

基于显著性先验和金字塔转换机制的 Trans-Sal Det 小目标油井检测模型

赵梓翔^{1,2,3}, 李佳慧^{2,4,5}, 步贤业^{2,3,4}, 穆树娟^{2,3,4}, 隋杨⁶

(1. 东北石油大学 陆相页岩油气成藏及高效开发教育部重点实验室, 黑龙江 大庆 163318; 2. 东北石油大学 人工智能能源研究院, 黑龙江 大庆 163318; 3. 东北石油大学 电气信息工程学院, 黑龙江 大庆 163318; 4. 东北石油大学 黑龙江省网络化与智能控制重点实验室, 黑龙江 大庆 163318; 5. 东北石油大学 数学与交叉科学研究中心, 黑龙江 大庆 163318; 6. 国家管网集团东北分公司 大庆维抢修中心, 黑龙江 大庆 163000)

摘要:为解决复杂地表环境与背景干扰条件下的油田遥感影像中油井小目标难以准确识别的问题,提出一种融合显著性先验与金字塔转换机制的 Trans-Sal Det 小目标油井检测模型。构建基于 Tiny-U-Net 的显著性生成模块,获取高分辨率显著性图,突出潜在油井区域并抑制冗余背景信息;将显著性图与原始遥感影像进行通道级拼接,输入金字塔式 Transformer 编码器,采用多尺度窗口自注意力机制,实现跨层特征建模,有效融合底层细节信息与高层语义特征;引入交叉尺度特征融合策略,增强模型对不同尺度油井目标的表征能力。在典型油田遥感数据集上开展对比与消融实验,对模型性能进行验证。结果表明:Trans-Sal Det 小目标油井检测模型在复杂背景、遮挡及低对比度场景下表现出更优的小目标检测能力,目标召回率与检测精度显著优于其他方法的。引入显著性先验可以有效引导注意力聚焦关键区域,提升 Transformer 对小目标的感知能力。该结果为油田自动化遥感检测与智能监测提供高效和可行的技术路径。

关键词:显著性先验;金字塔转换机制;交叉尺度融合;油井检测;动态显著性;Transformer;小目标检测

中图分类号:TE372 **文献标识码:**A **文章编号:**2095-4107(2026)02-0109-14

0 引言

早期油井识别主要依赖于人工判读与图像增强方法^[1],操作灵活,解释直观,但存在效率低、主观性强和批量推广难等问题。为提高自动化程度,人们采用基于规则的图像识别方法^[2-3],如边缘检测、形态学滤波和模板匹配等,可以识别规则清晰、目标特征明显的影像,但在地物遮挡、尺度变化和成像角度差异等情况下会出现识别偏差。通过引入支持向量机、随机森林等传统机器学习模型,构建特征空间与判别函数,实现目标识别并提升识别效果,但对人工特征的依赖较高,模型迁移能力和泛化能力不足。陈维力等^[4]对塔里木盆地某地区开展影像实验,使用灰度直方图阈值设定与边缘连接规则,实现近 80% 的初步检出率,但在植被干扰与设备混杂的区域误检率高。在苏里格气田山 1 段储层的某段影像中,因井场为非规则分布,边缘与背景无明显亮度对比,模板匹配法的准确率不足 40%^[5];在阿尔伯塔油区的某段冬季采集的影像中,积雪反光与建筑屋顶混淆,传统灰度差方法几乎失效^[6]。随深度学习的发展,显著性检测方法逐渐成为解决小目标定位与特征稀疏问题的重要途径。显著性网络通过模拟人类视觉注意机制,将显著区域从复杂背景中凸显出来,为后续目标检测提供先验信息。显著性图的卷积神经网络在遥感、医学与场景分割等领域显著提高检测性能。Wang C 等^[7]建立混合注意力显著性检测模型,在复杂背景下准确定位小目标区域。对于油井识别任务,显著性信息的引入强化空间注意力分布,显式抑制噪声干扰。这类基于 CNN(Convolutional Neural Network)的显著性增强方法受限于卷积的局部感受野^[8-9],难以实现全局特征依赖建模,在跨尺度、跨场景任务中性能不稳定。

近年来,Transformer 结构被广泛引入计算机视觉领域^[10-12],利用自注意力机制对全局上下文信息进行建模,突破传统卷积神经网络在感受野上的局限。在目标检测任务中,基于 Transformer 的检测框架实

收稿日期:2025-09-30;编辑:蔡田田

基金项目:国家自然科学基金面上项目(62573112);黑龙江省自然科学基金优秀青年基金项目(YQ2023F003);黑龙江省博士后特别资助项目(LBH-TZ2505)

作者简介:赵梓翔(1995—),男,博士研究生,主要从事油气信息与控制工程方面的研究。

现从特征提取到目标预测的端到端建模^[13],避免锚框设计与后处理步骤,为复杂场景下的目标识别提供新的技术路径。遥感影像具有分辨率高、地物类型复杂且尺度变化显著的特点,采用结合金字塔转换机制的 Transformer 模型,构建多尺度特征表示与分层编码机制,实现从局部细节到全局语义的逐级融合^[14],有效提升模型对不同尺度目标的感知能力。现有方法在特征建模过程中对显著性信息的利用不充分,注意力分布缺乏明确的空间引导,容易受复杂背景与噪声干扰,导致在小目标检测任务中出现目标响应弱、定位不稳定等问题,检测精度与鲁棒性需进一步提升。

笔者提出一种融合显著性先验与金字塔转换机制的端到端 Trans-Sal Det 油井检测模型。在显著性预处理阶段,设计轻量化 Tiny-U-Net 结构,生成高分辨率显著性图,根据显式建模潜在目标区域的空间分布,对原始遥感影像进行语义引导,强化目标响应并抑制无关背景干扰。在特征编码阶段,将显著性图与多通道遥感影像进行通道级融合,输入金字塔式 Transformer 编码器,构建多尺度 token 序列,利用分层窗口自注意力机制实现跨层特征交互与长距离依赖建模。在解码阶段,将显著性响应以动态先验形式注入查询向量,采用调制查询与键之间的相似度计算过程,使得注意力权重随空间显著性自适应调整,引导模型聚焦于潜在目标区域;在多头自注意力中引入信息熵约束项,对注意力分布进行正则化优化,在显著区域内收缩并提高分布稀疏性,从而有效抑制跨尺度冗余信息与背景噪声干扰。

1 模型建立

Trans-Sal Det 模型框架见图 1,在整体结构上可见数据流向。输入原始图像后,网络分为两条并行分支:显著性图生成模块分支采用轻量 Tiny-U-Net 编解码语义分割结构,生成高分辨率显著性图并提取显著性偏置,为网络提供前景目标的先验引导;主干网络分支采用 Res Net 或 Mobile Net 作为特征提取 backbone,完成深度特征提取与多层次语义特征抽取^[14],输出 1/4、1/8、1/16 尺度的多尺度特征图,覆盖图像的局部细节与全局语义信息。将显著性图生成模块分支输出的显著性偏置与主干网络分支的多尺度特征图进行元素级融合,利用显著性图对多尺度特征进行加权调制,实现前景目标区域特征响应的增强与背景干扰的抑制,完成显著性先验与语义特征的深度融合。深度融合的特征输入金字塔式 Transformer 编码器,利用多尺度自注意力机制对特征进行逐级编码,有效捕获特征的全局上下文依赖关系,提升网络对复杂场景下目标的特征表达能力。将显著性先验信息注入交叉注意力机制,结合可学习的查询向量,引导解码器聚焦前景目标区域^[15],强化目标特征,过滤背景噪声,实现准确的目标区域注意力聚焦。通过信息熵正则约束注意力分布,避免注意力退化为均匀分布或过度集中^[16],保证注意力机制的有效性;由预测头输出目标的边界框坐标与类别标签,完成端到端的目标检测任务。

1.1 多尺度特征及 token 分支

与传统方法将图像切分为统一尺寸的补丁^[17-18]不同,Trans-Sal Det 模型保留主干网络的原始空间金字塔信息,在不同尺度上独立建模并保证上下文依赖关系和空间聚焦能力。假设输入图像的高度为 H ,宽度为 W ,图像大小为 HW ,主干网络输出三组不同分辨率的多尺度特征^[19-20],其表达式为

$$\begin{cases} F^{(1)} \in \mathbf{R}^{\frac{H}{4} \times \frac{W}{4} \times C_1}; \\ F^{(2)} \in \mathbf{R}^{\frac{H}{8} \times \frac{W}{8} \times C_2}; \\ F^{(3)} \in \mathbf{R}^{\frac{H}{16} \times \frac{W}{16} \times C_3}. \end{cases} \quad (1)$$

