P&IDSeg-Net：一种面向管道仪器图的轻量化分割网络

王晨曦[1]，杨婷婷[2]，张涛[3]，邹雨辰[1]，邵会凯[1]，钟德星[1]
（1. 西安交通大学自动化科学与工程学院，710049，西安；2. 中国广核集团有限公司科技数字化部，518000，广东深圳；3. 中广核智能科技（深圳）有限责任公司，518000，广东深圳）

摘要：为了提高管道和仪器图（P&ID）智能分割的性能，提出了一种用于图纸分割的轻量化网络P&IDSeg-Net。首先，设计了通道减半的非局部注意力模块，在深层特征空间建立全局依赖，解决了图纸中管道与仪器的结构混淆问题；其次，用特征加法融合代替通道拼接进行跳跃连接，进一步减少参数量，适应小规模数据集；然后，基于P&ID中管道区域的低灰度特性，构建了视觉提示词启发的灰度掩膜加权损失，使网络聚焦前景关键区域；最后，收集和手工标注了P&ID_Pipe数据集，并在该数据集上进行了大量的实验。结果表明：所提出的P&IDSeg-Net网络仅需 2.181×10^7 的参数量，轻量化优势突出，同时平均交并比达到77.40%，相比于主流的分割方法，能够获得更优的性能。

关键词：图纸分割；非局部注意力；视觉提示词；轻量化网络
中图分类号：TP39 **文献标志码**：A
DOI：10.7652/xjtuxb202605022 **文章编号**：0253-987X(2026)05-0226-11

P&IDSeg-Net: A Lightweight P&ID Segmentation Network

WANG Chenxi[1], YANG Tingting[2], ZHANG Tao[3], ZOU Yuchen[1],
SHAO Huikai[1], ZHONG Dexing[1]
(1. School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. Department of Digital Technology, China General Nuclear Power Group, Shenzhen, Guangdong 518000, China;
3. CGN Intelligent Technology (Shenzhen) Co., Ltd., Shenzhen, Guangdong 518000, China)

Abstract: To enhance the performance of intelligent segmentation for piping and instrumentation diagrams (P&IDs), a lightweight segmentation network named P&IDSeg-Net is proposed. First, a channel-halved non-local attention module was designed to establish global dependencies in deep feature space, addressing the structural confusion between pipelines and instruments. Second, feature addition fusion was employed instead of channel concatenation for skip connections, further reducing the parameter count and adapting to small-scale datasets. Then, based on the low-grayness characteristic of pipeline regions in P&IDs, a visual prompt-inspired grayscale mask weighted loss was constructed to focus the network on key foreground areas. Finally, the P&ID_Pipe dataset was collected and manually annotated, and extensive experiments were conducted on this dataset. The results show that the proposed P&IDSeg-Net requires only 2.181×10^7 parameters, demonstrating significant lightweight advantages, while achieving a mean intersection over union of 77.40%. Compared with mainstream segmentation methods, it achieves superior performance.

Keywords: P&ID segmentation; non-local attention; visual prompt; lightweight network

收稿日期：2025-06-25。 作者简介：王晨曦（2002—），女，硕士生；邵会凯（通信作者），男，助理教授，硕士生导师。
基金项目：中央高校基本科研业务费资助项目（xzy012023061）。
网络出版时间：2025-12-31 网络出版地址：https://link.cnki.net/urlid/61.1069.T.20251230.1738.006

在工业设施的全生命周期管理中，管道和仪器图（P&ID）等工程图纸是整个流程的核心，其包含设备间的连接关系、工艺逻辑等重要信息，既是施工建造的基准蓝图，又是设施运维阶段工艺优化、安全评估和资产管理的关键依据。然而，当前工业实践中，大量 P&ID 仍以非结构化的 CAD 文件、PDF 文件、纸质文档等形式存档，同时存在元数据缺失、符号标注不规范等问题，严重制约了图纸信息的机器可读性，使得工艺追溯、设备关联分析等数字化应用难以有效开展。然而，P&ID 包含数以千计的异构符号、密集交错的管道网络以及多层标注信息，呈现出结构复杂、语义多样、视觉相似等特点，其数字化方法具有独特的研究价值。

P&ID 的数字化研究起步于传统方法阶段。Ishii 等[1]基于传统方法进行了符号检测等基本任务。此后有学者基于传统计算机视觉技术开展 P&ID 数字化工作。例如，Elyan 等[2]通过启发式符号检测，对 P&ID 中的组成元素进行分类；Kuchuganov 等[3]用聚类方法对工程图纸的机器零件进行自动分组。这一阶段的研究虽然有所成果，但受制于人工特征设计，在面对复杂工程图纸时效果不理想。Moreno-grcía 等[4]描述了工程图数字化的动机和挑战，认为工程图纸的自动分析和处理远未完成，并呼吁在该领域引入最新的深度学习方法。

近年来，大量研究将深度学习方法应用于 P&ID 图纸符号检测、组件识别等任务。在符号检测方面，学者通过改进开集目标检测 YOLO-World[5]、微调目标检测 YOLOX 模型[6]，有效提高了变电所布局图纸等工程图纸的图符检测精度。在字符识别方面，基于深度卷积网络和不变矩的识别方法[7]，有效解决了传统方法在识别效率和对倾斜、变形字符适应性不足的问题。在结构化解析方面，将工程图纸转化为图结构，结合图匹配算法[8]、图神经网络[9]进行组件关联分析、制造方法分类等，为工程师的图纸维护和自动化报价系统提供了技术支持。在区域分割方面，研究者采用 Transformer 模型[10]、图卷积网络[11]实现了对图纸中 2D/3D 视图、加工要求、标题栏等关键区域的分割，以及对轮廓线、尺寸标注等的语义分割。然而，相较于对符号和组件等离散元素的识别任务，针对 P&ID 图中承载物质与能量传输功能的管道这一连续性结构的分割与识别研究较少。由于管道结构直接决定工艺逻辑的完整性，其识别对于图纸整体结构的理解具有关键作用，因此亟需开展更加系统且深入的研究工作。

