

# 形状与位置先验信息引导的遥感影像建筑物域适应提取方法

赵绍轩<sup>1</sup>, 周晓光<sup>1</sup>, 侯东阳<sup>1</sup>, 郭静<sup>2,3</sup>, 刘华光<sup>2,3</sup>

1. 中南大学 地球科学与信息物理学院, 长沙 410083;

2. 实景三维建设与应用技术湖南省工程研究中心, 长沙 410000;

3. 湖南省第一测绘院, 长沙 410000

**摘要:** 深度学习为遥感影像建筑物提取提供了一种有效途径,但在训练数据(源域)与测试数据(目标域)分布不同的情况下,仅利用源域训练的模型直接应用在目标域时精度会显著下降。建筑物域适应提取方法可以克服域间数据分布差异,提升不同环境下建筑物信息的识别能力。针对建筑物域适应提取中目标域标签获取难度大、标注成本高,且当前方法未能充分利用建筑物不变特性提供跨域一致性约束的问题,本文提出了形状与位置先验信息引导的遥感影像建筑物域适应提取方法。首先,综合利用建筑物指数、Harris等方法自动提取目标域建筑物角点,并借助图像形态学方法提取源域标签建筑物边缘,作为两域的形状先验信息;然后,设计基于高斯转换的位置先验信息提取方法,将目标域OSM对象和源域标签对象分别转为双域的位置先验信息;最后,利用上述形状先验信息构建形状损失函数,同时在两域对建筑物目标提供训练约束,将双域位置先验信息作为独立附加通道叠合影像图层构成4通道输入,丰富目标域建筑物信息,设计基于对抗学习的建筑物域适应提取模型AU\_AdaptNet。结果表明:本文方法的IoU指标比未经过域适应的基础模型提高了15%,比未添加先验信息引导的域适应模型提高了6%,且在目标域无OSM数据时,仅依靠形状先验信息引导也能提高建筑物域适应提取精度,在目标域OSM数据完整性较低的条件下,也可与使用高质量目标域标签的半监督域适应方法取得相近效果。

**关键词:** 建筑物提取, 域适应, 遥感影像, 先验信息, OSM

**中图分类号:** P2

**引用格式:** 赵绍轩,周晓光,侯东阳,郭静,刘华光.2025.形状与位置先验信息引导的遥感影像建筑物域适应提取方法.遥感学报, 29(12): 3609-3625

Zhao S X, Zhou X G, Hou D Y, Guo J and Liu H G. 2025. Domain-adaptive building extraction method from remote sensing imagery that uses shape and position priors. National Remote Sensing Bulletin, 29(12): 3609-3625 [DOI:10.11834/jrs.20255067]

## 1 引言

地面建筑物是城市地理信息的重要内容之一,准确提取建筑物边缘轮廓对城市规划分析、灾害损失评估、违章建筑检测等具有重要意义(李德仁, 2018; 晏雄锋等, 2022)。因此,建筑物提取一直是遥感智能解译的重要研究方向(杜培军, 2020; 刘海兵等, 2023; 殷吉崇等, 2024)。

基于遥感影像的建筑物提取方法可以分为传统方法和深度学习方法。传统方法依托计算机视

觉原理,通过分析建筑物的光谱特征、空间结构特征以及纹理特征等,结合专家经验构建特征提取模型,利用阈值分割或角点(Cote和Saeedi, 2013)、边缘轮廓(孙金彦等, 2017; Wang等, 2015)等形态学处理实现建筑物提取。然而传统方法存在明显技术瓶颈,一方面高度依赖先验知识设计特征提取算子,另一方面基础图像处理算法难以实现高精度的建筑物识别,特别是处理大规模遥感影像数据时,其准确性和适应性均受到限制(刘异等, 2024)。

收稿日期: 2025-03-19; 预印本: 2025-08-23

基金项目: 国家重点研发计划(编号: 2022YFB390420501); 实景三维建设与应用技术湖南省工程研究中心开放课题(编号: 3DRS2024Y4)

第一作者简介: 赵绍轩,研究方向为遥感影像智能解译。E-mail: zsx2020@csu.edu.cn

通信作者简介: 侯东阳,研究方向为时空数据检索、挖掘与群智更新。E-mail: houdongyang1986@csu.edu.cn

近年来, 遥感影像建筑物提取逐渐实现从人工特征驱动向数据驱动的范式迁移(杨明旺等, 2024)。这是因为语义分割、实例分割等深度学习模式在大规模高质量训练样本驱动下可以自动提取具有较好表征能力的建筑物高层级特征(陈经纬等, 2024)。例如, 龙丽红等(2023)在UNet基础上设计D-UNet进行城市建筑物分割, Zuo等(2025)利用语义分割网络的多阶段特征图构建跨尺度建筑物特征融合网络CFF-Net。同时, 部分学者在深度神经网络架构基础上引入几何约束等传统关于建筑物特性的先验知识, 形成混合增强算法框架。这种技术路径不仅继承了传统方法的可解释性优势, 又充分发挥了深度学习的复杂特征提取优势。例如, 李治等(2023)利用形态学序列特征扩充高分遥感建筑物信息, 构建自适应分割模型。Yiming等(2024)结合建筑物轮廓设计了形状增强模块用于建筑物提取。

然而建筑物标签获取难度大、标注成本高, 且受区域、气候、风格、拍摄硬件等因素的共同影响, 训练样本难以完全覆盖所有的建筑物类型, 即训练样本(源域)与实际应用场景(目标域)间会存在不同的数据分布, 导致深度学习模型在超出训练样本多样性以外的新区域精度急剧下降, 如图1所示。目前, 域适应是解决上述模型泛化问题的主流思路, 其是通过挖掘源域与目标域数据间的特征相似性, 对齐双域的特征分布, 实现目标域的预测与推理(范苍宁等, 2021)。

域适应可以分为无监督和半监督方法。其中, 无监督域适应主要针对源域有样本标签、目标域无样本标签场景。例如, Tsai等(2018)基于对抗学习设计AdaptNet, 使用源域样本训练网络, 再对齐源域和目标域预测结果的输出分布差异。Shi等(2021)基于循环生成网络CycleGAN, 将源域图像转化为接近目标域图像的特征分布, 提出了一种综合像素级和表示级的联合域适应网络。Peng等(2022)结合源域和目标域在图像、特征和输出中的多层级分布, 设计了多级域适应网络FDANet进行建筑物域适应提取。然而, 由于目标域缺乏监督信息, 源域和目标域之间的数据分布可能存在较大差距, 导致无监督域适应方法精度较低, 且当前无监督域适应建筑物提取方法未能充分利用建筑物本身在源域和目标域的不变特性辅助特征对齐。半监督域适应则在目标域中引入

少量标签数据作为额外的监督信息, 用于指导双域的特征分布对齐。例如, Basak和Yin(2023)设计了一种半监督微调域适应方法SSFT, 首先挖掘两域数据的差异性内容和不变数据特征预训练编码器, 然后使用10%—50%数量比例的目标域样本微调模型。Yan等(2024)提出了一种半监督主动域适应框架SS-ADA, 同时使用源域标签和25%数量比例的目标域样本, 提高模型训练过程中的伪标签质量。Yang等(2024)在主动学习模式下, 从目标域选取与源域特征分布疏离的图像绘制建筑物轮廓, 作为目标域标签再输入模型中进行训练。上述半监督域适应方法提高了目标域建筑物提取精度, 但其有效性依赖于目标域中样本数据的数量与质量, 并且少量样本数据也更容易导致模型过拟合, 不利于其他未标记数据。

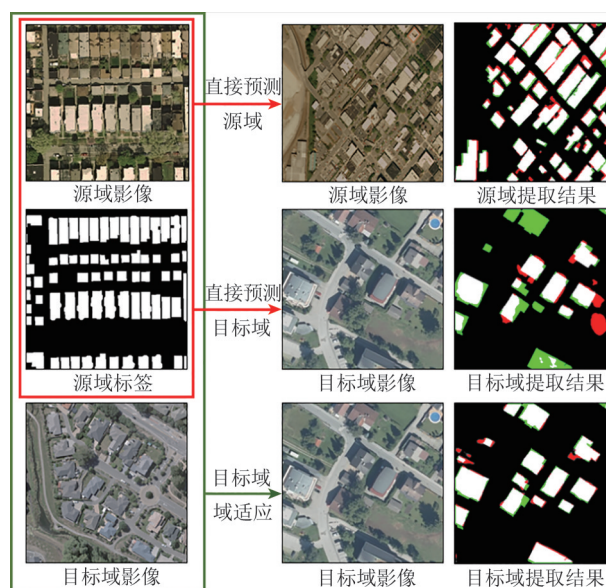


图1 有无域适应方法的提取结果对比(白色像素为正确提取;红色像素为误检;绿色像素为漏检)

Fig.1 Comparison of extraction results with and without domain adaptation method(White indicates correct extraction, red indicates false detection, and green indicates missed detection)

实际上, 已有研究发现由非专业人员众包贡献的众源地理数据可以为空间指引和样本扩充等领域提供全新的解决思路。例如: Kang等(2018)和Zhang等(2020)利用OSM(Open Street Map)数据辅助人工标注样本, 成功协助模型实现了单域建筑物提取。Cao和Huang(2023)在真实标签完全缺乏或有限的情况下, 借助众源建筑物矢量轮廓构建建筑物提取数据集。上述方

法仅仅在单域建筑物提取证明了众源地理数据的可用性, 尚未在建筑物域适应提取中应用。此外, 众源数据集可能在样本数量、质量和覆盖区域上存在显著差异, 导致无法充分表达所需的监督信息 (Viana 等, 2019)。

鉴于此, 本研究旨在无监督域适应场景下, 融合建筑物传统特征算子和众源地理数据等先验信息, 增强遥感影像建筑物特征泛化性, 弥补目标域无高质量标签引导特征对齐的不足, 提出形状与位置先验信息引导的建筑物域适应提取方法。首先, 综合利用建筑物指数、Harris、线段检测器 LSD (Line Segment Detector) 等方法自动提取目标域影像的建筑物角点, 并结合源域标签生成的建筑物边缘, 分别形成目标域与源域形状先验信息。然后, 设计高斯置信转换方法弱化众源数据可能存在的质量、覆盖率问题, 分别将源域一定比例的标签对象和目标域 OSM 对象转换成源域与目标域的位置先验。最后, 基于对抗学习设计域适应

提取模型, 利用上述形状先验信息构建损失函数, 以期改变传统无监督模型在目标域缺乏建筑物约束的训练方式, 并将位置先验信息作为独立附加通道与遥感影像叠合, 进一步丰富目标域建筑物信息。

## 2 研究方法

本文提出的方法主要包含建筑物角点提取模块、高斯转换模块和对抗学习提取模型 3 个部分, 具体技术路线如图 2 所示。其中建筑物角点提取模块综合利用建筑物指数、Harris 和 Stephens (1988) 算法自动实现目标域建筑物角点的准确提取; 高斯转换模块针对目标域的 OSM 对象和源域标签中抽取的建筑物对象, 基于轮廓最大内切圆和高斯函数进行置信转换, 形成模型输入的附加通道; 对抗学习提取模型将角点和边缘损失函数作为形状先验信息, 高斯置信转换结果作为位置先验信息, 融入分割网络协同引导模型训练。

