

# UGRoadUpd: An Unchanged-Guided Historical Road Database Updating Framework Based on Bi-Temporal Remote Sensing Images

Mingting Zhou<sup>1</sup>, Haigang Sui, Shanxiong Chen<sup>2</sup>, Xu Chen<sup>3</sup>, Wenqing Wang, Jianxun Wang, and Junyi Liu

**Abstract**—Timely updated road networks are the basis for many real-world applications such as intelligent navigation and traffic management. Existing road updating methods based on remote sensing images learn from historical road databases to update roads. Road extraction models learned from historical images however, are not easily applied to a current image due to spectral differences; and only changed roads need updating. In this paper, an Unchanged-Guided Road Updating (UGRoadUpd) framework is proposed to improve the quality of updated road networks by limiting the road updating range and learning from historical unchanged roads. The UGRoadUpd framework identifies road changes using a novel dual-task dominant-transformer-based neural network for road change detection (DT-RoadCDNet). DT-RoadCDNet executes road segmentation and change detection simultaneously, from bi-temporal remote sensing images. The Dominant-Transformer based Global Context Modeling module in DT-RoadCDNet globally models the contextual spatial structure for improved integrity in roads and road changes. Based on the discovery of road changes, an unchanged-guided road update strategy updates the roads in changed areas by learning from the prior information provided by unchanged roads in a historical road database. Experiments on two newly annotated road change detection and update datasets confirms the effectiveness of our UGRoadUpd framework.

**Index Terms**—Road network change detection, road network update, dominant-transformer, road change detection datasets.

## I. INTRODUCTION

**A**N UPDATED high-precision road network has applications in intelligent navigation, traffic management, urban

Manuscript received 12 October 2021; revised 16 April 2022; accepted 2 June 2022. Date of publication 4 July 2022; date of current version 7 November 2022. This work was supported in part by the National Natural Science Foundation of China Major Program under Grant 42192580 and Grant 42192583, in part by the National Natural Science Foundation of China under Grant 41771457, in part by the Young Scientists Fund of the National Natural Science Foundation of China under Grant 42101358, and in part by the Guangxi Science and Technology Major Project under Grant AA22068072. The Associate Editor for this article was D. F. Wolf. (Corresponding authors: Haigang Sui; Junyi Liu.)

Mingting Zhou, Haigang Sui, Jianxun Wang, and Junyi Liu are with the State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: mintyzhou@whu.edu.cn; haigang\_sui@263.net; wangjianxun@whu.edu.cn; liujunyi\_ljy@163.com).

Shanxiong Chen is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China, and also with the Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Kowloon, Hong Kong, China (e-mail: shanxiongchen@whu.edu.cn).

Xu Chen is with the Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China (e-mail: chenxu@ccnu.edu.cn).

Wenqing Wang is with the Collaborative Innovation Research Institute, University of Macau, Macau, China (e-mail: mc15408@um.edu.mo).

Digital Object Identifier 10.1109/TITS.2022.3181095

planning, and emergency response management. However, extracting a current and accurate road network from remotely sensed data is a challenging task. Updating roads involves the verification of roads in an old database and the extraction of new roads that must be integrated into a geographic information system [1]. Basic geospatial databases have undergone continuous development; thus, the focus of road network construction has gradually transitioned from the interpretation of the entire road network to the discovery of road changes and the updating of roads in changed areas. However, even with the historical geographic databases as the foundation, road network updating is still time-consuming, laborious, and inefficient. Hence, the development of automatic road network updating methods from remote sensing images is an urgent need.

Road extraction methods from remote sensing images collected at a current time are used to update road networks. Many different road extraction methods have been proposed in the past decades. These methods can be divided into traditional methods and deep learning-based methods. Traditional methods [2]–[4] use manually-designed features with expert knowledge to distinguish road pixels from the backgrounds. However, the manually-designed low-level and middle-level features used in these traditional methods are often overfit to a small region, and lack robustness when roads appear under different complex scenes [5]. Unlike traditional methods, deep learning-based road extraction methods [6]–[8] automatically learns hierarchical feature expression from a benchmark road dataset. The development of deep learning-based methods has greatly improved the accuracy of large-scale road network extraction. However, factors such as material changes, spectral confusions, and occlusions will reduce the completeness and correctness of the road semantic segmentation results. Complex post-processing procedures should be conducted on the whole image to smooth bristly road boundaries, link fragmented roads, and remove false roads before these road segmentation results are applied to real-world road network updates. Therefore, although road semantic segmentation methods based on deep learning can automatically extract roads, there are restrictions in directly using the road segmentation results for road updating. Road change detection between images collected at a historical time and a current time can limit the scope of road network update and thus improving the efficiency.

Road change detection can be used to determine the regions where roads should be updated, thereby reducing the workload

of road network update. Due to the lack of benchmark road change detection datasets, the existing deep learning-based road change detection methods are mainly based on general change detection neural networks [9], [10]. General change detection networks can effectively detect changes in land cover; however, complex post-processing steps are required to find road-related changes. Moreover, general change detection networks focus on spectral changes and lack geometric constraints on boundaries; thus, these networks are not sensitive to changes between objects such as buildings, roads, and paved spaces since they are constructed with similar materials and are spectral-alike. Hence, a specialized road change detection method is needed to improve the completeness of road change detection and restrict the regions of road network update; since the inherent features of roads can provide clues for road change detection from images collected at different times.

To address these problems, an Unchanged-Guided Road Updating (UGRoadUpd) framework is proposed to update a historical road database from bi-temporal remote sensing images. A dual task dominant-transformer CNN for road change detection (DT-RoadCDNet) was designed for the first stage of the proposed UGRoadUpd framework that discovers road changes from bi-temporal remote sensing images of the same area. DT-RoadCDNet executes road semantic segmentation and road change detection at the same time. This dual-task collaborative learning strategy improves road change detection accuracy by learning road structures from bi-temporal images. A Dominant-Transformer CNN was introduced into DT-RoadCDNet to model the spatial contextual structure globally thus alleviating the problem of discontinuity in road detection results. Based on the road change detection results, an unchanged-guided road update strategy updates roads in changed areas with the prior information provided by the unchanged roads in a historical road database. In addition to the proposed UGRoadUpd framework, another contribution of this paper is two new road change detection datasets, designed for testing and validating road extraction and change detection algorithms. These two datasets will be available for download at our website.

The remainder of this paper is organized as follows. Section II introduces the related road extraction and road change detection methods. Section III introduces an overview of the proposed road change detection and update framework. Section IV presents the experimental results and analysis. Section V demonstrates the ablation analyses. Conclusions are presented in Section VI.

## II. RELATED WORK

Road extraction and road change detection from high-resolution remote sensing images are the two main ways to update a historical road database. Deep learning has robust feature learning abilities for high quality road extraction and road change detection. In this section, we briefly present a review of the literature relevant to road extraction and road change detection algorithms based on deep learning techniques.

### A. Road Extraction

Deep learning-based road extraction methods automatically detect roads in a current image to update a historical road database. Research on road extraction using deep learning methods began with Mnih and Hinton's work on neural networks trained with expert-labelled data to detect roads [11]. Since then, numerous road extraction neural networks have been proposed [12]–[19]. Fully convolutional networks are the most commonly used road extraction backbones. Feature pyramid networks [8], [20], [21] and multi-scale feature fusion strategies [22]–[24] were integrated into fully convolutional networks to improve the accuracy of multi-width road extraction. However, the plain convolutional operations used in these networks do not capture long-range contextual features and thus cannot extract a complete road network when the roads in an image are occluded. Attention mechanisms can improve the incomplete road extraction results.

Attention-based road extraction methods identify relationships between pixels separated by shadows, trees, or buildings in an image and link fragmented road segments. For example, Mei [25] proposed a connectivity attention module and designed CoANet to explore the relationships between neighboring pixels to deal with occlusions. Tao *et al.* [26] designed a spatial information inference structure (SIIS-Net), enabling multidirectional message passing across the rows and columns of feature maps to guess occluded roads. Transformer-based methods [27]–[30] calculate the relationships between all pixels on a feature map using a self-attention module. In a self-attention module, each pixel has a global receptive field, thus supporting a stronger global-contextual structure reasoning ability than the conventional spatial attention modules found in CoANet and SIIS-Net, greatly improving the correctness and the topological completeness in road extraction results. However, they are computationally heavy and slow when processing large-scale remote sensing images. Moreover, updating a historical road database based on road extraction methods is constrained by the extensive updating range since only the changed roads need to be updated. In this paper, a new unchanged-guided road update (UGRoadUpd) framework is presented to update historical road databases. Road change detection and road extraction are integrated in the proposed UGRoadUpd framework, in which the road change detection procedure limits the range of roads in need of updating for higher road update efficiency, while the road extraction step updates roads in changed areas with the guidance of unchanged historical roads for higher road update accuracy.

