

通过交叉熵遮罩调整 SAM 以应对遥感变化检测中的类别不平衡

Humza Naveed^{*†}, Xina Zeng^{*†}, Mitch Bryson^{*†}, Nagita Mehrseresht[‡]

^{*}The University of Sydney [†]ARIAM [‡]NearMap

{ hanv0885, xzen3830, mitch.bryson } @uni.sydney.edu.au, nagita.mehrseresht@nearmap.com

Abstract—基础模型在计算机视觉的各个领域取得了显著成功。它们学习到的通用表示可以轻松转移到训练中未见的任务上。其中一个这样的基础模型是“Segment anything model (SAM)”，它可以准确地对图像中的对象进行分割。我们建议通过微调适应 SAM 编码器用于遥感变化检测 (RSCD)，并结合时空特征增强 (STFE) 和多尺度解码器融合 (MSDF)，以在多尺度上稳健地检测变化。此外，我们提出了一种新颖的交叉熵掩码 (CEM) 损失来处理变化检测数据集中严重的类别不平衡。我们的方法在四个变化检测数据集 Levir-CD、WHU-CD、CLCD 和 S2Looking 上的表现优于最新的 (SOTA) 方法。在大型复杂的 S2Looking 数据集上，我们的 F1 得分提高了 2.5 %。代码可在以下网址获取：<https://github.com/humza909/SAM-CEM-CD>

遥感变化检测 (RSCD) 通过识别由卫星或航拍影像在不同时刻拍摄的图像对之间的变化，正确地发现变化在未来预测、环境监测、土地覆盖和土地使用分析、以及灾害管理等应用中至关重要。RSCD 方法已经被广泛研究，从简单的特征差分方法到现代深度学习方法。变更检测的深度学习方法通常利用在相关视觉任务和领域训练的模型的预训练权重来提高性能。例如，预训练的 ResNet 骨干被用于训练双时态图像变换器 (BIT) 进行变化检测，预训练的 VGG 骨干被用于基于变化先验的变化检测 (CGNET)，以及数据高效的变换器 (deit) 被用于 Changevit。尽管这些骨干模型为不同领域训练，但其预学习的特征经过微调后非常适用于遥感变化检测，与从头开始训练这些模型相比，带来了显著的性能提升。最近，通过基础模型实现的性能提升吸引了研究人员将它们改编用于变化检测；例如，remote-clip 的编码器与变化检测架构结合使用以提高变化检测性能。上述大多数预训练模型都是为非密集预测任务而训练的，如图像分类而非图像分割。像素级变化检测是一个密集的分割任务，训练用于相似密集空间任务的编码器所提取的特征预计对于变化检测更为有效。为此，图像分割的基础模型 SAM (Segment Anything Model) [1], [2] 被改编用于变化检测。这项工作冻结了 FastSAM (SAM 的一个较小和更快的变体) [2] 编码器，并从头开始训练一个适配器和解码器，顺序地处理双时态影像并在分类前融合它们。

本文提出了一种新的变化检测架构，也利用了预训练的 FastSAM。我们从 FastSAM [2] 的特征金字塔网络中提取多尺度特征用于双时态影像。我们采用一个空间-时间特征增强 (STFE) 分支来学习每个尺度的跨时间关系。STFE 分支的输出被传递到一个 UNet 风格的解码器 [3]，然后经过多尺度解码器融合 (MSDF) 和一个残差头用于变化图分类。与 [4] 相对，我们架构中的编码器使用基础模型的编码器，以共享编码器/双胞胎风格连接。

我们对 FastSAM 编码器的微调使用在 [5], [6] 中，使用非密集预测模型的编码器的变化检测框架，表现优于其他

方法，这在我们的实验中经验上得以验证。

RSCD 是一个高度类别不平衡的问题，其中变化的像素比非变化像素明显少得多。这种类别不平衡在学习过程中引入了固有偏差，使模型更倾向于提高非变化区域的准确性而非变化区域。为了解决这个问题，我们提出了一种在损失计算中屏蔽非变化像素的新方法，称为交叉熵屏蔽 (CEM)，也称为交叉熵丢弃，如图 ?? 所示。

