



(12) 发明专利

(10) 授权公告号 CN 114529552 B

(45) 授权公告日 2025. 05. 06

(21) 申请号 202210202403.9

G06V 10/774 (2022.01)

(22) 申请日 2022.03.03

G06V 10/80 (2022.01)

(65) 同一申请的已公布的文献号

G06V 10/82 (2022.01)

申请公布号 CN 114529552 A

G06V 10/40 (2022.01)

G06V 10/764 (2022.01)

(43) 申请公布日 2022.05.24

(56) 对比文件

(73) 专利权人 北京航空航天大学

CN 112149585 A, 2020.12.29

地址 100191 北京市海淀区学院路37号

专利权人 中国科学院空间应用工程与技术中心

审查员 王从雷

(72) 发明人 刘庆杰 张明明 温奇 王蕴红

(74) 专利代理机构 北京天汇航智知识产权代理有限公司 (普通合伙) 11987

专利代理师 黄川 史继颖

(51) Int. Cl.

G06T 7/10 (2017.01)

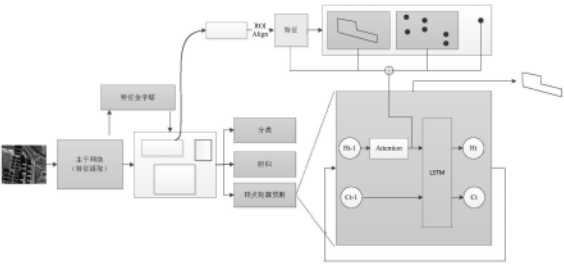
权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种基于几何轮廓顶点预测的遥感影像建筑物分割方法

(57) 摘要

本发明公开了一种基于几何轮廓顶点预测的遥感影像建筑物分割方法,本方法首先使用两阶段检测网络结构,通过该网络从遥感影像中提取建筑物的特征信息,利用ROI-Align操作对建筑物检测框进行操作得到建筑物的特征图,然后在该特征图上使用基于注意力机制的长短时记忆网络对建筑物轮廓进行建模预测,在轮廓预测过程中,先做边缘检测和关键点检测,将关键点得分最高的顶点作为开始顶点,输入到基于注意力机制的长短时记忆网络中开始预测,直到碰到结束标志或达到最大时序长度时结束预测,最后将预测的顶点结果映射到原图,并按照顺序相连得到建筑物的轮廓。



1.一种基于几何轮廓顶点预测的遥感影像建筑物分割方法,其特征在于,包括以下步骤:

S1:读取地区的遥感影像图,将所有数据预处理切分到 $512 \times 512$ 像素大小,获取训练集和测试集;

S2:构建网络模型,包括建筑物检测模块和轮廓预测模块;

S3:对网络模型进行训练,得到建筑物分割模型;

S4:将未分割的遥感图像输入到已训练完成的建筑物分割模型中,得到结果;

所述步骤S2中建筑物检测模块为两阶段检测模型,该模型

利用基于transformer的特征提取器提取特征,获取特征提取器每层提取到的特征得到特征金字塔;

使用区域生成网络提取候选框;

RCNN头部对候选框进行分类和回归,得到建筑物检测框;

所述步骤S2中轮廓预测模块执行以下步骤:

S2-1:将特征金字塔的多层特征进行融合得到更丰富的特征P,然后在该特征P上对检测框进行ROI-Align操作得到检测框对应的特征R,并将特征R映射到 $28 \times 28$ 大小,获取更高分辨率的特征;

S2-2:对检测框做边缘检测和关键点检测,将预测概率最大的关键点作为起始点,输入到轮廓预测模块中开始进行轮廓预测;

所述步骤S2-1中特征融合具体为:

特征金字塔网络产生的多层特征定义为 $P = \{P_2, P_3, P_4, P_5\}$ ,以特征 $P_3$ 为基础,将其他层特征信息通过池化操作或线性插值调整到与 $P_3$ 同等尺度,然后将特征相加并平均后得到更丰富的特征信息层 $P_3$ ,利用建筑物检测模块提取到的检测框在更丰富的特征信息层 $P_3$ 进行ROI-Align操作获得检测框对应的特征,将提取的特征映射到 $28 \times 28$ 大小,获得高分辨率特征 $R^{N \times 256 \times 28 \times 28}$ ;

所述步骤S2-2中起始点通过以下方式确定:

使用两个 $3 \times 3$ 卷积提取步骤S2-1获得的特征图特征,利用组归一化帮助训练,通过两个 $1 \times 1$ 卷积,分别获得建筑物的边界掩码E和建筑物顶点预测V,V经过sigmoid激活后,得到每一个建筑物顶点的概率,选择最大概率的点作为起始点S;

所述步骤S2-2中轮廓预测具体为:

当前时序的顶点位置 $y_t$ 的预测建模公式如下:

$$P(y_t | y_{t-1}, y_{t-2}, y_0) \quad (1)$$

其中, $y_0$ 为起始点位置, $y_{t-1}$ 为上一时刻预测位置, $y_{t-2}$ 为上上次时刻预测位置;

在步骤S2-1得到的特征图上拼接顶点的位置信息,包括x坐标信息和y坐标信息的两层特征图,其取值范围为 $(-1, 1)$ ,根据特征图尺度均匀取值,然后进行两个 $3 \times 3$ 卷积提取特征F;

对E和V使用sigmoid函数激活,获得参考预测信息;

拼接E,V,F和三个预测结果,使用两个维度为64的基于注意力机制的长短时记忆网络提取时空特征;

使用两个全连接网络分别预测类别和偏置,当网络感知到需要再次预测第一个位置

时,输出结束标志完成预测;