### B. Road Change Detection

Road change detection from bi-temporal remote sensing images [31]–[33] identifies roads that are newly-built, disappeared, damaged, or reconstructed based on images collected at different times. Deep learning-based road change detection methods are usually based on general detection systems since there are few publicly available benchmark road change detection datasets. Attention mechanisms that focus on regions of interest while ignoring irrelevant areas are the most commonly deployed module in general change

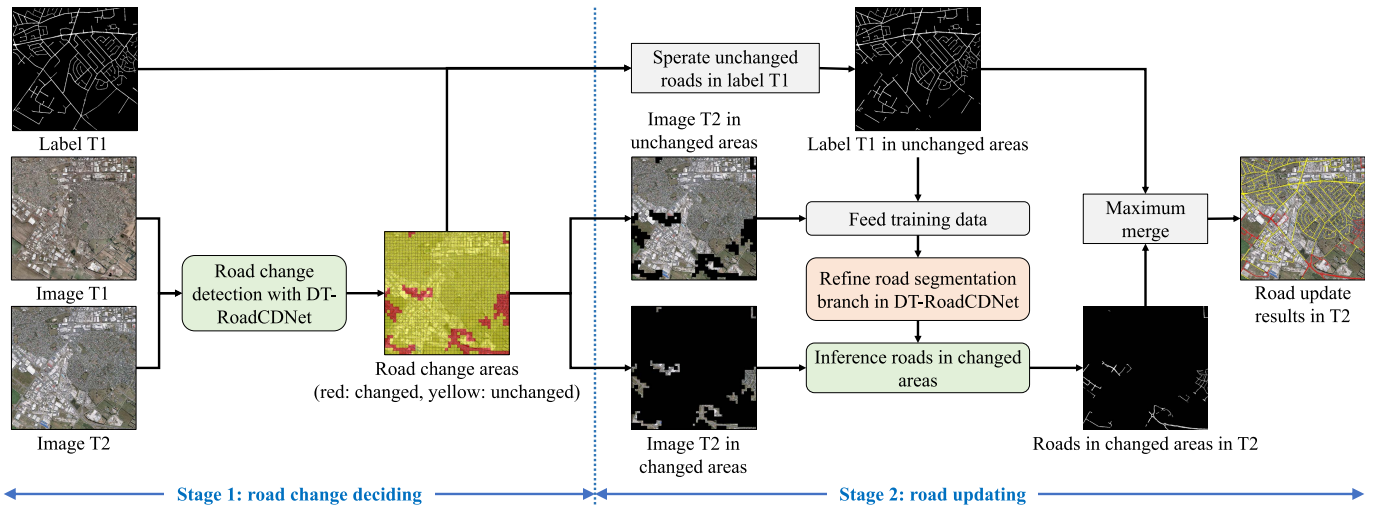


Fig. 1. Workflow of the proposed unchanged-guided road updating (UGRoadUpd) framework.

detection neural networks that model land cover changes from bi-temporal images. The spatial attention module and the channel attention module are integrated in DASNet [34] and DTCDCSN [35] to learn changes. DSAMNet [36] combines a convolutional block attention module with metric learning to learn more discriminative change features. BIT-CD [37] adopts a transformer-based encoder and decoder to relate long-range relationships between pixels in space-time. In addition to attention mechanisms, recurrent neural networks are also exploited to build temporal-spatial relationships between bi-temporal images to extract land cover changes. For example, EGRCNN [38] incorporates edge information with a recurrent convolutional neural network to improve the boundary accuracy of change detection results. UnetLSTM [39] combines U-net with long short-term memory for temporal modeling. These methods detect changes in land cover by capturing spectral differences between bi-temporal remote sensing images.

Based on the land cover change areas detected by these general change detection systems, expert knowledge is used to distinguish road changes from general land cover changes. Song *et al.* [9] proposed a spatially-adaptive denormalization based U-Net to detect general land cover changes; and object-oriented analysis was used to distinguish change types of different features including buildings, roads, farmlands, tiny houses, forests, and waterside areas. Han *et al.* [10] developed a convolutional Siamese network-based change detection method to identify the changed road region using time-series Unmanned Aerial Vehicle images. This approach provides a reference road map that can help engineers manage progress in the construction of physical roads. The post-processing steps used to separate road changes from general land cover changes are complex and lack robustness, thus leading to difficulties when updating a historical road database. A road change detection method based on the inherent features of roads is needed to improve the accuracy of road change detection and thus support automatic road updating.

### III. METHODOLOGY

A two-stage Unchanged-Guided Road Updating (UGRoadUpd) approach is proposed to update roads in basic geographic

database. The premise for the proposed UGRoadUpd framework is that the area of changed roads is far smaller than that of the unchanged roads; and thus, we can update roads in the changed area from prior knowledge provided by the unchanged roads labeled in historical databases. The workflow of the UGRoadUpd framework is shown in Fig. 1.

It can be seen from Fig. 1 that the proposed UGRoadUpd framework takes two images collected at different times and a historical reference road map as the input, and outputs an updated roads result. There are two steps in the proposed UGRoadUpd framework: a Dual-Task Road Change Detection and segmentation Network (DT-RoadCDNet) discovers regions where roads changed; and the road update strategy updates the road database in changed areas. In the second stage of the UGRoadUpd framework, our road update strategy only updates the roads in changed areas. Since a well-trained DT-RoadCDNet learns not only road changes but also roads, we used the road segmentation branch of DT-RoadCDNet to update changed roads from the current image in the second stage of our UGRoadUpd framework. However, there are spectral differences between the benchmark dataset used for training DT-RoadCDNet and the current image. Hence, to adapt the initial weights of the road segmentation network in DT-RoadCDNet to fit the current image, the road annotations covering the unchanged areas in the historical road database are combined with the current image to refine the weights of the road segmentation network. This optimization strategy allows the road segmentation network to learn the roads in the current image from the prior knowledge provided by the roads in the unchanged area, thus supporting high quality road extraction in the changed areas in a current image.

#### A. Dual-Task Learning for Road Change Detection and Segmentation

DT-RoadCDNet takes the images at two times as the input, and predicts both pixel-level road segmentation and change detection results. There are three branches in DT-RoadCDNet, including two road segmentation branches and one road change detection branch. The weights are shared among

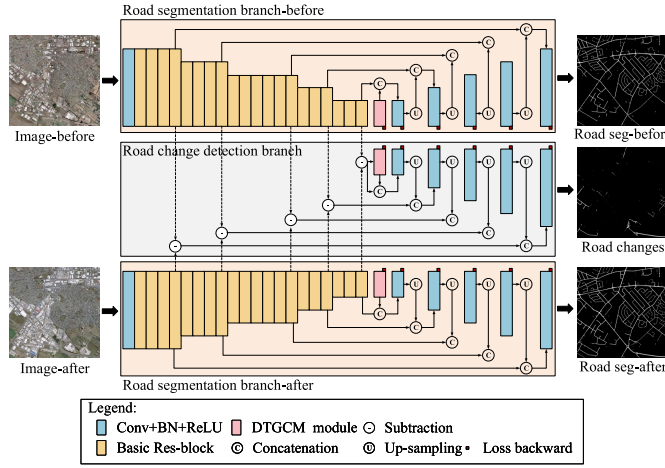


Fig. 2. The architecture of the dual-task road change detection and segmentation network (DT-RoadCDNet).

the three branches to reduce the number of parameters in DT-RoadCDNet. The architecture of DT-RoadCDNet is shown in Fig. 2, in which the two road segmentation branches are in the two big orange boxes, while the road change detection branch is in the gray box.

It can be seen from Fig. 2 that DT-RoadCDNet uses an architecture similar to the U-Net [40] with ResNet-34 [41] as the encoder. The difference between DT-RoadCDNet and U-Net lies in the fact that DT-RoadCDNet is a Siamese network with two encoders and three decoders, whereas U-Net has only one encoder and one decoder. The two encoders in DT-RoadCDNet allow users to feed bi-temporal images to the network so it can learn multi-scale road features from images of the same area collected at different times. The multi-scale road features extracted from bi-temporal images are sent to the road change detection branch to obtain the difference feature maps. Unlike the conventional U-Net that adopts a plain convolution operation to link its encoder and decoder, the output feature map of the last encoder of our DT-RoadCDNet is sent to a Dominant-Transformer based Global Context Modeling (DTGCM) module to calculate the spatial relationships between all road pixels in the whole feature map. The DTGCM module improves the completeness of road extraction and change detection results more than the plain convolution blocks used in the conventional U-Net. Details about the DTGCM module are described in Section III-B.

