M-AFDE-NET: NOVEL DEEP LEARNING-BASED BUILDING CHANGE DETECTION OF FRESHLY BUILT LOCALES FROM SATELLITE IMAGERY IN THE NILE VALLEY, EGYPT

Shimaa Holail[1], Tamer Saleh[1,2], Xiongwu Xiao[1,*], Zhenfeng Shao[1], Haigang Sui[1], Deren Li[1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China (xwxiao@whu.edu.cn, shimaa.younes17@feng.bu.edu.eg)
[2] Geomatics Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt

KEY WORDS: Building Change Detection, Deep Learning, High-resolution Satellite Imagery, Freshly Built Locales (FBLs), Convolutional Neural Network (CNN), Transformers.

ABSTRACT:

Urban land expansion is a defining characteristic of urbanization, necessitating the monitoring of this phenomenon and the detection of changes to promote sustainable land use and contribute to updating geospatial databases. Methods based on detecting changes in high-resolution satellite imagery have shown poor performance due to downsampling during image processing, resulting in the loss of boundary information. Furthermore, these methods struggle with complex backgrounds where the ground resembles building roofs. This paper delves into the investigation and evaluation of Freshly Built Locales (FBLs) using bi-temporal images through recently proposed computer vision networks. To address the limitations of existing approaches, we have introduced modifications to the AFDE-Net model, which include the novel residual pyramid attention fusion (RPAF) module. This enhancement enables more precise identification of intricate details in complex change detection scenarios. Our proposed model, M-AFDE-Net, has been evaluated on a newly captured dataset from the Nile Valley regions of Egypt, with a spatial resolution of 30 cm. Special attention has been given to New Mansoura and New Tiba as focal areas for analysis. The evaluation results reveal that the modified model, M-AFDE-Net, outperforms other state-of-the-art models in detecting FBLs. It achieves an impressive F1-score of approximately 89.2%, demonstrating its superiority and effectiveness.

1. INTRODUCTION

The development of new cities in Egypt is a crucial national strategy aimed at fostering comprehensive growth and addressing issues arising from population growth and economic activity. Over the past few decades, Egypt has witnessed a significant increase in urbanization, with an annual urban growth rate of about 2%. This translates to almost one million new citizens that Egyptian cities must accommodate each year. Greater Cairo alone added half a million new residents in 2017, earning it the title of the fastest-growing city in the world (The National, 2020). Urban land expansion is one of the defining characteristics of urbanization, making it essential to monitor the phenomenon and detect changes to promote sustainable land use. In this context, the updating of geospatial databases and monitoring of urbanization can significantly contribute to the development of smart cities. Recent advances in high-resolution satellite imagery have demonstrated their high efficiency in observing the Earth and detecting changes. This is due to the rapid response and wide coverage of Earth observations in record time (Bandara and Patel, 2022a); (Chen et al., 2021); (Shi et al., 2020). Building change detection (BCD) from satellite imagery has a critical role in shaping the character and livability of the urban environment. Methods based on pixel spectral variance analysis demonstrated poor performance in detecting changes due to limited contextual information (Liu et al., 2022). Conversely, convolutional neural networks (CNNs) have been shown to be highly capable of harvesting rich contextual information at different scales (Zheng et al., 2021). Some of the works, such as those by (Holail et al., 2023, Chen and Shi, 2020, Liang et al., 2022, Zhou et al., 2022, Shen et al., 2022, Chen et al., 2022), focused on improving feature extraction through the use of spatio-temporal attention and differential feature fusion, to overcome problems of spurious differences caused by seasonal fluctuations, building shadows, and illumination differences (Liu et al., 2021a). However, these methods run into limitations, particularly when dealing with complex backgrounds such as desert sands, which can obscure building surfaces and make detecting changes difficult. In addition, downsampling during image processing can lead to incomplete buildings and irregular borders, both of which reduce detection accuracy. Recently, transformers have been used in BCD to address the above challenges, such as in BiT (Chen et al., 2021), ChangeFormer (Bandara and Patel, 2022b), SwinTransformer (Liu et al., 2021b), and TransCD (Wang et al., 2021). These methods have a large receptive field that enables them to capture long-range dependencies and accurately detect building boundaries.

