# An Ensemble Learning Framework for Anomaly Detection of Important Geographical Entities

Haolin Li [1]*, Jiaojiao Tian [2], Shan Wang [1], Ping Song [3], Yi She [1]

[1] Sichuan Surveying and Mapping Product Quality Test & Control Center, MNR, Chengdu 610041, China – 871342021@qq.com
[2] Remote Sensing Technology Institute, German Aerospace Center (DLR)D-82234 Wessling, Germany – jiaojiao.tian@dlr.de
[3] Sichuan xingshu engineering survey and design group CO.LTD, Chengdu 610041, China – 514078371@qq.com

**Key Words:** Geographical entity, Anomaly detection, Change detection, Ensemble learning

**Abstract**

Due to the complex landforms and the limited resolution of remote sensing imagery, it is difficult to avoid the problem of incorrectly capturing geographical entities, such as buildings. Therefore, anomaly detection of important geographical entities is of great significance to ensure the authenticity and accuracy of geographical entity data. In this paper, we propose an ensemble learning framework for anomaly detection of geographical entity by aggregating the predicted labels generated by multiple deep learning models. In detail, we explore multiple change detection and semantic segmentation model and fully utilize the advantages of various deep learning neural network architectures. The proposed anomaly detection strategy of buildings has been performed on two benchmark datasets, including WHU Building change detection dataset and LEVIR building change detection dataset, the experimental results prove that the proposed method can achieve a more robust and better performance than using single change detection model in terms of quantitative performance and visual performance.

## 1. Introduction

In order to serve natural resource management and support economic development, government agencies worldwide are actively advocating for the creation of 3D real-world representations, with geographical entities being a crucial component of the dataset(Tambassi et al., 2021). Geographic entities can be used in urban planning, environmental monitoring, agricultural survey, disaster assessment and map revision. However, due to complex landforms and the limitations in remote sensing image resolution, the challenge of accurately capturing geographical entities persists. Hence, there arises a critical need for anomaly detection of geographical entities, particularly for urban infrastructure such as buildings(Chen et al., 2023).

Anomaly detection methods based on Convolutional Neural Network (CNN) have achieved great success in industrial sector(Pang et al., 2021). However, anomaly detection of geographic entities mainly relies on manual interpretation, which is inefficient and lacks objectivity. In recent years, some units have started to use change detection or semantic segmentation to detect anomalies in geographic entities (Lei et al, 2019). However, there are some reliability issues, including the low accuracy of labels obtained for change detection, easy omission or misidentifying crucial geographic entities, and the results of semantic segmentation cannot be used to detect updated geographic entities.

In this paper, we explore the performance of fusing two advanced change detection models for building change detection, including BIT(Chen et al., 2021) and P2V(Lin et al., 2023). Moreover, to enhance the integrity and boundary accuracy of buildings, we employ a state-of-art semantic segmentation model for precise delineation of building structure, which is HRNet (Wang et al., 2020). Two fusion techniques are tested, which are Union fusion and Intersect fusion. We assess the proposed methods on two building change detection benchmark dataset, WHU Building change detection dataset(Ji et al., 2019) and LEVIR building change detection dataset (Chen et al., 2020). By comparing the results, it has been illustrated that the fused predictions from two state-of-art change detection models exhiBIT a more robust performance. Additionally, the segmentation of buildings can be used to optimize prediction maps generated by the change detection models.

## 2. Related Work

### 2.1 Change Detection

The change detection of remote sensing imagery mainly uses multi-source images of different time periods to determine the changes of land features, including changes in position and range. Most recent supervised CD methods rely on a CNN-based structure to extract from each temporal image, high-level semantic features that reveal the change of interest, such as Faster R-CNN(Wang et al., 2018), such as STANet (Chen et al., 2022), SNUNet(Fang et al., 2021), CDNET(Yang et al., 2019), P2V(Lin et al., 2022), and FCCDN(Chen et al., 2022).

Moreover, methods anchored on transformers have further accelerated the advancement of this field, which is a new change detection route. which can obtain a more global perspective, such as BIT(Chen et al., 2021), ChangeFormer (Bandara al., 2022). Recently, some articles have begun to incorporate the general knowledge of visual foundational models into the task of change detection, for example TTP(Chen al., 2023).

