Deep learning framework for analyzing birefringence imaging by incorporating optical polarization overlap in stress-induced ferroelectric $SrTiO_3$

Hirotaka Manaka[a], Shoutarou Katayama[a], Soichiro Honda[a], and Yoko Miura[b]

[a]Graduate School of Science and Engineering, Kagoshima University, Korimoto, Kagoshima 890-0065, Japan; [b]National Institute of Technology, Suzuka College, Shiroko-cho, Suzuka, Mie 510-0294, Japan

ARTICLE HISTORY
Compiled September 13, 2025

1. Mahalanobis distance

Figure S1 shows the results of the elbow method, Silhouette score, and gap statistic for determining the optimal number of clusters using the dataset (3 λ, 1 Mh, 233 T) × 42,280 pixels. As expected, these methods yield inconsistent results, but $k = 2$–5 seem to be the reasonable choices. Figure S2 shows the spatial distribution of clustering results for $k = 2$ and 3. The stripe structures observed in Figure 3 are not clearly visible, regardless of k, in these results. This loss of structure is likely due to the reduction from the original variables of (3 λ, 2 θ, 2 ϕ, 233 T) to only (3 λ, 1 Mh, 233 T).

2. Long short-term memory (LSTM)

Figure S3 shows the $Mh(T)$ curves at 523 and 543 nm as well as the LSTM prediction results at the positions b1 and b2 in Figure 2(a). These predictions closely follow the $Mh(T)$ data. This shows that the LSTM model successfully captures the underlying T-dependent trends in the optical polarization (OP) states. Figure S4 and the accompanying video present distributions of the dataset containing (3 λ, 5 PCs, 233 T, 42,280 pixels) obtained using temperature series principal component analysis (TsPCA). This visualization reveals horizontal stripe structures that were not explicitly present in the original LSTM input dataset. This suggests that TsPCA effectively reconstructs underlying OP variations that may have been masked in the raw data. The K-shape multivariate clustering method was applied to the dataset (3 λ, 5 PCs, 233 T) × 42,280 pixels. Figure S5 shows the results of the elbow method and Silhouette score for determining the optimal number of clusters (k). Figure S6 shows the clustering results for $k = 4$ and $k = 5$. These results suggest that LSTM training enhances the diversity of extracted polarization features, thereby allowing for a more detailed characterization of temperature-dependent birefringence variations. The success of TsPCA-based reconstruction is confirmed by the preservation of key OP trends in the raw data.

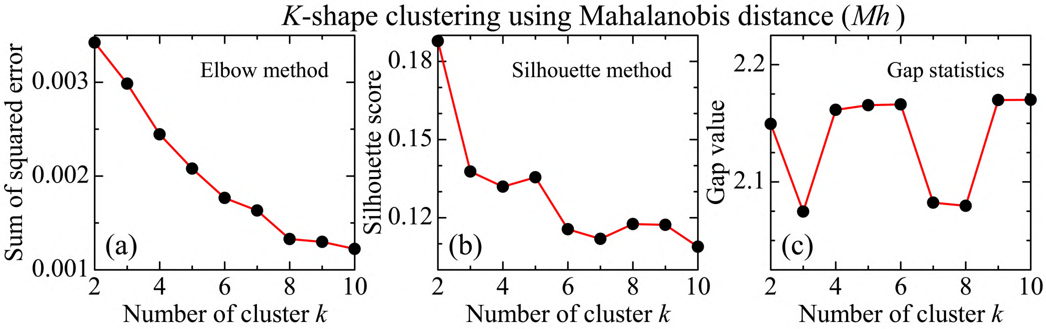


Figure S1. Evaluation of *K*-shape multivariate clustering with three variables applied to a dataset of (3 λ, 1 Mh, 233 T, 42,280 pixels) as T decreases from 130.0 to 14.1 K. The optimal number of clusters (k) is determined using (a) the elbow method, (b) Silhouette score, and (c) gap statistic.

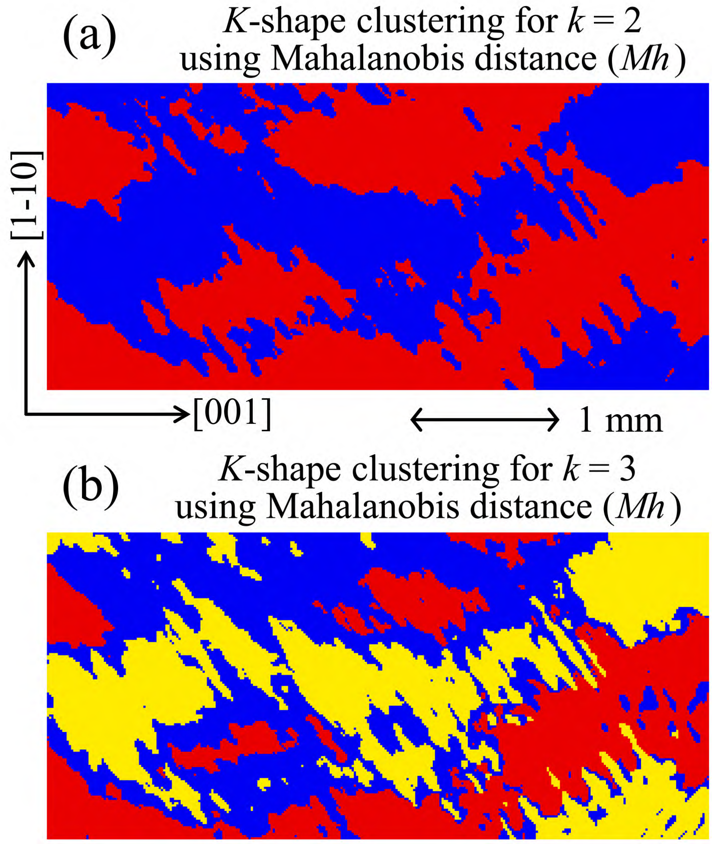


Figure S2. Results of *K*-shape multivariate clustering with three variables applied to a dataset of (3 λ, 1 Mh, 233 T, 42,280 pixels) for (a) $k = 2$ and (b) $k = 3$.

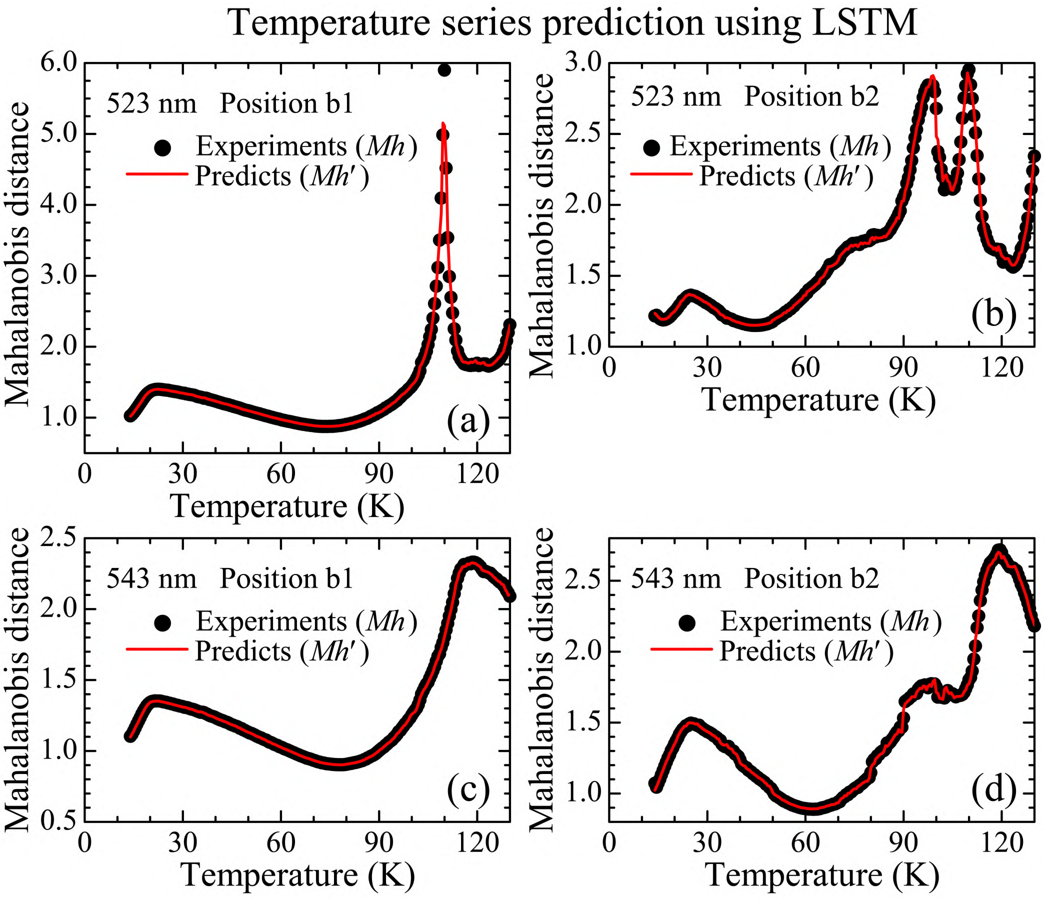


Figure S3. Temperature dependence of the Mahalanobis distance (Mh') predicted by the LSTM at (a) b1 and (b) b2 for 523 nm and at (c) b1 and (d) b2 for 543 nm. Circles indicate the Mahalanobis distance (Mh) derived from the experimental data, whereas solid lines indicate the LSTM predictions (Mh').

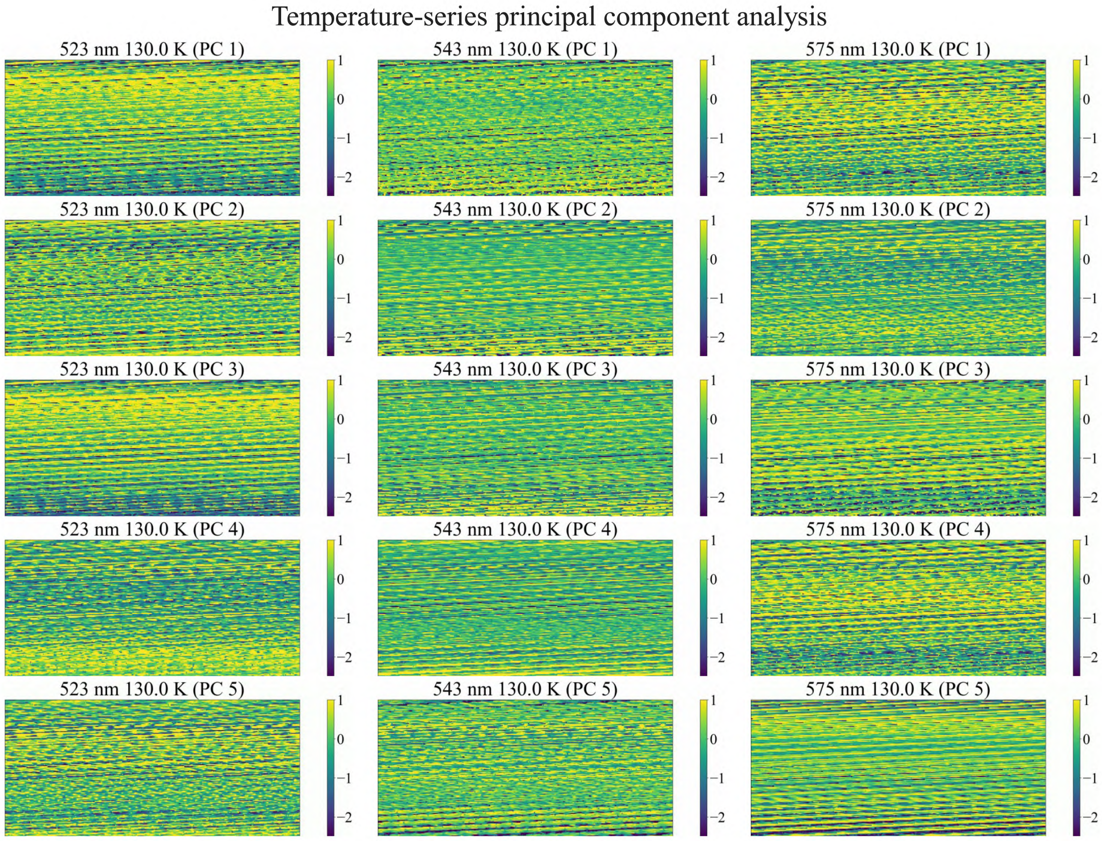


Figure S4. Distributions of the dataset (3 λ, 5 PCs, 233 T, 42,280 pixels) obtained via temperature-series principal component analysis (TsPCA). These results are also shown in the accompanying movie.

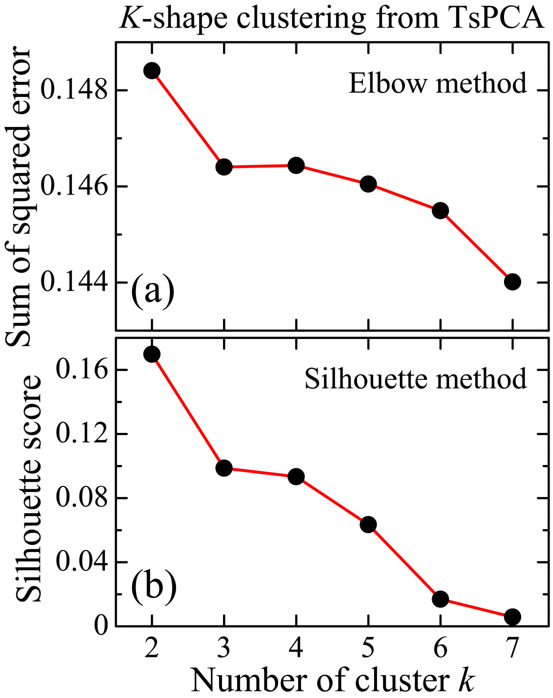
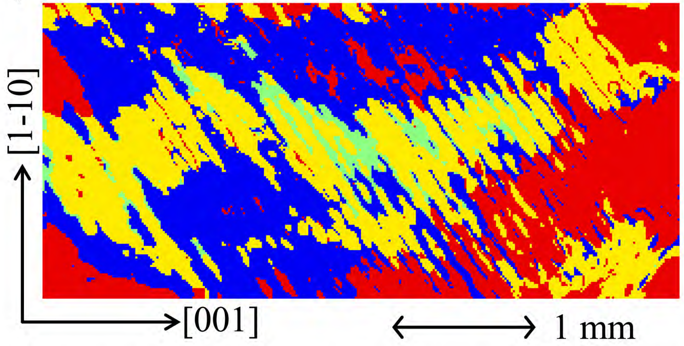


Figure S5. Evaluation results of *K*-shape multivariate clustering with 15 variables applied to a dataset containing (3 λ, 5 PCs, 233 T, 42,280 pixels). The optimal number of clusters k is determined using (a) the elbow method and (b) Silhouette score.

(a) K-shape clustering for $k = 4$ from TsPCA



(b) K-shape clustering for $k = 5$ from TsPCA

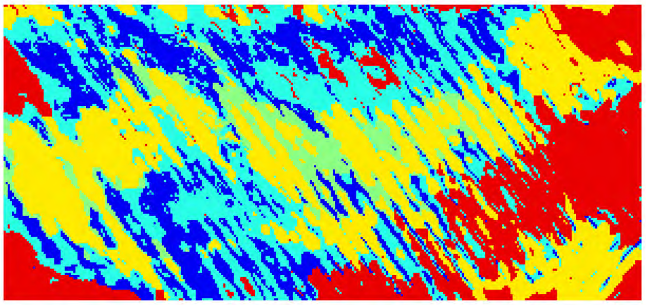


Figure S6. Results of K-shape multivariate clustering with 15 variables applied to a dataset (3 λ, 5 PCs, 233 T)×42,280 pixels shown for (a) $k = 4$ and (b) $k = 5$.

3. 3D convolutional autoencoder (3DCAE)

Figures S7 and S8 show the results obtained after applying a 3D convolutional autoencoder (3DCAE) to the 575 nm dataset, $(PC, T, W, H) = (5, 233, 302, 140)$. Figures S7(a) and S8(a) show the epoch-wise evolution of the loss function for the spatial receptive field (SRF) sizes of 5×5 and 10×10, respectively. These results indicate that the loss function decreases sufficiently, confirming stable training. Figures S7(b–c) and S8(b–c) show the prediction results at b1 and b2 for each SRF size. Despite identical input data, the predictions vary considerably with the SRF sizes. In particular, the range of variation with respect to T is smaller for the 5×5 case but larger for the 10×10 case. Although the sliding window width along the temperature axis is set to 2.5 K, the response for the 5×5 case is much flatter.

After transforming the five PC's obtained from the 3DCAE into a single Mahalanobis distance (Mh''), K-shape multivariate clustering was applied to the dataset structured as ($3\ \lambda$, $1\ Mh''$, $233\ T$) × 42,280 pixels for each SRF size. Figures S9–S11 show the evaluation of the optimal number of clusters (k) using the elbow method, Silhouette score, and gap statistic for SRF sizes of 3×3, 5×5, and 10×10. As determining a definitive optimal value of k is challenging, candidate values between $k = 2$ and 5 are considered. Figures S12–S15 show the detailed clustering results for each SRF size at $k = 2$–5, along with summary tables. The segmentation patterns indicate that the clustering results remain largely consistent across different SRF sizes. In all cases, the overall clustering trends are similar and effectively delineate the expected stress-concentration regions. The robustness of these clustering results across different SRF sizes indicates that the essential structural patterns are consistently extracted, reinforcing the validity of the clustering approach.

