Deep Change Monitoring: A Hyperbolic Representative Learning Framework and a Dataset for Long-term Fine-grained Tree Change Detection

Supplementary Material

1. Overview

This supplementary material includes details of data collection and annotation, ablation study, Tree Change (TC) visualizations and Cross-Domain Anti-Spoofing (CD-FAS) implementations and experiments, which are not included in the main paper due to page limitations.

In this supplementary material, we begin by detailing the tree data collection process using a UAV, as shown in Figure 1. Figure 2 presents the specialized software utilized for cropping individual tree regions. Additionally, Figure 3 provides examples of the data annotations. Furthermore, we offer a comprehensive ablation study to verify the effectiveness and generalizability of Hierarchical Siamese Networks (HSNs), as shown in Table 1. Finally, Figure 4 showcases additional visual examples to further illustrate the effectiveness of our proposed HSNs, and Figure 5 shows a failed example of Tree Change Detection (TCD).

1.1. Data collection process

To elucidate the tree data collection process, Figure 1 offers a detailed visual exposition of the procedures employed by the UAV. Figure 1 (a) captures the UAV in a live operational setting, meticulously gathering visual data with its integrated camera.

The UAV, engaged in the tree data collection task, captures a series of high-resolution images through its mounted camera system. This device is an integrated platform of advanced technology, as demonstrated in Figure 1 (b). The UAV encompasses a resilient battery, a precise GNSS antenna that ensures exact location tracking, a camera capable of capturing intricate details, and a state-of-the-art visual obstacle avoidance mechanism that affirms the vehicle's safe traversal over complex terrains.

Furthermore, the UAV's trajectory is controlled by a UAV remote controller, as shown in Figure 1 (c). The UAV remote controller is equipped with an array of navigational controls that are critical for the precise management of the UAV's trajectory (the green lines in the screen of the controller), ensuring that the UAV adheres to predetermined routes for systematic data collection. The images initially captured by the UAV are illustrated in Figure 1 (d). In this paper, around 220 images were recorded along the UAV's trajectory in each flight task. These images are used for constructing a Digital Orthophoto Model (DOM), a task undertaken with the aid of specialized DJI Terra software. This DOM combines accurate geographical data with high-fidelity photographic images, resulting in a product that went through an extensive registration process. This rigorous approach yielded an impressive geographic registration with a root mean square error of under 4 cm and a reprojection error confined to within a single pixel.

Figure 2 showcases the application of the 'Environment for Visualizing Images' (ENVI) software platform in delineating regions of interest (ROIs) for individual trees, a process enhanced by GPS data to ensure precise location mapping. ENVI is a software platform used for processing and analyzing geospatial imagery. It is widely used in remote sensing and the analysis of satellite and aerial imagery data.

1.2. Data annotation

Figure 3 illustrates the six distinct phenological tree states we annotated in the dataset: green foliage, yellowing leaves, branches, fallen leaves or sprouts, blossom, and destruction. Specifically, fallen leaves and sprouts are assigned a unified label as they both signify transitional phases, marking the point at which a tree begins to undergo seasonal changes. In the tree-change detection task, the inconsistent states indicate the presence of Tree Changes (TCs), otherwise is no change.

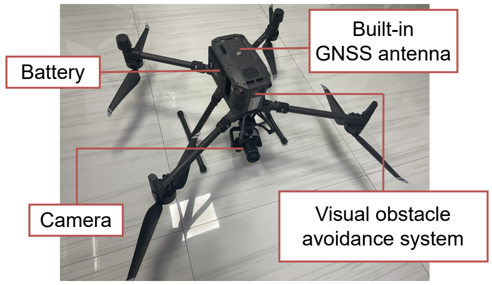
1.3. Ablation study

In order to verify the effectiveness of HSNs, the hyperbolic space is applied on the Siamese network backbone of ResNet18 [3], ResNet101, MobileViT [5], VGG16 [7], InceptionV3 [10], and MobileNet [8], respectively. According to the hyperparameter analysis in Section 5.3 in the main paper, for the deep Siamese network based on ResNet101+HS, we set c to 1.0; otherwise, c is set to 0.3. The ball dimension is set to eight. As shown in Table 1, the ResNet101+HS has achieved the best performance amongst all the methods, in terms of both accuracy and F1-score. Moreover, compared to Euclidean-based Siamese networks (ESNs), HSNs enhance the results by 13.5%, 24.0%, 23.1% ,6.8%, 15.6%, and 8.4% with ResNet18, ResNet101, MobileViT, VGG16, InceptionV3, and MobileNet, respectively, in terms of F1-score. The results indicate that our proposed HSNs are generalizable and can largely improve the performance of TCD. We can draw a conclusion that there is hyperbolicity in TCD and HSNs can better represent TCs with hyperbolicity.

1.4. TC Visualization

Figure 4 displays class activation maps of TCs with Grad-cam [6]. The specific states of the trees have been denoted

(a) The realistic scenario of data collection.