在训练过程中,对每次预测结果替换为真实结果,测试时使用预测结果,将预测结果映射为二值图;对于第一个时序,使用全为0的特征图代替;

模型的总体损失为:

$$L=L_{\text{edge}}+\lambda_{\text{vertex}}*L_{\text{vertex}}+L_{\text{offset}}+L_{\text{rnn}}+L_{\text{det}} \quad (2)$$

其中, $L_{\text{edge}}$ 表示建筑物边缘预测损失, $L_{\text{vertex}}$ 表示建筑物顶点预测损失, $L_{\text{offset}}$ 表示时序顶点预测中的偏置回归损失, $L_{\text{rnn}}$ 表示时序顶点预测中的分类损失, $L_{\text{det}}$ 表示建筑物检测损失,建筑物顶点预测损失权重 $\lambda_{\text{vertex}}=10$ ,利用总体损失监督训练建筑物分割模型;

建筑物顶点预测过程,为保证正负训练样本平衡,避免训练过程中正样本附近存在的拉扯现象,利用2D高斯分布建模衰减系数,得到公式如(3)所示,其中offsetx,offsety表示样本相对ground truth的偏移量,

$$\theta = e^{-\frac{\text{offsetx}^2 + \text{offsety}^2}{2}} \quad (3)$$

利用sigmoid激活函数,表示一个特征点是建筑物顶点的概率,利用衰减系数和Focal loss损失计算顶点预测损失 $L_{\text{vertex}}$ ,公式如(4)所示,

$$L_{\text{vertex}} = \begin{cases} (1-p)^{\gamma} \log(p), & \theta=1 \\ (1-\theta)^{\beta} p^{\gamma} \log(1-p), & \text{otherwise} \end{cases} \quad (4)$$

其中, $\gamma$ 为聚焦参数, $\beta$ 为权重因子, $p$ 表示预测样本属于1的概率, $p$ 的取值为0-1, $\theta$ 为衰减系数。

2.根据权利要求1所述的基于几何轮廓顶点预测的遥感影像建筑物分割方法,其特征在于,时序顶点预测使用基于注意力机制的长短时记忆网络依次预测顶点位置,将顶点量化后得到坐标(x,y),假设特征图尺度为 $w \times h$ ,按照如下公式(5)得到类别,

$$c = x + y \times w \quad (5)$$

使用多类交叉熵损失进行监督,公式如下:

$$Loss = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_c^{(n)} \log(p_c^{(n)})$$

其中, $N$ 表示样本数量, $C$ 表示类别数目, $y_c^{(n)}$ 符号函数,如果样本 $n$ 的标签为 $c$ 取1,否则取0, $p_c^{(n)}$ 表示样本 $n$ 属于类别 $c$ 的概率。

3.根据权利要求2所述的基于几何轮廓顶点预测的遥感影像建筑物分割方法,其特征在于,在时序顶点预测损失中添加一个偏置回归损失,损失目标如公式(6)所示,

$$\begin{aligned} o_{x'} &= x' - \lfloor x' \rfloor \\ o_{y'} &= y' - \lfloor y' \rfloor \end{aligned} \quad (6)$$

$o_{x'}$ 和 $o_{y'}$ 为学习的偏移参数, $x'$ 和 $y'$ 为样本点映射到特征图的浮点位置,目标区间在(0,1),通过Smooth L1损失进行监督学习。

## 一种基于几何轮廓顶点预测的遥感影像建筑物分割方法

### 技术领域

[0001] 本发明属于图像处理技术领域,尤其是一种基于序列关系的建筑物轮廓建模方法。

### 背景技术

[0002] 建筑物分割任务是指通过对遥感图像进行处理,对图像进行逐像素分类,分割出每个建筑物并获得建筑物的轮廓信息。作为计算机视觉领域的一个热门问题,建筑物分割是遥感图像解译的重要研究方向,提取的建筑物信息是基础地理数据获取和更新的重要数据支撑;而且建筑物分割模型在城市规划、地图更新、灾害评估等领域都具有广泛的应用。

[0003] 现有的建筑物分割方法通常是采用先分割再提取轮廓两阶段的方法,首先通过基于深度学习的图像分割方法获得建筑物的像素级别分割结果,得到每个建筑物的边框信息,然后对轮廓进行拟合或者建模得到精细的建筑物轮廓信息。尽管基于深度学习的方法可以获得极好的语义信息和边缘信息,但由于遥感影像分辨率相较于自然场景更低,且分割掩码提取到的建筑物轮廓复杂,不够平滑,不能表现建筑物的几何特性,不能直接用于下游应用,也不符合人的直接理解。建筑轮廓拟合算法则需要先验性设计的手工算法,流程复杂、计算时间长。轮廓拟合算法与之前的分割算法分阶段进行,不能端到端训练,且拟合结果严重依赖前面的分割结果。所以,建筑物轮廓拟合算法提取的建筑物轮廓不够稳定且通用性不强。建筑物轮廓建模算法需要事先采样固定数量的点,由于建筑物的角点数不一样,很容易出现多余点问题,使得边界线多余,从而不能得到建筑物最符合人理解的描述形式。上述这些问题导致建筑物分割模型提取的建筑物轮廓精度不高,有待进一步提高。

### 发明内容

[0004] 为了解决上述已有技术存在的不足,本发明提出一种基于几何轮廓顶点预测的遥感影像建筑物分割方法,可以解决建筑物分割任务中分割模型预测的建筑物边界模糊或不平滑、预测多余的建筑物点和边,而导致建筑物分割精度指标低的问题,本发明的具体技术方案如下:

[0005] 一种基于几何轮廓顶点预测的遥感影像建筑物分割方法,其特征在于,包括以下步骤:

[0006] S1:读取地区的遥感影像图,将所有数据预处理切分到 $512 \times 512$ 像素大小,获取训练集和测试集;

