

# 聚“星”成链，筑梦青春

2023 “吉林一号”杯卫星遥感应用青年创新创业大赛

比赛赛题 1：基于高分辨率卫星影像的耕地变化检测

## 技术方案

参赛队伍： Secret;Weapon

队长： 李丹阳

队员一： 李宇轩

指导老师： 李翔

# 目录

1 数据分析 .....	3
1.1 正负样本比例 .....	3
1.2 类别分布 .....	3
2 算法介绍 .....	5
2.1 整体框架 .....	5
2.2 数据增强 .....	5
2.2.1 随机图片旋转 .....	6
2.2.2 随机图片裁剪 .....	6
2.2.3 随机图片翻转 .....	6
2.2.4 随机双时图片交换 .....	7
2.2.5 随机图片色彩亮度、对比度调整 .....	7
2.3 Encoder 模块 .....	7
2.4 特征交换模块 .....	9
2.4.1 空间交换模块 .....	9
2.4.2 通道交换模块 .....	9
2.5 Decoder 模块 .....	10
2.5.1 Mix FFN .....	11
2.5.3 Flow make .....	11
3 实验细节 .....	12
3.1 详细配置 .....	12
3.2 模型复杂度 .....	13
3.3 实验结果 .....	13
4 参考文献 .....	14

# 1 数据分析

## 1.1 正负样本比例

为保证模型泛化性强，我们对所提供的数据集进行了分析，并按照像素级计数统计了正负样本比例，相关数据如表 1 所示：

表 1 数据分析

数据集	数据数量（组）	尺寸	正样本比例	负样本比例
Train	6000	256*256*3（image） 256*256*1（label）	0.28	0.72
test	2000	同上	—	—

## 1.2 类别分布

同时，我们还分别按照像素点类别计数和图片中所包含类别计数两种不同的计数方式，对类别分布进行了统计分析，如表 2、图 1 所示：

表 2 类别分布

Train									
Class	0	1	2	3	4	5	6	7	8
Img_number	5845	1384	4	571	1319	1809	630	241	740
Pixel_ratio(%)	72.05	4.35	0.01	1.57	7.94	7.20	2.03	0.44	4.40

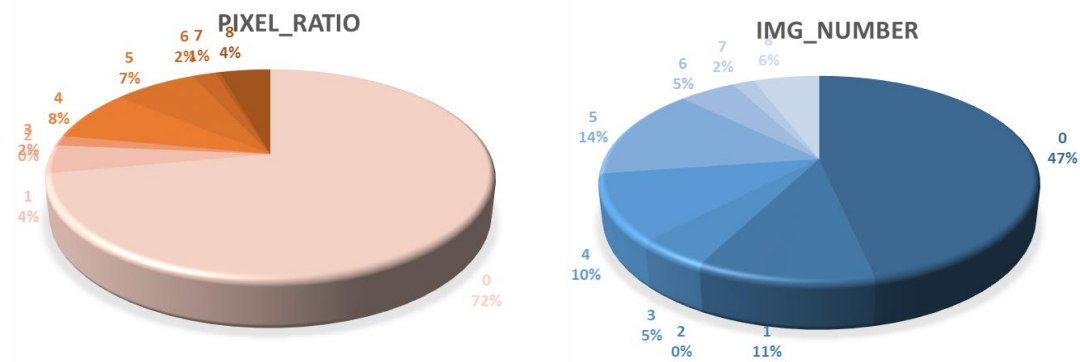


图 1 类别分布（左为像素数量分布，右为图片数量分布）

我们发现，样本类别存在严重的不平衡问题。因此，我们根据样本分布调整各类别的损

失权重。对于样本数量少的类别，我们增加了其损失权重，对于样本数量多的类别，我们降低了其损失权重。但由于盲目增加/减小类别损失权重可能会导致模型不收敛，适得其反。因此，我们设计如下公式，合理计算各类别损失权重。

$$weight = \sqrt{1/Pixel\_ratio}$$

像素类别占比的反比（ $1/Pixel\_ratio$ ）很好的反应了该类别所额外需要收到的关注。但是为了防止权重过高（如第 2 类像素类别的反比为 10000）导致模型不收敛，我们利用 $y = \sqrt{x}$ 单调递增，但导数单调递减的性质，对过大的权重进行抑制。

