街景影像弱监督语义变化检测方法及其城市更新动态制图研究

彭奕霖[1,2]，付迎春[2*]，邢汉发[1]，陈书其[2]，李贞豪[2]，张思[2]

1. 华南师范大学 北斗研究院, 佛山 528225;
2. 华南师范大学 地理科学学院, 广州 510631

摘　要：街景影像是感知城市物质环境的一种新型地理大数据。通过街景发现立面变化并识别变化语义类别是城市更新的重要感知手段。传统变化检测方法无法直接区分街景变化物体的时相归属（变化拆分），难以高效识别两个时相中变化区域的语义类别。本文提出Cross-C2PO （Cross-Combine 2 POssible change types）模型统一变化检测与时相拆分任务，有助于引入现有图像语义分割模型实现街景语义变化检测。在此基础上，文章构建了城市更新动态度指标的感知分析方法，以广州主城区2013-2019年更新变化监测为目标，进行街景全景综合变化感知，实现前后左右视四个视角感知的城市更新动态度制图，直观展现了城市更新的分布及物理环境变化强度，为街景与计算机视觉智能综合应用提供创新方法和案例研究。

关键词：城市更新，街景影像，语义变化检测，场景变化检测，弱监督，动态度

中图分类号：TP181

引用格式：彭奕霖，付迎春，邢汉发，陈书其，李贞豪，张思.XXXX.街景影像弱监督语义变化检测方法及其城市更新动态制图研究.遥感学报，XX(XX)：1-19

PENG Yilin, FU Yingchun, XING Hanfa, CHEN Shuqi, LI Zhenhao, ZHANG Si. XXXX. Weakly Supervised Semantic Change Detection in Street View Imagery and Its Application in Urban Renewal Dynamics Mapping. National Remote Sensing Bulletin, DOI:10.11834/jrs.20255171]

1 引言

街景影像（Street view imagery，SVI）是表达城市环境的一种新型的大数据源，其观测视角更接近于城市居民，表达内容丰富，已被广泛应用于城市分析中（张帆等，2021）。街景影像不但可以详尽地描绘城市物质空间环境，例如建筑物（Liu等，2024）、道路、自然地物（Badland等，2010；Liang等，2024）等，同时隐含有丰富的城市功能、社会经济与人类活动的信息（Ma等，2021），是联系人和地理环境相互作用的纽带，也是表达城市变化及地理知识的基本单元。相比而言，卫星遥感影像应用在城市研究具有大尺度对地观测的优势，但缺乏对微观建成环境的全面、整体和精细化分析（Zhang等，2019）。因此，表达场所及场所变化的深层内涵，形成对城市立面变化的视觉感知、计算与制图的智能方法，是对三维空间变化的重要维度拓展，对城市更新变化监测具有重要意义。

近年来，计算机视觉的发展促进了街景影像自动提取各种信息的应用场景，深度学习的应用也大幅提升了数据处理的效率与准确性（Badland等，2010；Berland等，2017；Goel等，2018）。城市建成环境及其变化与城市健康-福祉密切相关，好的建筑环境可以改善并提升人居环境质量与人类身心健康（Davison等，2020）。基于计算机视觉的街景影像采用如绿视率（Long等，2017）、天空开阔度（Miao等，2020）及界面合围度（Yin等，2016）等解译指标，利用大规模街景数据源包括百度街景（BSV），谷歌街景（GSV）等反映城市

收稿日期：2024-10-31；**预印本：**XXXX-XX-XX
基金项目：国家自然科学基金面上项目(No. 42071399)
第一作者简介：彭奕霖，研究方向为街景变化检测。E-mail：2023025220@m.scnu.edu.cn
通信作者简介：付迎春，主要研究方向为定量遥感与城市遥感。E-mail：fuyc@m.scnu.edu.cn

空间环境变化潜力（Ma 等，2021），显著减少实地调研需求（Badland 等，2010），街景影像指标实现了对环境变化整体的感知，但缺乏街景立面的精细变化及语义信息的检测。

街景影像变化检测旨在识别不同时刻车载相机拍摄的图像对间的变化信息，主要包含二值变化检测（Binary Change Detection， BCD）以及语义变化检测（Semantic Change Detection， SCD）两类，前者要求模型预测一个变化掩码以指示图像对中的变化与未变化（背景）区域，后者进一步区分变化区域所属时相（变化拆分）以及变化像素的语义类别。二值变化检测通常作为语义变化检测的一个步骤，早期研究利用手工提取特征或显式特征匹配生成变化掩码（Radke 等，2005；Sakurada 等，2017）。最近，得益于卷积神经网络的图像处理架构在街景影像分析中取得的良好性能（Ronneberger 等，2015；Shelhamer 等，2017；Szegedy 等，2014）。当前 BCD 方法逐渐转向更加模块化的设计，通常使用“编码器-特征融合-解码器”架构，其中特征融合算子处理两个不同来源的特征图以表征差异信息，被认为是关键组件。FC-Siam-conc（Caye 等，2018）直接将特征对沿通道拼接进行特征融合，FC-Siam-diff（Caye 等，2018）则采用特征对的差值与原始特征进行拼接。CSCDNet（Sakurada K 等，2020）引入用于光流估计的相关层以辅助定位变化位置。HPCFNet（Lei 等，2021）采用交叉特征堆叠充分利用多级特征，并使用不同尺寸的空洞卷积进行多样化的特征融合。DR-TANet（Chen 等，2021）引入注意力机制（Vaswani 等，2023）在固定范围内查找相似关系，并提出横向，纵向注意力机制优化条状物体的检测。C-3PO（Wang 等，2023）提出时序融合模块（Merge Temporal Features， MTF），通过不同子函数对变化进行类型建模以提升检测能力。这些精心设计的特征融合算子改善了 BCD 任务的性能。

然而，少有研究关注网络结构的改进，以上方法默认使用的单分支结构是“不完备”的：对于大部分的特征融合算子（MTF 除外），例如特征对拼接，特征对差值等。由于它们是不满足交换律的函数，在单分支结构中，交换图像对的输入顺序将生成不同的变化检测结果，这将产生歧义，并且这种单分支结构无法直接完成变化拆分，因此传统语义变化检测方法（Sakurada 等，2020）包含两个阶段：首先通过常规的二值变化检测模型生成一个联合变化掩码，第二阶段通过一个精心设计的模型将变化拆分以及语义分割耦合处理生成最终结果，这使得其无法直接利用现有最先进的语义分割模型（Chen 等，2017；Ravi 等，2024；Xie 等，2021），而需要从当前语义分割数据集中生成同时具备两个任务所需标签的合成数据集对所设计的模型进行从头训练，极大增加了研究成本，并且这一步骤包含了多个预处理过程，包括类别采样以及多种形态学变换，进一步增加了 SCD 任务的复杂度。因此，无法直接区分变化物体归属时相仍然是限制现有方法应用范围的重要瓶颈。

为此，本文提出一种新颖的双分支交叉融合结构 Cross-C2PO （Cross-Combine 2 POssible change types），首次实现了端到端的弱监督变化拆分，并结合语义分割模型简化了弱监督语义变化检测流程。在第一阶段同时进行二值变化检测与变化拆分并且无需引入任何额外的合成数据进行训练，第二阶段中直接利用现有语义分割模型及其预训练权重生成语义变化结果。值得注意的是，针对不同应用场景的分割模型通常是易于获取的，因此，这种设计在实际应用中具备良好的可扩展性以及较低的研究成本。

同时，Cross-C2PO 的结构是“完备”的，这意味着无论所使用的特征融合算子是否满足交换律，交换图像对的输入顺序都将产生一致的生成结果。Cross-C2PO 在多个变化检测数据集上相比多种主流方法表现出更优的性能，并且消融实验进一步表明了 Cross-C2PO 结构的稳健性与可迁移性。基于模型变化检测结果，本文提出了城市更新动态度指标计算与制图方法，将变化检测技术与更新应用场景联系起来，以广州海城区 2013-2019 年城市更新检测为目标，形成全景以及前视、后视、左视、右视四个视角方位感知下的城市更新动态度制图，进行多功能区的更新变化检测统计分析及可视化，直观展现了城市物理变化分布，以帮助开发者识别热点区域，为更新策略的制定提供参考，为街景与计算机视觉智能结合应用提供重要的方法和案例研究。

2 研究区和数据

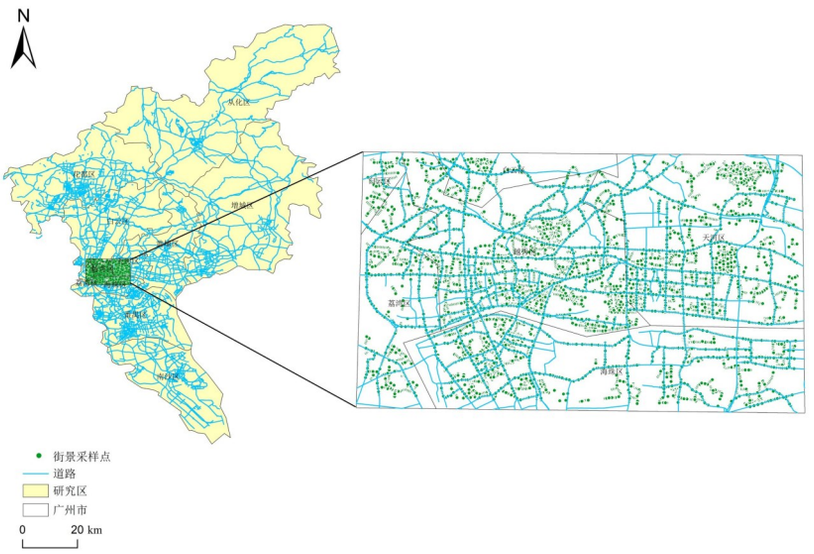

图1 研究区示意图

Fig.1 Study Area

2.1 研究区概况

广州作为较早开始更新改造的城市，分别经历了2009年以前的危房改造导向、2009—2014年的产业置换导向和2014年至今的集约用地导向（姚之浩等，2017）。“十三五”前采取拆除重建（拆建）为主，之后提倡整治保护为主的微改造（微改）。本文选取广州新旧中轴线所在的旧城区作为研究区，分布在东经113°13′27.61″至113°21′56.77″、北纬23°5′6.37″至23°9′58.56″，空间上覆盖越秀、荔湾、海珠和天河及白云区的部分区域，总面积约为150 km²。近二十年来，研究区进行了旧城与旧厂改造更新和历史街区的微更新，更新全面多样，更新时长、范围广且密集分布，占旧城区总更新区域面积的比例约为62%（付迎春等，2022；黄慧明，2013）。