[0007] S2:构建网络模型,包括建筑物检测模块和轮廓预测模块;

[0008] S3:对网络模型进行训练,得到建筑物分割模型;

[0009] S4:将未分割的遥感图像输入到已训练完成的建筑物分割模型中,得到结果。

[0010] 进一步的,步骤S2中建筑物检测模块使用两阶段检测模型,利用基于transformer的特征提取器提取特征,获取提取器每层提取到的特征得到特征金字塔;使用区域生成网络提取候选框;RCNN头部对候选框进行分类和回归,得到建筑物检测框。

[0011] 进一步的,步骤S2中轮廓预测模块包括以下步骤:

[0012] S2-1:将特征金字塔的多层特征进行融合得到更丰富的特征P,然后在该特征P上对检测框进行ROI-Align操作得到检测框对应的特征R,并将特征R映射到 $28 \times 28$ 大小,获取更高分辨率的特征;

[0013] S2-2:对检测框做边缘检测和关键点检测,将预测概率最大的关键点作为起始点,输入到轮廓预测模块中开始进行轮廓预测。

[0014] 步骤S2-1中特征融合具体为:

[0015] 特征金字塔网络产生的多层特征定义为 $P = \{P_2, P_3, P_4, P_5\}$ ,以特征 $P_3$ 为基础,将其其他层特征信息通过池化操作或线性插值调整到与 $P_3$ 同等尺度,然后将特征相加并平均后得到更丰富的特征信息层 $P_3$ ,利用建筑物检测模块提取到的检测框在特征 $P_3$ 进行ROI-Align操作获得检测框对应的特征,将提取的特征映射到 $28 \times 28$ 大小,获得高分辨率特征 $R^N \times 256 \times 28 \times 28$ 。

[0016] 进一步的,步骤S2-2中起始点预测具体为:

[0017] 使用两个 $3 \times 3$ 卷积提取S2-1获得的特征图特征,利用组归一化帮助训练,通过两个 $1 \times 1$ 卷积,分别获得建筑物的边界掩码E和建筑物的顶点预测V,V经过sigmoid激活后,得到每一个建筑物点的概率,选择最大概率的点作为起始点S。

[0018] 进一步的,步骤S2-2中轮廓预测具体为:

[0019] 当前时序的顶点位置 $y_t$ 的预测建模:

[0020]  $P(y_t | y_{t-1}, y_{t-2}, y_0)$  (1)

[0021] 其中, $y_0$ 为起始点位置, $y_{t-1}$ 为上一时刻预测位置, $y_{t-2}$ 为上上次时刻预测位置。

[0022] 在S2-1得到的特征图上拼接顶点的位置信息,包括x坐标信息和y坐标信息的两层特征图,其取值范围为 $(-1, 1)$ ,根据特征图尺度均匀取值,然后进行两个 $3 \times 3$ 卷积提取特征F;

[0023] 对E和V使用sigmoid函数激活,获得参考预测信息,拼接E,V,F和三个预测结果,使用两个维度为64的基于注意力机制的长短时记忆网络提取时空特征,使用两个全连接网络分别预测类别和偏置。当网络感知到需要再次预测第一个位置时,输出结束标志完成预测。

[0024] 在训练过程中,对每次预测结果替换为真实结果,测试时使用预测结果,将预测结果映射为二值图;对于第一个时序,使用全为0的特征图代替。

[0025] 进一步的,步骤S2中,模型的总体损失为:

[0026]  $L = L_{\text{edge}} + \lambda_{\text{vertex}} * L_{\text{vertex}} + L_{\text{offset}} + L_{\text{rnn}} + L_{\text{det}}$  (2)

[0027] 其中, $L_{\text{edge}}$ 表示建筑物边缘预测损失, $L_{\text{vertex}}$ 表示建筑物顶点预测损失, $L_{\text{offset}}$ 表示时序顶点预测中的偏置回归损失, $L_{\text{rnn}}$ 表示时序顶点预测中的分类损失, $L_{\text{det}}$ 表示建筑物检测损失,建筑物顶点预测损失权重 $\lambda_{\text{vertex}} = 10$ ,利用总体损失监督训练建筑物分割模型。

[0028] 进一步的,建筑物顶点预测损失 $L_{\text{vertex}}$ 具体为:

[0029] 建筑物顶点预测过程,利用2D高斯分布建模衰减系数,得到公式如(3)所示,其中offsetx,offsety表示样本相对ground truth的偏移量。

[0030]  $\theta = e^{-\frac{\text{offsetx}^2 + \text{offsety}^2}{2}}$  (3)

[0031] 利用sigmoid激活,表示一个特征点是建筑物顶点的概率,利用衰减系数和Focal

loss损失计算顶点预测损失 $L_{\text{vertex}}$ ,公式如(4)所示,

$$[0032] \quad L_{\text{vertex}} = \begin{cases} (1-p)^{\gamma} \log(p), & \theta = 1 \\ (1-\theta)^{\beta} p^{\gamma} \log(1-p), & \text{otherwise} \end{cases} \quad (4)$$

[0033] 其中, $\gamma$ 为聚焦参数, $\beta$ 为权重因子, $p$ 表示预测样本属于1的概率, $p$ 的取值为0-1, $\theta$ 为衰减系数。

[0034] 进一步的,时序顶点预测使用基于注意力机制的长短时记忆网络依次预测顶点位置,将顶点量化后得到坐标 $(x,y)$ ,假设特征图尺度为 $w \times h$ ,按照如下公式(5)得到类别,

$$[0035] \quad c = x + y \times w \quad (5)$$

[0036] 使用多类交叉熵损失进行监督,添加一个权重参数,除建筑物顶点数量和一个结束标志可训练外,权重设为1,其余不参与训练,设置为0。

[0037] 在时序顶点预测损失中添加一个偏置回归损失,损失目标如公式(6)所示,

$$[0038] \quad \begin{aligned} o_x &= x - \lfloor x \rfloor \\ o_y &= y - \lfloor y \rfloor \end{aligned} \quad (6)$$

