BTCDNet: Bayesian Tile Attention Network for Hyperspectral Image Change Detection

Junshen Luo, Jiahe Li, Xinlin Chu, Sai Yang, Lingjun Tao, and Qian Shi, *Senior Member, IEEE*

***Abstract*— Hyperspectral images (HSIs) provide detailed spectral information, which are effective for change detection (CD). Prior knowledge has been proven to improve the robustness of models in HSI processing. However, current CD methods do not fully use prior knowledge, and research on hyperspectral mangroves' CD is limited. In this letter, we propose a general hyperspectral CD model with Bayesian prior guided module (BPGM) and tile attention block (TAB) called BTCDNet. BPGM leverages prior information to steer the model training process under limited labeled samples condition, while TAB can reduce complexity and improve performance by tile attention. Moreover, a novel and restricted hyperspectral CD dataset Shenzhen has been annotated for hyperspectral mangroves' CD reference. Experiments demonstrate that our proposal achieves state-of-the-art (SOTA) performances on this dataset and two other public benchmark datasets. Our code and datasets are available at https://github.com/JeasunLok/BTCDNet**

***Index Terms*— Attention, Bayesian, change detection (CD), deep learning (DL), hyperspectral image (HSI), mangroves, prior knowledge.**

I. Introduction

AS ONE of the most common methods to capture changes on land surface, change detection (CD) has a great potential in monitoring, especially with rapid development of hyperspectral images (HSIs) which have continuous spectral bands and fine spectral resolution.

Traditional HSI CD methods can be generally categorized into three types.

1) Algebra-based methods, represented by change vector analysis (CVA) [1], is computationally efficient but sensitive to noise.
2) Transformation-based methods, such as principal component analysis (PCA) [2], mitigate noise problems by transforming the data into a fresh feature space but requires threshold selection manually.
3) Classification-based methods, use machine learning models such as support vector machine (SVM) for CD [3], but their performances depend heavily on classification accuracies.

Received 16 January 2025; revised 6 April 2025; accepted 9 April 2025. Date of publication 24 April 2025; date of current version 9 May 2025. This work was supported by the National Natural Science Foundation of China under Grant 42222106. *(Corresponding author: Qian Shi.)*

The authors are with Guangdong Provincial Key Laboratory for Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: luojsh7@mail2.sysu.edu.cn; lijh335@mail2.sysu.edu.cn; chuxlin@mail2.sysu.edu.cn; yangs37@mail2.sysu.edu.cn; taoly3@mail2.sysu.edu.cn; shixi5@mail.sysu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2025.3563897

With the development of deep learning (DL), models with powerful feature extraction capabilities have contributed a new outlook to HSI CD. Convolutional neural network (CNN)-based HSI CD methods have emerged [4], and recurrent neural network (RNN) excelling at handling sequential data has also obtained a good performance [5]. Furthermore, methods combining CNN and RNN have shown excellent results [6]. Recently, the success of transformer with self-attention mechanisms has inspired researchers to explore its applications in HSI CD [7], [8], [9], resulting in notable improvements. However, these methods rely on sufficient labeled samples and tend to overemphasize the feature extraction ability, while underestimating the strength of prior knowledge in HSI processing. This is particularly evident in anomaly detection, where its significance has been well-established [10], yet existing DL CD methods fail to fully incorporate it.

Mangroves are crucial for ecosystem services and coastal livelihoods, but highly vulnerable to human activities and climate change [11], which makes monitoring and protecting them an urgent need. However, current researches on mangroves' CD mainly depend on traditional methods using optical or multispectral data [12]. The use of DL models for hyperspectral mangroves' CD, including dataset construction, remains underexplored.

In this letter, we propose a Bayesian tile attention network called BTCDNet, which composes of Bayesian prior guided module (BPGM) and tile attention block (TAB), to explore the capabilities of DL models with prior knowledge for hyperspectral CD in complex scenarios and under limited labeled samples condition especially in mangroves. The main contributions of our work are as follows.