The three decoders in DT-RoadCDNet share the same structure and are symmetrical to the encoders, accomplishing the tasks of road extraction from the input bi-temporal images and road change detection. The segmentation of roads from bi-temporal images provides contextual information to find road changes, thus eliminating the influence of spectral variants for improved road change detection accuracy. There are five stages in each decoder, containing three plain convolution blocks in each stage. Eighteen side outputs are produced from the five decoders and the bridge part to supervise the training process of DT-RoadCDNet. A hybrid loss function  $l_{hybrid}$  designed in [5] is utilized for backward propagation to supervise the training of each side output. Since each branch

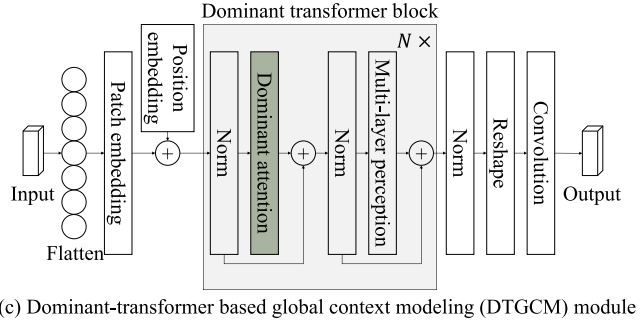
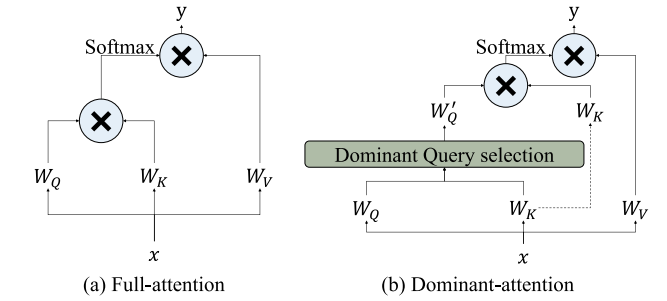


Fig. 3. Illustration of the DTGCM module, full-attention, and dominant-attention.

in the DT-RoadCDNet is supervised with six side outputs, the loss function of each branch is a summation of the six side outputs. To obtain road segmentation maps and road changes at high quality, losses of these three branches are integrated to train the network, and the overall loss is calculated as:

$$L = w_{seg}l_{seg1} + w_{seg}l_{seg2} + w_{change}l_{change} \quad (1)$$

in which  $w_{seg}$  is the weight of the outputs of the two road segmentation branches,  $w_{change}$  is the weight of the output of the change detection branch. The value of  $w_{seg}$  and  $w_{change}$  will be analyzed in Section V-A.

### B. The Dominant-Transformer Based Global Context Modeling (DTGCM) Module

Since plain convolution operations can only capture local spatial context within the receptive field, a Dominant-Transformer based Global Context Modeling (DTGCM) module was designed as the bridge to connect the encoders and decoders in DT-RoadCDNet and thus obtain the long-range contextual relationships between all road pixels in entire images for improved road segmentation and change detection integrity. The DTGCM module was developed based on Vision Transformer (ViT) [42]. In the conventional ViT, a full attention module extracts the global spatial context. However, a full attention block is computationally heavy and slow when processing large-scale remote sensing images. To speed up road extraction and change detection processing, dominant-attention is adopted as the kernel of the DTGCM module for efficient global contextual structure modeling. The architecture of the full-attention block, the dominant-attention block, and the DTGCM module are illustrated in Fig. 3.

The design of the dominant-attention module was inspired by the Informer method from natural language processing [43]. Informer was originally designed for long sequence time-series forecasting. We explored its potential in remote sensing

image processing to speed up road change detection and road extraction. Extensive experiments in Informer reveal that the dot-products of  $Q$  and  $K$  in the full-attention block are in a long-trail distribution, but only a few of them make a major contribution. Inspired by this observation, the dominant-attention module employs an approximate Kullback-Leibler divergence to evaluate the importance of each  $q_i$  in  $Q$ . The larger the value of the Kullback-Leibler divergence of  $q_i$ , the more important is the  $q_i$ . The dominant self-attention module selects  $U$  queries with the largest Kullback-Leibler divergence value to get  $Q'$ . The value of  $U$  is determined according the length of  $Q$  using the equation:

$$U = scale * \ln(L_Q) \quad (2)$$

in which  $scale$  is a pre-set scale factor to adjust the number of selected queries, which was set to five. The influence of the value of  $scale$  on the speed and accuracy of road change detection will be discussed in Section V-C.  $L_Q$  is the length of  $Q$ . The value of  $L_Q$  equals 1024. Top  $U$  queries with the largest Kullback-Leibler divergence values are selected from  $Q$  and we get the dominant queries  $Q'$ . Then the dominant-attention block calculates the dot-products between the dominant query  $Q'$  with  $K$ . In this way, we can not only use the Transformer's ability to model the global spatial context between long-range pixels to improve the continuity of road extraction/change detection results, but also reduce computational load by reducing the number of queries involved in the dot-product computation, thereby speeding up the road network change detection process.

### C. Unchanged-Guided Changed Roads Update

Although DT-RoadCDNet outputs road extraction results from a current remote sensing image, the accuracy of road extraction is limited since DT-RoadCDNet is trained on a benchmark dataset and there are differences between the benchmark dataset and a current image in reality. To alleviate this problem, this article proposes an unchanged-guided changed roads update strategy. This strategy considers that only roads in the changed areas need to be updated; and guiding the roads in changed area with the assistance of unchanged roads in a historical road database is the simplest way to update roads. Fig. 4 demonstrates the process of our unchanged guided road update strategy.

It can be seen from Fig. 4 that the proposed unchanged-guided changed road update strategy receives a current image (Image T2), a historical road database (Label T1), and a change mask produced from the DT-RoadCDNet as the input, and outputs a road update map. There are four steps in the unchanged-guided road update strategy, including reorganizing road extraction dataset, refining road segmentation network, inferencing roads in changed areas, and maximum merging to output the final road update result. As discussed in section III-A, a well-trained DT-RoadCDNet learns roads and road changes simultaneously. Hence, the segmentation branch of DT-RoadCDNet is used to infer roads in changed areas. However, the initialized segmentation branch of DT-RoadCDNet trained on a

benchmark road change detection dataset is not robust to the current image, since there are spectral differences between the benchmark dataset and the current image. Therefore, before inferencing roads in changed areas using the segmentation branch of DT-RoadCDNet, we refine the initial weights of the road segmentation branch of DT-RoadCDNet by learning from prior information provided by the unchanged roads in the historical database for improved road extraction accuracy.

The reorganization road extraction dataset and the road segmentation network refining procedures designed in our unchanged-guided road update strategy supports stronger adaptability of the road extraction network. The road extraction dataset reorganizing step automatically reproduces a new training dataset from the unchanged roads in current region to optimize the initialized road segmentation branch of DT-RoadCDNet. Based on the pixel level road change detection results from DT-RoadCDNet described in Section III-A, the proportion of changed pixels in each image patch is counted. If the pixel-level road change rate exceeds 1%, then road change has occurred in that image patch and this image patch is recorded as a changed patch; otherwise recorded as an unchanged patch. Subsequently, we use the historical road networks covering the unchanged patches with the new remote sensing images to reproduce the training dataset. In the road segmentation network refining step, the weights of the road segmentation branch of DT-RoadCDNet are refined using the newly produced training dataset. The refining process not only retains the feature distribution learned by the network on large benchmark datasets, but also can migrate the model to fit the current image. Based on the optimized road segmentation branch of DT-RoadCDNet, we predict the roads in the current images of the changed patches to update the road network in the changed area. A maximum merge operation is conducted to integrate the unchanged roads in the historical database and the updated roads in changed areas thus a complete road network on the current image can be obtained.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

### A. Experimental Setups

1) *Test Datasets and Training Details:* We manually annotated two road change detection datasets to evaluate the performance of the proposed unchanged-guided road updating (UGRoadUpd) framework. The first dataset is located in Christchurch, New Zealand, and named the Christchurch Road Change Detection (CRCD) dataset. The bi-temporal remote sensing images in the CRCD dataset were collected in the year 2012 and 2016 and were downloaded from here.<sup>1</sup> The second dataset we manually annotated is located in Jiangxia District, Wuhan, China, and named the Wuhan Road Change Detection (WRCD) dataset. We downloaded the remote sensing images of the WRCD dataset in the year of 2012, 2014, and 2016 from Google Earth.<sup>2</sup> We manually annotated the roads and road changes in the CRCD dataset and the WRCD dataset to train and test road change detection and road update algorithms. Both the CRCD and the

<sup>1</sup>[http://study.rsgis.whu.edu.cn/pages/download/building\\_dataset.html](http://study.rsgis.whu.edu.cn/pages/download/building_dataset.html)

<sup>2</sup><https://www.google.com/earth/>

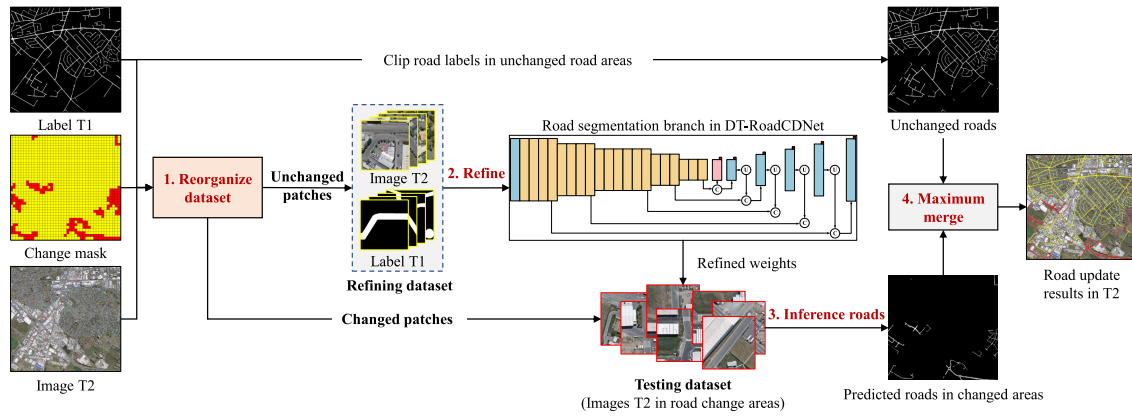


Fig. 4. Process of the unchanged-guided road update strategy.