我们工作的主要贡献是：

- 提出一种基于 SAM 的变化检测架构，通过 STFE 和 MSDF 学习跨时间多尺度特征关系，以实现精确的变化图分类。
- 我们建议结合使用预训练模型中的编码器进行密集图像分割，并对其进行调整以用于变化检测，而不是使用非密集预测模型。
- 我们提出了一种处理变化检测中类别不平衡的新方法，即通过在真实数据中掩盖或去除非变化像素，将其命名为交叉熵掩盖 (CEM)。

本文其余部分的组织如下：第 I 节讨论了所提方法，第 II 节讨论了实验和结果，接下来在第 III 节是总结。

I. 提出的方法

A. 暹罗快速 SAM 编码器

FastSAM [2] 是一个比 segment anything [1] 更小更快的实例分割模型。它仅使用 SAM 数据的 2 % 进行训练，但达到了相等的性能。FastSAM 包含编码器-解码器架构，其中编码器通过特征金字塔网络将输入图像编码为融合多尺度特征的压缩表示。由于密集图像分割训练，编码器学习到的特征表示比仅训练用于分类的 ResNet 或视觉转换器模型更容易转移到变化分割图中，如 [5], [7] 中所述。我们的方法通过细调 FastSAM 编码器，使其适应 Siamese 风格架构（共享权重）中的变化检测。

给定一组用于变化检测的遥感图像 I_1 和 I_2 ，我们通过编码器对图像进行处理。我们从编码器中提取四个尺度 $1/4$ 、 $1/8$ 、 $1/16$ 和 $1/32$ 的四个特征图 F_1 、 F_2 、 F_3 和 F_4 。尺度 $1/4$ 从 FastSAM 的 CNN 主干提取，其余特征图则取自特征金字塔网络。

我们连接时空表示，通过变换处理它们，以学习特征图中的跨时间关系和差异，如图 ?? 所示。我们的时空特征增强 (STFE) 模块由以下公式表示：

$$F_{STFE} = SiLU(BN(Conv2D_{3 \times 3}(F_i^{pre}, F_i^{post}))) \quad (1)$$

其中， F_{STFE} 是增强后的特征图， F_i^{pre} 和 F_i^{post} 分别是变化前和变化后的图像， $Conv2D$ 是 2D 卷积， BN 是批归一化， $SiLU$ 是 Sigmoid 线性单元激活函数。STFE 模块在从编码器提取的每个尺度的特征图上复制四次。增强的表示被转发到解码器以进行变化检测。

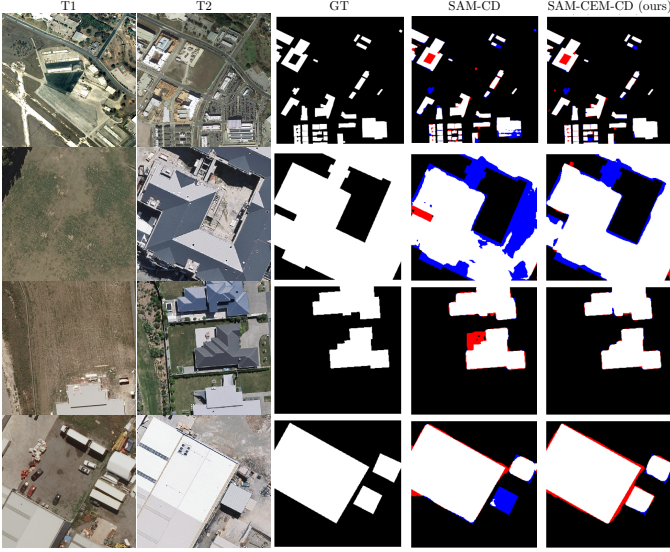


Fig. 1: 变化检测结果的比较。T1 和 T2 是在不同时间戳的输入图像，GT 是真实情况，SAM-CD [4] 和 SAM-CEM-CD (我们的方法) 是模型的预测结果。红色像素是误报，蓝色像素是漏报。

B. Unet 解码器

多尺度变化特征从 STFE 模块用于在类 UNet 解码器中重建变化图 [3]。解码器中的特征图分辨率随着跳跃连接逐渐增加，将来自 STFE 的特征与解码器的输出连接。解码器块表示为：

$$\begin{aligned} F_{up} &= \text{ConvTrans2D}_{2 \times 2}(F_{STFE}) \\ F_{out} &= \text{SiLU}(\text{BN}(\text{Conv2D}_{3 \times 3}(F_{up}, F_{dec}))) \end{aligned} \quad (2)$$