1) To the best of our knowledge, it is the first attempt to annotate and release a novel and restricted hyperspectral mangroves' CD dataset with high-quality annotations. The dataset called Shenzhen addresses the gap in hyperspectral mangroves' CD data and supports future researches.
2) From a novel Bayesian perspective, BPGM reexamines feature extraction in DL methods. After leveraging normalized spectral angle (SA) as prior knowledge, we model the likelihood probability at each layer, which is then used in Bayesian formula to compute the posterior probability as the next prior probability iteratively.
3) TAB is used in the proposal to reduce the computational cost of self-attention mechanisms, achieving better performance while lowering the complexity.

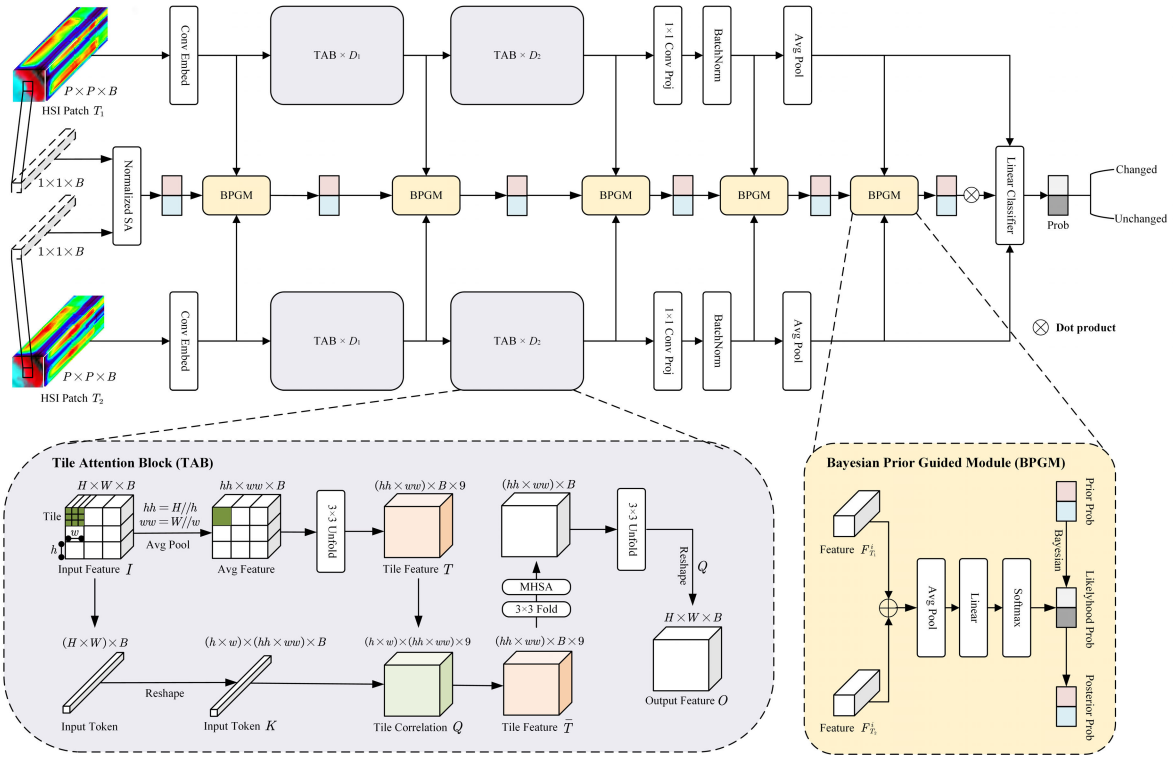

Fig. 1. Overall architecture of BTCDNet. Two HSI patches separately pass through a convolutional embedding layer, several TAB layers, and a convolutional projection layer with batch normalization and global average pooling to extract features. At the same time, two center pixel vectors from both the HSI patches produce initial prior probability by normalized SA, and then compute posterior probability by BPGM iteratively. Eventually, a fully connected layer is used to generate CD probability, incorporating the probability from the final BPGM by Bayesian formula then producing the ultimate CD result.