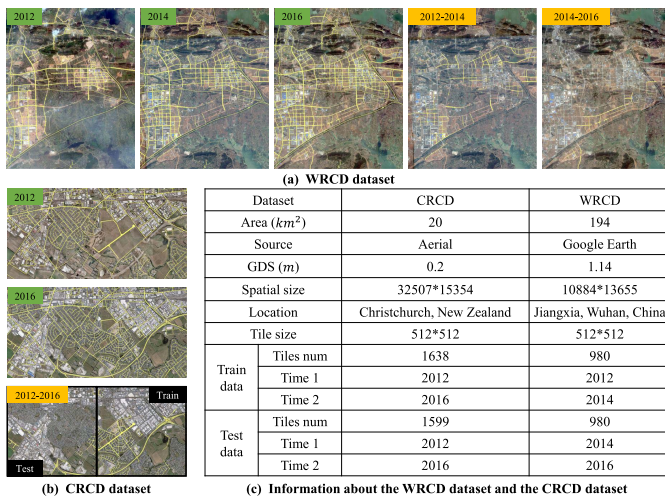


Fig. 5. Illustration of the CRCD and the WRCD datasets.

WRCD dataset will be publicly available for download at <http://www.lmars.whu.edu.cn/suihaigang/index>. An illustration of the WRCD and the CRCD datasets are shown in Fig. 5.

It can be seen from Fig. 5 that the CRCD dataset and the WRCD dataset cover a total area of approximately  $214 km^2$  with varied ground sampling distance (GSD) from 0.2 to 1.14 meters. All the images and the corresponding ground truths were cropped into  $512 \times 512$  tiles. There were 120 overlapping pixels between adjacent image blocks. The first 41 columns of CRCD dataset were taken as the test dataset with a total of 1599 samples; the other 1638 samples were used for training. Unlike the CRCD dataset, we collected images of Wuhan at three different times to verify the adaptability of the model across years. The images of 2012 and 2014 were taken as the training dataset in the WRCD dataset to train the models, and images from 2014 and 2016 as the test dataset to infer road changes and update roads. A total of 980 training samples and 980 test samples were obtained.

All the experiments were conducted on a server with one NVIDIA TITAN RTX GPU accelerator, with 24 GB GPU memory. Limited by the size of the GPU memory, the training batch size for training DT-RoadCDNet was set to two. We trained the network until the loss of the validation

dataset converged within 60 training epochs. For the WRCD dataset, the entire training process took about eleven hours. For the CRCD dataset, it took about twenty-four hours to converge. The prediction process took about 2.39 milliseconds to generate a road change result and two segmentation results for each sample pairs.

2) *Evaluation Metrics*: Precision, Recall, and Intersection over Union (IoU) were applied to evaluate road change detection and road update performance. Both road change detection and update aim to distinguish between targets and backgrounds. Precision measures the percentage of correctly classified target pixels among all predicted target pixels, while Recall measures the percentage of correctly classified target pixels among all actual target pixels. IoU is a comprehensive metric. It is the ratio of the overlapping area to the union area of the ground truth and the predicted map. They are defined as follows:  $Precision = TP / (TP + FP)$ ,  $Recall = TP / (TP + FN)$ , and  $IoU = TP / (TP + FP + FN)$ , where  $TP$ ,  $FN$  and  $FP$  are true positive, false negative, and false positive, respectively.

3) *Algorithms for Comparative Evaluation*: The proposed UGRoadUpd framework was compared with other state-of-the-art methods in two ways, a road change detection comparison and a road update comparison. To evaluate the effectiveness of our proposed DT-RoadCDNet on road change detection, DT-RoadCDNet was compared with DASNet [34], UnetLSTM [39], DTCDSN [35], BIT-CD [37], DSAM-Net [36], and EGRCNN [38] on the CRCD and WRCD datasets. The road update results of our UGRoadUpd framework were compared with the road segmentation results of the current image produced by DTCDSN [35] since DTCDSN is a dual-task collaborative networks that can extract roads and road changes simultaneously. In addition to DTCDSN, two road semantic segmentation algorithms including SIISNet [26], and CoANet [25] are also compared with the road update results of our UGRoadUpd framework. Both SIISNet and CoANet were trained with the historical remote sensing images and the corresponding road labels that cover the same region as the test dataset of CRCD and WRCD. The comparison between our UGRoadUpd framework with the three road segmentation methods verifies the effectiveness of our method on road update.

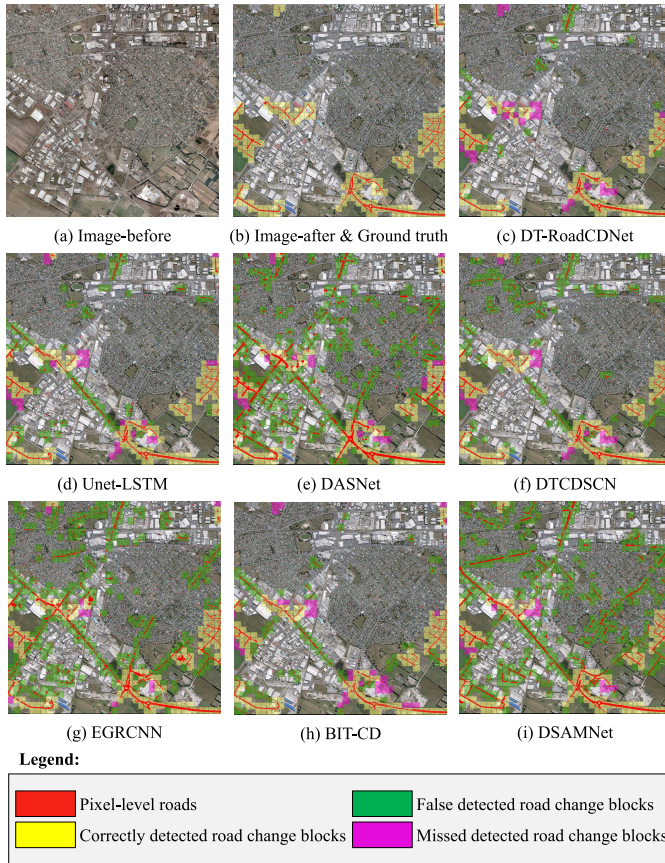


Fig. 6. Road change detection results on the CRCD dataset.

### B. Visual Results on the Christchurch Road Change Detection (CRCD) Dataset

Experiments on the CRCD dataset were conducted to verify the adaptability of the tested methods to update roads across regions. In this section, we will show the experimental results of the tested algorithms in discovering road changes on the CRCD dataset, as well as the road update results using our proposed unchanged-guided changed road update strategy, as shown in Section IV-B.1 and IV-B.2.

#### 1) Road Change Detection Results on the CRCD Dataset:

The change detection results on the CRCD dataset for all tested methods are shown in Fig. 6. In subfigures (c) to (i), the pixel-level road change detection results for the seven tested methods are shown in red. Blocks with different color marked are their patch-level road change detection results, in which the yellow blocks indicate correctly recognized road change patches, while the green and the purple blocks are false- and miss- detected road change patches.

It can be seen from Fig. 6 that the road change detection results from the proposed DT-RoadCDNet have less detection error than the other six tested change detection algorithms since there are fewer green blocks in subfigure (c) than subfigures (d) to (i). Less falsely alarmed road change areas indicate that the problem of pseudo changes caused by the spectral difference between the images at different times can be eliminated by our method. All of the six tested algorithms can recognize most of the road change blocks, as there are few purple blocks shown in Fig. 6. We selected five samples from

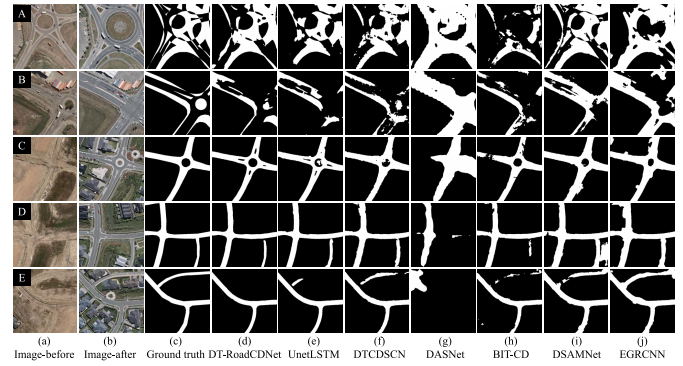


Fig. 7. Close-ups of road change detection results on the CRCD dataset.

the test dataset of the CRCD dataset to give a more intuitive comparison of our proposed method with the other three tested algorithms, as shown in Fig. 7.

