引文格式：李红玲,赵磊,吴查查,等. 顾及深度残差 U-Net 的遥感影像建筑物提取[J]. 地理空间信息, 2025, 23(10):5-9. doi:10. 3969/j.issn. 1672-4623.2025.10.002

顾及深度残差 U-Net 的遥感影像建筑物提取

李红玲[1]，赵　磊[1]，吴查查[1]，李晓琴[1]，王　巍[1]
（1. 国能数智科技开发（北京）有限公司，北京 100013）

摘　要：建筑物提取是遥感技术领域的一项核心研究议题。提出了深度残差 U-Net 模型，用于遥感影像中的建筑物自动提取，该模型融合了 U-Net 架构、残差学习、ASPP 多尺度分析及焦点损失函数等策略，旨在克服传统方法局限。通过改良的残差学习模块降低网络参数，缓解退化问题；利用 ASPP 捕获多尺度特征和上下文信息；引入焦点损失函数增强建筑物识别能力。在 WHU 影像数据集上的实验表明，本方法在提取精度、边界界定及多尺度信息处理上均具显著优势，为遥感影像处理领域的建筑物提取提供了坚实的理论与实践基础。

关键词：建筑物提取；卷积神经网络；U-Net 网络；残差学习；ASPP；损失函数
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2025）10-0005-05

Building Extraction from Remote Sensing Images Based on Deep Residual U-Net

LI Hongling[1], ZHAO Lei[1], WU Chacha[1], LI Xiaoqin[1], WANG Wei[1]
(1. China Energy Digital Intelligence Technology Development Co., Ltd., Beijing 100013, China)

Abstract: Building extraction is a core research topic in the field of remote sensing technology. We proposed a model named deep residual U-Net for automatic building extraction from remote sensing images. The model integrates strategies such as U-Net architecture, residual learning, ASPP multi-scale analysis, and focal loss function, aiming to overcome the limitations of traditional methods. By utilizing an improved residual learning module, the network parameters are reduced, and the degradation problem common in deep neural networks is alleviated. ASPP is employed to capture multi-scale features and contextual information. The focal loss function is introduced to enhance the building recognition capability. Experiments on the WHU image dataset demonstrate that this method exhibits significant advantages in extraction accuracy, boundary delineation, and multi-scale information processing, providing a solid theoretical and practical foundation for building extraction in the field of remote sensing image processing.

Key words: building extraction, convolutional neural network, U-Net network, residual learning, ASPP, loss function

从遥感影像中自动提取建筑物对城市规划等至关重要[1-3]。然而，传统方法依赖人工特征，泛化能力有限[4-6]。近年来，深度学习方法（如卷积神经网络 CNN、全卷积网络 FCN）通过自动学习特征[7-12]，显著提升了提取效果，但在边界精确度和多尺度适应性方面仍存在不足[13-18]。针对这些问题，本文提出了一种改进的残差 U-Net 模型。该模型引入了空洞空间金字塔池化（ASPP）模块，有效融合多尺度特征，增强了对不同尺寸建筑物的识别能力。同时，采用焦点损失函数（Focal Loss），缓解了遥感影像中建筑物与背景类别不平衡的问题，进一步提升了模型的精度与效率。

1　残差 U-Net 方法建筑物提取方法

本文设计了一种基于 U-Net 架构的建筑提取模型，所提出模型的具体架构如图 1 所示。该模型共分为 3 个部分：编码器、连接阶段和解码器单元。编码器提取低级特征，解码器提取相应的高级特征。在低级特征和相应的高级特征之间进行连接，以促进训练过程中的反向传播。编码器和解码器单元的每个块采用残差网络单元，并且在网络中融合 ASPP，进一步提高对建筑物特征的提取能力。

1.1　编码器

在网络中，编码器的目标是完成遥感影像的特征提取。由于遥感影像的分辨率已经达到亚米级，大型建筑通常会覆盖大量的切片单元。此外，背景信息在建筑影像提取过程中极为重要。为了提取更多信息，本文使用输入影像为 512×512 大小。由于较大的影像尺寸会增加训练时间。因此，本文提出了一种高效的编

项目来源：数智科技公司科技创新资助项目（YF20230004A）。
第一作者简介：李红玲（1982—），硕士研究生，高级工程师，主要从事生态环境监测与遥感应用方面的工作，E-mail：11685045@ceic.com。