Section II provides a detailed explanation of our proposed method. The datasets we used including our annotated hyperspectral mangroves' CD dataset Shenzhen, and the experimental results are presented in Section III. Eventually, the conclusions are drawn in Section IV.

II. Proposed Approach

A. Overall Architecture

The overall architecture of BTCDNet is shown in Fig. 1. First, two center pixel vectors from HSI patches of T_1 and T_2 are extracted to calculate normalized SA as initial prior probability. Next, the two HSI patches are separately processed by 3×3 convolutional embeddings for the first step of feature extraction. The extracted features are then successively passed through two TAB modules with depths D_1 and D_2, followed by an 1×1 convolution layer and a batch normalization layer to obtain advanced information. A global average pooling layer and a fully connected layer are subsequently applied to produce $P_{\text{Classifier}}$ eventually.

During the above-mentioned process, aside from the initial step where the normalized SA is used as the prior probability, the feature extraction results from each step are modeled as the likelihood probability by BPGM. The posterior probability from the previous step is iteratively considered as the prior probability for the current step, and the Bayesian formula is applied to compute the posterior probability P_{BPGM} at each stage. Finally, $P_{\text{Classifier}}$ and the ultimate P_{BPGM} are combined via dot product with Temperature coefficient to produce the CD outcome probability $\mathbf{P}$

$$\mathbf{P} = \text{Softmax}\left(\frac{P_{\text{BPGM}}}{\text{Temperature}} \cdot P_{\text{Classifier}}\right). \tag{1}$$

B. Tile Attention Block

TAB receives the input feature I with size $H \times W \times B$ and generates tile feature T by an average pooling layer and a 3×3 unfold layer with tile size $h \times w$, and I is transformed to token K meanwhile. Tile correlation Q is obtained by an attention-like way, defined as

$$Q = \text{Softmax}\left(\frac{K T^{\text{T}}}{\sqrt{d}}\right). \tag{2}$$

Subsequently, the deeper tile feature $\bar{T}$ is updated by the original tile feature T and the tile correlation Q

$$\bar{T} = K Q^{\text{T}}. \tag{3}$$

Since $\bar{T}$ is the simplified local representation of I, multihead self-attention (MHSA) is adopted to acquire global features

$$\text{MHSA}(\bar{T}) = \bigsqcup_i \left\{ \text{Softmax}\left[\frac{(W_Q^i \bar{T})(W_K^i \bar{T})^{\text{T}}}{\sqrt{d}}\right] W_V^i \bar{T} \right\} \tag{4}$$

where W_K^i, W_Q^i, and W_V^i indicate parameters of the ith self-attention head, $\sqcup_i$ denotes the concatenation operator of MHSA, and d denotes the dimension of input features.

Eventually, the output feature O of TAB is obtained by

$$O = \text{MHSA}(\bar{T}) Q^{\text{T}}. \tag{5}$$

TABLE I
DATASETS' INFORMATION

Dataset	Shenzhen	Farmland	Hermiston
Size	231×209	420×140	390×200
Spectral Bands	166	154	242
Date T_1	November 20^{th}, 2021	May 3^{rd}, 2006	May 1^{st}, 2004
Date T_2	December 10^{th}, 2022	April 23^{rd}, 2007	May 8^{th}, 2007
Changed Samples	426	18383	9986
Unchanged Samples	7047	40417	68014

C. Bayesian Prior Guided Module

The initial prior probability P^1_{prior} is based on normalized SA, which is calculated by

$$P^1_{\text{prior}} = 1 - \text{Normalized SA}\left(x_{T_1}, x_{T_2}\right) \tag{6}$$

$$\text{Normalized SA}\left(x_{T_1}, x_{T_2}\right) = \frac{\arccos\left(\frac{x_{T_1} \cdot x_{T_2}}{\|x_{T_1}\| \|x_{T_2}\|}\right)}{\pi} \tag{7}$$

where x_{T_1} and x_{T_2} denote the center pixel vectors of T_1 and T_2, respectively.