In Fig. 7, region A and B show the changes brought about by road reconstruction; region C to E show the changes from dirt to paved roads. As can be seen from region A and B that the changes in road reconstruction are complicated, especially the changes between the large-scale intersections shown in the figure, it is difficult to maintain the slender topological structure of the roads. The results obtained by DT-RoadCDNet, UnetLSTM, DTCDCSN and DSAMNet are similar to the ground truth, while the results of DASNet and EGRCNN have many incorrectly identified road changes. From region C to E in the figure, we can see that our proposed DT-RoadCDNet yielded the most complete and accurate results for roads changing from dirt to paved surfaces. The change detection results of UnetLSTM, DTCDCSN, and EGRCNN at the road boundary are less accurate than DT-RoadCDNet. The pixel-level road change detection results of our DT-RoadCDNet had greater integrity than DASNet. DT-RoadCDNet and DASNet are both methods that use the attention mechanisms. However, DASNet is not sensitive to changes from soil to road, resulting in discontinuous road changes. Visual results on the CRCD dataset demonstrate that our proposed DT-RoadCDNet can effectively extract road changes caused by dirt roads, and roads under construction. The area of false-alarm and miss-detected road changes are smaller than those from the comparative methods, demonstrating that the DT-RoadCDNet improves the efficiency of road change detection.

#### 2) Road Update Results on the CRCD Dataset:

Among the 1599 patch-level samples on the CRCD test dataset, we detected 222 patches with road changes and 1377 unchanged patches based on our proposed DT-RoadCDNet. These patch-level changed and unchanged samples were derived from the pixel-level road change detection results from section IV-B.1 at a change ratio of 1%. The patch-level road change rate in the CRCD test dataset is 13.89%. We take the current remote sensing images and road annotations in the old period of these 1377 unchanged samples to form a new dataset, named ref\_CRCDD. ref\_CRCDD was randomly divided into training dataset and validation dataset at a ratio of 4:1 to re-train the road segmentation branch of DT-RoadCDNet. Since the road segmentation branch of DT-RoadCDNet is optimized based on the target image, the

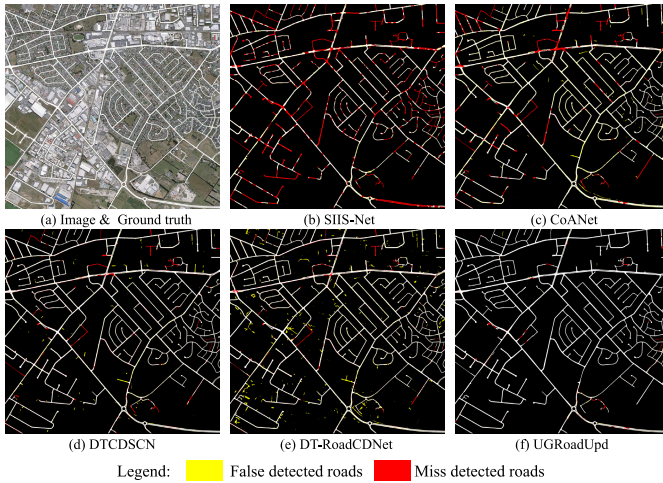


Fig. 8. Road update results on the CRCD dataset.

refined model is with more robust adaptability than the initial weights provided by DT-RoadCDNet. We inferred the roads from the current images of the 222 changed samples to update roads in changed areas. The road update result for the new image in the CRCD test dataset is shown in Fig. 8.

As can be seen from Fig. 8, in comparison with the segmentation results from the other four tested algorithms, the visual completeness and accuracy of the road update results produced by our UGRoadUpd framework shown in subfigure (f) were greatly improved. The road update results of SIIS-Net and CoANet are discontinuous as there are many red pixels in subfigures (b) and (c). Subfigure (e) is the road segmentation result obtained by DT-RoadCDNet, i.e. the first stage of our UGRoadUpd framework. Compared with the results of DT-RoadCDNet, there were less missed detected roads after adopting the unchanged guided roads refining process, verifying the effectiveness of the proposed unchanged-guided road update strategy.

The low visual integrity of SIIS-Net and CoANet demonstrates that it is challenging to generalize neural networks trained across time. Hence, learning from historical images and road database cannot guarantee the quality of road update from current images. The unchanged-guided road update strategy can not only restrict the range of road update for improved road update efficiency, but also improve the ability of the model to adapt to the current image. The capacity of the road extraction model to migrate is one of the biggest bottlenecks restricting the automaticity of road update methods. The strategy proposed in this paper effectively alleviates this bottleneck, thereby improving road network update automation.

### C. Visual Results on Wuhan Road Change Detection Dataset

Experiments on Wuhan road change detection (WRCD) dataset were conducted to evaluate the adaptability of the method across time. The imaging time of the training and the testing data from the WRCD dataset was collected during different years, unlike the CRCD dataset collected from different regions. Hence, there is a greater difference in radiation and spectrum between the training and testing dataset of the

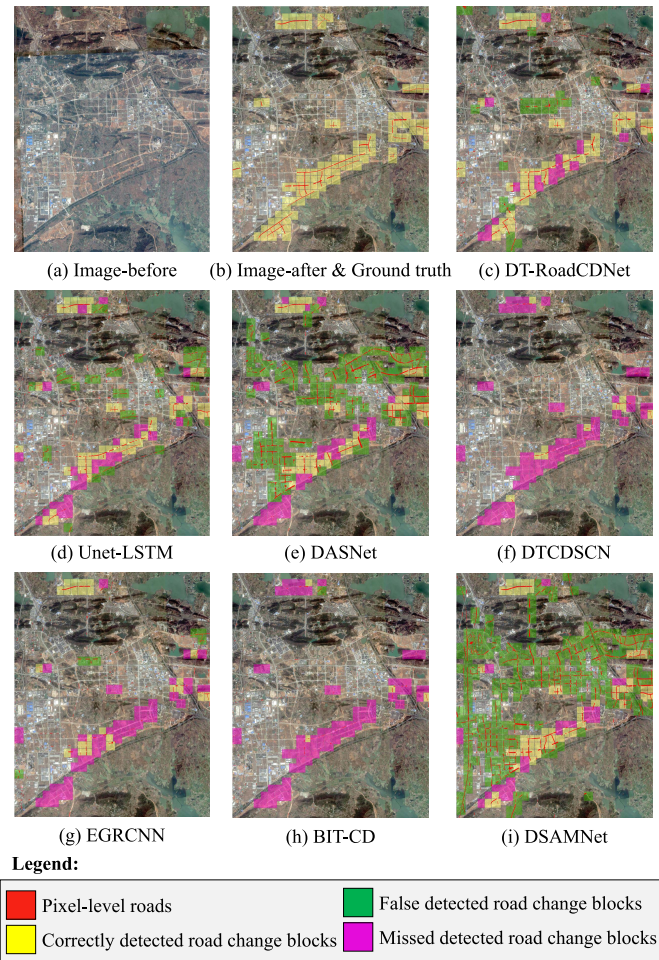


Fig. 9. Road change detection results on the WRCD dataset from 2014 to 2016.

WRCD dataset, requiring higher robustness from the tested algorithms.

1) *Road Change Detection Results on the WRCD Test Dataset:* The change detection results for the seven tested change detection methods on the WRCD dataset are shown in Fig. 9.

As can be seen from Fig. 9, DT-RoadCDNet, UnetLSTM, and DSAMNet can distinguish most of the road changes at the pixel-level and the patch-level. There were many miss-detected road changes in the results of DTCDCSN, DASNet, EGRCNN, and BIT-CD, especially for roads changed from muddy soil. There were fewer pseudo changes detected in the results from our DT-RoadCDNet than DASNet and DSAMNet. An excessively high false change detection rate will increase the workload in the basic geographic database updating process, because the roads in the changed area need to be updated. We selected five samples from the test dataset of the WRCD dataset to give a more intuitive comparison of our proposed method with the other six tested algorithms, as shown in Fig. 10.

It can be seen from the five detailed images in Fig. 10 that the continuity of changing roads detected by DT-RoadCDNet was stronger with fewer discontinuous roads as compared with the other six tested methods, indicating that the global attention mechanism for modeling spatial relations improved the



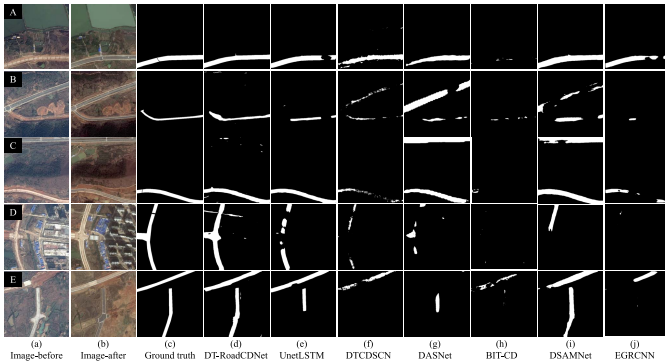


Fig. 10. Close-ups of road change detection results on the WRCD dataset from 2014 to 2016.

integrity of the changed road results. The integrity of changing roads was weak in the UnetLSTM results shown in subfigure (e), as there were missed detections at road intersections and in areas where the road material changed. DASNet and EGRCNN could not detect road changes between the roads under-construction and completed roads seen in region D and E, thus limiting its applicability in the management of a road construction. The completeness and correctness of BIT-CD on the WRCD dataset were the least visually convincing. Both BIT-CD and our proposed DT-RoadCDNet are transformer-based CNNs. BIT-CD employs transformer modules in the decoders and encoders; while our DT-RoadCDNet take transformer as a bridge to link CNN-based encoders and decoders to exploit the global-context modeling ability of transformers. Transformers in both encoders and decoders cannot be trained easily with a small number of road change labels, thus the migration ability of BIT-CD on images at different imaging times was less effective than our DT-RoadCDNet. In general, on the WRCD dataset, DT-RoadCDNet has the strongest migration ability for detecting changes between images of different times.