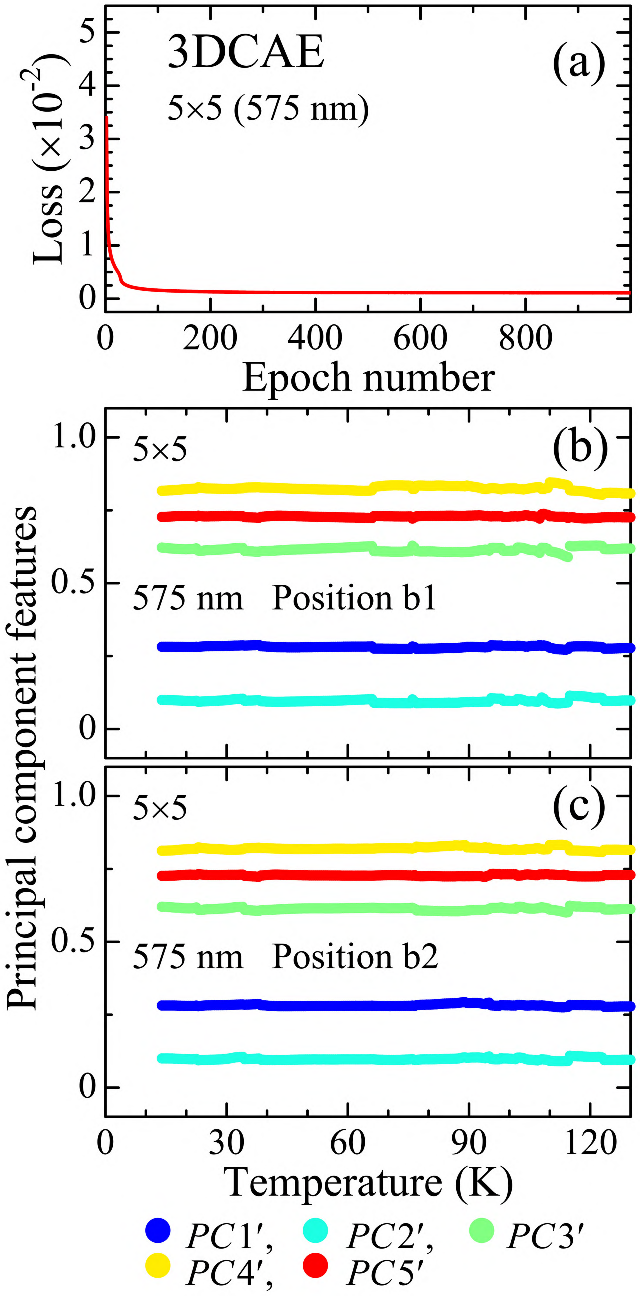


Figure S7. (a) Evolution of the loss function during the training of 3DCAE using the dataset (5 PCs, 233 T, 302 pixels, 140 pixels) at 575 nm for the 5×5 spatial receptive field (SRF) size. Temperature dependence of the five principal components (PC's) predicted by the 3DCAE at (b) b1 and (c) b2.

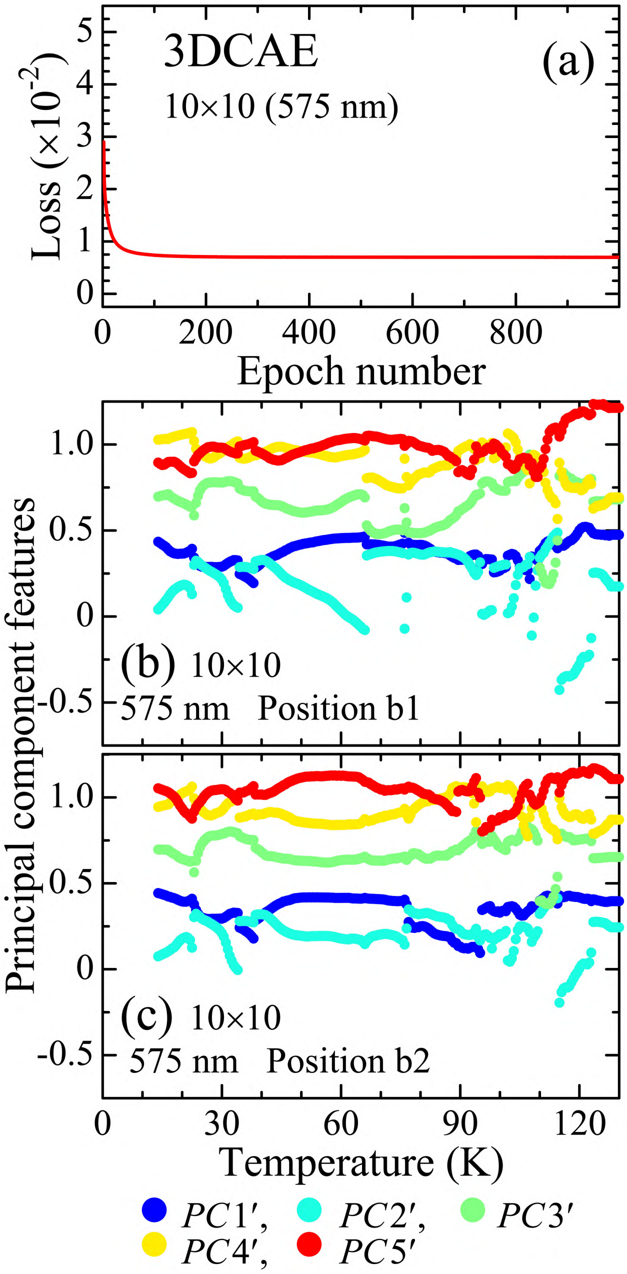


Figure S8. (a) Evolution of the loss function during the training of 3DCAE using the dataset (5 PCs, 233 T, 302 pixels, 140 pixels) at 575 nm for the 10×10 SRF size. Temperature dependence of the five PC's predicted by the 3DCAE at (b) b1 and (c) b2.

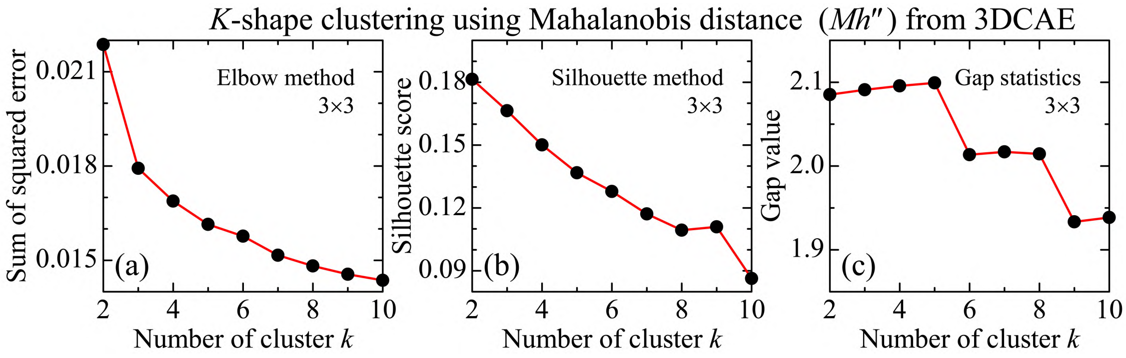


Figure S9. Evaluation results of K-shape multivariate clustering with three variables applied to a dataset containing (3 λ, 1 Mh'', 233 T) × 42,280 pixels for the 3×3 SRF size. The optimal number of clusters (k) is determined using (a) the elbow method, (b) Silhouette score, and (c) gap statistic.

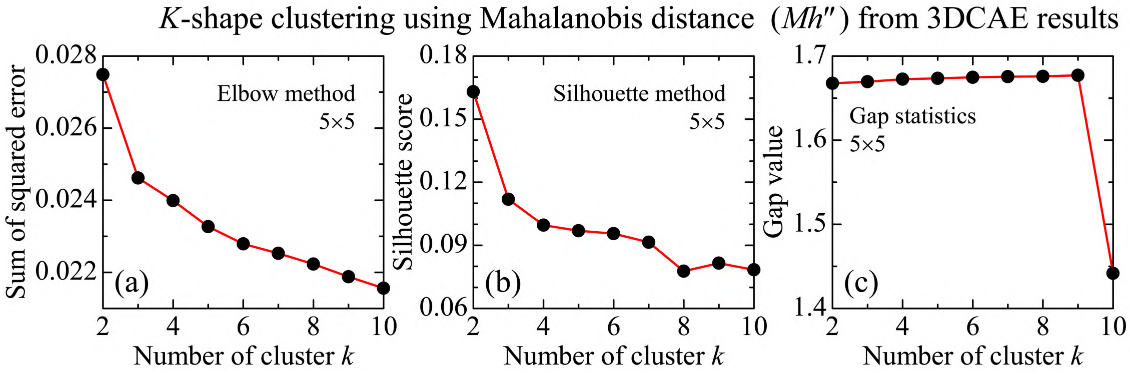


Figure S10. Evaluation results of K-shape multivariate clustering with three variables applied to a dataset containing (3 λ, 1 Mh'', 233 T) × 42,280 pixels for the 5×5 SRF size. The optimal number of clusters (k) is determined using (a) the elbow method, (b) Silhouette score, and (c) gap statistic.

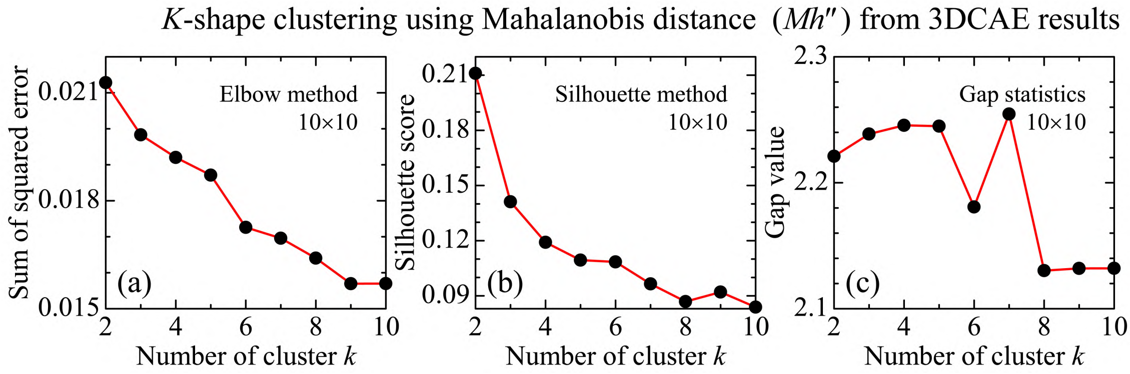
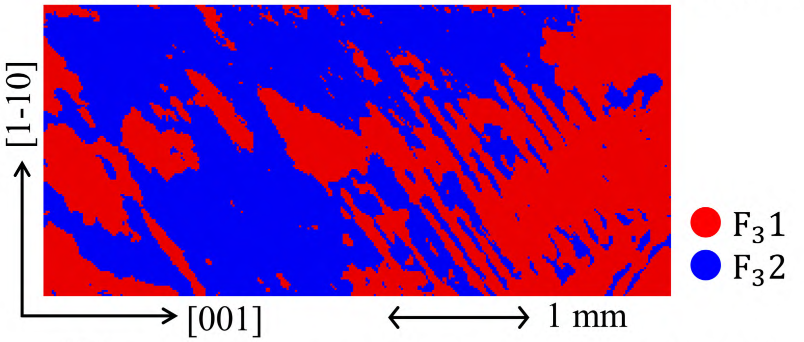


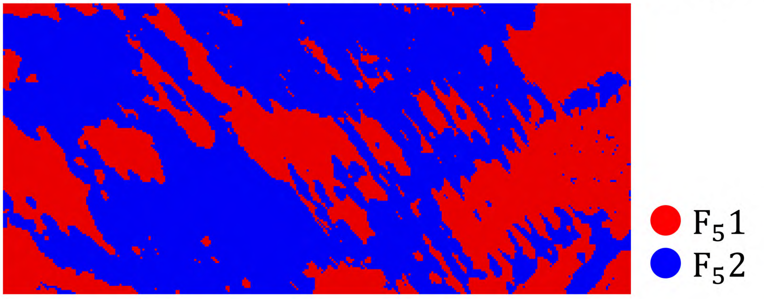
Figure S11. Evaluation results of K-shape multivariate clustering with three variables applied to a dataset containing (3 λ, 1 Mh'', 233 T) × 42,280 pixels for the 10×10 SRF size. The optimal number of clusters (k) is determined using (a) the elbow method, (b) Silhouette score, and (c) gap statistic.

(a) *K*-shape clustering for $k = 2$ from 3DCAE (3 × 3)



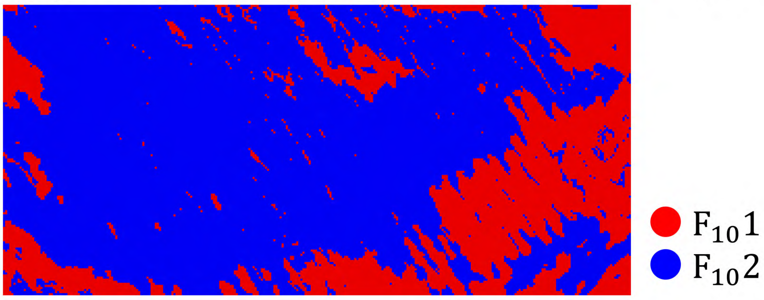
K-shape	
$F_3 1$	19,659
$F_3 2$	22,621

(b) *K*-shape clustering for $k = 2$ from 3DCAE (5 × 5)



K-shape	
$F_5 1$	16,795
$F_5 2$	25,485

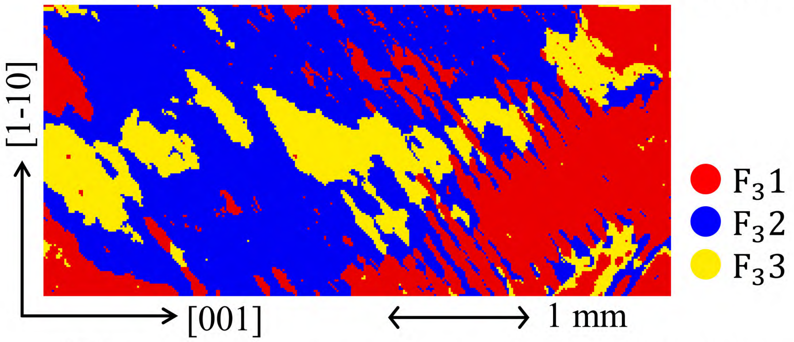
(c) *K*-shape clustering for $k = 2$ from 3DCAE (10 × 10)



K-shape	
$F_{10} 1$	11,350
$F_{10} 2$	30,930

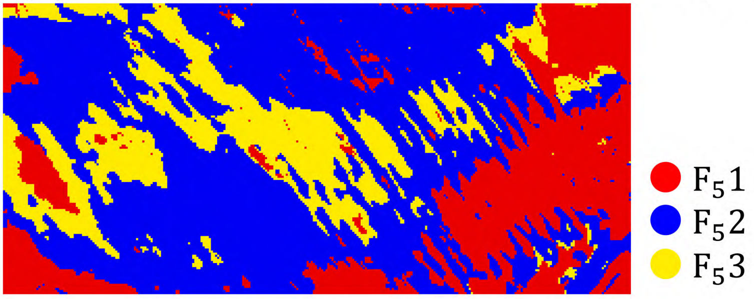
Figure S12. Results of K-shape clustering using the Mahalanobis distance (Mh'') derived from 3DCAE outputs at $k = 2$ for SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The corresponding tables showing the number of pixels in each cluster.

(a) *K*-shape clustering for $k = 3$ from 3DCAE (3 × 3)



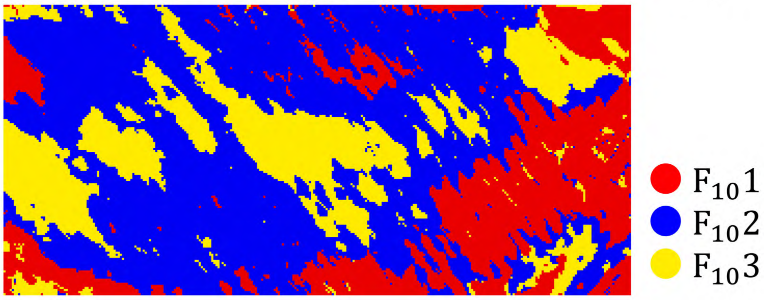
K-shape	
$F_3 1$	12,767
$F_3 2$	21,968
$F_3 3$	7,545

(b) *K*-shape clustering for $k = 3$ from 3DCAE (5 × 5)



K-shape	
$F_5 1$	10,346
$F_5 2$	23,897
$F_5 3$	8,037

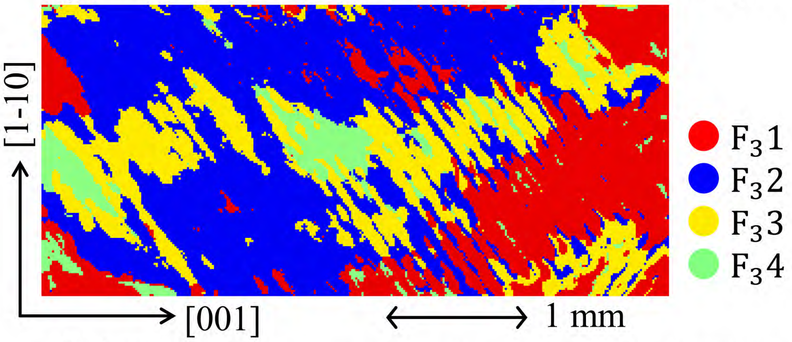
(c) *K*-shape clustering for $k = 3$ from 3DCAE (10 × 10)



K-shape	
$F_{10} 1$	9,810
$F_{10} 2$	23,069
$F_{10} 3$	9,401

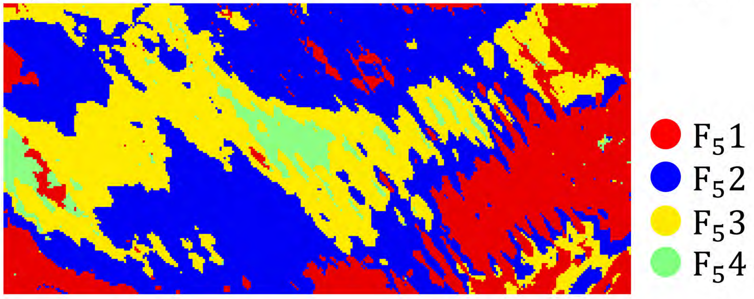
Figure S13. Results of *K*-shape clustering using the Mahalanobis distance (Mh'') derived from 3DCAE outputs at $k = 3$ for SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The corresponding tables showing the number of pixels in each cluster.