图像分割作为计算机视觉中的核心任务，广泛应用于医学影像、遥感分析等多个领域。传统方法如阈值分割[12]和区域生长[13]受限于人工设计的特征表达能力，只能进行粗粒度的分割。随着深度学习的发展，基于卷积神经网络的分割方法逐渐成为主流。从全卷积网络的端到端像素级预测[14]，到反卷积的像素级类概率预测[15]，到 U-Net 对称编解码结构的提出[16]，再到 DeepLab 的空洞卷积[17]、PSPNet 的多尺度特征融合[18]等，这些方法不断提升了分割性能。近年来，基于 Transformer 架构的视觉模型通过全局建模展现出更强的语义理解与边界分割能力。Swin Transformer[19]通过分层设计和移位窗口机制实现高效全局建模，SegFormer[20]结合轻量级编码器与 MLP 解码器，在保持计算效率的同时提升了长距离依赖建模能力。在此基础上，Segment Anything（SAM）模型[21]在 1 100 万图像、10 亿掩码的超大规模视觉语料库上训练，在零样本分割迁移任务中表现惊人。其后续版本 SAM 2[22]进一步实现了图像和视频的统一分割，其视频分割能力代表了分割任务从静态向时序建模的重要突破。但基于 Transformer 的模型普遍面临参数量大、计算资源需求高的挑战，且 P&ID 中细长管道结构与常规自然物体形态差异导致特征提取偏差，SAM 模型在特定领域分割中存在显著领域适配问题。为降低 SAM 微调成本，已有研究如 SAM Adapter[23]等轻量化微调方式，通过在 ViT 模型主干中插入任务特定的适配层，在冻结主干参数的基础上，训练少量新增参数以适应特定任务。但由于图纸与预训练模型中的数据集差异较大，提示驱动机制效果有限，难以直接迁移到工程图纸分割任务。因此，尚未有模型针对 P&ID 图纸分割任务进行适配，亟需突破。

综上，P&ID 管线分割直接决定了工艺流程的逻辑完整性，是实现结构化解析和智能运维的关键。为满足工业数字化对高精度分割的需求，本文提出了一种兼具长程依赖建模能力与轻量特性的分割网络 P&IDSeg-Net，主要贡献与创新点如下。

（1）针对 P&ID 图中管道结构与仪器等组件的视觉相似性问题，P&IDSeg-Net 网络在编码器深层引入轻量级非局部模块，以进行长程依赖建模，增强对结构相似目标的辨别能力。同时引入通道压缩以降低计算开销，并借助残差机制无缝插入预训练网络。

（2）为缓解传统 U-Net 跳跃连接采用通道拼接所带来的解码器参数冗余，采用特征加法融合策略整合浅层细节与深层语义信息，显著减少了解码器

参数量，提升了网络对小规模 P&ID 数据集的适应性。此外，每级跳跃连接均可展开为级联与并行的编解码模块，增强了对管道尺度变化的鲁棒性。

(3)针对前景(管道、仪器)稀疏与大面积背景之间的样本不均衡问题，网络结合二者的灰度分布差异，将灰度先验作为视觉提示词引入到损失函数中，引导网络关注前景区域、屏蔽背景干扰，有效缓解了前景与背景的极端类别不平衡问题。

1 数据集构建

1.1 数据收集与预处理

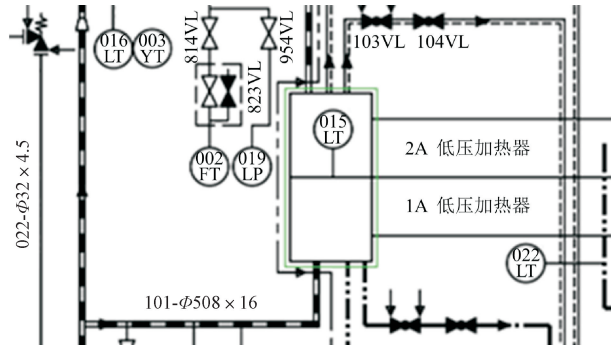
由于工程图纸涉及敏感信息且标注需专业知识，目前带标注的 P&ID 数据集较少。本文构建了 P&ID_Pipe 数据集，收集了 93 张 2019—2024 年多机组高分辨率图纸(平均分辨率为 4 962×3 508)，覆盖了给水、余热排出等多个关键子工作系统。

数据预处理用于消除图纸中的敏感信息，并对扫描的模糊图像进行增强。首先，采用光学字符识别(OCR)图中的项目代码、客户名称等敏感信息区域并填充背景脱敏；然后，针对清晰度有限的扫描图纸，采用非极大值抑制消除扫描伪影，输出无损 PNG 格式。图 1(a)为收集的图纸局部，其中数字和字母代表设备/管线型号。

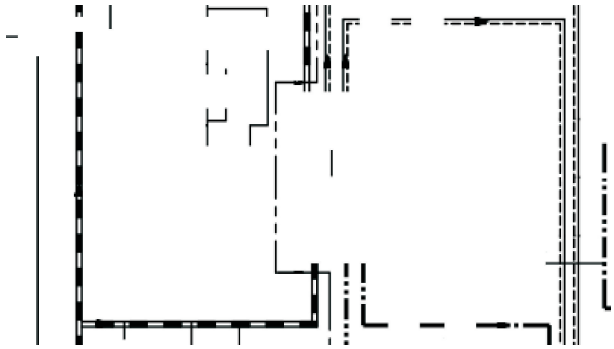
1.2 标签标注

机组 P&ID 包含泵、控制阀、流量计等设备符号，不同的设备符号通过管道段连接，组成不同的冷却循环、反应器进料等功能子系统。在图纸中，管道具有不同的样式，如粗管道、细管道、虚线管道、直线管道等，设备也分为大设备和小设备。特别地，大设备符号上往往存在较长的边缘，极易和管道混淆。为得到准确的 P&ID 管道标注图，本文采用双人独立标注-差异仲裁的工作流对图纸进行标注。所有标注者在正式标注前接受了标准化标注培训，整体标注时长近 500 h。最终的标签图仅包含管道对象，典型管道标注如图 1(b)所示。

本文构建的 P&ID_Pipe 数据集具有以下特点：①标签标注采用双人独立标注-差异仲裁，实现了精准管道区域标注；②大设备符号与管道之间易发生混淆，如图 1(a)中绿色标注所示，二者在形态特征上呈现高度相似性；③存在类别不平衡问题，经统计，P&ID 图纸中白色背景占比 91.86%，而管道区域仅占图像总面积的 1.55%，会直接影响图像分割模型对 P&ID 图纸管道结构的识别分割能力。



(a)P&ID 区域示例



(b)P&ID 标签示例

图 1 P&ID_Pipe 数据集中的图纸局部细节展示

Fig. 1 Local details of diagrams in P&ID_Pipe dataset

2 P&IDSeg-Net 网络架构

针对 P&ID 图纸的数据特点与任务需求，本文提出了 P&IDSeg-Net 网络，结构图如图 2 所示。P&IDSeg-Net 采用改进的 U-Net 编码器-解码器架构。首先，受限于数据规模，高参数量的 Transformer 结构在小数据集上存在过拟合风险，本文选用参数量适中的 ResNet34 作为编码器主干。其次，P&ID 图像结构线性特征显著，Transformer 的多层注意力机制在逐层传递特征时会丢失图像的细节线性信息，而卷积神经网络(CNN)结构可直接学习边缘、方向、形状等特征，保留线性特征，更适配 P&ID 线性结构特征。此外，针对管道实例中存在的显著多尺度特性，如介质管道的长程连续特征与信号线的短程分布特征，本文采用了跳跃连接的编码解码结构。通过并行多分支结构处理不同感受野的特征，捕捉到多尺度特性，同时可以有效减少解码器参数量。然后，针对特殊设备符号与管道容易产生混淆的问题，传统基于卷积神经网络的方法由于其局部感受野限制，难以建模长程依赖，所以本文创新性地在深层特征空间嵌入轻量化非局部注意模块，通过构建特征图内的全局关联，在不大幅增加计算复杂度的前提下，有效增强了网络对图像长程上下文信