在方程 2 中，来自差分分支 (或 STFE) 的特征图 F_{STFE} 通过转置卷积 F_{up} ，内核大小为 2×2 ，进行上采样以匹配解码器输出的空间分辨率。然后将 F_{up} 与解码器的前一输出连接，以便通过下一个解码器块进行处理，提供关于双时态影像变化的上下文信息。为了保持架构一致性，我们确保解码器块中的卷积、批归一化和激活函数的顺序与在方程 1 中的 FastSAM 编码器和 STFE 模块中的顺序相同。

C. 多尺度解码器融合与变化图

解码器的早期层包含丰富的全局语义信息，而最后的层具有局部变化结构细节。解码器在多空间分辨率下处理这些信息，考虑到不同尺度下时态图像中的对象和变化。我们收集每个解码器块的输出，将其融合，并将其转发给用于变化检测的分类头。我们的分类头包含残差块，随后是如图 ?? 所示的变化图卷积层。

D. 用于类别不平衡的交叉熵掩码

双时相变化检测数据集在变化和未变化区域上存在严重的不平衡。例如，Levir-CD 和 WHU-CD 有 95 % 未变化像素，而只有 5 % 变化像素。由于这种不平衡，变化检测中的交叉熵损失高度倾向于准确分类未变化像素，而非变化像素。为了确保模型同等强调变化像素和未变化像素，我们随机去掉二元交叉熵损失中的未变化像素。为此，我们计算如下提到的每个元素的二元交叉熵 (BCE)：

$$L_{bce} = -y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (3)$$

，其中 y_i 是真实值标签， $\hat{y}_i$ 是预测标签。为了在二元交叉熵中舍弃像素，我们通过保留变化像素的所有损失值并根据舍弃值舍弃未变化像素的值来创建一个掩膜。可以写成如下形式：

$$M_i = \begin{cases} 1 & \text{if } y_i = 1 \\ 1 & \text{if } y_i = 0 \text{ and } (R_i > \delta) \\ 0 & \text{if } y_i = 0 \text{ and } (R_i < \delta) \end{cases} \quad (4)$$

在方程 4 中， R_i 是一个与变化图同维度的矩阵，其值从均匀分布函数 $R_i \sim \text{Uniform}(0, 1)$ 中采样。方程 4 中的 δ 是确定从损失计算中去掉的不变化像素数量的丢弃概率。整体的带丢弃的二元交叉熵损失函数变为：

$$L_{drop} = \frac{\sum_{i=1}^N L_{bce} \cdot M_i}{\sum_{i=1}^N M_i} \quad (5)$$

II. 实验与结果

A. 实现细节

我们使用了 PyTorch 进行模型训练。我们采用随机梯度下降 (SGD) 优化器训练模型 35 个周期。我们使用 FastSAM 编码器并微调其预训练权重。学习率 (lr) 对于 Levir-CD 和 CLCD 为 0.01，对于 WHU-CD 和 S2Looking 为 0.001。我们在每次迭代中衰减学习率 $lr * (1 - \text{iterations}/50)^{\text{decay}}$ ，其中 Levir-CD 和 CLCD 的衰减值为 2.0，WHU-CD 和 S2Looking 的衰减值为 3.0。在训练过程中应用图像翻转和随机裁剪增强，并在推理时翻转图像 4 次以产生稳定的结果。更多细节可以在我们的 GitHub 仓库中查看：<https://github.com/humza909/SAM-CEM-CD>。

B. 变化检测数据集

以下数据集用于我们变化检测框架的训练和测试：

1) *Levir-CD* [16]：该数据集包含 637 对图像，含有 31,333 个标注的变化对象。数据在 2012 年至 2016 年间从谷歌地球中收集，覆盖了德克萨斯州的六个城市。包含 445 张训练图像，64 张验证图像和 128 张测试图像。

2) *WHU-CD* [17]：这是一个航拍图像数据集，拍摄于新西兰克赖斯特彻奇市，记录时期为 2011 年地震后的 2012 至 2016 年。我们使用了 [4] 提供的 6096 张训练图像、762 张验证图像和测试图像的划分。