### 2.2 ANOMALY DETECTION

The anomaly detection methods are used to build a model that distinguishes between ordinary and abnormal classes, and these technologies can be divided into two categories: machine learning based, and non-machine learning based. Lately, the machine learning based techniques are increasingly being used(Peterson et al., 2020), which can be split into three broad

* This work was supported by Sichuan Provincial Bureau of Surveying, Mapping and Geoinformation (Grant No. 2023KJ004 and 2023KJ003)

categories based on the training data function used to build the model, including supervised anomaly detection, Semi-supervised anomaly detection and unsupervised anomaly detection. It should be noted that these technologies are mainly applied in video, hyperspectral imagery etc., and have relatively few applications in geographic entity anomaly detection(Nassif et al., 2021). In details, remote sensing anomaly detection methods have been applied to water quality assessment and objects deviating from the background(Peterson et al., 2020, Li et al., 2023).

## 3. Methods

In this section, we present an ensemble learning framework wherein two fusion approaches are described and employed to combine the prediction results from BIT (Chen et al., 2021), P2V(Lin et al., 2023), HRNet (Wang et al., 2020) models.

The main steps of anomaly detection are as follows: firstly, vectorize the change detection results; secondly, compare the vectorized change detection results with the geographic entities of buildings to identify any missed or incorrectly collected buildings. In addition, We can also use the predicted map of change detection to evaluate the accuracy of entity ID and location ID of geographic entities.

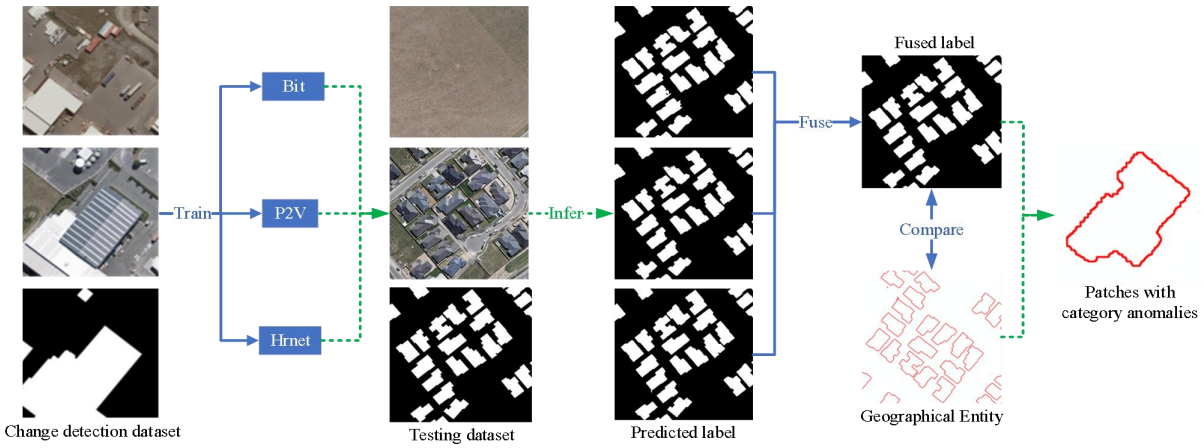



Figure 1. Flowchart of the proposed method

### 3.1 Building change detection

**Bitemporal image transformer (BIT) model** was proposed by Chen et al (2021). Firstly, it extracts high-level semantic features from input image pairs through a CNN backbone network, such as ResNet, and uses spatial attention to transform each temporal feature map into a compact set of semantic tokens. Then, a transformer encoder was used to model the context in two token sets, resulting in context-rich tokens that were re-projected into pixel level space by a connected transformer decoder to enhance the original pixel level features. Finally, feature difference images are calculated from two refined feature maps, and then sent to shallow CNN to generate pixel level change prediction. Notably, BIT-based model has been widely used in detecting changes in buildings and farmland, which can significantly outperform the purely convolutional baseline in terms of efficiency and accuracy(Jia et al., 2021).

**Pair-to-video change detection (P2V-CD) model** proposes a more explicit and sophisticated time modeling method. Firstly, the input image pair is constructed as a pseudo transition video carrying rich temporal information as input to the time encoder, interpreting CD as a problem of video understanding. Secondly, the stitching of dual time images is used as input, using a series of spatial blocks to construct a spatial encoder to capture spatial context that helps locate changing regions. The third is to construct a pseudo video frame sequence to obtain a more detailed temporal data view. Furthermore, the deep supervision technique is applied to accelerate the model training(Lin et al., 2023).