式中: C_1 、 C_2 、 C_3 为各尺度对应的通道维度, $C_1 \geq C_2 \geq C_3$ 。

将式(1)进行二维空间展平,构造二维序列表达式为

$$T^{(l)} = \text{Reshape}(F^{(l)}) \in \mathbf{R}^{N_l C_l}, N_l = \frac{H_l W_l}{s_l^2}, l \in \{1, 2, 3\}, \quad (2)$$

式中: $T^{(l)}$ 为第 l 层的 token 输入, $T^{(l)} \in \mathbf{R}^{N_l C_l}$; s_l 为第 l 层特征的下采样比例;Reshape(\cdot) 为张量重塑操作; N_l 为第 l 层 token 数量; H_l 、 W_l 分别为第 l 层图像的高度和宽度。

构造完成的三组 token(序列单元)分支为

$$\begin{cases} T^{(1)} \in \text{TransformerBlock}^{(1)}; \\ T^{(2)} \in \text{TransformerBlock}^{(2)}; \\ T^{(3)} \in \text{TransformerBlock}^{(3)}. \end{cases} \quad (3)$$

对 3 个不同尺度的 token 序列,分别送入对应的 TransformerBlock 并进行处理。

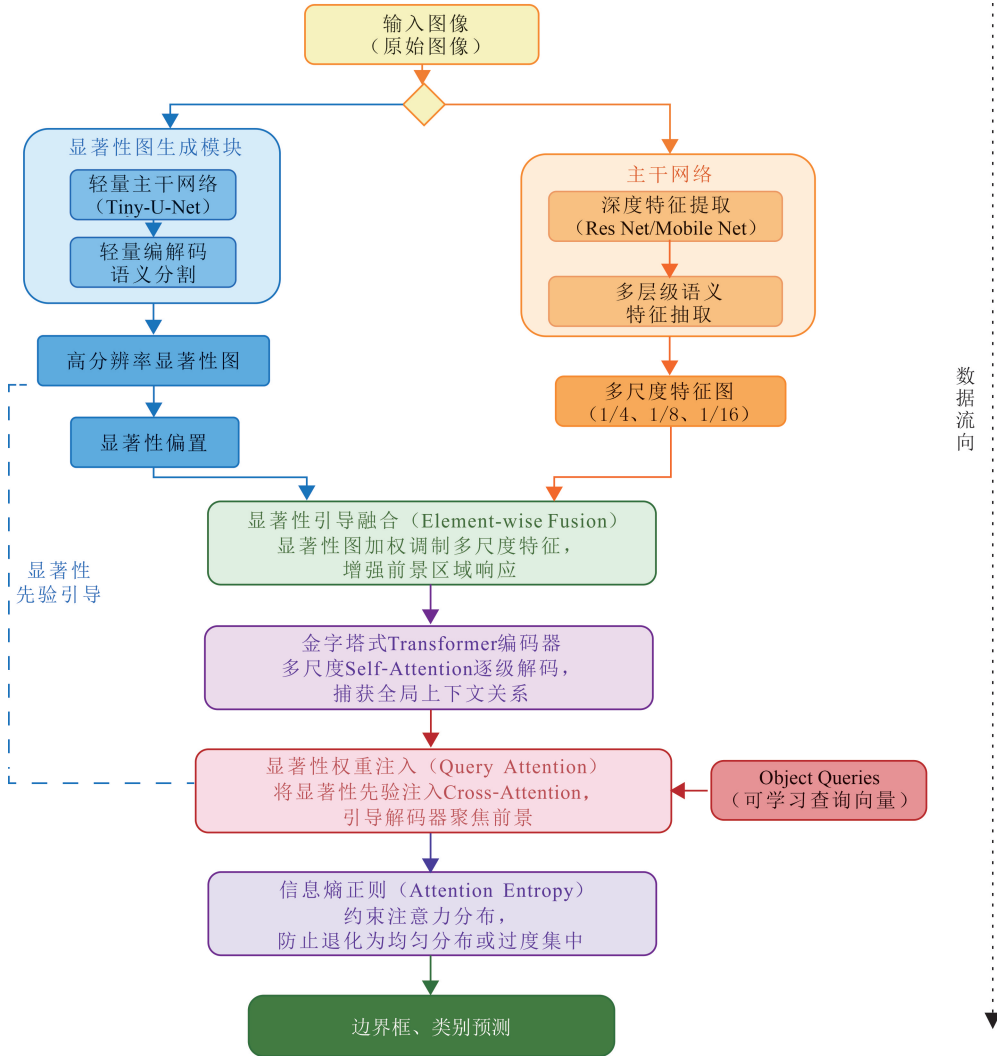


图 1 Trans-Sal Det 模型框架
Fig. 1 Trans-Sal Det model framework

1.2 多层显著性图嵌入机制

引入多层显著性图作为注意力引导项,构建可控的空间偏置。对于原始显著性图 $S \in R^{H \times W \times 1}$,分别下采样至与每层特征图空间尺寸一致,得到 3 个尺度的显著性图 $S^{(l)}$:

$$S^{(l)} = \text{Interp}_{\downarrow s_l}(S), S^{(l)} \in R^{\frac{HW}{s_l^2}}, l \in \{1, 2, 3\}, \quad (4)$$

式中: $\text{Interp}_{\downarrow}(\cdot)$ 为双线性插值下采样操作。将式(4)展平为显著性向量 $s^{(l)}$ 的形式:

$$s^{(l)} = \text{Reshape}(S^{(l)}) \in R^{N_l \times 1}. \quad (5)$$

假设有一个特征图 $F \in R^{N_l \times D}$,将显著性向量 $s^{(l)}$ 作为权重,与特征图 F 进行加权相乘,得到加权后的特征图 $\hat{F}^{(l)}$:

$$\hat{F}^{(l)} = F \odot s^{(l)}, \quad (6)$$

式中: \odot 为按元素乘法。

(1)特征加权层。对每一层 token 输入 $T^{(l)} \in \mathbf{R}^{N_l c_l}$, 将显著性图作为空间引导因子, 对每个 token 的特征向量进行点乘权重调整, 使用广播乘法可得

$$\tilde{T}^{(l)} = T^{(l)} \odot (1_{c_l} (1 + \gamma S^{(l)})), \quad (7)$$

式中: $\tilde{T}^{(l)}$ 为经过显著性加权后的第 l 层 token 特征; γ 为可学习的超参数; 1_{c_l} 为通道方向全 1 向量。

(2)注意力偏置层。在多头注意力计算中, 注意力偏置在 softmax 归一化前进行显式偏移调整。标准注意力得分 $A^{(l)}$ 的计算公式为

$$A^{(l)} = \frac{\mathbf{Q}^{(l)} (\mathbf{K}^{(l)})^T}{\sqrt{d}}, \quad (8)$$

式中: $\mathbf{Q}^{(l)}$ 、 $\mathbf{K}^{(l)}$ 分别为 Query 和 Key 的线性投影结果, $\mathbf{Q}^{(l)}$ 、 $\mathbf{K}^{(l)} \in \mathbf{R}^{N_l d}$; d 为每个头的特征维度。

引入一项基于显著性引导的加性偏置 $B^{(l)}$, 其表达式为

$$B_{i,j}^{(l)} = \eta \psi(S_j^{(l)}), \psi(x) = \log(1 + \varepsilon + x), \varepsilon > 0, \quad (9)$$

式中: $B_{i,j}^{(l)}$ 为第 l 层多头注意力中第 i 个 token 对第 j 个 token 的基于显著性的加性偏置项; $S_j^{(l)}$ 为第 l 层显著性图的第 j 个位置的显著性; η 为偏置强度参数; $\psi(\cdot)$ 为非线性映射函数; ε 为极小稳定平滑系数。