(a) *K*-shape clustering for $k = 4$ from 3DCAE (3 × 3)



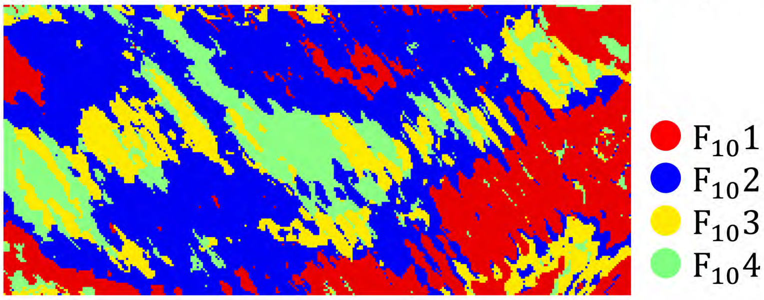
K-shape	
F$_3$1	10,356
F$_3$2	19,886
F$_3$3	9,380
F$_3$4	2,658

(b) *K*-shape clustering for $k = 4$ from 3DCAE (5 × 5)



K-shape	
F$_5$1	9,860
F$_5$2	18,402
F$_5$3	12,232
F$_5$4	1,786

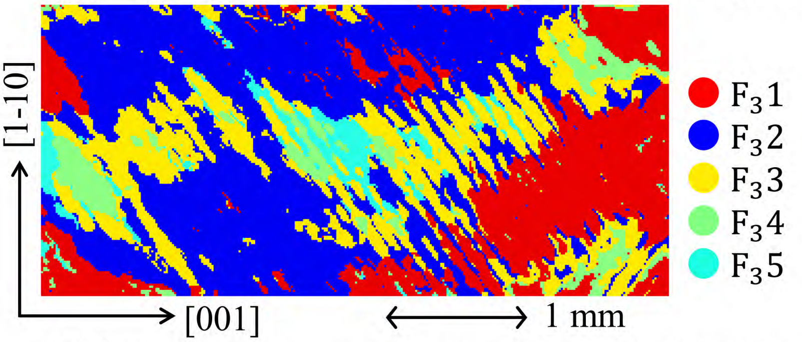
(c) *K*-shape clustering for $k = 4$ from 3DCAE (10 × 10)



K-shape	
F$_{10}$1	9,611
F$_{10}$2	19,580
F$_{10}$3	6,851
F$_{10}$4	6,238

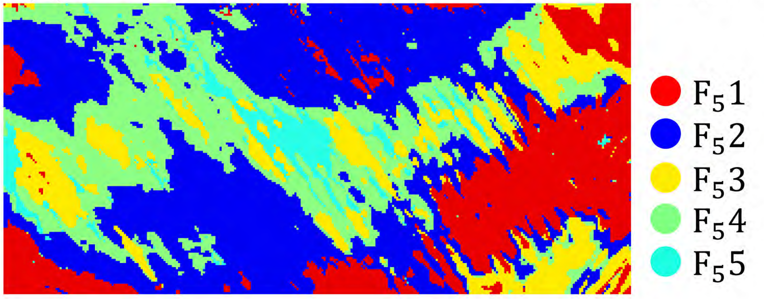
Figure S14. Results of *K*-shape clustering using the Mahalanobis distance (Mh'') derived from 3DCAE outputs at $k = 4$ for SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The corresponding tables showing the number of pixels in each cluster.

(a) *K*-shape clustering for $k = 5$ from 3DCAE (3×3)



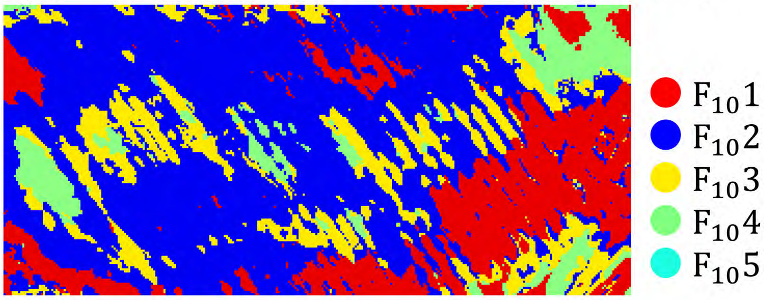
K-shape	
$F_3 1$	9,732
$F_3 2$	19,083
$F_3 3$	9,084
$F_3 4$	2,565
$F_3 5$	1,816

(b) *K*-shape clustering for $k = 5$ from 3DCAE (5×5)



K-shape	
$F_5 1$	8,044
$F_5 2$	17,004
$F_5 3$	5,546
$F_5 4$	9,748
$F_5 5$	1,938

(c) *K*-shape clustering for $k = 5$ from 3DCAE (10×10)



K-shape	
$F_{10} 1$	8,718
$F_{10} 2$	17,817
$F_{10} 3$	6,326
$F_{10} 4$	3,312
$F_{10} 5$	6,107

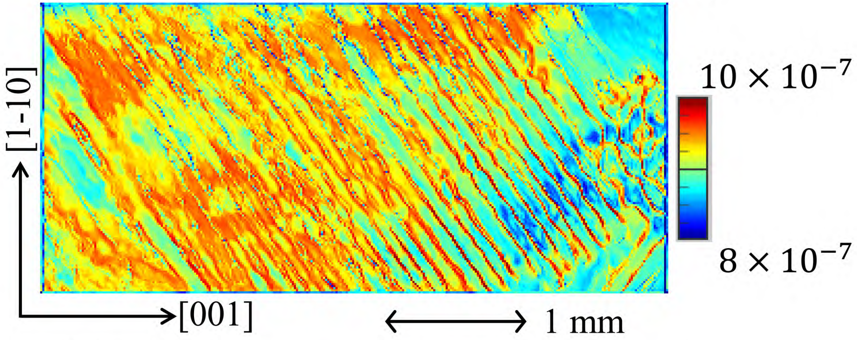
Figure S15. Results of *K*-shape clustering using the Mahalanobis distance (Mh'') derived from 3DCAE outputs at $k = 5$ for SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The corresponding tables showing the number of pixels in each cluster.

Figures S16 and S17 show the effective receptive field (ERF) results for 523 and 543 nm, respectively, and Figures S18 and S19 show the layer-wise relevance propagation (LRP) results for the same wavelengths. The ERF results for the 3×3 and 5×5 SRF sizes distinguish between stress-concentration and uniform regions, and the stripe structure is also learned. The 10×10 SRF size shows a different trend from other cases. The LRP results are also consistent with the ERF results under the same conditions, suggesting that the regions effectively used during training are considered significant during prediction. This improved both the stability and interpretability of the model. It also implied a lower tendency for overfitting, as the model learned meaningful features rather than redundant information.

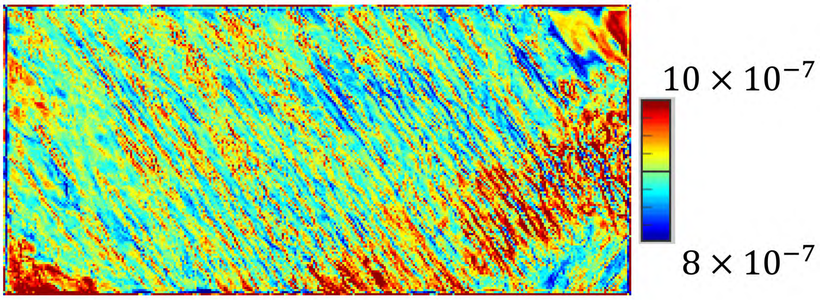
Figures S20 and S21 show the amplitude spectra of the 2D FFT ERF images for each SRF size at 523 and 543 nm, and Figures S22–S24 show the amplitude spectra of the 2D FFT LRP images at 523, 543, and 575 nm. The frequency components highlighted in red indicate statistically significant modes at the 95% confidence level. As the ERF and LRP results agree well with each other, the 2D FFT images show similar trends for each λ and SRF size. Although a significant mode for the 543 nm images is not clearly visible for the 5×5 case, the significant mode may likely appear in the top-right diagonal direction based on the results from the other λ. Figures S25 and S26 show the inverse Fourier-transform ERF images reconstructed using only the statistically significant frequency modes for each SRF size at 523 and 543 nm. Figures S27–S29 show the inverse Fourier-transform ERF images at 523, 543, and 575 nm, respectively, wherein stripe structures are visible, albeit with different results for the 10×10 SRF size.

Figures S30 and S31 show the results of occlusion sensitivity analysis (OSA) results at 523 and 543 nm and reveal similar trends at both wavelengths. Compared with the results obtained at 575 nm (Figure 20), the patterns for the 3×3 and 5×5 cases show an opposite trend; however, there is a clear distinction between stress-concentration and uniform regions. For the 10×10 case, the stress-concentration regions show increased strength; however, the overall trend is different from that in the 3×3 and 5×5 cases. Figures S32 and S33 show the feature maps and feature weight histograms at 523 and 543 nm, respectively. As shown in Figure 13, the convolution layer adopts a single-layer structure for the 3×3 SRF size, whereas a two-layer structure is used for the 5×5 and 10×10 cases. In these multilayer configurations, the sign of the features is reversed between the first and second layers, except for the 5×5 case at 543 nm. This strongly suggests that the network learns OP states with opposite characteristics, which enhances its ability to discriminate between stress-concentration and uniform regions. Finally, the weight histograms in Figures S32(f–h) and S33(f–h) show a symmetric distribution centered around zero, confirming that the model captures meaningful representations without overfitting. For 5×5 and 10×10 SRF sizes, the first layer learns global image trends by assigning a wide range of weights and the second layer refines the representation by focusing on localized structures with a narrower weight distribution.

(a) ERF image (3 × 3) at $\lambda = 523$ nm



(b) ERF image (5 × 5) at $\lambda = 523$ nm



(c) ERF image (10 × 10) at $\lambda = 523$ nm

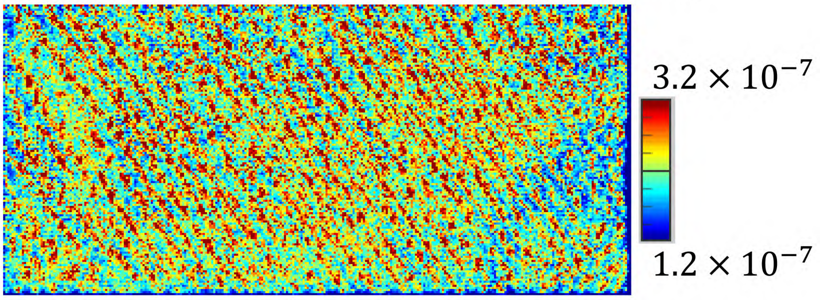
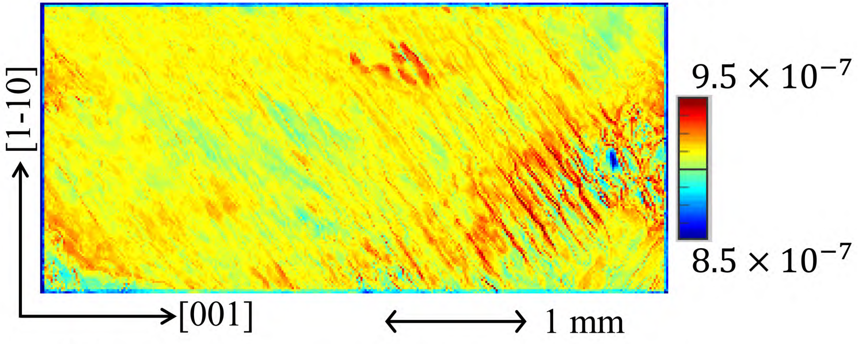
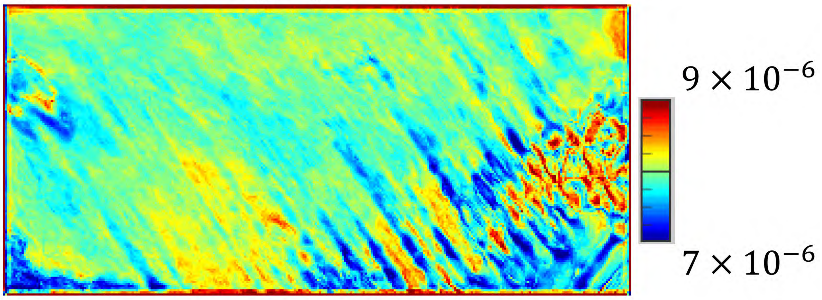


Figure S16. Comparison of the effective receptive field (ERF) images for different SRF sizes at 523 nm: (a) 3×3, (b) 5×5, and (c) 10×10.

(a) ERF image (3×3) at $\lambda = 543$ nm



(b) ERF image (5×5) at $\lambda = 543$ nm



(c) ERF image (10×10) at $\lambda = 543$ nm

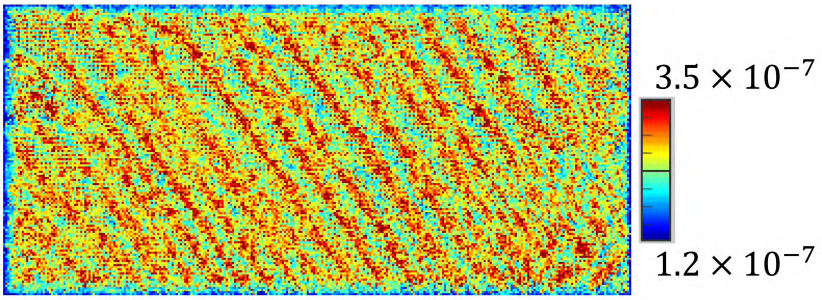


Figure S17. Comparison of the ERF images for different SRF sizes at 543 nm: (a) 3×3, (b) 5×5, and (c) 10×10.

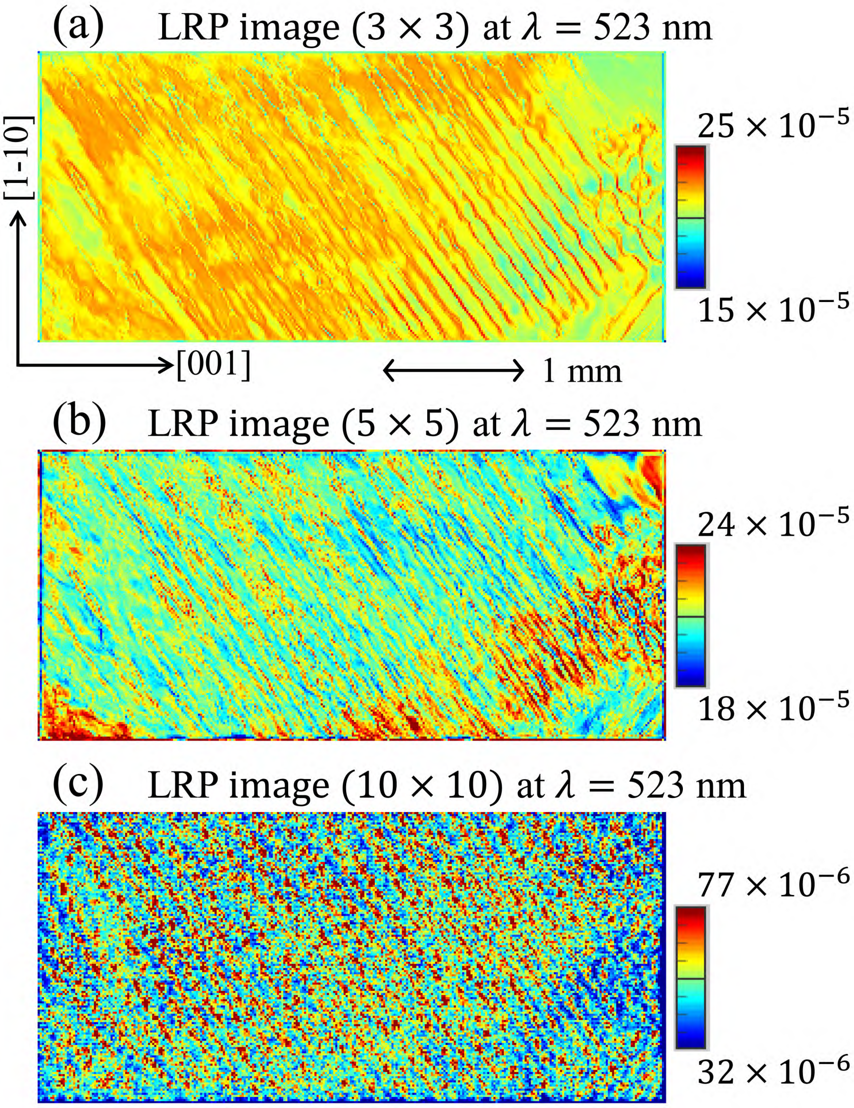
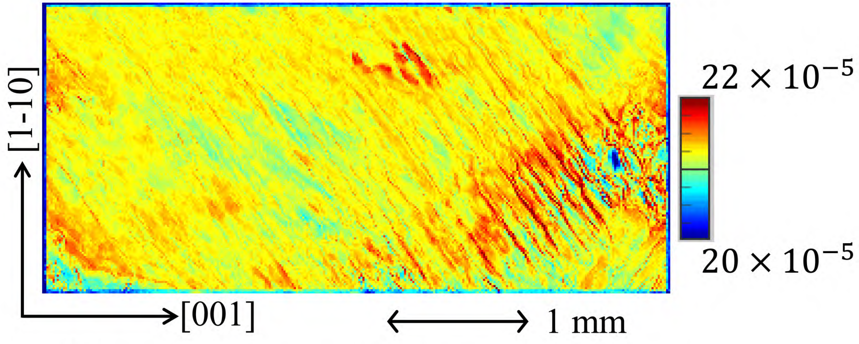
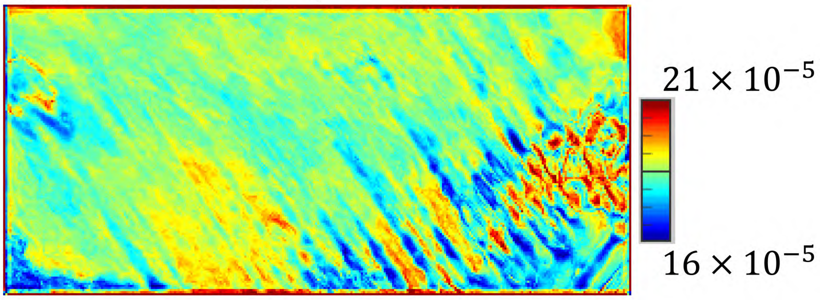


Figure S18. Comparison of the layer-wise relevance propagation (LRP) images for different SRF sizes at 523 nm: (a) 3×3, (b) 5×5, and (c) 10×10.