As mentioned in Section II-A, the features $F^i_{T_1}$ and $F^i_{T_2}$ from the ith feature extraction step are added, then passing through an average pooling layer and a fully connected layer in sequence to model P^i_{like}. We use Bayesian formula to calculate P^i_{post}, which also acts as P^{i+1}_{prior} for the next BPGM

$$P^i_{\text{post}} = \frac{P^i_{\text{like}} \cdot P^i_{\text{prior}}}{\left(1 - P^i_{\text{like}}\right) \cdot \left(1 - P^i_{\text{prior}}\right) + P^i_{\text{like}} \cdot P^i_{\text{prior}}} \tag{8}$$

$$P^{i+1}_{\text{prior}} = P^i_{\text{post}} \tag{9}$$

where P^i_{post} denotes the ith posterior probability by the ith BPGM.

III. EXPERIMENTS

A. Datasets

The detailed information of all the datasets are shown in Table I and Fig. 2. Acquired by the Hyperion sensor aboard EO-1, the Farmland and Hermiston datasets are public benchmark datasets which can be found in rslab.ut.ac.ir To explore the potentials of novel methods in hyperspectral mangroves' CD, we annotate Shenzhen dataset, which is located at Shenzhen, Guangdong Province, China, and collected by the AHSI sensor aboard ZY1E. In addition, the Shenzhen dataset undergoes preprocessing, including radiometric calibration and atmospheric correction.

B. Implementation Details

To further demonstrate the effectiveness of our proposed BTCDNet, we used six state-of-the-art (SOTA) CD methods for comparison including CVA [1], GETNET [4], ReCNN [6], BIT [7], CSANet [8], and SSTFormer [9].

For each method, only 100 samples from the changed and unchanged categories are used for training, and all the remaining samples are reserved for testing. All the datasets are split using a fixed random seed, and the accuracy results of each method correspond to the highest accuracy achieved over multiple repeated experiments. Besides, we use Adam algorithm

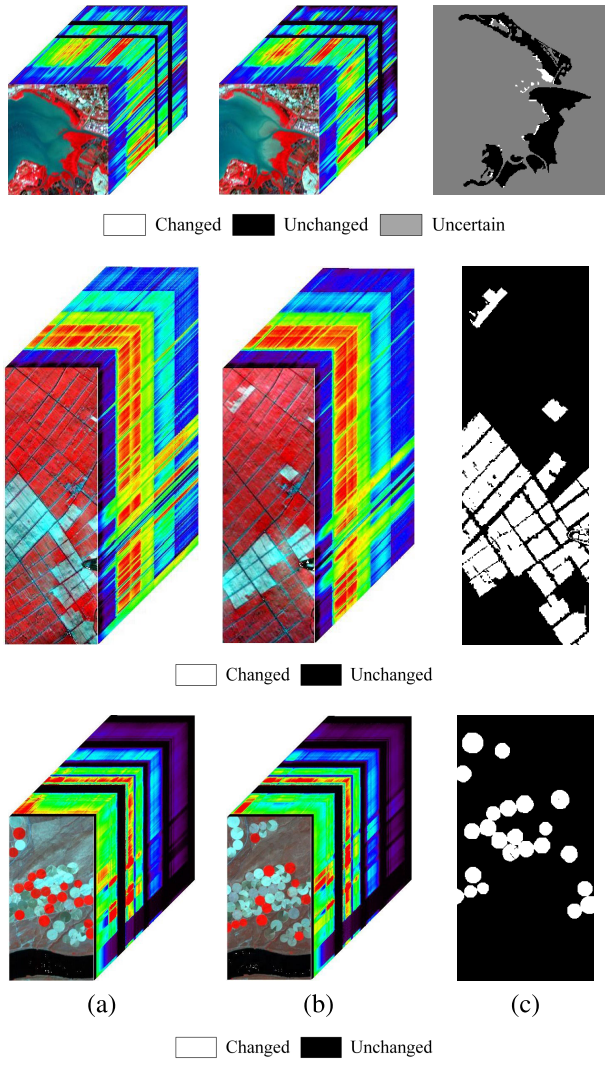

Fig. 2. (Top to bottom) Pseudocolor map of the Shenzhen dataset, Farmland dataset, and Hermiston dataset for CD. (a) HSI T_1. (b) HSI T_2. (c) Ground truth.