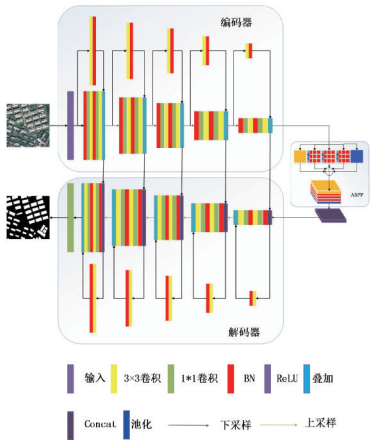


图1　本文U-Net网络

码器单元，使用残差神经网络来解决网络的退化问题，使网络能够进行更深的训练。残差神经网络由多个残差单元组成，用于完成身份映射。

$$y_l = h(x_l) + f(x_l, w_l)$$
$$x_{l+1} = f(y_l) \tag{1}$$

式中，x_l 和 x_{l+1} 分别为第 l 个残差单位函数的输入和输出；$h(x_l)$ 为残差学习的映射函数；w_l 为卷积用于调整 x_l 通道尺寸的操作；$F(,)$为残差函数，而 $f(y_l)$ 代表激活函数。结合残差学习的优势，本文设计的编码器具有促进训练和减少退化的优点。该编码器由5个残差单元构成，这些单元利用残差学习机制来加速网络的训练进程。如图2（左）所示，每个残差单元由2个卷积块和一个残差映射组合而成。具体而言，每个卷积块内部包含一个批量归一化（Batch Normalization，BN）层、一个ReLU激活层以及一个采用3×3卷积核的卷积层。与常见的特征恒等映射不同，本文中的残差单元映射包含一个3×3卷积核的卷积层和一个BN层。单元的输出 x_{l+1} 是由2个卷积块和残差映射的结果共同决定的。此外，每个单元的输出 x_{l+1} 会作为下一个单元的输入，并建立起通往解码器单元的特征通道连接。这样的设计不仅充分利用了残差学习的优势，还通过精细的卷积块和残差映射组合，进一步提升了编码器的性能，使其在处理复杂任务时能够表现出更高的效率和稳定性。

1.2　中间连接结构

在传统的U-Net模型架构中，编码器的下采样输出直接作为解码器上采样阶段的输入。为了拓宽网络的感受野并有效融入多尺度特性，本文提出了一种创新的中间连接结构，该结构精准地对接了编码器与解码器。在这一连接阶段的单元设计中，本文摒弃了传统的普通卷积，转而采用ASPP（atrous spatial pyramid pooling，ASPP）模块。ASPP模块内置了具有不同

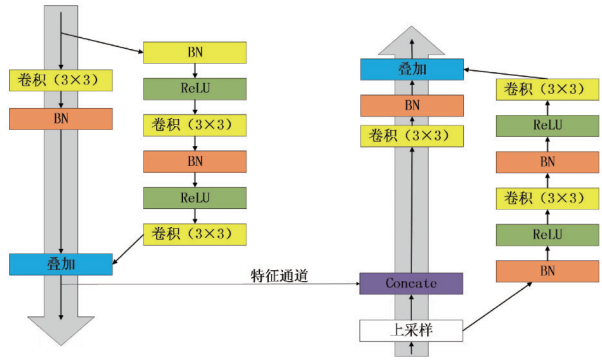


图2　残差结构

膨胀率的空洞卷积，这种设计策略在不增加模型参数规模的前提下，显著扩展了感受野的范围，使卷积层能够更有效地捕捉并整合多元化的空间特征。

此策略极大地促进了多尺度细节信息的表达，显著增强了模型对不同尺度特征的敏感度。在本文所提出的方法中，ASPP模块（图3）被专门引入，用于从高分辨率的建筑影像中精准提取多尺度特征。具体而言，本文精心设定了膨胀率为1、2、4、6的空洞卷积以及平均池化的组合，以满足在不同尺度下对象分割的精细需求。这种膨胀率的配置不仅确保了感受野能够广泛覆盖图像区域，而且有效避免了信息连续性的损失。