This paper delves into the investigation and evaluation of Freshly Built Locales (FBLs) using bi-temporal images through state-of-the-art computer vision networks that have been recently proposed. In addition, we have incorporated modifications to the AFDE-Net model (Holail et al., 2023) by introducing a novel residual pyramid attention fusion (RPAF) module. This enhancement enables more accurate identification of intricate details in complex change detection scenarios. Our focus lies on automatically detecting FBLs in the Nile Valley regions of Egypt, with specific attention given to New Mansoura and New Tiba, both of which have been recently established to bolster the local economy. The training set utilized in this study encompasses dual-time images obtained from Maxar data, boasting a spatial resolution of 30 cm, covering the New Mansoura and New Tiba regions. For the testing phase, we employed two

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1393-2023

images captured over New El Alamein in western Egypt, allowing us to analyze urban development between the years 2011 and 2022. The results of our study showcase the superiority of the modified AFDE-Net model, referred to as M-AFDE-Net, in detecting FBLs when compared to other models. The incorporation of the RPAF module plays a crucial role in the model's success, enabling enhanced performance and accuracy in FBL detection.

The remaining parts of this paper are organized as follows. In Section 2, the study area, dataset used, the proposed method, and evaluation metrics are listed. The experimental results is described in Section 3. Finally, some conclusions are drawn in Section 4.

2. MATERIALS AND METHODS

2.1 Study Area

This study selects New Mansoura and New Tiba as study areas because they represent areas of rapid urban development in the Nile Valley regions of Egypt. The study area is located between latitudes 24 and 32 degrees north and longitudes 28 and 34 degrees east. It is characterized by a mixture of rural buildings, farmland, and barren fields, as shown in Figure 1. In 2018, the new city of Mansoura was established to boost the economy in the Delta Nile region. The city is now covered with large-scale residential areas, highways, factories, and other surface features that can help detect changes in buildings. New Tiba is another area in the Nile Valley that is undergoing rapid development, with large construction and infrastructure projects underway.

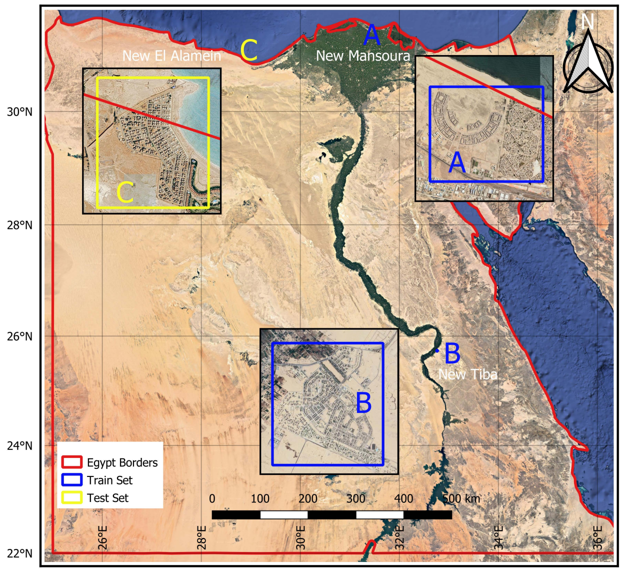
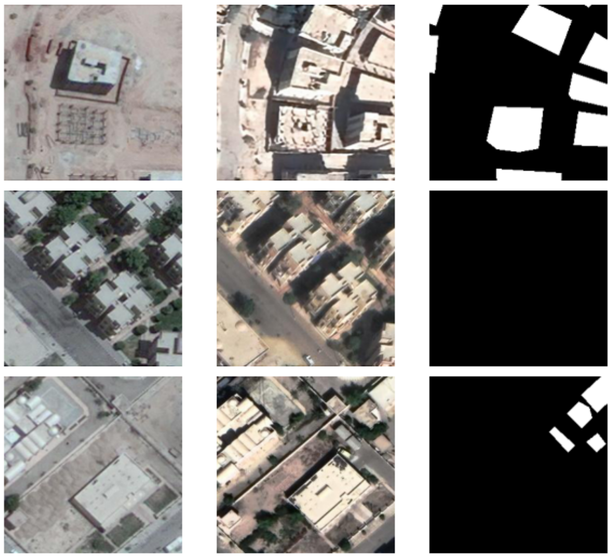


Figure 1. Rapidly developing study areas of New Mansoura and New Tiba in the Nile Valley, Egypt.