[0039]  $o_x$ 和 $o_y$ 为学习的偏移参数, $x$ 和 $y$ 为样本点映射到特征图的浮点位置,目标区间在 $(0,1)$ ,通过Smooth L1损失进行监督学习。

[0040] 本发明的有益效果:

[0041] 本方法将建筑物轮廓预测模块添加于现有的建筑物分割网络中,提高了建筑物分割的准确性,在建筑物分割任务中运用本分割模型预测的建筑物边界清楚且平滑,且不会预测出多余的建筑物点和边,建筑物分割精度(AP)指标高。

## 附图说明

[0042] 通过参考附图可更好地理解本发明。图中的构件不应视作按比例绘制,重点应放在示出本发明的原理上。

[0043] 图1基于几何轮廓顶点预测的建筑物分割模型框架图

[0044] 图2特征金字塔结构

[0045] 图3顶点预测建模——高斯分布

[0046] 图4建筑物轮廓预测流程示例

[0047] 图5建筑物轮廓预测结果

## 具体实施方式

[0048] 为了便于本领域普通技术人员理解和实施本发明,下面结合附图与具体实施方式对本发明作进一步详细描述:

[0049] 一种基于几何轮廓顶点预测的建筑物分割方法,流程为:

[0050] S1:读取地区的遥感影像图,将所有数据预处理切分到 $512 \times 512$ 像素大小,获取训练集和测试集;

[0051] S2:构建网络模型,包括建筑物检测模块和轮廓预测模块;

[0052] S3:对网络模型进行训练,得到建筑物分割模型;

[0053] S4:将未分割的遥感图像输入到已训练完成的建筑物分割模型中,得到结果。

[0054] 数据来源上,首先对影像数据进行划分训练集和测试集,防止训练集和测试集来源于同一幅影像,从而造成性能误导性上升。由于原始遥感影像尺度较大,不能一次性输入到网络中,所以将所有数据预处理切分到 $512 \times 512$ 像素大小。为尽可能保证建筑物实例的完整性,切分过程中有64像素的重叠,保留占比大于0.5的建筑物。

[0055] 建筑物分割模型是基于Pytorch框架进行训练的,在建筑物分割网络中除了主干网络部分,其他网络使用均值为0,方差为0.01的高斯分布进行随机初始化。训练过程使用了SGD优化器,初始学习率设置为0.005,动量设置为0.9,权重衰减参数为0.0005。数据增强使用了随机水平翻转,训练批次大小设置为2。模型一共训练24个迭代周期,分别在第16个和第22个迭代周期对学习率衰减10倍。图像大小设置输入尺度最小为800,最大为512进行训练,测试时使用同样的配置。

[0056] 具体的基于几何轮廓顶点预测的建筑物分割框架见图1,其中建筑物检测模块和轮廓预测模块具体为:

[0057] 建筑物检测模块,建筑物检测阶段使用两阶段检测模型,首先利用基于transformer的特征提取器提取特征,然后使用区域生成网络提取候选框,最后RCNN头部对候选框进行分类和回归,最终得到建筑物检测框。

[0058] 建筑物轮廓预测模块主要可以分为三个部分:

[0059] (1) 特征融合模块。特征金字塔网络会产生多层特征,定义为 $P = \{P_2, P_3, P_4, P_5\}$ 。底层特征 $P_2$ 含有更多的位置信息,对局部的梯度变化更加敏感;而高层特征 $P_5$ 含有更丰富的语义信息,有更大的感受野,能对整体建筑物进行建模。如果仅利用单层特征信息,可能会造成信息的缺失,对预测结果造成一定的影响。因此,以特征 $P_3$ 为基础,它是原图的1/8,能同时兼顾语义信息和位置信息,并将其他层特征信息通过池化操作或线性插值调整到与 $P_3$ 同等尺度,并将特征相加取平均,获得更丰富的特征信息层 $P_3$ 。然后利用上一阶段提取到的检测框在特征 $P_3$ 进行ROI-Align操作,与检测网络不同,点的预测需要更高的分辨率,所以将得到的特征映射到 $28 \times 28$ 大小,从而获得特征 $R^{N \times 256 \times 28 \times 28}$ 。

[0060] (2) 起始点预测。在上一阶段获得特征图使用两个 $3 \times 3$ 卷积提取特征,然后通过两个 $1 \times 1$ 卷积,分别获得建筑物的边界掩码E和建筑物的顶点预测V。V经过sigmoid激活后,可得到每一个建筑物点的概率,而后选择最大概率的点作为起始点S。

[0061] (3) 轮廓预测。首先,通过上一时刻预测位置 $y_{t-1}$ 和上上次时刻预测位置 $y_{t-2}$ 这两个预测位置,可以确定当前时序预测位置的方向。其次,当网络感知到需要再次预测第一个位置时,可以输出结束标志完成预测。因此,当前时序的顶点位置 $y_t$ 预测建模如公式(7)所示,流程示例如图4所示。

$$[0062] \quad P(y_t | y_{t-1}, y_{t-2}, y_0) \quad (7)$$

其中, $y_0$ 为起始点位置, $y_{t-1}$ 为上一时刻预测位置, $y_{t-2}$ 为上上次时刻预测位置。

[0063] 在ROI-Align得到的特征图基础上拼接位置信息,两层特征图分别表示x坐标信息和y坐标信息,取值范围为(-1,1),根据特征图尺度均匀取值,进行两个 $3 \times 3$ 卷积提取特征F,每个特征F都利用组归一化(Group Norm)帮助训练。训练过程,对每次预测结果替换为真实结果,测试时则使用预测结果,并将预测结果映射为二值图。对E和V使用sigmoid函数激活,获得参考预测信息,拼接E,V,F和三个预测结果,使用两个维度为64的基于注意力机制