对显著性图进行变换并生成偏置项 $B^{(l)}$:

$$\tilde{S}^{(l)} = \psi(S^{(l)}), B^{(l)} = \eta \tilde{S}^{(l)}. \quad (10)$$

除显著性引导的偏置外, 考虑位置偏置 $P^{(l)}$, 将位置偏置与显著性偏置相加, 得到更新后的注意力得分:

$$\tilde{\mathbf{A}}^{(l)} = \mathbf{A}^{(l)} + B^{(l)} + P^{(l)}. \quad (11)$$

对于注意力得分 $\mathbf{A}^{(l)}$ 及其更新形式 $\tilde{\mathbf{A}}^{(l)}$, 加性偏置的引入导致注意力分布发生变化。考虑第 i 行的注意力权重为

$$a_{i,j}^{(l)} = \frac{\exp(\tilde{\mathbf{A}}_{ij}^{(l)})}{\sum_x \exp(\tilde{\mathbf{A}}_{ik}^{(l)})} = \frac{\exp(\mathbf{A}_{ij}^{(l)}) \exp(B_{ij}^{(l)}) \exp(P_{ij}^{(l)})}{\sum_x \exp(\mathbf{A}_{ik}^{(l)}) \exp(B_{ik}^{(l)}) \exp(P_{ik}^{(l)})}, \quad (12)$$

式中: $a_{i,j}^{(l)}$ 为第 l 层多头注意力中第 i 个 token 对第 j 个 token 的最终注意力权重; $B_{ij}^{(l)}$ 为第 l 层多头注意力中第 i 个 token 对第 j 个 token 的加性偏置; $P_{ij}^{(l)}$ 为第 l 层多头注意力中第 i 个 token 对第 j 个 token 的其他类型偏置; $\mathbf{A}_{ij}^{(l)}$ 为第 l 层多头注意力中第 i 个 token 对第 j 个 token 的原始注意力得分; $\tilde{\mathbf{A}}_{ij}^{(l)}$ 为第 l 层多头注意力中第 i 个 token 对第 j 个 token 的更新注意力得分。

假设备置项 $B^{(l)}$ 相对于原始注意力得分 $\mathbf{A}^{(l)}$ 较小 (即 η 较小), 可以通过泰勒展开近似计算加性偏置对注意力得分的影响:

$$a_{i,j}(\eta) = \frac{\exp(\mathbf{A}_{ij}) \exp(\eta b_j)}{\exp(\mathbf{A}_{ik}) \exp(\eta b_k)} \approx a_{i,j}^0 (1 + \eta (b_j - E_k[b_k])), \quad (13)$$

式中: $a_{i,j}(\eta)$ 为引入偏置强度参数的第 i 个 token 对第 j 个 token 的注意力权重; $a_{i,j}^0$ 为无偏置的注意力得分; $E_k[b_k]$ 为偏置的加权均值; b_k 为第 k 个 token 对应的加性偏置; b_j 为第 j 个 token 的偏置项。

引入位置偏置 $P^{(l)}$, 注意力得分 $a_{i,j}$ 的计算公式为

$$a_{i,j} = \frac{\exp(\mathbf{A}_{ij}) \exp(B_{ij}) \exp(P_{ij})}{\sum_x \exp(\mathbf{A}_{ik}) \exp(P_{ik})}, \quad (14)$$

式中: \mathbf{A}_{ij} 为无偏置时的原始注意力得分; B_{ij} 为第 i 个 token 对第 j 个 token 的显著性加性偏置; P_{ij} 为第 i 个 token 对第 j 个 token 的位置偏置。

在训练过程中, 采用标准的反向传播算法更新偏置参数 η 和 $\psi(S)$ 。这种动态显著性注入机制强调显著性信息在注意力生成过程中的主动作用。在注意力映射生成阶段实时引入显著性响应, 注意力权重根据输入图像中不同区域的重要性进行自适应分配。

显著性偏置的梯度表达式为

$$\frac{\partial L}{\partial b_j} = \sum_i \sum_x \frac{\partial L}{\partial a_{i,k}} \frac{\partial a_{i,k}}{\partial b_j}, \quad (15)$$

式中: L 为损失函数; $a_{i,k}$ 为第*i*个 token 对第*k*个 token 的注意力权重; $\frac{\partial a_{i,k}}{\partial b_j}$ 为注意力权重对显著性偏置的敏感度,表达形式由 softmax 函数的导数决定。

显著性偏置注入机制示意图 2,其中 d_i 为键的维度, \sqrt{d} 为缩放因子。在多头自注意力计算过程中,将显著性图的空间响应作为附加偏置项,与标准注意力权重矩阵进行加性融合。显著性模块输出的响应图先通过归一化映射生成显著性权重矩阵,矩阵与查询键相似度矩阵按元素加权叠加,在注意力分布中,对高显著区域赋予更高响应值,对低显著区域实施抑制。该机制在空间上强化模型对潜在目标区域的聚焦能力,使得注意力分布更加稀疏且方向性更强,为信息熵正则化提供先验约束,在复杂背景下提升模型的稳定性与鲁棒性。

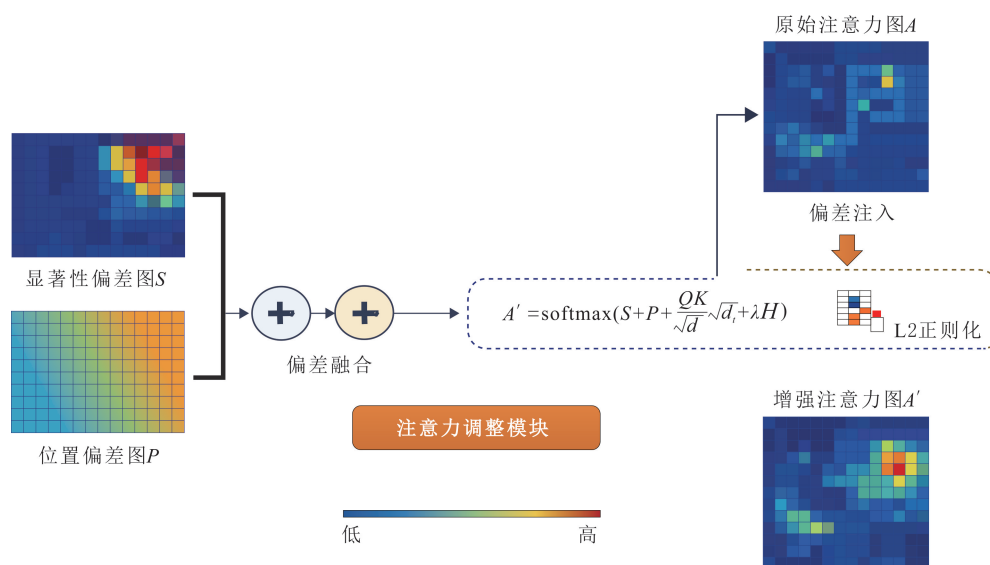


图2 显著性偏置注入机制示意

Fig.2 Schematic diagram of saliency bias injection mechanism

1.3 分层注意力模块结构与输出格式

每个转换分支由多个编码块堆叠而成^[21],每96个编码块(block)包含多头自注意力(Multi-Head Self-Attention, MHSA)模块、前馈神经网络(Feedforward Neural Network, FFN)模块,以及标准的残差连接与归一化结构。

$$\begin{cases} Z_0^{(l)} = \tilde{T}^{(l)}; \\ Z_i^{(l)} = \text{LayerNorm}(Z_{i-1}^{(l)} + \text{MHSA}(Z_{i-1}^{(l)})); \\ Z_i^{(l)} = \text{LayerNorm}(Z_i^{(l)} + \text{MHSA}(Z_i^{(l)})). \end{cases} \quad (16)$$

式中: $Z_i^{(l)}$ 为第*l*个转换分支中第*i*个编码快的输出特征, $i=1,2,\dots,N_{\text{blk}}$; N_{blk} 为每层堆叠的编码器数量,一般取2~4; $\text{MHSA}(\cdot)$ 为多头自注意力模块; $\text{LayerNorm}(\cdot)$ 表示层归一化。

将 token 表示投影到多个子空间,实现多个子空间的并行注意力计算,再进行拼接与线性变换,变换公式为

$$\text{MHSA}(X) = \text{Concat}(m_1, m_2, \dots, m_n)W_o; \quad (17)$$

$$m_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_i}}\right)V_i. \quad (18)$$

式(17-18)中: X 为输入特征; m_i 为第*i*个注意力头的输出; W_o 为输出线性变换; V_i 为将注意力权重加权到 value 向量; B_i 为通过显著性图生成的偏置; $\text{Concat}(\cdot)$ 表示所有头的输出拼接; Q_i 为第*i*个注意力头