to minimize the cross-entropy loss with four NVIDIA GeForce RTX 2080Ti GPUs, and epoch, batch size, learning rate, the minimum of learning rate, weight decay, patch size, and Temperature coefficient are set to 200, 64, 1×10^{-4}, 1×10^{-6}, 0, 5, and 0.5, respectively. The depths of the two TABs (D_1 and D_2) are both set to 2. Two tile sizes and two embedding dimensions of our proposed BTCDNet are set to 2×2 and 1×1 and 96 and 192, respectively. One thing should be noted that all the hyperparameters including Temperature coefficient and two tile sizes were optimally set based on sensitivity analyses.

C. Experimental Results and Comparison

Overall accuracy (OA), kappa coefficient (Kappa), Recall (Rec), Precision (Pre), $F1$ Score ($F1$), model parameters (Params), floating point operations (FLOPs), training time (TT), and inference time (IT) are used to validate all the methods. The quantitative assessments are shown in Tables II–IV, while the visualizations are shown in Figs. 3–5.

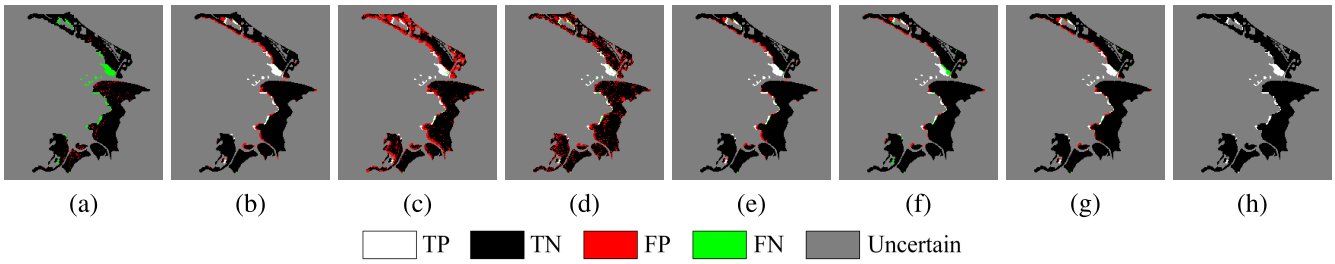

Fig. 3. CD maps of different compared methods and our proposed BTCDNet on the Shenzhen dataset. (a) CVA. (b) GETNET. (c) ReCNN. (d) BIT. (e) CSANet. (f) SSTFormer. (g) BTCDNet. (h) Ground truth.