**HRNet** is an earlier semantic image segmentation network structure from Microsoft research (Wang et al., 2020). It enables the high-resolution representations through the interaction of the high-to-low resolution convolution streams in parallel. In particular, it can repeatedly exchange information across high-level and low-level presentations. The benefit is that the resulting representation is semantically richer and spatially more precise, until now it has been used in a wide range of applications, including human pose estimation, semantic segmentation, and object detection. It has also a good performance in building extraction (Seong et al., 2021, Cheng et al., 2020).

### 3.2 Fusion methods

We explore two fusion approaches to produce a final prediction. Each output from the change detection models can be presented as a predicted map, which is a binary map and clearly indicating whether each pixel belongs to the changed building class.

**3.2.1 Union**: In union fusion, we sum up the predicted maps that are generated by BIT and P2V model. It is defined as Equation (1).

$$Y\_Union = Pred_{BIT} \cup Pred_{P2V} \quad (1)$$

Where $Y\_Union$, $Pred_{BIT}$ and $Pred_{P2V}$ denote the fused map, predicted map of BIT model, predicted map of P2V model.

**3.2.2 Intersect**: In the Intersect approach, we firstly union the predicted maps generated by different building change detection, which are generated by BIT and P2V model. Then, we generate a changed building mask by calculating the intersection of Union and the predicted maps generated by semantic segmentation model. It's defined as Equation (2).

$$\text{Y\_Intersect} = \text{Y\_Union} \cap \text{Pred}_{\text{HRNet}} \quad (2)$$

Where $\text{Y\_Intersect}$ and $\text{Pred}_{\text{HRNet}}$ denotes the fused map and predicted map generated by HRNet model.

## 4. Experimental Results

### 4.1 Descriptions of Datasets

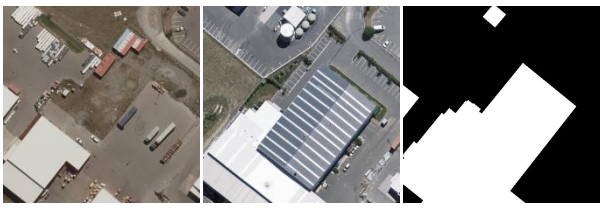
To verify the effectiveness and efficiency of the proposed method, LEVIR building change detection (LEVIR-CD) dataset (Chen et al.2020) and WHU building change detection dataset (Ji et al., 2019) are employed in the experiment.

**LEVIR Building Change Detection (LEVIR-CD) Dataset.** LEVIR-CD consists of 637 very high-resolution(VHR, 0.5m/pixel) Google Earth image patch pairs with a size of 1024 × 1024 pixels. These BITemporal images with time span of 5 to 14 years have significant land-use changes, especially the construction growth. LEVIR-CD covers various types of buildings, such as villa residences, tall apartments, small garages and large warehouses. The fully annotated LEVIR-CD contains a total of 31,333 individual change building instances (Chen et al.2020).

**WHU Building Change Detection Dataset.** The dataset comprises two aerial images with a resolution of 0.2m/pixel, and total image size of 15354×32507 pixels . As ground truth, the dataset provides a change vector, a change raster map, and two corresponding building vectors of these two aerial images (Ji et al., 2019).

Pre-event Post-event Change label

**Figure 2**. Example images of the LEVIR-CD datasets.



Pre-event Post-event Change label

**Figure 3**. Example images of the WHU-CD building datasets.

### 4.2 Experiment setup and training details

To verify the accuracy and reliability of the proposed method based on the LEVIR-CD and WHU-CD dataset, all images are cropped in the 512×512 pixel patches, which results in a total of 2548 tiles for LEVIR-CD dataset and 2046 tiles for WHU-CD dataset. Meanwhile, we use the officially recommended method to divide the dataset into training, testing and validation set, and we compute the mIoU just based on the testing set. It should be noted that buildings with extremely small areas are generally not the focus of quality inspection, and buildings with less than 400 pixels in the dataset are filtered out. The proposed method is implemented under the PaddleRS framework, and all the experiments were conducted on 2 GeForce RTX 4060 GPUs.