(a) LRP image (3 × 3) at $\lambda = 543$ nm



(b) LRP image (5 × 5) at $\lambda = 543$ nm



(c) LRP image (10 × 10) at $\lambda = 543$ nm

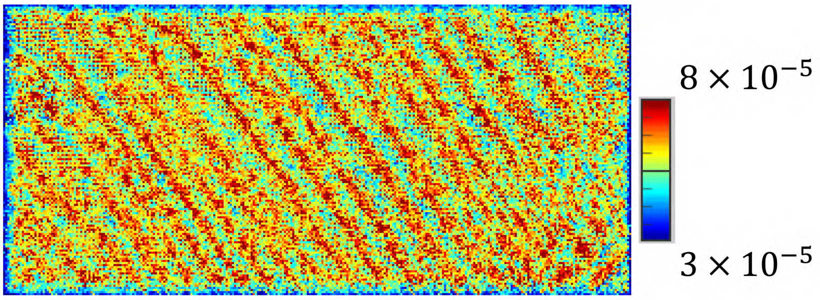
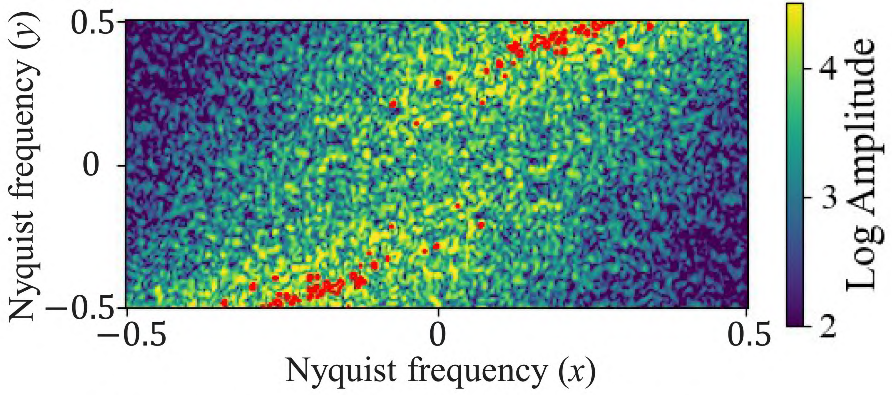
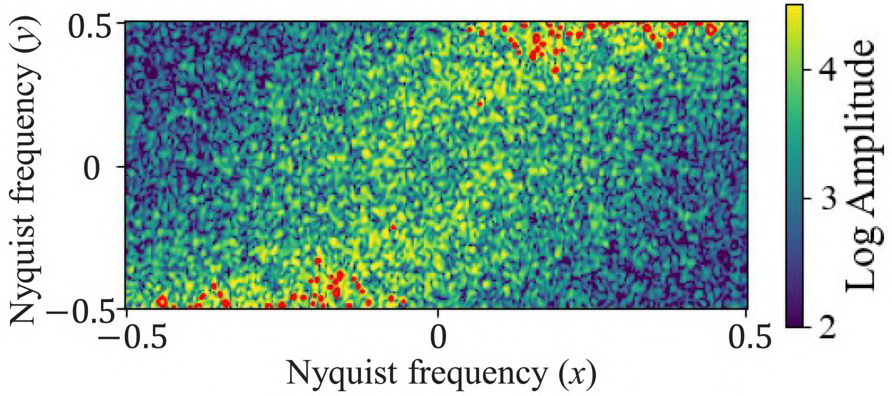


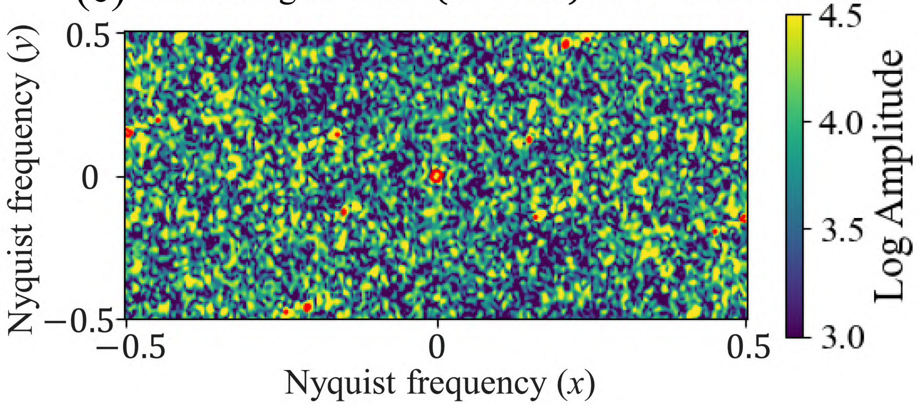
Figure S19. Comparison of the LRP images for different SRF sizes at 543 nm: (a) 3×3, (b) 5×5, and (c) 10×10.



(a) FFT image of ERF (3 × 3) at $\lambda = 523$ nm

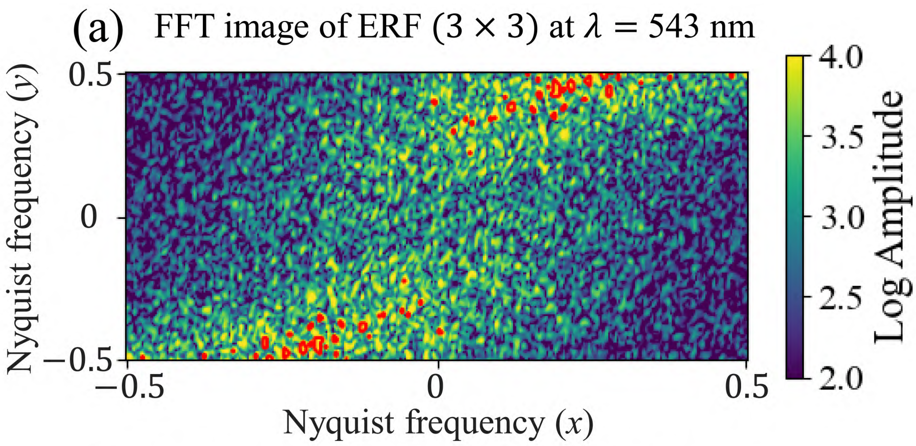


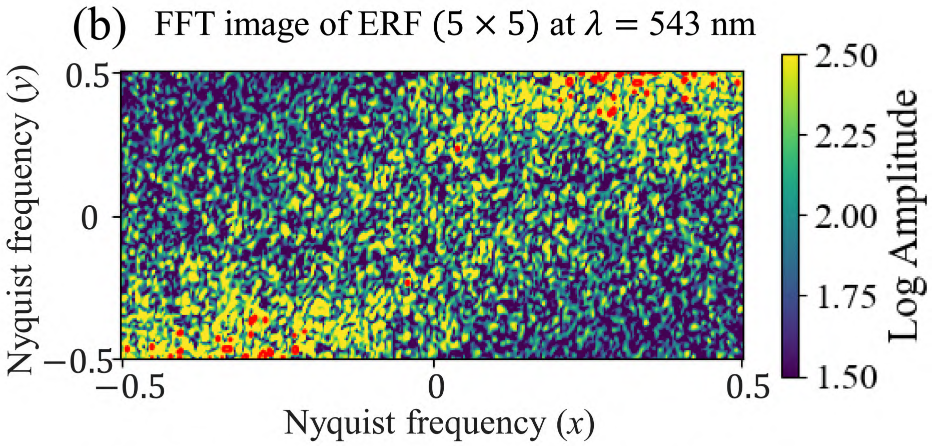
(b) FFT image of ERF (5 × 5) at $\lambda = 523$ nm



(c) FFT image of ERF (10 × 10) at $\lambda = 523$ nm

Figure S20. Comparison of the amplitude spectra of 2D fast Fourier-transform (FFT) ERF images at 523 nm for different SRF sizes: (a) 3×3, (b) 5×5, and (c) 10×10. Red circles indicate frequency components identified as significant at the 95% confidence level.


(a) FFT image of ERF (3 × 3) at $\lambda = 543$ nm


(b) FFT image of ERF (5 × 5) at $\lambda = 543$ nm

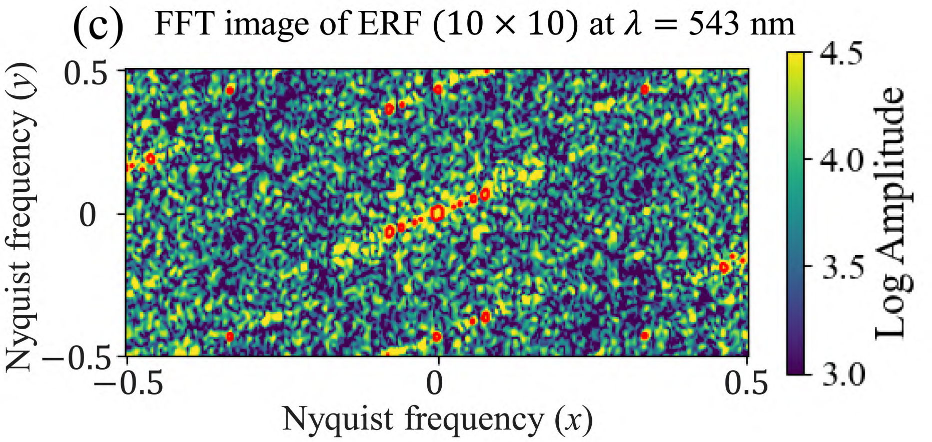
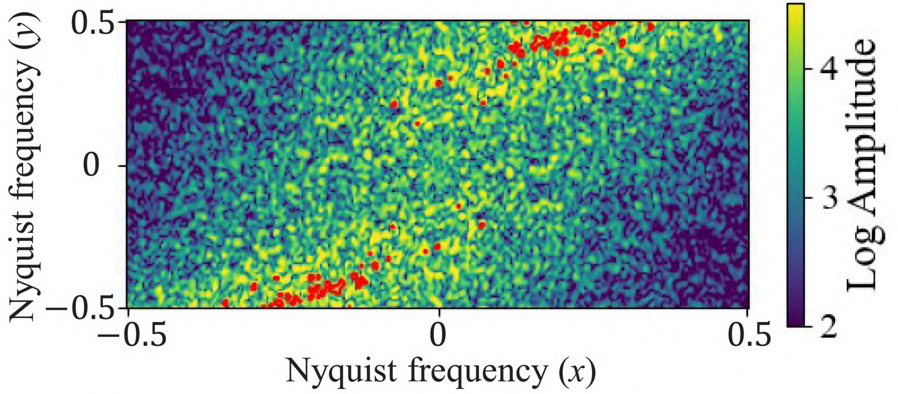
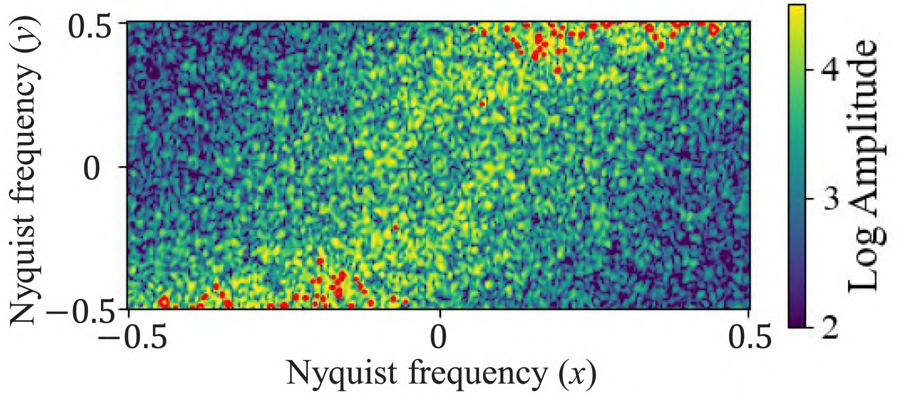

(c) FFT image of ERF (10 × 10) at $\lambda = 543$ nm

Figure S21. Comparison of the amplitude spectra of 2D FFT ERF images at 543 nm for different SRF sizes: (a) 3×3, (b) 5×5, and (c) 10×10. Red circles indicate frequency components identified as significant at the 95% confidence level.

(a) FFT image of LRP (3 × 3) at λ = 523 nm



(b) FFT image of LRP (5 × 5) at λ = 523 nm



(c) FFT image of LRP (10 × 10) at λ = 523 nm

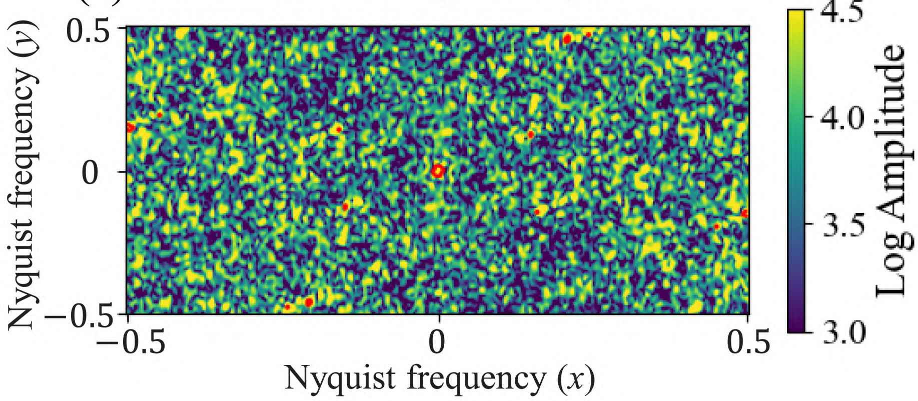
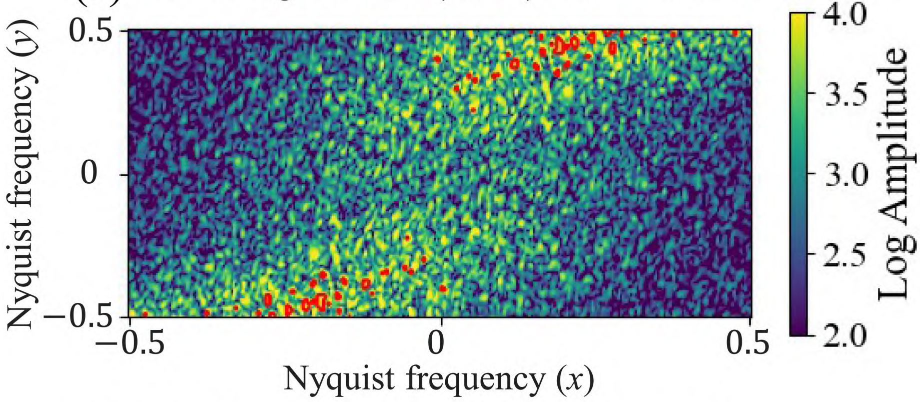
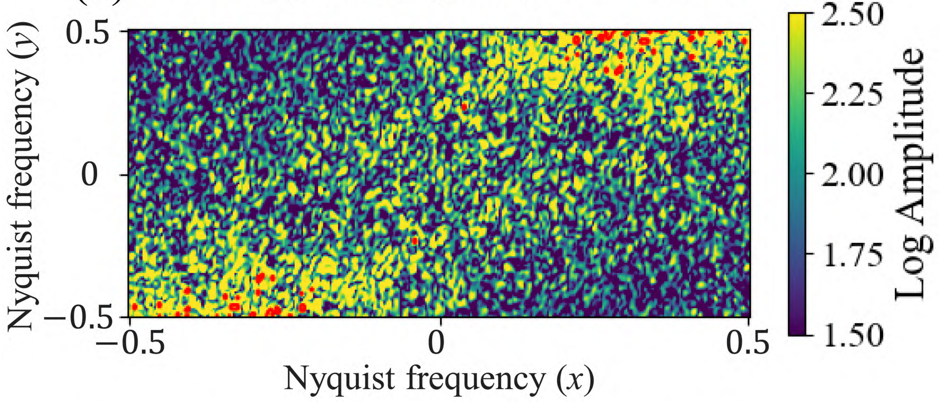


Figure S22. Comparison of the amplitude spectra of 2D FFT LRP images at 523 nm for different SRF sizes: (a) 3×3, (b) 5×5, and (c) 10×10. Red circles indicate frequency components identified as significant at the 95% confidence level.

