



基于无人机遥感影像的AI解译研究 -若干应用案例

汇报人：唐攀攀

南湖实验室
北京大数据研究中心

2021.11 广州

实验室概况

Overview of Laboratory

南湖实验室成立于2020年6月，地处浙江省嘉兴市美丽的七星湖畔，是嘉兴市为贯彻国家创新驱动发展战略和长三角区域一体化战略而主导创设的重大科技创新平台。实验室以建设世界一流新型科研机构为目标，以开拓高端前沿领域科技“无人区”为追求，以国家重大科技创新战略需求为导向，汇聚海内外顶尖科学家和青年英才，开展前沿技术基础研究和技术创新，攻关前瞻性、颠覆性、引领性重大科学问题，打造新一轮科技革命产业革命策源地和创新高地。



1 研究背景

2 无人机影像AI解译应用案例

3 存在问题及思考



1

研究背景

定义：无人机遥感，也就是我们所说的通过将无人驾驶的飞行器与遥感传感器结合在一起，再加上通讯技术和GPS定位技术，将获取资源和获取信息的过程简单化、智能化，通过获取的信息使现在的工作更简单便捷的技术。

无人机遥感 VS 卫星遥感

◆ 更高的分辨率（空间、光谱）

– 完成更精细的测绘、探测任务（真正取代了很多地面作业模式）

– 更多应用场景

◆ 改变了数据获取方式（从被动到主动）

– 获取时间、地点、波段、分辨率都可以自由选择

– 时效性（应急救援）、满足更丰富多样的需求

◆ 天气条件影响（云）

◆ 成本越来越低，走进了千家万户





遥感AI解译技术 下一个十年变革值得期待

- ◆ 海量数据
- ◆ 传统解译技术速度慢，精度低，始终依赖人工
- ◆ 深度学习神经网络为解决这个问题提供了可能

无人机遥感 VS 卫星遥感

- ◆ 更高的分辨率 (空间、光谱)
 - 意味着更复杂的场景和更大的解译难度
- ◆ 开源的样本更少

2

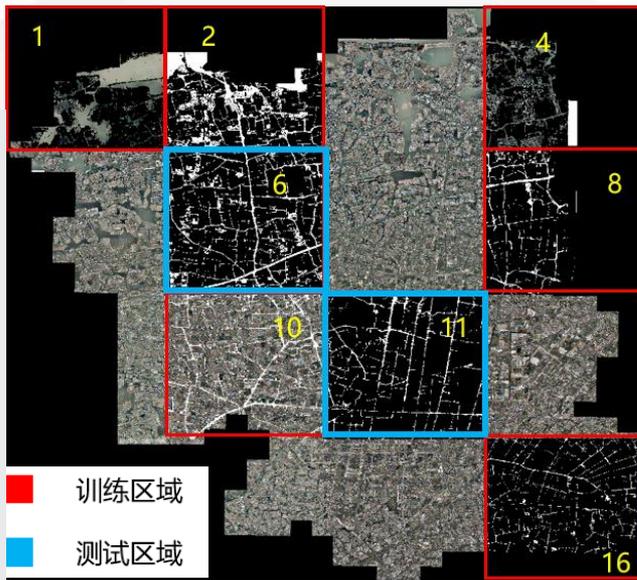
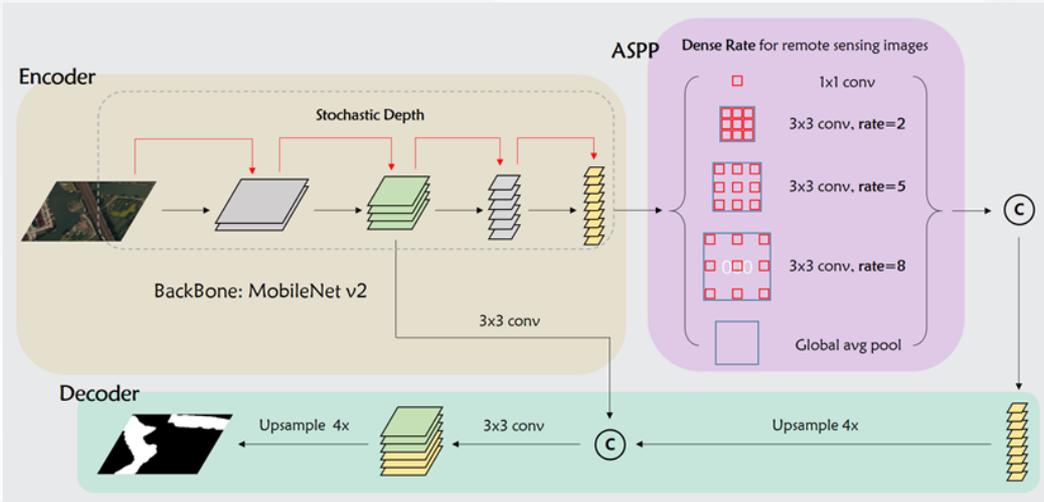
无人机遥感影像AI解译案例



一、长三角地区的水体分割

水体模型

分割模型采用DeepLab v3+架构，其中Backbone网络MobileNet v2，最终训练模型在42m bit左右
 8G GPU：吞吐速度 288 样本/sec，3.99 km2/sec
 能够在18min左右完成整个嘉兴市的水体识别



Recall (召回率) : 0.951
 F1-score (F1-分数) : 0.950

Precision (精度) : 0.950
 IoU (交并比) : 0.906

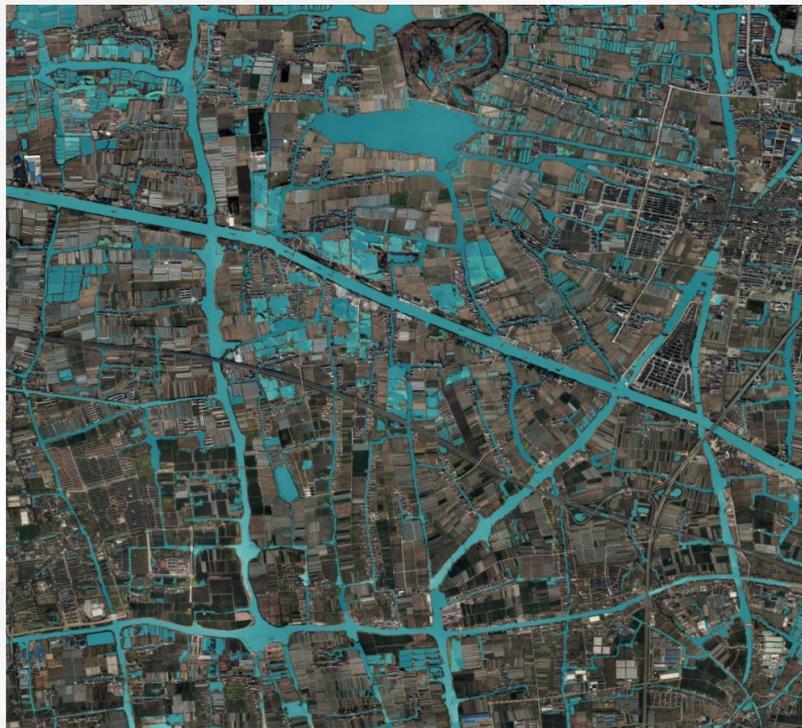
样本数量 (256 x 256, RGB) :
 训练: 10481 测试: 3571