### 4.3 Evaluation Method

To evaluate the accuracy of the extracted building segments, three parameters are computed: Mean Intersection-Over Union (mIoU), the total number of omitted buildings and the total number of incorrectly identified buildings. The mIoU is defined as Equation (6) .

$$\text{mIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{\text{TP}}{\text{TP+FP+FN}} \quad (7)$$

Where TP, FP, and FN denote the pixel numbers of True Positives, False Positives, and False Negatives, respectively. K represents the number of categories. Omitted buildings refer to buildings that have undergone changes, but the change detection model is unable to recognize them. Incorrectly identified buildings refer to buildings where the change detection model mistakenly identifies other features as changing buildings. Note that higher mIoU and lower number of omitted or incorrectly identified buildings denote better overall performance.

### 4.4 Experimental Results

The aim of this section is to evaluate the fused building change detection approach by comparing them to single change detection models. It should be noted that the core of the accuracy of building geographic entity anomaly detection lies in the accuracy of change detection, so the experimental section will not repeat the explanation of the accuracy of anomaly detection.

**4.4.1 Comparison with different methods**: Table 1 and Table 2 summarize the mIoU metrics, the number of omitted buildings and the number of incorrectly identified buildings yielded by single change detection models and fusion method, including BIT, P2V, as well as different fusion approaches. The visually comparison of building change detection maps are shown in Figure 4 and Figure 5, respectively. Due to the lack of corresponding semantic segmentation labels for buildings in the LEVIR-CD dataset, semantic segmentation cannot be performed. Therefore, the Intersect method was not tested on the LEVIR-CD dataset.

| Model/Method | LEVIR-CD | WHU-CD |
|---|---|---|
| | mIoU [%] | mIoU[%] |
| BIT | 88.86 | 86.15 |
| P2V | 87.90 | 85.97 |
| Union(BIT,P2V) | 89.28 | 86.84 |
| Intersect(Union(BIT,P2V),HRNet) | / | 89.50 |

Table 1. Summary of the mIoU obtained by different methods

**Case 1: LEVIR-CD**: Firstly, comparing to P2V, BIT has a generally better performance on this dataset. However, combining the predictions generated by BIT and P2V can still further improve the accuracy. As Table 1 shows, the proposed Union approach archives the increase of mIoU by 0.42% and 0.96% comparing with BIT and P2V, respectively. Secondly, the proposed Union approach can reduce the number of omitted buildings from 207 and 301 to 94, while also not increasing the number of buildings that were incorrectly identified.

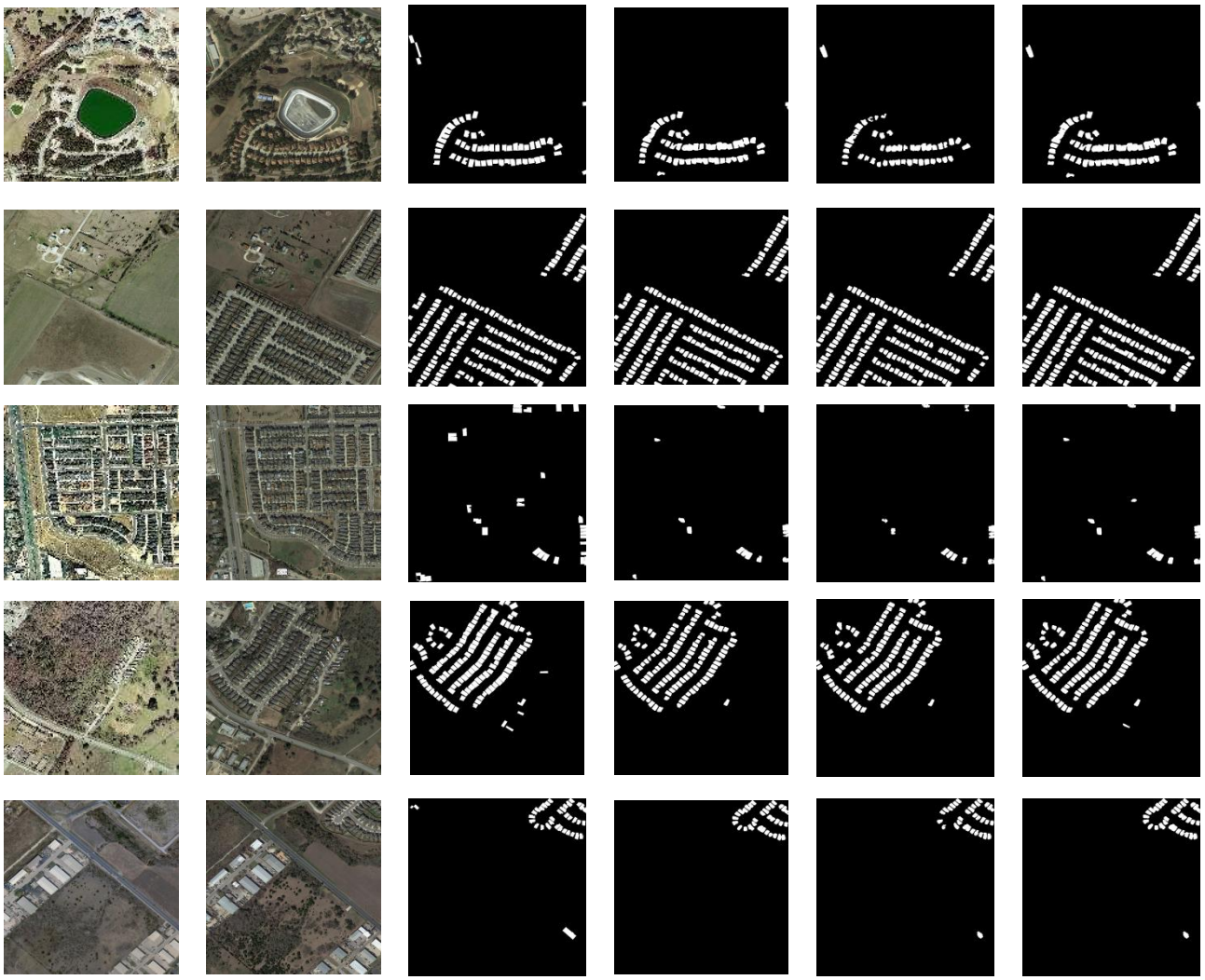
**Case2: WHU-CD:** In this example, comparing to P2V, BIT also has a generally better performance when using WHU-CD dataset. The Union approach still outperforms BIT and P2V with an mIoU gain of 0.69% and 0.87%, combining the