由此，我们获得了各类别的损失权重，如表 3 所示。

表 3 各类别损失权重

Class	0	1	2	3	4	5	6	7	8
Pixel_ratio(%)	72.05	4.35	0.01	1.57	7.94	7.20	2.03	0.44	4.40
Class_weight	1.18	4.79	100.00	7.98	3.55	3.73	7.02	15.08	4.77

# 2 算法介绍

## 2.1 整体框架

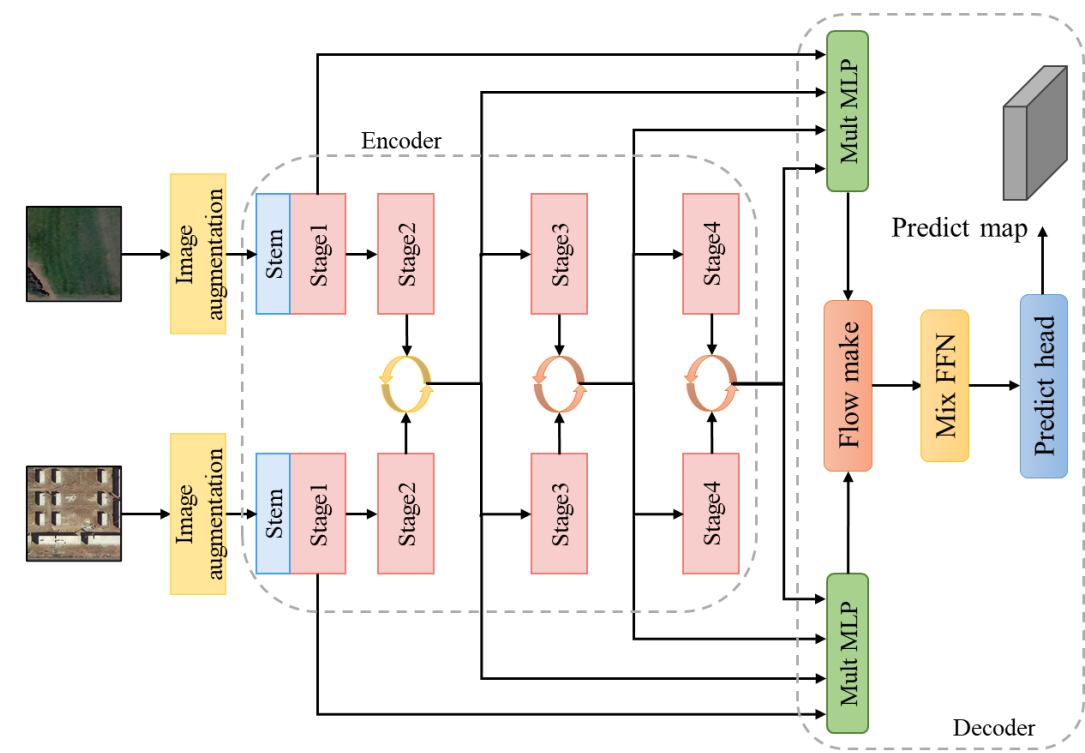


图 2 网络整体框架图

如图 2 所示，本次比赛的网络主要由两部分构成：

1. Encoder：包含了两个权重共享的特征提取网络和一系列的特征交换模块；
2. Decoder：多级 MLP 融合提取到的多尺度特征，后经由 Flow make、Mix FFN 和 Predict head[1]对融合之后的特征进行预测。

## 2.2 数据增强

我们全面采用了多种数据增强策略，旨在提高模型对于各种图像变化的适应能力。这一系列数据增强方案的综合应用，包括随机图片旋转、随机图片裁剪、随机图片翻转、随机双时图片交换以及随机调整图片色彩亮度和对比度等操作，有效地为模型提供更加广泛、多样的训练数据，从而增强其对复杂环境和图像变化的鲁棒性。