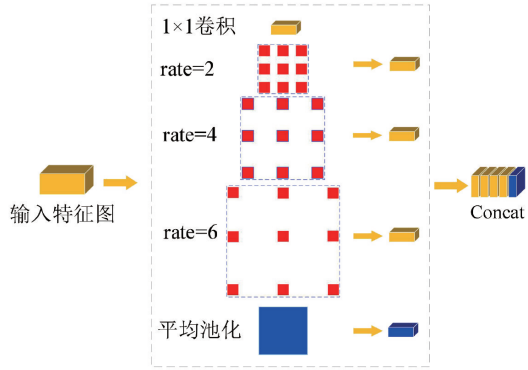


图3　ASPP模块

1.3　解码器

解码器的架构与编码器保持一致。解码器由5个单元组成，这些单元利用残差学习技术来促进网络的训练。如图2（右）所示，每个单元均包含一个上采样模块、一个拼接操作、2个卷积块以及一个残差映射。上采样操作可通过反卷积实现，拼接操作负责接收来自编码器单元的通道信息。每个卷积块均包含批量归一化（BN）层、ReLU激活层以及一个3×3卷积核的卷积层。残差映射则由一个3×3卷积核的卷积层和BN层组成。解码器的输出由2个卷积块和残差映射的输出共同构成。在模型的最终输出之前，采用了一个1×1大小的卷积操作，并使用sigmoid作为激活函数，

以构建出预期的分段多通道特征图。

本文采用了武汉大学（WHU）航空影像数据集作为实验数据。为了尽可能多地获取影像信息，输入影像的大小被设定为512×512，各个块的输出及参数详情如表1所示。

表1　残差U-Net网络结构

	模块	卷积层	卷积核	步长	输出大小
输入					512 × 512 × 3
编码层	模块1	Conv1–3	[3 × 3,16] × 3	1	512 × 512 × 16
	模块2	Conv4–6	[3 × 3,32] × 3	1	256 × 256 × 32
	模块3	Conv7–9	[3 × 3,64] × 3	1	128 × 128 × 64
	模块4	Conv10–12	[3 × 3,128] × 3	1	64 × 64 × 128
	模块5	Conv13–15	[3 × 3,256] × 3	1	32 × 32 × 256
中间连接阶段	模块6	Conv1	[3 × 3,256] × 3	1	
		Conv16	[3 × 3,256] × 1	1	
		Conv17	[3 × 3,256] × 1	6	
		Conv18	[3 × 3,256] × 1	12	
		Conv19	[3 × 3,256] × 1	18	32 × 32 × 1 024
解码层	模块7	Conv20–22	[3 × 3,256] × 3	1	32 × 32 × 256
	模块8	Conv23–25	[3 × 3,128] × 3	1	64 × 64 × 128
	模块9	Conv26–28	[3 × 3,64] × 3	1	128 × 128 × 64
	模块10	Conv29–31	[3 × 3,32] × 3	1	256 × 256 × 32
	模块11	Conv32–34	[3 × 3,16] × 3	1	512 × 512 × 16
输出		Conv35	[1 × 1,1] × 1		512 × 512 × 1

其中，编码层（模块1~5）作用是逐层提取多尺度特征，通过残差连接缓解梯度消失；中间连接阶段（模块6）作用是利用ASPP模块融合多尺度上下文信息；解码层（模块7~11）作用是通过上采样与残差学习恢复空间分辨率。

1.4　损失函数

为解决建筑物目标与背景类别间以及正样本与负样本间的不平衡问题，本文采用了焦点损失（focal loss，FL）方法作为损失函数。在本文中，该方法被用于处理建筑物与其他类别间物种分布的不平衡问题。焦点损失源自二分类的交叉熵（cross entropy，CE）损失，二分类的交叉熵损失表达如下：

$$CE(p,y)=\begin{cases}-\log(p) & y=1\\ -\log(1-p) & y=\text{其他}\end{cases} \tag{2}$$

式中，$y\in\{0,1\}$ 为真实标签；$p\in[0,1]$ 为模型预测为正例的概率值。为了简化表述，本文进行如下变换：

$$p_t=\begin{cases}p & y=1\\ 1-p & y=\text{其他}\end{cases} \tag{3}$$

二元交叉熵（Binary Cross-Entropy）的最终形式为：

$$CE(p,y)=CE(p_t)=-\log(p_t) \tag{4}$$

在众多研究中发现，即便是简单的样本也会产生相对较高的损失值，这一损失值在一定程度上掩盖了正样本在训练过程中的影响。为了有效减轻简单样本对模型训练的干扰，本文引入了均衡因子，该因子通常包含了对立类别的比重信息。为了便于表述和计算，本文类比于 p_t 定义了 α_t，并据此定义了均衡交叉熵。

$$CE(p_t)=-\alpha_t\log(p_t) \tag{5}$$

具体而言，对于困难样本，其均衡交叉熵值相对较小，但权重较大，这一设计使得网络在训练过程中更倾向于利用这些具有挑战性的样本进行参数更新。此外，这一权重是动态变化的。随着模型对复杂样本的区分能力逐渐增强，这些样本对训练过程的影响也会逐渐减弱。最终将均衡交叉熵与焦点损失（focal loss，FL）相结合，形成了改进后的损失函数，以进一步提升模型在复杂场景下的性能。