(a) FFT image of LRP (3 × 3) at λ = 543 nm



(b) FFT image of LRP (5 × 5) at λ = 543 nm



(c) FFT image of LRP (10 × 10) at λ = 543 nm

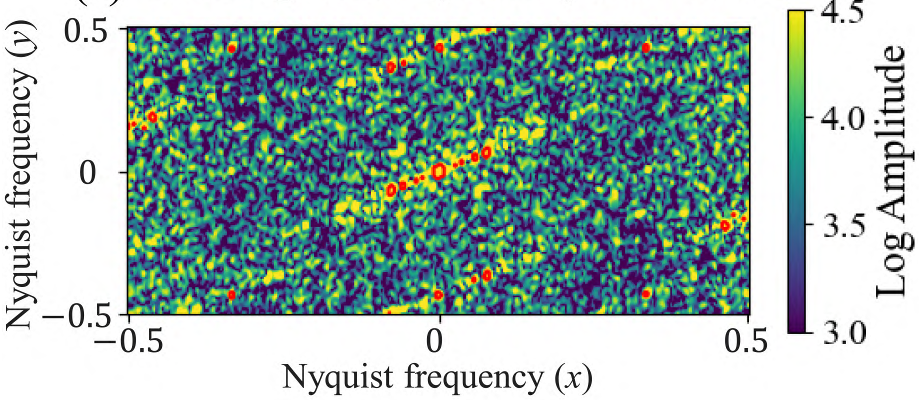
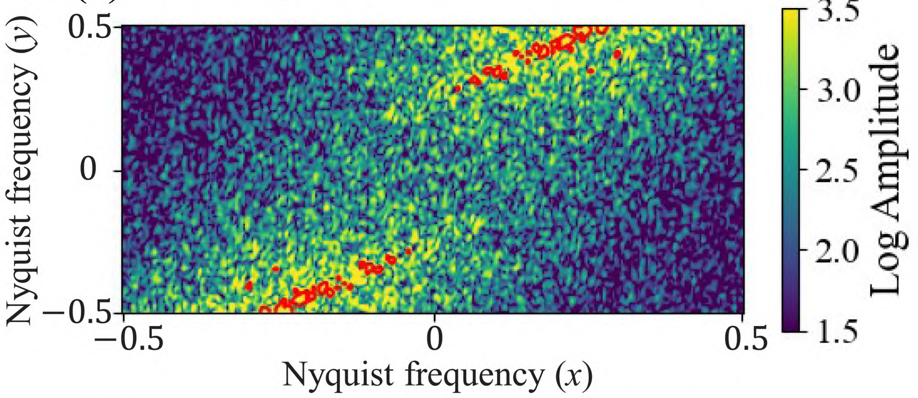
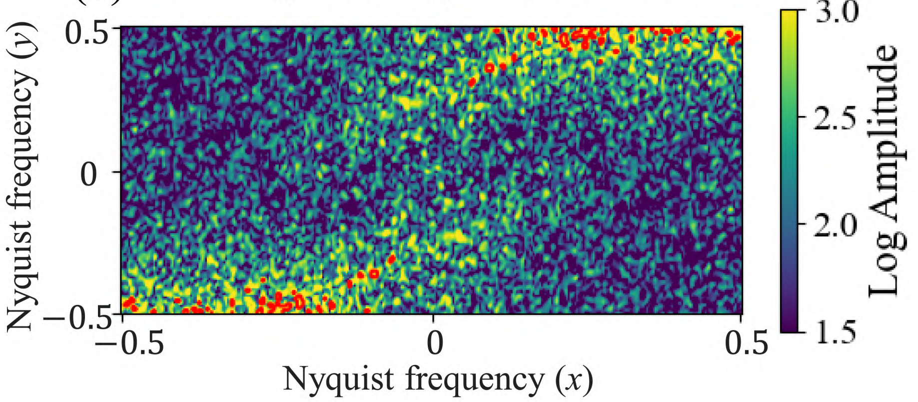


Figure S23. Comparison of the amplitude spectra of 2D FFT LRP images at 543 nm for different SRF sizes: (a) 3×3, (b) 5×5, and (c) 10×10. Red circles indicate frequency components identified as significant at the 95% confidence level.

(a) FFT image of LRP (3 × 3) at λ = 575 nm



(b) FFT image of LRP (5 × 5) at λ = 575 nm



(c) FFT image of LRP (10 × 10) ats λ = 575 nm

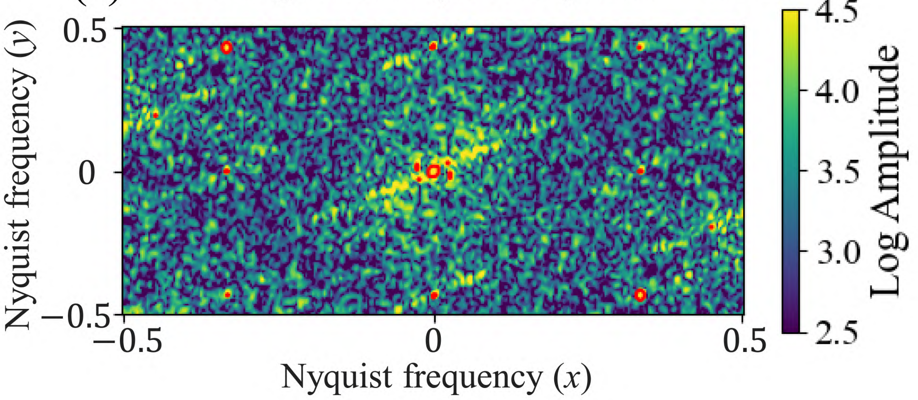
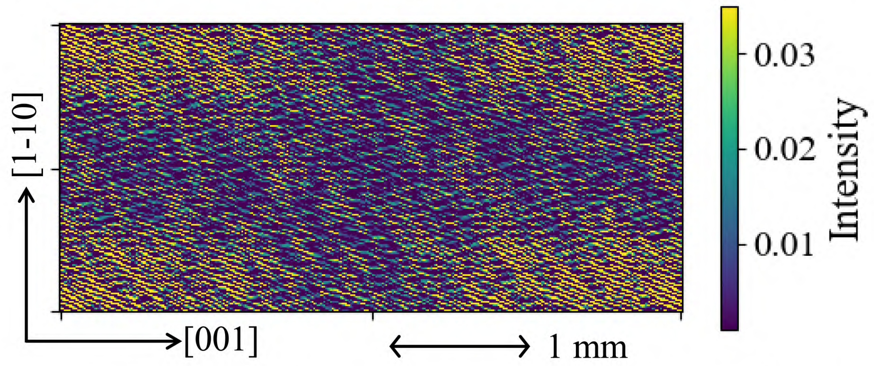
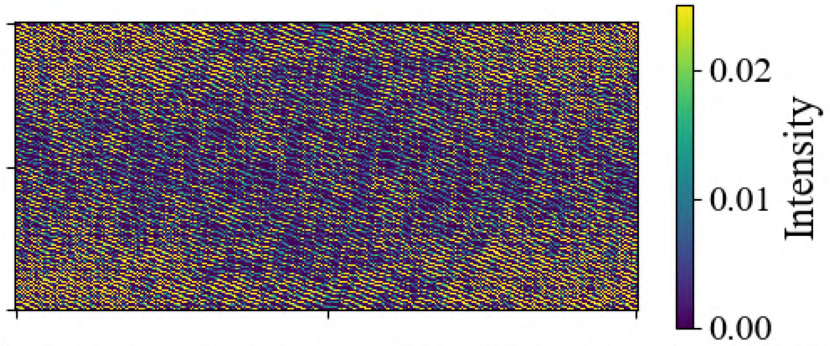


Figure S24. Comparison of the amplitude spectra of 2D FFT LRP images at 575 nm for different SRF sizes: (a) 3×3, (b) 5×5, and (c) 10×10. Red circles indicate frequency components identified as significant at the 95% confidence level.

(a) Reconstructed image of ERF (3 × 3) at λ = 523 nm



(b) Reconstructed image of ERF (5 × 5) at λ = 523 nm



(c) Reconstructed image of ERF (10 × 10) at λ = 523 nm

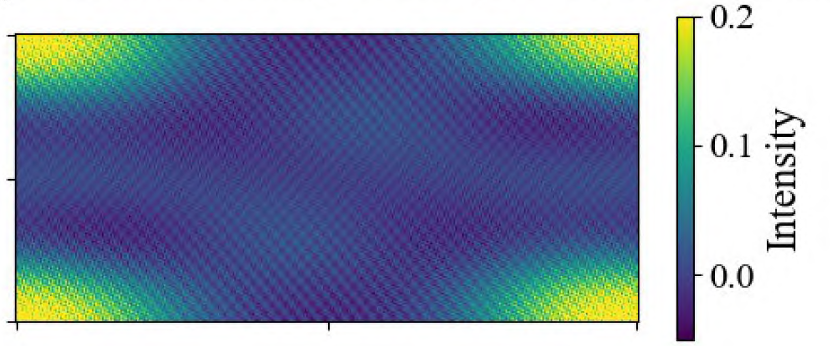
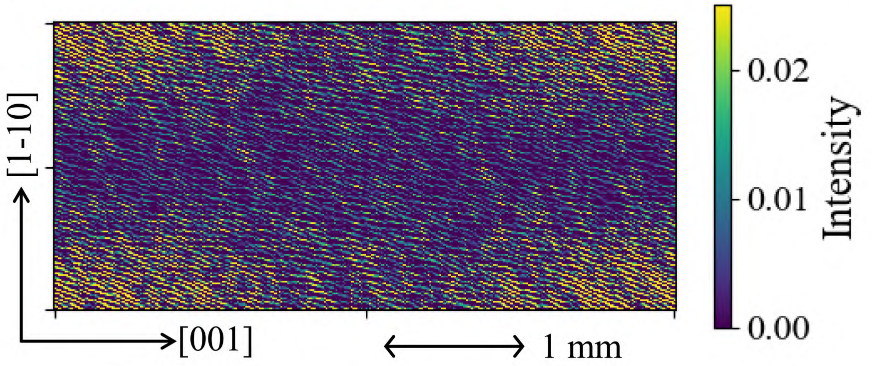
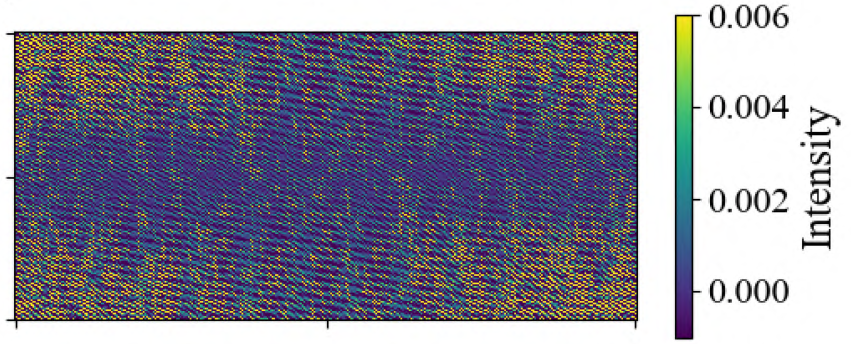


Figure S25. Comparison of inverse Fourier-transform ERF images at 523 nm for different SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10 using only the statistically significant frequency modes shown in Figure S20.

(a) Reconstructed image of ERF (3 × 3) at λ = 543 nm



(b) Reconstructed image of ERF (5 × 5) at λ = 543 nm



(c) Reconstructed image of ERF (10 × 10) at λ = 543 nm

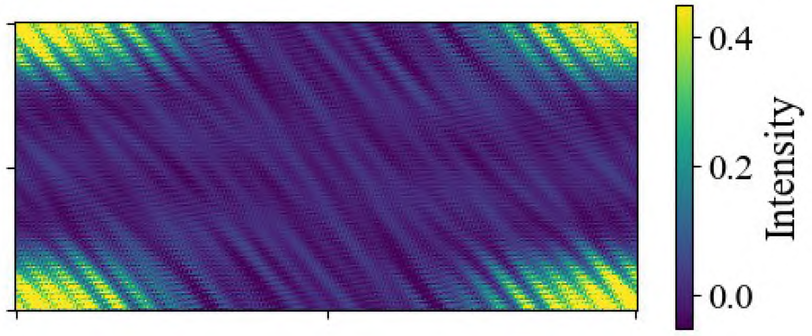
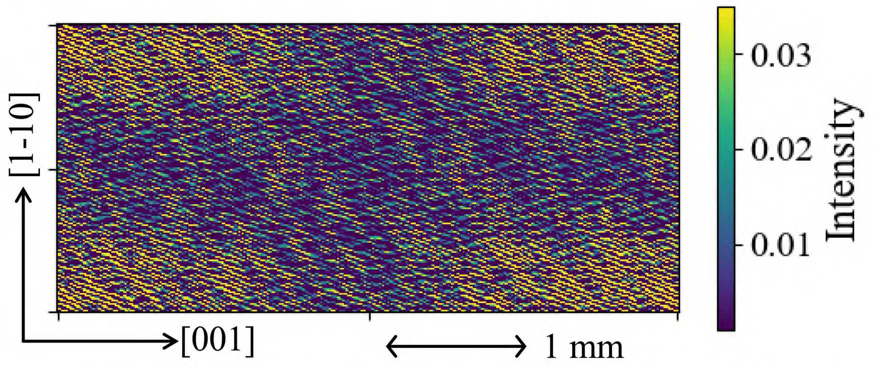
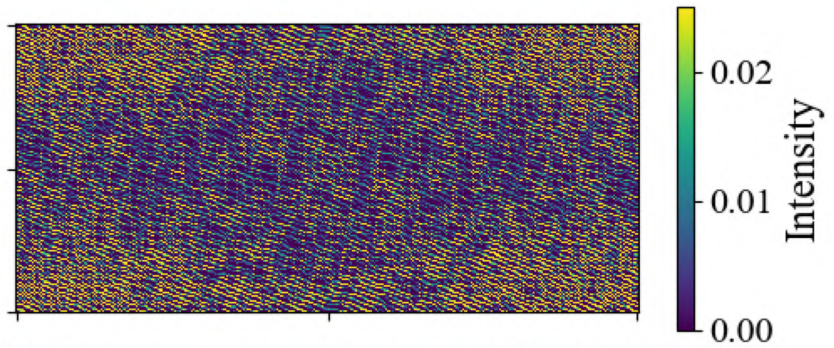


Figure S26. Comparison of inverse Fourier-transform ERF images at 543 nm for different SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10 using only the statistically significant frequency modes shown in Figure S21.

(a) Reconstructed image of LRP (3 × 3) at λ = 523 nm



(b) Reconstructed image of LRP (5 × 5) at λ = 523 nm



(c) Reconstructed image of LRP (10 × 10) at λ = 523 nm

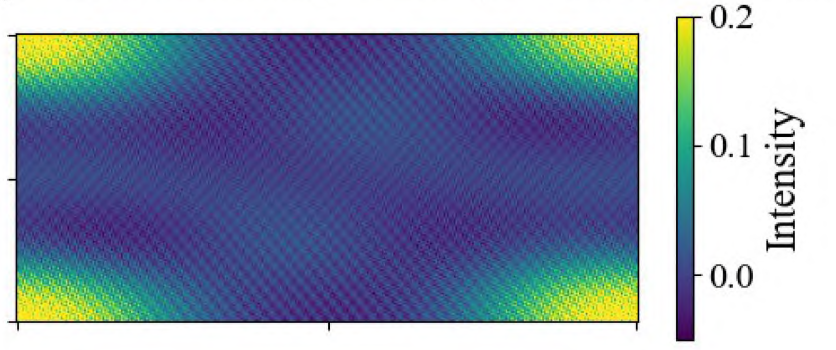
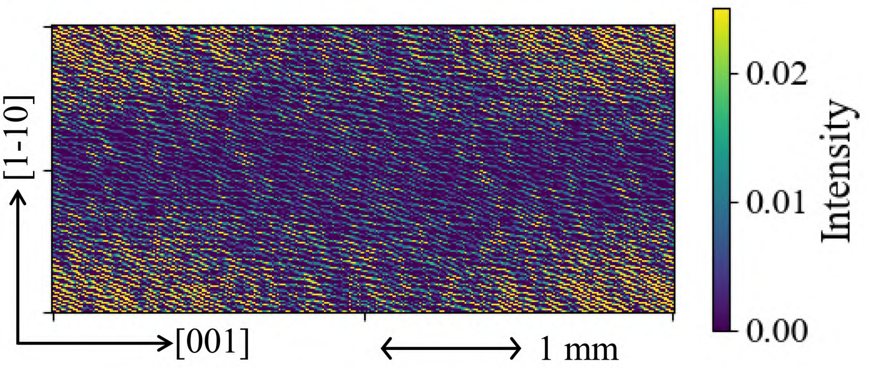
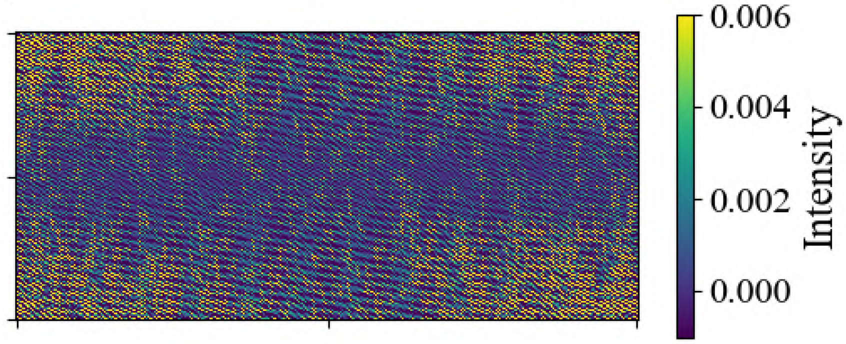


Figure S27. Comparison of inverse Fourier-transform LRP images at 523 nm for different SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10 using only the statistically significant frequency modes shown in Figure S22.