2.2 街景影像获取与结果处理

本文所用街景影像源自百度全景地图。基于2021年广州市道路设施数据构建的中心城区路网，以50米间隔布设采样点，分别获取2013年以及2019年相同地理位置上的360°全景影像。为确保数据可比性，对于无法同时获取两个时相影像的采样点（存在数据缺失情况），均视为无效点位予以剔除。最终采集包含越秀区、海珠区、荔湾区、天河区、白云区部分区域上的11431对全景影像，影像大小为1024 × 2048像素（数据总存储量约18GB）。百度全景地图采用等距长方形投影，为了获取街道四个方位的城市更新动态度，本文将全景影像的变化检测结果转换为立方体投影，以获得前后左右上下六个视角的透视图，这种转换有效减少了街景的图像畸变，使影像更接近真实视角（Orhan，2022）。其中“上”视角影像主要包含天空，不包含或仅包含少量建筑，“下”视角影像仅包含拍摄车辆，因此，本文仅采用前后左右四个视角的结果图进行计算，如图2结果处理所示。同时，论文使用的辅助数据包括该时段的广州市功能区分类数据（Gong等，2020）以及广州市行政区划矢量数据。

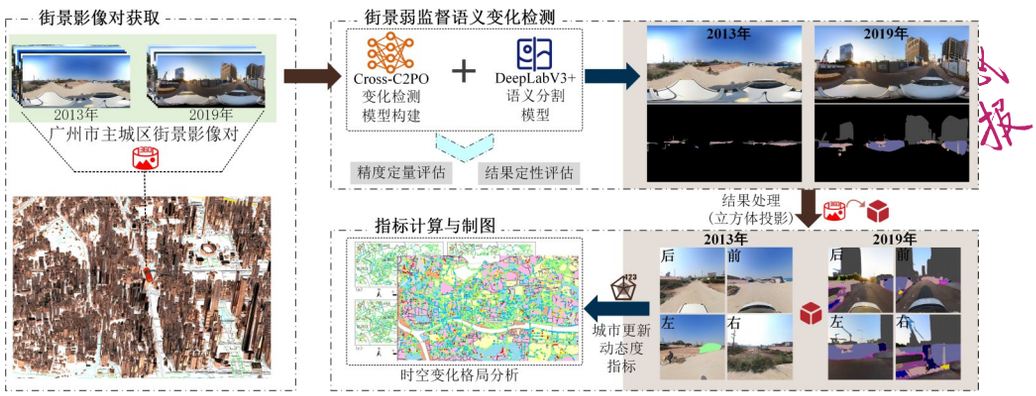

图2　城市更新检测与分析整体流程

Fig. 2　Overall flowchart of Urban Renewal Detection and Analysis

3　研究方法

研究整体框架包括图2所示三个部分：街景影像对获取，街景弱监督语义变化检测以及指标计算与制图，通过下述4个步骤实现：（1）沿路网生成采样点，分别获取每个采样点上2013年，2019年车载相机拍摄的全景影像；（2）使用本文提出的语义变化检测流程对全景影像进行处理，以获取影像对中的变化物体及其类别信息；（3）利用立方体投影，将全景视角下的语义变化检测结果转换为前后左右四个透视视图以获取符合人眼观察规律的变化感知；（4）提出城市更新动态度指标，利用全景以及四个透视视图检测结果评估采样点的更新程度，进行时空变化格局分析与制图。

3.1　街景弱监督语义变化检测

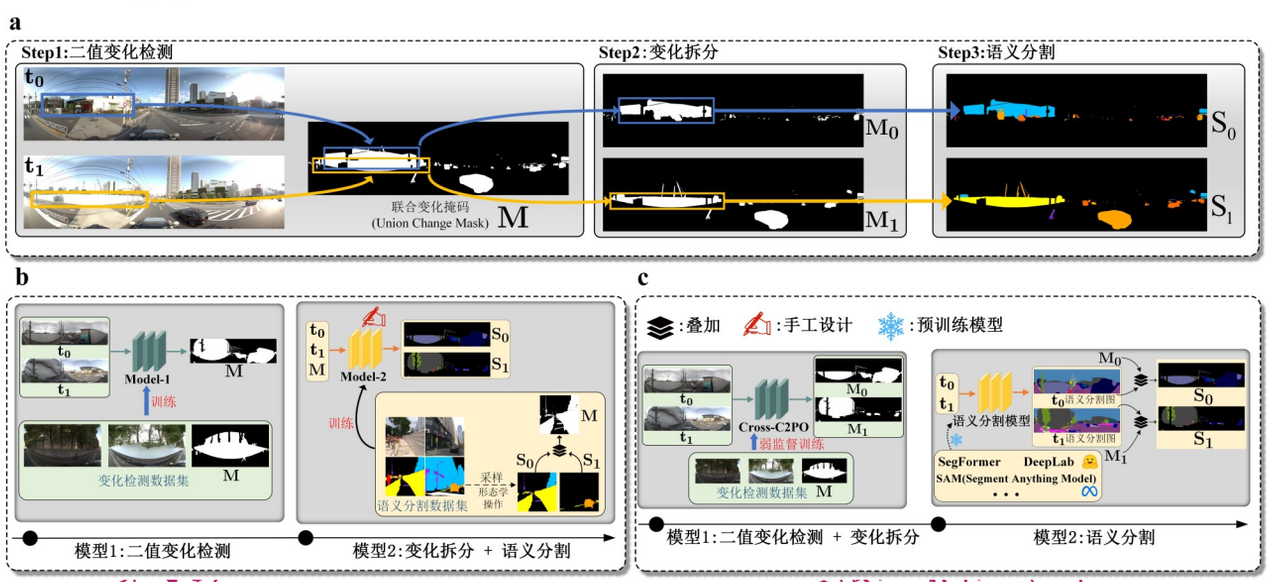

图3　(a)弱监督街景语义变化检测示例;(b)传统方法流程;(c)本文流程

Fig.3　(a) Example of weakly supervised semantic change detection in street scenes; (b) Workflow of traditional methods; (c) Workflow of our proposed approach

弱监督语义变化检测可以避免为不同场景创建对应的数据集，同时提供变化物体类别信息（Sakurada等，2020），面向城市更新应用目标，本文利用这种类别信息，去除流动要素（行人，车辆等）的干扰以实现对城市建筑街道等物理环境的进一步分析。如图3a所示，其通常包含3个步骤：首先检测得到联合变化掩码**M**，**M**指示了图像对中所有变化物体所在区域的并集，这也是传统二值变化检测模型的目标。由于重叠的变化区域可能包含两个不同类别的物体（图3a蓝色与黄

色框），无法直接赋予语义类别，因此需要区分变化物体的时序归属得到$\mathbf{M}_0$，$\mathbf{M}_1$，分别指示了t_0，t_1时刻影像中的变化区域，这一步骤称为变化拆分。最后将$\mathbf{M}_0$，$\mathbf{M}_1$中的变化区域赋予对应物体的语义类别，得到最终的语义变化检测结果S_0，S_1。

由于传统二值变化检测方法无法完成变化拆分步骤，以往的弱监督语义变化检测方法（Sakurada等，2020）需要通过一个精心设计的网络将变化拆分与语义分割两个子问题耦合处理（图3b），然而这种耦合使得网络的设计困扰与解决一般的分割问题，同时需要创建一个包含两个子问题所需标签的合成数据集进行从头训练，这一步骤包含类别采样，以及多种形态学操作进一步增加了复杂度。

本文提出新的弱监督语义变化检测流程如图3c所示，首先通过Cross-C2PO模型同时处理二值变化检测与变化拆分，Cross-C2PO是一个易于迁移的架构，传统二值变化检测方法可以方便的迁移到此架构上实现端到端的弱监督变化拆分，而无需引入$\mathbf{M}_0$，$\mathbf{M}_1$的标签信息以及其它辅助数据，然后充分利用最先进的语义分割模型及其预训练权重，例如DeepLab（Chen等，2017），SegFormer（Xie等，2021），SAM（Ravi等，2024）得到图像对的语义分割结果，最后仅保留由Cross-C2PO得到的变化拆分掩码$\mathbf{M}_0$，$\mathbf{M}_1$所指示的变化区域位置，得到最终的语义变化检测结果S_0，S_1。

3.2 Cross-C2PO模型结构

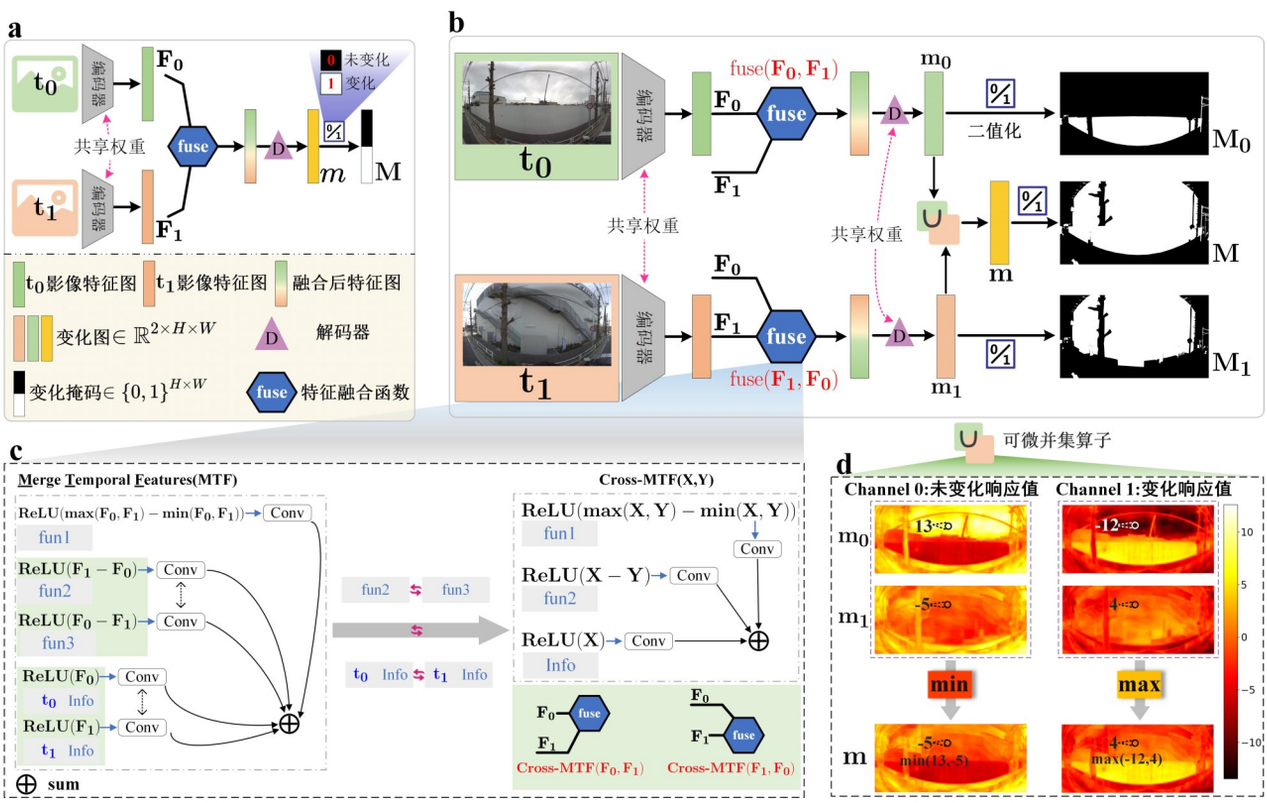

图4 （a）传统二值变化检测模型架构；（b）Cross-C2PO模型架构；（c）Cross-MTF特征融合算子；（d）可微并集算子。

Fig.4 (a) Traditional binary change detection model architecture; (b) Cross-C2PO model architecture; (c) Cross-MTF feature fusion operator; (d) Differentiable union operator.