2.2 Dataset

All experiments in this study were conducted using bi-temporal images obtained from Google Earth[1] and Maxar Technologies[2] with a spatial resolution of 30 cm. For the training phase, one image was acquired on May 30, 2014, and another on June 30, 2022, specifically from the New Mansoura region. Additionally, two images were used, captured in the years 2007 and 2022, respectively, of the new city of Tiba in southern Egypt. Each image consisted of a full RGB color representation. Table 1 provides information about the data used in this paper. To prepare the data for training the deep learning models, manual annotation was performed, generating a binary format label dataset. The labels categorized pixels into two classes: buildings (1) and no buildings (0). Figure 2 illustrates sample datasets and labels used for training the deep learning models. To facilitate model training, the image patches were created with dimensions of 256 × 256 and subsequently divided into train, validation, and test sets, constituting 80%, 10%, and 10% of the dataset, respectively. For the external testing phase, a single image of the New El Alamein region, taken on July 12, 2022, was utilized. It's worth noting that all selected data had cloud cover of less than 5%, ensuring their suitability for accurate interpretation of terrestrial objects.



(a) (b) (c)

Figure 2. A sample of the dataset used, showing: (a) The reference image captured before any changes were made, (b) An image taken after the construction of new buildings, and (c) The ground truth.

	New Mansoura	New Tiba
Date of Acquisition	June 30, 2022	January 24, 2022
Image Size (Pixel)	10496 × 7680	8960 × 8448
Resolution (cm)	30	30
No. of Patches	1230	1155
Patch Size (Pixel)	256 × 256	256 × 256
Latitude (dd)	31.449508	25.733520
Longitude (dd)	31.494342	32.764698

Table 1. Remote Sensing Data used for Freshly Built Locales in the Nile Valley, Egypt.

2.3 Method

2.3.1 Overview This paper investigates the detection of Freshly Built Locales from bi-temporal images using state-of-the-art computer vision networks proposed recently. In this study, we have made modifications to the AFDE-Net (Holail et al., 2023), which incorporates the features of the differential image and ensemble spatial-channel attention fusion (ESCAF)

[1] https://earth.google.com/
[2] https://www.maxar.com/

module. To address the loss of spatial information, we have introduced a novel residual pyramid attention fusion (RPAF) module, enabling better identification of fine details in complex change detection scenarios (see Figure 3). Initially, a pair of bi-temporal images is fed into a Siamese network that shares the same parameters. The ResNet50 backbone serves as the feature extractor, providing the $X1$ and $X2$ feature maps. Subsequently, these feature maps are passed through the proposed RPAF module, resulting in updated attention feature maps $Z1$ and $Z2$. Finally, $Z1$ and $Z2$ are fed into the loss function for change prediction.

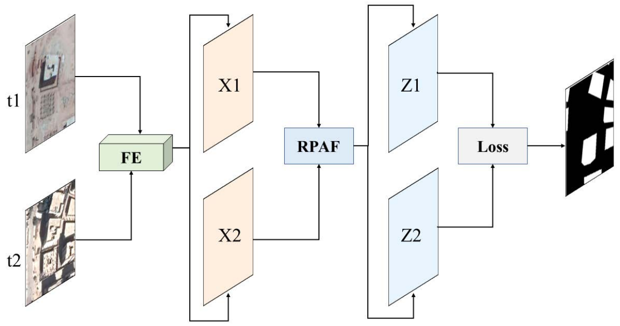


Figure 3. Overview of the proposed method for building change detection from satellite imagery.