嘉兴地区无人机影像的水体分割

嘉善县

海盐县

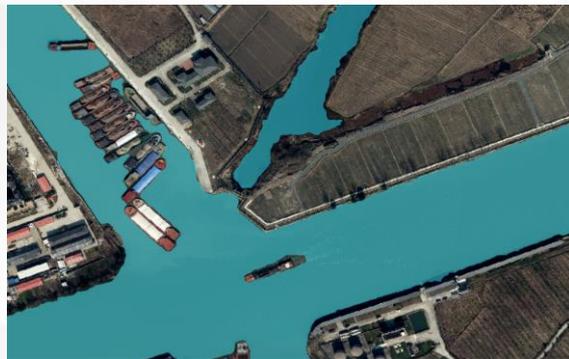
■ 水体



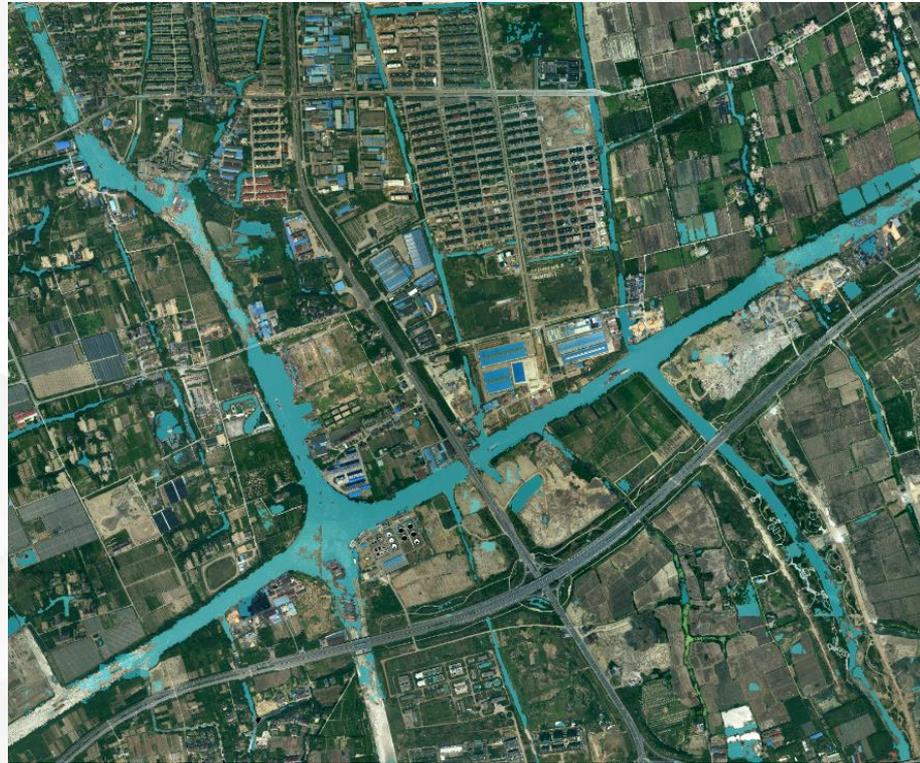
嘉兴地区无人机影像的水体分割

平湖市

■ 水体



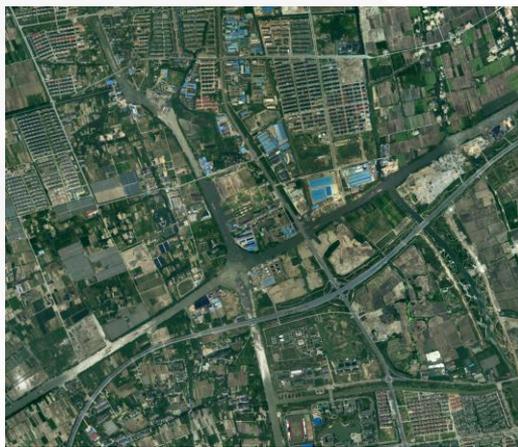
无人机影像：嘉兴



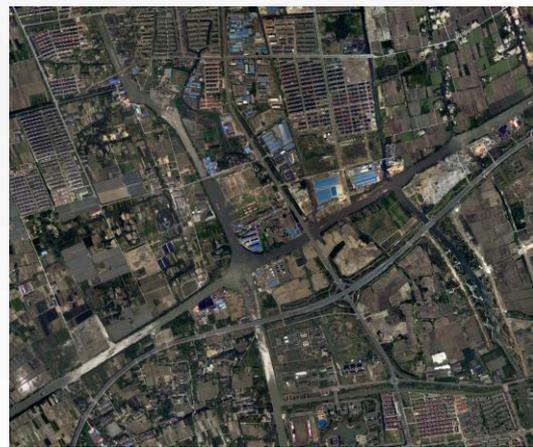
泛化时，如果能保证影像的色调一致就好了



基准图 (训练样本)



预测区域



匀色后

直方图匹配



二、建筑物屋顶提取

建筑物提取算法—数据集统计

1.武大建筑物数据集-Satellite dataset I (204 512*512)

QuickBird, Worldview, IKONOS, 资源三号等光学数据, 分辨率范围0.3m-2.5m, RGB三通道

571 0.3552 (building:9808620 background:27612436)

163 0.3287 (building:2642739 background:8039629)

82 0.3321 (building:1339893 background:4034059)

2.中国典型城市建筑物实例数据集 (5985 500*500)

RGB三通道光学数据, 分辨率为0.29m,数据时间范围: 2017-2019年,地理区域: 北京、上海、深圳及武汉

16758 0.2440 (building:215397904 background:882854384)

4788 0.2496 (building:62672748 background:251113620)

2394 0.2502 (building:31401024 background:125492160)

3.高分辨率可见光图像中建筑物普查数据集(4000 512*512)

高分二号、吉林一号光学数据, 分辨率优于1m, RGB三通道

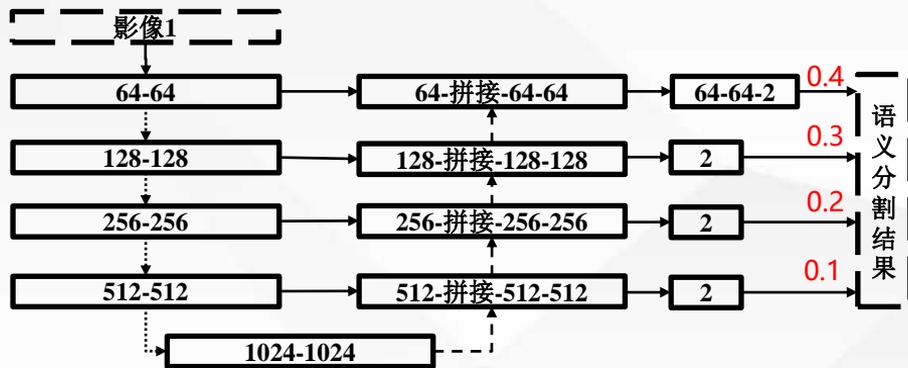
11117 0.09345 (building:62287518 background:666276194)

3176 0.0922 (building: 17566657 background: 190575679)

1589 0.0983 (building: 9317789 background: 94818915)

建筑物模型设计

1. 网络结构设计



↓ 下采样 ↑ 密集上采样 输出通道数 卷积模块(卷积, 标准化, 激活)

基于UNet网络的改进模型

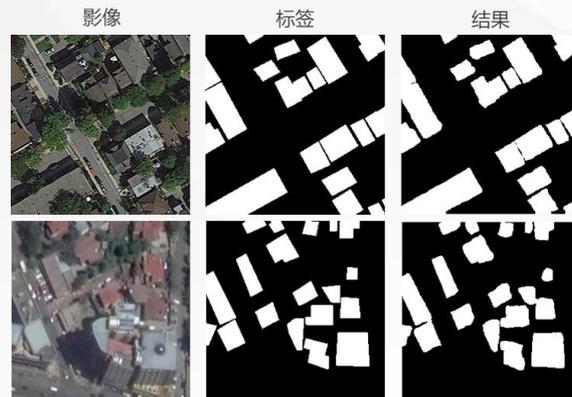
网络优化:

1. 密集上采样 (DUC模块) : 恢复损失的细节信息
2. 多输出支路: 多尺度预测结果更新权重, 强化不同尺度的映射关系
3. 加权交叉熵: 基于数据集统计结果计算类型权重, 减轻样本不平衡; 多输出设置输出权重(红色)。

模型精度

loss	Precision	OA	MIoU	Recall
Dice	0.8137	0.8498	0.6369	0.7610
Cross entropy	0.9195	0.9242	0.7833	0.9128
Cross entropy (Without weights)	0.8698	0.9430	0.8069	0.7629
Dice+cross entropy	0.9143	0.9272	0.7876	0.8956
Dice+focal	0.8724	0.9446	0.8115	0.7669
Cross entropy +focal	0.8891	0.9437	0.8148	0.8093
Dice+cross entropy+focal	0.8685	0.9408	0.8015	0.7629

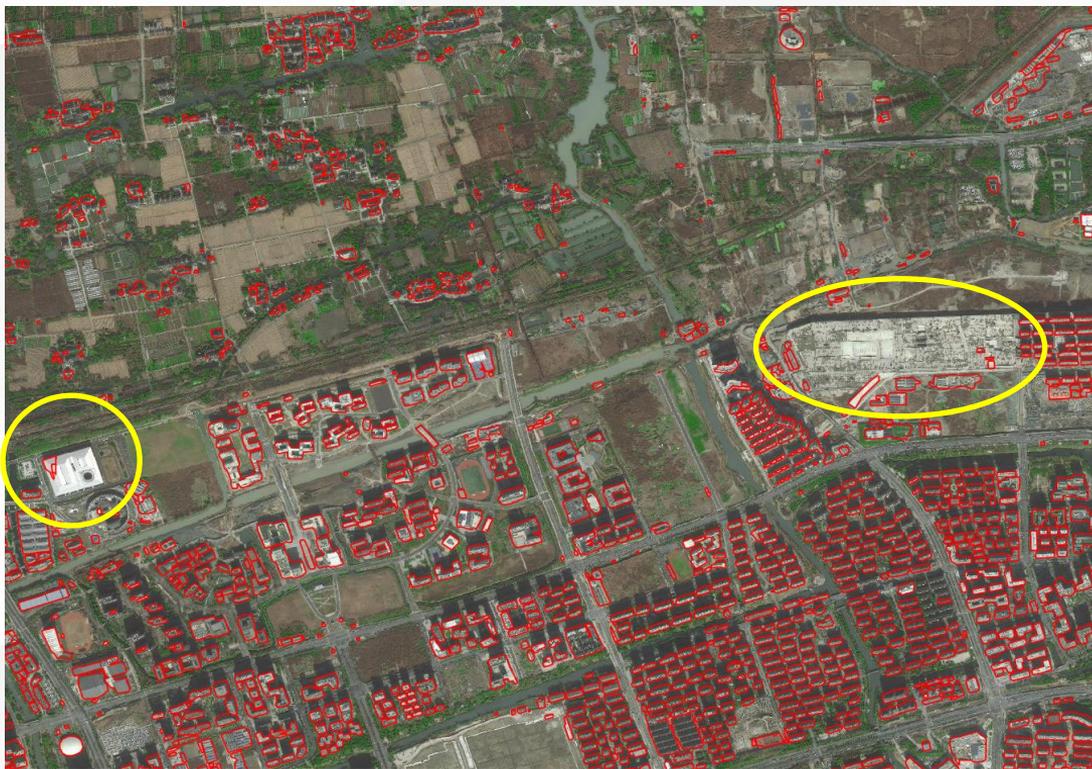
模型样本预测结果



建筑提取部分结果



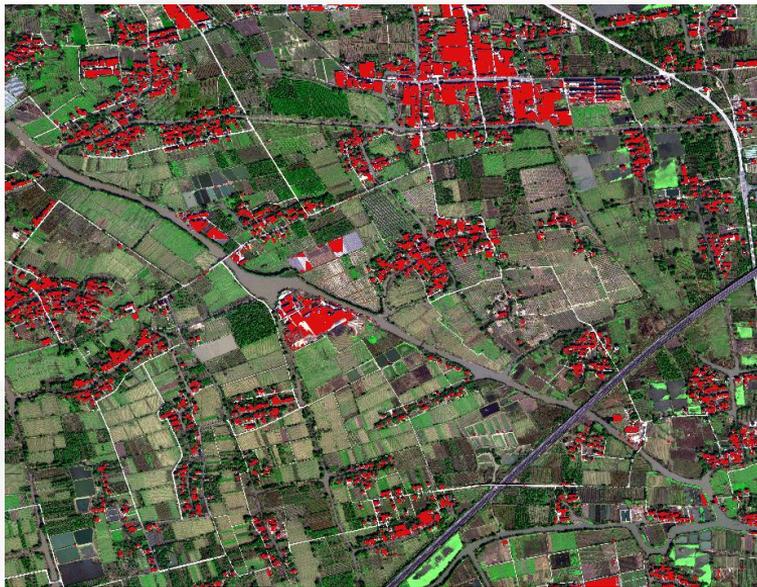
阿里达摩院-建筑提取结果



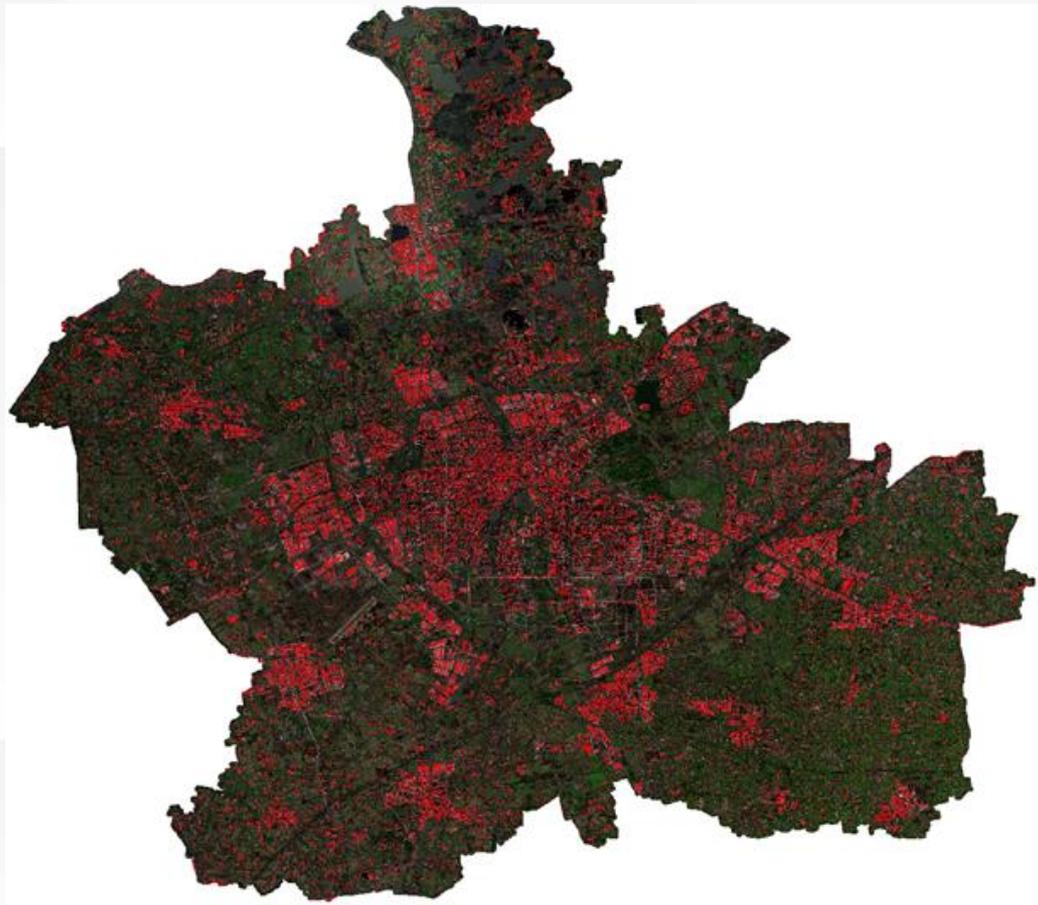
商汤科技-建筑提取结果



嘉兴乡村建筑物提取结果



乡村区域直接迁移预测结果



- 部分道路、广场、河流被错分