传统二值变化检测模型采用单个检测分支的“编码器-特征融合-解码器”架构（图4a）。首先使用编码器得到图像对的多尺度特征图，然后特征融合算子也即fuse函数在单个分支中融合特征图对以表征差异信息，最后融合后的特征图经由解码器得到检测结果，这种设计并未对时间维度的信息加以区分，难以分离变化物体的时相归属。本文重新审视变化检测任务，提出了一种新颖的双分支架构以实现端到端的弱监督变化拆分，并将C-3PO（Wang等，2023）迁移到此架构上最终形成所提出的Cross-C2PO（图4b）。

具体来说，本文将变化检测任务定义为交叉

对比寻找相对变化的过程：t_0与t_1时刻影像对比代表了t_0相对t_1的变化，以表征t_0时刻影像中的变化物体，反之，t_1与t_0时刻影像对比代表了t_1相对t_0的变化，以表征t_1时刻影像中的变化物体，这意味着将两个不同的对比顺序视为不同的检测过程。这种双路径对比机制将传统单一路径的全局差异检测解耦为两个方向敏感的变化表征过程，从而显式建模时相间的非对称变化关系。Cross-C2PO采用VGG-16（Simonyan等，2015）作为编码器提取输入图像对特征，分别生成从1/4到1/32分辨率的多尺度特征图$\mathbf{F}_0$，$\mathbf{F}_1$。为有效捕捉时相间的相对变化设立了两个具有不同输入顺序的分支，在上分支中，使用$\mathrm{fuse}(\mathbf{F}_0, \mathbf{F}_1)$进行特征融合以表征$\mathbf{F}_0$相对$\mathbf{F}_1$的特征差异，在下分支中则使用$\mathrm{fuse}(\mathbf{F}_1, \mathbf{F}_0)$表征$\mathbf{F}_1$相对$\mathbf{F}_0$的特征差异，为确保两个分支能够区分不同的变化方向，特征融合函数必须显式地不满足交换律（$\exists \mathbf{X}, \mathbf{Y} \in R^{C\times H\times W}$使得$\mathrm{fuse}(\mathbf{X}, \mathbf{Y}) \neq \mathrm{fuse}(\mathbf{Y}, \mathbf{X})$），这一特性避免了分支输出的同质化，是实现准确变化拆分的关键。作为C-3PO的架构迁移版本，Cross-C2PO继承了其解码器设计，采用特征金字塔网络（FPN）（Lin等，2017）融合不同空间尺度的特征，FPN有效改善了不同尺寸物体的检测能力，已被广泛应用在目标检测以及语义分割等任务中（He等，2017；Kirillov等，2019）。融合后的特征图分别经过解码器得到两个时相的变化图（Change map）$\mathbf{m}_0$，$\mathbf{m}_1 \in R^{2\times H\times W}$。

二值变化检测将每个像素分为两个类别，第0类为未变化（背景），第1类为变化，$\mathbf{m}_0$，$\mathbf{m}_1$包含的两个通道分别对应这两个类别的响应值。图像对的变化掩码（Change Mask）$\mathbf{M}_0$，$\mathbf{M}_1 \in \{0, 1\}^{H\times W}$由下式得到

$$\mathbf{M}_{0,1} = argmax\left(\mathbf{m}_{0,1}\right)\#(1)$$

其中argmax操作沿通道维度选取响应值最大的类别索引，实现二值化。为了变化掩码间保持正确的一致性关系：M中的变化区域等于$\mathbf{M}_0$与$\mathbf{M}_1$中变化区域的并集。本文设计了一种用于生成联合变化图$\mathbf{m}$的可微并集算子，通过维持这种一致性关系以使梯度正确传播实现弱监督训练（图4d）：

$$\boldsymbol{m} = \left[\min\left(\mathbf{m}_0^0, \mathbf{m}_1^0\right), \max\left(\mathbf{m}_0^1, \mathbf{m}_1^1\right)\right]\#(2)$$

其中$[\cdot]$代表沿通道方向拼接，$\mathbf{m}_{\{0,1\}}^i$代表$\mathbf{m}_{\{0,1\}}$的第i个通道。式（2）将每个位置上$\mathbf{m}_0$，$\mathbf{m}_1$的最小未变化响应值作为$\mathbf{m}$的未变化响应值，最大变化响应值作为$\mathbf{m}$的变化响应值，确保$\mathbf{m}$中的变化区域能够覆盖$\mathbf{m}_0$和$\mathbf{m}_1$中所有预测为变化的区域。

C-3PO提出的MTF模块通过三个不同的子函数隐式建模不同类型的变化（图4c）：

$$fun1 = ReLU\left(\max\left(\mathbf{F}_0, \mathbf{F}_1\right) - \min\left(\mathbf{F}_0, \mathbf{F}_1\right)\right)\#(3)$$

$$fun2 = ReLU\left(\mathbf{F}_0 - \mathbf{F}_1\right)\#(4)$$

$$fun3 = ReLU\left(\mathbf{F}_1 - \mathbf{F}_0\right)\#(5)$$

最终MTF被表述为：

$$MTF\left(\mathbf{F}_0, \mathbf{F}_1\right) = Conv(\mathrm{fun1}) + Conv(\mathrm{fun2} + \mathrm{fun3}) + Conv\left(t_0\mathrm{Info} + t_1\mathrm{Info}\right)\#(6)$$

其中Conv代表窗口大小为3×3的卷积操作，$t_0\mathrm{Info} = \mathrm{ReLU}(\mathbf{F}_0)$，$t_1\mathrm{Info} = \mathrm{ReLU}(\mathbf{F}_1)$用于提供辅助信息。然而，MTF因fun2与fun3，t_0Info与t_1Info的对称性而满足交换律也即$\mathrm{MTF}(\mathbf{F}_0, \mathbf{F}_1) = \mathrm{MTF}(\mathbf{F}_1, \mathbf{F}_0)$，这与所提架构的约束相矛盾。本文在此基础上提出Cross-MTF作为网络的fuse函数（图4c）。通过将对称操作合并为两个函数：

$$fun2(\mathbf{X}, \mathbf{Y}) = ReLU(\mathbf{X} - \mathbf{Y})\#(7)$$

$$Info(\mathbf{X}, \mathbf{Y}) = ReLU(\mathbf{X})\#(8)$$

Cross-MTF最终表述为：

$$Cross - MTF(\mathbf{X}, \mathbf{Y}) = Conv(\mathrm{fun1}) + Conv\left(\mathrm{fun2}(\mathbf{X}, \mathbf{Y})\right) + Conv\left(\mathrm{Info}(\mathbf{X}, \mathbf{Y})\right)\#(9)$$

此时，原始MTF所包含的子函数将被分配到上下两个不同输入顺序的分支中。这种设计既保留了MTF的建模能力，又通过分支特化破坏交换律以满足结构约束。

3.3 语义分割模型

本文选用DeeplabV3+（Chen等，2017）作为所提出的弱监督语义变化检测流程中的语义分割模型。DeeplabV3+是一个经典的语义分割模型，由于其出色的性能表现以及细节还原能力，已被广泛应用与许多实际应用场景中（刘春亭等，2022；赵通等，2024）。使用在Cityscapes数据集上训练和测试（交并比MIOU=76.2%）的DeepLabV3+（https://github.com/VainF/DeepLabV3Plus-Pytorch）对研究区街景影像对进行语义分割。Cityscapes（Cordts等，2016）是一个专为城市环境中的语义分割、实例分割和物体检测任务设计的大型数据集，其主要包含来自50个不同城市的街道场景，具有5000张精细标注的图像以及

20000张粗略标注的图像（共有19个类别）已被广泛应用与城市相关研究中并体现了良好的通用性（Orhan，2022；Yue等，2024）。图5展示了研究区域内几个样本的语义分割结果，从视觉上看，大部分地物的分割结果是准确的，例如车辆，建筑，树木，道路，围栏。准确的分割结果可以为后续城市更新动态度模型的构建和计算提供可靠依据。

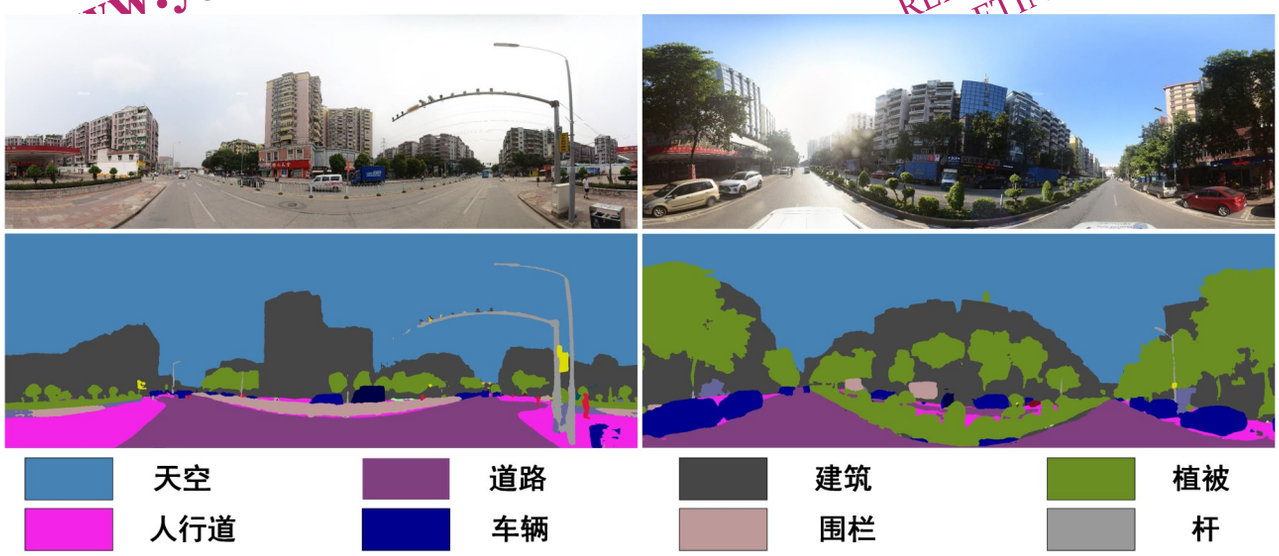

图5 语义分割结果示例

Fig.5 Example of Semantic Segmentation Results

3.4 城市更新动态度模型

土地利用动态度用于表示土地资源变化速率，以描述土地利用/土地覆盖变化程度，其主要分为单一土地利用类型动态度，综合土地利用动态度（王秀兰等，1999）。综合土地利用动态度表达的是某研究区土地利用总体变化率。本文提出城市更新动态度模型，将综合土地利用动态度模型引入街道级更新感知中，以感知城市立面的变化程度。