2.3.2 Feature Extractor

For the feature extractor, we begin by inserting the input image and applying a convolutional layer, followed by a max pooling layer. This down-sampling process is then repeated four times for each Res-Block layers. Afterward, a convolutional layer is applied to each Res-Block layer, incorporating up-sampling operations, and the resulting feature maps are merged through concatenation. Finally, we apply two fully connected layers to obtain the feature tensors $X1$ and $X2$ (see Figure 4).

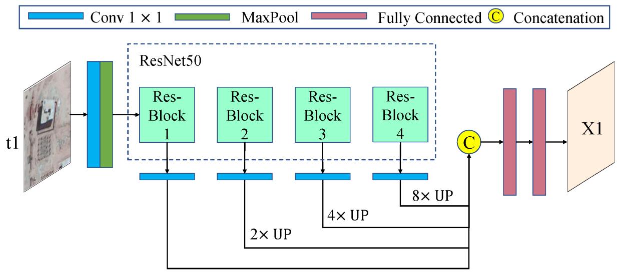


Figure 4. Multi-Layered Feature Extraction with Convolutional and ResNet Blocks.

2.3.3 Residual Pyramid Attention Fusion

In Figure 5, we combine the feature tensors $X1$ and $X2$ into a single feature tensor X. This tensor X is then divided into four pyramid branches: $P1$, $P2$, $P3$, and $P4$, corresponding to the resolutions $1 \times 1, 2 \times 2, 4 \times 4$, and 8×8, respectively. Each pyramid branch utilizes four ESCAFs to generate four new residual pyramid branches: $Y1$, $Y2$, $Y3$, and $Y4$. The purpose of ESCAF is to compensate for the loss of spatial information in deeper layers, and for more detailed information, referred to (Holail et al., 2023). Next, the four feature tensors are combined through concatenation followed by a convolutional layer operation, to obtain the tensor Y, and the first feature tensor X is added to it, resulting in updated attention feature maps Z. These Z feature maps are then split into $Z1$ and $Z2$ tensors. Finally, $Z1$ and $Z2$ are inputted into a loss function, considering resizing them to the same size as the input images, and used for predicting the changes during the training of the dataset. The enhanced AFDE-Net offers several notable advantages, including improved semantic levels and reduced loss of spatial information through the integration of RPAF and ESCAF modules.

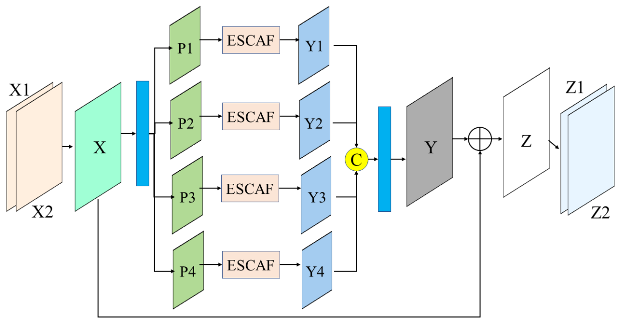


Figure 5. Integration of Residual Pyramid and ESCAF Modules for Improved Feature Extraction and Change Prediction.

2.4 Implementation Details

The experiments in this study were conducted on a virtual machine desktop running a 64-bit Windows 10 Pro operating system. The software configuration included the Python programming language, PyTorch 1.7.1, CUDA 10.1, and cuDNN 7.6.1. To ensure optimal hardware capabilities, we utilized an NVIDIA GRID RTX8000-8Q with 8GB memory, an Intel(R) Xeon(R) CPU E5-2687W v4 @ 3.00GHz, and 32.0 GB of GPU memory. In order to achieve model convergence, all methods were trained for 200 epochs using the Adam optimizer with an initial learning rate of 0.0001. The input images were resized to dimensions of 256×256, and a batch size of 8 was employed. The cross-entropy loss was used when adjusting model weights during training. Validation was performed after each training epoch to assess the model's effectiveness. The best-performing model on the validation set was saved and subsequently used for evaluation on the test set.