- 大型平整的屋顶易被漏检
- 极其复杂建筑区域会漏检

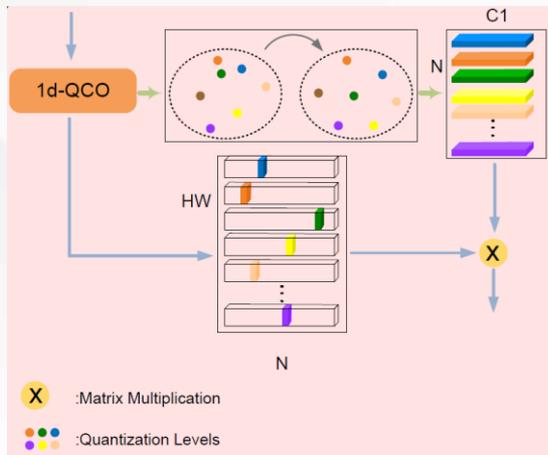
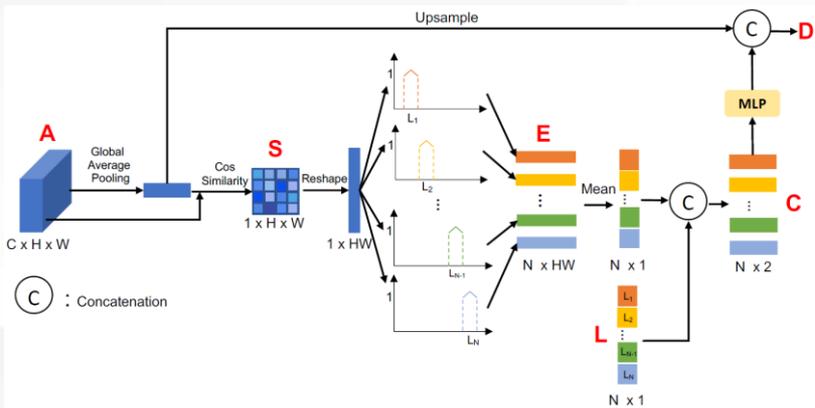
986平方公里面积
1个多小时
普通工作站

网络优化

1d-QCO(Quantization and Counting Operator)+TEM (Texture Enhance Module)

- 基本思想:**1) 利用左图1d-QCO模块达到类似直方图统计的目的，近似得到纹理表征统计结果。
 2) 对每个灰度级进行重建，在对每个级别进行重建的过程中需要感知其他级数的统计信息

FCNs、U-Net变种等各类型语义分割网络单纯的传递浅层信息，并未对纹理特征进行分析，通过添加上述模块测试建筑图提取边界优化效果。

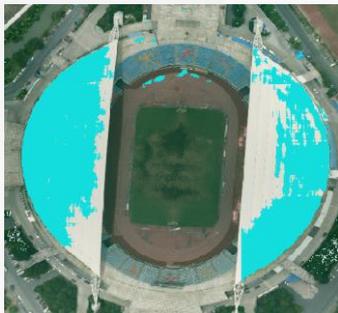


建筑物提取改进结果

训练数据：高分辨率可见光图像中建筑物普查数据集

验证数据：嘉兴航拍影像

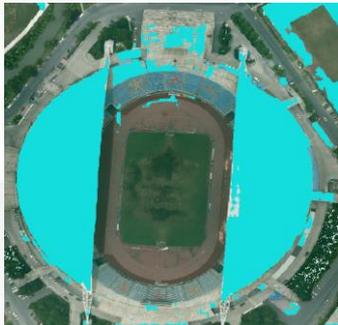
无TEM



优点：显著改善大型建筑的提取效果。

缺点：建筑物会显得无比庞大，密集区域会连成一片。

有TEM





三、变化检测



用途：变化无处不在，无时不刻，最需要高效率的自动算法

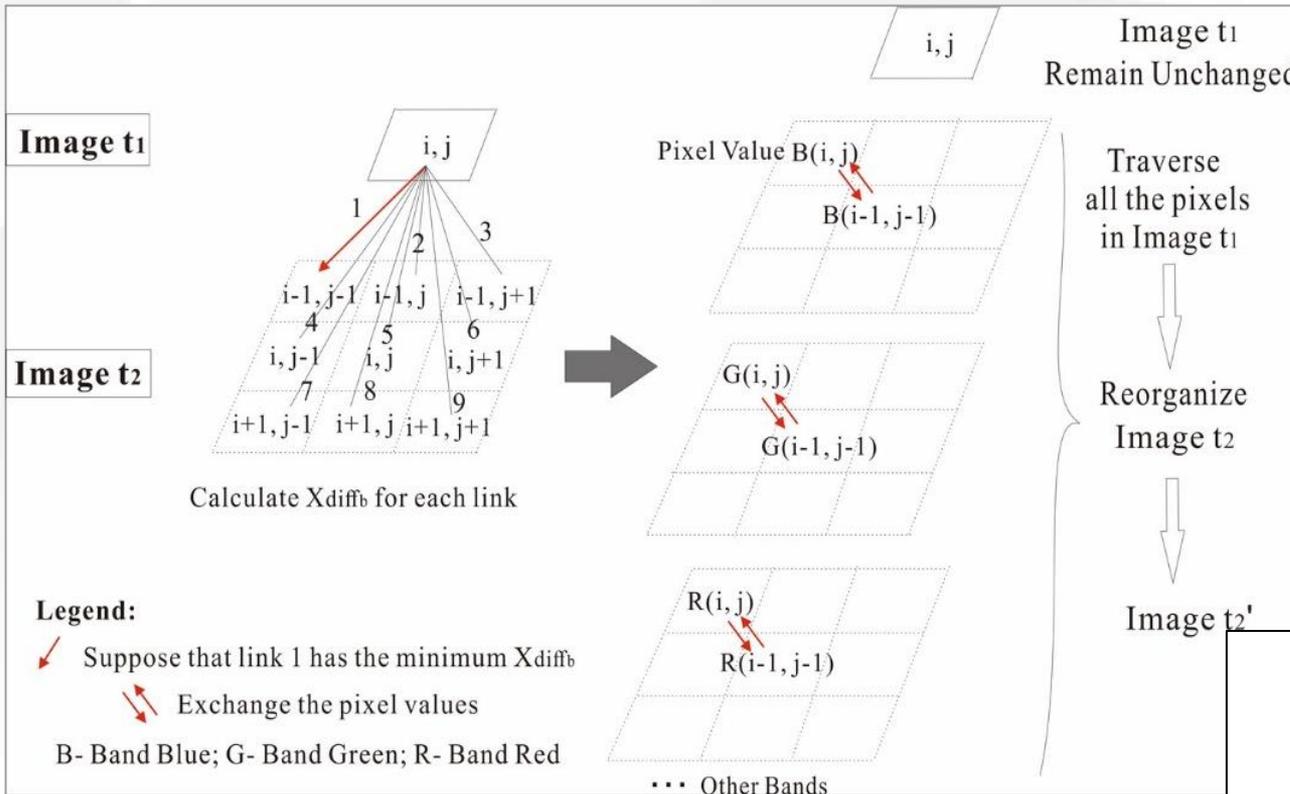
灾害监测、应急、评估
城市变迁、城镇化、基础设施建设监测
自然资源的土地利用变化
生态环境变化
国防军事