### 2.2.1 随机图片旋转

通过引入随机图片旋转，我们使模型能够更好地适应图像在不同方向上的变化，提高其对旋转变换的鲁棒性。并且，随机图片旋转有助于使模型具备方向不变性，即无论图像中的物体是正立还是倒立，模型都能够正确地识别。这对于处理实际场景中各种角度和方向的图像非常重要，尤其是本次比赛使用的是遥感图像，存在着不同的拍摄角度。

我们的随机图片旋转示意图如图 3 所示：



图 3 随机图片旋转示意图（左：未旋转；右：旋转后）

### 2.2.2 随机图片裁剪

随机图片裁剪通过引入不同裁剪尺寸和位置的变化，增加了模型对于局部特征的捕捉能力，提高了其在处理部分遮挡或局部变化的情况下的性能表现。同时，由于随机图片裁剪引入了对图像进行不同位置和尺寸裁剪的随机性，这有助于减轻模型对于训练数据中特定位置和尺寸的过拟合风险，使其在面对未见过的数据时更为鲁棒。

我们的随机图片裁剪示意图如图 4 所示：

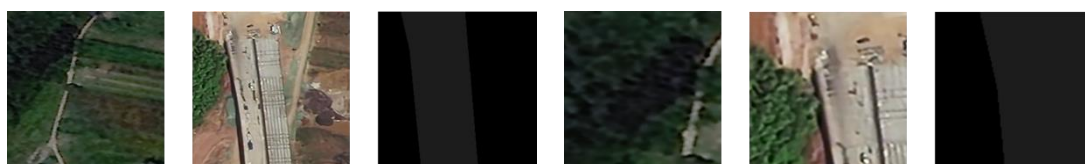


图 4 随机图片裁剪示意图（左：未裁剪；右：裁剪后）

### 2.2.3 随机图片翻转

我们使用了水平翻转和垂直翻转两种翻转方式，图片翻转的使用有助于拓展模型对于镜像变换的适应能力，使其在处理对称性不同的场景时更具优势。

我们的随机图片翻转示意图如图 5 所示：





图 5 随机图片翻转示意图（按顺时针顺序，先后为原图，竖直翻转，水平翻转）

### 2.2.4 随机双时图片交换

双时图片交换可以使模型在训练过程中能够接触到更多的变化模式。这对于变化检测任务来说尤为重要。因为实际场景中的变化可能具有多样的形式，包括光照变化、季节变化、遮挡等，通过引入这样的数据变换，可以提高模型对不同变化的敏感性。示意图如图 6 所示。



图 6 随机双时图片交换示意图

### 2.2.5 随机图片色彩亮度、对比度调整

最后，我们引入了对图片色彩、亮度和对比度的随机调整，以模拟真实世界中光照和色彩的变化。这不仅使模型更具有鲁棒性，同时增加了其在各种环境下的适应性，为比赛中面对不同场景的数据变化提供了更为全面的训练。

我们的随机图片色彩亮度、对比度调整示意图如图 7 所示：



图 7 随机图片色彩亮度、对比度调整示意图

## 2.3 Encoder 模块

Encoder 部分主要用来提取双时图像中所包含的变化地物特征[2]。输入的双时图像共享 encoder 中的参数。同时，我们设置了 stage1-4 四个阶段，从中取出相对于原图像下采样倍数分别为 1/4, 1/8, 1/16 和 1/32 的特征图用来构成多尺度特征。提取多尺度特征可以

使模型对图片中的大小物体均有良好的识别性能。同时，我们在 stage2-4 阶段均设置了也正交换模块，用来捕捉双时图像间存在的差异性。我们使用了 ResNet18[3] 作为我们的 backbone，我们的特征提取网络流程如图 8 所示（取一路为例）。