综合土地利用动态度模型公式表达如下：

$$LC = \frac{\sum_{i=0}^{n}\sum_{j=0}^{n} U_{ij}}{2\sum_{i=0}^{n} U_i} \times \frac{1}{T} \times 100\% \#(10)$$

其中：

$$U_{ij} = \begin{cases} U_{ij}, if\ i \neq j \\ 0, \quad if\ i = j \end{cases} \#(11)$$

U_{ij}代表监测时段内第i类土地利用类型转变为第j类土地利用类型的面积总和，U_i代表研究时段初期第i类土地类型的面积，T代表研究持续时长，若T单位为年，则LC代表该研究区综合土地利用年变化率。式（11）可以看出土地利用动态度模型不研究类别内的变化，但在城市更新研究中，包含许多同类别间的变化，例如：广告牌内容变化，建筑翻新等，因此在城市更新动态度模型中应去除此限制。

本文提出的城市更新动态度（City Dynamic）模型公式表达如下：

$$CD = \frac{C}{2K_0} \times \frac{1}{T} \times 100\% \#(12)$$

为了去除流动要素的干扰，对t_0，t_1语义变化检测结果去除车辆，行人类别后转换为二值变化掩码，并求并集得到去除流动要素后的联合变化掩码，C代表其变化部分的像素面积，K_0代表t_0图像语义分割结果去除天空类别后的像素面积，T代表研究持续时长，若T单位为年，则CD代表该研究区城市地物年变化率。本文研究2013，2019年广州市主城区城市更新年动态度，式（12）中T为7，各变量与式（10）对应关系如下：

$$\begin{cases} C \sim \sum_{i=1}^{n}\sum_{j=0}^{n} U_{ij} \\ K_0 \sim \sum_{i=0}^{n} U_i \end{cases} \#(13)$$

4 实验设置

4.1 数据集及预处理

本文在VL-CMU-CD（Alcantarilla等，2018），PSCD（Sakurada等，2020）两个常用的公开变化检测数据集上对Cross-C2PO进行了评估。VL-CMU-CD包含1362个经过配准矫正的透视图像对，遵循以往方法将图像调整为512×512大小，并将其中933对划分为训练集，429对划分为测试集，训练集通过旋转增强扩充至3732对（Wang等，2023）。VL-CMU-CD仅包含联合变化标签，不提供变化拆分标签。

PSCD包含770对全景街景图像，进一步将图像调整为1024×224大小，然后在宽上进行步长为56，窗口大小为224×224的滑动裁剪，并将生成的图像块调整到256×256分辨率，最后对每个图像块进行旋转增强。实验采用5折交叉验证进行训练与测试，每个训练集将包含36960对256×256大小的图像块，每个测试集将包含154对1024×256大小的图像。PSCD提供联合变化与变化拆分标签。

4.2 精度评价指标

与以往研究相同（Chen等，2021），本文使用F1分数评估变化检测算法的性能。F1分数是精确率（Precision）与召回率（Recall）的调和平均，能够综合衡量模型在准确性和全面性方面的表现，其值范围为0到1，更高的F1值代表更高的精确率与召回率。F1的公式表达如下：

$$F1-score=\frac{2\times Precision\times Recall}{Precision+Recall}\#(14)$$

在二值变化检测任务中，“变化”类别代表正，“未变化”类别代表负。给定真正TP（True Positive）、真负TN（True Negative）、假正FP（False Positive）、假负FN（False Negative），有$Precision=\frac{TP}{TP+FP}$，$Recall=\frac{TP}{TP+FN}$。

4.3 对比方法

为全面评估模型性能，本文将所提出的Cross-C2PO模型与8个基线模型进行比较：

（1）FC-EF（Caye等，2018）：FC-EF首先将图像对按通道方向融合，然后利用UNet网络进行变化检测，参数量少，简单有效，是常用的基准方法。

（2）Siam-conc（Caye等，2018）：Siam-conc是FC-EF的变体，使用一个孪生网络首先提取图像对特征，然后将特征对沿通道拼接进行融合，相较FC-EF，其在街景变化检测中通常展现出更为优异的性能（Chen等，2021）。

（3）Siam- diff（Caye等，2018）：Siam- diff是FC-EF的另一个变体，采用特征对的差值绝对值与原始特征图进行拼接融合检测变化，是常用的对比方法。

（4）Changenet（Varghese等，2019）：Changenet使用平行权重共享网络进行特征提取并结合了不同水平的卷积层的输出，以捕获物体的粗糙信息和细节信息改善检测效果，在街景变化检测中展现出优异的性能。

（5）CSCDNet（Sakurada等，2020）：CSCDNet在孪生神经网络中引入了用于光流估计的相关层改善了相机视角不同所带来的估计误差，在很多研究中作为对比基准。

（6）HPCFNet（Lei等，2021）：HPCFNet有效地利用密集融合架构进行多层次特征融合并提出一种适应位置和尺度变化的多部分特征学习策略，解决了变化区域的空间分布和尺度差异问题，是常用的变化检测基准方法。

（7）DR-TANet（Chen等，2021）：DR-TANet将注意力机制引入变化检测任务，以提升检测不同尺寸与形状目标的能力，并提出横向，纵向注意力机制进一步优化条状物体的检测。

（8）C-3PO（Wang等，2023）：C-3PO提出新的变化检测范式，充分利用语义分割模型改善变化检测性能，并隐式建模不同的变化类型进一步改善检测能力，在最近的研究中表现出优异的性能。

4.4 实现细节

训练阶段，使用ImageNet预训练权重初始化VGG-16。本文使用Adam优化器（Kingma等，2017）对模型进行优化，初始学习率为10^{-4}，并采用余弦退火调度策略（cosine annealing schedule）（Loshchilov等，2017）动态调整学习率。为了方便复现，对VL-CMU-CD数据集，batch size为4，对PSCD数据集，batch size为16，所有对比模型均训练100 epochs。我们使用1个NVIDIA GeForce

RTX 4090 GPU进行所有实验。

二值变化检测任务将每个像素位置分为“变化”或“未变化”两个类别其中之一，这实际上可以被视为一个二分类语义分割任务，因此，我们使用交叉熵作为损失函数训练所提出的Cross-C2PO。需要注意的是，通常的街景变化检测任务中，变化区域往往只占一小部分。为了缓解这种类别不平衡带来的影响，受（Wang等，2023）的启发，本文采用带权重的交叉熵损失函数，每个像素位置上的带权重交叉熵损失公式如下：

$$L = -\sum_{i=0}^{N-1} w_i y_i log p_i \#(15)$$

N代表类别数，在二值变化检测任务中$N=2$，其中第0类代表背景，也即未变化，第1类代表变化。y_i，p_i分别代表联合变化掩码真值以及模型预测值m的第i类。w_i代表第i类的权重，权重值通过统计训练集的各类别占比得到：

$$w_0 = 1 - w_1 = \frac{n_c}{n_a} \#(16)$$

n_c代表训练集中变化区域的像素数，n_a代表总像素数。

5 实验结果

由于传统方法无法完成端到端的变化拆分，本文首先对Cross-C2PO与主流变化检测基线模型在二值变化检测任务上的表现进行了定量评估（5.1节）。然后将这些基线模型迁移到本文提出的双分支架构中，迁移后的模型可以实现端到端的弱监督变化拆分，对迁移后模型在变化拆分任务上的表现进行了定量评估（5.2节）。最后，通过可视化分析验证语义变化检测流程的有效性。为深入解析模型组件贡献，在5.2节中进一步设计了系列消融实验。

5.1 对比实验

如表1所示，在VL-CMU-CD数据集上，Cross-C2PO相比C-3PO，F1分数提高1.6%，相比CSCDNet，HPCFNet以及DR-TANet分别提高了5%，6.4%，6.5%，总体来说相较基线模型在二值变化检测任务上有较大提升。同时表1还展现了不同方法的参数量以及在512×512大小图像上的推理速度（FPS）、计算复杂度（GFLOPs）。

表1 Cross-C2PO与基线模型在VL-CMU-CD数据集上的F1分数对比

Table 1 Comparison of F1 Scores between Cross-C2PO and Previous Models on the VL-CMU-CD Dataset

对比方法	F1(%)▮	编码器	参数量▮(M)	FPS▮	复杂度▮(GFLOPs)
FC-EF	44.6	U-Net	**1.35**	**354.8**	**24.9**
Siam-conc	65.6	U-Net	1.55	245.5	38.8
Siam-diff	65.3	U-Net	1.35	253.1	34.0
Changenet	60.3	ResNet-50	51.31	120.7	86.5
CSCDNet	76.6	ResNet-18	94.20	23.6	336.6
HPCFNet	75.2	VGG-16	-	-	-
DR-TANet	75.1	ResNet-18	33.39	75.7	56.3
C-3PO	80.0	VGG-16	55.41	25.6	950.5
Cross-C2PO	**81.6**	VGG-16	55.41	16.6	1500.0

如表2所示，Cross-C2PO在PSCD数据集上F1分数同样达到最优，相较于第二名C-3PO提升1%，相比CSCDNet，DR-TANet分别提高了6.7%，5.6%。图6展示了Cross-C2PO与C-3PO的可视化结果，可以看出，本文方法表现出更加准确的检测结果，并且在容易误检的区域表现良好，例如在右列中本文方法成功判断出窗户未发生改变，建筑颜色发生改变，而标签则错误的将窗户区域视为变化，C-3PO则没有检测出建筑颜色的改变。定量与可视化的定性结果表明我们的方法在二值变化检测任务下具有优良的性能。