难点：变没变难以定义？

诸多因素影响（成像角度和天气差异、信息遮挡、季节变化、配准误差）
场景极其复杂（目视难以判断）
样本稀缺和质量不高

现状：现有的方法（传统+AI），精度低、泛化性差
偏学术、难以推广应用

1, 预处理之自动配准实现



自动配准原理:

不同于传统的 SIFT 和模板匹配算法团队提出一种新的、依赖于像元间相似性的自动配准算法, 其通过重组后时相影像的局部像元顺序来实现配准

相似度计算:

$$x_{diff_a}(j, k) = \min_{(p \in [j-w, j+w], q \in [k-w, k+w])} \left\{ \sqrt{\sum_{i=1}^N (t_2^i(j, k) - t_1^i(p, q))^2} \right\}$$

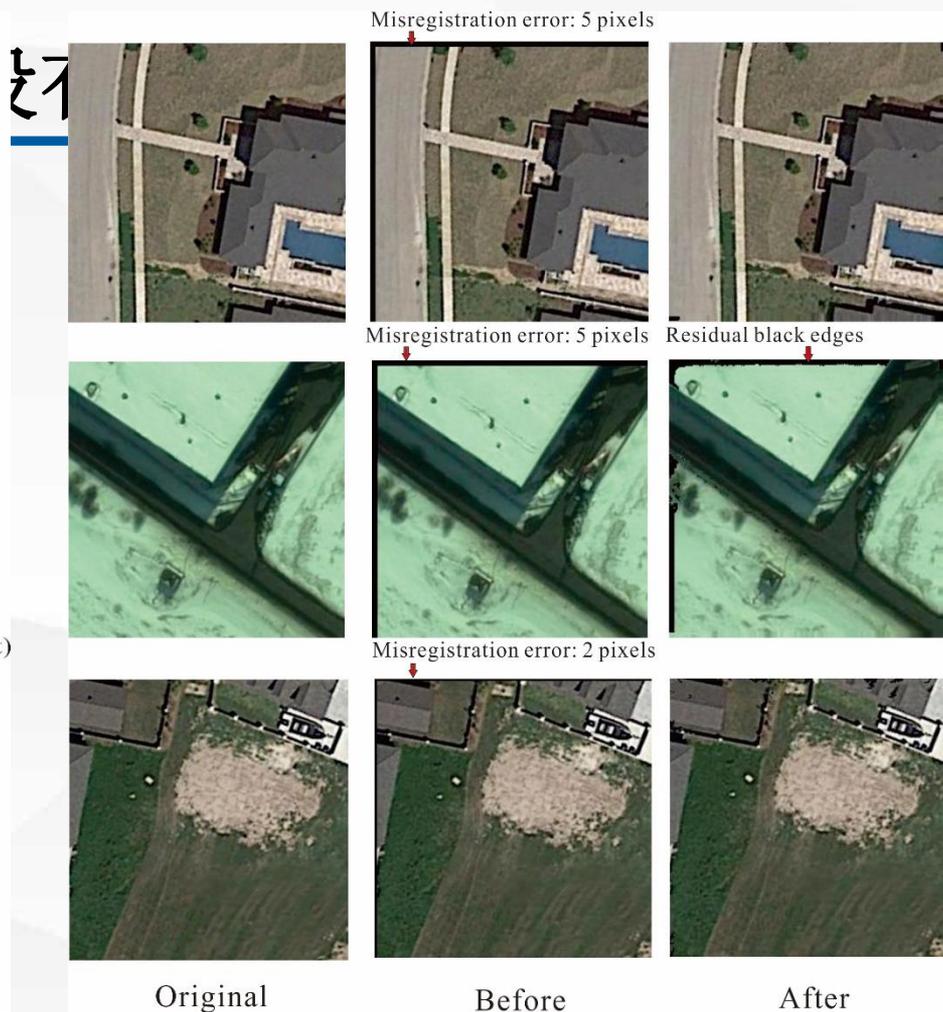
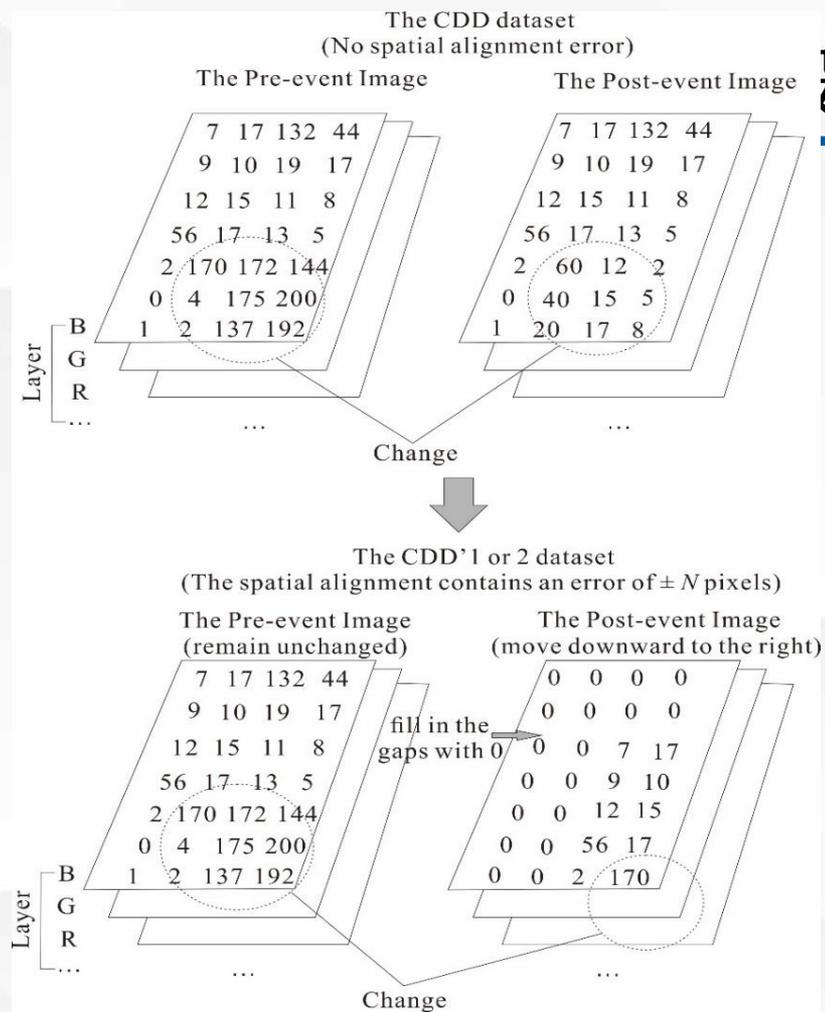
$$x_{diff_b}(j, k) = \min_{(p \in [j-w, j+w], q \in [k-w, k+w])} \left\{ \sqrt{\sum_{i=1}^N (t_1^i(j, k) - t_2^i(p, q))^2} \right\}$$

Where N is the number of image bands.