的查询; \mathbf{K}_i 为第 i 个注意力头的键。

前馈神经网络模块采用两层全连接结构,使用非线性激活函数:

$$\text{FFN}(X) = \text{ReLU}(XW_1 + b_1)W_2 + b_2, \quad (19)$$

式中: $\text{FFN}(X)$ 为前馈神经网络对输入 X 的输出结果; $\text{ReLU}(\cdot)$ 为激活函数; W_1, b_1 分别为第一层全连接的权重与偏置; W_2, b_2 分别为第二层全连接的权重与偏置。中间维度一般为原始维度的 4 倍,以提升非线性变换能力。

各尺度转换的输出公式为

$$\begin{cases} O^{(1)} = Z_{N_{\text{blk}}}^{(1)} \in \mathbf{R}^{N_1 C_1}; \\ O^{(2)} = Z_{N_{\text{blk}}}^{(2)} \in \mathbf{R}^{N_2 C_2}; \\ O^{(3)} = Z_{N_{\text{blk}}}^{(3)} \in \mathbf{R}^{N_3 C_3}. \end{cases} \quad (20)$$

式中: $O^{(1)}, O^{(2)}, O^{(3)}$ 分别为 3 个不同分支的 token 序列; $Z_{N_{\text{blk}}}^{(i)}$ 为第 i 个尺度经过 N_{blk} 个编码块后的输出。

通过通道匹配后加权平均或池化映射为

$$O_f = \sum_{l=1}^3 \mu \text{Proj}_d(O^{(l)}), \quad (21)$$

式中: O_f 为多尺度融合的最终输出; μ 为各尺度的可学习融合权重; $\text{Proj}_d(\cdot)$ 为通道对齐变换操作。

2 基于信息熵的注意力正则操作设计

引入一种轻量级的注意力约束策略,在各层注意力图中对注意力权重分布实施信息熵正则化。通过在查询键注意力矩阵上施加熵约束,使得注意力权重分布更加稀疏,引导模型将注意力集中于有限且关键的显著区域。在训练阶段引入额外的正则损失项,实现稀疏化过程,梯度可直接作用于 Transformer 内部 Query-Key 的计算路径,从而实现对注意力收敛行为的自适应调控。

2.1 Attention 权重提取与归一化方法

假设某一尺度的 token 序列输入特征 $X \in \mathbf{R}^{N \times C}$, 在第 l 个注意力中, Query、Key、Value 的计算过程分别为

$$\mathbf{Q}^{(l)} = XW_l^Q; \mathbf{K}^{(l)} = XW_l^K; \mathbf{V}^{(l)} = XW_l^V; \mathbf{Q}^{(l)}, \mathbf{K}^{(l)} \in \mathbf{R}^{N \times d}. \quad (22)$$

式中: W_l^Q, W_l^K, W_l^V 分别为第 l 个注意力头对应的 Query、Key、Value 的线性投影矩阵, $W_l^Q, W_l^K, W_l^V \in \mathbf{R}^N$, N 为序列长度。

为构造更具稳定性的可正则注意力图,定义原始得分 $\tilde{a}_{i,j}^{(l)}$ 为

$$\tilde{a}_{i,j}^{(l)} = \frac{1}{\tau} (\mathbf{q}_i^{(l)} \mathbf{k}_j^{(l)} + \delta_{i,j} \omega), \quad (23)$$

式中: τ 为软化温度系数; $\mathbf{q}_i^{(l)}$ 为第 l 层第 i 个 query 向量; $\mathbf{k}_j^{(l)}$ 为第 l 层第 j 个 key 向量; $\delta_{i,j}$ 为 Kronecker delta 项; ω 为抑制自注意影响的常数偏移项。

归一化注意力权重采用平滑 softmax 变体,引入控制曲率的幂次归一策略,得到最终注意力权重为

$$\hat{a}_{i,j}^{(l)} = \frac{\exp(\tilde{a}_{i,j}^{(l)\theta})}{\sum_{k=1}^N \exp(\tilde{a}_{i,k}^{(l)\theta}), \theta \in (0.5, 1.5), \quad (24)$$

式中: $\hat{a}_{i,j}^{(l)}$ 为第 l 层多头注意力中第 i 个 query 对第 j 个 key 的最终注意力权重; θ 为幂指数控制参数。当 $\theta < 1$ 时,可抑制极值分布过陡;当 $\theta > 1$ 时,可强化注意力权重的集中性,在训练前期可设 $\theta = 0.8$,后期平滑调升至 1.2,形成注意力权重聚集调控路径。

对于每一组多头注意力得分 $A^{(l)}$,保留空间维度并将所有注意力头的得分进行平均或最大池化,得到聚合注意力图 $\mathbf{A}^{(\text{agg})}$:

$$\mathbf{A}^{(\text{agg})} = \text{Aggregate}_l(A^{(l)}), \mathbf{A}^{(\text{agg})} \in \mathbf{R}^{N \times N}, l \in [1, L], \quad (25)$$

式中: $\text{Aggregate}_i(\cdot)$ 为自定义可学习加权池化。矩阵中每一行 $A^i \in \mathbf{R}^{NN}$ 视作第 i 个查询的注意力图分布概率, 成为熵正则作用对象。

2.2 信息熵计算路径设计与实现

对每层转换中注意力矩阵 $\mathbf{A}^{(\text{agg})} \in \mathbf{R}^{NN}$ 进行归一化处理, 将矩阵的每一行 $x_i \in \mathbf{R}^N$ 视作第 i 个查询 token 的注意力概率分布。

对于每个概率向量 $\mathbf{x}_i = [a_{i,1}, a_{i,2}, \dots, a_{i,n}]$, 定义标准熵函数:

$$H_i = - \sum_{j=1}^n a_{i,j} \log(a_{i,j} + \delta), \quad (26)$$

式中: δ 为防止数值溢出的平滑常数, 通常为 10^{-6} 。

将熵函数重新构造为

$$\hat{H}_i = - \sum_{j=1}^n \left(\frac{a_{ij}^{1/\tau_i}}{\sum_{k=1}^n a_{ik}^{1/\tau_i}} \right) \log \left(\frac{a_{ij}^{1/\tau_i}}{\sum_{k=1}^n a_{ik}^{1/\tau_i}} + \epsilon \right), \quad (27)$$

式中: \hat{H}_i 为第 i 个查询对应的归一化注意力分布的信息熵; n 为元素总数。当 $\tau_i \rightarrow 1$ 时, 式(27)退化为标准熵函数; 当 $\tau_i < 1$ 时, 强化头部 token 权重。

令 $\omega \in \mathbf{R}^N$ 表示 Query 权重(可来源于显著性图平均值、token 自信度或可学习门控), 正则项的表达式为

$$L_{\text{entropy}} = \lambda \frac{1}{\sum_i \omega_i} \sum_i \omega_i \hat{H}_i, \quad (28)$$

式中: L_{entropy} 为熵正则项损失; ω_i 为第 i 个 token 对应的 Query 权重; λ 为熵正则权重。

将正则项加入总损失, 与分类损失 L_{cls} 和边界框损失 L_{box} 并行优化, 总损失 L_{total} 的表达式为

$$L_{\text{total}} = L_{\text{cls}} + L_{\text{box}} + \alpha L_{\text{saliency}} + \beta L_{\text{entropy}}, \quad (29)$$

式中: α 为显著性损失项的超参数权重; β 为熵正则项的超参数权重, 一般取为 $0.05 \sim 0.20$; L_{saliency} 为显著性损失。

3 目标解码与边界回归设计

通过显著性先验与信息熵约束机制, Trans-Sal Det 模型分别实现特征空间的引导与注意力分布的优化。前者通过 Tiny-U-Net 生成高分辨率显著性图, 模型在输入阶段能够聚焦潜在目标区域; 后者利用信息熵正则化约束多头自注意力矩阵, 注意力在训练过程中趋向显著区域集中, 抑制背景干扰与冗余响应。为在解码阶段有效融合二者结果, 引入显著性引导的目标查询(Object Query)机制与边界回归结构。输入显著性增强后的多尺度特征, 将注意力优化结果映射到目标空间中, 实现从显著性引导特征学习到目标级显著性检测的自然过渡, 形成 Trans-Sal Det 模型的完整闭环。