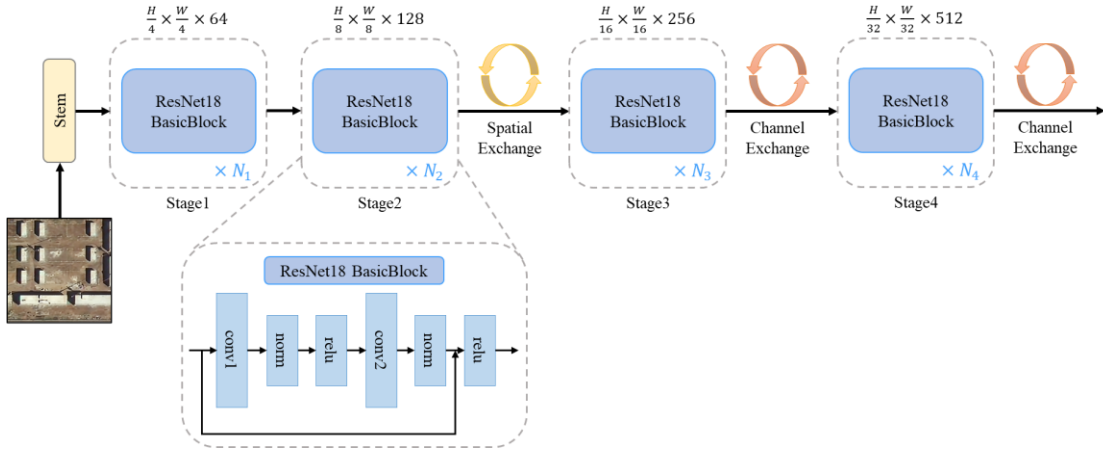


图 8 Encoder 结构图

具体参数如下表所示：

表 4 Encoder 参数表

Encoder			
stage	operation	[out_c,in_c,k_h,k_w]	stride
stem	Conv	[32,3,3,3]	1
		[32,32,3,3]	
		[64,32,3,3]	
stage1	Conv	[64,64,3,3]	1
		[64,64,3,3]	
		[64,64,3,3]	
		[64,64,3,3]	
stage2	Conv	[128,64,3,3]	2
	Residual downsample	[128,128,3,3]	
		[128,64,1,1]	
		Conv	
		[128,128,3,3]	
Feature exchange		spatial exchange	-
stage3	Conv	[256,128,3,3]	2
	Residual downsample	[256,256,3,3]	
		[256,128,1,1]	
		Conv	
		[256,256,3,3]	
Feature exchange		channel exchange	-
stage4	Conv	[512,256,3,3]	2
	Residual downsample	[512,512,3,3]	
		[512,256,1,1]	



	Conv	[512,512,3,3]	
		[512,512,3,3]	
	Feature exchange	channel exchange	-

## 2.4 特征交换模块

特征交换是双时图片变化检测任务中的重要过程[4]。通过特征交换，我们可以捕捉到来自于双时图像中的同源/异质信息。这些信息能够有效地使模型学习到相比于前时相图片，后时相图片究竟发生了哪些变化。具体来说，通过特征交换和之后卷积层的相互学习，模型可以感知双时间特征的上下文信息。通过特征交换和后续的网络层，两个分支的特征之间的分布更加相似，并在一定程度上实现了双时态域之间的自动域自适应。同时，我们的特征交换模块不引入额外的计算量但大幅提高了模型精度。以下是我们模型中所使用的两种特征交换模块：空间交换模块和通道交换模块。