3) *升力系数* [18]：农作物变化检测数据集是在中国广东省 2017 至 2019 年间获取的。它由高分二号卫星收集的 600 对图像构成，划分为 320 对训练图像，120 对验证和测试图像。

4) *S2Looking* [19]：这是一个大规模变化检测数据集，包含 5000 对卫星图像，来源包括高分、SuperView 和北京二号等。数据集提供了诸如失准和姿态变化等各种挑战。

TABLE I: 不同 RSCD 方法在 Levir-CD 和 WHU-CD 数据集上的定量结果，结果在% 中。最佳结果已被突出显示。

Method	Levir-CD					WHU-CD				
	<i>Pre</i>	<i>Rec</i>	<i>OA</i>	<i>mF₁</i>	<i>mIoU</i>	<i>Pre</i>	<i>Rec</i>	<i>OA</i>	<i>mF₁</i>	<i>mIoU</i>
FC-Siam-diff [8]	92.93	87.02	98.15	89.73	82.67	94.52	89.32	98.70	91.74	85.65
FC-Siam-conc [8]	92.03	89.82	98.28	90.89	84.34	93.94	92.83	98.90	93.37	88.19
SNUNet [9]	93.83	90.11	98.50	91.88	85.84	92.61	83.26	98.10	87.30	79.40
BIT [6]	90.27	83.37	97.60	86.46	78.23	87.92	93.41	98.29	90.46	83.72
ChangeFormer [10]	91.46	86.31	97.95	88.69	81.21	96.28	92.95	99.12	94.55	90.08
CTD-Former [11]	93.60	91.85	98.62	92.71	87.11	96.74	97.03	99.50	96.89	94.11
CGNet [5]	95.95	94.95	99.13	95.44	91.58	98.04	96.14	99.52	97.07	94.44
EATDer [12]	96.29	92.70	98.87	94.41	89.84	95.57	93.06	99.01	94.28	89.64
SAM-CD [4]	95.87	95.14	99.14	95.50	91.68	96.91	95.42	99.36	96.15	92.81
SAM-CEM-CD (ours)	96.12	95.84	99.23	95.98	92.51	97.45	97.19	99.54	97.23	94.73

TABLE II: 不同 RSCD 方法在 CLCD 数据集上的定量结果，结果以% 为单位。

Method	Accuracy Metrics				
	<i>Pre</i>	<i>Rec</i>	<i>OA</i>	<i>mF₁</i>	<i>mIoU</i>
FC-Siam-diff [8]	83.13	74.60	94.73	78.10	68.19
FC-Siam-conc [8]	81.65	79.94	94.84	80.77	70.98
SNUNet [9]	85.92	82.62	95.84	84.19	75.11
BIT [6]	81.42	70.27	94.17	74.39	64.51
ChangeFormer [10]	84.41	81.36	95.47	82.80	73.40
CTD-Former [11]	87.29	83.17	96.12	85.08	76.24
CGNet [5]	85.87	85.30	96.00	85.58	76.83
EATDer [12]	87.16	77.18	93.84	81.16	71.18
SAM-CD [4]	88.25	85.65	96.26	86.89	78.53
SAM-CEM-CD (ours)	88.71	86.89	96.71	87.76	79.77

TABLE III: 针对变更类别，在 S2Looking 数据集上的不同 RSCD 方法的定量结果，结果以% 表示。

Method	Accuracy Metrics			
	<i>Pre</i>	<i>Rec</i>	<i>F₁</i>	<i>IoU</i>
FC-Siam-diff [8]	83.49	32.32	46.60	30.38
FC-Siam-conc [8]	68.27	18.52	13.54	-
SNUNet [9]	45.26	50.60	47.78	31.39
BIT [6]	70.26	56.53	62.65	45.62
ChangeFormer [10]	72.82	56.13	63.39	-
CGNet [5]	70.18	59.38	64.33	47.41
EATDer [12]	65.85	54.74	59.78	42.64
SAM-CD [4]	72.80	58.92	65.13	48.29
SAM-CEM-CD (ours)	73.20	63.02	67.73	51.19