息的建模能力。最后，针对管道等前景与背景区域类别的极度不平衡问题，本文在损失函数中引入基于 P&ID 图像特性的视觉提示词，使网络更加聚焦前景目标区域。

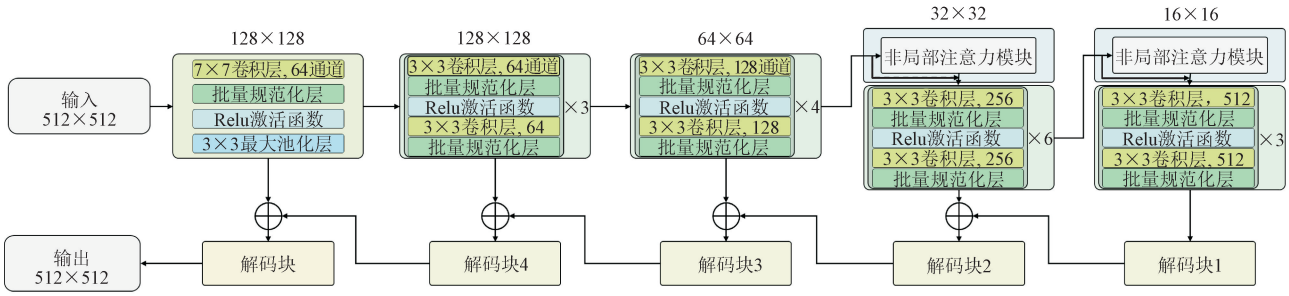


图 2　本文提出的 P&ID-IDSeg-Net 整体网络架构

Fig. 2　Architecture of the proposed P&ID-IDSeg-Net

2.1　整体架构

如图 2 所示，编码器使用 ResNet34 为主干网络，其占用的内存空间更小同时能够保持较优线性特征提取性能。具体地，该方法基于 ResNet34 前 3 阶段在浅层提取管道连接点、仪器边缘等局部细节信息，在第 4 和 5 阶段引入非局部注意力模块，依赖其强大的长程依赖建模能力，以解决 P&ID 图纸分割中管道结构与特殊仪器符号间的易混淆问题。在更深层引入非局部注意力模块，可以映射更大的感受野，以更好地利用上下文获取深层语义信息，促进分割易混淆的管道和部分特殊仪器符号。同时，深层特征图低分辨率更小（输入尺寸分别为原图的 1/8、1/16 尺寸），能够显著降低计算复杂度。解码器采用转置卷积，严格保持编码-解码阶段的数量对称性，通过 5 次上采样逐步恢复空间分辨率。编码器与解码器之间采用改进的跳跃连接机制，在提取 P&ID 图像多尺度特征信息的同时有效减少参数量，减少传统拼接方式带来的解码器参数开销。

2.2　非局部注意力模块

如图 1 绿色框所示，P&ID 图像中部分管道结构与特殊仪器符号存在高度视觉相似性，而传统方法往往依赖局部信息进行分类，因而处理此类高相似区域时，常出现误判或漏检。为解决这一问题，本文引入了非局部操作[24]来增强网络对长程依赖关系的建模能力，提升网络对结构相似区域（如管道与仪器）之间差异的建模能力。具体来说，该操作将输出特征图中每一个位置的响应表示为输入特征图中所有位置特征的加权求和，从而实现跨空间位置的全局信息交互。输出特征图为

$$\boldsymbol{y}_i = \frac{1}{C(x)}\sum_{j=1}^{N} f(\boldsymbol{x}_i,\boldsymbol{x}_j)g(\boldsymbol{x}_j) \tag{1}$$

式中：$C(x)$ 为归一化因子，$C(x)\in\mathbf{R}^+$；N 为输入特征图的位置素数（空间维度）；$i\in N$ 是输出特征图的位置索引；$j\in N$ 是所有可能输入位置的索引；$\boldsymbol{x}\in\mathbf{R}^{N\times c_1}$ 是输入特征图；$\boldsymbol{y}\in\mathbf{R}^{N\times c_2}$ 是输出特征图，其中 c_1、c_2 分别为输入和输出通道数；$\boldsymbol{x}_i,\boldsymbol{x}_j\in\mathbf{R}^{c_1}$ 表示输入特征图在位置 i 和位置 j 的特征向量；$\boldsymbol{y}_i\in\mathbf{R}^{c_2}$ 是输出特征图在位置 i 的特征向量；双变量函数 f：$\mathbf{R}^{c_1}\times\mathbf{R}^{c_2}\rightarrow\mathbf{R}$ 计算位置 i 和 j 特征之间的相关性标量；一元函数 g：$\mathbf{R}^{c_1}\rightarrow\mathbf{R}^{c_2}$，将输入特征 x_j 变换为 c_2 通道。最后，由 $C(x)$ 进行归一化，确保稳定性。

函数 f 和 g 有几种选择。函数 g 采用简单线性嵌入形式：$g(\boldsymbol{x}_j)=\boldsymbol{W}_g\boldsymbol{x}_j$。其中，$\boldsymbol{W}_g\in\mathbf{R}^{c_2\times c_1}$ 为可学习的权重矩阵。对于双变量函数 f，点积关系如下

$$f(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{x}_i)v(\boldsymbol{x}_j)=(\boldsymbol{W}_u\boldsymbol{x}_i)^{\mathrm{T}}(\boldsymbol{W}_v\boldsymbol{x}_j) \tag{2}$$

式中：$\boldsymbol{W}_u,\boldsymbol{W}_v\in\mathbf{R}^{c_2\times c_1}$ 是权重矩阵；u，v：$\mathbf{R}^{c_1}\rightarrow\mathbf{R}^{c_2}$，此时使用 $C=N$ 作为点积成对函数。高斯函数为

$$f(\boldsymbol{x}_i,\boldsymbol{x}_j)=\mathrm{e}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j} \tag{3}$$

将二者融合，即对 f 的高斯函数使用点积相似性，嵌入式高斯函数

$$f(\boldsymbol{x}_i,\boldsymbol{x}_j)=\mathrm{e}^{\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{x}_i)v(\boldsymbol{x}_j)}=\mathrm{e}^{(\boldsymbol{W}_u\boldsymbol{x}_i)^{\mathrm{T}}(\boldsymbol{W}_v\boldsymbol{x}_j)} \tag{4}$$