5.2 消融实验

本研究将基线模型迁移至提出的双分支架构上，由于其下两个分支采用共享权重，因此不会引入额外参数，如表3所示，迁移后的模型均实现了端到端的弱监督变化拆分，其中本文方法Cross-C2PO作为C-3PO的迁移版本。横向对比来看，迁移后所有基线模型在二值变化检测任务中性能都得到了提升：对比表2，平均提升1.3%，其中FC-

EF提升2%，CSCDNet提升1.9%。纵向对比来看，本文方法在变化拆分任务上取得了最优的表现：相比第二名CSCDNet（C），两个时相的变化拆分性能分别提升5.3%、5.6%。图7展示了本文方法在PSCD数据集上的变化拆分可视化结果，可以看到其具有良好的视觉效果，基本实现了正确的变化物体所属时相区分。定量评估与良好的可视化结果表明本文所提出的架构具有良好的可迁移性与稳健性，可以在带来普遍的性能改进的同时实现变化拆分。同时表3还展现了迁移后模型的推理速度与计算复杂度，FC-EF（C）具有最快的推理速度，但检测精度较低，适用于具有稠密采样点，对精度要求较低，实时性要求较高的需要快速验证的应用场景。Cross-C2PO方法具有最高检测精度，实时性相对低，更契合街景影像更新周期长（天/月）的应用需求，适合用于常规的城市更新监测应用场景中。

表2　Cross-C2PO与基线模型在PSCD数据集上的F1分数对比

Table 2　Comparison of F1 Scores between Cross-C2PO and Previous Models on the PSCD Dataset

对比方法	F1（%）▮	编码器	参数量▮（M）	FPS▮	复杂度▮（GFLOPs）
FC-EF	64.5±1.1	U-Net	**1.35**	**381.3**	**24.9**
Siam-conc	72.3±0.5	U-Net	1.55	256.7	38.8
Siam-diff	71.8±0.5	U-Net	1.35	268.5	34.0
Changenet	59.6±0.7	ResNet-50	51.31	116.4	86.5
CSCDNet	75.8±0.3	ResNet-18	94.20	25.7	336.6
HPCFNet	–	VGG-16	–	–	–
DR-TANet	76.9±1.2	ResNet-18	33.39	78.3	56.3
C-3PO	81.5±0.1	VGG-16	55.41	29.5	950.5
Cross-C2PO	**82.5±0.2**	VGG-16	55.41	18.2	1500.0

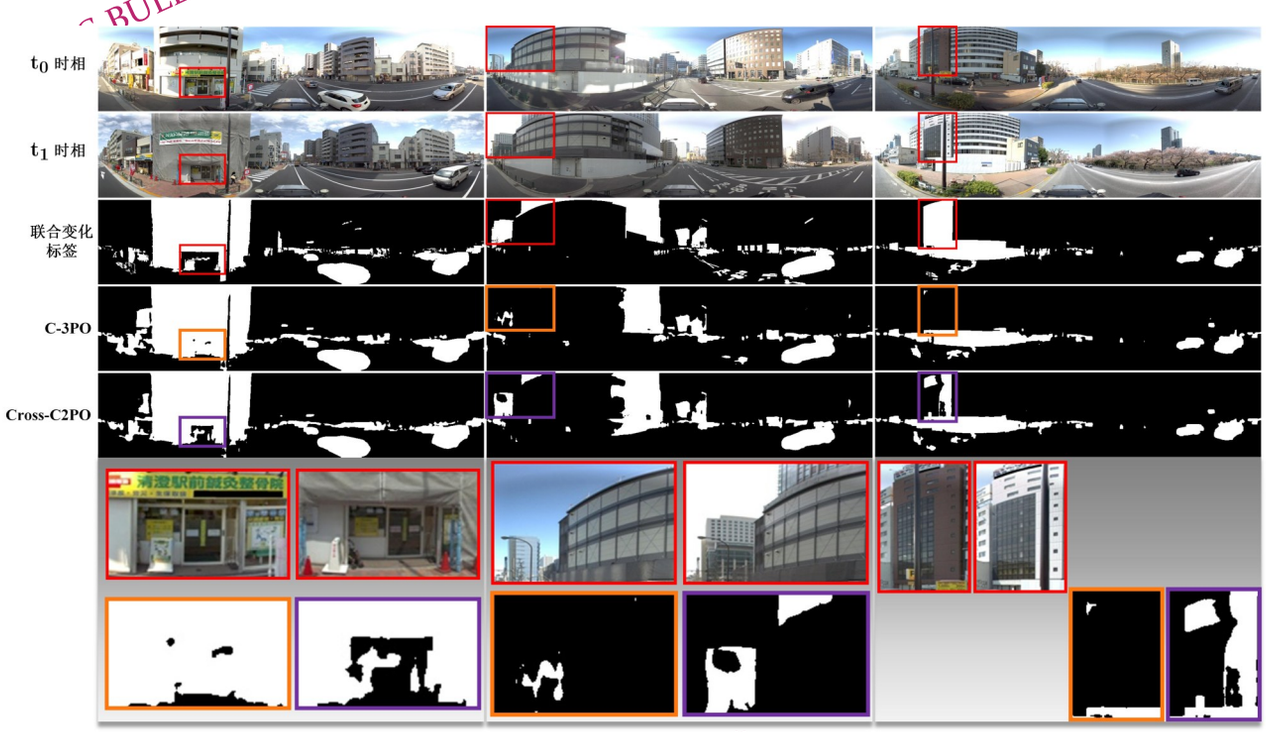

图6　PSCD数据集变化检测结果对比

Fig.6　Comparison of Change Detection Results on the PSCD Dataset

为了验证可微并集算子的有效性，本文设计了四种可微并集算子，用以生成联合变化图$\mathbf{m}$。“+”代表$\mathbf{m}=\mathbf{m}_0+\mathbf{m}_1$，max代表$\mathrm{m}=\max(\mathbf{m}_0,\mathbf{m}_1)$，min代表$\mathrm{m}=\min(\mathbf{m}_0,\mathbf{m}_1)$，其中

[min，max]代表上述式（2）。如表4所示，[min，max]相比其它算子在二值变化检测与变化拆分任务上都达到了最优的性能。

表3 基线模型架构迁移消融实验（PSCD数据集）：（C）代表原始模型的迁移版本，M，M_0，M_1分别代表联合变化掩码以及两个变化拆分掩码的F1分数。

Table 3 Baseline Model Architecture Transfer Ablation Experiment (PSCD Dataset): (C) represents the transfer version of the original model, and M, M_0, M_1 represent the F1 scores of the union change mask and the two change split masks, respectively.

对比方法	F1（%）			编码器	参数量（M）	FPS ↑	复杂度 ↓（GFLOPs）
	M	M_0	M_1				
FC-EF（C）	66.5±0.9	41.9±0.4	46.3±0.8	U-Net	**1.35**	**183.0**	49.8
Siam-conc（C）	73.6±0.6	47.0±0.4	50.0±0.5	U-Net	1.55	169.5	58.9
Siam-diff（C）	73.0±0.3	49.6±0.1	54.4±0.3	U-Net	1.35	178.3	**49.3**
CSCDNet（C）	77.7±0.3	67.5±0.2	68.0±0.3	ResNet-18	94.20	13.8	631.5
DR-TANet（C）	77.3±0.7	59.6±0.4	62.0±0.9	ResNet-18	33.39	46.5	74.7
Cross-C2PO	**82.5**±0.2	**72.7**±0.1	**73.6**±0.3	VGG-16	55.41	18.2	1500.0

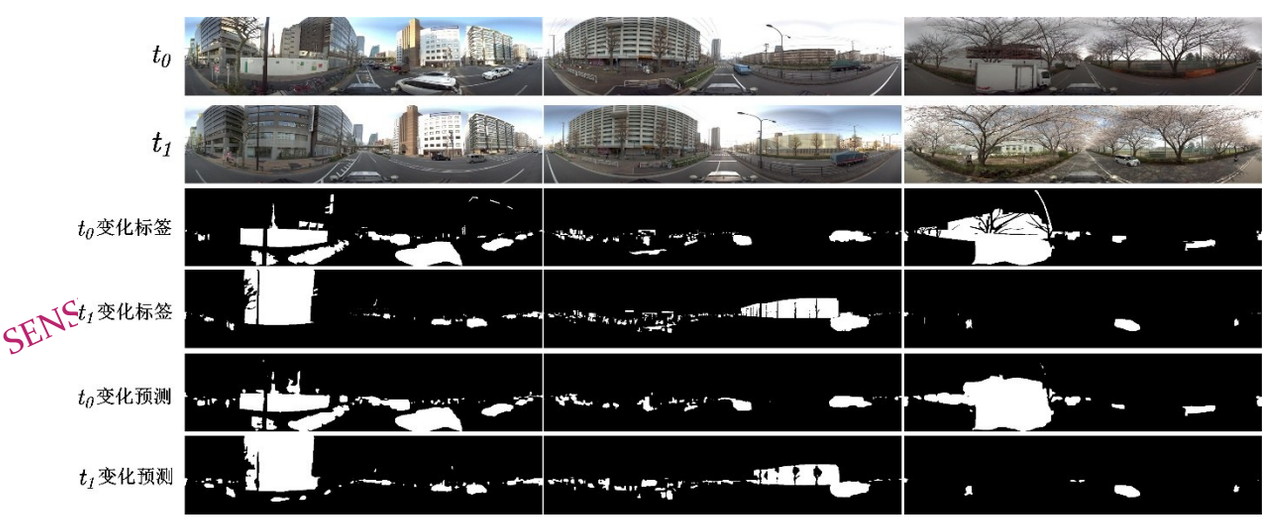

图7 Cross-C2PO在PSCD数据集上的变化拆分结果展示

Fig. 7 Visualization of Change Decomposition Results by Cross-C2PO on the PSCD Dataset

表4 可微并集算子消融实验（PSCD数据集）

Table 4 Differentiable Union Operator Ablation Experiment (PSCD Dataset)

可微并集算子	F1（%）		
	M	M_0	M_1
+	80.1	65.7	67.7
max	80.6	71.3	72.2
min	80.3	71.0	72.4
[min, max]	**82.5**	**72.7**	**73.6**

在表5中，通过比较两种监督策略，本文使用的弱监督训练策略超越了直接使用变化拆分标签进行训练的强监督策略，这表明所提出的可微并集算子实现了正确的反向传播并维持了良好的一致性关系，同时具有优异的泛化性能。

表4-5已经证明了所提出的可微并集算子有助于变化拆分的正确实现，接下来我们研究fuse函数，在表6中比较了直接使用原始MTF（式6）以及使用Cross-MTF（式9）作为Cross-C2PO中上下分支的fuse函数。实验结果指示使用Cross-MTF的变化拆分任务性能显著优于MTF（在两个时相上分别提升14.9%，12.1%），进一步证明了本文提出的观点：通过交叉对比实现相对变化检测时，特征融合函数必须设计为不满足交换律的形式，以确保时序信息的有效区分。

PSCD数据集包含多种城市场景与本文研究区

数据具有很高的相似性，因此本文使用PSCD数据集预训练的Cross-C2PO进行广州主城区的街景变化检测，并采用图3c所示的语义变化检测流程。最后，图8展示了研究区内样本去除行人，车辆后的最终检测结果。可以看出，发生变化的街景静止要素被很好的区分保留。语义变化检测结果基本正确，但仍存在部分噪声，例如最左列中，围栏与墙体分类混乱，这可能是由于围栏与墙体的结构相似性导致，但是这并不影响式（12）城市更新动态度计算的准确性。

表5　监督策略消融实验（PSCD数据集）：

Table 5　Supervision Strategy Ablation Experiment (PSCD Dataset): The strong supervision strategy directly uses change split labels for training, while the weak supervision strategy trains using only union change labels.