的长短时记忆网络提取时空特征。对于第一个时序,由于没有 $y_{t-2}$ ,所以使用全为0的特征图代替,最后使用两个全连接网络分别预测类别和偏置。在轮廓预测阶段值得注意的是,由于在上一阶段对所有点同等训练,所以任意一个点都可以作为起始点,故首先对真值样本点数进行随机旋转,从而能应对这种情况。

[0064] 建筑物检测模块使用两阶段检测模型,建筑物检测模块的损失为建筑物检测损失,包括区域生成网络损失和检测头部分损失两部分。轮廓预测模块损失主要包含建筑物边缘预测损失、建筑物顶点预测损失和时序顶点预测损失三部分,其中时序顶点预测损失包括偏置回归损失和分类损失。利用检测损失去监督特征提取,可以获得更抽象更具代表的建筑物特征。

[0065] 建筑物边缘预测过程中,将建筑物边缘映射到特征图上,将其认为是正样本,其余作为背景,利用sigmoid激活函数,判断当前特征点是建筑物边缘的概率,所以建筑物边缘预测损失使用二分类交叉熵损失。

[0066] 建筑物顶点预测过程中,真实顶点映射到特征图尺度并量化的位置作为正样本,其余作为负样本。正样本相比于负样本仍然是非常少的,为了保证训练平衡,利用Focal loss损失,公式如(8)所示。

$$[0067] \quad L_f = \begin{cases} -\alpha(1-p)^\gamma \log(p), & \text{label} = 1 \\ -(1-\alpha)p^\gamma \log(1-p), & \text{otherwise} \end{cases} \quad (8)$$

[0068] 其中, $\gamma$ 为聚焦参数, $\alpha$ 为权重因子, $p$ 表示预测样本属于1的概率, $p$ 取值为0-1, label表示样本标签, label的取值为 $\{+1, -1\}$ 。 $\alpha=0.25, \gamma=2, p$ 为经过sigmoid激活后概率。

[0069] 通过上述损失,可以防止大量简单样本损失占据主要地位,从而关注到困难样本。除此之外,因为在真实顶点附近预测到较高得分是可接受的且合理的,所以如果惩罚所有背景样本,在训练过程中在正样本附近可能会存在拉扯现象,使得对于正样本预测困难。因此,对背景样本不进行同等惩罚,设置一个衰减系数,越靠近真实顶点位置,衰减系数越大,减少真实顶点附近预测得分高的影像。具体地,利用2D高斯分布建模衰减系数,得到公式如(3)所示,其中offsetx,offsety表示样本相对ground truth的偏移量。

$$[0070] \quad \theta = e^{-\frac{\text{offsetx}^2 + \text{offsety}^2}{2}} \quad (3)$$

[0071] 模拟一个 $10 \times 10$ 的特征图,并设置位置(5,5)顶点进行建模,将其可视化。结果如图3所示,位置(5,5)的系数最高,其余位置逐渐衰减直至为0。同样利用sigmoid激活,表示一个特征点是建筑物顶点的概率,利用衰减系数和Focal loss计算计算顶点预测损失 $L_{\text{vertex}}$ ,公式如(4)所示,

$$[0072] \quad L_{\text{vertex}} = \begin{cases} (1-p)^\gamma \log(p), & \theta = 1 \\ (1-\theta)^\beta p^\gamma \log(1-p), & \text{otherwise} \end{cases} \quad (4)$$

[0073] 其中, $\gamma$ 为聚焦参数, $\beta$ 为权重因子, $p$ 表示预测样本属于1的概率, $p$ 取值范围为0-1, $\theta$ 为衰减系数,越靠近正样本 $\theta$ 越高,从而使得尽管预测概率最大,仍然能获得一个较低的损失,从而防止正样本附近出现拉扯现象。



[0074] 使用多类交叉熵损失进行监督,如公式(8),针对建筑物顶点数量和一个结束标志进行训练,

$$[0075] \quad Loss = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_c^{(n)} \log(p_c^{(n)}) \quad (8)$$

[0076] 其中,N表示样本数量,C表示类别数目, $y_c^{(n)}$ 符号函数,如果样本n的标签为c取1,否则取0, $p_c^{(n)}$ 表示样本n属于类别c的概率。

[0077] 在时序顶点预测中,使用基于注意力机制的长短时记忆网络依次预测顶点位置,需要将顶点量化,从而得到预测类别。假设特征图尺度为 $w \times h$ ,将顶点量化后得到坐标(x,y),则按照如下公式(5)得到类别:

$$[0078] \quad c = x + y \times w \quad (5)$$

[0079] 然后使用多类交叉熵损失进行监督。训练过程需要批处理,而建筑物顶点数量是不一样的,因此,我们添加一个权重参数,除建筑物顶点数量和一个结束标志可训练外,权重设为1,其余不参与训练,设置为0。

[0080] 除上述的分类损失外,顶点预测过程中,需要量化到具体位置,但是样本位置一般是浮点数,这会对定位精度产生一定的损失。因此,在时序顶点预测损失中添加一个偏置回归损失,从而得到更精确的定位,损失目标如公式(6)所示,

$$[0081] \quad \begin{aligned} o_x &= x - \lfloor x \rfloor \\ o_y &= y - \lfloor y \rfloor \end{aligned} \quad (6)$$