对于(3)和(4)，归一化因子均为 $C=\sum_{j=1}^{N}f(\boldsymbol{x}_i,\boldsymbol{x}_j)$。

为减少信息损失，网络通过残差连接将非局部操作的结果与原始输入特征融合，构成完整的非局部注意力模块，表达式为

$$\boldsymbol{z}_i=\boldsymbol{W}_z\boldsymbol{y}_i+\boldsymbol{x}_i \tag{5}$$

式中：$\boldsymbol{y}_i$ 由式(1)计算得到；$\boldsymbol{W}_z\in\mathbf{R}^{c_1\times c_2}$ 是权重矩阵，用于调整通道维度。同时，残差连接的引入可以使非局部注意力模块能够灵活地嵌入至任意预训练网络的中间层，实现即插即用。在实现上，通过将 $\boldsymbol{W}_z$ 初始化为零，即可确保插入非局部注意力模块不改变网络行为，从而实现平滑集成与稳定训练，便于嵌入预训练模型，增强模块的可扩展性。

图 3 是非局部注意力模块结构示意图，图中 C_h、H_t、W_d 分别为特征图的通道数、高度、宽度。由于 P&ID 图像是二维数据，所有权重矩阵变换 $\boldsymbol{W}_u$、$\boldsymbol{W}_v$、$\boldsymbol{W}_g$、$\boldsymbol{W}_z$ 采用 1×1 卷积。此外，为了在网络中引入更多非局部注意力模块的同时控制计算开销，本文设置了一个通道压缩超参数 $c_2 = c_1/2$，即在非局部操作内部，将通道数减半。该策略能够显著降低模块的计算复杂度，使在整体网络架构中能够部署多个非局部注意力模块。

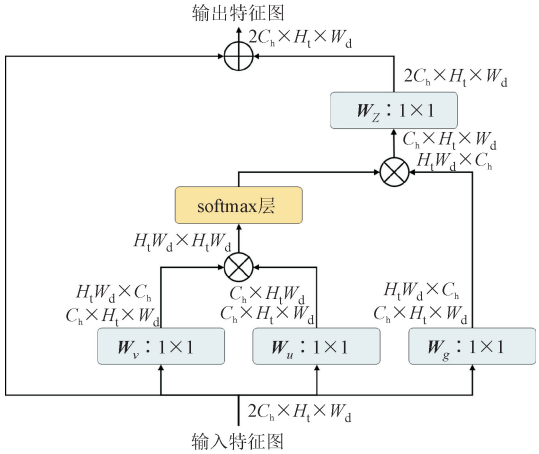


图 3 非局部注意力模块结构

Fig. 3 Non-local attention module structure

2.3 基于特征图加法融合的跳跃连接

传统 U-Net 架构采用通道拼接进行跳跃连接，导致解码器输入通道数翻倍，进而增加解码器参数量。为进一步实现 P&ID 图纸分割的网络结构轻量化，本文摒弃了传统编码器-解码器结构中跳跃连接所普遍采用的通道拼接策略，改用特征图加法融合机制，加法融合保持通道数不变，能够大大减少解码器额外的参数，进一步适应小规模 P&ID 数据集。

同时，编码器与解码器结构的对称性，确保了特征映射在空间维度上的一致性，使网络避免了由于层间深度差异导致的信息不匹配问题，相反可有效抑制恢复过程中的信息损失，有助于管道结构的拓扑连续性与完整性。此外，P&ID 图纸存在显著多尺度特征，网络中的每一跳跃连接均可展开如图 4 所示的子结构，包含具有级联模式与并行结构的编码与解码模块。网络通过多条路径引入不同感受野，可以联合提取多尺度特征信息，增强网络在处理尺度变化管道实例时的适应能力。

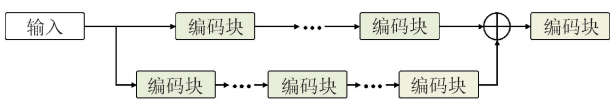


图 4 具备级联与并行的跳跃连接示意图

Fig. 4 Schematic of the skip connections with cascade and parallel structures

2.4 基于视觉提示词的损失函数

在语义分割任务中，常用的损失函数为二值交叉熵损失 L_{BCE}，其形式为

$$L_{BCE} = -[t\lg p + (1-t)\lg(1-p)] \tag{6}$$

式中：$p \in [0,1]$ 表示网络预测的像素概率；$t \in \{0,1\}$ 表示真实标签。

然而，如 1.3 节所述，P&ID 图像中前景（如管道、仪器）、背景区域存在 1∶59.3 的像素级极端类别不平衡问题，类别不平衡比率 γ 为

$$\gamma = \frac{N_{fg}}{N_{bg}} \approx 0.016\,9 \tag{7}$$

式中：N_{fg}、N_{bg} 分别表示前景（管道 / 仪器）与背景的像素数。背景像素在总像素中占据绝对多数，在这种情况下，直接使用二值交叉熵损失（BCE）会因背景主导梯度更新，而难以有效学习前景区域的语义信息。

视觉提示词[25]是一类用于引导模型关注特定区域或语义信息的先验信息或引导机制，多应用于迁移学习与视觉问答等任务中。具体地，在迁移学习中，视觉提示词可用于快速调整预训练模型的关注焦点，使其适应新的数据分布。在视觉问答任务中，视觉提示词则引导模型关注具有回答语义的区域，以提升推理能力与回答准确率。

相应地，启发于视觉提示词引导模型注意关键区域的核心思想，本文将其引入到图像分割任务类别不平衡问题建模中，提出了一种基于灰度值的视觉提示词机制，用于调节损失函数的关注区域，使分割模型关注少数类前景中的特征学习以更好区分管道及仪器。P&ID 图像中，管道、仪器等前景目标在灰度图中呈现出低灰度值（接近黑色），而背景区域则多为高灰度值（接近白色），基于这一先验知识，本文利用灰度图中的像素值构建损失权重掩码，仅对低灰度值区域（前景）计算损失，从而显著增强了网络对少数却关键的前景区域的感知能力，抑制大量无关的高灰度背景区域，从而有效应对前景与背景之间的极端类别不平衡问题。具体地，掩码的构建方式如下

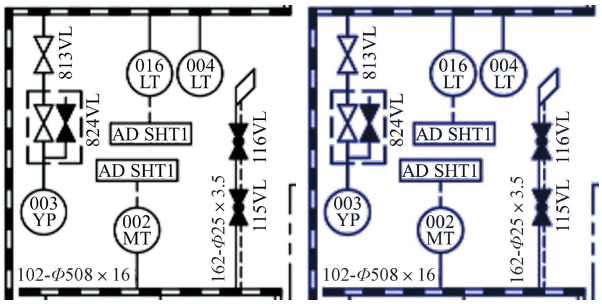
$$M(i,j) = \begin{cases} 1, & I(i,j) < \tau \\ 0, & \text{其他} \end{cases} \tag{8}$$

式中：$I(i,j)$ 表示图像中第 i 行第 j 列像素的灰度值；$\tau \in [0,1]$ 表示设置的灰度阈值，本文设置 $\tau = 0.9$；$M(i,j)$ 为掩膜中像素 (i,j) 对应的位置。