C. 评价指标

为了评估我们模型的性能，我们采用了变化检测中广泛使用的指标，包括精度 (Prec)、召回率 (Rec)、 F_1 分数、整体精度 (OA) 以及交并比 (IoU)。其公式与 [4] 中所示相同。

我们的方法，SAM-CEM-CD，在文献中的其他 SOTA 方法中表现出色。为了确保结果的正确性，我们对每个数据集进行了三次实验并取平均值。表 I、II 和 III 中其他方法的结果来自 [4]。表 I 显示了在 Levir-CD 数据集上的 mF_1 和 $mIoU$ 分别增加了 0.48 % 和 0.83 %。同样地，对于 WHU-CD 数据集，我们实现了 97.23 % 的 mF_1 和 94.73 % 的 $mIoU$ ，相比之下，次佳的 CGNET [5] 提供了 97.07 % 的 mF_1 和 94.44 % 的 $mIoU$ 。我们重新计算了 WHU-CD 的 SAM-CD 结果，如表 I 所示。表 II 显示了 CLCD 相对于其他方法在 mF_1 和 $mIoU$ 大约提高了 1 %，而对于 S2Looking，我们与表 III 中的 SOTA 方法相比，性能提高了 2.5 %。

TABLE IV: 在 Levir-CD 数据集上使用我们架构中不同模块的消融研究，结果在% 中展示。STFE 代表时空特征增强，D 代表 Unet 解码器，MSDF 代表多尺度解码器融合，CEM 代表交叉熵掩码。

Method	<i>mF₁</i>	<i>mIoU</i>
SAM + STFE + D	95.87	92.32
SAM + STFE + D + MSDF	95.92	92.38
SAM + STFE + D + MSDF + CEM	95.98	92.51

TABLE V: 在 Levir-CD 数据集上使用不同掩蔽概率值的消融研究，结果位于%，其中 δ 是掩蔽比率。

Metric	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$	$\delta = 0.5$	$\delta = 0.6$
<i>mF₁</i>	95.93	95.98	95.94	95.86	95.80
<i>mIoU</i>	92.41	92.51	92.44	92.29	92.19

D. 可视化

我们在图 1 中直观比较了 SAM-CD 和 SAM-CEM-CD 的变化图。蓝色预测是 FN，红色预测是 FP。可以看到，图中通过 SAM-CEM-MCD 减少了 FN 和 FP，表明我们的方法相比其他 SOTA 方法的重要性。

E. 消融研究

在本节中，我们研究了不同架构组件对性能的影响。

1) 交叉熵掩码比率：交叉熵掩膜平衡了变化检测数据集中高度不平衡的类别比重的影响。然而，为了达到最佳网络性能，找到一个合理的掩膜比率非常重要。过高的掩膜值可能导致网络学习的信息过少，而过低的值可能导致交叉熵掩膜的表现与交叉熵相等。我们在表 V 中展示了不同掩膜比率的分析，发现掩膜概率 δ 等于 0.3 是最优的。

2) 架构模块：我们比较了架构中各种模块的影响。我们在表 IV 中展示了由于 STFE、MSDF 和 CEM 带来的性能提升，其中包含所有模块的最终架构比其他所有配置表现更好。

3) 类别不平衡损失：我们将交叉熵掩膜 (CEM) 与广泛用于类别不平衡的其他损失进行了比较，包括表 [20] 中的 focal loss、二元交叉熵和 dice 损失的组合 [21]、以及加权二元交叉熵 [21]。与这些相比，CEM 效果最好，而加权 BCE 是第二好的。

4) 预训练模型比较：表 VII 中展示了用于非密集预测任务 [13]–[15] 与密集预测任务 [1], [2] 的预训练模型的对比。我们用各种预训练模型的编码器替换了图 ?? 中的编码器。可以看到，从图像分割训练的模型表现优于其他。

TABLE VI: 使用不同损失函数对 Levir-CD 数据集进行类别不平衡消融研究, 结果在% 中。

Method	mF_1	$mIoU$
BCE ($= 0.7$) + Dice ($= 0.3$)	95.74	92.08
Focal ($\alpha = 0.5$, $\gamma = 2.0$)	95.78	92.16
Weighted-BCE ($w_0 = 0.7$, $w_1 = 1.0$)	95.87	92.31
Weighted-BCE ($w_0 = 0.3$, $w_1 = 0.7$)	95.66	91.95
CEM ($\delta = 0.3$)	95.98	92.51