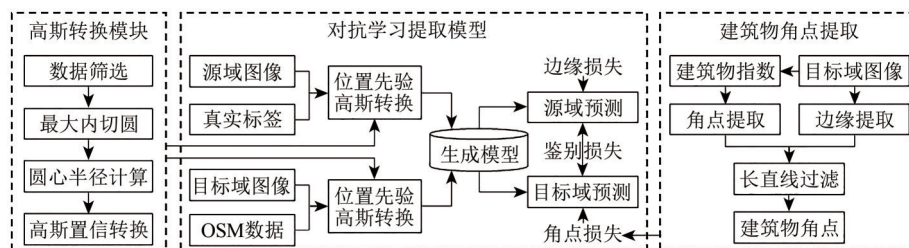


图2 技术路线图

Fig. 2 Technical roadmap

### 2.1 基于角点和边缘的形状先验信息提取

遥感影像中的人造构筑物与自然地物有显著差异 (李治 等, 2023)。建筑物在局部对比度、规则类矩形和线性边缘等方面有明显特征。本文凝练建筑物形状特性, 提取目标域建筑物角点和源域建筑物边缘作为形状先验信息。

#### 2.1.1 目标域建筑物角点提取

目标域影像建筑物角点由形态学建筑物指数 MBI (Morphological Building Index) (Huang 和 Zhang, 2011) 计算得到。建筑物指数是高分辨率影像中建筑物特点与形态学算子特性的结合, 相比原始影像可以更具针对性地在影像中反映建筑物强度信息。为了进一步减少非建筑物角点对提取结果的影响, 计算影像边缘信息后, 利用建筑物角点

一般分布近长直边缘线的特性, 对角点再次过滤。本文目标域影像建筑物角点的自动获取步骤如图 2 (c) 所示, 具体如下:

(1) 计算形态学建筑物指数 (MBI)。影像经过形态学白帽重构, 计算微分形态学剖面 DMP。具体公式如下:

$$DMP_{WTH}(d, s) = |WTH(d, (s + \Delta s)) - WTH(d, s)| \quad (1)$$

式中,  $WTH$  为形态学白帽重构结果,  $d$  和  $s$  分别代表线性结构元素的方向和尺度。MBI 指数计算公式如下:

$$MBI = \frac{\sum_{d,s} DMP_{WTH}(d, s)}{D \times S} \quad (2)$$

式中,  $S = (S_{max} - S_{min})/\Delta S + 1$ , 表示多尺度形态学运算时结构元素尺度的范围, 由下限  $S_{min}$  与上限  $S_{max}$  和步长  $\Delta S$  决定,  $D$  为计算建筑物剖面时的方向数。

(2) 角点提取。利用 Harris 角点检测法 (Harris 和 Stephens, 1988) 在 MBI 图像中按照一定方向滑动窗口, 根据窗口内灰度的变化情况计算矩阵特征值, 提取明显的特征点。角点响应函数公式如下:

$$R = \det(\mathbf{M}) - k(\text{tr}(\mathbf{M}))^2 = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (3)$$

式中,  $R$  为角点响应程度,  $\mathbf{M}$  表示窗口滑动前后像素灰度值形成的自相关矩阵,  $\det$  表示行列式,  $\text{tr}$  表示矩阵的迹,  $k$  是人为设置的经验常数, 本文参照一般案例默认设置为 0.05,  $\lambda_1$ 、 $\lambda_2$  为矩阵  $\mathbf{M}$  的两个特征值。

(3) 影像边缘信息提取。利用 Canny 算法 (Canny, 1986) 搜索原始影像中的边缘潜在目标, 将灰度梯度方向上的极大值保留为边缘点以获得边缘。图像的梯度幅度值  $G(x, y)$  和方向  $\theta(x, y)$  分别由式 (4) 和式 (5) 确定。

$$G(x, y) = \sqrt{H(x, y)^2 + V(x, y)^2} \quad (4)$$

$$\theta(x, y) = \arctan\left(\frac{V(x, y)}{H(x, y)}\right) \quad (5)$$

式中, 利用图像灰度的一阶导数获取梯度幅度和方向, 以  $H(x, y)$  为水平方向  $x$  导数, 以  $V(x, y)$  为垂直方向  $y$  导数。

(4) 建筑物角点过滤。对边缘结果进行 LSD 直线检测 (Von Gioi 等, 2010) 操作, 经过线段生成、合并、评估的步骤后提取目标域影像直线段。约束线段长度去除过短线段, 将剩余部分视作不够完整但相对精确的建筑物边缘。基于建筑物角点多分布在建筑物边缘上的先验认知, 本文在长直线段上建立单向 3 个栅格宽度的缓冲区。缓冲区内分布的角点作为正确建筑物角点, 缓冲区外部角点被过滤删除。

### 2.1.2 源域建筑物边缘提取

建筑物边缘难以在影像中自动准确提取, 本文应用的边缘信息由源域标签数据获得, 仅在源域训练中使用。取原始掩膜向外膨胀 3 个像素与向内腐蚀 3 个像素的差值作为建筑物边缘标签, 用于衡量与预测边缘的差异程度。

本文综合考虑目标域建筑物角点和源域建筑物边缘作为形状先验信息, 可以在跨域特征对齐过程中建立双重约束条件: 一方面基于目标域的

角点分布构建空间区域约束, 提示建筑物大致区域与范围; 另一方面利用源域的边缘响应特征形成轮廓形态约束, 促进预测结果规则化。这种方式能够有效增强域适应网络对不同数据域中建筑物形态特征的联合感知能力。

## 2.2 基于高斯转换的位置先验信息融合

### 2.2.1 基于最大内切圆的高斯置信转换

针对 OSM 数据可能存在的类别属性错误问题, 本文借鉴魏东升和周晓光 (2019) 样本过滤方法计算 OSM 对应影像对象的异常程度, 剔除可能标注错误的 OSM 数据。

受限于 OSM 平台的标注方式和坐标转换过程, OSM 数据还容易存在位置偏移问题。样本更接近边界会有更高的不确定性水平和更明显的异常表现 (Viana 等, 2019)。针对上述问题, 本文引入最大内切圆心作为位置空间信息的关注焦点。多边形的最大内切圆指的是能够在多边形内部与多边形边缘相切的最大圆。因此, 建筑物轮廓的最大内切圆必定满足在轮廓内部, 且接近对应区域的中心位置。OSM 数据可能存在的位置偏移现象也最大程度地符合越靠近圆心属于前景目标的置信度越高的期望。部分图形的最大内切圆数量不唯一 (例如, 标准矩形的最大内切圆有无数个), 本文对一个 OSM 数据对象仅使用一个最大内切圆, 不考虑不唯一的情况。基于此, 本文建立最大内切圆圆心到圆上的先验信息置信度高斯映射关系。其中, 最大内切圆的圆心靠近对象中心, 可视作置信度最大的位置, 从圆心向圆上依照高斯函数逐渐递减, 直至到达圆上的置信度接近 0。具体映射关系如式 (6) 所示, 转换过程如图 3 所示。

$$Q(x) = \frac{A}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (6)$$

式中,  $Q$  表示 OSM 数据的先验信息置信强度,  $x$  表示周围像素点到最大内切圆心的距离, 用  $r$  表示对应 OSM 数据轮廓最大内切圆半径长度, 令  $r=3\sigma$ ,  $\sigma$  作为高斯函数中的方差。A 为延展系数, 使函数的取值范围扩展到  $[0, 255]$  的像元灰度范围。

### 2.2.2 位置先验信息融合

本文使用高斯函数将 OSM 建筑物轮廓数据转换为神经网络更易于处理的信息强度图, 作为附加通道与影像叠加, 共同输入分割网络进行训练,

可以为目标域提供一定的建筑物位置和大小信息,如图4所示。若图像区域无OSM数据,融合的附加通道相应赋值为0。训练中,源域的对数据由标签直接生成,用相同方式与影像做通道融合,以满足模型相同的数据输入维度。

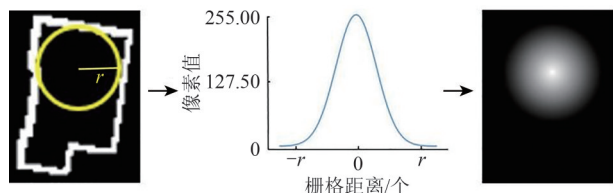


图3 建筑物轮廓高斯置信转换过程

Fig. 3 Gaussian confidence transformation process of building contour

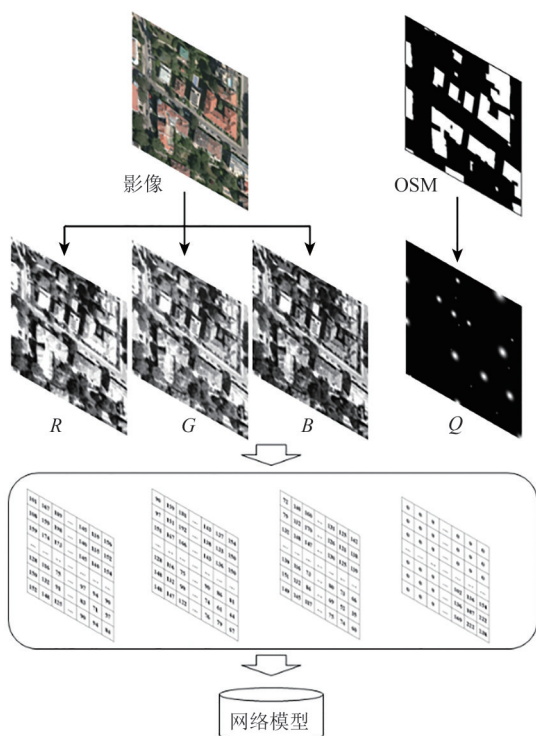


图4 附加通道的融合方式

Fig. 4 Integration method of additional channel

为了充分使用OSM数据的建筑物信息,方法在最后的提取结果中尝试补充模型计算遗漏的正确范围。根据OSM建筑物数据与影像中对应预测建筑物的交集匹配重叠率,优化建筑物提取的完整度与准确率,如图5所示。顾及数据质量问题,利用重叠过滤方法判断OSM建筑物和模型计算建筑物之间的重叠比例。具体计算公式如下:

$$\frac{\text{Inter}(\text{Area}_{\text{OSM}}, \text{Area}_{\text{res}})}{\text{Area}_{\text{OSM}}} \geq T_{\text{int}} \quad (7)$$

式中,  $\text{Area}_{\text{OSM}}$  为OSM代表的区域,  $\text{Area}_{\text{res}}$  为提取计算结果代表的区域,  $\text{Inter}$  代表交集匹配操作,  $T_{\text{int}}$  表示数据交集匹配重叠率。当  $T_{\text{int}}$  越大时,表示OSM数据区域内包含的预测建筑物区域越多。考虑到OSM数据可提供的完善数据与边界优化作用,和数据偏移等质量问题的矛盾性,本文将  $T_{\text{int}}$  设置为0.8。