predictions generated by BIT and P2V can still further improve the accuracy. Moreover, the proposed Intersect approach outperforms BIT, P2V and Union method with an mIoU gain of 3.35%, 3.53% and 2.66%.

As Table 2 shows, the proposed Union approach can reduce the number of omitted buildings from 70 and 87 to 55, while slight increasing the number of buildings that were incorrectly identified. The proposed Intersect approach can achieve similar performance. Moreover, it significantly reduces the number of incorrectly identified buildings from 111 to 43 compared to the Union approach.

| Method | LEVIR-CD (Number of buildings) | | | WHU-CD (Number of buildings) | | |
|---|---|---|---|---|---|---|
| | Total | Omitted | Incorrect | Total | Omitted | Incorrect |
| Ground Truth | 6325 | / | / | 680 | / | / |
| BIT | 5270 | 207 | 46 | 742 | 70 | 93 |
| P2V | 5656 | 301 | 87 | 580 | 87 | 52 |
| Union(BIT,P2V) | 5262 | 94 | 66 | 749 | 55 | 111 |
| Intersect(Union(BIT,P2V),HRNet) | / | / | / | 640 | 55 | 43 |

Table 2. Comparison of the accuracy of proposed fusion methods with other methods



Pre-event Post-event Change label BIT P2V Union(BIT,P2V)

Figure.4 Examples of building change detection maps obtained by different methods for the Case1 LEVIR-CD.