2) *Road Update Results on the WRCD Dataset:* Among the 980 samples on the test dataset of the WRCD dataset, we detected 185 samples with road changes and 795 unchanged samples using the proposed DT-RoadCDNet. We predicted the roads from the current images of the 185 changed samples to update the historical road database as guided by 795 unchanged samples. The road update results for the current image in the WRCD test dataset is shown in Fig. 11. The overview road update result of our UGRoadUpd framework is shown in subfigure (a). Details about the visual comparisons between the five tested methods for regions marked in subfigure (a) are magnified and shown in subfigures (b) to (h).

It can be seen from Fig. 11 that the overview road update results of our UGRoadUpd framework are visually convincing. The details of small roads were retained in the results of our UGRoadUpd framework. Details shown in subfigures (b) to (h) reveal that the road network updated by our UGRoadUpd maintains the connectivity while the results of CoANet, SIIS-Net, DTCDCSCN, and DT-RoadCDNet may be disconnected. In Region A, our UGRoadUpd method can accurately identify the road with material changes, but DTCDCSCN and

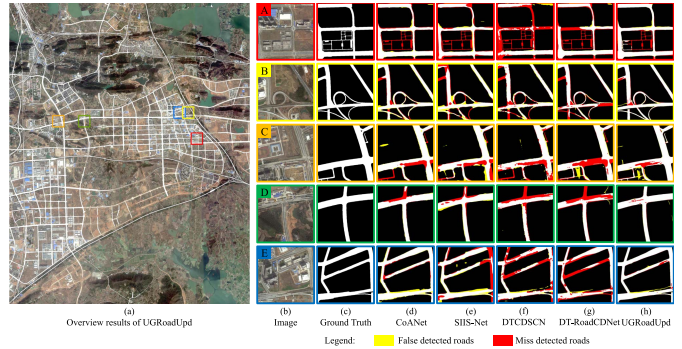


Fig. 11. Road updating results on the WRCD dataset in the year 2016.

DT-RoadCDNet cannot extract these roads at high integrity. In Regions B to D, the roads in the middle of the image are under construction with heterogeneous road surface; the results of DTCDCSCN and DT-RoadCDNet failed to preserve the connectivity, but the result obtained by UGRoadUpd is more consistent with the ground truth. Region E demonstrates the roads occluded by shadows on the image. All four comparative methods failed to extract occluded roads, while our UGRoadUpd approach obtained a complete road network.

*D. Quantitative Analysis*

1) *Road Change Detection Results:* TABLE I makes a quantitative comparison between our method and six other change detection methods on the CRCDD and the WRCD datasets. Pixel-level and patch-level metrics were displayed in the table. Line five to eleven of the table show the metrics for the seven tested algorithms, and the twelfth to the seventeenth rows show the difference between our proposed DT-RoadCDNet and six other tested methods.

As can be seen from TABLE I, our proposed DT-RoadCDNet achieved the highest IoU scores on both CRCDD and WRCD datasets at pixel-level and patch-level, indicating that our method can keep the balance between the precision and completeness of road change detection. The road change detection metrics of all the tested methods on the WRCD test dataset were lower than the scores on the CRCDD test dataset. This is because the imaging time of the train and the test WRCD dataset are different. Lower evaluation scores of road change detection results on the WRCD test dataset indicate that it is a bottleneck for the change detection neural networks to generalize across time. The pixel-level Recall scores of DT-RoadCDNet on both tested datasets are lower than DSAMNet. Considering both the visual results in Fig. 6 and Fig. 9, our method has fewer omissions as compared with DSAMNet visually. Moreover, our method showed a significant improvement on Precision and IoU as compared to DSAMNet, demonstrating that there are fewer false detected road change pixels in the results of our DT-RoadCDNet than that of the DSAMNet. In terms of patch level evaluation, compared with the six comparative methods, our method delivered a noticeable improvement of 15.87% to 44.91% on Precision on the CRCDD dataset, an increase of 2.02% to 53.60% on Recall on the WRCD dataset, and at least a 9.60% improvement on the comprehensive indicator IoU on both

TABLE I  
QUANTITATIVE ANALYSIS OF ROAD CHANGE DETECTION

Methods		Dataset											
		Christchurch road change detection (CRCD) dataset						Wuhan road change detection (WRCD) dataset					
		Pixel level metrics (%)			Patch level metrics (%)			Pixel level metrics (%)			Patch level metrics (%)		
		Recall	Precision	IoU	Recall	Precision	IoU	Recall	Precision	IoU	Recall	Precision	IoU
The whole test area	DAS [34]	77.38	25.30	23.55	89.37	31.79	30.63	34.15	10.83	8.96	66.67	24.00	21.43
	DTC [35]	85.09	55.78	50.81	90.34	48.20	45.83	16.07	29.25	23.37	35.35	50.00	26.12
	ULS [39]	81.65	59.99	52.86	87.44	57.46	53.08	38.35	23.37	16.99	76.77	32.90	29.92
	BIT [37]	67.89	57.03	44.91	85.02	54.83	50.00	13.93	34.08	10.97	29.23	74.36	26.61
	DSM [36]	86.57	29.06	27.81	91.79	29.01	28.27	51.80	9.70	8.90	80.81	19.66	18.78
	EGR [38]	83.49	30.80	29.03	93.24	28.42	27.85	23.24	32.35	15.64	44.44	50.57	30.99
	DTR	79.74	77.67	64.87	85.02	73.33	64.94	49.23	31.68	23.88	82.83	44.32	40.59
Diff	DTR-DAS	2.36	52.37	41.32	-4.35	41.54	34.31	15.08	20.85	14.92	16.16	20.32	19.16
	DTR-DTC	-5.35	21.89	14.06	-5.32	25.13	19.11	33.16	2.43	0.51	47.48	-5.68	14.47
	DTR-ULS	-1.91	17.68	12.01	-2.42	15.87	11.86	10.88	8.31	6.89	6.06	11.42	10.67
	DTR-BIT	11.85	20.64	19.96	0.00	18.50	14.94	35.30	-2.40	12.91	53.60	-30.04	13.98
	DTR-DSM	-6.83	48.61	37.06	-6.77	44.32	36.67	-2.57	21.98	14.98	2.02	24.66	21.81
	DTR-EGR	-3.75	46.87	35.84	-8.22	44.91	37.09	25.99	-0.67	8.24	38.39	-6.25	9.60

DTR, DTC, ULS, DAS, BIT, DSM, and EGR are short for DT-RoadCDNet, DTCDCSN [35], UnetLSTM [39], DASNet [34], BIT-CD [37], DSAMNet [36], and EGRCNN [38]. Diff is the short for Difference.

TABLE II  
QUANTITATIVE ANALYSIS OF ROAD UPDATE RESULTS

Dataset	Method	Metrics (%)		
		Recall	Precision	IoU
CRCD	DT-RoadCDNet	90.77	92.30	84.38
	DTCDCSN [35]	92.21	84.62	78.98
	CoANet [25]	82.61	84.77	71.93
	SIIS-Net [26]	60.86	91.16	57.47
	UGRoadUpd	97.18	98.56	95.82
WRCD	DT-RoadCDNet	75.43	86.40	67.43
	DTCDCSN [35]	69.52	80.91	59.73
	CoANet [25]	80.73	83.32	69.50
	SIIS-Net [26]	80.12	77.53	65.02
	UGRoadUpd	94.11	95.48	90.10

datasets. These results confirm that our method can reduce false changes and greatly improve the accuracy of change detection, thus improving the road update efficiency.

2) *Road Update Results*: TABLE II shows the quantitative comparison of the three tested road update results on the CRCD dataset and the WRCD dataset shown in Fig. 8 and Fig. 11.

As can be seen from TABLE II, compared with the road update results of the other tested algorithms, the result of our UGRoadUpd framework yielded at least an improvement of 3% in Recall, 6% in Precision, and 11% in IoU scores on both CRCD and WRCD datasets. Although the metrics of the road change detection results of our DT-RoadCDNet on the WRCD test dataset were lower than the CRCD test dataset shown in TABLE I, the recall, precision, and IoU scores of the road update results on the WRCD test dataset were higher than 90%. These road update scores show comparable performance with results on the CRCD test dataset. The road update results on the WRCD test dataset are produced based on the road change detection results of DT-RoadCDNet and the unchanged-guided road update strategy proposed in this paper. The UGRoadUpd is consisted of a road change detection process using DT-RoadCDNet and a road update process using unchanged-guided refining strategy. Compared with DT-RoadCDNet, the UGRoadUpd boost the recall scores of the road update results from 75.43% to 94.11%, precision scores from 86.40% to 95.48%, and IoU scores from 67.43% to 90.10% on the WRCD dataset. These improvements in the three evaluation indicators further validate the significance of

the unchanged-guided road network update strategy proposed in this paper.