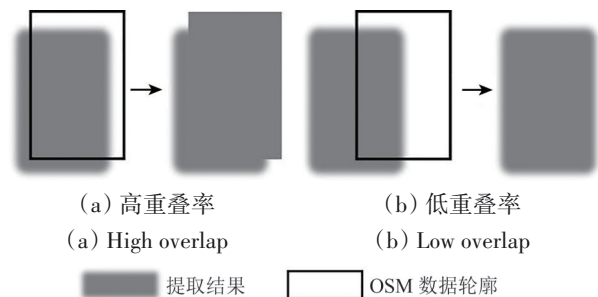


图5 OSM数据叠加示意图

Fig. 5 Diagram of OSM data overlap

## 2.3 基于对抗学习的域适应提取模型构建

无论图像来自源域还是目标域,它们的分割提取结果均为建筑物对象,在全局和局部上具有很强的相似性。对抗学习旨在不断循环鉴别器训练阶段和生成器训练阶段,相互提高鉴别器和生成器的能力,最终最小化图像在生成器中的损失,同时最大化源域和目标域的输出相似度。本文设计了基于对抗学习的建筑物域适应提取模型AU\_AdaptNet,由生成器和鉴别器两部分组成。其中,分割网络AU\_Net作为生成器,负责影像到建筑物提取结果的映射过程,鉴别器负责判断源域与目标域计算结果的分布一致性。

### 2.3.1 AU\_AdaptNet 总体架构

建筑物域适应提取模型AU\_AdaptNet如图6所示。其中,  $x^s$  表示源域数据,  $x^t$  表示目标域数据,  $y_{\text{seg}}$  和  $y_{\text{edge}}$  分别表示建筑物和边缘的模型计算结果,  $\tilde{y}_{\text{corner}}$  和  $\tilde{y}_{\text{edge}}$  分别表示2.1节中提取的角点和边缘信息,  $\tilde{y}_{\text{seg}}$  表示建筑物标签。生成器AU\_Net用  $G$  表示,鉴别器用  $D$  表示,  $L_{\text{adv}}$  表示对抗损失,  $L_{\text{seg}}$  表示源域分割损失,  $L_{\text{edge}}$  表示源域边缘损失,  $L_{\text{corner}}$  表示目标域角点损失。源域利用建筑物掩膜和边缘的预测结果分别计算损失,虽然目标域同样输出2通道,但仅利用建筑物预测掩膜参与对抗学习。

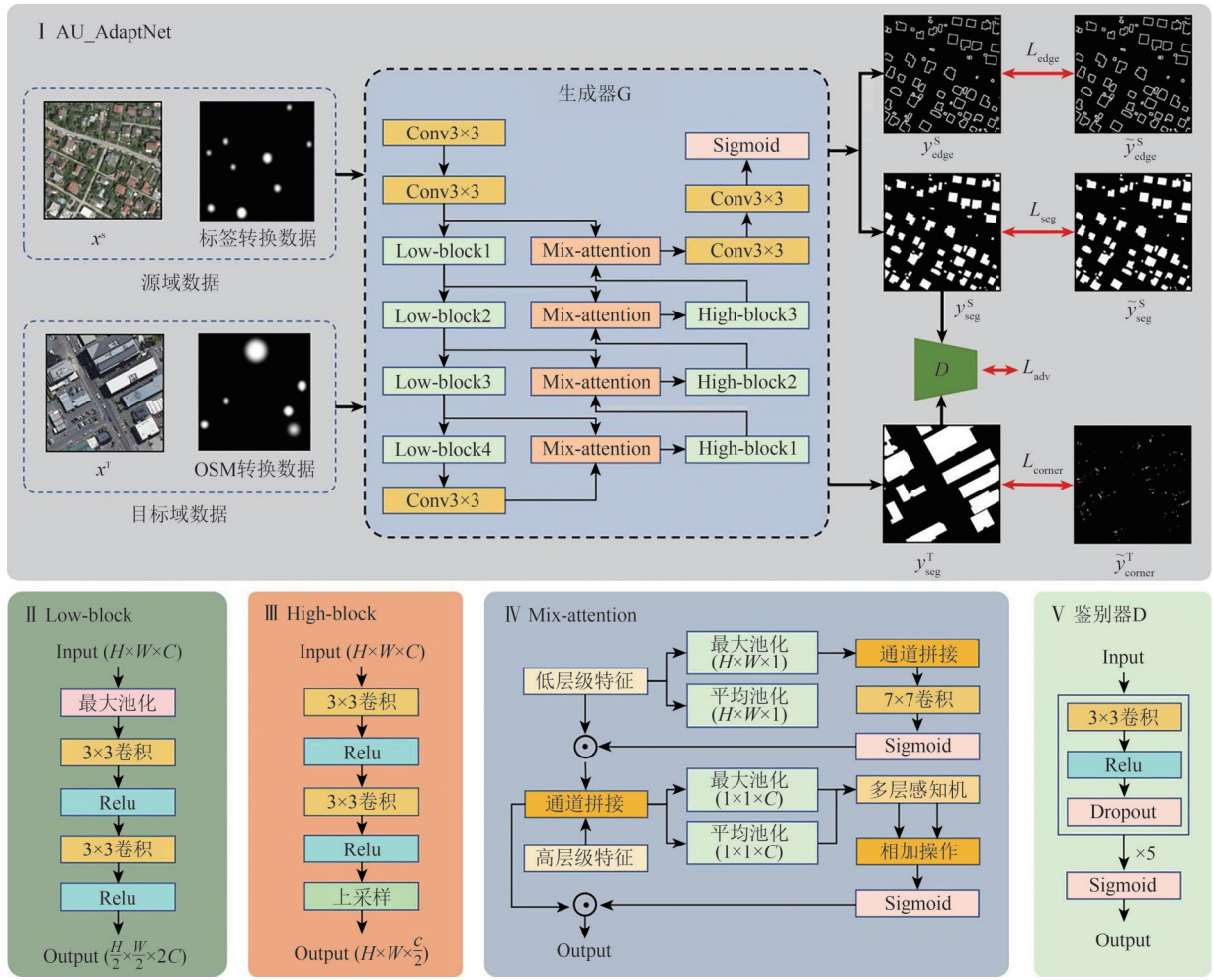


图6 AU\_AdaptNet模型架构

Fig. 6 AU\_AdaptNet architecture

AU\_AdaptNet的训练过程可以分为鉴别器训练和生成器训练两个阶段的不断循环。其中，用 $P_s$ 表示源域的建筑物掩膜网络输出， $P_T$ 表示目标域的建筑物掩膜网络输出， $h$ 和 $w$ 分别表示张量的高和宽， $D(P_s)^{(h,w)}$ 和 $D(P_T)^{(h,w)}$ 分别表示源域和目标域预测结果经由鉴别器生成的结果。 $P^{(h,w)}(1)$ 和 $P^{(h,w)}(0)$ 表示与 $D(P_s)^{(h,w)}$ 和 $D(P_T)^{(h,w)}$ 形状一致，数值全为1或0的张量矩阵。本文规定鉴别器输出结果越接近全为1的张量矩阵，代表输入数据更近似源域数据；鉴别器输出结果越接近全为0的张量矩阵，代表输入数据更近似目标域数据。其中与1和0的近似程度用BCE损失函数计算。

鉴别器训练阶段，当鉴别器中输入源域数据时，可得源域鉴别损失 $L_D(P_s)$ 如下：

$$L_D(P_s) = L_{BCE} \left( D(P_s)^{(h,w)}, P^{(h,w)}(1) \right) = - \sum_{i=1}^h \sum_{j=1}^w \log \left( D(P_s)^{(i,j)} \right) \quad (8)$$

当鉴别器中输入目标域数据时，可得目标域鉴别损失 $L_D(P_T)$ 如下：

$$L_D(P_T) = L_{BCE} \left( D(P_T)^{(h,w)}, P^{(h,w)}(0) \right) = - \sum_{i=1}^h \sum_{j=1}^w \log \left( 1 - D(P_T)^{(i,j)} \right) \quad (9)$$

源域和目标域数据各进行一次输入，鉴别器一次完整训练过程中的鉴别损失 $L_D(P)$ 如下：

$$L_D(P) = L_D(P_s) + L_D(P_T) = - \sum_{i=1}^h \sum_{j=1}^w \left( \log \left( D(P_s)^{(i,j)} \right) + \log \left( 1 - D(P_T)^{(i,j)} \right) \right) \quad (10)$$

生成器训练阶段，为了使目标域经生成器的输出结果分布更接近源域，设定仅针对目标域的对抗损失 $L_{adv}$ 如下：

$$L_{adv} = L_{BCE} \left( D(P_T)^{(h,w)}, P^{(h,w)}(1) \right) = - \sum_{i=1}^h \sum_{j=1}^w \log \left( D(P_T)^{(i,j)} \right) \quad (11)$$

生成器对源域图像的分割过程由分割损失引

导, 用BCE损失函数设计分割损失如下:

$$L_{\text{seg}} = -\left(\tilde{y}_{\text{seg}} \log(y_{\text{seg}}) + (1 - \tilde{y}_{\text{seg}}) \log(1 - y_{\text{seg}})\right) \quad (12)$$

式中,  $y_{\text{seg}}$  表示源域图像经生成器的预测值,  $\tilde{y}_{\text{seg}}$  表示源域图像的标签像素值。

依据所提角点位于建筑物预测区域的比例, 设计角点损失函数, 该损失函数仅在目标域训练过程中使用。损失函数  $L_{\text{corner}}$  如式 (13) 所示。

$$L_{\text{corner}} = 1 - \frac{\text{Sum}_{\text{inter}}}{\text{Sum}_{\text{all}}} \quad (13)$$

式中,  $\text{Sum}_{\text{all}}$  表示由 2.1.1 节得到的全部角点数量,  $\text{Sum}_{\text{inter}}$  表示其中目标域预测结果覆盖的角点数量。

针对边缘损失函数  $L_{\text{edge}}$  的计算, 本文也使用 BCE 损失函数, 如式 (14) 所示。

$$L_{\text{edge}} = -\left(\tilde{y}_{\text{edge}} \log(y_{\text{edge}}) + (1 - \tilde{y}_{\text{edge}}) \log(1 - y_{\text{edge}})\right) \quad (14)$$

式中,  $y_{\text{edge}}$  表示网络输出的目标预测值,  $\tilde{y}_{\text{edge}}$  表示 2.1.2 节得到的真实建筑物边缘标签。

对抗损失  $L_{\text{adv}}$  结合分割损失  $L_{\text{seg}}$ 、角点损失  $L_{\text{corner}}$ 、边缘损失  $L_{\text{edge}}$ , 构建总的生成训练过程损失函数如式 (15) 所示,  $\lambda_{\text{seg}}$ 、 $\lambda_{\text{corner}}$ 、 $\lambda_{\text{edge}}$ 、 $\lambda_{\text{adv}}$  为各损失项的权重。

$$L_{\text{total}} = \lambda_{\text{seg}} L_{\text{seg}} + \lambda_{\text{corner}} L_{\text{corner}} + \lambda_{\text{edge}} L_{\text{edge}} + \lambda_{\text{adv}} L_{\text{adv}} \quad (15)$$