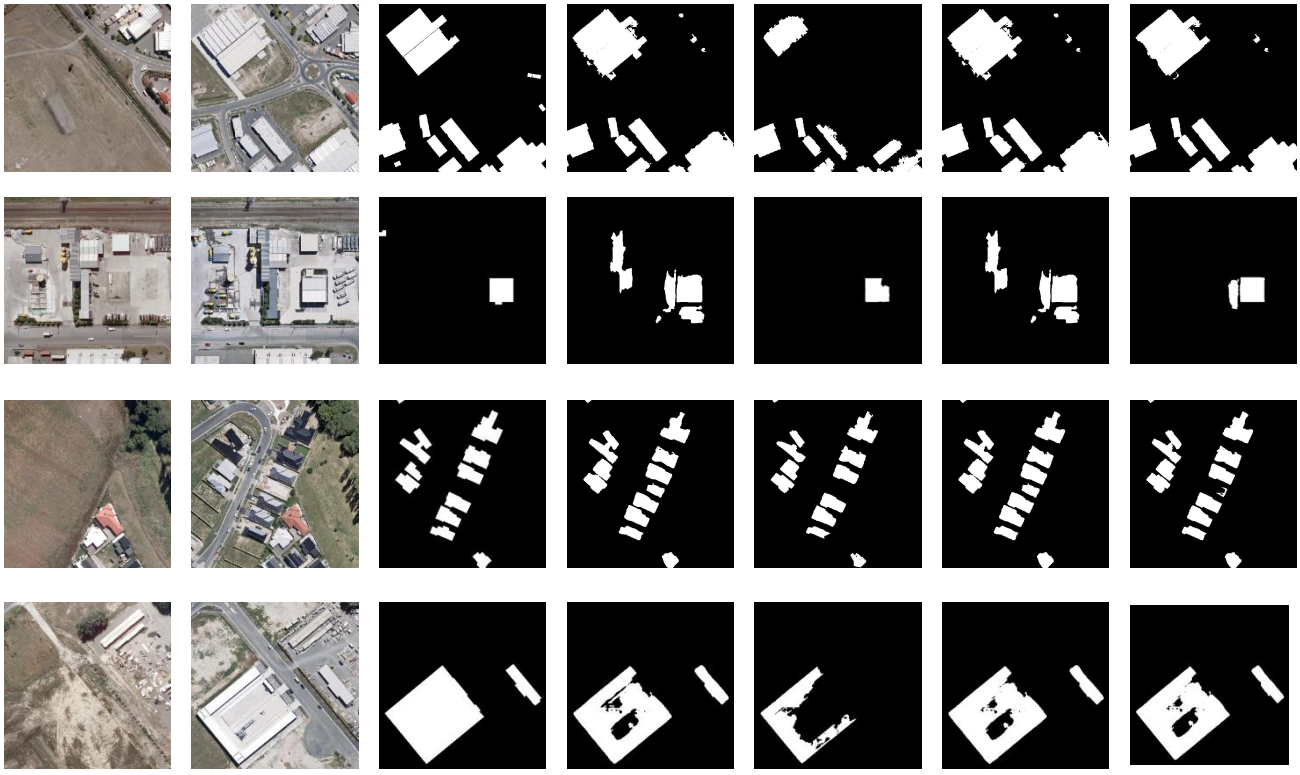



Figure.5 Examples of building change detection maps obtained by different methods for the Case 2 WHU-CD.

**4.4.2 Visual effect comparison:** For the qualitative visual comparison, we have selected five image patches for each case and presented in Figure. 4 and Figure. 5, respectively. The buildings derived by BIT, P2V and the best results from each fusion approach method are presented together with the ground truth. As can be seen, all models perform well on WHU building change detection dataset and LEVIR building change detection dataset. Building change detection results generated by Intersect and Union method outperform BIT and HRNet models in terms of building integrity and edge matching. It should be noted that the two methods proposed in this article have good stability and fewer missing buildings.

As presented in the first row of Figure. 4, the changed buildings obtained by Union method are almost identical to the ground truth, and it missed fewer buildings than the labels predicted by BIT and P2v model. The second row of Fig.4 illustrates that the integrity and edge matching of buildings obtained by Union method are better than other labels. In the third row of Figure.4, the Union method can help reduce the leakage rate of changed buildings. The last two row of Fig.4 clearly demonstrates that the Union method is capable of identifying small sized buildings, and it can help to correct some recognition errors.

In Figure 5, as can be seen, the Union method can fully utilize the advantages of BIT and P2V models. The changed buildings from Union and Intersect approach have better integrity than results from other single models, especially for the first and five examples. The second row of Figure.5 illustrates that the fused label generated by Intersect method has more precise edge than the buildings predicted by BIT, P2V and Union method. At the fourth row, we can observe that the proposed Intersect method can help to correct some recognition errors.

## 5. CONCLUSION

Building geographical entities are important foundational data for economic and social development, and the quality of data is extremely important. Therefore, employing change detection technology to detect anomalies in building geographic entities and improve data quality is a very meaningful attempt. The anomaly detection work of buildings requires high accuracy and stability of change detection results. However, a single change detection model often misses detecting buildings that have undergone significant changes, despite the rapid development of change detection technology. In this paper, we have proposed an ensemble learning framework to combine the prediction of state-of-art change detection and building segmentation. Under this framework two fusion techniques are explored and evaluated on two building change detection benchmark datasets. Our experiments have shown that combing the predicted labels from more change detection models can bring a considerable improvement. As we have used the latest change detection structures, our fusion approach has outperformed individual change detection methods. The Union approach enhances building change detection performance, and helps to correct some recognition errors. The Intersect approach has achieved the highest accuracy compare to other approaches, with relatively few missed or misidentified buildings. Importantly, it is worth noting that additional change detection models or large remote sensing models can also be fused using the method proposed by this article.

