Supplementary Material of: Dual-stage Hyperspectral Image Classification Model with Spectral Supertoken

Peifu Liu, Tingfa Xu†, Jie Wang, Huan Chen, Huiyan Bai, and Jianan Li†

Beijing Institute of Technology

S1 Experiment

S1.1 Experiment on Salient Object Detection

Data and Experimental Settings. In order to evaluate the transferability of our DSTC to other visual tasks, we have conducted experiments on a hyperspectral salient object detection dataset, HS-SOD [2]. This dataset comprises 60 hyperspectral images, each spanning a spectral range from 380 to 780 nm, at 5 nm intervals, and featuring a spatial resolution of 768×1024 pixels. We partition the HS-SOD dataset manually into training and testing subsets, allocating 48 HSIs for training and the remaining 12 for testing. During the training phase, we set the initial learning rate to 1×10^{-3} and extend the training duration to 150 epochs. We select ResNet18 as the backbone network.

Compared Methods. We benchmark our model against several established models, including Itti's model [3] and conventional hyperspectral salient object detection (HSOD) methods proposed by Liang *et al.* [5]. These conventional methods encompass spectral angle distance (SAD), spectral Euclidean distance (SED), and spectral grouping (SG). Additionally, we included comparisons with deep learning-based HSOD method SUDF [1], as well as RGB SOD methods like EDN [7], TRACER [4], and ABiU_Net [6].

The performance of the above models is evaluated using several metrics, namely Mean Absolute Error ($\mathcal{M}$), E-measure (E_ξ), F-measure (F_β), Area Under Curve (AUC), and Cross-Correlation (CC).

Quantitative Results. The quantitative results of our DSTC on the HS-SOD dataset are detailed in Tab. S1, which indicates that DSTC surpasses both traditional hyperspectral salient object detection methods and contemporary deep learning-based approaches, as well as SOD techniques, in most of the evaluated metrics. DSTC demonstrates outstanding performance across a range of metrics, achieving a $\mathcal{M}$ score of 0.094, an E_ξ of 0.786, an F_β of 0.608, and an AUC of 0.925. The only area where DSTC exhibits a slight deficit is in CC, lagging behind EDN by a marginal difference of 0.035. These results underscore the robust transferability of DSTC to the saliency detection task.

† Correspondence to: Tingfa Xu and Jianan Li.

Table S1: Quantitative Results on the HS-SOD Dataset.

Methods	$\mathcal{M} \downarrow$	$E_\xi \uparrow$	$F_\beta \uparrow$	AUC ↑	CC ↑
Itti [3]	0.259	0.539	0.207	0.783	0.225
SAD [5]	0.205	0.546	0.197	0.778	0.223
SED [5]	0.133	0.577	0.258	0.793	0.200
SG [5]	0.197	0.563	0.234	0.808	0.268
SUDF [1]	0.242	0.554	0.256	0.723	0.250
TRACER [4]	0.158	0.610	0.393	0.868	0.465
ABiU_Net [6]	0.119	0.620	0.391	0.846	0.472
DSTC (Ours)	**0.094**	**0.786**	**0.608**	**0.925**	**0.590**

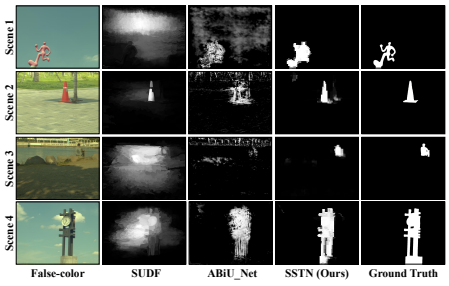

Fig. S1: Qualitative results on HS-SOD. DSTC's results closely align with the ground truth, demonstrating the robust transferability of DSTC.

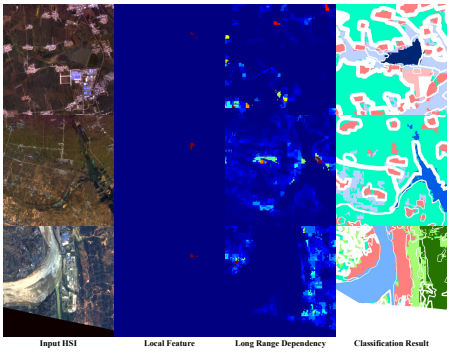

Fig. S2: Visualization of Local Feature and Long Range Dependency.

Qualitative Results. Fig. S1 presents the comparative performance of our DSTC against other deep learning methods, including SUDF and ABiU_Net. DSTC stands out for its remarkable ability to maintain edge clarity, an advantage stemming from the incorporation of a pre-classification stage. This stage clusters similar pixels, thereby preserving sharp edge details. The outputs generated by DSTC align closely with the ground truth across various scenes, conclusively demonstrating its robust transferability and effectiveness in the context of hyperspectral salient object detection.