3.1 Object Query 显著性融合

解码阶段输入 M 个查询对象组成的矩阵为 $\mathbf{Q} \in \mathbf{R}^{Md}$, 对应编码器的多尺度输出 token 为 $K, V \in \mathbf{R}^{dN}$ 。输入图像对应的显著性图为 $S \in \mathbf{R}^{HW}$, 重塑一维向量 $s \in \mathbf{R}^N$, 通过线性映射得到注意力偏差项 $b \in \mathbf{R}^N$:

$$b = \sigma(W_b s + b_o), \quad (30)$$

式中: W_b 为显著性权重线性层, $W_b \in \mathbf{R}^{1 \times 1}$; $\sigma(\cdot)$ 为 Sigmoid 激活函数; b_o 为线性层的偏置项。在交叉注意力模块中, 基于标准的注意力得分公式为

$$A_{i,j} = \frac{q_i k_j}{\sqrt{d}} + b_j, \quad (31)$$

式中: b_j 为第 j 个位置的偏置。将偏置项 b_j 加入 key 方向的注意力模块, 放大显著区域 token 的吸引力。通过 softmax 归一化得到注意力得分 $A \in \mathbf{R}^{MN}$, 用于生成每个 query 的输出 q_i :

$$q_i = \sum_{j=1}^n \text{softmax } A_{i,j} V_j, \quad (32)$$

式中: $\text{softmax } A_{i,j}$ 为第 i 个 Query 对第 j 个 Key 的注意力权重。

将每个查询的输入嵌入与显著性图的全局平均响应构建门控向量 \mathbf{g}_i :

$$\mathbf{g}_i = \phi(\text{MLP}([q_i^*, \bar{s}])) , \bar{s} = \frac{1}{N} \sum_{j=1}^n S_j, \quad (33)$$

式中: $\phi(\cdot)$ 为 Sigmoid 函数; $\text{MLP}(\cdot)$ 为多层感知机, 用于对拼接的特征进行非线性变换; $[\cdot^*, \cdot]$ 为拼接操作; \bar{s} 为显著性图的全局平均响应。query 最终被调制为

$$\hat{q}_i = \mathbf{g}_i q_i. \quad (34)$$

3.2 多尺度编码特征显著性融合

Trans-Sal Det 模型的转换编码模块输出来自 3 个尺度 (1/4、1/8、1/16 分辨率) 的 token 序列 $\{Q^{(1)}, Q^{(2)}, Q^{(3)}\}$, 分别对细粒度边缘信息、中层结构信息与深层语义关系建模。

将 S 下采样至与每层特征图大小一致, 得到 3 个尺度的显著性图 (见式 (4))。将显著性图重塑为一维向量后对 token 进行加权, 在输出阶段引入显式加权操作, 以保证显著性信息与 token 语义的一致性, 从而得到融合的 token 表示。为实现三层编码 token 的多尺度融合, 采用注意力引导的特征对齐机制, 对每层 token 应用通道映射函数 $\text{Proj}^{(l)}: \mathbf{R}_l^c \rightarrow \mathbf{R}^d$, 统一输出维度为

$$\hat{O}^{(l)} = \text{Proj}^{(l)}, \hat{O}^{(l)} \in \mathbf{R}^{N_l \times d}, \quad (35)$$

式中: $\hat{O}^{(l)}$ 为第 l 个分支投影的输出维度。

为保持空间分层差异性, 使用软门控权重控制不同尺度的贡献度:

$$K_{\text{final}} = \sum_{l=1}^3 \kappa_l \hat{O}^{(l)}, \sum_{l=1}^3 \kappa_l = 1, \quad (36)$$

式中: K_{final} 为最终融合的特征结果; $\{\kappa_1, \kappa_2, \kappa_3\}$ 为可学习或显著性引导的融合权重, 通过全局显著性池化得到。

显著性融合模块采用自低向高的层间对齐策略。首先, 将低层显著性特征通过卷积调整通道维度后上采样, 与中层特征在空间尺寸上保持一致; 然后, 将二者进行逐元素加权融合, 得到中低层显著性对齐结果; 最后, 对融合结果进行线性映射, 与高层显著性特征对齐, 通过注意力权重矩阵自适应分配不同层次显著性特征的重要性, 生成最终的融合显著性图。

4 仿真结果分析

4.1 实验配置与参数设置

为验证 Trans-Sal Det 模型在遥感油井目标检测任务中的性能表现, 对两个具有代表性的数据集进行实验评估, 分别为 NEPU-OWOD V1.0 和 JLU-Oil Well Dataset。两个数据集聚焦于小尺度、高密度和弱边界目标的检测场景, 具有背景高复杂性与目标显著退化的特征, 适合作为小目标显著性检测算法的评估平台。

NEPU-OWOD V1.0 数据集包含 8 500 幅遥感影像, 涵盖沙漠、戈壁、丘陵与城镇等典型地表类型, 其中 6 000 幅用于训练, 1 500 幅用于验证, 1 000 幅用于测试; JLU-Oil Well Dataset 数据集收录 5 200 幅高分辨率影像, 训练集、验证集与测试集按 7:2:1 比例划分, 分别包含 3 640、1 040 和 520 幅影像。两组数据的井场数量与空间分布保持均衡, 以确保训练与测试阶段的统计一致性和评估公正性。为保证模型的可复现性, 引入 Tiny-U-Net 子网络的结构与参数配置说明。Tiny-U-Net 作为显著性先验生成模块, 整体结构遵循“编码—瓶颈—解码”的对称结构, 总参数量约为标准 U-Net 的 30%。网络输入为 $H \times W \times 3$ 的三通道遥感影像, 输出为 $H \times W \times 1$ 的单通道显著性概率图。编码阶段采用两级卷积单元, 每级包含两层 3×3 卷积与批量归一化操作, 使用最大池化 (2×2) 实现特征下采样; 瓶颈层维持 256 维通道深度, 以聚合全局上下文特征; 解码阶段采用双线性插值进行上采样, 每级解码单元包含 3×3 卷积和 ReLU 激活, 以恢

复空间分辨率并融合浅层特征。最终输出层通过 1×1 卷积与 Sigmoid 激活生成显著性概率图。该结构在计算复杂度与分辨率显著性之间实现平衡,能够在较低显存开销下提取高质量的显著性先验,为 Transformer 模块提供空间权重指导。实验过程中, Tiny-U-Net 的权重通过随机初始化进行端到端联合训练,学习率设为 1×10^{-4} ,优化器采用 AdamW,批量大小为 8,训练轮次为 200 轮。

4.2 量化指标分析

选取精度、召回率、F1 得分及平均精度(Average Precision, AP)等指标对模型性能进行评估,从检测准确性与综合性能方面量化分析模型。其中,精度、召回率与 F1 得分由测试集预测结果计算得到;AP、AP50、AP75(AP50 表示 IoU 阈值为 0.50 时的平均精度;AP75 表示 IoU 阈值为 0.75 时的平均精度)指标是基于目标检测标准评估协议,通过不同 IoU 阈值的精度—召回曲线积分计算获得。所有对比模型在相同训练与测试数据划分条件下进行实验,以保证结果的公平性与可比性。

不同模型在油井检测任务上的综合性能见表 1。由表 1 可知, Trans-Sal Det 模型的精度、召回率及 F1 得分评价指标优于其他方法的。与 YOLOv8-SGE 模型相比, Trans-Sal Det 模型在精度上提升 1.1%, 召回率提升 1.2%, F1 得分提升至 0.890。 Trans-Sal Det 模型在 F1 得分上较 Query Det 模型的提升 0.5%, 在保证检测精度的同时具有更高的召回率,表明显著性引导机制在复杂遥感场景下有效增强小目标响应能力。在联合测试集上, Trans-Sal Det 模型在主流检测框架中表现优异, AP、AP50、AP75 指标高于其他方法的。这说明转换模块在引入结构性引导与显式注意力调控后,建模能力有效转化为检测性能的稳步提升。

表 1 不同模型在油井检测任务上的综合性能

Table 1 Overall performance of different models in oil well inspection tasks

模型	主干网络	精度	召回率	F1 得分	AP/%	AP50/%	AP75/%
Faster R-CNN	R50-FPN	0.684	0.919	0.784	52.941	89.547	56.974
	R50-C4	0.742	0.875	0.803	48.768	85.785	51.982
	R50-DC5	0.738	0.897	0.810	50.230	89.469	48.969
	R101-FPN	0.724	0.879	0.794	49.177	84.863	49.202
	R101-C4	0.734	0.924	0.818	52.292	88.720	56.103
	R101-DC5	0.764	0.928	0.838	50.034	87.408	52.935
Cascade R-CNN	R50-FPN	0.710	0.890	0.792	51.111	88.214	53.232
Mask R-CNN	R50-FPN	0.688	0.920	0.792	51.587	87.948	53.497
SSD	Darknet-53	0.807	0.674	0.734	43.298	85.853	36.103
YOLOv3	VGG16	0.166	0.267	0.205	3.698	7.653	3.061
RetinaNet	R50-FPN	0.693	0.903	0.787	49.765	88.276	50.384
MS-BSNet	—	0.823	0.890	0.855	55.628	98.764	57.868
YOLOv8-SGE	—	0.850	0.910	0.879	58.912	99.224	60.574
QueryDet	Swin-Tiny	0.856	0.917	0.885	59.231	99.351	60.812
Trans-Sal Det	PVT-Tiny+Tiny-U-Net	0.861	0.922	0.890	59.802	99.631	61.428