强监督策略直接使用变化拆分标签进行训练，弱监督策略仅使用联合变化标签训练。

监督策略	F1(%)		
	M	M_0	M_1
强监督	80.7	70.9	73.5
弱监督	**82.5**	**72.7**	**73.6**

表6　fuse函数消融实验（PSCD数据集）

Table 6　Ablation Experiment of fuse function (PSCD Dataset)

fuse函数	F1(%)		
	M	M_0	M_1
MTF	81.7	57.8	61.5
Cross-MTF	**82.5**	**72.7**	**73.6**

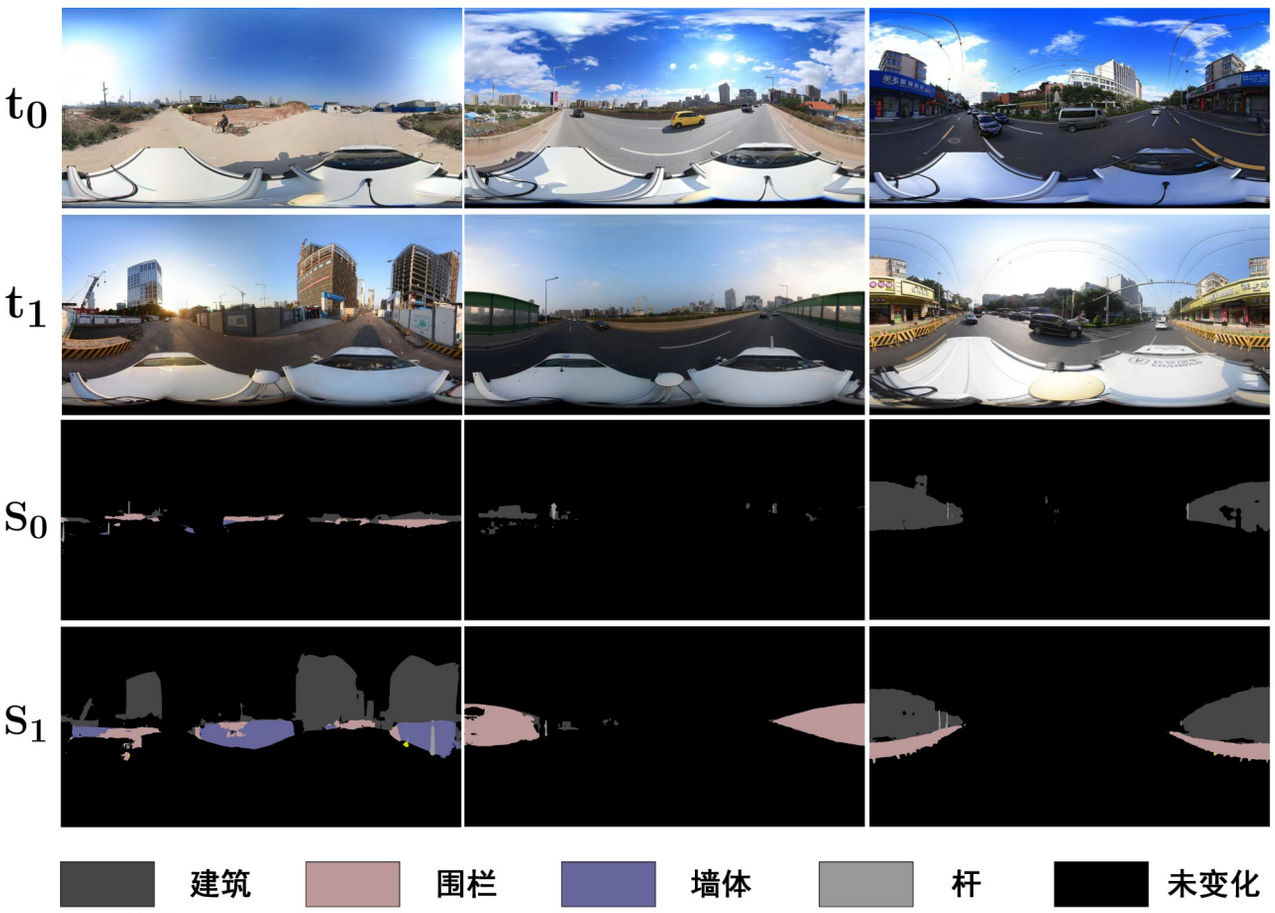

图8 研究区语义变化检测结果展示，t_0代表2013年影像，t_1代表同一地点2019年影像，S_0代表t_0的语义变化检测结果，S_1代表t_1的语义变化检测结果。

Fig. 8 Visualization of Semantic Change Detection Results in the Study Area. t_0 represents the 2013 image, t_1 represents the 2019 image of the same location, S_0 represents the semantic change detection results of t_0, and S_1 represents the semantic change detection results of t_1.

6 城市更新动态度指标制图与分析

图9展示了基于全景街景感知的研究区立面更新动态度及其分布，使用自然断点分级法将动态度分为高（1.86%–4.57%）占比4.13%、较高（1.02%–1.86%）占比13.78%、中（0.46%–1.02%）占比32.38%和低（0%–0.46%）占比49.71%四个等级，将街景动态叠置在2013年城市功能区六类用地上发现：高强度更新变化的街景主要分布在荔湾区与白云区的工业用地范围和海珠区与越秀区的公共服务与住宅用地类型；较高强度动态更新主要集中在海珠区与越秀区的少量带连续路段，而中等动态变化强度街景主要分布在住宅用地相关路段，推测以广告和门招等微更新变化为主，总体上高与较高更新空间与总体占比近似服从二八定律，说明了城市更新政策的有序执行，从城市边缘的大拆建逐渐转向有机更新特点。图9为理解城市发展的动态变化提供了重要依据，能够帮助城市规划和管理者识别变化热点及相对稳定区域，为未来的决策提供数据支持。

图10揭示了街景图像在前视、后视、左视、右视四个视角的更新动态变化强度与分布，因为前后视图与左右视图的视野不同，按不同强度等级制图并依次从低到高分为四个等级。如图10（a），（b）所示，前向与后向视图的高强度动态分布红色点位相似，直方图（e）与（f）的强度等级占比相似也说明了这一点；而左右视角的街景动态度显示出明显的分布格局差异，直方图的右视图（d）比左视图（c）具有约2倍多的高与较高强度变化，主要聚集发生在图10（d）中位于越秀区的广州城市中轴线分布，高与较高级更新点位叠加主要分布在交通与主干道接口，这跟广州市

实施历史街区活力更新有密切关系，相比全景下的更新动态度图9，四个方向视图显示了更加丰富的局部更新，如结合语义将进一步揭示街景变化类别的动态信息。

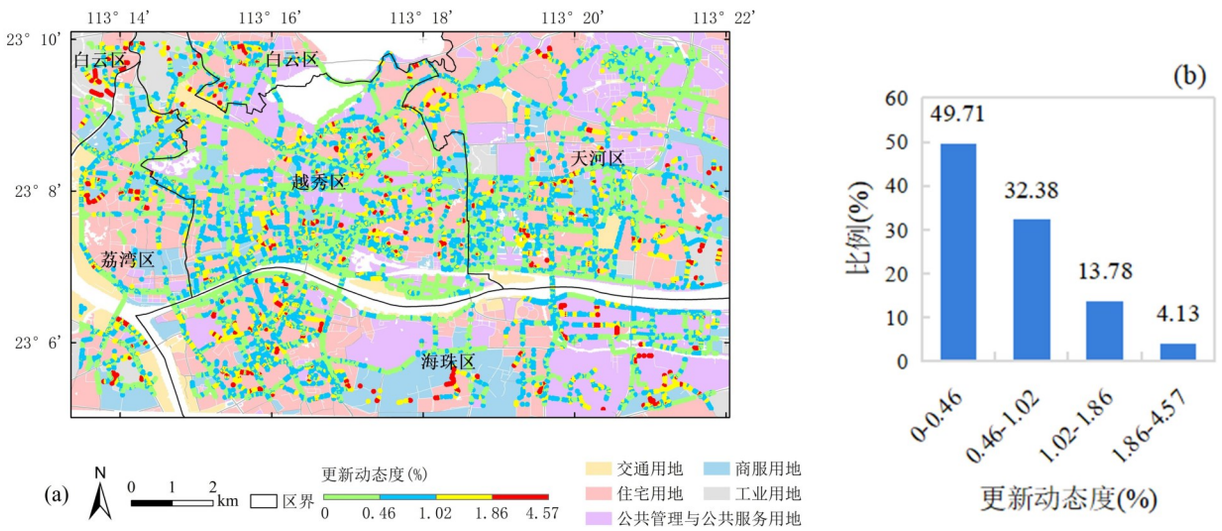

图9　基于全景街景感知的城市更新动态度图(a)及直方图(b)

Fig. 9　Urban Renewal Dynamics Map (a) and Histogram (b) Based on Panoramic Street View Perception