综上, 对抗学习的最终目标是最小化生成器中的损失, 同时最大化目标域预测被认为是源域预测的概率。

### 2.3.2 生成器结构

融合注意力机制的分割网络 AU\_Net 以 U 型网络为基础, 低层级特征计算模块 (Low-block) 逐步降低图像的空间维度, 同时增加特征通道数, 捕获图像的上下文信息, 如图 6 II 所示。高层级特征计算模块 (High-block) 逐步恢复图像的空间维度和细节信息, 如图 6 III 所示。跳跃连接过程中引入注意力机制 (Woo 等, 2018) 构建注意力混合模块 (Mix-attention), 如图 6 IV 所示。每层编码器输往解码器的特征层送入空间注意力结构, 经过权重调整后与解码器的对应层叠合, 再将叠合后的部分送入通道注意力结构, 完成注意力的计算过程。通过引入注意力混合模块, 网络能够有效整合全局信息和多通道信息, 增强建筑物的有效特征, 提高网络对目标域建筑物对象的关注程度。最终分割网络输出 2 层通道, 分别对应建筑物掩膜和边缘轮廓。

### 2.3.2 鉴别器结构

针对建筑物域适应提取, 鉴别器的具体任务是区分输入数据来自源域还是目标域。本文设计的鉴别器网络由 5 个卷积层组成, 核数为  $4 \times 4$ , 步长为 2, 通道数分别为  $\{64, 128, 256, 512, 1\}$ 。除最后一层外, 每个卷积层后面都有一个 ReLU 激活函数。

## 3 实验数据与设置

本研究共设计了 4 个实验, 以评价本文所述形状与位置先验信息引导的建筑物域适应提取方法。第 1 个实验评价了本文方法的建筑物域适应提取效果, 并与其他相关无监督和半监督方法进行了对比。第 2 个实验分析了形状先验信息为目标域建筑物带来的提升效果, 并展示与其他基础模型结合的灵活性。第 3 个实验评价了目标域建筑物角点自动提取的准确度情况, 及角点提取方法在建筑物域适应提取中的消融性能。第 4 个实验分析了位置先验信息为目标域建筑物提供的空间辅助作用。

### 3.1 评价指标

为了评价本文方法精度, 实验共涉及准确率 (Precision)、查全率 (Recall)、 $F$  度量 (F-measure) 和交并比 IoU (Intersection over Union) 4 个定量评价指标。定义分别如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (16)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (17)$$

$$F = \frac{(1 + \alpha) \times \text{Precision} \times \text{Recall}}{\alpha \times \text{Precision} + \text{Recall}} \quad (18)$$

$$\text{IoU} = \frac{\text{TP}}{\text{FP} + \text{TP} + \text{FN}} \quad (19)$$

式中, TP 和 FP 分别表示提取结果中正确和错误的像素数量, FN 表示被遗漏的正确像素数量。 $F$  指标中  $\alpha$  的值设为 1 (即 F1 指标), 表示均衡考虑准确率与查全率的影响。

### 3.2 数据集介绍

本文实验共用到 5 个开源数据集。其中, Inria 航空图像标签数据集 IAILD (Inria Aerial Image Labeling Dataset) 是一个面向城市建筑物检测的遥感图像数据集, 像素空间分辨率为 0.3 m, 包括 180 张空间分辨率为  $5000 \times 5000$  的训练图像和相应像素级标记图像 (Maggiori 等, 2017), 如图 7 (a)

所示。WHU 航空影像数据集包含了多个城市和乡村地区的影像与标签数据，空间分辨率为 0.3 m，每幅影像的大小为 512×512 像素，共计 8188 张（Ji 等，2019），如图 7（b）所示。WHU 全球城市卫星数据集的影像来自多种遥感资源，包含 204 张影像，每幅影像的大小为 512×512 像素，空间分辨率

从 0.3 m 到 2.5 m 不等（Ji 等，2019），如图 7（d）所示。Potsdam 数据集包括建筑群、街道和密集聚落等（Gerke，2014），针对本文建筑物提取的任务，将标签数据中的建筑物赋为白色，其他类型赋为黑色，如图 7（c）所示。

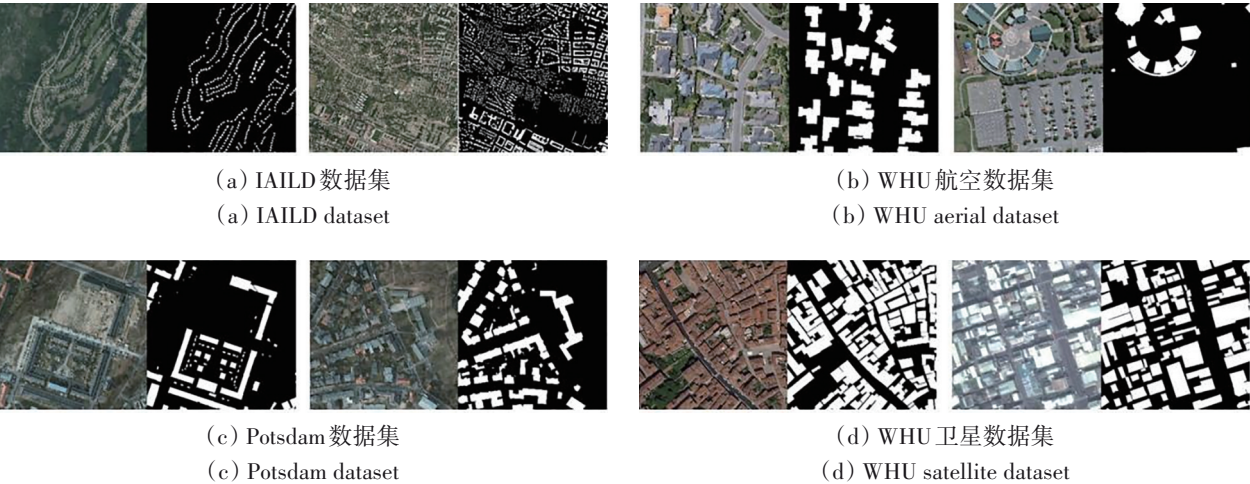


图7 数据集示例  
Fig.7 Examples of datasets

此外，本文使用了航空图像分割数据集 AISD（Aerial Image Segmentation Dataset）。该数据集涵盖柏林、芝加哥、巴黎和苏黎世等城市的不同区域，影像来自谷歌地图的航空遥感图像，标签数据完全来自 OpenStreetMap 建筑、道路和背景的像素级标报信息（Kaiser 等，2017）。AISD 数据集内的建筑物是标签数据的红色区域，数据示例如图 8 所示。

般性质量问题，如建筑物的遗漏标注、错误标注、与其他地物类型的混淆等。在数据集包括的 4 个城市中各选取一幅影像，示例其问题数据及标签如图 9 所示。其中，柏林区域对应的影像标签右下方的建筑物缺失，且图像中部的铁路沿线及盖顶也被归为建筑物；芝加哥区域左方的密集建筑物部分缺失且有一定的位置偏移，右方大型建筑物整体缺失；巴黎区域建筑物标注错误；苏黎世区域道路标签的错误标注影响了建筑物的标注结果。图 9 中的示例进一步证明了 OSM 数据的质量问题，以及对 OSM 数据进行质量控制的必要性。

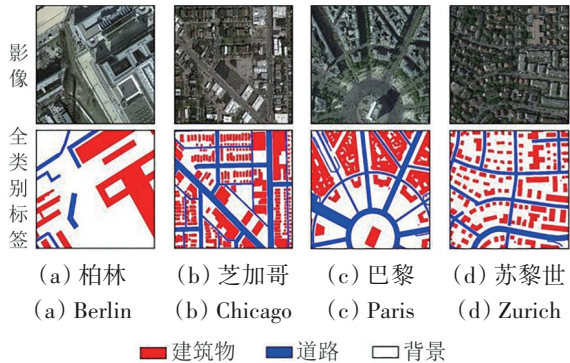


图8 AISD数据集示例  
Fig.8 Examples of AISD dataset

观察 AISD 数据集内的图像与标签，可以发现由 OSM 数据制作的数据集确实存在 OSM 数据的一

### 3.3 实验设置

实验训练硬件环境为 Intel Core i9-10900K 的 CPU，NVIDIA GeForce RTX 3060 显卡，具备 12 GB 显存，软件使用 Pycharm 社区版开发平台，Python3.8 开发语言和 Pytorch1.12 的深度学习框架。训练 AU\_AdaptNet 时， $\lambda_{seg}$ 、 $\lambda_{corner}$ 、 $\lambda_{edge}$  全部设置为 1，赋予模型训练中各损失项相同的重要性。这种等权重策略旨在避免因人为调整权重而引入的偏差，使模型能够平等优化各项损失，更全面地学习任务中的不同特征。生成器和鉴别器的优化器都选

用 $\beta = (0.5, 0.999)$ 的Adam，初始学习率为0.0002，当评价指标停止提升时学习率降低10%。

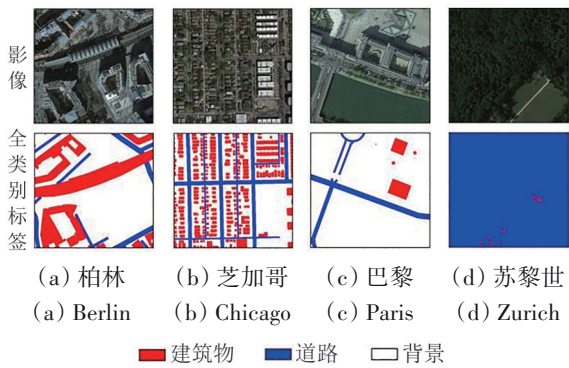


图9 OSM问题标签示例  
Fig. 9 Examples of OSM problem label

4 实验结果与分析

4.1 本文方法与无监督和半监督方法的对比实验

本实验采用 IAILD 数据集作为源域数据集，AISD 数据集作为目标域数据集，对比了源域预训练分割网络在目标域的直接测试、两个经典的无监督域适应网络 CycleGAN 和 AdaptNet、两个半监督域适应网络 SSFT 和 SS-ADA、本文方法、目标域数据集内训练和测试的结果。源域标签符合源域图像中建筑物的真实分布，而目标域 OSM 数据在现实中存在不完整性和不均匀性。为了避免源域的位置先验信息相比目标域的位置先验信息详尽太多使得模型训练中注意力偏差，本文方法在源域标签中抽取 50% 比例的建筑物对象与源域图像融合，在目标域标签中抽取 0—50% 随机比例的建筑物对象模拟 OSM 数据与目标域图像融合。两个半监督域适应模型 SSFT 和 SS-ADA 首先依照 Basak 和 Yin (2023) 和 Yan 等 (2024) 所述，均在目标域使用 25% 训练图像数量比例的高质量样本 (图像中全部建筑物在标签中均有准确对照)，又采取与本文方法中目标域相同抽样方式，使用同等数量但每张标签中抽取 0—50% 随机比例建筑物的低质量样本，与本文方法对比。

实验结果的精度如表 1 所示。AU\_AdaptNet 加入形状损失的结果与仅用源域数据训练网络的结果 F1 指标由 0.6921 提升到了 0.8120，提升了接近 12%，进一步融合 OSM 数据后 F1 指标达到了 0.8317。并且该实验条件仅在目标域融合 0—50%