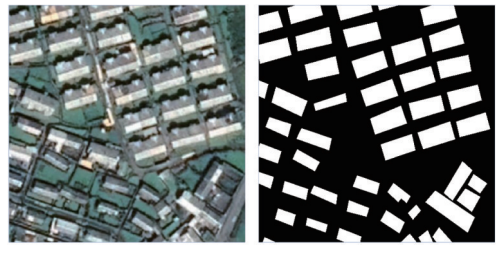
$$FL(p_t)=-\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{6}$$

2　实验与结果

2.1　数据集

WHU Building数据集是武汉大学于2019年公开发布的一项专注于建筑物识别的遥感数据集，其中仅包含建筑物这一种遥感地物类型，具有高度的专业性和针对性。该数据集规模庞大，共收录了17 388张遥感影像，为科研和实际应用提供了丰富的资源。在本次实验中，选取了其中的600张影像作为研究样本。遥感影像每张的大小均为512×512，空间分辨率为0.45 m，能够清晰地呈现出建筑物的细节特征。影像的数据源涵盖了QuickBird、Worldview series、IKONOS、ZY-3等多种高分辨率遥感卫星，数据具有多样性和广泛适用性。为了更科学地进行实验和分析，本文选取的遥感数据影像集随机划分为3个部分：训练数据集（包含480张影像）、验证数据集（同样包含60张影像）以及测试数据集（包含60张影像）。此划分有助于本文全面评估模型的性能，并对其进行有效的优化和调整。

下图展示了本文所采用的建筑物数据集中的两张代表性影像，分别为图4a和图4b。其中，图4a是一张典型的遥感影像，清晰地展示了建筑物的分布和形态；而图4b则是与图4a相对应的标签影像。



a 影像　　b 标签

图4　影像及标签

2.2 实验环境及精度评价

为了有效评估模型的分割性能，探讨不同模型在遥感影像语义分割中的优缺点，在训练参数方面，设置了epochs为300，batch size为4。使用精度评价指标包括召回率（Recall）、总体精度（overall accuracy，OA）和准确率（Precision）系数，其计算公式如下：

$$\text{Recall}=\frac{TP}{TP+FN} \tag{7}$$

$$OA=\frac{TP+TN}{TP+FN+FP+TN} \tag{8}$$

$$\text{Precision}=\frac{TP}{TP+FN} \tag{9}$$

式中，TP 为被正确分类的正样本数；FN 为被错误分类的正样本数；FP 为被错误分类的负样本数；TN 为被正确分类的负样本数；P_0 为每一类被正确分类的像素数之和除以总像素个数；P_e 为每一类真实样本个数与预测样本个数乘积的和除以总样本个数的平方；Recall为预测为建筑物的像素与真实建筑物像素的比值；OA 为预测为建筑物和背景的正确像素与所有像素的比值；Precision为建筑物被正确分为建筑物的概率。

2.3 对比实验

为了验证本文算法在处理建筑物遥感影像方面的有效性和优越性，针对影像中建筑物的分布、形状等特点选择具有代表性的测试影像进行实验。通过使用U-Net、SegNet、PSPNet、Deeplab v3+以及本文算法对测试影像进行建筑物提取，并对实验结果进行对比分析。图5中为影像的部分实验结果，其中，图a1~g1分别为原始影像、影像对应标签、U-Net建筑物提取结果、SegNet建筑物提取结果、Deeplab v3+建筑物提取结果、PSPNet建筑物提取结果以及本文算法建筑物提取结果，以便进行可视化比较分析。