(b) The UAV platform and camera.

(c) The UAV remote controller.

(d) Examples of captured images during the flight mission.

Figure 1. The illustration of the tree data collection process with UAV. (a) the realistic scenario of data collection, (b) the UAV platform and camera, (c) the final UAV remote controller, and (d) examples of captured pictures during the flight mission.

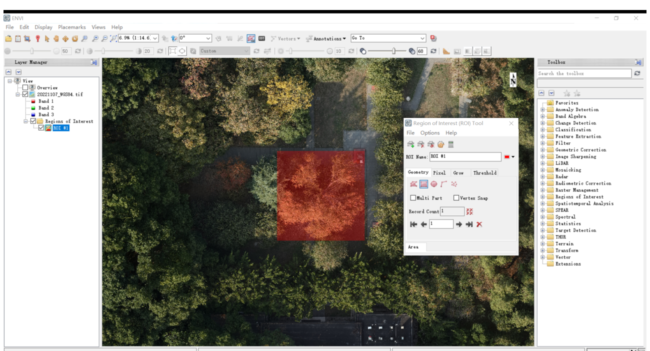
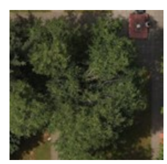


Figure 2. The user interface of ENVI software for tree cropping.



(a) Green foliage

(b) Yellowing leaves

(c) Branches

(d) Fallen leaves or sprouts

(e) blossom

(f) Destruction

Figure 3. The illustration of tree annotation.

in the figure. The inconsistent states indicate the existence of TC. From Figure 4, it can be seen that the HSN appears to effectively concentrate on the tree regions, successfully avoiding misinterpretations caused by illumination and background. For instance, in the first row, the HSN is able to mitigate the impact of shadows and concentrates on extracting features from the tree regions. Additionally, in the examples featuring blossoms, the HSN demonstrates its effectiveness by focusing more accurately on the flower regions compared to the ESN. This enhanced capability of the HSN illustrates its potential for more precise and accurate feature extraction in varied environmental conditions.

Moreover, Figure 5 illustrates a scenario where the TCD process failed. This failure can be attributed to the significant influence of shadows, which underscores the challenges faced in accurately detecting TCs in certain environmental conditions. This example highlights the need for further refinement in the TCD to handle complex scenarios more effectively.

1.5. CD-FAS implementations and experiments

To further prove the effectiveness and impacts of HSN, we evaluate the HSN on cross-domain face anti-spoofing (CD-FAS) tasks, showcasing its significance in AI. CD-FAS is a task that discerns between genuine human faces and counterfeit representations in biometric authentication systems.

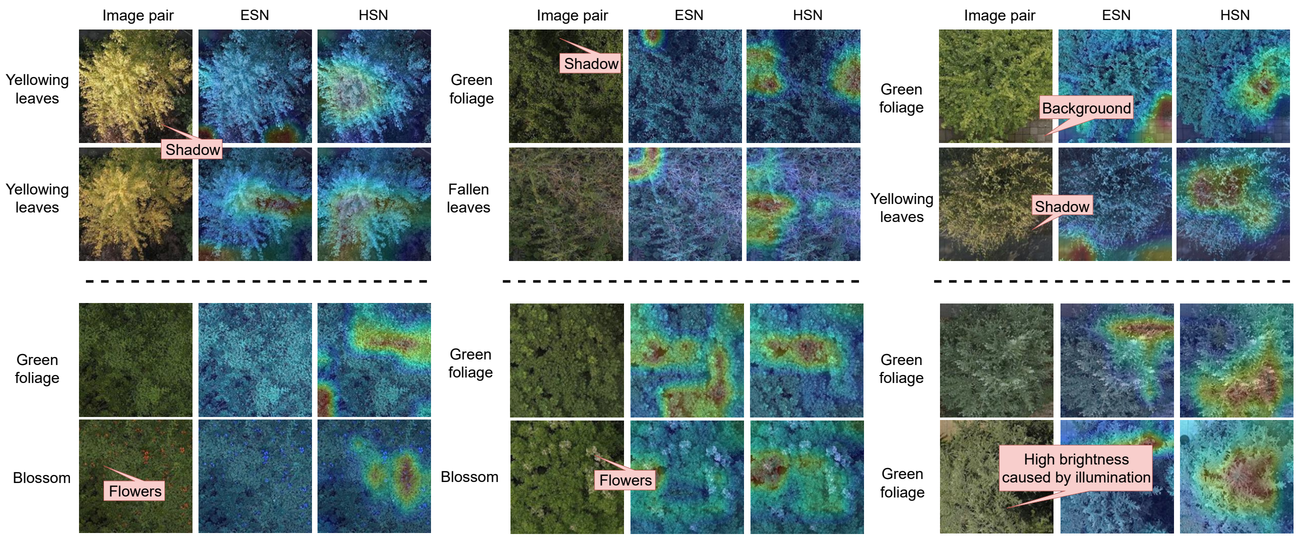


Figure 4. The visualizations of tree change with Gradcam.