2.5 Evaluation Metrics

To compare the ground truth and predicted change map, five metrics were utilized to validate the accuracy and effectiveness of the proposed method. These metrics include Recall (R), Precision (P), IoU, Overall Accuracy (OA); and F1-score. The formulas for these metrics are provided below:

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \\ \text{Precision} &= \frac{TP}{TP + FP} \\ \text{OA} &= \frac{TP + TN}{TP + FN + TN + FP} \\ \text{IoU} &= \frac{TP}{TP + FP + FN} \\ \text{F1-score} &= \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \end{aligned} \tag{1}$$

where TP = total number of true positives for all classes
FN = false negatives
FP = false positives
TP = true positives

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1393-2023

3. EXPERIMENTAL RESULTS

3.1 Comparison Methods

Five network models, namely DASMNet (Shi et al., 2021), DASNet (Chen et al., 2020), SNUNet (Fang et al., 2021), BiT (Chen et al., 2021), and ChangeFormer (Bandara and Patel, 2022b), were employed to assess their performance in detecting changes in FBLs. SNUNet and DASNet are CNN methods that extract features from bi-temporal images using downsampling, max-pooling, and convolutional layers. They employ a Siamese architecture with weight sharing and incorporate dense skip connections for preserving fine-grained features. DSAMNet, on the other hand, uses convolutional block attention modules and a deep supervision module to enhance learning. It employs a metric module based on deep metric learning to learn change maps, capturing long-term dependencies and improving feature representation. Transformer-based models like BiT and ChangeFormer focus on modeling context in space-time using tokens that represent input images. They use convolutional blocks to obtain feature maps, which are converted into semantic tokens and processed by the transformer to leverage global semantic relations. A Siamese transformer decoder is employed to optimize the context-rich tokens at the pixel level, resulting in improved change detection compared to purely CNN-based structures.

3.2 Overall Performance

Generally, a higher F1-score indicates better model performance, as shown in Figure 6. The proposed M-AFDENet ranked first as the most stable from epoch 60 and achieved an F1-score of about 89.2%, which is higher than the other models. It was followed by SNUNet and AFDENet. For DASMNet, its lowest value is about 77.9%, and the F1 score curve of DASMNet fluctuates wildly, with its value being the lowest among the models.

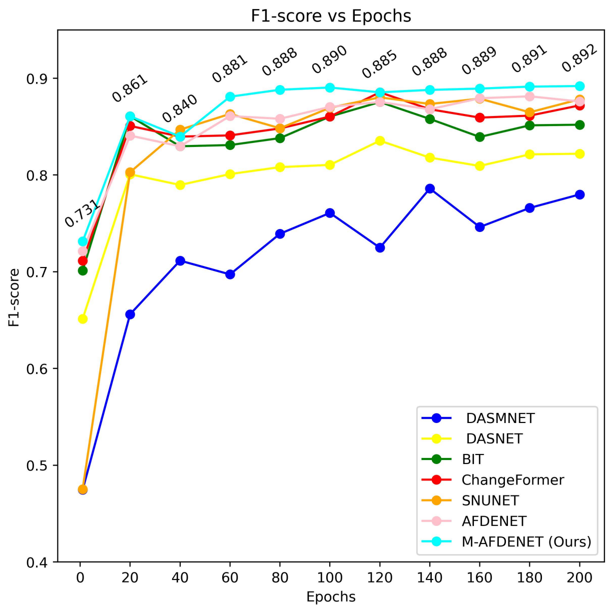


Figure 6. Tracking F1-score across 200 epochs during cross-validation training runs.

Table 2 clearly illustrates the performance of different models. Notably, M-AFDENet stands out with the highest precision (P) value of 91.5%, indicating its superior ability to accurately predict change regions compared to the total samples predicted as truth. Following closely behind are BIT, ChangeFormer and SNUNet with precision values of 91.1%, 90.8%, and 90.7% respectively. In contrast, DASMNet exhibits the lowest precision value among the models. When it comes to recall (R) values, M-AFDENet secures the first-highest recall score, while DASMNet lags behind with a score of 75.6%, indicating room for improvement. BIT comes in second with a score of 85.5%, showcasing its exceptional capability to correctly identify change regions. The F1-score, a weighted harmonic mean of precision and recall, provides a comprehensive evaluation of the models. Here, M-AFDENet takes the lead with an impressive F1-score of 89.2%, outperforming the second-place model, SNUNet, by 1.4 points. These metrics collectively demonstrate M-AFDENet's optimal performance, as it not only secures the top rank in F1-score and IoU but also achieves a commendable second place in recall.