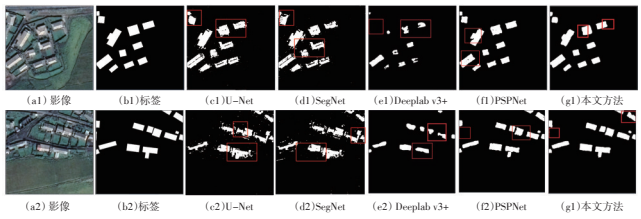


图5 对比方法实验结果

从图5（a1）可以观察到，在对应的实验结果中，无论是U-Net还是SegNet法，均存在错误识别其他地类的问题，且在边缘细节的处理上显得不够精细。相较于这2种方法，Deeplab v3+和PSPNet在提取准确性上有所提升。然而Deeplab v3+模型在描述建筑物边缘细节时表现欠佳，尤其是在小型建筑物的识别上存在遗漏现象。SegNet模型虽然能处理部分边缘细节，但在建筑物内部提取时会出现空洞，导致建筑物提取不完整，且小型建筑物同样未能被准确识别。相比之下，本文模型在遥感影像建筑物边缘细节的处理上表现出色，识别准确度高，几乎能够识别到每一个建筑物，包括那些非常小的建筑物，且没有错误地提取其他地物。总体上，本文模型的精度非常高。与其他方法相比，本文模型在提取效果上更为优越。

从图5（a2）可以观察到，在对应的实验结果中，本文可以进一步确认U-Net和SegNet方法，在建筑物信息的准确识别以及边缘细节的处理上仍存在不足。PSPNet模型在边缘处理上与本文模型有相似之处，但在建筑物内部提取时仍会出现空洞现象。不过，与第一组相比，空洞数量已经有所减少。然而，PSPNet模型在小型建筑物的提取上仍然不完整，存在像素丢失的情况，小型建筑物甚至未被识别到，因此其精度相对较低。与本文模型的实验结果相比，PSPNet模型的结果仍然存在一定的差距。

表2 各方法精度评价

方法	U-Net	SegNet	Deeplab v3+	PSPNet	本文方法
总体精度/%	0.812 2	0.832 4	0.853 2	0.912 6	0.957 6
召回率/%	0.865 2	0.876 5	0.925 5	0.952 4	0.987 9
精确度/%	0.915 4	0.913 4	0.924 6	0.931 6	0.988 0
参数量/M	7.8	29.4	41.2	45.6	31.5
推理时间/ms	28.6	34.1	42.3	48.7	36.2

根据表2所示的数据可以看出，在总体精度方面，U-Net和SegNet模型的表现均未达到90%的阈值，显示出其在处理遥感影像建筑物提取任务时存在一定的局限性。具体而言，U-Net模型的总体精度为0.812 2，虽然其精确度较高（0.915 4），但整体性能仍有待提升。SegNet模型则在精度上有所提升（0.832 4），精确度也维持在相近水平（0.913 4），表明其在影像分割任务上具有一定的潜力，但仍有改进空间。相较于U-Net和SegNet，Deeplab v3+和PSPNet模型在总体精度上实现了显著提升。Deeplab v3+模型的总体精度为0.853 2，而PSPNet模型更是达到了0.912 6的高水平。这2个模型在召回率和精确度方面也表现出色，进一步证明了它们在处理复杂影像分割任务时的有效性。然而，本文模型在遥感影像建筑物提取任务中展现出了卓越的性能。其总体精度高达0.957 6，远远超过了其他对比模型。在召回率和精确度方面，本文模型同样以显著优势位居榜首（召回率：0.987 9，精确度：0.988 0）。这一结果表明，本文模型在细节处理、地物识别以及整体精度上均表现出色，具有较高的实际应用价值。在参数量上，本文

模型较U-Net增加15%，但显著低于Deeplab v3+和PSPNet，主要得益于残差模块的参数共享机制。综上所述，提出的模型在遥感影像建筑物提取任务上表现出色，具有较高的实际应用价值。其卓越的性能不仅体现在总体精度、召回率和精确度等关键指标上，还体现在对细节处理的准确性和对地物识别的准确性上。在参数量上，通过残差学习与ASPP模块的协同优化，在保持高精度的同时，实现了计算资源的高效利用。其参数量与推理速度均优于主流语义分割模型，这一成果为遥感影像建筑物提取任务提供了新的思路和方法。