Table 1. Ablation study of HSN in terms of F1-score.

Method	F1
VGG16	0.5716
VGG16+HSN	0.6106
InceptionV3	0.5673
InceptionV3+HSN	0.6556
MobileNet	0.4592
MobileNet+HSN	0.4979
ResNet18	0.5331
ResNet18+HSN	0.6053
MobileViT	0.5655
MobileViT+HSN	0.6910
ResNet101	0.5947
ResNet101+HSN	**0.7372**

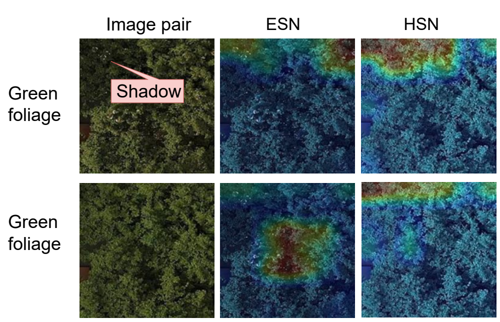


Figure 5. A failed example of tree-change detection.

This task essentially involves identifying pairwise changes, similar to the process of detecting changes in a tree. Four popular benchmark datasets was used, including Oulu-NPU (O) [1], CASIA (C) [15], Idiap Replay Attack (I) [2], and MSU-MFSD (M) [13]. Following prior works, we treat each dataset as one domain and apply the leave-one-out test protocol to evaluate their cross-domain generalization. Specifically, the OCI→M is referred as the protocol that trains on Oulu-NPU, CASIA, Idiap Replay attack and tests on MSU-MFSD. OMI→C, OCM→I and ICM→O are defined in a similar fashion.

Table 2. Evaluation upon convergence: Evaluation of cross-domain face anti-spoofing among CASIA (C), Idiap Replay (I), MSU-MFSD (M), and Oulu-NPU (O) databases in terms of HTER and AUC (**HTER↓/AUC↑**). Methods are compared at their std performance based on the last 10 epochs.

Method (%)	OCI → M	OMI → C	OCM → I	ICM → O
SSDG-R [4]	1.21 / 1.35	0.89 / 1.10	1.14 / 1.31	1.29 / 0.96
SSAN-R [12]	3.68 / 3.78	2.91 / 2.83	8.04 / 9.03	3.74 / 4.69
PatchNet [11]	1.13 / 0.87	1.98 / 1.89	2.76 / 1.35	1.80 / 1.92
SA-FAS [9]	0.92 / 0.82	0.72 / 0.58	0.85 / 0.99	0.55/ 0.59
Ours	0.85 / 0.32	1.01 / 0.28	1.22 / 0.78	0.79/ 0.24

The input images are cropped using MTCNN [14] and resized to 256×256. For fair comparisons with state-of-the-art (SoTA) methods [4, 9, 11, 12], we use the same ResNet-18 backbone. We train the network with the SGD optimizer and an initial learning rate of 5×10^{-3}, which is decayed by 2 at epoch 40 and 80, and the total training epoch is 100. We set the weight decay as 5×10^{-4} and the batch size as 96 for each training domain.

For fair comparisons with state-of-the-art (SoTA) methods [4, 9, 11, 12], we adapt the learning strategy and evaluation protocol of SA-FAS [9], reporting the average performance over the last 10 epochs after convergence. Three metrics are employed: Half Total Error Rate (HTER) and Area Under the Curve (AUC). While HTER and AUC assess theoretical performance, TPR at a specific FPR effectively reflects practical model performance.

In addition to the mean accuracy shown in Table 2, Table 2 illustrates the standard deviation, indicating the stability of each method's performance. Most methods converge to a relatively stable state, especially our method. SSAN-R [12] incorporating adversarial loss exhibit a relatively larger standard deviation, highlighting the instability inherent in adversarial learning.

References

[1] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, pages 612–618. IEEE, 2017. 3

[2] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, pages 1–7. IEEE, 2012. 3

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[4] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8484–8493, 2020. 3

[5] Sachin Mehta and Mohammad Rastegari. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*, 2021. 1

[6] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 1

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[8] Debjyoti Sinha and Mohamed El-Sharkawy. Thin mobilenet: An enhanced mobilenet architecture. In *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*, pages 0280–0285. IEEE, 2019. 1

[9] Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24563–24574, 2023. 3

[10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 1

[11] Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai. Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20281–20290, 2022. 3

[12] Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang. Domain generalization via shuffled style assembly for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4123–4133, 2022. 3, 4

[13] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015. 3

[14] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23 (10):1499–1503, 2016. 3

[15] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face antispoofing database with diverse attacks. In *2012 5th IAPR international conference on Biometrics (ICB)*, pages 26–31. IEEE, 2012. 3