在此基础上，本文提出的基于视觉提示词的损失函数可表示为

$$L_{\text{masked}}=\frac{\sum_{i,j}L_{\text{BCE}}(p_{i,j},t_{i,j})M(i,j)}{\sum_{i,j}M(i,j)} \tag{9}$$

如图 5 所示，本文所提出的掩码机制有效引导网络聚焦于管道、仪器及周边关键区域（图中蓝色所示区域），从而在损失计算中有选择性地屏蔽大量无关的背景区域。该策略显著提升了训练过程中的关注集中性，使网络更加专注于前景目标的精确分割，有效提升了分割性能与训练收敛效率。



(a)P&ID 区域示例　　(b)损失函数作用区域示例

图 5　损失函数作用范围

Fig. 5　Scope of the loss function

3　实验与结果分析

3.1　数据集与数据扩充

本文所使用的数据集为本文构建的 P&ID_Pipe 数据集，共收集了 93 张高分辨率（4 962×3 508）P&ID 图纸。为了保证训练样本的丰富性和数量，数据集按照 7∶1∶2 的比例随机划分为训练集、验证集和测试集。

虽然在训练过程中，整幅图像会被划分为小块送入网络中训练，数据集数量大于原始样本数量，但由于原始样本数量有限，容易导致网络在训练过程中难以充分提取管道特征，出现过拟合问题。为缓解该问题，本文在训练阶段对训练集样本引入了多种数据增强方法，包括随机裁剪、随机放大、随机旋转（0°、90°、180°、270°，概率各为 0.25）、随机翻转（概率为 0.5）。上述操作同时作用于原图与对应标注，确保标签一致性。

3.2　实验环境与参数配置

实验所采用的软硬件环境如表 1 所示。为确保实验结果的可比性，所有实验均在统一的计算平台上进行。

表 1　实验硬件与软件环境配置

Table 1　Hardware and software environment configuration for the experiments

硬件或软件	版本
操作系统	Linux（Ubuntu 24.04）
CPU	Intel（R）Xeon（R）Gold 6148 CPU @ 2.40 GHz
GPU	NVIDIA GeForce RTX 3090（24 GB）
编程语言	Python 3.10
CUDA	11.8
深度学习框架	PyTorch 2.3.1

网络实现基于 PyTorch 框架，在训练过程中，采用 Adam 优化器，初始学习率设置为 0.001，主干网络采用在 ImageNet 上预训练的 ResNet34 权重进行初始化。输入图像尺寸为 512×512 的分辨率，批量大小设为 16，总训练轮数为 300，以确保网络充分收敛。此外，背景与前景的类别不均衡问题是数据的固有特性，因此所有对比模型在训练阶段均采用了本文提出的基于视觉提示词机制的损失函数，从而提升在前景类别上的分割性能，保证各模型在相同损失范式下的公平比较。

3.3　测试方法

由于图像分割任务不仅依赖于单个像素的灰度值，还高度依赖其周边像素的上下文信息，因此边缘区域往往由于上下文信息不足而导致分割精度下降。为缓解该问题，本文设计并采用了一种重叠滑窗测试策略，以提升模型在边缘区域的预测性能与整体分割结果的稳定性。

如图 6 所示，首先在待测图像四周各补充 128 像素的空白区域，以使边缘区域也处于子图的中心。随后，将整张图像划分为多个尺寸为 512×512 的重叠子图（与训练尺寸相同），相邻子图（横向与纵向）之间设定为 256 像素的重叠区域。在完成每个小块的预测后，仅保留其中心区域（大小为 256×256）作为有效输出区域进行拼接，来重建整幅图像的预测结果。该策略通过舍弃子图缺乏上下文信息的边缘区域，仅使用中心区域作为有效输出，在保证图像完整覆盖的同时，还有效缓解了因边缘上下文信息缺失所带来的预测误差，从而提升了网络整体分割性能。同样地，图像边缘问题也是所有模

型普遍面临的问题，因而本实验中所有模型在测试阶段均统一采用该策略，以确保评估结果的公平性与可比性。

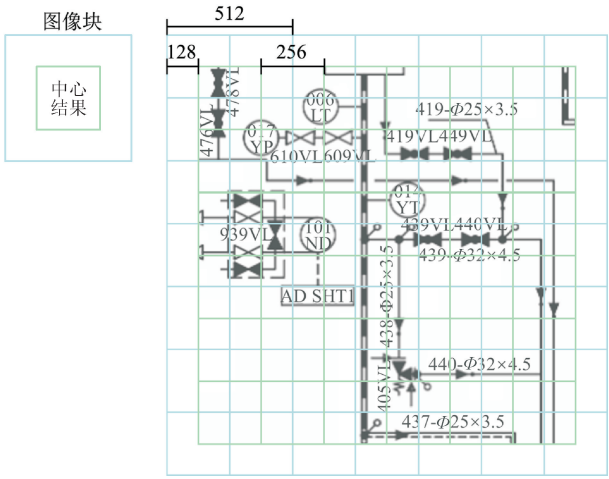


图 6 重叠滑窗测试策略

Fig. 6 Overlapping sliding window test strategy

为进一步提升测试鲁棒性与精度，本文在推理阶段引入了基于置信度加权的测试时增强（TTA）策略。该策略通过模拟训练过程中的增强操作，在保持网络结构不变的前提下，提升模型对图像变换的泛化能力。具体而言，TTA 在推理过程中对每张输入图像施加多种与训练阶段一致的增强方式，并将增强后的图像分别输入模型进行预测，获得多个预测结果。随后依据每个预测图的置信度对所有预测结果进行加权融合，从而得到更为稳定、鲁棒的最终预测。

假设在 TTA 过程中共获得 L 个增强预测结果 $\{P_1, P_2, \cdots, P_L\}$，其中每个预测图 P_l 的置信度定义为其所有像素预测值的绝对平均值

$$C_l = \frac{1}{H_t W_d} \sum_{i=1}^{H} \sum_{j=1}^{W} | P_l(i,j) | \tag{10}$$

式中：H_t 和 W_d 表示图像的高与宽。所有置信度经过 softmax 操作得到归一化权重

$$w_l = \frac{\exp(C_l)}{\sum_{l=1}^{L} \exp(C_l)}, l = 1, 2, \cdots, L \tag{11}$$

最终预测图 $\hat{P}$ 通过各增强预测图加权求和得到

$$\hat{P}(i,j) = \sum_{l=1}^{L} w_l P_l(i,j) \tag{12}$$

由此，该方法有效提升了预测的稳定性与精度。

3.4 实验结果与分析

为验证本文所提出方法的有效性，特别是非局部注意模块在建模远程依赖关系方面的作用，本文设计了一组基于不同输入尺寸的对比实验。鉴于非局部机制在感受野较大的情况下能够更有效地捕获图像中的全局上下文信息，实验分别选取了 3 种输入图像块（patch）尺寸：256×256、512×512 和 1 024×1 024，对所提出的 P&IDSeg-Net 网络进行训练与测试，并使用平均交并比作为性能评估指标。