[0082]  $o_x$ 和 $o_y$ 为学习的偏移参数,x和y为样本点映射到特征图的浮点位置。目标区间在(0,1),通过Smooth L1损失进行监督学习。

[0083] 综上,模型的总体的损失为如下公式(2)所示,

$$[0084] \quad L = L_{\text{edge}} + \lambda_{\text{vertex}} * L_{\text{vertex}} + L_{\text{offset}} + L_{\text{rnn}} + L_{\text{det}} \quad (2)$$

[0085] 其中, $L_{\text{edge}}$ 表示建筑物边缘预测损失, $L_{\text{vertex}}$ 表示建筑物顶点预测损失, $L_{\text{offset}}$ 表示时序顶点预测中的偏置回归损失, $L_{\text{rnn}}$ 表示时序顶点预测中的分类损失, $L_{\text{det}}$ 表示建筑物检测损失,建筑物顶点预测损失权重 $\lambda_{\text{vertex}}$ 设置10,其余为1。模型在总体损失函数L的约束下,可以实现端到端的训练。

[0086] 建筑物分割模型将轮廓结果转为掩码结果,使用MS COCO提出的ap,ap@50,ap@75,ap@small,ap@medium,ap@large评价指标作为参考。其中,ap@50反映建筑物基本被分割正确;ap@75反映的一个更高的定位标准下的效果;ap@small为小物体的精度;ap@medium为中等物体的精度;ap@large为大物体的精度;ap是ap@50到ap@95的一个加权结果,反映的是总体的性能指标。

[0087] 在建筑物轮廓预测过程中需要注意几个重要的参数。首先,由于通过ROI-Align后的特征后续会被用来提取时序特征以及预测位置信息(类别),所以ROI-Align后的特征尺度大小会影响轮廓预测的精度。其次,由于使用基于注意力机制的长短时记忆网络进行顶点预测,所以要综合考虑网络的时序长度的影响。最后,如果一个点附近预测得分较高,则它附近点应该获得较高的分数,所以要充分利用像素的位置相关信息。

[0088] 为了验证池化尺度、时序长度、位置信息、偏置回归分支、平均核操作参数对模型

性能的影响,进行了相关实验。

#### [0089] (1) 池化尺度

[0090] 池化尺度是指通过ROI-Align后的特征尺度,该特征后续会被用来提取时序特征以及预测位置信息(类别)。当池化尺度越小,量化后的精度损失越大,同样可表示的范围会变少,当进行逆变换映射到原图时,会带来更高的精度。如下表1所示,实验采用五个不同的尺度(7,14,16,28,32)。在池化尺度为7时,ap@75非常低,说明量化位置后的定位效果难以满足需求。当池化尺度逐渐变大时,在ap@75上带来了大约有20个点的提升。随着尺度越来越大,性能变饱和,甚至下降,是由于更大的池化尺度带来了更大的优化空间,使得优化困难。因此模型参数选用28作为标准池化尺度。

#### [0091] 表1不同池化尺度效果对比

|        | 池化尺度 | ap   | ap@50 | ap@75 | ap@small | ap@medium | ap@large |
|--------|------|------|-------|-------|----------|-----------|----------|
| [0092] | 7    | 19.2 | 53.4  | 4.2   | 17.3     | 23.0      | 21.3     |
|        | 14   | 27.7 | 57.7  | 23.3  | 24.0     | 34.0      | 35.3     |
|        | 16   | 28.4 | 58.2  | 24.5  | 24.8     | 34.9      | 36.9     |
| [0093] | 28   | 29.3 | 56.9  | 27.3  | 24.3     | 37.7      | 40.2     |
|        | 32   | 29.2 | 56.3  | 27.3  | 24.4     | 37.0      | 39.4     |

#### [0094] (2) 时序长度

[0095] 时序越长越能保留复杂建筑物的轮廓点数,但却会带来训练时间增长,以及对复杂建筑物的过度建模,使得损失过大,优化陷入局部最优,影响建筑物的检测精度。如下表2所示,当时序长度为5时,性能与其他时序长度性能有较大差别;其余时序长度对精度影响不大,表明已经可以很好的覆盖建筑物顶点数量。因此模型参数使用20作为标准时序长度。

#### [0096] 表2不同时序尺度效果对比

|        | 时序长度 | ap   | ap@50 | ap@75 | ap@small | ap@medium | ap@large |
|--------|------|------|-------|-------|----------|-----------|----------|
| [0097] | 5    | 28.1 | 54.7  | 25.6  | 22.5     | 36.7      | 37.9     |
|        | 10   | 29.1 | 56.5  | 27.0  | 24.1     | 37.5      | 38.2     |
|        | 20   | 29.3 | 56.9  | 27.3  | 24.3     | 37.7      | 40.2     |
|        | 30   | 29.1 | 56.6  | 26.8  | 24.5     | 36.8      | 39.4     |

#### [0098] (3) 位置信息

[0099] 在轮廓预测过程中,预测点的位置间有一定的相关性。本方法建立两个显式的位置信息特征图,分别表示x坐标和y坐标的相对位置,取值范围为(-1,1),根据特征图尺度均匀取值。通过在第一个卷积过程中加入两层特征,来显式的融入位置信息。实验结果如下表3所示,添加位置信息在ap上有0.7%的增益。

#### [0100] 表3位置信息性能影响

|        | 位置信息 | ap   | ap@50 | ap@75 | ap@small | ap@medium | ap@large |
|--------|------|------|-------|-------|----------|-----------|----------|
| [0101] |      | 29.3 | 56.9  | 27.3  | 24.3     | 37.7      | 40.2     |
|        | √    | 30.0 | 57.7  | 28.1  | 25.1     | 38.1      | 38.3     |