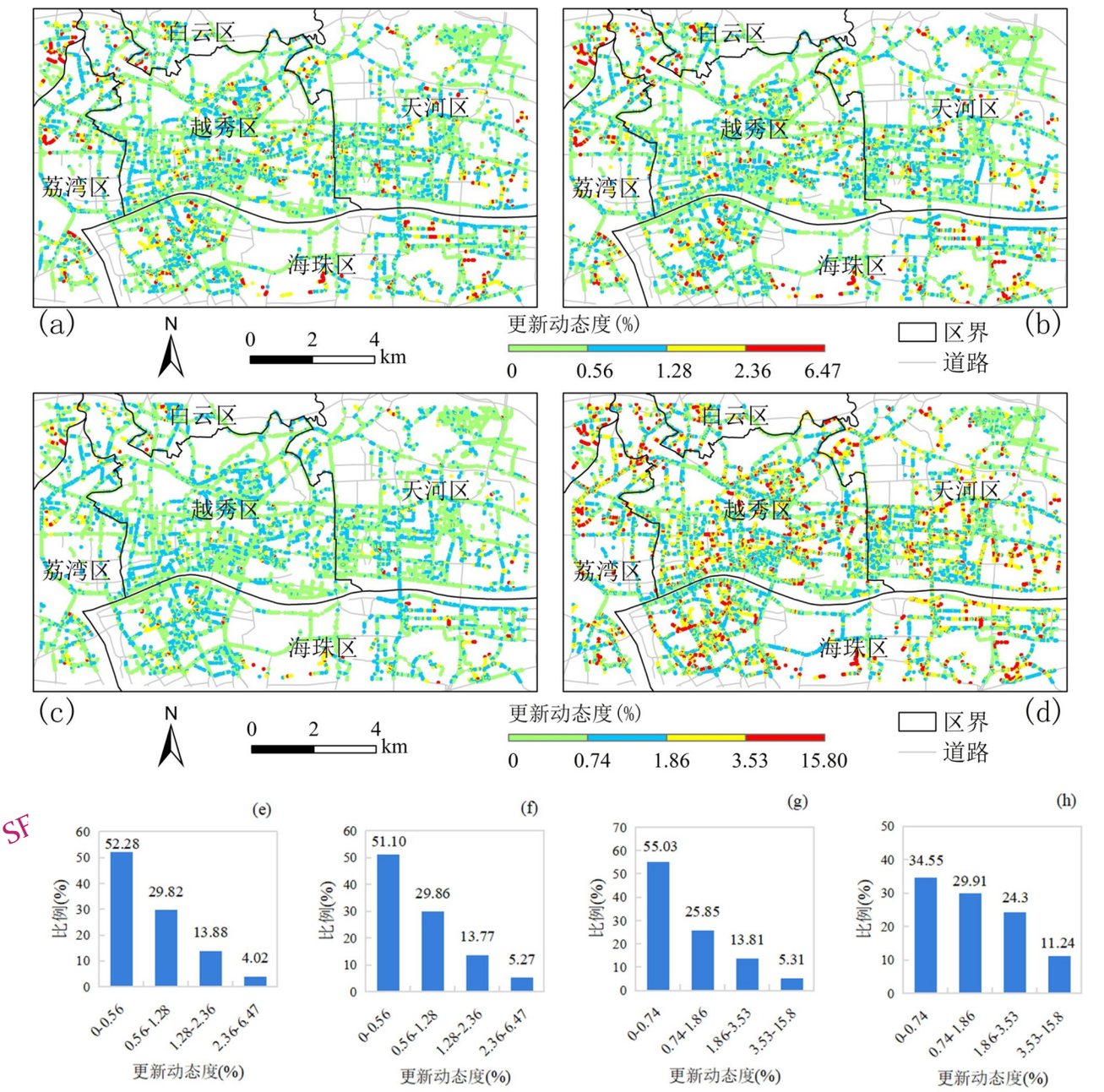

图10 基于全景街景感知的城市更新四方向视图(前(a),后(b),与左(c),右(d))及相应直方图(e~h)

Fig.10 Four-Directional Views of Urban Renewal Based on Panoramic Street View Perception (Front (a), Back (b), Left (c), and Right (d)) and Corresponding Histograms (e~h)

7 讨论与结论

本文提出了一种更加简洁高效的弱监督街景语义变化检测流程以及城市更新动态度指标计算框架，以百度街景影像为数据源，实现了2013年，2019年广州市主城区的城市更新变化监测并进行多功能区的更新变化检测统计分析及可视化，是面向低成本，自动化，精细化的城市更新监测的一次有益尝试，为街景与人工智能视觉结合应用提供重要的方法和案例研究，结合街景影像大规模数据源与变化检测模型扩展城市立面空间检测维度，刻画了城市更新阶段的不同地物更新活力与时空变化格局，该方法具有较好的迁移性与高效性。研究结论如下：

（1）本文通过一种新颖的相对角度审视街景变化检测问题，以交叉融合为基础，结合可微并集算子提出新的Cross-C2PO变化检测方法，实现端到端的弱监督变化拆分，并大幅简化语义变化

检测流程以及实现难度。实验结果表明，Cross-C2PO方法相比以往基准算法取得更为优异的检测精度以及稳定性，特别在变化拆分任务上，以往模型无法或难以实现。应用分析表明，本文提出的弱监督语义变化检测流程具有良好的稳健性以及易用性，并且具有较强的灵活性；在实际应用中，可根据目标区域的地物特点灵活选择更为合适的语义分割网络，可进一步扩展到更大规模的区域研究上。尽管当前主流方法可以迁移到Cross-C2PO架构中实现弱监督变化拆分，然而由于架构依赖与非对称的双路径检测机制，仍然会导致计算复杂度的提升以及推理速度的降低，未来研究将通过模型剪枝，路径建模统一化进一步降低计算复杂度以期满足轻量化应用需要。另一方面，将针对语义分割与变化检测紧密关联的特点，进一步优化语义变化检测流程，以期实现端到端的弱监督语义变化检测。

（2）本文提出一种街景变化的城市更新动态度指标及计算方法，综合街景变化点位与动态度形成城市立面变化特征的多维表征，开展对城市功能更新强度的全景地图变化格局与前视、后视、左视、右视图的局部分析，明确了城市街景变化与指标计算的内涵和外延；研究方法应用在2013年到2019年广州城市中心，得出和城市更新政策相互印证的结论：较高强度的街景全景变化主要分布在荔枝湾工业区和海珠区，中等水平更新主要分布在越秀区，而四个视角指标提供了空间分异的更新动态度感知，总体上高与较高更新空间与总体占比近似服从二八定律，城市更新政策的有序执行，从城市边缘的大拆建逐渐转向有机更新的政策执行特点。本研究存在一些潜在的局限。受车载移动采集方式的制约，无法覆盖道路变迁区域（新增/消失道路）以及非道路区域的更新动态。同时，受限于街景数据采集时间间隔，对不同时间跨度的敏感性尚未进行深入分析。这些因素对方法的可扩展潜力可能存在影响。未来研究将通过多源数据融合和多维指标体系构建予以改进并进一步揭示街景变化类别的动态信息，以期实现城市更新的街景视觉智能感知。

参考文献（References）

Alcantarilla P F, Stent S, Ros G, Arroyo R and Gherardi R. 2018. Street-view change detection with deconvolutional networks. Autonomous Robots, 42(7): 1301-1322 [DOI: 10.1007/s10514-018-9734-5].

Badland H M, Opit S, Witten K, Kearns R A and Mavoa S. 2010. Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? Journal of Urban Health, 87(6): 1007-1016 [DOI: 10.1007/s11524-010-9505-x].

Berland A and Lange D A. 2017. Google Street View shows promise for virtual street tree surveys. Urban Forestry & Urban Greening, 21: 11-15 [DOI: 10.1016/j.ufug.2016.11.006]

Caye Daudt R, Le Saux B and Boulch A. 2018. Fully Convolutional Siamese Networks for Change Detection. 2018 25th IEEE International Conference on Image Processing (ICIP). 4063-4067 [DOI: 10.1109/ICIP.2018.8451652]

Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv [DOI: 10.48550/arXiv.1706.05587]

Chen S, Yang K and Stiefelhagen R. 2021. DR-TANet: Dynamic Receptive Temporal Attention Network for Street Scene Change Detection. 2021 IEEE Intelligent Vehicles Symposium (IV). 502-509 [DOI: 10.1109/IV48863.2021.9575362]

Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S and Schiele B. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society: 3213-3223 [DOI: 10.1109/CVPR.2016.350]

Davison G, Ferris D, Pearson A and Shach R. 2020. Investments with returns: a systematic literature review of health-focused housing interventions. Journal of Housing and the Built Environment, 35 (3): 829-845 [DOI: 10.1007/s10901-019-09715-6].

FU Y C, GUO B Y, WANG M and QIN X L. 2022. Spatial resilience evolution of green space from the perspective of social-ecological system adaptive governance: A case study for urban renewal in Guangzhou, China. JOURNAL OF NATURAL RESOURCES, 2022, 37(8): 2118-2136 (付迎春, 郭碧云, 王敏, 覃小玲. 2022. 社会—生态系统适应性治理视角下绿地空间恢复力的演化——广州旧城区更新案例. 自然资源学报, 37(8): 2118-2136 [DOI: 10.31497/zrzyxb.20220813])

Goel R, Garcia L M T, Goodman A, Johnson R, Aldred R, Murugesan M, Brage S, Bhalla K and Woodcock J. 2018. Estimating city-level travel patterns using street imagery: A case study of using Google Street View in Britain. PLOS ONE, 13(5): e0196521 [DOI: 10.1371/journal.pone.0196521].

Gong P, Chen B, Li X, Liu H, Wang J, Bai Y, Chen J, Chen X, Fang L, Feng S, Feng Y, Gong Y, Gu H, Huang H, Huang X, Jiao H, Kang Y, Lei G, Li A, Li Xiaoting, Li Xun, Li Y, Li Zhilin, Li Zhongde, Liu Chong, Liu Chunxia, Liu M, Liu S, Mao W, Miao C, Ni H, Pan Q, Qi S, Ren Z, Shan Z, Shen S, Shi M, Song Y, Su M, Ping Suen H, Sun B, Sun F, Sun J, Sun L, Sun W, Tian T, Tong X, Tseng Y, Tu Y, Wang H, Wang L, Wang X, Wang Z, Wu T, Xie Y, Yang Jian, Yang Jun, Yuan M, Yue W, Zeng H, Zhang K, Zhang N, Zhang T, Zhang Y, Zhao F, Zheng Y, Zhou Q, Clinton N, Zhu

Z and Xu B. 2020. Mapping essential urban land use categories in China (EULUC-China): preliminary results for 2018. Science Bulletin, 65(3): 182-187 [DOI: 10.1016/j.scib.2019.12.007]

HUANG H M. 2013. The characteristics of urban morphological transformations and development mechanisms: A case study of Guangzhou since 1949. Guangzhou: South China University of Technology. (黄慧明 . 2013. 1949年以来广州旧城的形态演变特征与机制研究 . 广州: 华南理工大学)

He K, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV). 2980-2988 [DOI: 10.1109/ICCV.2017.322]

Kingma D P and Ba J. 2017. Adam: A Method for Stochastic Optimization. arXiv [DOI: 10.48550/arXiv.1412.6980]