(a) Reconstructed image of LRP (3 × 3) at λ = 543 nm



(b) Reconstructed image of LRP (5 × 5) at λ = 543 nm



(c) Reconstructed image of LRP (10 × 10) at λ = 543 nm

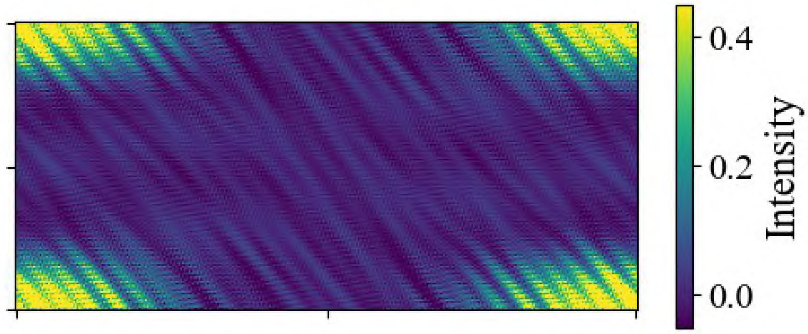
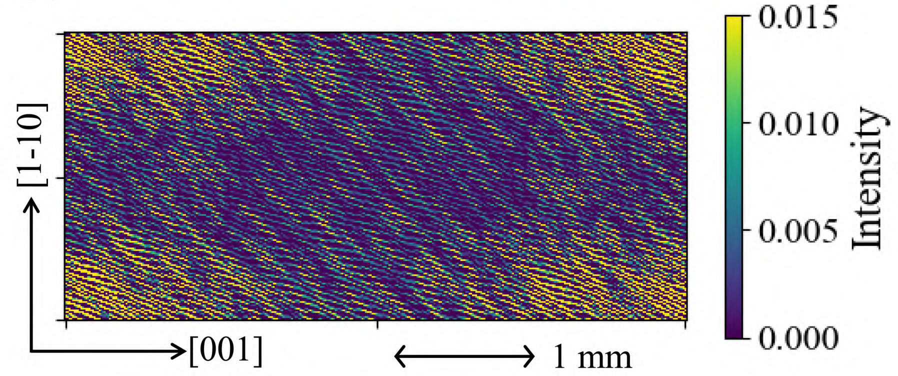
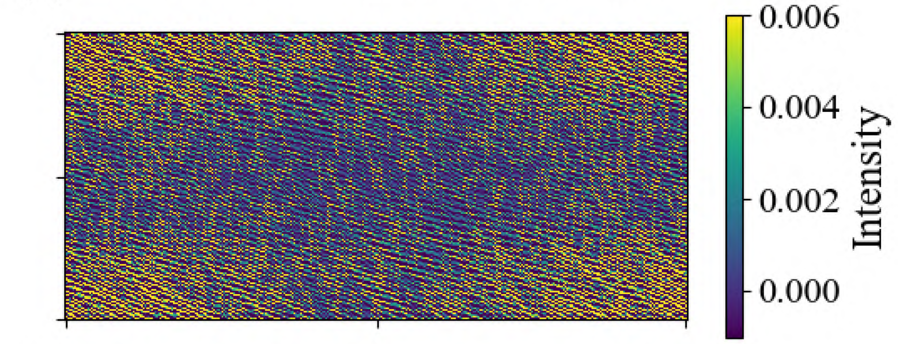


Figure S28. Comparison of inverse Fourier-transform LRP images at 543 nm for different SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10 using only the statistically significant frequency modes shown in Figure S23.

(a) Reconstructed image of LRP (3 × 3) at λ = 575 nm



(b) Reconstructed image of LRP (5 × 5) at λ = 575 nm



(c) Reconstructed image of LRP (10 × 10) at λ = 575 nm

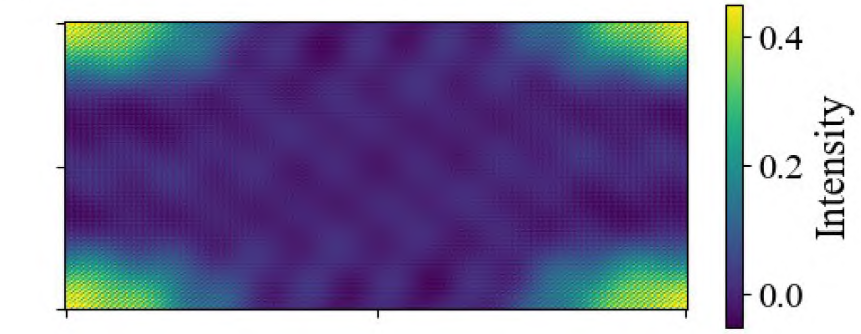
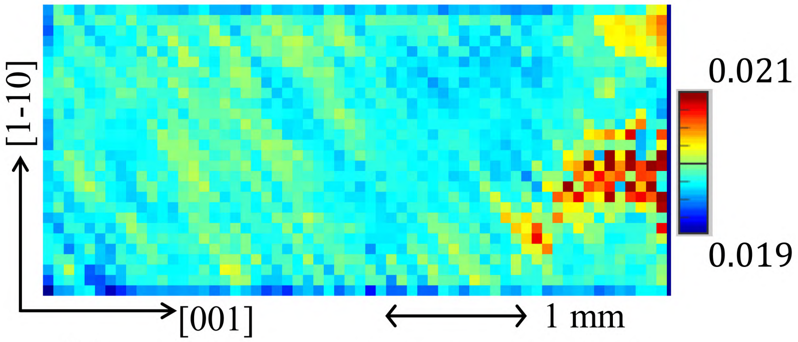
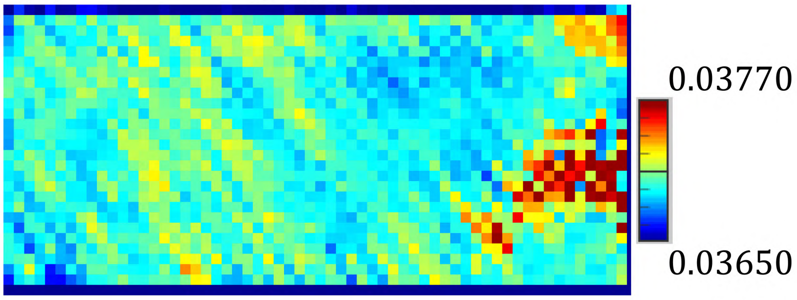


Figure S29. Comparison of inverse Fourier-transform LRP images at 575 nm for different SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10 using only the statistically significant frequency modes shown in Figure S24.

(a) OSA image (3 × 3) at λ = 523 nm



(b) OSA image (5 × 5) at λ = 523 nm



(c) OSA image (10 × 10) at λ = 523 nm

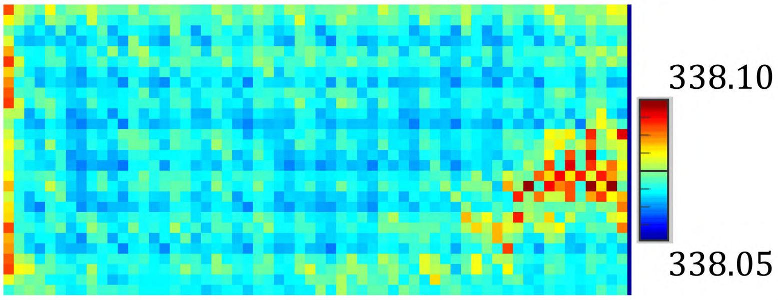
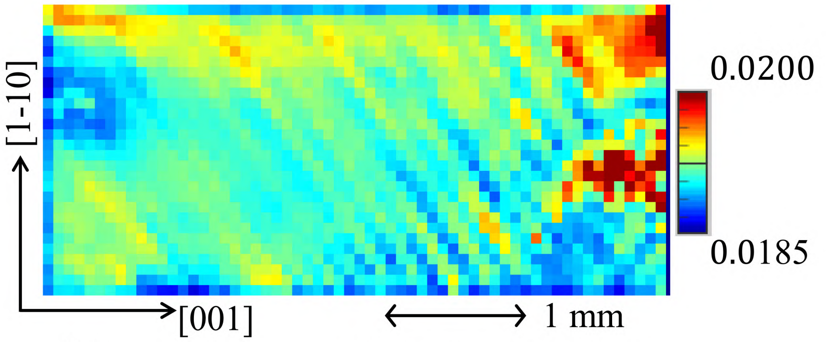
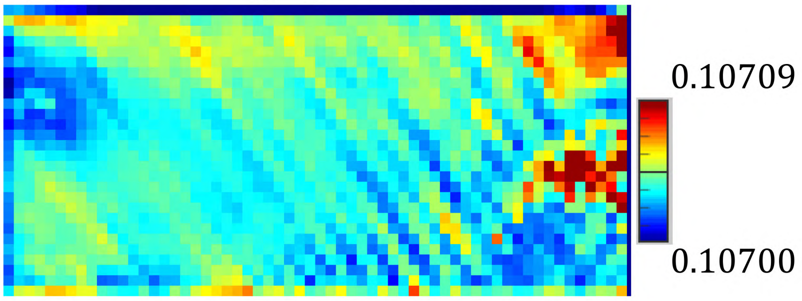


Figure S30. Comparison of occlusion sensitivity analysis (OSA) images for different SRF sizes at 523 nm: (a) 3×3, (b) 5×5, and (c) 10×10.

(a) OSA image (3×3) at $\lambda = 543$ nm



(b) OSA image (5×5) at $\lambda = 543$ nm



(c) OSA image (10×10) at $\lambda = 543$ nm

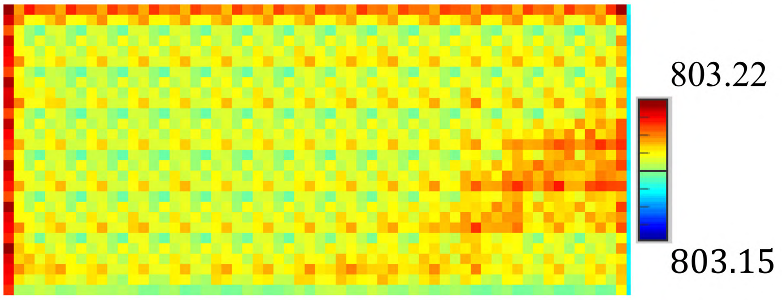


Figure S31. Comparison of OSA images for different SRF sizes at 543 nm: (a) 3×3, (b) 5×5, and (c) 10×10.

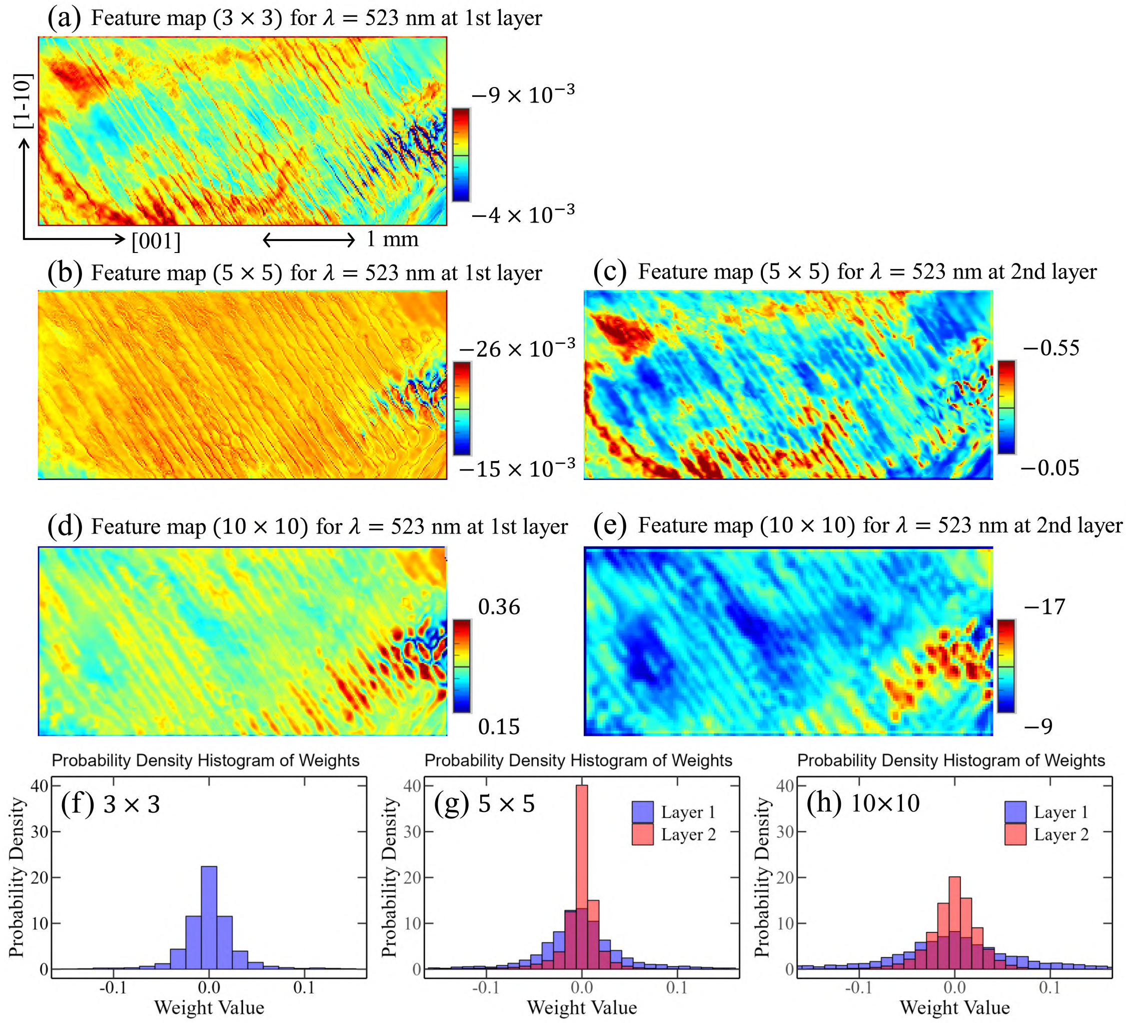


Figure S32. Comparison of feature maps at 523 nm for different SRF sizes: (a) 3×3, (b–c) 5×5, and (d–e) 10×10. The 3×3 case includes a single-layer structure, whereas the 5×5 and 10×10 cases include two layers. Histograms of the feature weights are shown for (f) 3×3, (g) 5×5, and (h) 10×10.

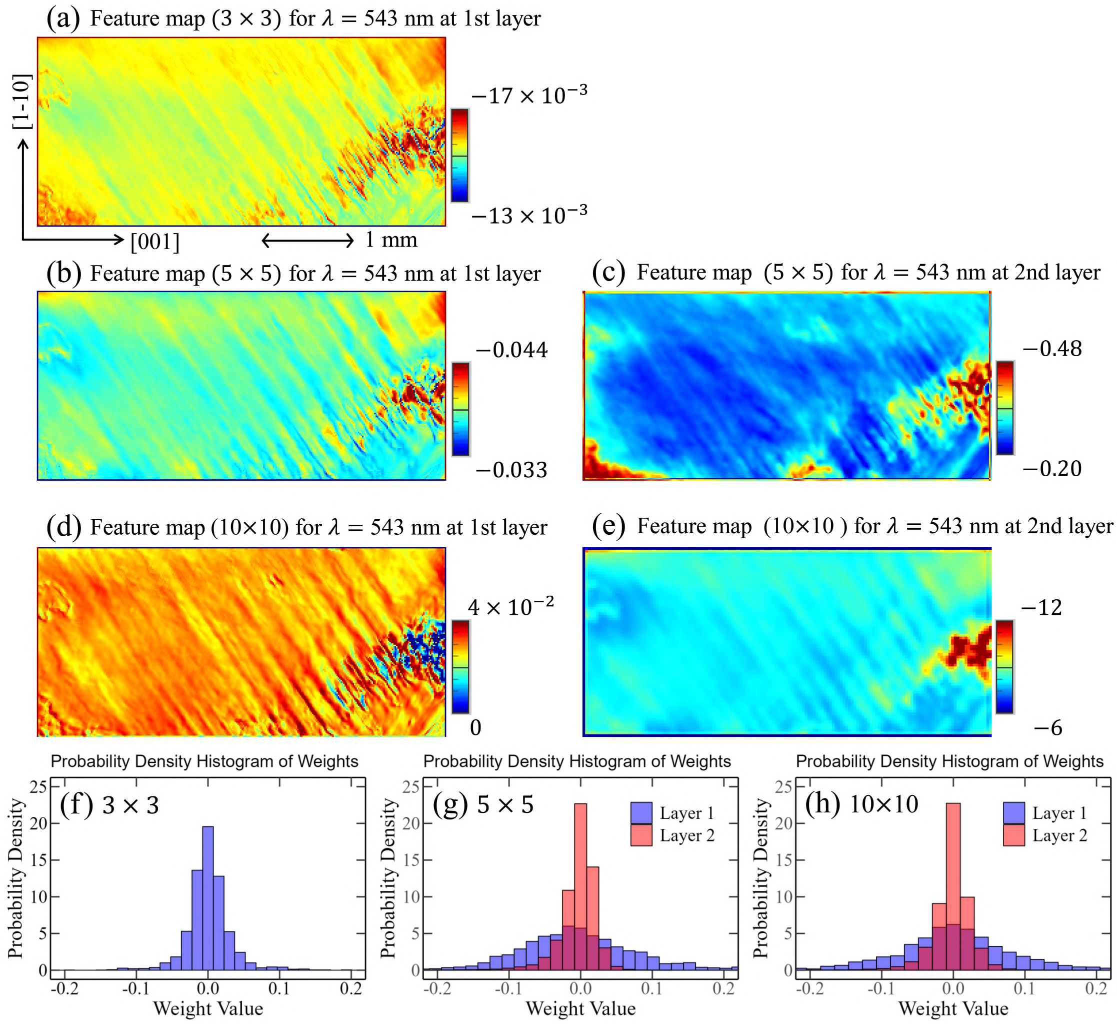


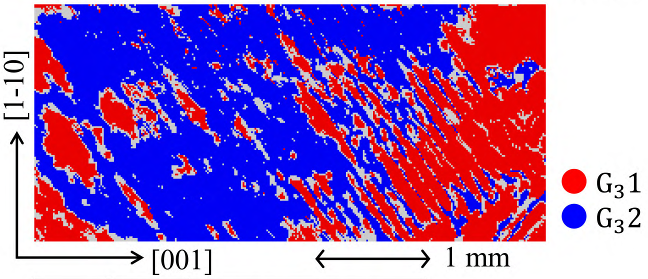
Figure S33. Comparison of feature maps at 543 nm for different SRF sizes: (a) 3×3, (b–c) 5×5, and (d–e) 10×10. The 3×3 case includes a single-layer structure, whereas the 5×5 and 10×10 cases include two layers. Histograms of the feature weights are shown for (f) 3×3, (g) 5×5, and (h) 10×10.

