System. Long Beach, USA: Curran Associates Inc., 2017: 6000-6010.
- [15] SARKER I H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions[J]. SN Computer Science, 2021, 2(6): 420.
- [16] ZUO C, QIAN J M, FENG S J, et al. Deep learning in optical metrology: a review[J]. Light: Science & Applications, 2022, 11(1): 39.
- [17] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: a survey[J]. Computational Visual Media, 2022, 8(3): 331-368.
- [18] ATES H C, NGUYEN P Q, GONZALEZ-MACIA L, et al. End-to-end design of wearable sensors[J]. Nature Reviews Materials, 2022, 7(11): 887-907.
- [19] 马富齐, 王 波, 董旭柱, 等. 面向电力生产精细化风险解译的高度相似防护工具智能检测技术研究[J]. 中国电机工程学报, 2024, 44(3): 971-980. MA Fuqi, WANG Bo, DONG Xuzhu, et al. Research on intelligent detection technology of highly similar protection tools for refined risk interpretation of power production[J]. Proceedings of the CSEE, 2024, 44(3): 971-980.
- [20] 季佳仲, 吴俊勇, 王彦博, 等. 基于深度残差网络的电力系统暂态电压稳定评估[J]. 电网技术, 2022, 46(7): 2500-2509. JI Jiazhong, WU Junyong, WANG Yanbo, et al. Power system transient voltage stability assessment based on deep residual network[J]. Power System Technology, 2022, 46(7): 2500-2509.
- [21] 李海英, 孙 越, 张 笑, 等. 模态时频图与 ResNet50 融合的真空接触器故障诊断方法[J]. 高电压技术, 2023, 49(5): 1831-1840. LI Haiying, SUN Yue, ZHANG Xiao, et al. Fault diagnosis of vacuum contactor based on modal time-frequency diagram and ResNet50[J]. High Voltage Engineering, 2023, 49(5): 1831-1840.
- [22] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//16th European Conference on Com-

- puter Vision. Glasgow, UK: Springer, 2020: 213-229.
- [23] 舒胜文, 陈阳阳, 张梓奇, 等. 基于多维能力和知识图谱-多层感知机的变压器运行状态画像构建方法[J]. 电网技术, 2024, 48(2): 750-759.
- SHU Shengwen, CHEN Yangyang, ZHANG Ziqi, et al. Construction method for transformer operating state portrait based on multi-dimensional capability and knowledge graph-multilayer perceptron[J]. Power System Technology, 2024, 48(2): 750-759.
- [24] 张 驰, 郭 媛, 黎 明. 人工神经网络模型发展及应用综述[J]. 计算机工程与应用, 2021, 57(11): 57-69.
- ZHANG Chi, GUO Yuan, LI Ming. Review of development and application of artificial neural network models[J]. Computer Engineering and Applications, 2021, 57(11): 57-69.
- [25] ZAIDI S S A, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514.
- [26] BHARTI K, CERVERA-LIERTA A, KYAW T H, et al. Noisy intermediate-scale quantum algorithms[J]. Reviews of Modern Physics, 2022, 94(1): 015004.
- [27] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 9992-10002.
- [28] 何 敏, 秦 亮, 赵 峰, 等. 面向电力系统现场作业的安全风险管控智能检测算法[J]. 高电压技术, 2023, 49(6): 2442-2457.

HE Min, QIN Liang, ZHAO Feng, et al. Intelligent detection algorithm of security risk management and control for power system on-site operation[J]. High Voltage Engineering, 2023, 49(6): 2442-2457.



MA Fuqi
Ph.D.

Associate professor

马富齐

1994—, 男, 博士, 副教授
研究方向为防灾减灾与安全、电力安全影像解译和边缘智能
E-mail: xautmfq@xaut.edu.cn



LIU Yongwen
Corresponding author

刘永文(通信作者)

1998—, 男, 硕士生
研究方向为电力深度视觉关系检测
E-mail: lyw783013352@163.com

收稿日期 2024-05-11 修回日期 2024-07-24 编辑 卫李静