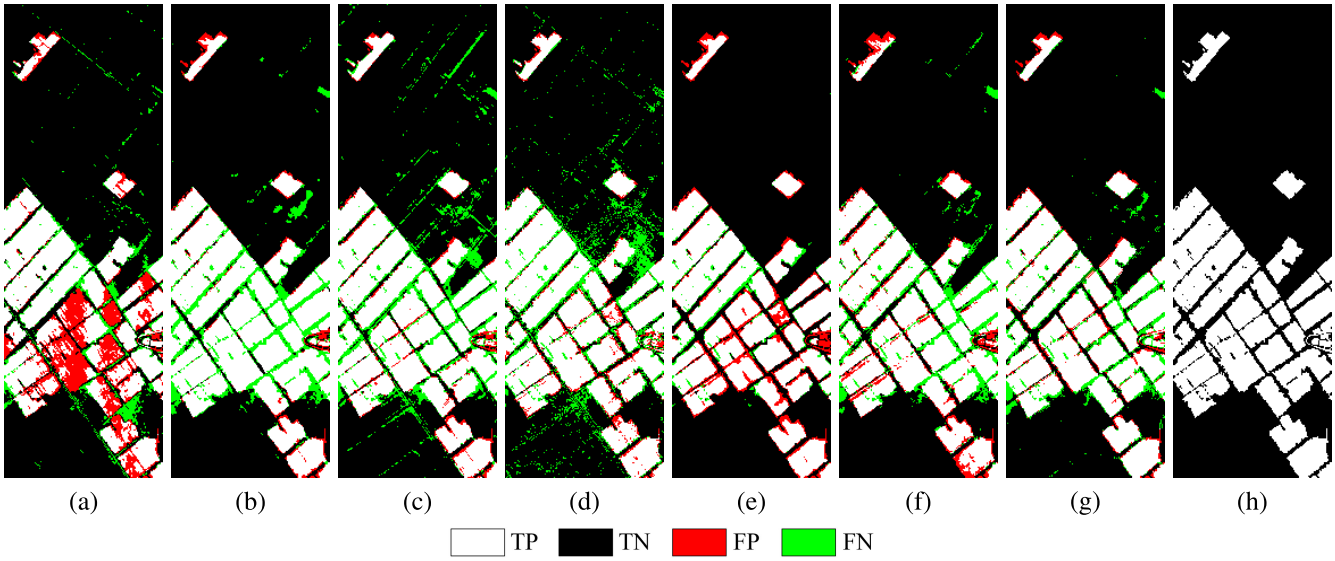

Fig. 4. CD maps of different compared methods and our proposed BTCDNet on the Farmland dataset. (a) CVA. (b) GETNET. (c) ReCNN. (d) BIT. (e) CSANet. (f) SSTFormer. (g) BTCDNet. (h) Ground truth.

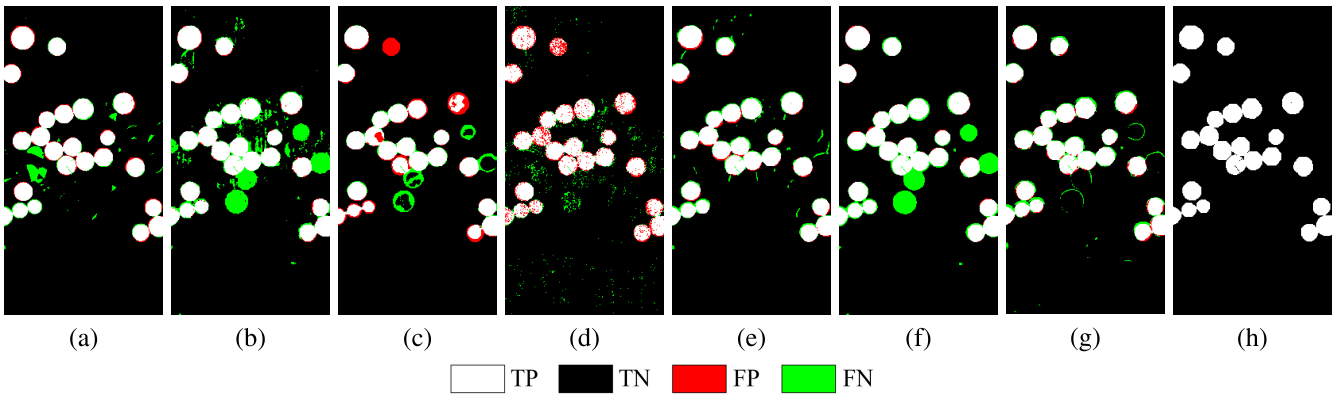

Fig. 5. CD maps of different compared methods and our proposed BTCDNet on the Hermiston dataset. (a) CVA. (b) GETNET. (c) ReCNN. (d) BIT. (e) CSANet. (f) SSTFormer. (g) BTCDNet. (h) Ground truth.