矫正权重 λ 参数敏感性实验结果见表 2。由表 2 可知,矫正权重对模型性能具有显著影响:当 $\lambda < 0.05$ 时,约束强度不足,注意力分布趋于扩散,导致显著性引导效果减弱;当 $\lambda > 0.20$ 时,注意力集中过度,模型出现轻微欠拟合,整体性能下降;当 $\lambda = 0.10$ 时,精度、召回率、F1 得分及 AP 指标达到最优,其中 AP50 和 AP75 分别约提升 0.7% 和 1.8%。适度的信息矫正则项能有效平衡注意力稀疏性与特征覆盖度,提升模型在复杂场景下的小目标检测鲁棒性。

4.3 模块贡献对比

在各模型变体中使用模块缩写:SM 表示 Saliency Map 显著性图模块;TW 表示 Token-Level Weighting,为基于显著性图的 token 加权操作;AB 表示 Attention Bias,为显著性图引导下的注意力偏置注入;

ER 表示 Entropy Regularization, 用于收缩注意力分布的信息熵正则项; MS 表示 Multi Scale Token Branches 多尺度转换分支结构; GF 表示 Gated Fusion, 为门控融合的多尺度熵控制模块。

表 2 熵正则权重 λ 参数敏感性实验结果

Table 2 Sensitivity experiment results of entropy regularization weight λ parameter

λ	精度	召回率	F1 得分	AP50/%	AP75/%
0(无正则)	0.842	0.908	0.874	98.921	59.602
0.05	0.853	0.915	0.883	99.132	60.144
0.10(最佳)	0.861	0.922	0.890	99.631	61.428
0.15	0.857	0.918	0.887	99.427	60.982
0.20	0.848	0.905	0.876	99.103	59.731

Trans-Sal Det 模型结构各模块消融实验结果见图 3。从图 3(a)可以看出, Trans-Sal Det 模型的各组成模块在油井检测任务中具有明确的性能提升作用, 整体架构设计具有较强的模块互补性与协同增强效果。基础模型仅限变压器基线, 依靠最小化结构(无显著性引导、无注意偏置、无熵约束), 精度为 0.801, 召回率为 0.867, F1 得分为 0.832, 明显低于完整模型的。

为验证 Trans-Sal Det 模型各模块对模型性能的贡献, 对平均精度指标开展消融实验(见图 3(b))。实验中逐步去除模型的关键模块, 分析每个模块的去除对 AP 指标的影响。依次去除 Trans-Sal Det 模型的关键模块, 评估对整体检测性能的独立贡献。显著性引导(SM+TW)对提升小目标召回率具有核心作用, 注意力偏置(AB)与信息熵正则(ER)有效增强注意力分布的稳定性与聚焦性, 多尺度分支(MS)与门控融合(GF)能够在复杂背景下显著降低误检率。

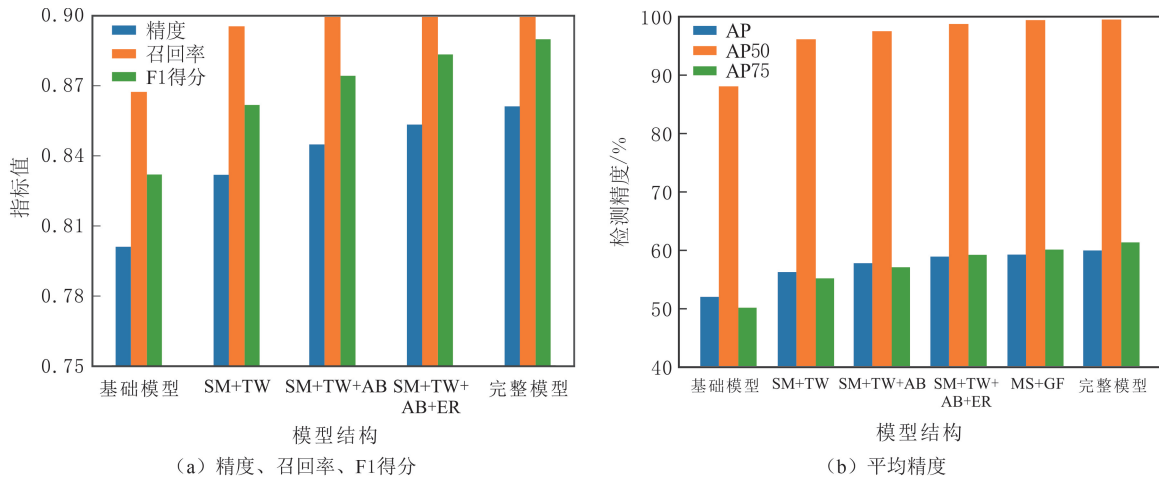


图 3 Trans-Sal Det 模型结构各模块消融实验结果
Fig. 3 Ablation experiment results of each module in Trans-Sal Det model

4.4 可视化分析

在相同测试数据集上对各模型进行目标检测实验, 检测结果可视化, 不同模型的检测效果见图 4, 其中, 红色框表示检测框, 绿色框表示真实框。由图 4 可以看出, 不同模型在目标检测任务中的表现存在明显差异, 尤其是目标框的精度和置信度方面。Trans-Sal Det 模型的检测框与真实框高度重合, 定位误差极小, 置信度稳定且较高。在复杂背景或目标存在重叠的情况下, Trans-Sal Det 模型能够保持较高的检测精度和良好的鲁棒性, 明显优于其他模型的。

Mask R-CNN 模型在一些测试图像上检测效果良好, 但对于小目标和复杂背景的样本, 定位框的边界偏移量远大于 Trans-SalDet 模型的, 且置信度分布较散乱; Faster R-CNN 和 Cascade R-CNN 模型在高密度目标区域检测时漏检率高, 检测框和真实框之间的 IoU 较低, 难以达到 AP75 的阈值要求。Trans-Sal Det 模型显著降低密集小目标的漏检率, 检测框能够更精准地对齐目标的真实边缘与轮廓, 在复杂背景干扰下能保持稳定的目标定位能力。对于具有复杂空间分布和丰富上下文信息的场景, Transformer