Method	OA	IoU	F1	P	R
DASMNet	94.3	69.8	77.9	80.4	75.6
DASNet	96.5	81.6	82.2	89.6	75.9
BIT	96.9	83.1	85.2	<u>91.1</u>	<u>85.5</u>
ChangeFormer	<u>97.1</u>	83.3	87.2	90.8	83.9
AFDENet	96.4	84.1	87.6	90.4	84.9
SNUNet	96.3	<u>84.5</u>	<u>87.8</u>	90.7	85.0
M-AFDENet (Ours)	**97.7**	**85.9**	**89.2**	**91.5**	**87.0**

Table 2. Performance Analysis of Different Models on Dataset used for Building Change Detection. A bold color indicates the best result, while the underlined values indicate a close second.

3.3 Freshly Built Locales (FBL) Detection

Figure 7 presents a comprehensive visual comparison of the change maps obtained from the evaluated methods, offering a more intuitive understanding of their performance. In the first row, it is apparent that the majority of the methods successfully detect a previously overlooked changed building, which was not identified in the ground truth. Notably, our proposed M-AFDENet model exhibits superior performance in capturing the edges within the change region, resulting in sharper boundaries, as exemplified in column (j). Additionally, the buildings depicted in the change maps closely resemble the ground truth, highlighting the effectiveness of the residual pyramid attention fusion (RPAF) module. By leveraging global spatio-temporal context, this module excels in recognizing fine details of color changes. ChangeFormer demonstrates commendable performance in detecting sharp edges. However, it is prone to false alarms, which is a common challenge encountered by other models as well. Similarly, the BiT model is more susceptible to variations in the color of building surfaces, as illustrated in column f of the third row. Despite incorporating the transformer mechanism, which has exhibited remarkable capabilities in hyperspectral image classification, both the BiT model and ChangeFormer manifest false positive detections within our experimental analysis.

3.4 FBL of the New El Alamein region

The New El Alamein City project has held significant importance since its inception, serving as a crucial opportunity to address the issue of overcrowding in Egypt by leveraging the northern coast as a residential destination and attracting year-round tourism. It is imperative to ensure the proper monitoring of Freshly Built Locales (FBLs) for sustainable growth. In