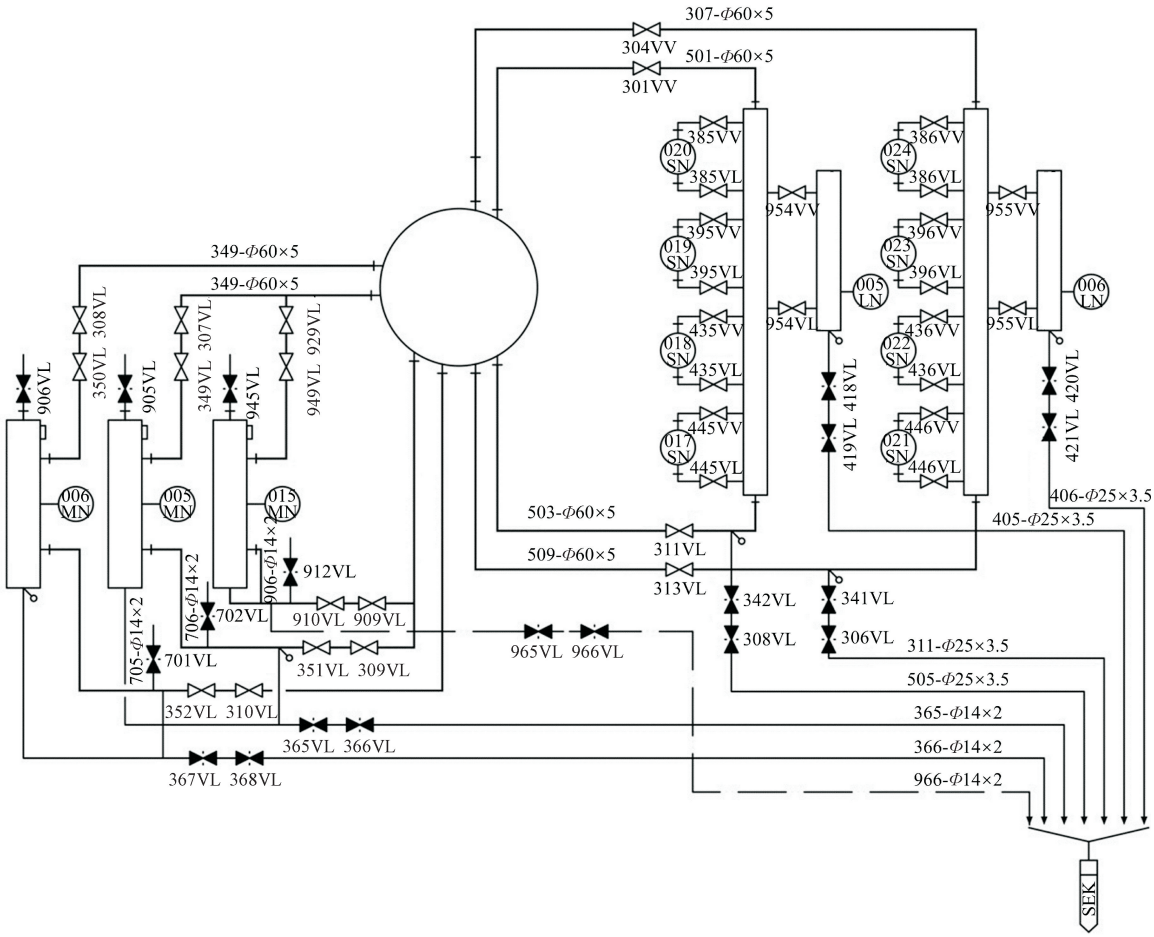
实验结果如表 2 所示。可以看出，随着输入图像块尺寸的逐步增大，P&IDSeg-Net 在未使用置信度加权 TTA 的情况下平均交并比呈持续上升趋势。其中，当图像块从 256×256 扩大至 512×512 后，平均交并比提升了 15.94%，表明小感受野限制了非局部注意模块对全局上下文的建模能力。而图像块从 512×512 扩至 1 024×1 024 时平均交并比达到最高值 79.01%，说明充足的感受野能够最大化发挥非局部机制优势。在引入置信度加权 TTA 后，各尺寸的平均交并比进一步提升，其中对 256×256尺寸上补偿效果最为明显（平均交并比提升 6.67%），验证了其在一定程度上弥补特征提取不足的能力。

表 2 不同图像块尺寸的实验结果

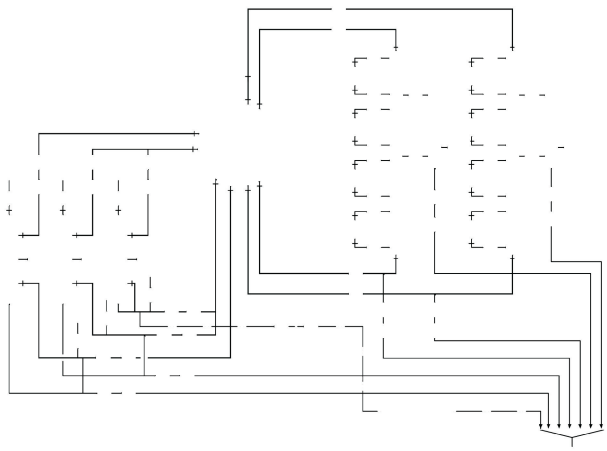
Table 2 Results of patches with different sizes

图像块尺寸	是否测试时增强	平均交并比/%
256×256	否	61.46
	是	68.13
512×512	否	77.40
	是	77.73
1 024×1 024	否	79.01
	是	79.30

为进一步验证所提出的重叠滑窗策略的有效性，同时整体展示本文模型的分割效果，采用重叠滑窗策略、选择输入图像块尺寸为 512×512 进行推理测试，该尺寸下的分割效果较 256×256 尺寸有显著提升，同时相比于 1 024×1 024 尺寸，又有较高的计算效率。如图 7(a)、(b)分别是同一张 P&ID 图纸的原图、本文模型的分割结果图。总体来说，模型能够准确区分管道与设备，整体分割结果具有较高的完整性和准确性，验证了所提出的重叠滑窗策略能够有效保障整幅图像的分割效果。



(a)P&ID 图纸原图



(b)P&ID 图纸本文模型分割结果

图 7　本文模型分割结果示例

Fig. 7　Segmentation results example by proposed model

3.5　对比实验

为验证本文所提出的 P&IDSeg-Net 在工业图纸管道分割任务中的有效性，本文对比了多种主流语义分割网络。由于尚无针对该任务的专用网络，故选取了具有结构相似性和任务相近性的代表性方法：用于医学影像血管分割的 U-Net[16] 及其变体（深层网络、浅层网络、在 bottleneck 中引入金字塔空洞卷积结构的网络）、用于遥感影像路网提取的 CoANet[26]、具有 Transformer 编码器与轻量级多尺度特征聚合的 Segformer[20]、融合 CNN 与 Transformer 的代表性架构 TransUNet[27]。考虑到上述实验结果，即 256×256 尺寸在感受野方面存