#### [0102] (4) 偏置回归分支

[0103] 轮廓预测中,需要将标注数据映射到池化的特征图并量化,从而得到需要预测的

类别。在量化过程存在位置损失,影响定位的精度。为解决上述问题,在基于注意力机制的长短时记忆网络的预测部分,添加一个回归分支去预测量化过程的损失。实验结果如下表4所示,添加回归分支对于ap@50并没有带来收益,主要体现在ap@75指标上,即对定位精度有更大的提升;对于物体尺度,性能提高主要体现在中大物体上,这是由于中大物体量化过程相比于小物体的损失更加明显导致的。

[0104] 表4偏置回归分支性能影响

|        | 偏置回归分支 | ap   | ap@50 | ap@75 | ap@small | ap@medium | ap@large |
|--------|--------|------|-------|-------|----------|-----------|----------|
| [0105] |        | 29.3 | 56.9  | 27.3  | 24.3     | 37.7      | 40.2     |
|        | ✓      | 29.9 | 56.9  | 27.6  | 24.0     | 38.8      | 40.7     |

[0106] (5) 平均核操作

[0107] 如果一个点附近预测得分较高,则它附近点应该获得较高的分数。为充分利用这种特性,增加推断过程的稳定性,对预测结果附近点的得分进行加权后,选择最高得分作为当前时序的预测结果。加权方法主要使用高斯加权和求和加权,该实验测试不同的加权及加权范围对精度的影响,加权通过卷积来实现,其中高斯核是二维高斯核函数,常数则表示全部用1进行加权。实验结果如下表5所示。第一行不使用求和加权,可以认为是使用核尺度大小为1的高斯核或者常数核。从表中可以发现,常数核会严重降低性能,是由于此时不考虑中心位置差别,导致得分分布发生严重的变化,当核尺度变大时,分布变化更加剧烈,从而导致后续的预测出现问题。通过使用高斯核,中心位置的权值高,附近的权值低,从而突出预测点的中心位置,获得更稳定的结果,在高斯核尺度是3或5的情况下,效果提升差不多,为简化计算,使用高斯核尺度3。

[0108] 表5不同核操作影响

|        | 方法 | 核尺度 | ap   | ap@50 | ap@75 | ap@small | ap@medium | ap@large |
|--------|----|-----|------|-------|-------|----------|-----------|----------|
|        |    |     | 29.3 | 56.9  | 27.3  | 24.3     | 37.7      | 40.2     |
| [0109] | 高斯 | 3   | 30.4 | 57.2  | 29.4  | 25.6     | 38.5      | 38.5     |
|        | 常数 | 3   | 28.5 | 56.2  | 26.0  | 25.2     | 34.8      | 33.0     |
|        | 高斯 | 5   | 30.4 | 57.2  | 29.4  | 25.6     | 38.5      | 38.6     |
|        | 常数 | 5   | 20.9 | 50.8  | 10.9  | 20.9     | 23.5      | 16.7     |

[0110] 根据以上实验结果分析,选择使用池化尺度28和时序长度20的实验配置进行消融实验,结果如下表6所示:

[0111] 表6各模块性能对比

|        | 位置信息 | 偏置回归分支 | 平均核 | ap   | ap@50 | ap@75 |
|--------|------|--------|-----|------|-------|-------|
|        |      |        |     | 29.3 | 56.9  | 27.3  |
| [0112] | ✓    |        |     | 30.0 | 57.7  | 28.1  |
|        |      | ✓      |     | 29.9 | 56.9  | 27.6  |
|        |      |        | ✓   | 30.4 | 57.2  | 29.4  |
|        | ✓    | ✓      | ✓   | 31.5 | 58.0  | 30.5  |

[0113] 5) 与其他方法的对比

[0114] 表7报告了所提方法与Mask RCNN在公开数据集上的结果对比。为了比较建筑物轮廓的拟合结果,对Mask RCNN的分割结果进行后处理,表7中Mask RCNN(矩形)表示将多边形

结果使用最小外接矩形表示,Mask RCNN(多边形)表示利用拟合算法将掩码结果拟合成多边形。本方法通过基于注意力机制的长短时记忆网络可以直接得到多边形结果,不需要经过后处理操作。从表中可以看出,本方法精度相比于拟合算法要更高,大约提高了6.1%,由此证明提出方法的有效性。

[0115] 表7所提方法与其他方法的对比

| [0116] | 方法              | ap   | ap@50 | ap@75 | ap@small | ap@medium | ap@large |
|--------|-----------------|------|-------|-------|----------|-----------|----------|
|        | Mask RCNN (矩形)  | 31.5 | 58.1  | 32.3  | 21.0     | 38.8      | 35.0     |
|        | Mask RCNN (多边形) | 31.9 | 58.0  | 30.5  | 26.6     | 39.6      | 41.9     |
|        | 我们的方法           | 36.4 | 59.9  | 39.9  | 26.6     | 44.4      | 36.6     |

[0117] 以上所述仅为本发明的优选实施例而已,并不用于限制本发明,对于本领域的技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