We combine the two difference images in a second calculation as:

$$M = x_{diff_b}(j, k), \text{ if } x_{diff_a}(j, k) \geq x_{diff_b}(j, k)$$

$$M = x_{diff_a}(j, k), \text{ if } x_{diff_a}(j, k) < x_{diff_b}(j, k)$$



在开源的 CDD 数据集上人工制造配准误差

人工制造的配准误差及自动配准效果

在开源的 CDD 数据集上的对比实验：CDD 即随季节变化的遥感图像数据集，其具有16000个图像集，图像尺寸为256x256像素（10000个训练集以及3000个测试和验证集），空间分辨率为3至100 cm / px。

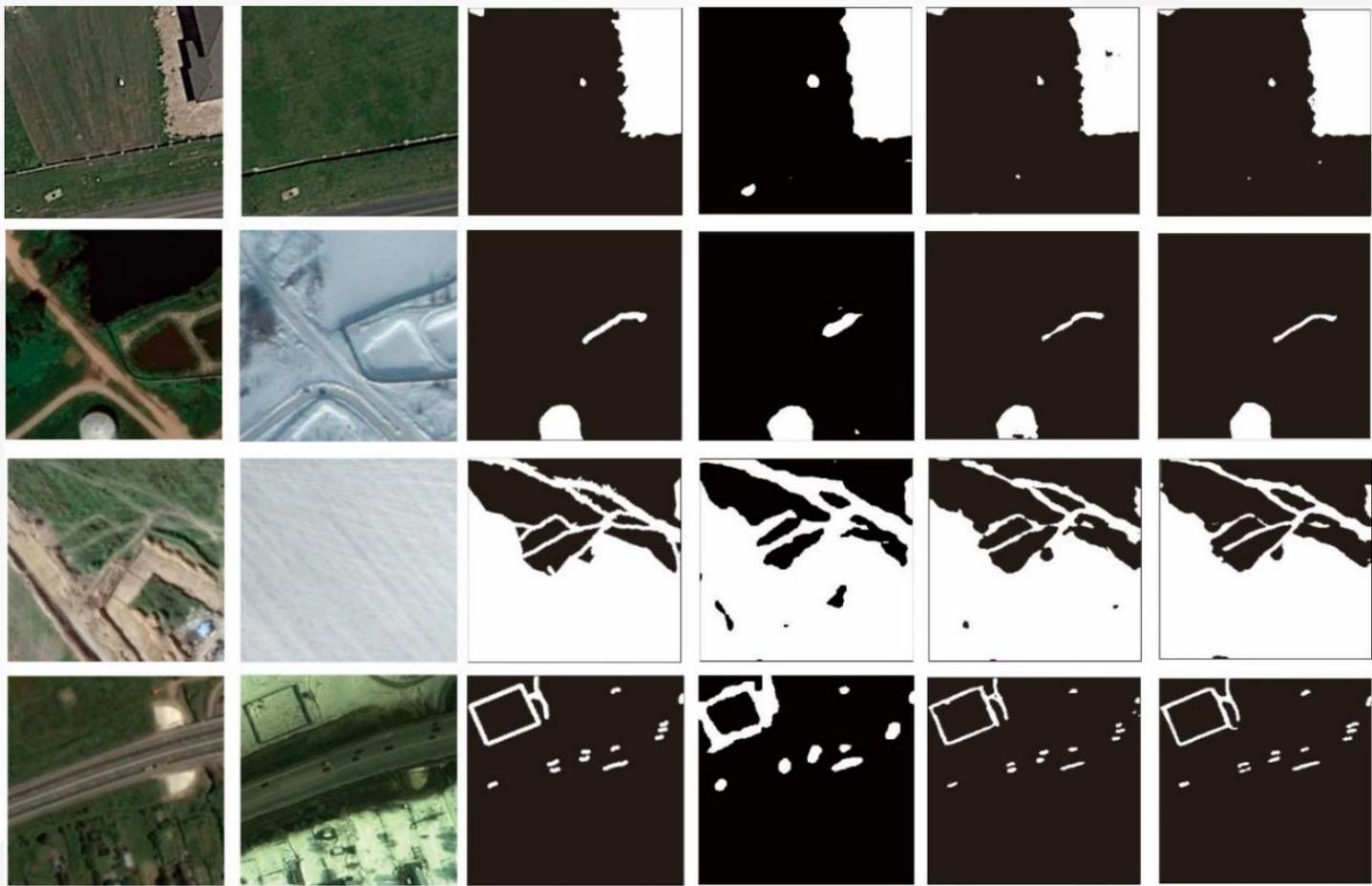
Table 1 Performance comparison on CDD data set

	Method	Parameters(M)	Precision	Recall	F1-Score
Baseline	FC-EF	1.35	0.609	0.583	0.592
	FC-Siam-Conc	1.54	0.709	0.603	0.637
	FC-Siam-Diff	5.39	0.783	0.626	0.692
带空间-通道双注意力机制	DASNet	16.25	0.914	0.925	0.919
带自注意力机制	STANet*	16.93	0.852	0.892	0.872
	UNet++_MSOF	11.00	0.895	0.871	0.876
	IFN	35.72	0.950	0.861	0.903
基于UNet++的变检模型	SNUNet-CD	12.03	0.956	0.949	0.953
消融实验	SiUNet3+-I*	26.99	0.940	0.858	0.897
	SiUNet3+-II*	26.99	0.943	0.857	0.898
	SiUNet3+-III*	26.99	0.948	0.867	0.906
	SiUNet3+-CD*	27.00	0.950	0.870	0.908
	SiUNet3+-CA-CD*	27.00	0.949	0.866	0.906

带通道注意力机制的 SiUNet3+模型



The Second Best



(a)

(b)

真实值

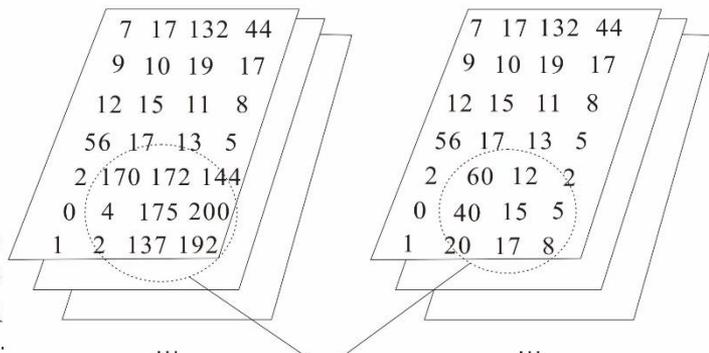
注意力机制

无深监督

最终模型

The BCDD dataset
(No spatial alignment error)

The Pre-event Image The Post-event Image



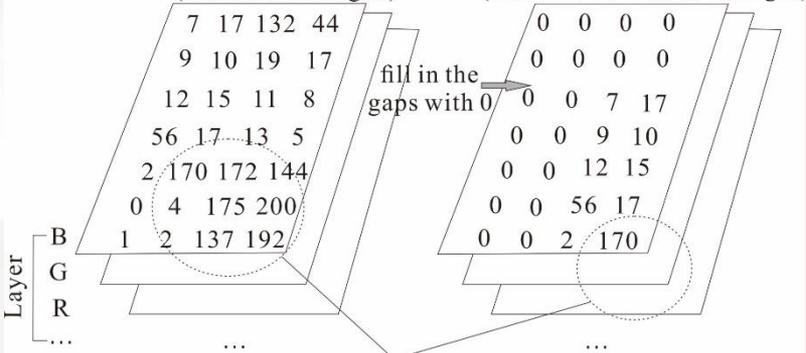
Change



The BCDD' dataset

(The spatial alignment contains an error of ± 2 pixels)

The Pre-event Image (remain unchanged) The Post-event Image (move downward to the right)