比例建筑物对象的位置先验信息下，随着目标域可利用的位置先验信息增加，提取效果有更高的潜力。两种无监督域适应模型中，AdaptNet 的 F1 指标相比直接预测结果提升了 0.05，CycleGAN 由于仅在图像级进行风格转换，并未对其他特征进行对齐操作，精度甚至出现略微下降。两种半监督域适应模型 SSFT 和 SS-ADA 的 F1 指标分别达到了 0.84 和 0.85，比本文所述方法的 0.83 略高，但它们在训练过程中使用了目标域高质量样本。当目标域标签中建筑物抽样至 50% 覆盖率以下时，SS-ADA 无法依据现有数据生成高质量伪标签，F1 和 IoU 指标分别下降到 0.7542 和 0.6213，SSFT 采用目标域微调策略，低质量样本甚至使得它的精度低于源域预训练模型的直接预测。实验证明，本文所述方法相对传统无监督方法有明显精度提升。与半监督方法相比，本文方法也可以利用较低建筑物覆盖率的低质量众源数据，辅助提升建筑物域适应提取的效果。

表 1 本文提出方法与不同无监督和半监督方法的对比实验精度

| Table 1 Accuracy comparison with unsupervised and semi-supervised methods |                       |               |               |
|---------------------------------------------------------------------------|-----------------------|---------------|---------------|
| 方法类型                                                                      | 方法名称                  | 方法精度          |               |
|                                                                           |                       | F1            | IoU           |
| 源域训练→<br>目标域测试                                                            | AU_Net                | 0.6921        | 0.5534        |
| 无监督方法                                                                     | CycleGAN              | 0.6879        | 0.5514        |
|                                                                           | AdaptNet              | 0.7405        | 0.6058        |
| 半监督方法                                                                     | SSFT(目标域高质量样本)        | 0.8403        | 0.7246        |
|                                                                           | SS-ADA(目标域高质量样本)      | <b>0.8521</b> | <b>0.7428</b> |
|                                                                           | SSFT(目标域低质量样本)        | 0.6138        | 0.5484        |
|                                                                           | SS-ADA(目标域低质量样本)      | 0.7542        | 0.6213        |
| 本文方法                                                                      | AU_AdaptNet           | 0.7726        | 0.6381        |
|                                                                           | AU_AdaptNet+shape     | 0.8120        | 0.6796        |
|                                                                           | AU_AdaptNet+shape+OSM | <b>0.8317</b> | <b>0.7012</b> |
| 目标域中训练<br>和测试                                                             | AU_Net                | 0.8937        | 0.8035        |

注：加粗数值代表半监督方法和本文方法中的最高精度。

4.2 形状先验信息通用性实验

本实验采用 IAILD、Potsdam 和 WHU 数据集分别作为源域和目标域。使用方法包括：源域预训练模型在目标域测试；源域预训练和目标域测试中添加形状先验信息 (Source only+shape)；两大经

典域适应架构 CycleGAN 和 AdpatNet；顾及边缘形状的 SegAD 架构（Zheng 等，2019）；CycleGAN 和 AdpatNet 添加形状先验信息；目标域内训练和测试。不同数据集分别作为源域和目标域的精度如表 2 所示，部分实验结果如图 10 至图 12 所示。

表 2 跨数据集形状先验通用性实验结果  
Table 2 Results of shape prior generalization across all dataset pairs

| 方法                | IAILD→Potsdam |               | IAILD→WHU     |               | WHU→IAILD     |               | WHU→Potsdam   |               | Potsdam→WHU   |               | Potsdam→IAILD |               |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                   | F1            | IoU           | F1            | IoU           | F1            | IoU           | F1            | IoU           | F1            | IoU           | F1            | IoU           |
| 源域训练→目标域测试        | 0.7027        | 0.5519        | 0.7238        | 0.5763        | 0.6271        | 0.4605        | 0.6536        | 0.4872        | 0.7037        | 0.5464        | 0.5709        | 0.4096        |
| Source only+shape | 0.7384        | 0.5630        | 0.7552        | 0.5894        | 0.6615        | 0.4927        | 0.7252        | 0.5760        | 0.7233        | 0.5641        | 0.6193        | 0.4429        |
| CycleGAN          | 0.6916        | 0.5349        | 0.4976        | 0.3790        | 0.5846        | 0.4208        | 0.6546        | 0.4828        | 0.5644        | 0.3938        | 0.4512        | 0.2952        |
| AdaptNet          | 0.7843        | 0.6248        | 0.7966        | 0.6559        | 0.7439        | 0.5916        | 0.8035        | 0.6417        | 0.7823        | 0.6211        | 0.6547        | <u>0.5178</u> |
| SegAD             | 0.8002        | 0.6637        | <u>0.8229</u> | 0.6753        | 0.7431        | 0.6058        | 0.8109        | 0.6667        | 0.7844        | 0.6253        | 0.6514        | 0.5128        |
| CycleGAN+shape    | 0.7251        | 0.5778        | 0.6665        | 0.4917        | 0.6218        | 0.4459        | 0.7058        | 0.5325        | 0.6814        | 0.5207        | 0.5606        | 0.4730        |
| AdaptNet+shape    | <u>0.8128</u> | <u>0.6676</u> | 0.8213        | <u>0.6755</u> | <u>0.7848</u> | <u>0.6483</u> | <u>0.8385</u> | <u>0.6784</u> | <u>0.7982</u> | <u>0.6336</u> | <u>0.6649</u> | 0.5104        |
| 目标域中训练和测试         | <b>0.9218</b> | <b>0.8653</b> | <b>0.9194</b> | <b>0.8668</b> | <b>0.9075</b> | <b>0.8036</b> | <b>0.9218</b> | <b>0.8653</b> | <b>0.9194</b> | <b>0.8668</b> | <b>0.9075</b> | <b>0.8036</b> |

注：表中数值加粗代表最高精度，下划线代表次高精度。

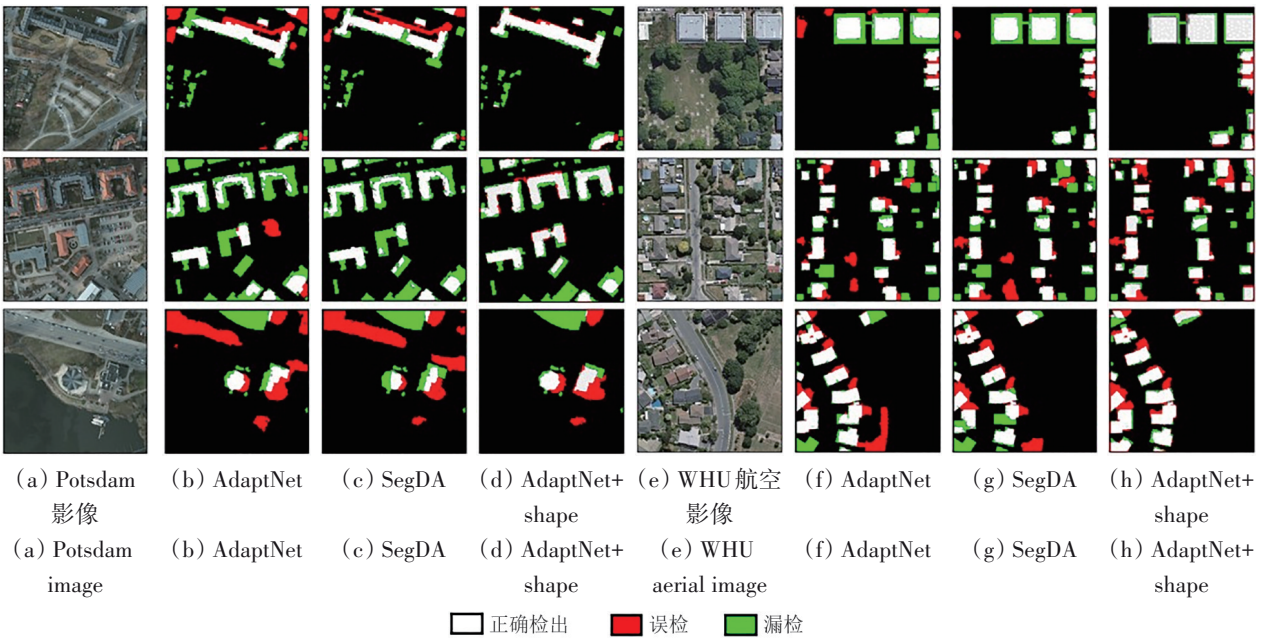


图 10 IAILD 数据集作源域的形状先验通用性实验可视化

Fig.10 Visualization of shape prior generalization experiment using IAILD dataset as source domain

由表 2 可见，IAILD 数据集作源域，Potsdam 和 WHU 数据集分别作为目标域时，添加形状先验信息后的 AdaptNet 模型在 F1 和 IoU 上比未添加的模型均提高了约 4%。同时，添加了形状先验信息的模型比 SegAD 也表现出更高的精度。此外，即使不使用域适应网络，仅仅在原始语义分割网络中添加形状先验信息也获得了一定的精度提升。WHU 数据集作为源域，Postdam 和 IAILD 数据集分别作为目标域的提取结果中，由于 WHU 和 Postdam、

IAILD 数据集之间的数据差异，直接预测的 IoU 只有 0.49 和 0.46。AdaptNet 的 IoU 分别提升到 0.64 和 0.59，SegAD 表现更好，达到了 0.67 和 0.61。AdaptNet 加入形状先验信息后精度进一步提升，相比直接预测的 IoU 均提升了 0.19。由 Potsdam 数据集作源域，IAILD 和 WHU 数据集作为目标域时，Potsdam 和 IAILD 两个数据集初始预测的 F1 和 IoU 仅有 0.57、0.41。SegAD 预测结果的 F1 和 IoU 为 0.65、0.51，有一定精度提升。AdaptNet 添加形状先验信息后，

F1 和 IoU 达到了 0.67、0.51。在数据较为接近的 Potsdam 和 WHU 数据集上，最终的 F1 和 IoU 已经达到了 0.80、0.63。

图 10 展示了由 IAILD 数据集作为源域的建筑物域适应提取结果。IAILD 数据集本身具有丰富的建筑物场景，在 AdaptNet 和 SegAD 这两个模型上都表现出不错的可视化效果。添加形状先验信息的 AdaptNet 模型相比原始模型都减少了漏检和误检区域，尤其是真实道路的误检几乎消失，且建筑物边缘处的漏检更少。图 11 为 WHU 航空数据集作为源域的建筑物域适应提取结果。AdaptNet 加

入形状先验信息后可视化效果提升明显，建筑物对象几乎都被发现，预测结果更加贴合真实标签，房屋阴影和停车场等复杂场景部分的误检也更少。图 12 展示了 Potsdam 数据集作为源域，IAILD 和 WHU 航空数据集作为目标域的提取结果。可见，添加形状先验信息的模型在视觉上表现出了更好的效果。具体表现在非建筑物要素对结果的扰动更少，红色误检部分几乎只存在于建筑物边缘处，建筑物识别性能更好，除个别特殊样式建筑物外均被检出。