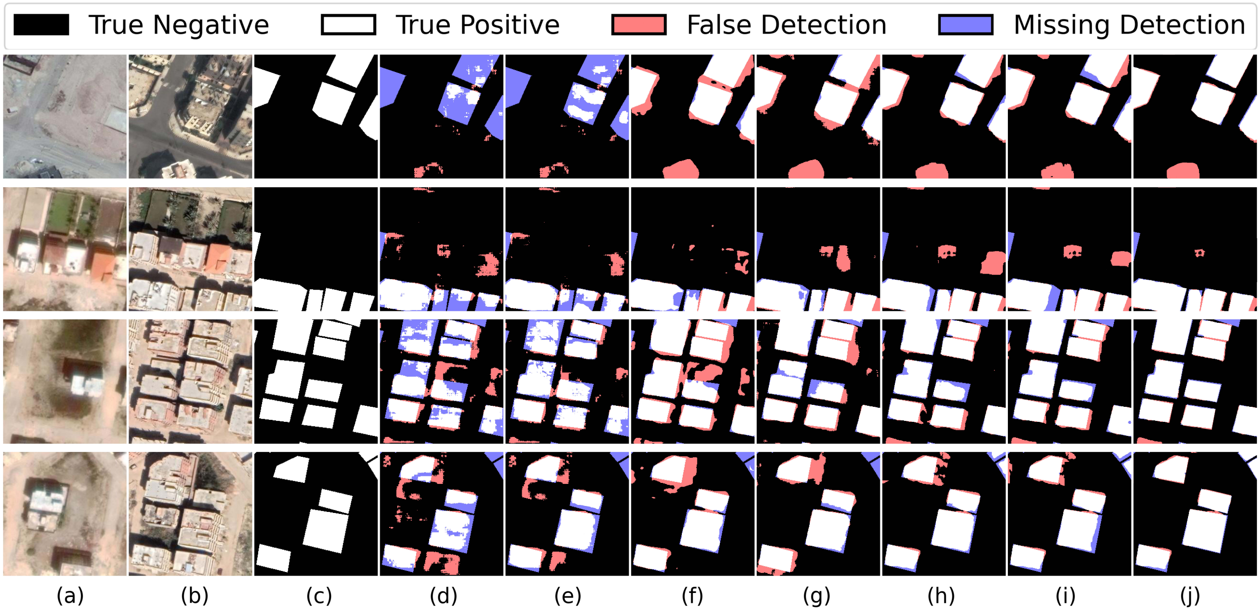


Figure 7. Showcases a variety of representative change detection results. Sub-figures (a) to (j) present the pre-image, post-image, ground truth, prediction results obtained from DASMNet, DASNet, BIT, ChangeFormer, AFDENet, SNUNet, and M-AFDENet (Ours), respectively.

line with our proposed method, we conducted change detection experiments in the North Coast region of New El Alamein, spanning from 2011 to 2022. The change detection results are presented in Figure 8. Overall, the study area exhibits extensive expansion of FBLs. Upon comparison with the original 2011 image, it becomes evident that the majority of the detected areas were previously barren desert lands or comprised only a few relatively small-sized chalets. This transformation underscores the significant development and growth that has taken place in the region over the assessed period, signifying the successful realization of the New El Alamein City project and its contribution to overcoming overcrowding challenges in Egypt.

4. CONCLUSIONS

In this paper, we undertake a comprehensive evaluation and comparison of the change detection performance exhibited by several state-of-the-art models, namely DASMNet, DASNet, AFDENet, SNUNet, BiT, and ChangeFormer. We utilize high-resolution satellite imagery acquired in the Nile Valley regions of Egypt as the primary data source for our analysis. To enhance the change detection capabilities further, we propose a novel residual pyramid attention fusion (RPAF) module. This module is integrated into the AFDENet model, effectively addressing the issue of spatial information loss. Through rigorous experimentation and evaluation, taking into account metrics such as the F1 score and precision, we arrive at a definitive conclusion. The modified M-AFDE-Net model outperforms all competing models under the specific experimental conditions we set forth. In order to provide a practical application of our findings, we proceed to apply the modified M-AFDE-Net model to a dataset encompassing the years 2011 and 2022. Specifically, we aim to identify Freshly Built Locales (FBLs) in the coastal area of New El Alamein City, covering a substantial area of 1.15 km2. The resulting prediction from our analysis reveals the presence of 136 newly built structures. Notably, these structures encompass a diverse range of functionalities, including administrative buildings, youth housing, hospitals, and tourist chalets. By