架构在建模遥感图时具有优越性。

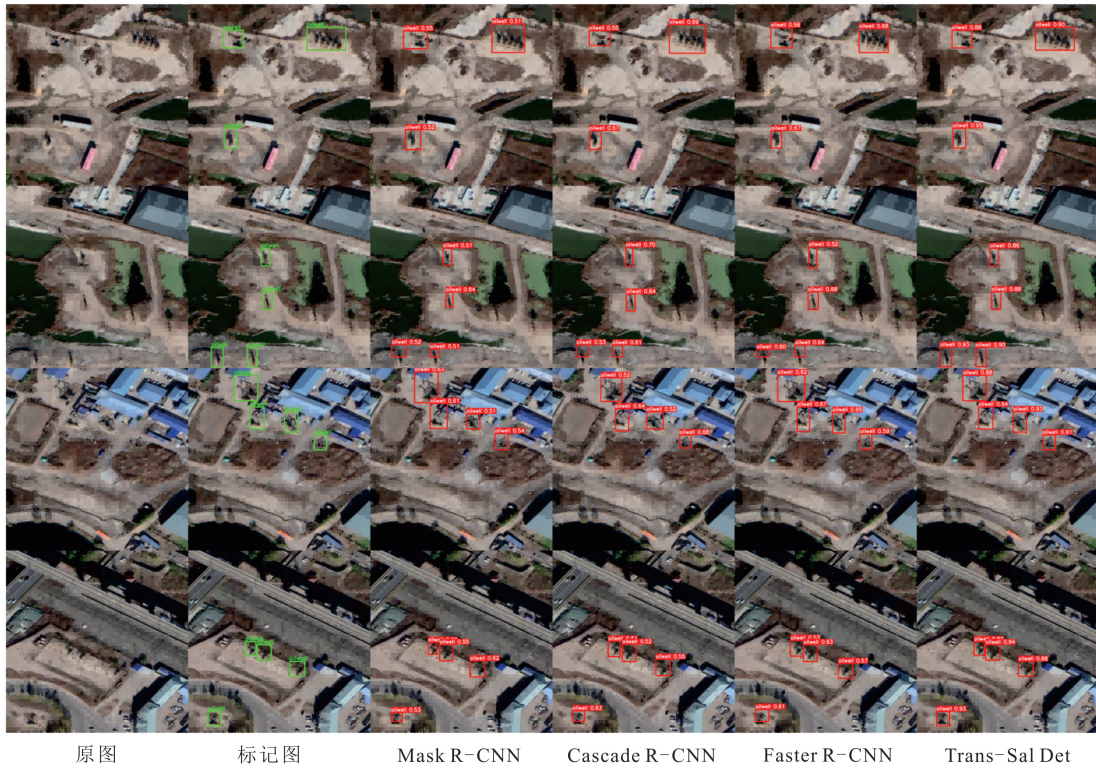


图 4 不同模型检测效果

Fig. 4 Comparison of detection performance across different models

Trans-Sal Det 模型在训练过程中的损失曲线见图 5。由图 5 可以看出,随训练轮次的增加,训练损失不断下降,表明模型有效学习数据特征并优化参数。验证损失在训练初期存在波动,随训练的进行逐渐稳定,表明模型在验证集上的表现不断提升。Trans-Sal Det 模型在训练过程中逐渐避免过拟合或欠拟合,在学习过程中达到平衡并提升对新数据的泛化能力,性能更稳定。

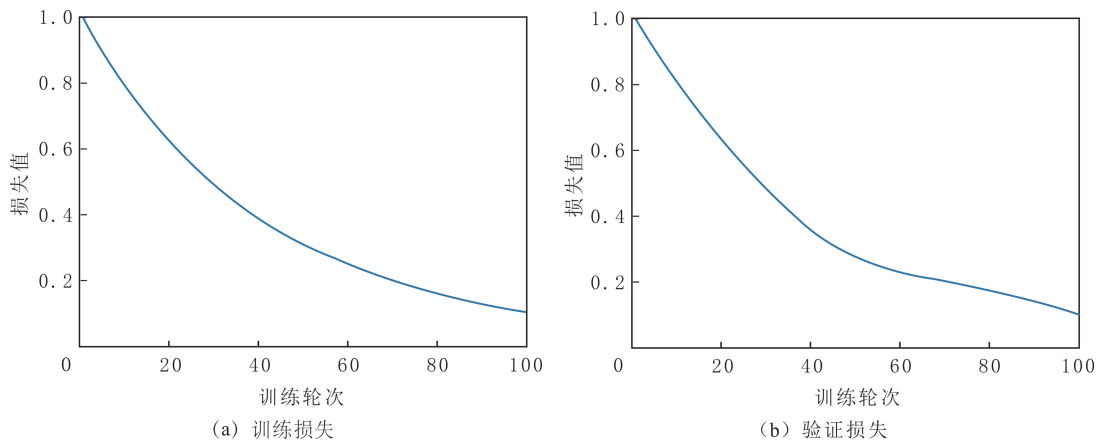


图 5 Trans-Sal Det 模型训练损失与验证损失曲线

Fig. 5 Training loss and validation loss curves of Trans-Sal Det model

4.5 模型复杂度分析

为评估 Trans-Sal Det 模型性能与效率的平衡性,比较各模型的参数量、计算量及单张图像的平均推理时间,模型复杂度对比实验结果见表 3。由表 3 可以看出,Trans-Sal Det 模型在保持较高精度和 F1 得分的同时,计算量为 192.7×10^9 次/s,参数量为 39.8×10^6 ,较 YOLOv8-SGE 模型的略高 9.9%,但性能提升 1.1%;平均推理时间为 74 ms/张,在保证实时性的前提下显著提高检测精度。Trans-Sal Det 模型

在显著性先验与金字塔转换机制融合的基础上,既提升检测效果,又保持较优的计算效率,实现性能与复杂度的良好平衡。

表 3 模型复杂度对比实验结果

Table 3 Experimental results of model complexity comparison

模型	参数量/ 10^6	计算量/ $(10^9 \text{ 次} \cdot \text{s}^{-1})$	平均推理时间/ $(\text{ms} \cdot \text{张}^{-1})$
Faster R-CNN	41.1	207.3	115
YOLOv3	61.9	155.4	72
RetinaNet	38.7	204.6	108
MS-BSNet	52.4	225.8	96
YOLOv8-SGE	36.2	187.5	68
Trans-Sal Det	39.8	192.7	74

4.6 统计显著性验证

为验证 Trans-Sal Det 模型性能统计可靠性,对 5 种模型进行 5 次独立训练与测试实验,报告主要指标的平均值与标准差(平均值 \pm 标准差),采用配对 t -检验评估模型间的显著性差异。模型性能统计显著性分析结果见表 4。由表 4 可以看出,Trans-Sal Det 模型在精度、召回率和 F1 得分 3 个指标表现最优,且波动性最小(标准差 ≤ 0.005),具有较强的稳定性与收敛一致性。与其他模型的性能差异具有统计显著性,与 YOLOv8-SGE 模型的显著性水平 p 小于 0.05,与传统检测器(如 Faster R-CNN、Retina Net)的 p 小于 0.01。这验证 Trans-Sal Det 模型的显著性引导与金字塔转换机制,在性能提升上具有统计可靠性与可重复性。

表 4 模型性能统计显著性分析结果

Table 4 Statistical significance analysis results of model performance

模型	精度	召回率	F1 得分	p
Faster R-CNN	0.724 ± 0.013	0.879 ± 0.010	0.794 ± 0.009	< 0.01
RetinaNet	0.693 ± 0.017	0.903 ± 0.012	0.787 ± 0.014	< 0.01
MS-BSNet	0.823 ± 0.008	0.890 ± 0.009	0.855 ± 0.010	< 0.05
YOLOv8-SGE	0.850 ± 0.006	0.910 ± 0.007	0.879 ± 0.008	< 0.05
Trans-Sal Det	0.861 ± 0.005	0.922 ± 0.004	0.890 ± 0.005	—

4.7 错误类型分析

为分析 Trans-Sal Det 模型的检测误差类型,在 JLU-Oil Well 测试集上统计不同模型的漏检率、误检率及总体错误率(见表 5)。由表 5 可以看出,Trans-Sal Det 模型的 3 个指标优于其他模型的,漏检率较 YOLOv8-SGE 模型的降低 1.9%,误检率降低 2.4%。传统模型在低对比度与密集井场区域容易发生漏检,Trans-Sal Det 模型在显著性先验和注意力偏置机制作用下,能够显著抑制伪响应区域,提高目标定位准确度,验证 Trans-Sal Det 模型在复杂背景下的鲁棒性与判别能力。

表 5 不同模型的错误类型统计结果

Table 5 Statistics of error types for different models

模型	漏检率/%	误检率/%	总体错误率/%
Faster R-CNN	8.7	6.5	15.2
RetinaNet	7.9	6.8	14.7
MS-BSNet	6.1	5.4	11.5
YOLOv8-SGE	5.4	4.8	10.2
Trans-Sal Det	3.5	2.4	5.9

5 结论

(1)对于遥感图像中小目标尺寸小、特征弱及易受复杂背景干扰,提出一种融合显著性先验与金字塔转换机制的 Trans-Sal Det 小目标检测框架,在复杂背景、遮挡及光照变化条件下,从特征增强和注意力引导方面提升遥感影像中小目标检测能力,增强模型的泛化能力与实际应用的适应性。

(2)Trans-Sal Det 模型在多个数据集上的表现优于其他方法的性能,AP50 和 AP75 指标提升更显著;模型在精度、召回率及 F1 得分方面表现更高的稳定性与一致性,验证模型在复杂场景与目标重叠情况下的有效性与鲁棒性。

(3)Trans-Sal Det 模型在性能上具有明显优势,但在计算复杂度及对数据规模的依赖方面存在局限性,可围绕模型轻量化、弱监督或少样本学习、跨模态信息融合及模型泛化能力提升等方面进一步研究,以增强实用性与适应性。

参考文献(References):