S2 More Visualization on WHU-OHS Dataset

Visualization of Local Feature and Long Range Dependency. In the DSTC framework, we initially aggregate local features through spectrum-derivative-based pixel clustering and semantic feature aggregation. Subsequently, in the second phase, global information is modeled using a Transformer. In this subsection, we randomly selected one spectral supertoken for visualization. As depicted in

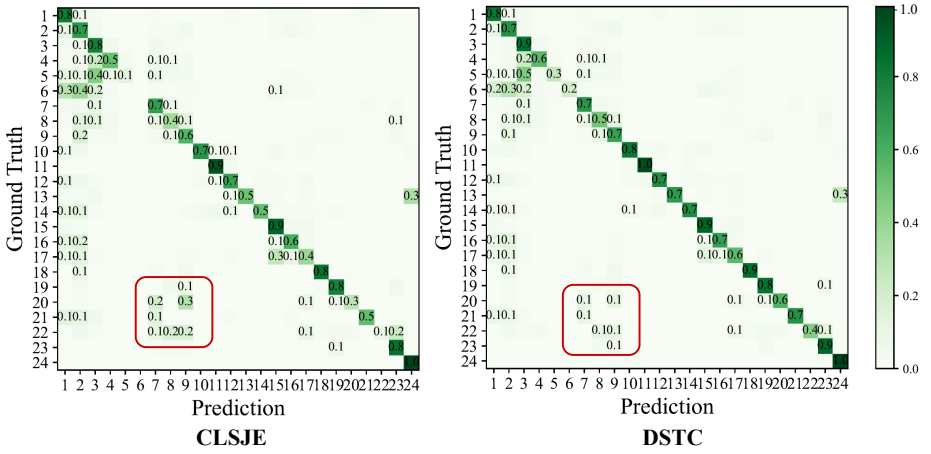

Fig. S3: Full Confusion Matrices.

Figure S2, this visualization elucidates the interconnections between this specific spectral supertoken and other supertokens. It is evident that employing the Transformer for modeling global dependencies plays a pivotal role.

Full Confusion Matrices. The complete confusion matrices are presented in Fig. S3. For clarity, only values greater than 0.1 are displayed, which is why the class probabilities do not sum to 1. Highlighted with red boxes, our DSTC shows robust performance across various classes.

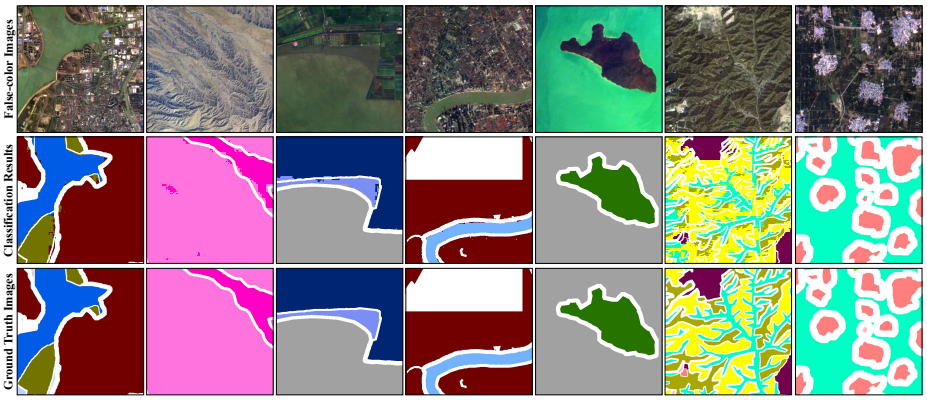

Fig. S4: More qualitative results on WHU-OHS dataset.

More Qualitative Results. Additional typical scenarios are selected for further illustration, as showcased in Fig. S4.

S3 Method

We summarize the pseudo-code of generating class-proportion-based soft labels as Algorithm S1.

Algorithm S1 Pytorch Pseudo Code of Generating CPSL.

```
# gt: ground truth image, [H, W]
# asso_mat: association matrix, [N, M]
# H, W: height and width of ground truth image
# C: number of classes
# N: number of pixels (N=H*W)
# M: number of center points

from einops import rearrange

# filtering ground truth
gt_reshape = rearrange(gt, "h w -> (h w)")
gt_filt = gt_reshape.unsqueeze(-1) * asso_mat
gt_filt = gt_filt.flatten(start_dim=3)

# calculating class probability
count = zeros((N, C + 1))
count.scatter_add_(dim=1, index=gt_filt, src=ones_like(
    gt_filt))
return rearrange(count[:, 1:], "n c -> c n")
```

References

1. Imamoglu, N., Ding, G., Fang, Y., Kanezaki, A., Kouyama, T., Nakamura, R.: Salient object detection on hyperspectral images using features learned from unsupervised segmentation task. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019)
2. Imamoglu, N., Oishi, Y., Zhang, X., Ding, G., Fang, Y., Kouyama, T., Nakamura, R.: Hyperspectral image dataset for benchmarking on salient object detection. In: 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX). pp. 1–3 (2018). https://doi.org/10.1109/QoMEX.2018.8463428
3. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**, 1254–1259 (1998)
4. Lee, M.S., Shin, W., Han, S.W.: Tracer: Extreme attention guided salient object tracing network (student abstract). In: AAAI. vol. 36, pp. 12993–12994 (2022)
5. Liang, J., Zhou, J., Bai, X., Qian, Y.: Salient object detection in hyperspectral imagery. 2013 IEEE International Conference on Image Processing (2013)
6. Qiu, Y., Liu, Y., Zhang, L., Lu, H., Xu, J.: Boosting salient object detection with transformer-based asymmetric bilateral u-net. IEEE Transactions on Circuits and

Systems for Video Technology pp. 1–1 (2023). https://doi.org/10.1109/TCSVT.2023.3307693

7. Wu, Y.H., Liu, Y., Zhang, L., Cheng, M.M., Ren, B.: Edn: Salient object detection via extremely-downsampled network. IEEE Transactions on Image Processing **31**, 3125–3136 (2022)