4. Temperature series forest (Tsf)

The initial annotations for the bagging-based temperature series forests (Tsf) are derived from the K-shape multivariate clustering results shown in Figures S12–S15. A pixel is classified as a "consistent cluster" if it receives the same winning cluster label across all four analysis conditions listed in Table 1. Figures S34–S37 show the clustering results based only on "consistent clusters" for each initial cluster number (k) and SRF size, along with a transition matrix that tracks changes from the initial annotations. These results show that stress-concentration and uniform regions remain stable, whereas other clusters show greater variability and tend to merge with clusters in uniform regions. After applying Tsf with cluster confidence-based adaptive learning (CCAL), which incorporates cluster confidence-based adaptive learning, the final clustering results for each k and SRF size are shown in Figures S38–S41. In addition, the transition matrix tracks the label changes between initial annotations, "consistent clusters", and the final cluster labels.

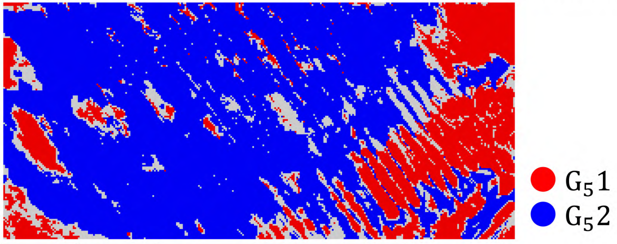
To evaluate the clustering performance results shown in Figures S38–S41, we computed the Silhouette score for two types of T-series datasets: LSTM input features (3 λ, 1 Mh, 233 T, 42,280 pixels), as shown in Figure 5, and 3DCAE output features (3 λ, 5 PC's, 233 T, 42,280 pixels), as shown in Figures 14(b–c). Figure S42 shows the variation in the Silhouette score as a function of the initial cluster number (k) using a sliding T window width of 19 T points (9.5 K). The highest Silhouette score was observed at $k = 2$ for the SRF sizes of 3×3 and 10×10. For the 5×5 case, the highest score was obtained at $k = 3$, but the Tsf process merged three clusters into two, resulting in a consistent two-group classification across all SRF sizes. Based on the two-group classifications obtained from the 3×3 case at $k = 2$ shown in Figure S38(a), Figures S43 and S44 show the distributions of the retardance (δ) and fast-axis direction (ψ) at $T = 14.1$, 40.0, 90.0, and 130.9 K divided into these two groups. The results for the cases 5×5 with $k = 3$ in Figure S39(b) and 10×10 with $k = 2$ in Figure S38(c) are also shown in Figures S45–S48.

(a) Consistent clusters (3 × 3) for initial annotation of $k = 2$



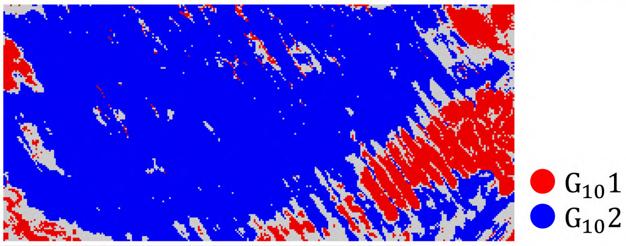
		Consistent clusters		
		$G_{3}1$	$G_{3}2$	N/A
Initial annotations	$F_{3}1$	12,870	2,866	3,923
	$F_{3}2$	688	19,813	2,120

(b) Consistent clusters (5 × 5) for initial annotation of $k = 2$



		Consistent clusters		
		$G_{5}1$	$G_{5}2$	N/A
Initial annotations	$F_{5}1$	7,954	5,162	3,679
	$F_{5}2$	457	23,000	2,028

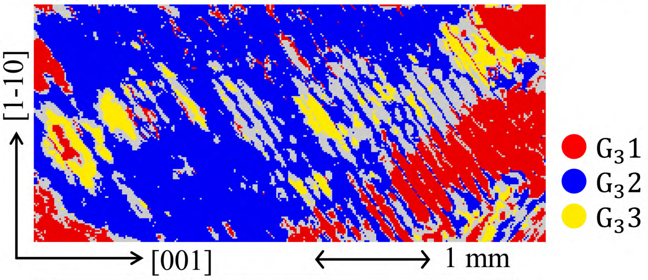
(c) Consistent clusters (10 × 10) for initial annotation of $k = 2$



		Consistent clusters		
		$G_{10}1$	$G_{10}2$	N/A
Initial annotations	$F_{10}1$	5,138	1,987	4,225
	$F_{10}2$	299	28,782	1,849

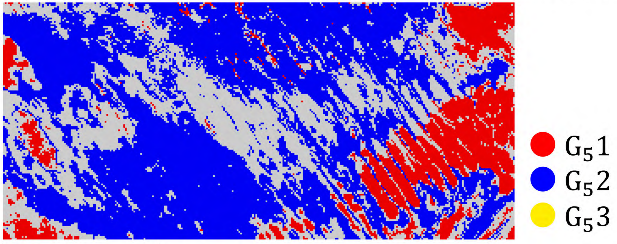
Figure S34. Distributions of "consistent clusters" derived from Tsf using bagging-based learning for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations ($k = 2$) obtained from K-shape clustering (Figure S12) are used as the reference. The transition matrix between the initial annotations and the "consistent cluster" labels is included. Pixels not classified as "consistent clusters" are labeled as N/A ("not applicable").

(a) Consistent clusters (3 × 3) for initial annotation of $k = 3$



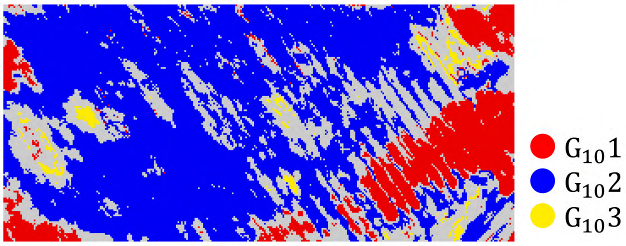
		Consistent clusters			
Initial annotations		$G_{3}1$	$G_{3}2$	$G_{3}3$	N/A
	$F_{3}1$	7,188	1,444	176	3,959
	$F_{3}2$	106	19,568	148	2,146
	$F_{3}3$	400	1,720	2,291	3,134

(b) Consistent clusters (5 × 5) for initial annotation of $k = 3$



		Consistent clusters			
Initial annotations		$G_{5}1$	$G_{5}2$	$G_{5}3$	N/A
	$F_{5}1$	5,799	942	0	3,605
	$F_{5}2$	197	19,502	0	4,198
	$F_{5}3$	41	2,041	0	5,955

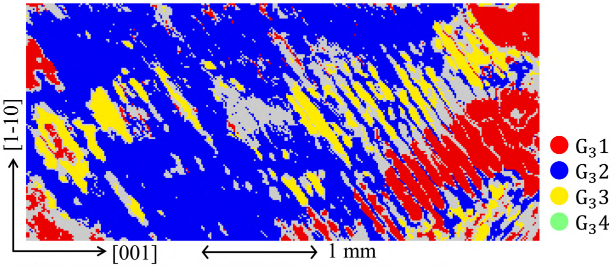
(c) Consistent clusters (10 × 10) for initial annotation of $k = 3$



		Consistent clusters			
Initial annotations		$G_{10}1$	$G_{10}2$	$G_{10}3$	N/A
	$F_{10}1$	5,028	951	42	3,789
	$F_{10}2$	125	20,267	1	2,676
	$F_{10}3$	484	3,467	637	4,813

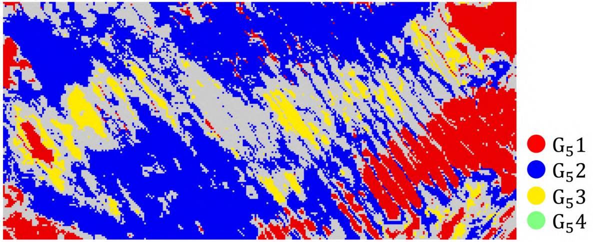
Figure S35. Distributions of "consistent clusters" derived from Tsf using bagging-based learning for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations (k = 3) obtained from K-shape clustering (Figure S13) are used as the reference. The transition matrix between the initial annotations and the "consistent cluster" labels is included. Pixels not classified as "consistent clusters" are labeled as N/A ("not applicable").

(a) Consistent clusters (3 × 3) for initial annotation of $k = 4$



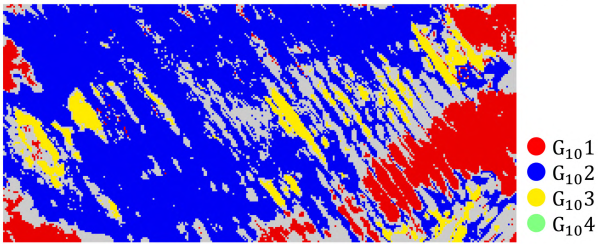
		Consistent clusters				
		$\mathrm{G}_{3}1$	$\mathrm{G}_{3}2$	$\mathrm{G}_{3}3$	$\mathrm{G}_{3}4$	N/A
Initial annotations	$\mathrm{F}_{3}1$	5,544	1,017	294	0	3,501
	$\mathrm{F}_{3}2$	24	18,354	79	0	1,429
	$\mathrm{F}_{3}3$	173	2,577	3,416	0	3,214
	$\mathrm{F}_{3}4$	342	728	266	0	1,322

(b) Consistent clusters (5 × 5) for initial annotation of $k = 4$



		Consistent clusters				
		$\mathrm{G}_{5}1$	$\mathrm{G}_{5}2$	$\mathrm{G}_{5}3$	$\mathrm{G}_{5}4$	N/A
Initial annotations	$\mathrm{F}_{5}1$	6,298	580	111	0	2,871
	$\mathrm{F}_{5}2$	146	15,198	135	0	2,923
	$\mathrm{F}_{5}3$	58	2,412	2,137	0	7,625
	$\mathrm{F}_{5}4$	90	364	156	0	1,176

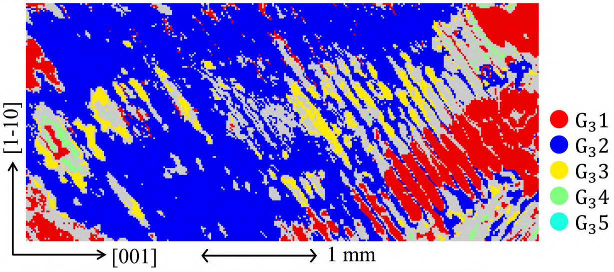
(c) Consistent clusters (10 × 10) for initial annotation of $k = 4$



		Consistent clusters				
		$\mathrm{G}_{10}1$	$\mathrm{G}_{10}2$	$\mathrm{G}_{10}3$	$\mathrm{G}_{10}4$	N/A
Initial annotations	$\mathrm{F}_{10}1$	4,959	781	98	0	3,773
	$\mathrm{F}_{10}2$	113	17,010	190	0	2,267
	$\mathrm{F}_{10}3$	160	2,517	2,062	0	2,112
	$\mathrm{F}_{10}4$	475	3,413	256	0	2,094

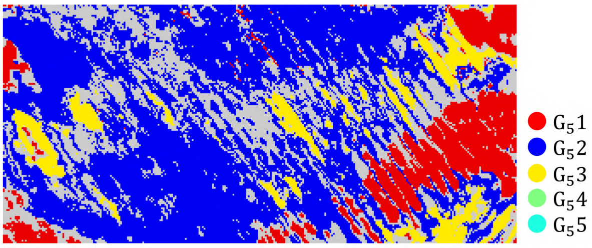
Figure S36. Distributions of "consistent clusters" derived from Tsf using bagging-based learning for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations ($k = 4$) obtained from K-shape clustering (Figure S14) are used as the reference. The transition matrix between the initial annotations and the "consistent cluster" labels is included. Pixels not classified as "consistent clusters" are labeled as N/A ("not applicable").

(a) Consistent clusters (3 × 3) for initial annotation of *k* = 5



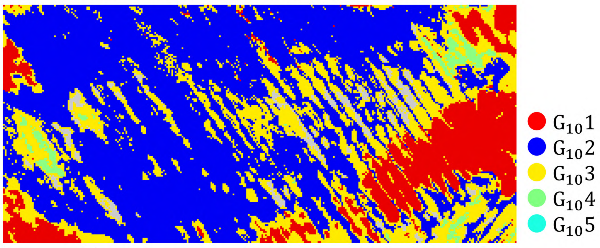
		Consistent clusters					
		$G_3 1$	$G_3 2$	$G_3 3$	$G_3 4$	$G_3 5$	N/A
Initial annotations	$F_3 1$	5,684	841	132	73	0	3,002
	$F_3 2$	37	17,944	53	0	0	1,049
	$F_3 3$	205	3,211	1,861	24	0	3,783
	$F_3 4$	278	331	207	247	0	1,502
	$F_3 5$	18	1,008	6	0	0	784

(b) Consistent clusters (5 × 5) for initial annotation of *k* = 5



		Consistent clusters					
		$G_5 1$	$G_5 2$	$G_5 3$	$G_5 4$	$G_5 5$	N/A
Initial annotations	$F_5 1$	5,076	547	21	0	0	2,400
	$F_5 2$	75	15,023	47	0	0	1,859
	$F_5 3$	86	894	2,905	0	0	1,661
	$F_5 4$	14	4,969	148	0	0	4,617
	$F_5 5$	15	623	0	0	0	1,300

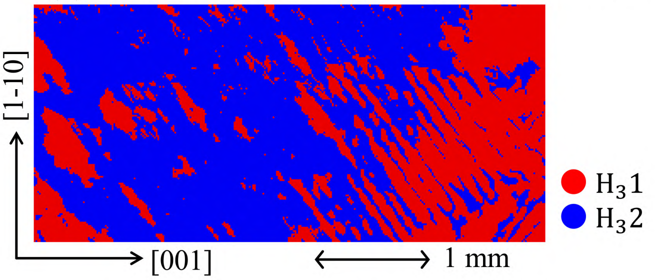
(c) Consistent clusters (10×10) for initial annotation of *k* = 5



		Consistent clusters					
		$G_{10} 1$	$G_{10} 2$	$G_{10} 3$	$G_{10} 4$	$G_{10} 5$	N/A
Initial annotations	$F_{10} 1$	4,780	749	6	22	0	3,161
	$F_{10} 2$	71	16,011	34	0	0	1,701
	$F_{10} 3$	97	2,818	970	20	0	2,421
	$F_{10} 4$	319	429	34	688	0	1,842
	$F_{10} 5$	362	4,186	0	0	0	1,559

Figure S37. Distributions of "consistent clusters" derived from Tsf using bagging-based learning for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations ($k = 5$) obtained from K-shape clustering (Figure S15) are used as the reference. The transition matrix between the initial annotations and the "consistent cluster" labels is included. Pixels not classified as "consistent clusters" are labeled as N/A ("not applicable").

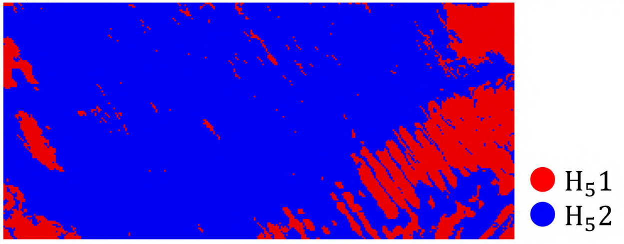
(a) Final clustering result (3 × 3) for initial annotation of $k = 2$



		Final clusters	
		$H_3 1$	$H_3 2$
Initial annotations	$F_3 1$	14,422	5,237
	$F_3 2$	1,010	21,611

		Final clusters	
		$H_3 1$	$H_3 2$
Consistent clusters	$G_3 1$	13,495	63
	$G_3 2$	0	22,679
	N/A	1,937	4,106

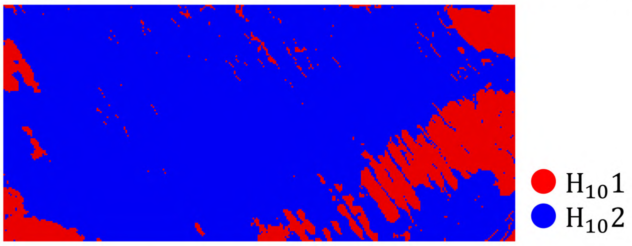
(b) Final clustering result (5 × 5) for initial annotation of $k = 2$



		Final clusters	
		$H_5 1$	$H_5 2$
Initial annotations	$F_5 1$	8,616	8,179
	$F_5 2$	913	24,572

		Final clusters	
		$H_5 1$	$H_5 2$
Consistent clusters	$G_5 1$	8,180	231
	$G_5 2$	0	28,162
	N/A	1,349	4,358

(c) Final clustering result (10 × 10) for initial annotation of $k = 2$

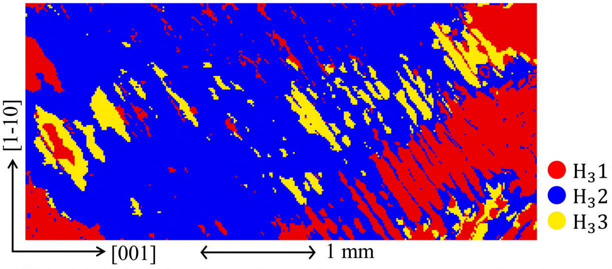


		Final clusters	
		$H_{10} 1$	$H_{10} 2$
Initial annotations	$F_{10} 1$	6,462	4,888
	$F_{10} 2$	508	30,422

		Final clusters	
		$H_{10} 1$	$H_{10} 2$
Consistent clusters	$G_{10} 1$	5,405	32
	$G_{10} 2$	0	30,769
	N/A	1,565	4,509

Figure S38. Final clustering results obtained using cluster confidence-based adaptive learning (CCAL) for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations ($k = 2$) derived from the K-shape multivariate clustering (Figure S12) are used as the reference. The transition matrix between the initial annotations, "consistent cluster" labels, and final cluster labels is included. Pixels not classified as the "consistent clusters" during the bagging-based learning phase are labeled as N/A ("not applicable").