- [1] Gui S, Song S, Qin R, et al. Remote sensing object detection in the deep learning era: a review[J]. Remote Sensing, 2024,16(2):28.
- [2] Teodosio L, Emanuele C, Irina C, et al. A MODIS-based robust satellite technique(RST) for timely detection of oil spilled areas[J]. Remote Sensing, 2017,9(2):128.
- [3] 江灏,张微,陈静,等.基于改进 YOLOv8 算法的去雾遥感图像电力目标检测[J].遥感技术与应用,2025,40(5):1243-1254.
Jiang Hao, Zhang Wei, Chen Jing, et al. Power target detection based on improved YOLOv8 algorithm in dehazing remote sensing image[J]. Remote Sensing Technology and Application, 2025,40(5):1243-1254.
- [4] 陈维力,潘永帅,范坤宇,等.塔里木盆地库车坳陷博孜一大北地区差异成藏过程及控藏因素[J].东北石油大学学报,2025,49(1):61-76.
Chen Weili, Pan Yongshuai, Fan Kunyu, et al. Pore structure differential accumulation process and reservoir controlling factors in Bozi-Dabei Area, Kuqa Depression, Tarim Basin[J]. Journal of Northeast Petroleum University, 2025,49(1):61-76.
- [5] 彭先锋,胡笑非,张焯毓,等.苏里格气田山1段储层致密化成因及控制因素[J].东北石油大学学报,2018,42(1):60-67.
Peng Xianfeng, Hu Xiaofei, Zhang Yeyu, et al. Densification and controlling factors of tight reservoirs in Shan 1 Formation of Sulige Gas Field[J]. Journal of Northeast Petroleum University, 2018,42(1):60-67.
- [6] Ral A, Kim J M. A novel pipeline leak detection approach independent of prior failure information[J]. Measurement, 2021, 167: 108284.
- [7] Wang C, Wang Z, Dong H, et al. Fusionformer: a novel adversarial transformer utilizing fusion attention for multivariate anomaly detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025,36(8):14479-14492.
- [8] Mahmood Y, Afrin T, Huang Y, et al. Sustainable development for oil and gas infrastructure from risk, reliability, and resilience perspectives[J]. Sustainability, 2023,15(6):4953.
- [9] 石颖,李莹,王维红,等.线性 Radon 变换噪声压制法及其在古龙断陷中的应用[J].东北石油大学学报,2012,36(4):116-120.
Shi Ying, Li Ying, Wang Weihong, et al. Approach of linear noise suppression using Radon transform and its application in Gulona fault depression[J]. Journal of Northeast Petroleum University, 2012,36(4):116-120.
- [10] Cordes E E, Jones D O B, Schlacher T A, et al. Environmental impacts of the deep-water oil and gas industry: a review to guide management strategies[J]. Frontiers in Environmental Science, 2016,4:58.
- [11] Vekeen S T, Balogun A L. Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment[J]. Remote Sensing, 2020,12(20):3416.
- [12] 李睿.油气管道内检测技术与数据分析方法发展现状及展望[J].油气储运,2024,43(3):241-256.
Li Rui. Current progress and prospects of in-line inspection techniques and data analysis methods for oil and gas pipelines[J]. Oil & Gas Storage and Transportation, 2024,43(3):241-256.
- [13] 高胜,王妍,任永良,等.大型复杂油田注水系统优化运行关键技术与智能化展望[J].东北石油大学学报,2020,44(4):91-98.
Gao Sheng, Wang Yan, Ren Yongliang, et al. Key technologies and intelligent prospect for optimized operation of large complex water injection systems in oilfields[J]. Journal of Northeast Petroleum University, 2020,44(4):91-98.
- [14] Brown K E, Lea J F. Nodal systems analysis of oil and gas wells[J]. Journal of Petroleum Technology, 1985,37(10):1751-1763.
- [15] 苏健,叶文强.复杂环境下的小目标交通标志检测[J].计算机系统应用,2025,34(11):202-211.
Su Jian, Ye Wenqiang. Traffic sign detection for small objects in complex environment[J]. Computer Systems & Applications, 2025,34(11):202-211.

- [16] Li Z, Wang Y, Zhang N, et al. Deep learning-based object detection techniques for remote sensing images: a survey[J]. Remote Sensing, 2022,14(10):2385.
- [17] Yang Y, Wang J, Liao J, et al. Abundance and diversity of soil petroleum hydrocarbon-degrading microbial communities in oil exploring areas[J]. Applied Microbiology and Biotechnology, 2015,99:1935-1946.
- [18] Zhou X, Chang N B, Li S. Applications of SAR interferometry in earth and environmental science research[J]. Sensors, 2009,9(3):1876-1912.
- [19] 王晓丽,邓达康,孟祥龙,等. 基于领域本体的油气勘探开发知识获取模式及实现[J]. 东北石油大学学报,2016,40(4):74-79.
Wang Xiaoli, Deng Dakang, Meng Xianglong, et al. Oil and gas exploration and production knowledge processing mode and implementation based on domain ontology[J]. Journal of Northeast Petroleum University, 2016,40(4):74-79.
- [20] Behera D K, Pujar G S, Kumar R, et al. A comprehensive approach towards enhancing land use land cover classification through machine learning and object-based image analysis[J]. Journal of the Indian Society of Remote Sensing, 2025,53(3):731-749.
- [21] Yang C, Huang Z, Wang N. QueryDet: cascaded sparse query for accelerating high-resolution small object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans: IEEE, 2022:13668-13677.

《东北石油大学学报》加入 SciEngine 平台

2025年5月,《东北石油大学学报》编辑部与中国科技出版传媒股份有限公司(科学出版社)签署期刊网络发布许可协议,《东北石油大学学报》加入 SciEngine 平台(<https://www.sciengine.com/>)。Sci-Engine 平台是由中国科技出版传媒股份有限公司(科学出版社)打造的科技期刊全流程数字出版与知识服务平台,是国家科技期刊出版基地的运营主体,通过技术创新和资源整合,为科研工作者提供高效、便捷的学术交流环境,助力中国科技期刊走向国际。



Trans-Sal Det small object detection model based on saliency prior and pyramid transformation mechanism/

2026,50(2):109-122

Zhao Zixiang^{1,2,3}, Li Jiahui^{2,4,5}, Bu Xianye^{2,3,4}, Mu Shujuan^{2,3,4}, Sui Yang⁶

(1. *Key Laboratory of Shale Oil and Gas Accumulation and Efficient Development, Ministry of Education, Northeast Petroleum University, Daqing, Heilongjiang 163318, China*; 2. *Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing, Heilongjiang 163318, China*; 3. *College of Electrical and Information Engineering, Northeast Petroleum University, Daqing, Heilongjiang 163318, China*; 4. *Key Laboratory of Networking and Intelligent Control in Heilongjiang Province, Northeast Petroleum University, Daqing, Heilongjiang 163318, China*; 5. *Mathematics and Interdisciplinary Science Research Center, Northeast Petroleum University, Daqing, Heilongjiang 163318, China*; 6. *Daqing Emergency Repair Center, National Pipeline Network Group Northeast Branch, Daqing, Heilongjiang 163000, China*)

Abstract: To address the difficulty of accurately detecting small oil well targets in oilfield remote sensing images under complex surface environments and background interference, this paper proposes a Trans-Sal Det small object detection model that integrates saliency priors with a pyramid transformation mechanism. A saliency generation module based on Tiny-U-Net is constructed to produce high-resolution saliency maps, highlighting potential oil well regions while suppressing redundant background information. The saliency maps are concatenated with the original remote sensing images at the channel level and fed into a pyramid Transformer encoder. Through a multi-scale window self-attention mechanism, cross-layer feature modeling is achieved, effectively integrating low-level detailed features with high-level semantic information. A cross-scale feature fusion strategy is introduced to enhance the model's representation capability for oil well targets of different scales. Comparative and ablation experiments are conducted on typical oilfield remote sensing datasets to systematically validate the model performance. The results demonstrate that Trans-Sal Det exhibits superior small object detection performance in complex backgrounds, occlusion, and low-contrast scenarios, with significantly improved recall and detection accuracy compared to mainstream methods. The introduction of saliency priors effectively guides attention to focus on key regions, enhancing the Transformer's perception of small targets. The proposed method provides an efficient and feasible technical approach for automated oilfield remote sensing detection and intelligent monitoring, with strong potential for practical applications.

Key words: saliency prior; pyramid transformation mechanism; cross-scale feature fusion; oil well detection; dynamic saliency; Transformer; small object detection