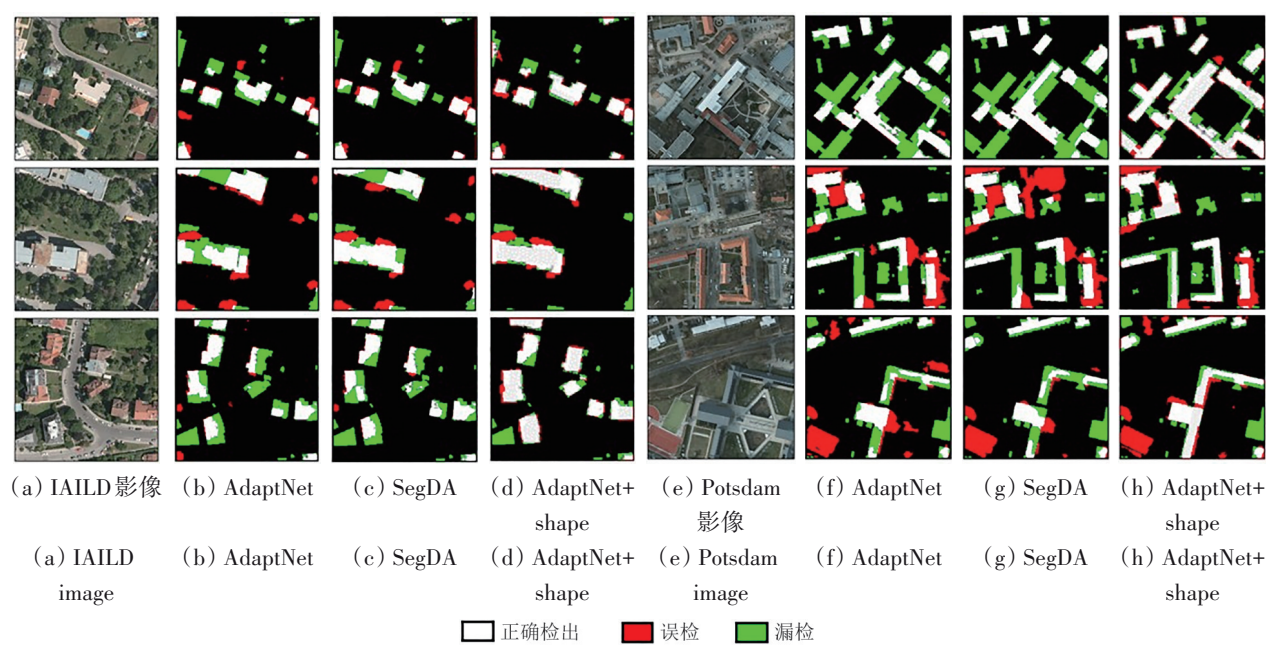


图 11 WHU 数据集作源域的形状先验通用性实验可视化

Fig. 11 Visualization of shape prior generalization experiment using WHU dataset as source domain

4.3 角点自动提取准确性及消融实验

准确的建筑物角点才能用于计算角点损失，在模型训练反向传播时起到提供目标域建筑物先验信息的作用。针对目标域建筑物角点提取的完整率评价中，一方面建筑物角点的提取没有规范的验证标准及数据，难以利用完整度进行评估，另一方面本研究没有在目标域提取所有建筑物角点的绝对需求。因此，本实验侧重分析建筑物角点提取结果的准确率及对本文域适应模型性能的影响。首先，实验分别在 3 个数据集中各抽取 100 张影像，对单一及组合方法的 4 种情况进行实验。角点的提取结果如图 13 所示，黄色线框表示真实建

筑物轮廓，白色像素表示提取的建筑物角点。标签建筑物对象膨胀 3 个像素后，所有提取角点中位于建筑物内部的结果视为正确提取，角点提取准确率（式（16））如表 3 所示，表示角点正确提取的比例。然后，以 IDILD 和 WHU 两个数据集为例，在 AU\_AdaptNet 中仅使用角点先验，不使用边缘和位置先验对比不同角点提取方法时的目标域建筑物 IoU 指标，如表 4 所示。最后，为了分析本文 3 个栅格宽度作为角点过滤阈值的合理性，对比不同栅格宽度对实验结果的影响，对不使用角点先验（设置为 0）与 1—5 个栅格宽度逐次实验，统计 IoU 结果如图 14 所示。

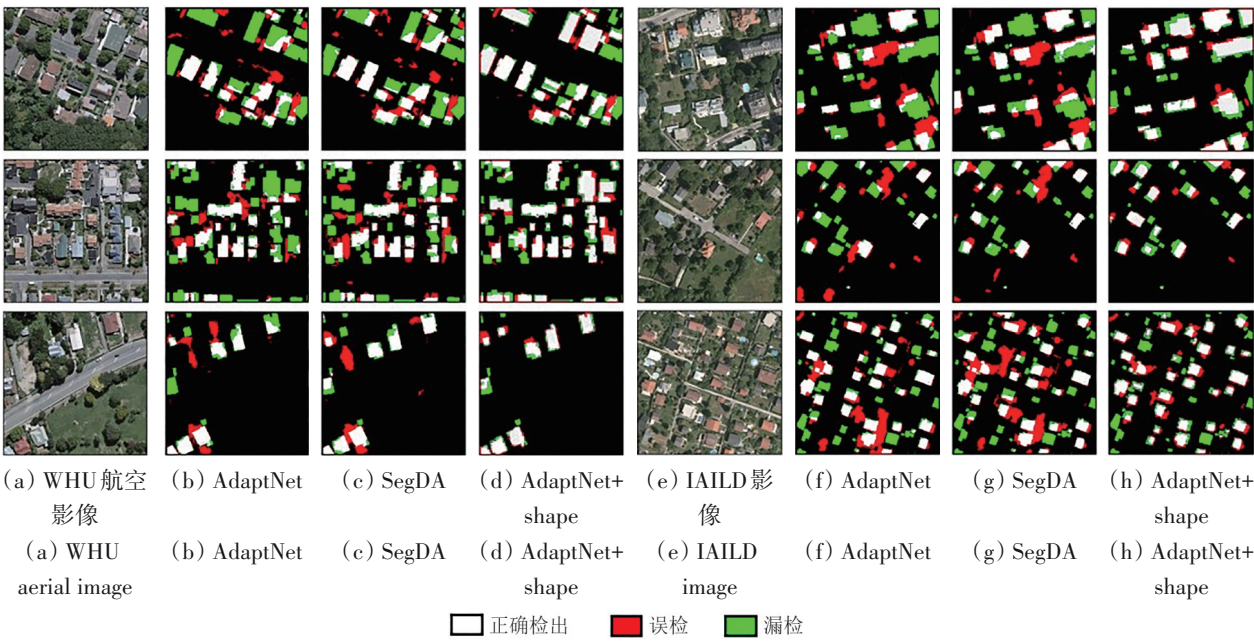


图 12 Potsdam 数据集作源域的形状先验通用性实验可视化  
Fig. 12 Visualization of shape prior generalization experiment using Potsdam dataset as source domain

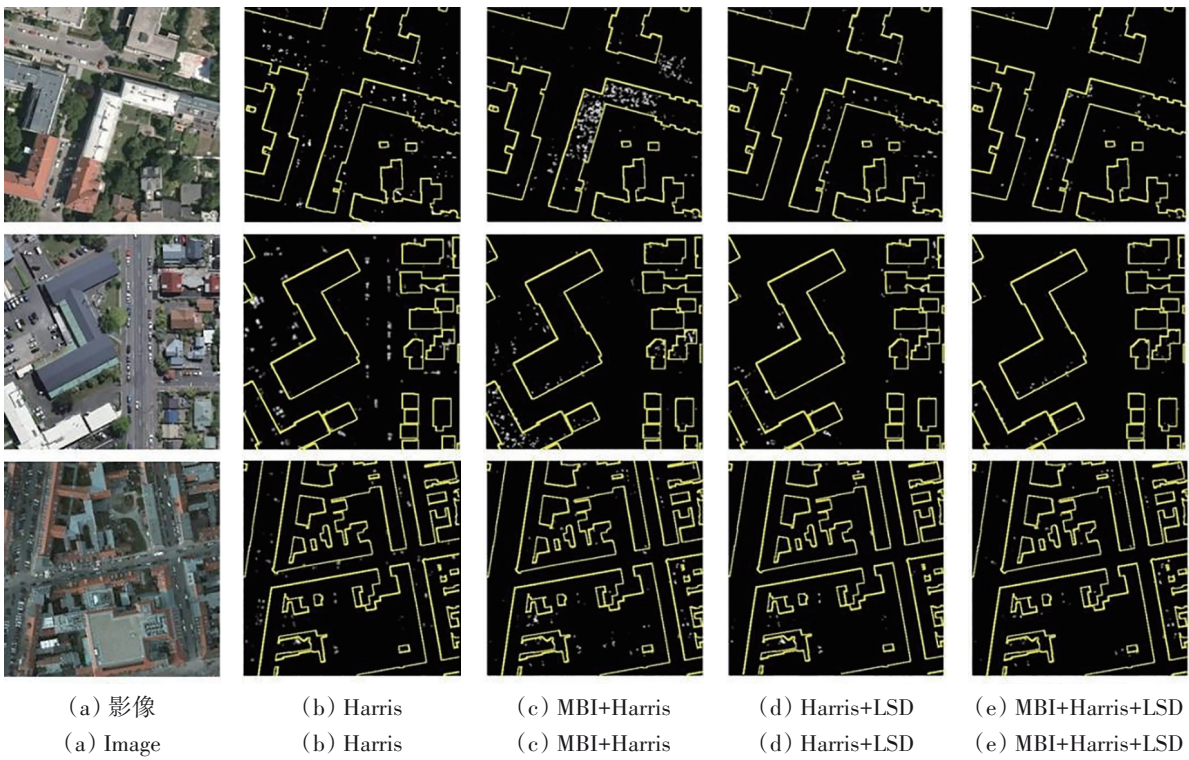


图 13 不同方法下建筑物角点提取结果可视化  
Fig. 13 Visualization of building corner points extraction using different methods

由表 3 和图 13 可见：本研究结合了 MBI、Harris、LSD 算法的角点提取准确率最高，在 3 个数据集中分别达到了 90.54%、90.24%、91.72%。仅基于原始影像提取的角点数量最多，但几乎包括所有地物要素的拐角，建筑物角点所占比例太

低；建筑物指数图提取的角点准确率相比原始影像直接提取有明显提升，但是亮度突出且形状规则的内容都易被误认为建筑物。如图 13（c）所示，车辆相对于道路背景具有亮度突出的特点和类似建筑物的规则形状，导致许多车辆得到了较

高的建筑物指数，降低了建筑物角点的提取准确率。长直线缓冲过滤原始影像角点的方式对于剔除非建筑物角点有效，但依然会受到道路等大型规则地物的影响。从视觉效果和精度评价中都可以看出，长直线缓冲过滤建筑物指数图提取的角点是上述几种方法中最好的。一方面，其他地物角点因建筑物指数强度较低而被忽略；另一方面，车辆等要素也因为不在长直线缓冲区中而被过滤。

表 3 不同方法下建筑物角点提取准确率对比  
Table 3 Comparison of building corner extraction precision using different methods

| 方法             | IAILD        | WHU          | Potsdam      |
|----------------|--------------|--------------|--------------|
| Harris         | 43.75        | 32.53        | 38.49        |
| MBI+Harris     | 83.54        | 76.42        | 64.82        |
| Harris+LSD     | 79.28        | 72.95        | 88.61        |
| MBI+Harris+LSD | <b>90.54</b> | <b>90.24</b> | <b>91.72</b> |