在明显不足，而 1 024×1 024 尺寸又带来过多的计算负担，因此所有对模型验统一采用 512×512 的输入尺寸，并不启用 TTA 增强，确保对比的公平性。

如表 3 所示，P&IDSeg-Net 取得了最高的平均交并比，达到了 77.40%，性能优于专注于道路结构提取的 CoANet（76.93%）与基线模型 U-Net（76.92%），同时优于引入深层结构、浅层结构或空洞金字塔结构的各类 U-Net 变体，也优于两个 Transformer 分割架构网络。

表 3 不同方法的对比实验结果

Table 3 Comparison results of different methods

方法	平均交并比/%
U-Net	76.92
深层 U-Net	76.88
浅层 U-Net	76.57
U-Net(金字塔空洞卷积)	76.32
CoANet	76.93
Segformer	76.91
TransUNet	76.18
P&IDSeg-Net	77.40

具体地，深层 U-Net 虽具备更强的特征表达能力，但在面对工业图纸中大量重复且密集的线状结构时，反而容易产生特征冗余，进而影响性能表现。浅层结构则由于特征提取层级较少，难以充分捕捉复杂语义关系，整体性能略低于标准 U-Net；引入 ASPP 模块虽然在一定程度上提升了全局上下文建模能力，但空洞卷积的使用可能破坏局部连续性，丢失细节信息；Transformer 架构网络虽具备强大的多尺度表达与长程依赖建模能力，但由于工程图纸结构高度规则、细节精密，其全局注意力机制在训练样本有限时较难准确聚焦于关键目标区域，整体性能未能超过本文方法，其中轻量级网络 Segformer 表现相对较好，而结构复杂的 TransUNet 参数量更大，训练更依赖数据规模，性能甚至低于U-Net。相比之下，P&IDSeg-Net 通过引入非局部注意机制，有效增强了网络对远程依赖关系的建模能力，使网络能够更准确地识别并分割细长、连续的管道目标，展现出更优的性能。实验证明了非局部机制在工业图纸语义分割中的有效性。

3.6 可视化实验

为了更加直观地展示本文提出的 P&IDSeg-Net 模型在 P&ID 图纸管道分割任务的有效性，选取不同方法的分割效果进行可视化对比，在模型推理图中，黑色表示模型正确预测的管道部分，红色区域表示漏检，即未能正确预测的管道，而蓝色区域则表示误判，即将仪器等错误识别为管道。

如图 8 所示，P&IDSeg-Net 有如下优势：P&IDSeg-Net 凭借非局部注意力机制对长程依赖的建模，能够在分割结果中有效保持管道结构的连续性。如案例 1 中，只有本文方法完整保留了该管道结构，Segformer 网络次之，但在管道连接处、边缘仍存在提升空间；P&IDSeg-Net 能获取更丰富的全局上下文信息，从而在易混区域中具备更强的判别能力。如案例 2，本文方法能够精准区分管道与形态相似的特殊仪器，Segformer 网络存在少量误判，而其他方法出现了较多的误判或漏检。P&IDSeg-Net 在处理斜向管道时仍存在改进空间，如案例 3 所示，本文方法的结果中斜向管道存在断裂问题，Segformer 虽然对斜向管道保存也较为良好，但存在较多的误判，而 CoANet 采用了方向自适应条状卷积模块，在该类目标的连续性分割方面更具优势，P&IDSeg-Net 当前的非局部操作则对方向敏感性不足。

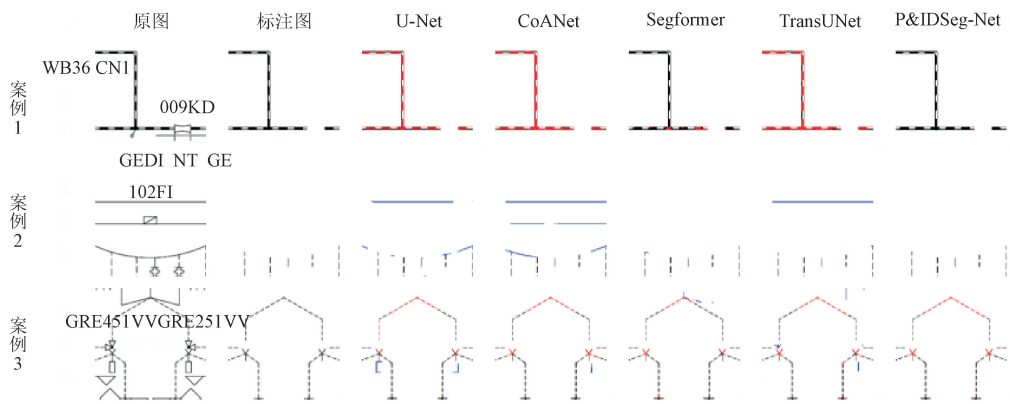


图 8 不同方法的性能可视化比较

Fig. 8 Visual results of performance for different methods

3.7　轻量化分析

为了评估不同方法在实际部署中的资源消耗和可用性，本文从参数量和浮点运算次数两个维度对比分析它们的计算复杂度，结果如表 4 所示。

表 4　不同方法的复杂度

Table 4　Complexity of different methods

方法	参数量/10^6	浮点运算次数/10^9
U-Net	31.04	218.65
深层 U-Net	124.37	270.74
浅层 U-Net	7.70	166.55
U-Net(金字塔空洞卷积)	17.14	169.78
CoANet	61.77	287.77
Segformer	13.71	13.76
TransUnet	88.07	111.10
P&IDSeg-Net	21.81	31.03

整体来看，P&IDSeg-Net 仅需 2.181×10^7 参数量，并可达到 3.103×10^{10} 的浮点运算次数，显著优于其他对比模型，展现出突出的轻量化优势。相比之下，深层 U-Net 由于网络层级较深，参数量高达 $1.243\,7\times10^8$，浮点运算次数达 $2.707\,4\times10^{11}$，网络模型庞大；浅层 U-Net 虽然参数量仅为 7.70×10^6，是所有模型中最少的，但浮点运算次数高达 $1.665\,5\times10^{11}$，轻量性并不理想，且前文实验显示其精度不高；U-Net(金字塔空洞卷积)在引入空洞金字塔结构后，参数量为 1.714×10^7，浮点运算次数接近 1.7×10^{11}，表明该结构虽然在模型规模上有所优化，但仍无法兼顾效率与精度；CoANet 作为采用方向自适应机制的网络，拥有较强的建模能力，但有 6.177×10^7 的参数量和 $2.877\,7\times10^{11}$ 的浮点运算次数，复杂度最高；U-Net 则处于中等水平，参数量为 3.104×10^7，浮点运算次数达 $2.186\,5\times10^{11}$，虽然结构简洁，但相比 P&IDSeg-Net 在资源消耗上依旧存在差距；SegFormer 作为轻量级 Transformer 的代表，具有最小的计算复杂度，在资源占用上具备显著优势。然而，其精度未能超过 P&IDSeg-Net；TransUNet 属于融合型重结构，参数量达 8.807×10^7，浮点运算次数为 1.111×10^{11}，远高于 P&IDSeg-Net，计算效率较差。