Change

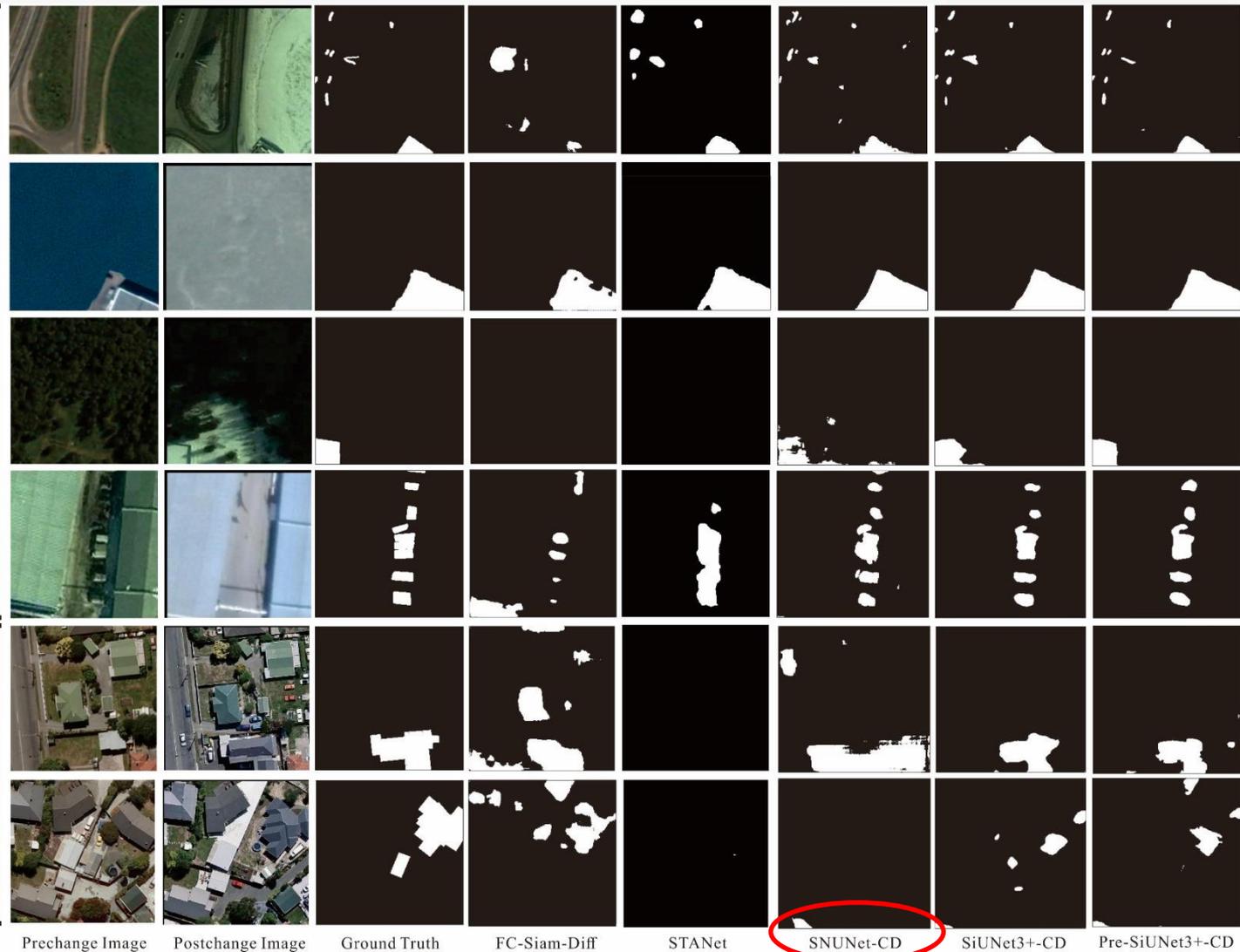
Table 2 Performance comparison on CDD'1 and CDD'2 data sets

Dataset	Method	Precision	Recall	F1-Score
CDD'1 配准误差为 2个像素	FC-Siam-Diff*	0.764	0.583	0.662
	STANet*	0.840	0.871	0.855
	SNUNet-CD*	0.895	0.824	0.858
	SiUNet3+-CD*	0.932	0.863	0.897
	Pre-SiUNet3+-CD*	0.943	0.868	0.904
CDD'2 配准误差为 5个像素	FC-Siam-Diff*	0.694	0.569	0.625
	STANet*	0.842	0.827	0.835
	SNUNet-CD*	0.886	0.810	0.846
	SiUNet3+-CD*	0.911	0.836	0.872
	Pre-SiUNet3+-CD*	0.940	0.868	0.902
BCDD' 配准误差为 5个像素 建筑物变化检测数据集	FC-Siam-Diff*	0.398	0.540	0.453
	STANet*	0.797	0.626	0.696
	SNUNet-CD*	0.783	0.743	0.761
	SiUNet3+-CD*	0.822	0.731	0.772
	Pre-SiUNet3+-CD*	0.853	0.747	0.796

Best results are highlighted in bold. Pre-SiUNet3+-CD denotes the SiUNet3+-CD model with pre-processing. Other notes in this table are consistent with those in Table 1. * means our re-implemented results. To ensure the fairness, we adopt the same batch size (2), epochs (60) and original learning rate (0.001) for all the re-implemented algorithms, and the relevant BN layers are all replaced by GN layers, and ReLU by Leaky ReLU.

CCD'2

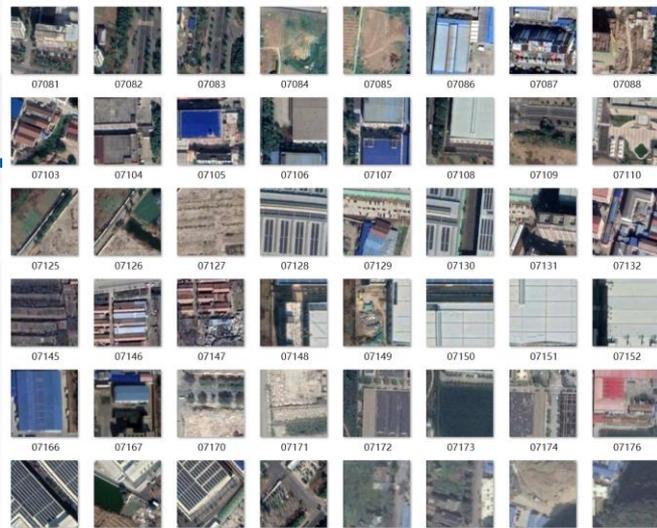
BCCD'