## V. PARAMETER SETTINGS AND ABLATION ANALYSIS

As described in Section III, there are two hyperparameters in our proposed DT-RoadCDNet, including the weights of the road change detection and road segmentation branches, and the scale parameter to control the number of the dominant queries. The effect of the DTGCM module on the integrity of road change detection is also analyzed in this section. To increase the speed of the experiments for parameter settings and ablation analysis, we created a small dataset based on images and labels from the year of 2012 and 2014 in the Wuhan Road Change Detection dataset, called the Mini Wuhan Road Change detection (MWRCD) dataset. All the images and the corresponding ground truths in the MWRCD dataset were cropped into  $512 \times 512$  tiles. There were 120 overlapping pixels between adjacent image blocks. Half of MWRCD dataset was taken as the training dataset, while the rest was the test dataset. A total of 490 training samples and 490 test samples were obtained. The entire training process took about five hours until the loss of the validation dataset converged within 60 training epochs. It saves about 90 hours of time by conducting the fifteen ablation experiments with the MWRCD dataset rather than the WRCD dataset.

### A. Influence of Dual Task Collaborative Learning

DT-RoadCDNet is a dual task neural network with three output branches. We conducted experiments of different weight setting of the change detection branch and the two road segmentation branches, to evaluate the contributions of different branches for road change detection. The quantitative comparisons for the MWRCD dataset are presented in TABLE III.

It can be seen from TABLE III that the Exp 1 outperformed the other parameter settings. Compared with Exp 5, Exp 1 shows an improvement of 6.84% on recall, 6.90% on precision, and 7% on IoU. In Exp 5, only the change detection branch was used to supervise the network training process; while in Exp 1, the change branch and the segmentation branches were regarded as equally important. The improved

TABLE III

COMPARISONS OF DIFFERENT WEIGHT SETTINGS OF THE CHANGE DETECTION BRANCH AND THE TWO SEGMENTATION BRANCHES

ID	Weights		Metrics (%)		
	$w_{change}$	$w_{seg}$	Recall	Precision	IoU
1	1	1	79.08	73.60	60.10
2	1	0.8	69.27	72.26	54.72
3	1	0.5	69.76	71.70	54.70
4	1	0.2	73.87	55.62	46.48
5	1	0	72.24	66.70	53.10

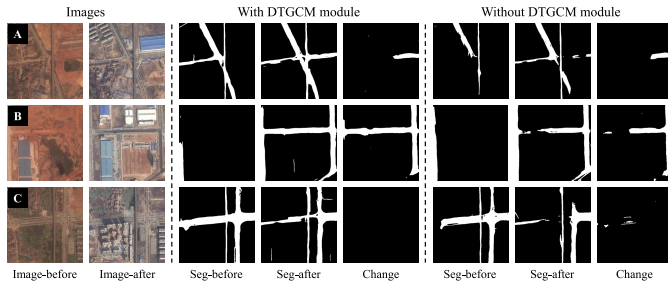


Fig. 12. Illustration of results on the Wuhan road change detection dataset with and without DTGCM module.

evaluation indicators of Exp 1 compared with Exp 5 shows that the supervision of road segmentation on the images collected at different times effectively supports the discovery of road changes. In Exp 2 to Exp 4, different weights for the two road segmentation branches were tested to discover how road segmentation influences the effect of road change detection. It can be seen that as the weights of the two segmentation branches increase, so do the precision and IoU scores of road change detection, indicating that the introducing of road segmentation branches can reduce the false-detected road changes. Therefore, we set the same weights for the road change detection branch and the two road semantic segmentation branches for improved performance of road change detection.

### B. Effect of the Dominant-Transformer Based Global Context Modeling (DTGCM) Module

To evaluate the influence of the proposed DTGCM module on road change detection and segmentation accuracy, we replaced the DTGCM module in DT-RoadCDNet with plain convolutional blocks, and conducted experiments on the MWRCD dataset. Fig. 12 demonstrates three representative experimental results.

It can be seen in Fig. 12 that the DTGCM module produces more complete results for roads and changed areas. Without the global spatial context modeling by the DTGCM module, complete roads and road changes cannot be maintained when there are sudden material changes, and the occluded roads cannot be reasonably estimated. The quantitative evaluation results of road change detection on the MWRCD dataset of DT-RoadCDNet with and without the DTGCM module are shown in TABLE IV.

It can be seen in TABLE IV that all the evaluation metrics of road change detection increases with the introduce of the DTGCM module. DT-RoadCDNet with DTGCM module shows an improvement of 5.33% on Recall, 1.51% on Precision, and 2.85% on IoU than DT-RoadCDNet without the

TABLE IV

QUANTITATIVE EVALUATION ON ROAD CHANGE DETECTION RESULTS WITH OR WITHOUT THE DTGCM MODULE

Methods	Metrics(%)		
	Recall	Precision	IoU
With DTGCM	79.08	73.60	60.10
Without DTGCM	73.75	72.09	57.25

TABLE V

INFLUENCE OF THE *Scale* PARAMETER ON THE PERFORMANCE OF ROAD CHANGE DETECTION

ID	<i>scale</i>	$U$	Time (ms)	Recall	Precision	IoU
1	1	7	2.32	69.36	68.93	52.84
2	3	21	2.32	55.51	77.04	47.63
3	5	35	2.39	79.08	73.60	60.10
4	7	49	2.39	64.45	69.88	50.44
5	9	63	2.53	78.36	65.80	55.69
6	10	70	2.53	72.71	75.39	58.76
7	30	210	3.04	72.46	70.59	55.66
8	50	350	3.11	71.25	74.96	57.55
9	90	630	3.49	63.74	77.24	53.66
10	full	1024	3.39	45.94	85.23	42.55

DTGCM module. These improvements verify the efficiency of the DTGCM module when extracting roads and road changes under complex environments.

### C. Effect of the Scale Parameter of Dominant Transformer

The *scale* parameter is a coefficient number that controls the number of the dominant queries in our proposed Dominant-Transformer block. We set the value of the *scale* parameter to [1, 3, 5, 7, 9, 10, 30, 50, 90] and tested their performance. The evaluation indicators of different *scale* values are shown in TABLE V. The third column in the table is the value of  $U$ , the number of dominant queries.  $U$  is automatically calculated using equation (2) and is related to the value of *scale*. The units of  $U$  and  $L_Q$  are number of “nodes,” while *scale* has no unit. The third column demonstrates the inference time of each image for different *scale* settings. To evaluate the effectiveness of the dominant attention module proposed in this paper, the last row of the table shows the road change detection accuracy using the full attention module found in Vision Transformer [42].

It can be seen from TABLE V that when *scale* is set to five, the recall and IoU of the change detection result are the highest. It took 2.39 milliseconds to predict the change detection result of an image pair with 35 dominant queries in Exp 3, while it took 3.39 milliseconds to use the full attention module, indicating that the dominant attention mechanism designed in this paper can simultaneously improve the completeness and the speed of change detection. In Exp 1, when the *scale* parameter was set to one, a total of seven dominant queries were selected. However, the three-evaluation metrics dropped significantly compared with Exp 3 in which the *scale* parameter was set to five, demonstrating that too small numbers of dominant queries will affect the effect of change detection. Exp 10 with full attention mechanism got the highest precision score. However, comparing with Exp 3, the recall and IoU of the full attention mechanism were reduced by 33.14% and 17.55%, indicating that selecting dot-product pairs that contribute to the major contribution is helpful to improve the completeness of road change detection. It can be seen from

Exp 9 that when *scale* is set to 90, the inference time of the Dominant Attention module on each image is longer than that of Full Attention. The result of Exp 9 shows that we should control the number of dominant queries to less than 60% of the original queries to improve the completeness and efficiency of road change detection. Based on the above analysis, we set *scale* to five for improved road change detection speed and accuracy.

## VI. CONCLUSION

In this paper, a two stage Unchanged-Guided Road Updating (UGRoadUpd) framework is proposed that automatically updates road networks. The proposed UGRoadUpd framework is based on the premise that only changed roads need to be updated; and we can update roads in the changed area by exploiting prior knowledge provided by the unchanged roads labeled in historical databases. A dual task dominant-transformer CNN for road change detection (DT-RoadCDNet) was designed as the first stage of the UGRoadUpd framework to discover road changed areas. DT-RoadCDNet collaboratively learns both road semantic segmentation and road change detection tasks thus reducing the high false- and miss-detected changes caused by the complex remote sensing imaging mechanisms as well as by the variation in the appearance of ground objects. Roads are topologically connected; but the common road detection networks capture contextual information only within the local receptive field, leading to discontinuous road segmentation results. To address this problem, a Dominant-Transformer module was introduced to model the spatial contextual structure globally, thus improving the integrity of road change detection and road updating. Unlike a conventional Transformer, the proposed Dominant-Transformer is more efficient when dealing with large-scale remote sensing images. In the second stage, an unchanged-guided road update strategy was designed to update roads in the changed area with prior information provided by roads in an unchanged area. Extensive experiments on two newly-annotated benchmark datasets confirm the effectiveness of the proposed UGRoadUpd framework. In the future, we will focus on the use of remote sensing images collected at a current time to directly update a historical road network since it is difficult to obtain historical remote sensing images that match the time of the historical road maps in some cases.