Several conclusions can be drawn from the results. CVA, as a traditional method, performs well on the Hermiston dataset but struggles in complex scenarios, particularly on the Shenzhen dataset where it exhibits severe errors. GETNET, ReCNN, and CSANet, which primarily use convolution for feature extraction, achieve high Rec but overlook Pre. In contrast, BIT and SSTFormer relying on self-attention mechanisms exhibit the opposite behavior. BTCDNet achieves a promising balance, attaining the best accuracies across all the three datasets. Compared with the closest performing methods, CSANet and SSTFormer, it shows a significant reduction in Params and FLOPs, as well as a better efficiency in the whole training and inference process.

Due to our restrictive use of only 100 changed and unchanged samples for model training, the results on all the datasets come across considerable noise by all the methods. However, BTCDNet's tile attention effectively addresses this issue without the spatial aggregation errors seen in GET-

TABLE II

QUANTITATIVE ASSESSMENT OF DIFFERENT COMPARED METHODS AND OUR PROPOSED BTCDNET ON THE SHENZHEN DATASET

Method	OA(%)	Kappa	Rec	Pre	F1	Params(M)	FLOPs(M)	TT(s)	IT(s)
CVA	91.66	-0.028	0.007	0.015	0.010	-	-	-	-
GETNET	93.44	0.516	0.880	0.400	0.546	**0.290**	16.672	26	10
ReCNN	76.71	0.199	**0.917**	0.152	0.261	0.603	**3.870**	**25**	**8**
BIT	80.86	0.214	0.802	0.164	0.272	12.498	71.292	72	25
CSANet	95.00	0.589	0.883	0.469	0.613	40.427	832.996	230	15
SSTFormer	94.90	0.528	0.706	0.455	0.554	2.538	944.780	66	66
BTCDNet	**95.05**	**0.591**	0.880	**0.472**	**0.615**	1.567	78.946	60	27

TABLE III

QUANTITATIVE ASSESSMENT OF DIFFERENT COMPARED METHODS AND OUR PROPOSED BTCDNET ON THE FARMLAND DATASET

Method	OA(%)	Kappa	Rec	Pre	F1	Params(M)	FLOPs(M)	TT(s)	IT(s)
CVA	79.22	0.707	0.741	0.850	0.792	-	-	-	-
GETNET	91.68	0.815	**0.956**	0.811	0.878	**0.290**	16.672	37	10
ReCNN	90.44	0.788	0.937	0.794	0.859	0.603	**3.870**	**35**	**9**
BIT	90.20	0.784	0.944	0.787	0.858	12.498	71.292	100	29
CSANet	93.81	0.852	0.856	**0.940**	0.896	40.427	832.996	296	19
SSTFormer	91.42	0.802	0.883	0.848	0.865	2.538	944.780	102	40
BTCDNet	**93.73**	**0.856**	0.932	0.875	**0.903**	1.567	78.946	90	33

TABLE IV

QUANTITATIVE ASSESSMENT OF DIFFERENT COMPARED METHODS AND OUR PROPOSED BTCDNET ON THE HERMISTON DATASET

Method	OA(%)	Kappa	Rec	Pre	F1	Params(M)	FLOPs(M)	TT(s)	IT(s)
CVA	97.57	0.894	0.934	0.883	0.908	-	-	-	-
GETNET	94.91	0.799	0.964	0.726	0.828	**0.290**	16.672	51	19
ReCNN	96.10	0.821	0.828	0.860	0.844	0.603	**3.870**	**49**	**18**
BIT	97.07	0.860	0.873	0.881	0.877	12.498	71.292	121	45
CSANet	97.63	**0.896**	**0.941**	0.881	**0.910**	40.427	832.996	256	32
SSTFormer	95.59	0.825	0.982	0.749	0.850	2.538	944.780	129	58
BTCDNet	**97.67**	**0.896**	0.921	**0.898**	**0.910**	1.567	78.946	111	49