### 2.4.1 空间交换模块

空间交换模块通过设计交换 mask 使双时图像特征图在 W 维度上进行交换，该模块在 stage2 后使用。具体交换示意图如图 9 所示。

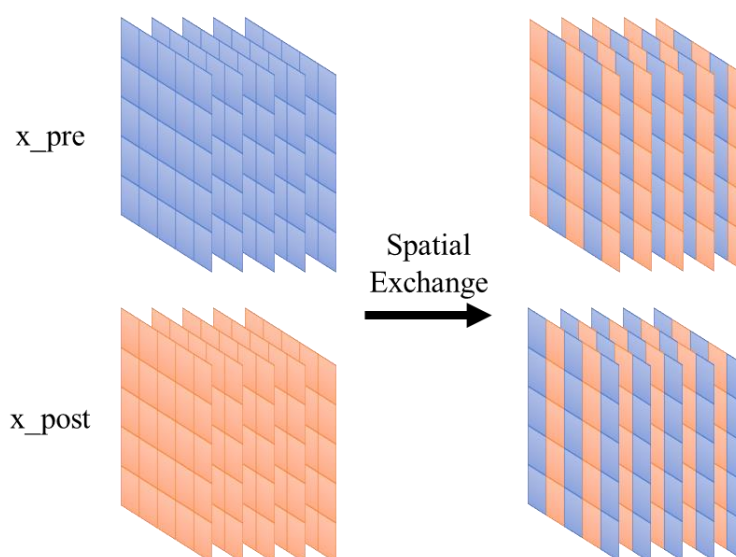


图 9 空间交换模块

### 2.4.2 通道交换模块

同样的，我们使用了类似于空间交换的 mask，但其被用于在 channel 维度上对双时图

像特征图进行交换。具体交换示意图如下。

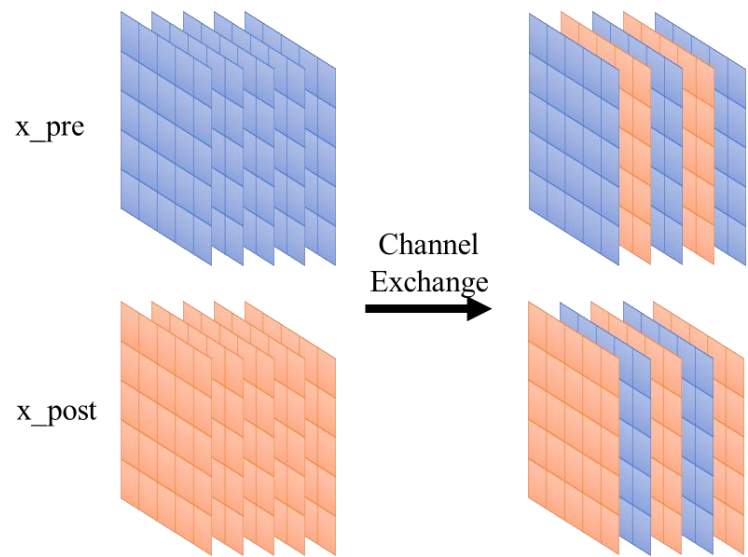


图 10 通道交换模块

## 2.5 Decoder 模块

Decoder 部分在收到来自于 Encoder 4 个 stage 的特征后,将分别使用  $1 \times 1$  卷积做 MLP[5] 操作,四张特征图的通道最后均为 128。接着将四张特征图在 channel 维度进行拼接。此时,特征图已经蕴含了来自不同阶段的丰富信息。为了更好地融合这些信息,我们使用 Fusion conv 模块将四张特征图的信息进行融合。这一步骤旨在保留每个阶段特征的重要信息,并生成更为综合和具有上下文信息的特征表示。Flow make 和 Mix FFN 分别完成了双时图像特征的对准和位置信息嵌入工作。最后经由 predict head 输出最后的预测结果。Decoder 结构图如下。