2.4　消融实验

为了验证本章在构建本文方法时，引用残差学习、ASPP模块和焦点损失函数的有效性和必要性，在本章数据集上进行了消融实验。在实验中，对比所提算法以及其他2种变体方法的表现。在本文中，方案一是以原始U-Net网络作为基础网络，引用残差学习；方案二是在方案一的基础上引用ASPP模块。本文方法为方案二的基础上引用焦点损失函数。在图6（a）~（e）中展示了原始影像、影像对应的标签、方案一提取结果、方案二提取结果以及本文算法提取结果。

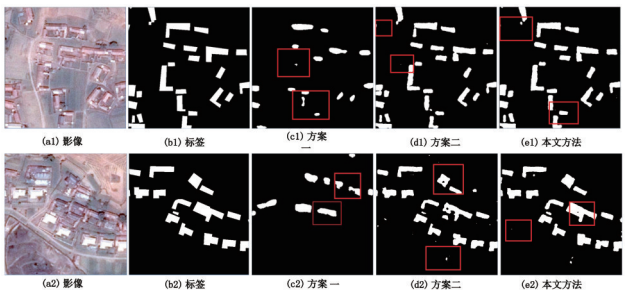


图6　消融实验结果

通过观察图6中的实验结果图，可以发现，方案一在提取结果上存在边缘处理效果不理想的情况，而在方案二中，这种情况得到了改善，这得益于ASPP的不同尺度的识别能力，然而本文方法达到了最佳效果，在建筑物提取任务中展现出了卓越的性能，无论是在细节处理、轮廓提取还是整体精度上，都远远超过了其他方法，说明了本文方法改进部分的有效性和必要性。

表3　消融实验精度对比

指标	方案一	方案二	本文方法
OA	0.901 2	0.931 5	0.957 6
Recall	0.902 5	0.951 5	0.987 9
Precision	0.922 2	0.953 0	0.988 0

根据表3可以看出，在总体精度上方案一的总体精度为0.901 2，是三者中最低的。方案二的总体精度提升至0.931 5，显示出在方案一的基础上引用ASPP模块时建筑物提取性能得到了提升。本文方法的总体精度最高，达到了0.957 6，比方案一和方案二都有显著的提升。这表明本文提出的方法经过改进后，在影像建筑物提取时具有更好的整体性能。从召回率方面看，方案一的召回率为0.902 5，方案二的召回率提升至0.951 5，相比方案一有了明显的提升，本文方法的召回率最高，达到了0.987 9。从精确度上来看，方案一的精确度为0.922 2，虽然相对较高，但仍有提升的空间。方案二的精确度提升至0.953 0，比方案一有了明显的提升。然而，本文方法的精确度同样达到了最高水平，为0.988 0，比方案二还有所提升。这表明本文方法的改进部分的有效性。

综上所述，本文改进部分在影像建筑物提取上表现出色，无论是总体精度、召回率还是精确度都优于方案一和方案二，这表明本文方法改进部分具有有效性和必要性。

3　结　语

本文提出一种基于U-Net的改进模型，通过引入残差单元缓解网络退化，采用ASPP模块融合多尺度特征，并利用焦点损失函数解决类别不平衡问题。实验表明，该模型在建筑物提取精度、召回率和细节处理上均优于主流模型，具有较高的应用价值。未来将探索Transformer架构以进一步提升性能。

参考文献

[1]　赵华，窦虎，张华. 多尺度特征增强CNN的遥感影像建筑物提取[J]. 遥感信息，2024，39(6):38-47

[2]　高青. 卷积神经网络在遥感影像城市建筑物提取中的应用[J]. 经纬天地，2024，38(5):57-60

[3]　陶从辉，高青山，赵梦琳. 改进U-Net网络的高分辨率遥感影像建筑物提取方法[J]. 地理空间信息，2025，23(1):9-12

[4]　郭英琦，张翔，李若兰，等. 顾及地理尺度的矢量建筑物变化检测方法[J]. 地理空间信息，2024，22(7):44-49

[5]　刘卓涛，龚循强，夏元平，等. KU-Net:改进U-Net的高分辨率遥感影像建筑物提取方法[J]. 遥感信息，2024，39(5):121-131

[6]　Dunaeva A V，Kornilov F A. Specific Shape Building Detection from Aerial Imagery in Infrared Range[J]. Ural Federal University Named after the First President of Russia B. N. Yeltsin (Mira Str)，2017，19(8):84-100

[7]　王顺，陈小月，林中杰，等. 运用U-Net++模型的遥感影像

（下转第19页）