## References

Tambassi, T., 2021. The philosophy of geo-ontologies: applied ontology of geography. Cham: Springer.

Chen, R., Lei, J., Yao, H., Li, T. and Li, S., 2023. Anchor-Enhanced Geographical Entity Representation Learning. IEEE Transactions on Neural Networks and Learning Systems.

Pang, G., Shen, C., Cao, L. and Hengel, A.V.D., 2021. Deep learning for anomaly detection: A review. ACM computing surveys (CSUR), 54(2), pp.1-38.

Lei, T., Zhang, Y., Lv, Z., Li, S., Liu, S. and Nandi, A.K., 2019. Landslide inventory map** from BITemporal images using deep convolutional neural networks. IEEE Geoscience and Remote Sensing Letters, 16(6), pp.982-986.

Chen, H., Qi, Z. and Shi, Z., 2021. Remote sensing image change detection with transformers. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-14.

Lin, M., Yang, G. and Zhang, H., 2022. Transition is a process: Pair-to-video change detection networks for very high resolution remote sensing images. IEEE Transactions on Image Processing, 32, pp.57-71.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X. and Liu, W., 2020. Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10), pp.3349-3364.

Ji, S., Wei, S. and Lu, M., 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Transactions on geoscience and remote sensing, 57(1), pp.574-586.

Chen, H. and Shi, Z., 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sensing, 12(10), p.1662.

Wang, Q., Zhang, X., Chen, G., Dai, F., Gong, Y. and Zhu, K., 2018. Change detection based on Faster R-CNN for high-resolution remote sensing images. Remote sensing letters, 9(10), pp.923-932.

Chen, H. and Shi, Z., 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sensing, 12(10), p.1662.

Fang, S., Li, K., Shao, J. and Li, Z., 2021. SNUNet-CD: A densely connected Siamese network for change detection of VHR images. IEEE Geoscience and Remote Sensing Letters, 19, pp.1-5.

Yang, J., Guo, J., Yue, H., Liu, Z., Hu, H. and Li, K., 2019. CDnet: CNN-based cloud detection for remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 57(8), pp.6195-6211.

Chen, P., Zhang, B., Hong, D., Chen, Z., Yang, X. and Li, B., 2022. FCCDN: Feature constraint network for VHR image change detection. ISPRS Journal of Photogrammetry and Remote Sensing, 187, pp.101-119.

Bandara, W.G.C. and Patel, V.M., 2022, July. A transformer-based siamese network for change detection. In IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium (pp. 207-210). IEEE.

Chen, K., Liu, C., Li, W., Liu, Z., Chen, H., Zhang, H., Zou, Z. and Shi, Z., 2023. Time Travelling Pixels: BITemporal Features Integration with Foundation Model for Remote Sensing Image Change Detection. ar**v preprint ar**v:2312.16202.

Peterson, K.T., Sagan, V. and Sloan, J.J., 2020. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. GIScience & Remote Sensing, 57(4), pp.510-525.

Nassif, A.B., Talib, M.A., Nasir, Q. and Dakalbab, F.M., 2021. Machine learning for anomaly detection: A systematic review. Ieee Access, 9, pp.78658-78700.

Li, J., Wang, X., Zhao, H., Zhang, L. and Zhong, Y., 2023. A Unified Remote Sensing Anomaly Detector Across Modalities and Scenes via Deviation Relationship Learning. ar**v preprint ar**v:2310.07511.

Jia, B., Cheng, Z., Wang, C., Zhao, J. and An, N., 2023. CA-BIT: A Change Detection Method of Land Use in Natural Reserves. Agronomy, 13(3), p.635.

Seong, S. and Choi, J., 2021. Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates. Remote Sensing, 13(16), p.3087.

Cheng, Z. and Fu, D., 2020, September. Remote sensing image segmentation method based on HRNET. In IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium (pp. 6750-6753). IEEE.