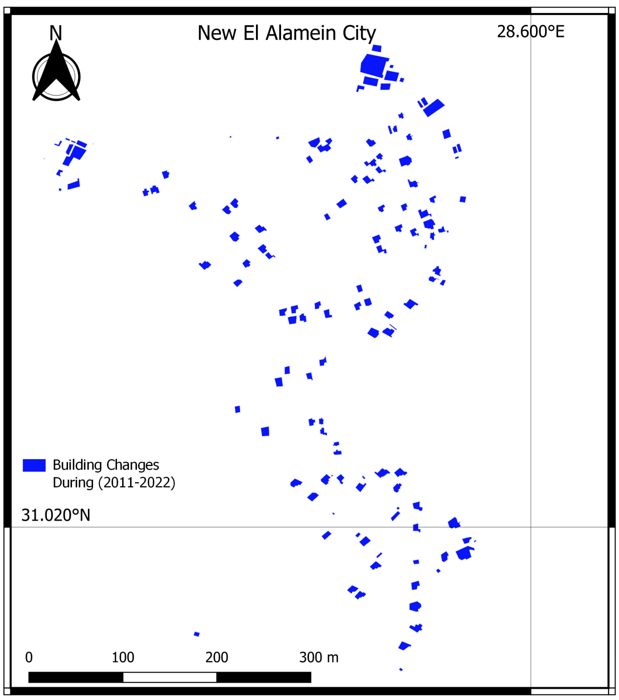


Figure 8. The map of Freshly Built Locales of the coastal area of New El Alamein City from 2011 to 2022, detected by our proposed method.

conducting this comprehensive evaluation, proposing innovative enhancements, and providing practical application results, our research significantly contributes to the field of change detection from high-resolution satellite imagery, specifically in the context of the Nile Valley regions of Egypt.

ACKNOWLEDGMENT

This paper was supported by the National Natural Science Foundation of China (Grant Nos. 42101449, 42090012, 61825103), the Natural Science Foundation of Hubei Province, China (Grant Nos. 2022CFB773, 2020CFA001), the Key Research and Development Project of Jinzhong City, China (Grant No. Y211006), the Key Research and Development Program of Hubei Province, China (Grant No. 2022BAA048), the Chutian Scholar Program of Hubei Province, the Yellow Crane Talent Scheme and LIESMARS Special Research Funding.

REFERENCES

Bandara, W. G. C., Patel, V. M., 2022a. Revisiting consistency regularization for semi-supervised change detection in remote sensing images. *arXiv preprint arXiv:2204.08454*.

Bandara, W. G. C., Patel, V. M., 2022b. A transformer-based siamese network for change detection. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 207–210.

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

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

Chen, J., Fan, J., Zhang, M., Zhou, Y., Shen, C., 2022. MSF-Net: A Multiscale Supervised Fusion Network for Building Change Detection in High-Resolution Remote Sensing Images. *IEEE Access*, 10, 30925–30938.

Chen, J., Yuan, Z., Peng, J., Chen, L., Huang, H., Zhu, J., Liu, Y., Li, H., 2020. DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 1194–1206.

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

Holail, S., Saleh, T., Xiao, X., Li, D., 2023. AFDE-Net: Building Change Detection Using Attention-Based Feature Differential Enhancement for Satellite Imagery. *IEEE Geoscience and Remote Sensing Letters*, 20(6006405), 1-5.

Liang, Z., Zhu, B., Zhu, Y., 2022. High resolution representation-based Siamese network for remote sensing image change detection. *IET Image Processing*, 16(9), 2506–2517.

Liu, J., Zhang, W., Liu, F., Xiao, L., 2021a. A probabilistic model based on bipartite convolutional neural network for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–14.

Liu, M., Shi, Q., Chai, Z., Li, J., 2022. PA-Former: learning prior-aware transformer for remote sensing building change detection. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021b. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Shen, Q., Huang, J., Wang, M., Tao, S., Yang, R., Zhang, X., 2022. Semantic feature-constrained multitask siamese network for building change detection in high-spatial-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 189, 78–94.

Shi, Q., Liu, M., Li, S., Liu, X., Wang, F., Zhang, L., 2021. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE transactions on geoscience and remote sensing*, 60, 1–16.

Shi, W., Zhang, M., Zhang, R., Chen, S., Zhan, Z., 2020. Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote Sensing*, 12(10), 1688.

The National, 2020. Egypt's rapid urban encroachment a hot topic after ei sisi threatens to resign.

Wang, Z., Zhang, Y., Luo, L., Wang, N., 2021. TransCD: scene change detection via transformer-based architecture. *Optics Express*, 29(25), 41409–41427.

Zheng, Z., Zhong, Y., Wang, J., Ma, A., Zhang, L., 2021. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. *Remote Sensing of Environment*, 265, 112636.

Zhou, Y., Wang, F., Zhao, J., Yao, R., Chen, S., Ma, H., 2022. Spatial-Temporal Based Multihead Self-Attention for Remote Sensing Image Change Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10), 6615–6626.