注：加粗数值代表最高精度。

表 4 不同角点提取方法的建筑物提取结果  
Table 4 IoU of building extraction with different corner extraction methods

| 方法             | WHU->IAILD    | IAILD->WHU    |
|----------------|---------------|---------------|
| Harris         | 0.4273        | 0.4416        |
| MBI+Harris     | 0.6004        | 0.6127        |
| Harris+LSD     | 0.5886        | 0.5834        |
| MBI+Harris+LSD | <b>0.6412</b> | <b>0.6758</b> |

注：加粗数值代表最高精度。

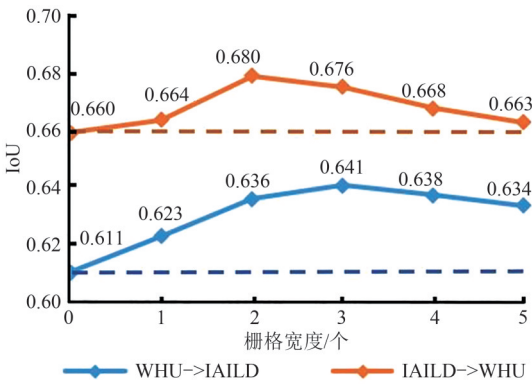


图 14 不同栅格宽度的建筑物提取结果折线图

Fig.14 Line chart of extraction IoU for buildings with different grid widths

如表 4 数据所示，本文设计的角点提取组合方法在两组实验中均取得了最高精度，且不同角点提取方法下的域适应精度高低与表 3 中对应目标域

建筑物角点提取准确度基本呈正相关关系。实验结果意味着本文设计的组合方法可以提供较高的角点准确度，继而为域适应提取模型提供有利的先验信息。

综合图 14 的结果可以看出，加入目标域建筑物角点作为先验信息对实验结果有积极作用，且随着栅格宽度增加，精度整体上呈现先提高后降低的趋势。可以分析得到，当栅格宽度过小时，部分可用角点被过滤剔除导致角点提取数量较少；当栅格宽度过大时，降低了真实角点准确度，甚至更大的栅格宽度会失去过滤的原始意义。实验结果说明，本文使用的 3 个栅格是一种较为合适的角点过滤宽度。

4.4 位置先验信息融合比例分析实验

OSM 数据在不同地区具有不均匀性，无法保证网络训练和预测时的辅助信息比例一致。因此，需要分析不同 OSM 融合比例对实验结果的影响差异，甚至目标域无 OSM 数据条件下的实验情况。本实验以 IAILD 数据集作源域，以 WHU 卫星数据集作目标域。首先，在源域融入 50% 标签建筑物对象的附加通道，仅用源域数据训练模型，分别比较目标域融入 0%（附加通道像素数值全为 0）、25%、50%、75% 不同比例建筑物对象附加通道的预测情况。其次，不融合附加通道，直接用源域数据训练模型在目标域数据上测试，作为基础网络（BaseNet）的实验结果，与先前目标域融入 0% 位置先验信息的实验结果对比，分析目标域无 OSM 数据条件下，仅源域附加通道含有一定位置先验信息是否对方法产生负面影响。建筑物跨域提取中融合不同 OSM 比例的测试精度如表 5 所示，可视化结果如图 15 所示。

表 5 跨域提取中融合不同 OSM 比例的测试结果

Table 5 Accuracy with different OSM fusion ratios in cross-domain extraction

| 融合比例    | Recall/% | Precision/% | F1/%  |
|---------|----------|-------------|-------|
| 0%      | 58.20    | 67.16       | 62.36 |
| 25%     | 62.51    | 67.68       | 64.99 |
| 50%     | 66.18    | 68.11       | 67.13 |
| 75%     | 69.19    | 70.12       | 69.65 |
| BaseNet | 55.39    | 67.19       | 60.72 |

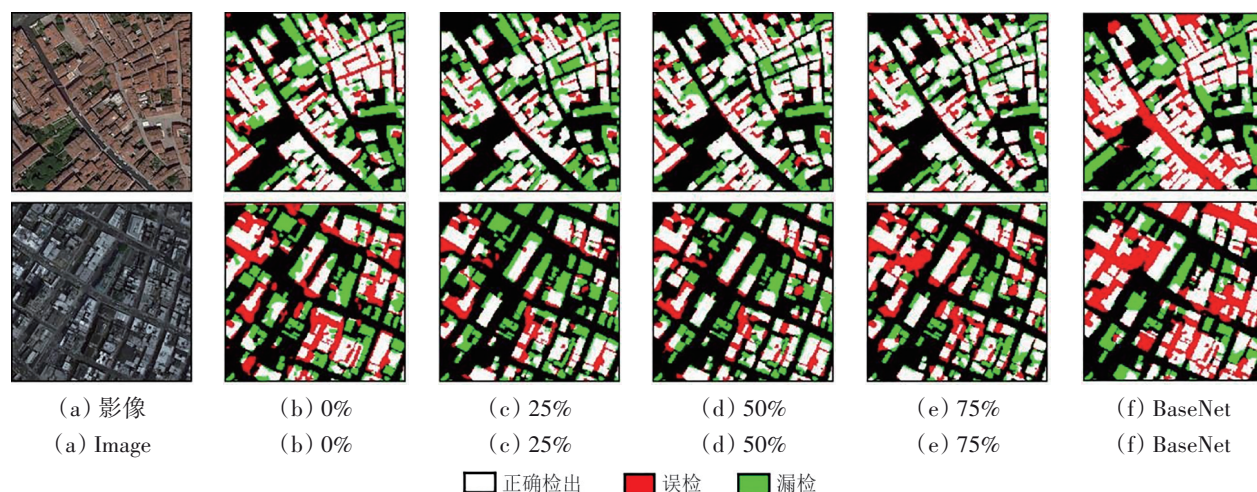


图 15 融合不同比例位置先验信息的建筑物提取结果可视化

Fig.15 Visualization of building extraction with different proportional position prior information

由表 5 和图 15 可见，即使训练中源域融入位置先验信息的比例固定，随着目标域添加位置先验信息的比例提高，F1 指标也分别提升了 1.64%、4.27%、6.41%、8.93%，其中 Recall 指标提升得格外明显，图 13 可视化结果中也可以看出白色区域明显增多。并且，融入 0% 位置先验信息的预测结果依然比 BaseNet 的 F1 指标高了 1.68%，图 13 中可见 BaseNet 的误检区域更多，道路被误认为建筑物的可能性更高。由此可以说明，先验信息对建筑物域适应提取过程是有意义的，而目标域没有任何位置先验信息时，仅源域融入部分位置先验信息也不会产生负面结果。

## 5 结 论

为解决遥感影像建筑物跨域提取中因数据分布差异导致的模型性能下降问题，本研究采用挖掘建筑物几何不变特征与引入众源数据相结合的技术路线，提出了一种形状先验与位置先验信息协同引导的建筑物域适应提取方法。主要结论如下：

(1) 通过角点与边缘信息构建的形状先验，能有效为源域和目标域提供一致的几何结构引导，显著增强了模型对建筑物本质特征的识别与保持能力，克服了外观差异的干扰；

(2) 高斯转换模块成功融合了目标域 OSM 数据的位置先验信息，并有效弱化了众源数据在标签质量与覆盖范围上的不确定性，提升了模型在真实复杂数据条件下的适应性；

(3) 基于对抗学习的域适应框架在形状与位置双先验的协同引导下，显著缩小了域间特征分布差异，最终在目标域 OSM 数据存在一定质量问题的情况下，实现了与需要高质量标签的半监督方法相近的精度。同时，该方法组件灵活、易于迁移，展现出良好的应用潜力。

## 参考文献 (References)

- Basak H and Yin Z Z. 2023. Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning//Proceedings of 26th International Conference on Medical Image Computing and Computer Aided Intervention - MICCAI 2023. Vancouver: Springer: 260-270 [DOI: 10.1007/978-3-031-43901-8\_25]
- Canny J. 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(6): 679-698 [DOI: 10.1109/TPAMI.1986.4767851]
- Cao Y X and Huang X. 2023. A full-level fused cross-task transfer learning method for building change detection using noise-robust pretrained networks on crowdsourced labels. Remote Sensing of Environment, 284: 113371 [DOI: 10.1016/j.rse.2022.113371]
- Chen J W, Li Y, Chen J and Zhang H Q. 2024. Building extraction method based on MFF-Deeplabv3+ network for high-resolution remote sensing images. Journal of University of Chinese Academy of Sciences, 41(5): 654-664 (陈经纬, 李宇, 陈俊, 张洪群. 2024. 基于 MFF-Deeplabv3+ 网络的高分辨率遥感影像建筑物提取方法. 中国科学院大学学报(中英文), 41(5): 654-664) [DOI: 10.7523/j.ucas.2023.010]
- Cote M and Saeedi P. 2013. Automatic rooftop extraction in nadir aerial imagery of suburban regions using corners and variational level set evolution. IEEE Transactions on Geoscience and Remote Sensing, 51(1): 313-328 [DOI: 10.1109/TGRS.2012.2200689]