前时相

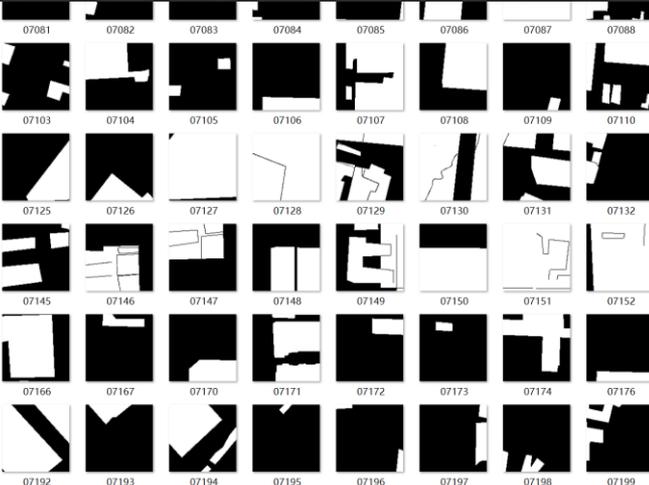


后时相

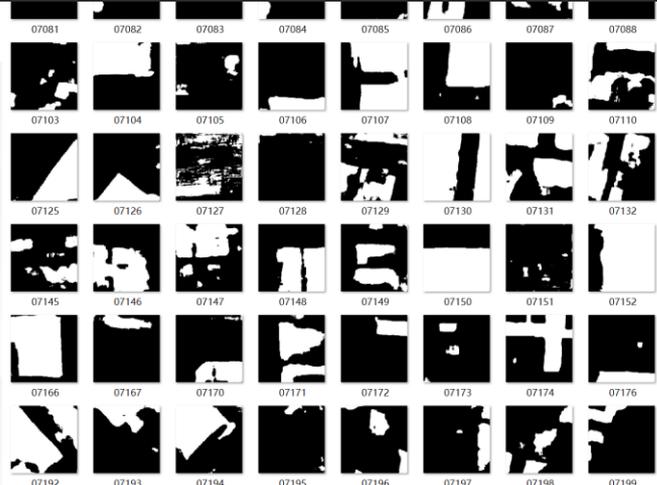


平均的一预测精度、召回率、F1得分、OA、Kappa 系数、IOU 分别为: [0.79505653 0.78257341 0.78873232 0.89573888 0.7193943]

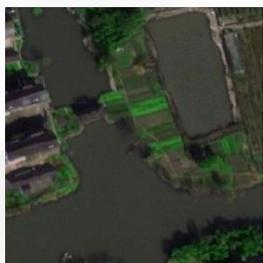
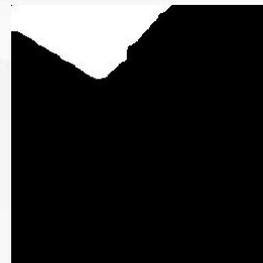
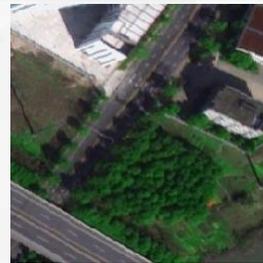
实际值



预测值



无人机影像上的预测结果（嘉兴主城区，0.5m 分辨率）



无人机影像上的预测结果2（嘉兴主城区，0.5m 分辨率）

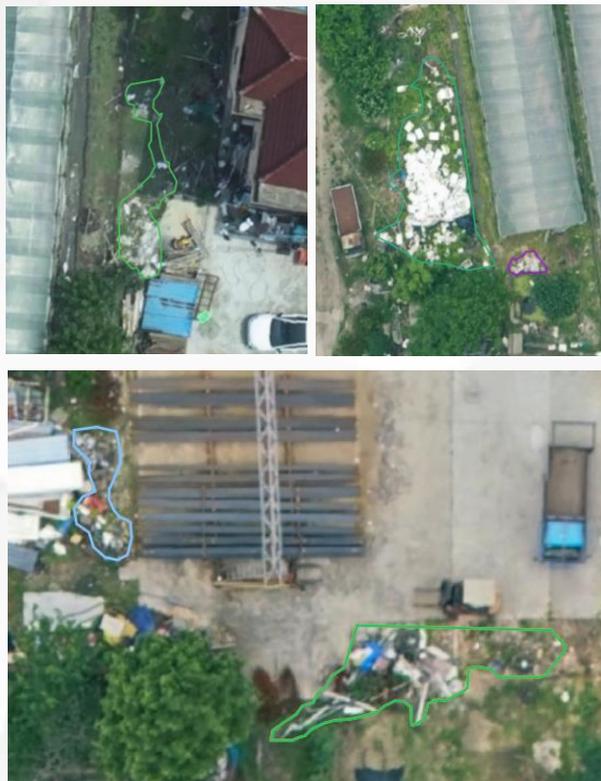




四、其他应用案例



deeplabV3 + 目视解译 + 野外调查



地裂缝

地表裂缝，也称为地裂缝，，被定义为在地面上出现的线性拉伸结构，缝的两侧可能存在一定的垂直偏移

地
裂
缝
灾
害

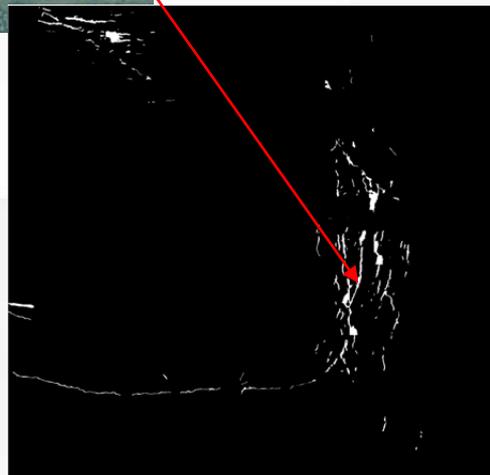
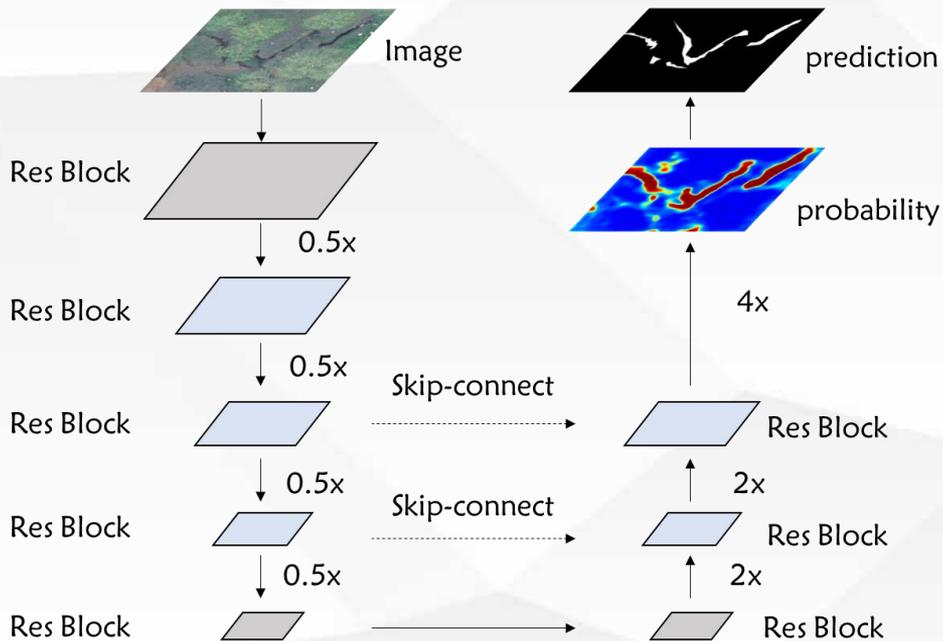
道路故障

房屋破损

地下隧道塌陷

水坝出现裂隙





增加了层与层之间的快速连接
提高特征的复用率



3

存在问题及思考

我们模型的局限性

- 水体模型：区域精度很高（95%左右），全国不同区域泛化性差
- 建筑物模型：精度中等，复杂区域会失效，边界捕捉是研究热点
- 变化检测模型：精度较低，市场上能拿之即用的模型不多

原因

- 样本问题（量少、质量差、样本特征差异）
- 遥感场景非常复杂（算法难度高）
- 技术原因限制（从事人员，监督网络特性）

行业现状 – 科研和实用脱钩严重

- 整体以科研为主、论文里的模型大都不可工程化
- 模型不知道怎么用，没人用（比如道路、建筑）
- 有需求的模型难度大，精度达不到（比如变化检测）
- 数据源的限制

卫星模型直接迁移到无人机上，是非常困难的

基于无人机影像打造智能解译系统，解决实际问题需要，非常有前景。



感谢指导！