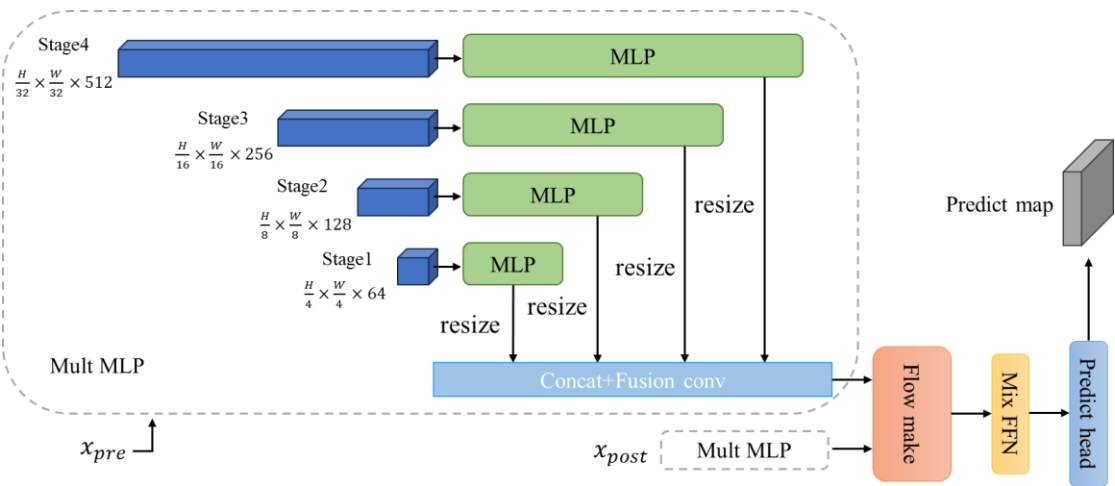


图 11 Decoder 模块结构图

具体参数如下表所示。

表 5 Decoder 参数表

Decoder			
stage	operation	[out_c,in_c,k_h,k_w]	stride
Mult MLP&Fusion	MLP	[128,64,1,1]	1
		[128,128,1,1]	
		[128,256,1,1]	
		[128,512,1,1]	
	Fusion conv	[64,512,1,1]	1
neck layer	Flow make	[128,1,5,5]	1
		[4,128,1,1]	
discriminator	Mix FFN	[128,128,1,1]	1
		[128,1,3,3]	
		[128,128,1,1]	
predict head	cls seg	[9,128,1,1]	1

## 2.5.1 Mix FFN

Mix FFN[6]通过直接在前馈网络（FFN）中使用  $3 \times 3$  卷积来考虑零填充对泄漏位置信息的影响。Mix FFN 可以表示为：

$$x_{out} = MLP(GELU(Conv_{3 \times 3}(MLP(x_{in})))) + x_{in}$$

结构图如下：

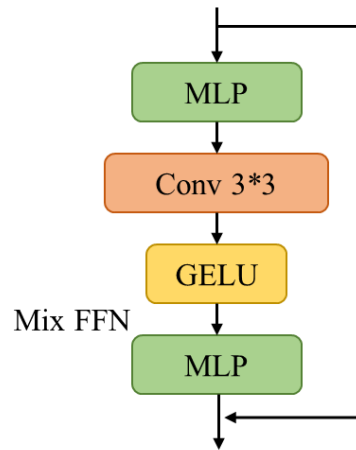


图 12 Mix FFN 结构图

## 2.5.3 Flow make

图像配准是变化检测预处理的重要组成部分。然而，由于拍摄、预处理等原因，训练数

据总存在或多或少的错位或侧视问题。我们通过 Flow make 对双时间特征图进行重新采样，以对它们进行对准矫正[7]。然后，我们分别获取到另一个原始特征图的距离，并将其输入到后续的正向传播中。这样就将对象提取任务被转换为了变化检测任务。整个流程可以通过如下公式表示。

$$x = Concat([x_0(p + \Delta p) - x_1, x_1(p + \Delta p) - x_0])$$

$$\Delta p = FlowNet(Concat([x_0, x_1]))$$

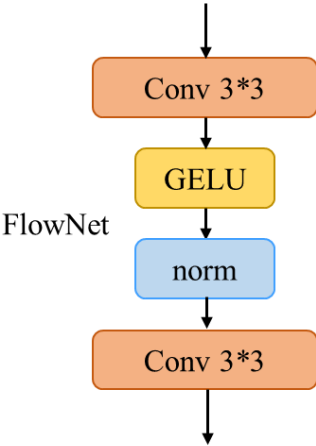


图 13 FlowNet 结构图

## 3 实验细节

### 3.1 详细配置

本次比赛使用 AdamW 优化器，学习率设置为 0.001，权重衰减系数设置为 0.05；采用 Linear 和 Poly 学习率调节策略，Linear 开始于第 1iter，结束于第 1000iter，用于进行初始的学习率 warm-up。PolyLR 开始于第 1001iter，结束于第 40000iter，用于更细致的学习率调整。Iteration 设置为 40000，batch\_size 设置为 32。除此之外，本次比赛使用的软硬件参数如下表所示。