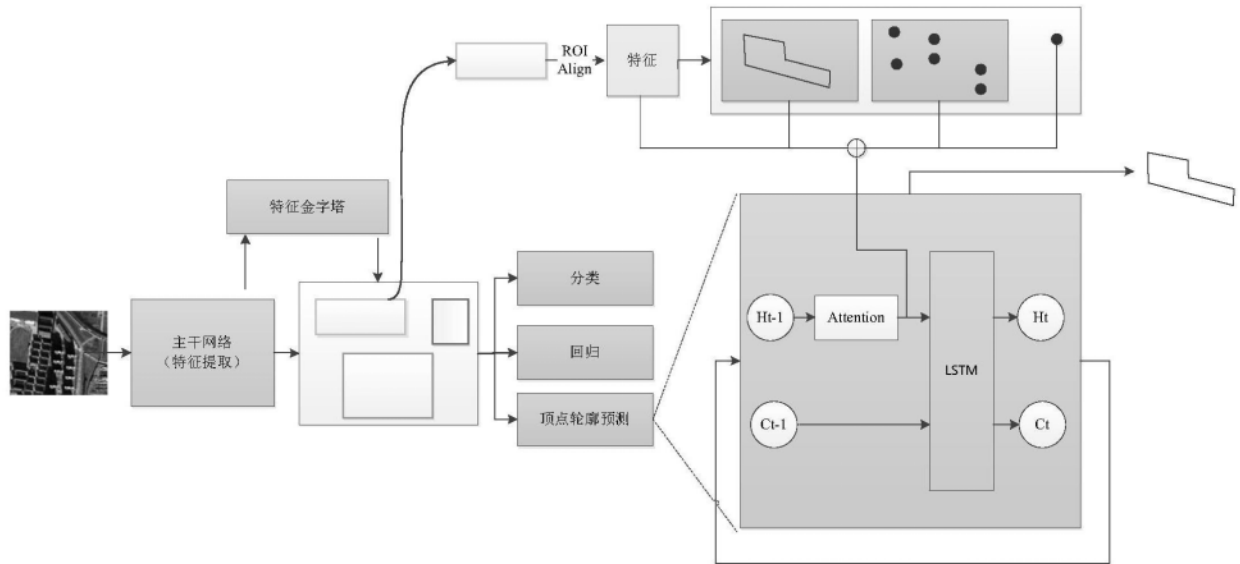


图1

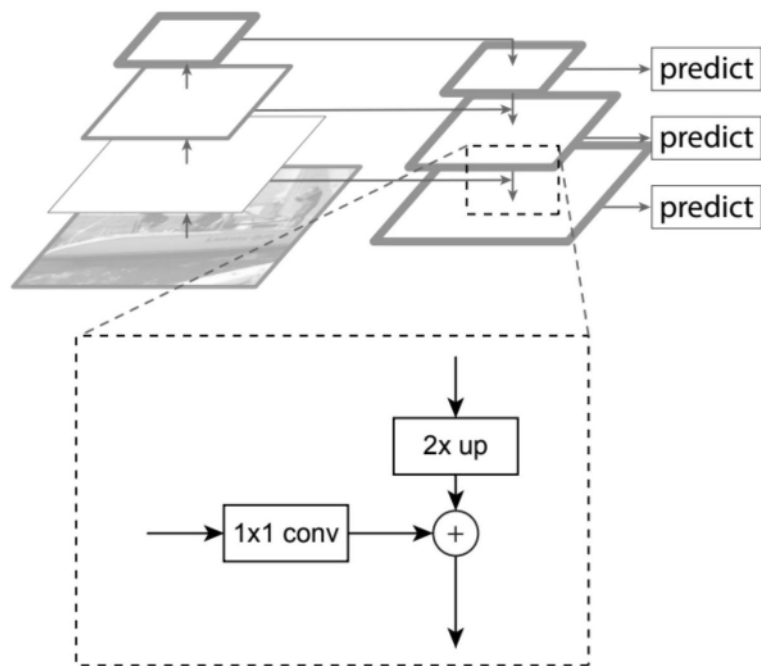


图2

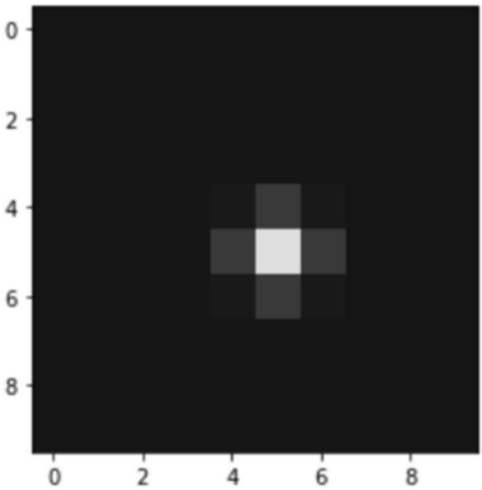


图3

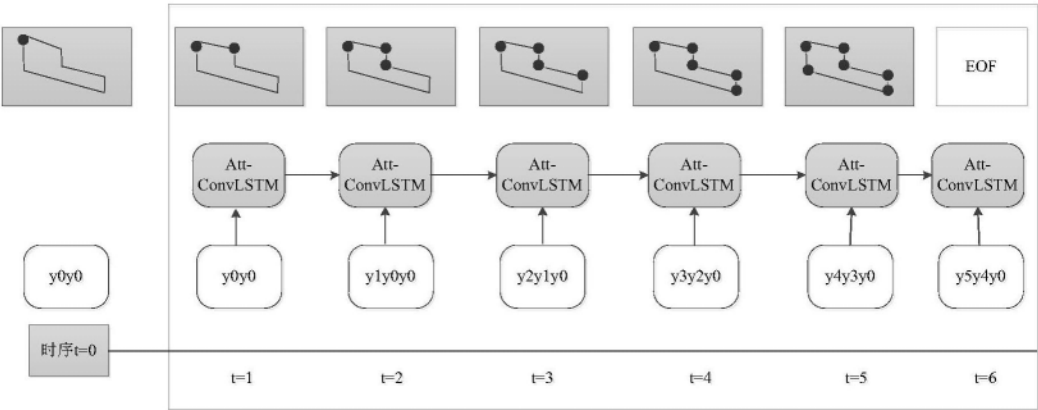


图4



图5