综上，P&IDSeg-Net 在保持精度领先的同时，以较低的参数量和计算开销实现了更高效的推理能力，充分体现了其结构设计的高效性与实用性，具备良好的工程部署潜力。

4　结　论

(1)本文提出的 P&IDSeg-Net 网络系统性适配图纸分割任务。引入非局部注意机模块，显著提升了网络的长程依赖建模能力，在保持结构连续性的同时，实现了对管道和特殊仪器符号的准确分割；用特征图加法融合代替通道拼接进行跳跃连接，有效减少解码器参数量，提高了网络在数据规模受限场景中的性能；设计了掩码加权损失，通过视觉提示机制引导网络聚焦于前景区域，有效缓解了类别不平衡问题。

(2)本文方法兼具精度与轻量化，具备工程应用潜力。所提出的网络参数量仅为 2.181×10^7，相较于 U-Net、TransUNet 等主流分割模型，参数量显著降低，同时具备更优的性能，其平均交并比最多高出 1.22%，在精度与可部署性之间取得了良好平衡，在工业图纸智能解析中具备良好的应用潜力与轻量化优势。

本研究可为工业图纸结构化解析提供参考，未来也可迁移到电气线路图、工艺流程图等工程图纸的智能解析。

参考文献：

[1] ISHII M, ITO Y, YAMAMOTO M, et al. An automatic recognition system for piping and instrument diagrams [J]. Systems and Computers in Japan, 1989, 20(3): 32-46.

[2] ELYAN E, GARCIA C M, JAYNE C. Symbols classification in engineering drawings [C]//2018 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ, USA: IEEE, 2018: 1-8.

[3] KUCHUGANOV V N, KUCHUGANOV A V, KASIMOV D R. Clustering algorithm for a set of machine parts on the basis of engineering drawings [J]. Programming and Computer Software, 2020, 46(1): 25-34.

[4] MORENO-GARCÍA C F, ELYAN E, JAYNE C. New trends on digitisation of complex engineering drawings [J]. Neural Computing and Applications, 2019, 31(6): 1695-1712.

[5] 范家斌，王宝会，陈继轩. 基于文本-图像多模态融合的变电所布局图纸图符检测方法 [J/OL]. 计算机科学. (2025-05-15)[2025-06-01]. https://link.cnki.net/urlid/50.1075.tp.20250514.1843.016.
FAN Jiabin, WANG Baohui, CHEN Jixuan. Method for symbol detection in substation layout diagrams based on text-image multimodal fusion [J/OL]. Computer Science. (2025-05-15)[2025-06-01]. https://link.cnki.net/urlid/50.1075.tp.20250514.1843.016.

[6] 章喻龙，罗壮强，石凌峰. 基于 YOLOX 算法的工程图纸标题栏识别 [J]. 数字技术与应用，2024，42(3)：198-202.
ZHANG Yulong, LUO Zhuangqiang, SHI Lingfeng.

外文标题 [J]. Digital Technology and Application, 2024, 42(3): 198-202.

[7] 侯凯, 梅诗妍. 基于特征提取和模板匹配的电网工程图纸字符识别技术 [J]. 沈阳工业大学学报, 2025, 47(2): 176-182.
HOU Kai, MEI Shiyan. Character recognition technology of power engineering drawings based on feature extraction and template matching [J]. Journal of Shenyang University of Technology, 2025, 47(2): 176-182.

[8] RICA E, ÁLVAREZ S, SERRATOSA F. Group of components detection in engineering drawings based on graph matching [J]. Engineering Applications of Artificial Intelligence, 2021, 104: 104404.

[9] XIE Liuyue, LU Yao FURUHATA T, et al. Graph neural network-enabled manufacturing method classification from engineering drawings [J]. Computers in Industry, 2022, 142: 103697.

[10] 汪伟. 基于人工智能的机械图纸图像切割技术的研究 [J]. 精密制造与自动化, 2025(1): 23-27.
WANG Wei. Research on mechanical drawing image segmentation technology based on artificial intelligence [J]. Precise Manufacturing & Automation, 2025(1): 23-27.

[11] ZHANG Wentai, JOSEPH J, YIN Yue, et al. Component segmentation of engineering drawings using Graph Convolutional Networks [J]. Computers in Industry, 2023, 147: 103885.

[12] OTSU N. A threshold selection method from gray-level histograms [J]. IEEE Transactions on Systems, Man and Cybernetics, 1979, 9(1): 62-66.

[13] HOJJATOLESLAMI S A, KITTLER J. Region growing: a new approach [J]. IEEE Transactions on Image Processing, 1998, 7(7): 1079-1084.

[14] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 3431-3440.

[15] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2015: 1520-1528.

[16] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham, Switzerland: Springer International Publishing, 2015: 234-241.

[17] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional Nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[18] ZHAO Hengshuang, SHI Jianping, QI Xiaojuan, et al. Pyramid scene parsing network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6230-6239.

[19] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 9992-10002.

[20] XIE Enze, WANG Wenhai, YU Zhiding, et al. SegFormer: simple and efficient design for semantic segmentation with transformers [C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2021: 12077-12090.

[21] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 3992-4003.

[22] RAVI N, GABEUR V, HU Y T, et al. SAM 2: segment anything in images and videos [EB/OL]. (2024-10-28) [2025-06-25]. https://arxiv.org/abs/2408.00714.

[23] CHEN Tianrun, ZHU Lanyun, DING Chaotao, et al. SAM-adapter: adapting segment anything in underperformed scenes [C]//2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway, NJ, USA: IEEE, 2023: 3359-3367.

[24] WANG Xiaolong, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7794-7803.

[25] JIA Menglin, TANG Luming, CHEN B C, et al. Visual prompt tuning [C]//Computer Vision-ECCV 2022. Cham, Switzerland: Springer Nature Switzerland, 2022: 709-727.

[26] MEI Jie, LI Roujing, GAO Wang, et al. CoANet: connectivity attention network for road extraction from satellite imagery [J]. IEEE Transactions on Image Processing, 2021, 30: 8540-8552.

[27] CHEN Jieneng, LU Yongyi, YU Qihang, et al. TransUNet: transformers make strong encoders for medical image segmentation [C]//2021 Interpretable Machine Learning in Healthcare. San Diego, CA, USA: ICML, 2021: 1-13.

(编辑 亢列梅)