## ACKNOWLEDGMENT

The authors would like to thank a reviewers and an editors for their valuable comments and efforts that helped improve this paper. They also would like to thank Stephen C. McClure for linguistic assistance.

## REFERENCES

- [1] A. Baumgartner, C. Steger, C. Wiedemann, H. Mayer, W. Eckstein, and H. Ebner, "Update of roads in GIS from aerial imagery: Verification and multi-resolution extraction," *Int. Arch. Photogramm. Remote Sens.*, vol. 31, pp. 53–58, Jul. 1996.
- [2] S. Leninisha and K. Vani, "Water flow based geometric active deformable model for road network," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 140–147, Apr. 2015.
- [3] H. Zhou, H. Kong, L. Wei, D. Creighton, and S. Nahavandi, "On detecting road regions in a single UAV image," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1713–1722, Jul. 2017.
- [4] J. Dai, T. Zhu, Y. Wang, R. Ma, and X. Fang, "Road extraction from high-resolution satellite images based on multiple descriptors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 227–240, 2020.
- [5] M. Zhou, H. Sui, S. Chen, J. Wang, and X. Chen, "BT-RoadNet: A boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 288–306, Oct. 2020.
- [6] Z. Chen, W. Fan, B. Zhong, J. Li, J. Du, and C. Wang, "Corse-to-fine road extraction based on local Dirichlet mixture models and multiscale-high-order deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4283–4293, Oct. 2020.
- [7] L. Ma, Y. Li, J. Li, J. M. Junior, W. N. Goncalves, and M. A. Chapman, "BoundaryNet: Extraction and completion of road boundaries with deep learning using mobile laser scanning point clouds and satellite imagery," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5638–5654, Jun. 2022.
- [8] H. Guan, Y. Yu, D. Li, and H. Wang, "RoadCapsFPN: Capsule feature pyramid network for road extraction from VHR optical remote sensing imagery," *IEEE Trans. Intell. Transp. Syst.*, early access, Aug. 6, 2021, doi: [10.1109/TITS.2021.3098855](https://doi.org/10.1109/TITS.2021.3098855).
- [9] C. Song, W. Wahyu, J. Jung, S. Hong, D. Kim, and J. Kang, "Urban change detection for high-resolution satellite images using U-Net based on spade," *Korean J. Remote Sens.*, vol. 36, nos. 2–6, pp. 1579–1590, 2020.
- [10] D. Han, S. Lee, M. Song, and J. Cho, "Change detection in unmanned aerial vehicle images for progress monitoring of road construction," *Buildings*, vol. 11, no. 4, p. 150, Apr. 2021.
- [11] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Proc. Eur. Conf. Comput. Vis.* New York, NY, USA: Springer, 2010, pp. 210–223.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* New York, NY, USA: Springer, 2016, pp. 630–645.
- [13] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1591–1594.
- [14] R. Lian and L. Huang, "DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1905–1916, 2020.
- [15] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [16] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018.
- [17] Y. Li, B. Peng, L. He, K. Fan, and L. Tong, "Road segmentation of unmanned aerial vehicle remote sensing images using adversarial network with multiscale context aggregation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2279–2287, Jul. 2019.
- [18] P. Shamsolmoali, M. Zareapoor, H. Zhou, R. Wang, and J. Yang, "Road segmentation for remote sensing images using adversarial spatial pyramid networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4673–4688, Jun. 2021.
- [19] X. Zhang, X. Han, C. Li, X. Tang, H. Zhou, and L. Jiao, "Aerial image road extraction based on an improved generative adversarial network," *Remote Sens.*, vol. 11, no. 8, p. 930, Apr. 2019.
- [20] H. He, D. Yang, S. Wang, S. Wang, and Y. Li, "Road extraction by using atrous spatial pyramid pooling integrated encoder–decoder network and structural similarity loss," *Remote Sens.*, vol. 11, no. 9, p. 1015, 2019.
- [21] Y. Lin, D. Xu, N. Wang, Z. Shi, and Q. Chen, "Road extraction from very-high-resolution remote sensing images via a nested SE-deeplab model," *Remote Sens.*, vol. 12, no. 18, p. 2985, Sep. 2020.
- [22] Y. Li, L. Guo, J. Rao, L. Xu, and S. Jin, "Road segmentation based on hybrid convolutional network for high-resolution visible remote sensing image," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 613–617, Apr. 2019.
- [23] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning," *Remote Sens.*, vol. 10, no. 9, p. 1461, Sep. 2018.
- [24] X. Tan, Z. Xiao, Q. Wan, and W. Shao, "Scale sensitive neural network for road segmentation in high-resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 533–537, Mar. 2020.

- [25] J. Mei, R.-J. Li, W. Gao, and M.-M. Cheng, "CoANet: Connectivity attention network for road extraction from satellite imagery," *IEEE Trans. Image Process.*, vol. 30, pp. 8540–8552, 2021.
- [26] C. Tao, J. Qi, Y. Li, H. Wang, and H. Li, "Spatial information inference net: Road extraction using road-specific contextual information," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 155–166, Dec. 2019.
- [27] L. Ding *et al.*, "Looking outside the window: Wide-context transformer for the semantic segmentation of high-resolution remote sensing images," 2021, *arXiv:2106.15754*.
- [28] Z. Zhang, C. Miao, C. Liu, and Q. Tian, "DCS-TransUpperNet: Road segmentation network based on CSwin transformer with dual resolution," *Appl. Sci.*, vol. 12, no. 7, p. 3511, Mar. 2022.
- [29] T. Chen, D. Jiang, and R. Li, "Swin transformers make strong contextual encoders for VHR image road extraction," 2022, *arXiv:2201.03178*.
- [30] Z. Sun, W. Zhou, C. Ding, and M. Xia, "Multi-resolution transformer network for building and road segmentation of remote sensing image," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 3, p. 165, Feb. 2022.
- [31] Z. Qiaoping and I. Couloigner, "Automatic road change detection and GIS updating from high spatial remotely sensed imagery," *Geo-Spatial Inf. Sci.*, vol. 7, no. 2, pp. 89–95, 2004.
- [32] M. Dong, H. Zhang, X. Zhu, and C. Chen, "Change detection of road networks based on remote sensing image," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 34, no. 2, pp. 178–182, 2009.
- [33] Z. Lu, X. Zuo, L. Huang, and J. Liu, "Road change detection using object-oriented projective interactive partition," *Remote Sens. Land Resour.*, vol. 3, pp. 60–64, Dec. 2012.
- [34] J. Chen *et al.*, "DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1194–1206, 2020.
- [35] Y. Liu, C. Pang, Z. Zhan, X. Zhang, and X. Yang, "Building change detection for remote sensing images using a dual-task constrained deep Siamese convolutional network model," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 811–815, May 2021.
- [36] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.
- [37] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.
- [38] B. Bai, W. Fu, T. Lu, and S. Li, "Edge-guided recurrent convolutional neural network for multitemporal remote sensing image building change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021.
- [39] M. Papadomanolaki, S. Verma, M. Vakalopoulou, S. Gupta, and K. Karantzas, "Detecting urban changes with recurrent neural networks from multitemporal sentinel-2 data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 214–217.
- [40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* New York, NY, USA: Springer, 2015, pp. 234–241.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [42] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [43] H. Zhou *et al.*, "Informer: Beyond efficient transformer for long sequence time-series forecasting," 2020, *arXiv:2012.07436*.



**Mingting Zhou** received the B.S. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, in 2015. She is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. Her research interests include high spatial resolution remote sensing image road detection, image processing, and deep learning.



**Haigang Sui** received the B.S. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1996 and 2002, respectively. He is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His major research interests include change detection in remote sensing, target recognition, and disaster analysis.



**Shanxiang Chen** received the B.S. degree in remote sensing science and technology and the Ph.D. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2015 and 2022, respectively. His research interests include object extraction, change detection, and deep learning.



**Xu Chen** received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2021. She is currently a Post-Doctoral Researcher with the Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan. Her major research interests include vehicle re-identification, object tracking, and deep learning.



**Wenqing Wang** received the B.S. degree in geodesy and geomatics engineering from the School of Geodesy and Geomatics, Wuhan University, Wuhan, China, in 2020. She is currently pursuing the M.S. degree in artificial intelligence with the Collaborative Innovation Research Institute, University of Macau, Macau, China. Her research interests include face detection and deep learning.



**Jianxun Wang** received the M.A. degree in agricultural informatization from Huazhong Agricultural University, Wuhan, China, in 2020. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing. His research interests include natural resources monitoring, change detection, and deep learning for remote sensing.



**Junyi Liu** received the B.S. degree in geodesy and geomatics engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University in 2002 and 2008, respectively. She is currently an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. Her research interests include synthetic aperture radar image segmentation, image registration, and change detection.