TABLE V

ABLATION STUDIES OF OUR PROPOSED BTCDNET ON DIFFERENT DATASETS

Dataset	TAB	BPGM	OA(%)	Kappa	Rec	Pre	F1	Params(M)	FLOPs(M)
Shenzhen	×	×	94.16	0.551	**0.893**	0.427	0.578	**1.566**	82.013
	×	✓	94.43	0.563	**0.893**	0.440	0.590	**1.566**	82.040
	✓	×	94.34	0.555	0.877	0.435	0.581	1.567	**78.917**
	✓	✓	**95.05**	**0.591**	0.880	**0.472**	**0.615**	1.567	78.946
Farmland	×	×	92.00	0.820	0.931	0.833	0.879	**1.566**	82.013
	×	✓	**93.78**	**0.857**	0.927	**0.880**	**0.903**	**1.566**	82.040
	✓	×	92.07	0.821	**0.933**	0.833	0.880	1.567	**78.917**
	✓	✓	93.73	0.856	0.932	0.875	**0.903**	1.567	78.946
Hermiston	×	×	94.36	0.770	0.903	0.722	0.803	**1.566**	82.013
	×	✓	96.88	0.859	0.873	0.880	0.877	**1.566**	82.040
	✓	×	95.03	0.801	**0.950**	0.736	0.829	1.567	**78.917**
	✓	✓	**97.67**	**0.896**	0.921	**0.898**	**0.910**	1.567	78.946

NET, ReCNN, and SSTFormer on the Hermiston dataset (see Fig. 5).

In particular, on the Shenzhen dataset, where most methods fail due to limited labeled samples and high complexity, large areas of misclassification occur. This is because the changes in mangroves are slight and subtle. As shown in Fig. 3, BTCDNet alleviates this problem to some extent, which strongly demonstrates the superiority of prior knowledge.

D. Ablation Studies

We conduct ablation studies to validate the effectiveness of our proposed TAB and BPGM, and the results are presented in Table V. BTCDNet without TAB adopts MHSA in standard transformer attention as a replacement for comparison. It can be observed that the model without proposed modules obtains poor performance on all the datasets especially in Pre; however, the addition of BPGM significantly boosts the Pre of the original model with negligible increasing computational cost, strongly highlighting the advantages of prior knowledge. Integrating TAB into the model results in both reduced computational cost and improved accuracy. BTCDNet with both TAB and BPGM effectively balances Pre and Rec, attaining the highest accuracies on the Shenzhen and Hermiston datasets, with only a very slight lag on the Farmland dataset.

IV. CONCLUSION

In this letter, we have proposed BTCDNet for hyperspectral CD especially focused on mangroves. BTCDNet consists of TAB and BPGM. The former reduces the complexity of computing self-attention by adopting a tile approach, while the latter provides a Bayesian prior perspective to guide the model's training. The effectiveness of BTCDNet has been demonstrated on a novel and restricted annotated hyperspectral mangroves' CD dataset Shenzhen and two publicly available benchmark datasets under limited labeled samples' condition. The Bayesian prior perspective may be further explored in few-shot even zero-shot learning in the future.

REFERENCES

[1] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, "Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4363–4378, Aug. 2015.

[2] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008.

[3] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.

[4] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2018.

[5] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, p. 506, Jun. 2016.

[6] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral–spatial–temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.

[7] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5607514.

[8] R. Song, W. Ni, W. Cheng, and X. Wang, "CSANet: Cross-temporal interaction symmetric attention network for hyperspectral image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[9] Y. Wang et al., "Spectral–spatial–temporal transformers for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536814.

[10] X. Shen, H. Liu, J. Nie, and X. Zhou, "Matrix factorization with framelet and saliency priors for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504413.

[11] A. F. Rahman, D. Dragoni, K. Didan, A. Barreto-Munoz, and J. A. Hutabarat, "Detecting large scale conversion of mangroves to aquaculture with change point and mixed-pixel analyses of high-fidelity MODIS data," *Remote Sens. Environ.*, vol. 130, pp. 96–107, Mar. 2013.

[12] X. Zhou, A. R. Armitage, and S. Prasad, "Mapping mangrove communities in coastal wetlands using airborne hyperspectral data," in *Proc. 8th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Aug. 2016, pp. 1–5.