TABLE VII: 在 Levir-CD 数据集上使用我们的方法对不同预训练模型进行消融研究, 有无 CEM, 结果见于%。

Method	mF_1	$mIoU$	mF_1 (CEM)	$mIoU$ (CEM)
VGG-16 [13]	95.10	91.00	95.33	91.38
ResNet-101 [14]	95.08	90.96	95.31	91.34
ResNet-152 [14]	95.09	90.98	95.39	91.48
WideResNet-101-2 [15]	95.00	90.83	95.01	90.84
Ours	95.92	92.38	95.98	92.51

III. 结论

本文提出采用密集预测 SAM 基础模型进行变化检测。我们提出了空间-时间特征增强 (STFE) 和多尺度解码器融合 (MSDF) 架构模块, 以提高检测性能。我们的新型交叉熵掩码 (CEM) 损失相较于其他为类别不平衡问题建议的损失实现了显著的提升。总体而言, 我们展示了利用预训练模型的编码器进行密集预测任务结合 CEM, 是一种有效的架构设计, 优于现有的最先进方法。

IV.

致谢 此研究得到 ARC 智能机器人系统实时资产管理研究中心 (IH210100030) 的支持。

REFERENCES

- [1] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, “Segment anything,” *arXiv preprint arXiv:2304.02643*, 2023.
- [2] X. Zhao, W. Ding, Y. An, Y. Du, T. Yu, M. Li, M. Tang, and J. Wang, “Fast segment anything,” *arXiv preprint arXiv:2306.12156*, 2023.
- [3] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [4] L. Ding, K. Zhu, D. Peng, H. Tang, K. Yang, and L. Bruzzone, “Adapting segment anything model for change detection in vhr remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, 2024.
- [5] C. Han, C. Wu, H. Guo, M. Hu, J. Li, and H. Chen, “Change guiding network: Incorporating change prior to guide change detection in remote sensing imagery,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.
- [6] H. Chen, Z. Qi, and Z. Shi, “Remote sensing image change detection with transformers,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.
- [7] D. Zhu, X. Huang, H. Huang, Z. Shao, and Q. Cheng, “Changevit: Unleashing plain vision transformers for change detection,” *arXiv preprint arXiv:2406.12847*, 2024.
- [8] R. C. Daudt, B. Le Saux, and A. Boulch, “Fully convolutional siamese networks for change detection,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 4063–4067.
- [9] S. Fang, K. Li, J. Shao, and Z. Li, “Snunet-cd: A densely connected siamese network for change detection of vhr images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
- [10] W. G. C. Bandara and V. M. Patel, “A transformer-based siamese network for change detection,” pp. 207–210, 2022.

- [11] K. Zhang, X. Zhao, F. Zhang, L. Ding, J. Sun, and L. Bruzzone, “Relation changes matter: Cross-temporal difference transformer for change detection in remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [12] J. Ma, J. Duan, X. Tang, X. Zhang, and L. Jiao, “Eatder: Edge-assisted adaptive transformer detector for remote sensing change detection,” *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [15] S. Zagoruyko and N. Komodakis, “Wide residual networks,” *arXiv preprint arXiv:1605.07146*, 2016.
- [16] H. Chen and Z. Shi, “A spatial-temporal attention-based method and a new dataset for remote sensing image change detection,” *Remote sensing*, vol. 12, no. 10, p. 1662, 2020.
- [17] S. Ji, S. Wei, and M. Lu, “Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set,” *IEEE Transactions on geoscience and remote sensing*, vol. 57, no. 1, pp. 574–586, 2018.
- [18] J. Yang and X. Huang, “30 m annual land cover and its dynamics in china from 1990 to 2019,” *Earth System Science Data Discussions*, vol. 2021, pp. 1–29, 2021.
- [19] L. Shen, Y. Lu, H. Chen, H. Wei, D. Xie, J. Yue, R. Chen, S. Lv, and B. Jiang, “S2looking: A satellite side-looking dataset for building change detection,” *Remote Sensing*, vol. 13, no. 24, p. 5094, 2021.
- [20] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [21] S. Jadon, “A survey of loss functions for semantic segmentation,” in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*. IEEE, 2020, pp. 1–7.