Kirillov A, He K, Girshick R, Rother C and Dollár P. 2019. Panoptic Segmentation. arXiv [DOI: 10.48550/arXiv.1801.00868]

Lei Y, Peng D, Zhang P, Ke Q and Li H. 2021. Hierarchical Paired Channel Fusion Network for Street Scene Change Detection. IEEE Transactions on Image Processing, 30: 55-67 [DOI: 10.1109/TIP.2020.3031173]

Liang C, Jiang H, Yang S, Tian P, Ma X, Tang Z, Wang H and Wang W. 2024. Characterizing street trees in three metropolises of central China by using Street View data: From individual trees to landscape mapping. Ecological Informatics, 80: 102480 [DOI: 10.1016/j.ecoinf.2024.102480]

Lin T Y, Dollár P, Girshick R, He K, Hariharan B and Belongie S. 2017. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society: 936-944 [DOI: 10.1109/CVPR.2017.106]

Liu C and Song W. 2024. Mapping property redevelopment via GeoAI: Integrating computer vision and socioenvironmental patterns and processes. Cities, 144: 104644 [DOI: 10.1016/j.cities.2023.104644]

Liu C T, Feng Q L, Liu J T, Wang Y, Shi T G, Li Y, Gong J H and Zhao H H. 2022. Urban green plastic cover extraction and spatial pattern changes in Jinan city based on DeepLabv3+ semantic segmentation model. National Remote Sensing Bulletin, 26(12): 2518-2530 (刘春亭, 冯权泷, 刘建涛, 王莹, 史同广, 李毅, 龚建华, 赵辉辉 . 2022. DeepLabv3+语义分割模型的济南市防尘绿网提取及时空变化分析 . 遥感学报, 26(12): 2518-2530 [DOI: 10.11834/jrs.20220101])

Long Y and Liu L. 2017. How green are the streets? An analysis for central areas of Chinese cities using Tencent Street View. PLOS ONE, 12(2): e0171110 [DOI: 10.1371/journal.pone.0171110].

Loshchilov I and Hutter F. 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. arXiv [DOI: 10.48550/arXiv.1608.03983]

Ma X, Ma C, Wu C, Xi Y, Yang R, Peng N, Zhang C and Ren F. 2021. Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. Cities, 110: 103086 [DOI: 10.1016/j.cities.2020.103086]

Miao C, Yu S, Hu Y, Zhang H, He X and Chen W. 2020. Review of methods used to estimate the sky view factor in urban street canyons. Building and Environment, 168: 106497 [DOI: 10.1016/j.buildenv.2019.106497]

Orhan S. 2022. Semantic Segmentation of Panoramic Images and Panoramic Image Based Outdoor Visual Localization. Izmir Institute of Technology (Turkey)

Radke R J, Andra S, Al-Kofahi O and Roysam B. 2005. Image change detection algorithms: a systematic survey. IEEE Transactions on Image Processing, 14(3): 294-307 [DOI: 10.1109/TIP.2004.838698]

Ravi N, Gabeur V, Hu Y T, Hu R, Ryali C, Ma T, Khedr H, Rädle R, Rolland C, Gustafson L, Mintun E, Pan J, Alwala K V, Carion N, Wu C Y, Girshick R, Dollár P and Feichtenhofer C. 2024. SAM 2: Segment Anything in Images and Videos. arXiv [DOI: 10.48550/arXiv.2408.00714]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv [DOI: 10.48550/arXiv.1505.04597]

Sakurada K, Shibuya M and Wang W. 2020. Weakly Supervised Silhouette-based Semantic Scene Change Detection. 2020 IEEE International Conference on Robotics and Automation (ICRA). 6861-6867 [DOI: 10.1109/ICRA40945.2020.9196985]

Sakurada K, Wang W, Kawaguchi N and Nakamura R. 2017. Dense Optical Flow based Change Detection Network Robust to Difference of Camera Viewpoints. arXiv [DOI: 10.48550/arXiv.1712.02941]

Sampson R J. 2017. Urban sustainability in an age of enduring inequalities: Advancing theory and ecometrics for the 21st-century city. Proceedings of the National Academy of Sciences, 114(34): 8957-8962 [DOI: 10.1073/pnas.1614433114]

Shelhamer E, Long J and Darrell T. 2017. Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]

Simonyan K and Zisserman A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv [DOI: 10.48550/arXiv.1409.1556]

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2014. Going Deeper with Convolutions. arXiv [DOI: 10.48550/arXiv.1409.4842]

Varghese A, Gubbi J, Ramaswamy A and Balamuralidhar P. 2019. ChangeNet: A Deep Learning Architecture for Visual Change Detection. Computer Vision – ECCV 2018 Workshops. Cham: Springer International Publishing: 129-145 [DOI: 10.1007/978-3-030-11012-3_10]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2023. Attention Is All You Need. arXiv [DOI: 10.48550/arXiv.1706.03762]

Wang G H, Gao B B and Wang C. 2023. How to Reduce Change Detection to Semantic Segmentation. Pattern Recognition, 138: 109384 [DOI: 10.1016/j.patcog.2023.109384]

Wang X L and Bao Y H. STUDY ON THE METHODS OF LAND USE DYNAMIC CHANGE RESEARCH. PROGRESS IN GE-

OGRAPHY, 1999, 18(1): 81-87 (王秀兰, 包玉海 . 1999. 土地利用动态变化研究方法探讨 . 地理科学进展, 18(1): 81-87 [DOI: 10.11820/dlkxjz.1999.01.012])

Xie E, Wang W, Yu Z, Anandkumar A, Alvarez J M and Luo P. 2021. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. arXiv [DOI: 10.48550/arXiv.2105.15203]

Yao Y H and Tian L. 2017. Transition of Pattern and Modes of Governance for Urban Renewal in Guangzhou since the 21st Century. Shanghai Urban Planning Review, (5): 29-34 (姚之浩, 田莉 . 2017. 21世纪以来广州城市更新模式的变迁及管治转型研究 . 上海城市规划, (5): 29-34)

Yin L and Wang Z. 2016. Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. Applied Geography, 76: 147-153 [DOI: 10.1016/j.apgeog.2016.09.024]

Yue Z, Lo C Y, Wu R, Ma L and Sham C W. 2024. Urban Aquatic Scene Expansion for Semantic Segmentation in Cityscapes. Urban Science, 8(2): 23 [DOI: 10.3390/urbansci8020023].

Zhang F, Wu L, Zhu D and Liu Y. 2019. Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. ISPRS Journal of Photogrammetry and Remote Sensing, 153: 48-58 [DOI: 10.1016/j.isprsjprs.2019.04.017]

Zhang F and Liu Y. 2021. Street view imagery: Methods and applications based on artificial intelligence. National Remote Sensing Bulletin, 25(5): 1043-1054 (张帆, 刘瑜 . 2021. 街景影像——基于人工智能的方法与应用 . 遥感学报, 25(5): 1043-1054 [DOI: 10.11834/jrs.20219341])

Zhao T, Zhang S C, He X N, Xue B W and Zha F K. 2024. Improved DeepLabV3+ model for landslide identification in high-resolution remote sensing images after earthquakes. National Remote Sensing Bulletin, 28(9): 2293-2305 (赵通, 张双成, 何晓宁, 薛博维, 查富康 . 2024. 改进的DeepLabV3+模型用于震后高分遥感影像滑坡识别 . 遥感学报, 28(9): 2293-2305 [DOI: 10.11834/jrs.20243393])

Weakly Supervised Semantic Change Detection in Street View Imagery and Its Application in Urban Renewal Dynamics Mapping

PENG Yilin[1,2], FU Yingchun[2*], XING Hanfa[1], CHEN Shuqi[2], LI Zhenhao[2], ZHANG Si[2]

1.Beidou Research Institute, South China Normal University, Foshan 528225, China;
2.College of Geography Science, South China Normal University, Guangzhou 510631, China

Abstract: Objective Street view imagery (SVI) has emerged as an important geospatial big data source for perceiving the urban built environment. Accurately detecting facade-level changes and identifying their semantic categories is essential for monitoring urban renewal dynamics. However, existing change detection approaches struggle to separate temporal ownership of changed objects (change decomposition) and to directly provide semantic change information, leading to complex workflows and high data preparation costs. This study aims to develop a weakly supervised semantic change detection framework that integrates change decomposition and semantic labeling, and to apply it to dynamic mapping of urban renewal in Guangzhou, China.Method We propose Cross-C2PO, a novel dual-branch architecture designed to achieve end-to-end weakly supervised change decomposition. Unlike traditional single-branch models, Cross-C2PO introduces a cross-comparison mechanism to explicitly model asymmetric temporal differences, ensuring completeness and consistency regardless of input order. The model integrates a differentiable union operator to maintain consistency constraints during weak supervision and employs the proposed Cross-MTF feature fusion function to break commutativity for accurate temporal differentiation. Building on Cross-C2PO outputs, we design a semantic change detection workflow that leverages state-of-the-art segmentation models (e. g., DeepLabV3+) without requiring synthetic datasets. Finally, we introduce an urban renewal dynamic index to quantify facade-level changes and visualize renewal patterns across panoramic and directional views (front, back, left, right) for Guangzhou's central districts (2013–2019), based on 11,431 pairs of Baidu street view panoramas.Result Traditional change detection methods fail to achieve end-to-end change decomposition and rely on complex multi-stage pipelines, limiting scalability and flexibility. In contrast, the proposed Cross-C2PO framework enables one-stage weakly supervised change decomposition and can seamlessly integrate with mainstream architectures, granting them both improved detection accuracy and the ability to perform temporal ownership splitting without additional labels. Experiments on multiple benchmark datasets demonstrate that our method consistently achieves state-of-the-art performance, outperforming existing approaches in both binary change detection and decomposition tasks. Ablation studies further validate the contribution of the cross-branch structure, Cross-MTF fusion, and the differentiable union operator. Applied to Guangzhou street view imagery, the workflow successfully produced urban renewal dynamic maps, revealing high-intensity updates clustered in Liwan and Baiyun industrial areas, while moderate changes dominate residential zones. Directional view analysis additionally highlights local disparities and micro-scale renewal patterns. Conclusion The proposed Cross-C2PO framework offers a simple yet effective solution for weakly supervised semantic change detection, enabling accurate change decomposition without additional synthetic labels. Combined with an interpretable urban renewal dynamic index, it provides a scalable and cost-effective approach for urban facade change analysis. This study bridges street view imagery and AI-based computer vision for urban analytics, offering new insights into spatiotemporal renewal dynamics. Future work will focus on optimizing

computational efficiency and extending the method to multi-source data integration for large-scale applications.

Key words：urban renewal, street view imagery, semantic change detection, scene change detection, weak supervision, dynamic index

Supported by Supported by National Natural Science Foundation of China (No.42071399)