表 6 软硬件参数表

Configuration	Detail
CPU	Intel(R) Xeon(R) Platinum 8358 CPU @ 2.60GHz
RAM	420G
Graphics Card	NVIDIA GeForce RTX 3090 * 9
Operating System	Linux x86_64
CUDA	11.1

3.2 模型复杂度

下表为模型的复杂度分析。参数量和 FLOPs 均由 opencd 的 get\_flops.py 统计得到。显卡为单块 RTX3090。

表 7 模型复杂度

Input shape	FLOPs	Params
(256,256)	5.959G	11.392M

3.3 实验结果

由于比赛数据集中未提供 val 数据集，因此，我们从 6000 组训练集图片中划分出 500 组图片用于 val，剩下的 5500 张图片用于 train。当我们得到 best model 后，保存 checkpoint。然后将 500 组 val 中的图片恢复至 train，在之前保存的 checkpoint 基础上进行微调。实验结果如下表所示（指标在我们划分出的 val 上测得）。

表 8 实验结果（在人为划分的 val 上测试）

	aAcc	mFscore	mPrecision	mRecall	mScore1	mIOU	mAcc
Our_model	96.72	93.37	93.25	93.58	91.81	87.7	93.58

对 test 数据预测结果的部分可视化如下图所示。

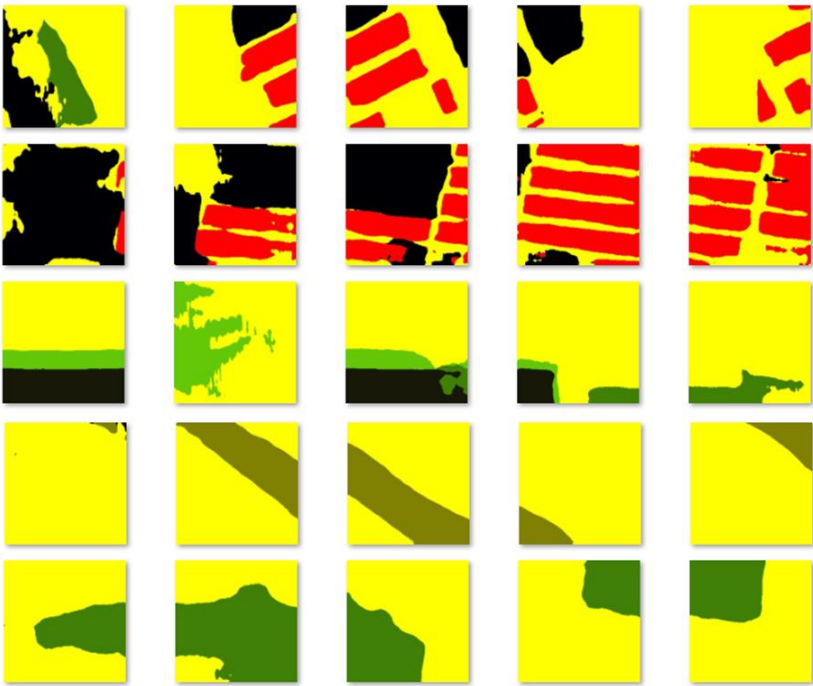


图 14 预测结果可视化

## 4 参考文献

- [1] Fang S, Li K, Li Z. Changer: Feature interaction is what you need for change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023.
- [2] Bandara W G C, Patel V M. A transformer-based siamese network for change detection[C]//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022: 207-210.
- [3] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016: 630-645.
- [4] Zhao S, Zhang X, Xiao P, et al. Exchanging Dual-Encoder–Decoder: A New Strategy for Change Detection With Semantic Guidance and Spatial Localization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-16.
- [5] Lin M, Yang G, Zhang H. Transition is a process: Pair-to-video change detection networks for very high resolution remote sensing images[J]. IEEE Transactions on Image Processing, 2022, 32: 57-71.
- [6] Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
- [7] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.