- Du P J. 2020. Progress of high resolution remotely sensed image progressing and urban application examples. *Modern Surveying and Mapping*, 43(1): 1-9 (杜培军. 2020. 高分辨率遥感影像处理进展与城市应用若干实例. *现代测绘*, 43(1): 1-9) [DOI: 10.3969/j.issn.1672-4097.2020.01.001]
- Fan C N, Liu P, Xiao T, Zhao W and Tang X L. 2021. A review of deep domain adaptation: general situation and complex situation. *Acta Automatica Sinica*, 47(3): 515-548 (范苍宁, 刘鹏, 肖婷, 赵巍, 唐降龙. 2021. 深度域适应综述: 一般情况与复杂情况. *自动化学报*, 47(3): 515-548) [DOI: 10.16383/j.aas.c200238]
- Gerke M. 2014. Use of the stair vision library within the ISPRS 2D semantic labeling benchmark (Vaihingen)[EB/OL]. <https://research.utwente.nl/en/publications/use-of-the-stair-vision-library-within-the-isprs-2d-semantic-labe>. [DOI: 10.13140/2.1.5015.9683]
- Harris C and Stephens M. 1988. A combined corner and edge detector// *Proceedings of the 4th Alvey Vision Conference*. Manchester: 147-151
- Huang X and Zhang L P. 2011. A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery. *Photogrammetric Engineering and Remote Sensing*, 77(7): 721-732 [DOI: 10.14358/PERS.77.7.721]
- Ji S P, Wei S Q and Lu M. 2019. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1): 574-586 [DOI: 10.1109/TGRS.2018.2858817]
- Kaiser P, Wegner J D, Lucchi A, Jaggi M, Hofmann T and Schindler K. 2017. Learning aerial image segmentation from online maps. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11): 6054-6068 [DOI: 10.1109/TGRS.2017.2719738]
- Kang L, Wang Q and Yan H W. 2018. Building extraction based on OpenStreetMap tags and very high spatial resolution image in urban area. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3: 715-718 [DOI: 10.5194/isprs-archives-XLII-3-715-2018]
- Li D R. 2018. Brain cognition and spatial cognition: on integration of geo-spatial big data and artificial intelligence. *Geomatics and Information Science of Wuhan University*, 43(12): 1761-1767 (李德仁. 2018. 脑认知与空间认知——论空间大数据与人工智能的集成. *武汉大学学报(信息科学版)*, 43(12): 1761-1767) [DOI: 10.13203/j.whugis.20180411]
- Li Z, Sui Z W, Fu Q Y, Zheng J J and Bu T. 2023. High-resolution remote sensing extraction of urban buildings based on morphological sequences and multi-source a priori information. *National Remote Sensing Bulletin*, 27(4): 998-1008 (李治, 隋正伟, 傅俏燕, 郑珊珊, 卜桐. 2023. 基于形态学序列和多源先验信息的城市建筑物高分遥感提取. *遥感学报*, 27(4): 998-1008) [DOI: 10.11834/jrs.20221077]
- Liu H B, Qu Y J, Yan Q S and Deng F. 2023. The method of variable scale mesh primitive extraction for monomeric and structural reconstruction of buildings. *Acta Geodaetica et Cartographica Sinica*, 52(11): 1953-1961 (刘海兵, 曲英杰, 颜青松, 邓非. 2023. 建筑物单体结构化重建的变尺度网格基元提取法. *测绘学报*, 52(11): 1953-1961) [DOI: 10.11947/j.AGCS.20220275]
- Liu Y, Zhang Y J, Ao Y, Jiang D L and Zhang Z R. 2024. Building extraction from remote sensing images via the multiscale information fusion method under the Transformer architecture. *National Remote Sensing Bulletin*, 28(12): 3173-3183 (刘昇, 张寅捷, 敖洋, 江大龙, 张肇睿. 2024. Transformer架构下跨多尺度信息融合的遥感影像建筑提取. *遥感学报*, 28(12): 3173-3183) [DOI: 10.11834/jrs.20233017]
- Long L H, Zhu Y T, Yan J W, Liu J J and Wang Z Y. 2023. New building extraction method based on semantic segmentation. *National Remote Sensing Bulletin*, 27(11): 2593-2602 (龙丽红, 朱宇霆, 闫敬文, 刘敬瑾, 王宗跃. 2023. 新型语义分割D-UNet的建筑物提取. *遥感学报*, 27(11): 2593-2602) [DOI: 10.11834/jrs.20211029]
- Maggiori E, Tarabalka Y, Charpiat G and Alliez P. 2017. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark//2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth: IEEE: 3226-3229 [DOI: 10.1109/IGARSS.2017.8127684]
- Peng D F, Guan H Y, Zang Y F and Bruzzone L. 2022. Full-level domain adaptation for building extraction in very-high-resolution optical remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-17 [DOI: 10.1109/TGRS.2021.3093004]
- Shi L K, Wang Z Y, Pan B and Shi Z W. 2021. An end-to-end network for remote sensing imagery semantic segmentation via joint pixel- and representation-level domain adaptation. *IEEE Geoscience and Remote Sensing Letters*, 18(11): 1896-1900 [DOI: 10.1109/LGRS.2020.3010591]
- Sun J Y, Huang Z J, Zhou S G, Xu N, Qian H M and Wang C L. 2017. Building outline vectorization from high spatial resolution imagery. *National Remote Sensing Bulletin*, 21(3): 396-405 (孙金彦, 黄祚继, 周绍光, 徐南, 钱海明, 王春林. 2017. 高分辨率遥感影像中建筑物轮廓信息矢量化. *遥感学报*, 21(3): 396-405) [DOI: 10.11834/jrs.20176127]
- Tasar O, Giros A, Tarabalka Y, Alliez P and Clerc S. 2021. DAUGNet: unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2): 1067-1081 [DOI: 10.1109/TGRS.2020.3006161]
- Tsai Y H, Hung W C, Schuster S, Sohn K, Yang M H and Chandraker M. 2018. Learning to adapt structured output space for semantic segmentation//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE: 7472-7481 [DOI: 10.1109/CVPR.2018.00780]
- Viana C M, Encalada L and Rocha J. 2019. The value of OpenStreetMap historical contributions as a source of sampling data for multi-temporal land use/cover maps. *ISPRS International Journal of Geo-Information*, 8(3): 116 [DOI: 10.3390/ijgi8030116]
- Von Gioi R G, Jakubowicz J, Morel J M and Randall G. 2010. LSD: a fast line segment detector with a false detection control. *IEEE transactions on pattern analysis and machine intelligence*, 32(4): 722-732 [DOI: 10.1109/TPAMI.2008.300]
- Wang J, Yang X C, Qin X B, Ye X and Qin Q M. 2015. An efficient

- approach for automatic rectangular building extraction from very high resolution optical satellite imagery. *IEEE Geoscience and Remote Sensing Letters*, 12(3): 487-491 [DOI: 10.1109/LGRS.2014.2347332]
- Wei D S and Zhou X G. 2019. Automatic sampling of remote sensing image change detection samples based on prior information of vector data. *National Remote Sensing Bulletin*, 23(3): 464-475 (魏东升, 周晓光. 2019. 遥感影像变化检测样本自动抽样. *遥感学报*, 23(3): 464-475 [DOI: 10.11834/jrs.20198013])
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//*Proceedings of 15th European Conference on Computer Vision - ECCV 2018*. Munich: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2\_1]
- Yan W H, Qian Y Q, Li Y Y, Wang C X and Yang M. 2024. SS-ADA: a semi-supervised active domain adaptation framework for semantic segmentation. *arXiv preprint arXiv: 2407.12788* [DOI: 10.48550/arXiv.2407.12788]
- YanX F, YuanT, YangM, KongB and LiuP C. 2022. An adaptive building simplification approach based on shape analysis and representation. *Acta Geodaetica et Cartographica Sinica*, 51(2): 269-278 (晏雄锋, 袁拓, 杨敏, 孔博, 刘鹏程. 2022. 建筑物形状特征分析表达与自适应简化方法. *测绘学报*, 51(2): 269-278) [DOI: 10.11947/j.AGCS.2022.20210302]
- Yang H L, Makkar N, Laverdiere M and Rose A. 2024. Boundary-aware adversarial learning domain adaption and active learning for cross-sensor building extraction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 9122-9138 [DOI: 10.1109/JSTARS.2024.3385380]
- Yang M W, Zhao L K, Ye L F, Jiang H W and Yang Z. 2024. A review of convolutional neural networks related methods for building extraction from remote sensing images. *Journal of Geo-information Science*, 26(6): 1500-1516 (杨明旺, 赵丽科, 叶林峰, 蒋华伟, 杨震. 2024. 基于卷积神经网络的遥感影像建筑物提取方法综述. *地球信息科学学报*, 26(6): 1500-1516) [DOI: 10.12082/dqxxkx.2024.240057]
- Yiming T, Tang X Y and Shang H B. 2024. A shape-aware enhancement Vision Transformer for building extraction from remote sensing imagery. *International Journal of Remote Sensing*, 45(4): 1250-1276 [DOI: 10.1080/01431161.2024.2307325]
- Yin J C, Wu F, Zhai R J, Qiu Y, Gong X Y and Hang R X. 2024. Two-stream boundary constraints and relativistic generation adversarial network for building contour regularization. *Acta Geodaetica et Cartographica Sinica*, 53(7): 1444-1457 (殷吉崇, 武芳, 翟仁健, 邱越, 巩现勇, 行瑞星. 2024. 面向建筑物轮廓规则化的双路径边界约束与相对论生成对抗网络. *测绘学报*, 53(7): 1444-1457) [DOI: 10.11947/j.AGCS.2024.20230056]
- Zhang Z H, Guo W W, Li M J and Yu W X. 2020. GIS-supervised building extraction with label noise-adaptive fully convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, 17 (12): 2135-2139 [DOI: 10.1109/LGRS.2019.2963065]
- Zheng Q, Chen J, Wang Z Y, Jiang J J and Liang C. 2019. Deep segmentation domain adaptation network with weighted boundary constraint. *IEEE Access*, 7: 93909-93918 [DOI: 10.1109/ACCESS.2019.2927599]
- Zuo X L, Shao Z F, Wang J M, Huang X and Wang Y. 2025. A cross-stage features fusion network for building extraction from remote sensing images. *Geo-Spatial Information Science*, 28(2): 387-401 [DOI: 10.1080/10095020.2024.2307922]

## Domain-adaptive building extraction method from remote sensing imagery that uses shape and position priors

ZHAO Shaoxuan<sup>1</sup>, ZHOU Xiaoguang<sup>1</sup>, HOU Dongyang<sup>1</sup>, GUO Jing<sup>2,3</sup>, LIU Huaguang<sup>2,3</sup>

1. School of Geosciences and Info-Physics, Central South University, Changsha 410083, China;

2. Hunan Engineering Research Center of 3D Real Scene Construction and Application Technology, Changsha 410000, China;

3. The First Surveying and Mapping Institute of Hunan Province, Changsha 410000, China

**Abstract:** Deep learning is an effective approach for building extraction from remote sensing images. However, when the distributions of training (source domain) and test (target domain) data differ, the accuracy of a model trained only in the source domain substantially decreases when the model is applied directly in the target domain. The domain adaptation method can overcome differences in data distribution between domains, thereby improving the recognition of building information in different environments. It increases the efficiency of urban planning, post disaster reconstruction, and land use management. However, the difficulty and cost of obtaining target domain labels in building domain adaptation extraction are high.

Current methods do not fully utilize the invariant properties of buildings to provide cross-domain consistency constraints. This study proposes a domain-adaptive building extraction method from remote sensing imagery guided by shape and position prior information. The approach has several main components. First, building corners in the target domain are automatically extracted through the comprehensive application of methods, such as the building index and Harris operator, and the edges of labeled buildings in the source domain are obtained using image morphological processing. These extracted features serve as shape priors for both domains. Second, a Gaussian transformation-based method is designed to convert the OSM objects in the target domain and the labeled objects in the source domain to position prior

information in the dual domain. Finally, a shape loss is constructed using the aforementioned shape prior information, and training constraints are provided for buildings in both domains. The dual-domain position prior information is used as independent additional channels to superimpose image layers to form a four-channel input, enriching the building information in the target domain. A building domain adaptation extraction model called AU\_adaptNet is also designed based on adversarial learning.

Experimental results show that the intersection over union metrics of the proposed method can be improved by 15% compared with that of the base generative model without domain adaptation and by 6% compared with that of the domain-adaptation model without added prior information guidance. When the target domain has no OSM data, domain adaptation accuracy can be improved by relying only on shape prior information guidance. Under conditions with low OSM data integrity in the target domain, comparable results can be achieved with semi-supervised domain adaptation methods by using high-quality target-domain labels.

The following conclusions are derived. (1) The proposed method extracts invariant features (e.g., corners and edges) from source and target domain buildings as shape prior while simultaneously incorporating crowdsourced data from the target domain as position prior, resulting in a building extraction framework guided by shape and position prior for domain adaptation. (2) When OSM data in the target domain exhibit low completeness, the proposed method's performance is comparable to that of semi-supervised domain adaptation approaches that utilize high-quality target-domain labels.

**Key words:** building extraction, domain adaptation, remote sensing imagery, prior information, OSM

**Supported by** National Key Research and Development Program (No. 2022YFB390420501); Real 3D Construction and Application Technology Open Project of Hunan Engineering Research Center (No. 3DRS2024Y4)