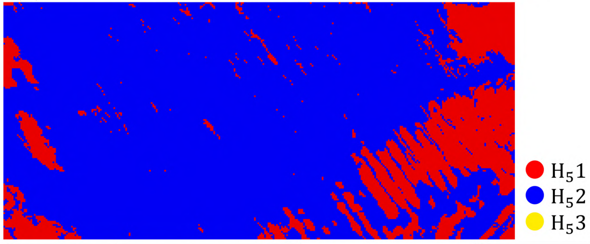
(a) Final clustering result (3 × 3) for initial annotation of k = 3



		Final clusters		
		$H_3$1	$H_3$2	$H_3$3
Initial annotations	$F_3$1	8,431	3,694	642
	$F_3$2	146	21,420	402
	$F_3$3	660	3,853	3,032

		Final clusters		
		$H_3$1	$H_3$2	$H_3$3
Consistent clusters	$G_3$1	7,694	0	0
	$G_3$2	0	22,732	0
	$G_3$3	0	0	2,615
	N/A	1,543	6,235	1,461

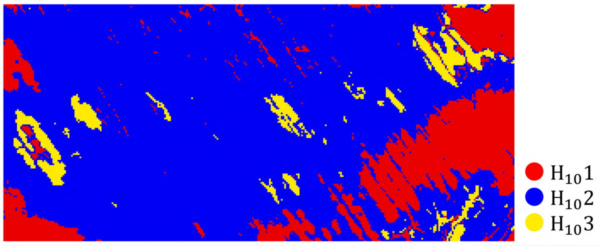
(b) Final clustering result (5 × 5) for initial annotation of k = 3



		Final clusters		
		$H_5$1	$H_5$2	$H_5$3
Initial annotations	$F_5$1	7,505	2,841	0
	$F_5$2	372	23,525	0
	$F_5$3	71	7,966	0

		Final clusters		
		$H_5$1	$H_5$2	$H_5$3
Consistent clusters	$G_5$1	6,037	0	0
	$G_5$2	0	22,485	0
	$G_5$3	0	0	0
	N/A	1,911	11,847	0

(c) Final clustering result (10 × 10) for initial annotation of k = 3

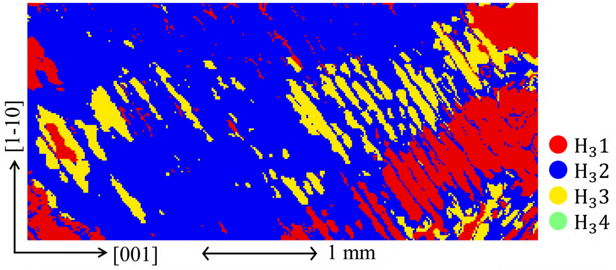


		Final clusters		
		H_{10}1	H_{10}2	H_{10}3
Initial annotations	F_{10}1	6,414	3,314	82
	F_{10}2	248	22,745	76
	F_{10}3	644	6,779	1,978

		Final clusters		
		H_{10}1	H_{10}2	H_{10}3
Consistent clusters	G_{10}1	5,637	0	0
	G_{10}2	0	24,685	0
	G_{10}3	0	0	680
	N/A	1,669	8,153	1,456

Figure S39. Final clustering results obtained using CCAL for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations ($k = 3$) derived from the K-shape multivariate clustering (Figure S13) are used as the reference. The transition matrix between the initial annotations, "consistent cluster" labels, and final cluster labels is included. Pixels not classified as the "consistent clusters" during the bagging-based learning phase are labeled as N/A ("not applicable").

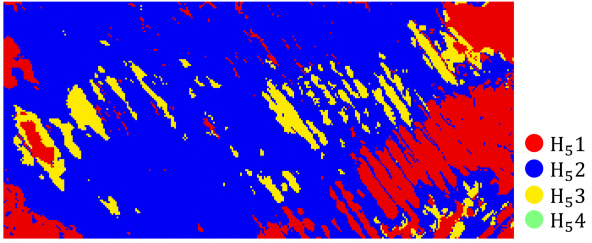
(a) Final clustering result (3 × 3) for initial annotation of $k = 4$



		Final clusters			
		$H_3 1$	$H_3 2$	$H_3 3$	$H_3 4$
Initial annotations	$F_3 1$	6,703	2,966	687	0
	$F_3 2$	41	19,667	178	0
	$F_3 3$	245	4,647	4,488	0
	$F_3 4$	503	1,852	303	0

		Final clusters			
		$H_3 1$	$H_3 2$	$H_3 3$	$H_3 4$
Consistent clusters	$G_3 1$	6,083	0	0	0
	$G_3 2$	0	22,676	0	0
	$G_3 3$	0	0	4,055	0
	$G_3 4$	0	0	0	0
	N/A	1,409	6,456	1,601	0

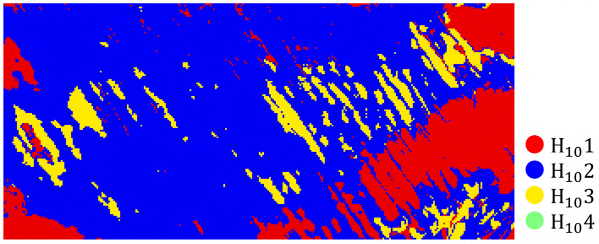
(b) Final clustering result (5 × 5) for initial annotation of $k = 4$



		Final clusters			
		$H_5 1$	$H_5 2$	$H_5 3$	$H_5 4$
Initial annotations	$F_5 1$	7,296	2,141	423	0
	$F_5 2$	348	17,528	526	0
	$F_5 3$	171	9,273	2,788	0
	$F_5 4$	98	1,481	207	0

		Final clusters			
		$H_5 1$	$H_5 2$	$H_5 3$	$H_5 4$
Consistent clusters	$G_5 1$	6,592	0	0	0
	$G_5 2$	0	18,554	0	0
	$G_5 3$	0	0	2,539	0
	$G_5 4$	0	0	0	0
	N/A	1,321	11,869	1,405	0

(c) Final clustering result (10 × 10) for initial annotation of $k = 4$

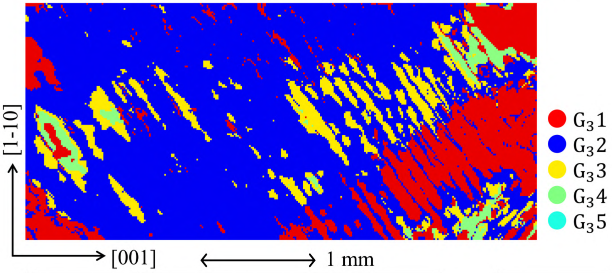


		Final clusters			
		$H_{10} 1$	$H_{10} 2$	$H_{10} 3$	$H_{10} 4$
Initial annotations	$F_{10} 1$	6,272	3,110	229	0
	$F_{10} 2$	193	18,842	545	0
	$F_{10} 3$	197	3,647	3,007	0
	$F_{10} 4$	641	5,293	304	0

		Final clusters			
		$H_{10} 1$	$H_{10} 2$	$H_{10} 3$	$H_{10} 4$
Consistent clusters	$G_{10} 1$	5,707	0	0	0
	$G_{10} 2$	0	23,721	0	0
	$G_{10} 3$	0	0	2,606	0
	$G_{10} 4$	0	0	0	0
	N/A	1,596	7,171	1,479	0

Figure S40. Final clustering results obtained using CCAL for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations ($k = 4$) derived from the K-shape multivariate clustering (Figure S14) are used as the reference. The transition matrix between the initial annotations, "consistent cluster" labels, and final cluster labels is included. Pixels not classified as the "consistent clusters" during the bagging-based learning phase are labeled as N/A ("not applicable").

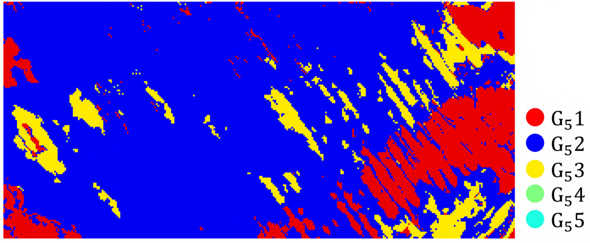
(a) Final clustering result (3 × 3) for initial annotation of $k = 5$



		Final clusters				
		H_31	H_32	H_33	H_34	H_35
Initial annotations	F_31	6,895	2,424	310	103	0
	F_32	55	18,915	113	0	0
	F_33	277	5,071	3,576	160	0
	F_34	443	676	462	984	0
	F_35	39	1,766	11	0	0

		Final clusters				
		H_31	H_32	H_33	H_34	H_35
Consistent clusters	G_31	6,222	0	0	0	0
	G_32	0	23,335	0	0	0
	G_33	0	0	2,259	0	0
	G_34	0	0	0	344	0
	G_35	0	0	0	0	0
	N/A	1,487	5,517	2,213	903	0

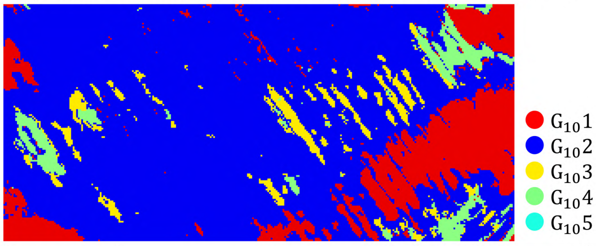
(b) Final clustering result (5 × 5) for initial annotation of $k = 5$



		Final clusters				
		H_51	H_52	H_53	H_54	H_55
Initial annotations	F_51	5,744	2,139	161	0	0
	F_52	184	16,667	153	0	0
	F_53	205	1,922	3,417	2	0
	F_54	50	9,485	208	5	0
	F_55	16	1,922	0	0	0

		Final clusters				
		H_51	H_52	H_53	H_54	H_55
Consistent clusters	G_51	5,266	0	0	0	0
	G_52	0	22,056	0	0	0
	G_53	0	0	3,121	0	0
	G_54	0	0	0	0	0
	G_55	0	0	0	0	0
	N/A	933	10,079	818	7	0

(c) Final clustering result (10 × 10) for initial annotation of $k = 5$



		Final clusters				
		$H_{10}1$	$H_{10}2$	$H_{10}3$	$H_{10}4$	$H_{10}5$
Initial annotations	$F_{10}1$	5,385	3,228	16	89	0
	$F_{10}2$	120	17,552	145	0	0
	$F_{10}3$	132	4,331	1,644	219	0
	$F_{10}4$	575	1,038	157	1,542	0
	$F_{10}5$	417	5,687	2	1	0

		Final clusters				
		$H_{10}1$	$H_{10}2$	$H_{10}3$	$H_{10}4$	$H_{10}5$
Consistent clusters	$G_{10}1$	5,629	0	0	0	0
	$G_{10}2$	0	24,193	0	0	0
	$G_{10}3$	0	0	1,044	0	0
	$G_{10}4$	0	0	0	730	0
	$G_{10}5$	0	0	0	0	0
	N/A	1,000	7,643	920	1,121	0

Figure S41. Final clustering results obtained using CCAL for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The initial annotations ($k = 5$) derived from the K-shape multivariate clustering (Figure S15) are used as the reference. The transition matrix between the initial annotations, "consistent cluster" labels, and final cluster labels is included. Pixels not classified as the "consistent clusters" during the bagging-based learning phase are labeled as N/A ("not applicable").

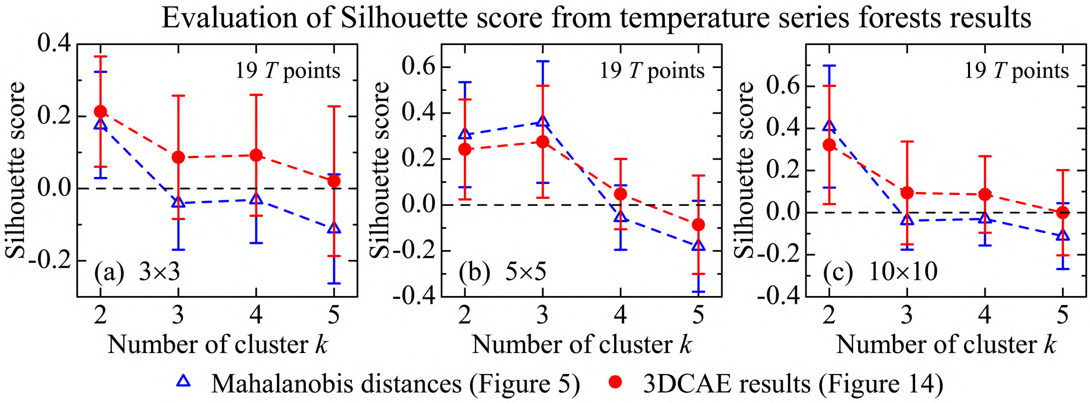
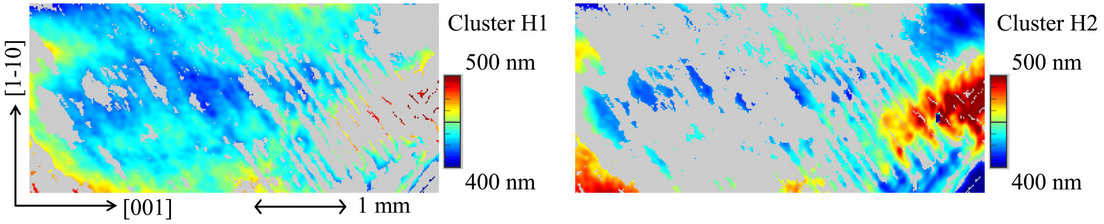
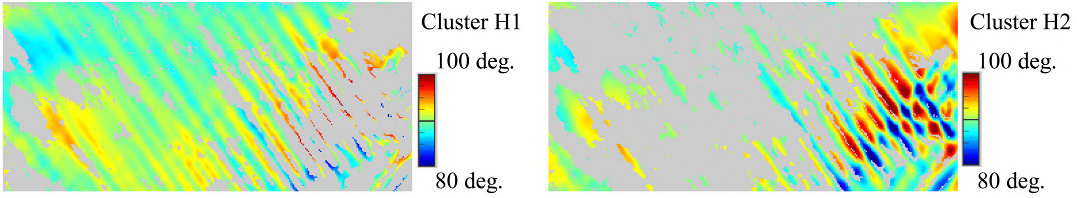


Figure S42. Silhouette scores for different initial numbers of clusters (k) computed using temperature series forests (Tsf) for the SRF sizes of (a) 3×3, (b) 5×5, and (c) 10×10. The sliding temperature window width is fixed at 19 points (9.5 K).

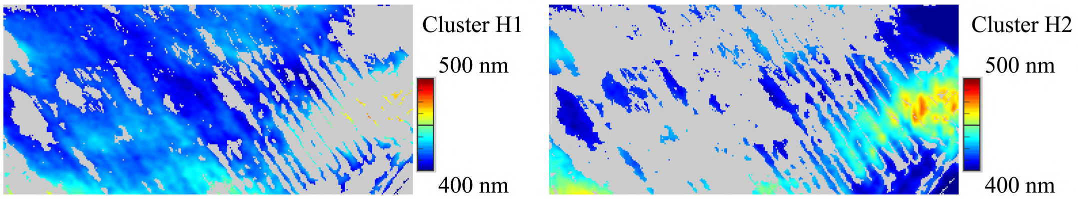
(a) Retardance δ at 14.1 K for $\lambda = 575$ nm from final clustering result with (3×3)



(b) Fast-axis direction ψ at 14.1 K for $\lambda = 575$ nm from final clustering result with (3×3)



(c) Retardance δ at 40.0 K for $\lambda = 575$ nm from final clustering result with (3×3)



(d) Fast-axis direction ψ at 40.0 K for $\lambda = 575$ nm from final clustering result with (3×3)

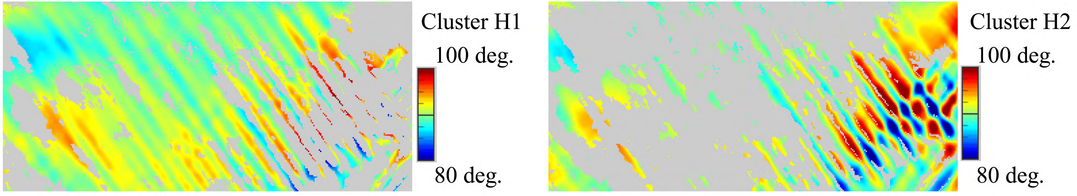
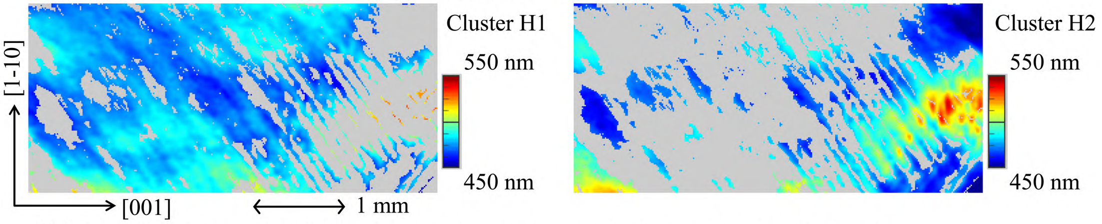
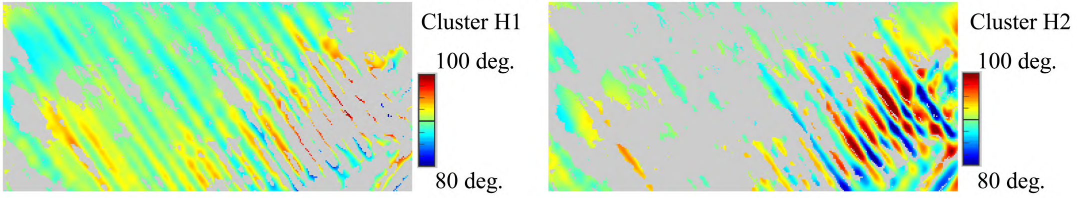


Figure S43. Based on the final clustering results of Tsf with the SRF size of 3×3 and an initial annotation of $k = 2$, the distribution of retardance (δ) and fast-axis direction (ψ) is categorized into two groups: (a) δ and (b) ψ at 14.1 K for 575 nm; (c) δ and (d) ψ at 40.0 K for 575 nm.

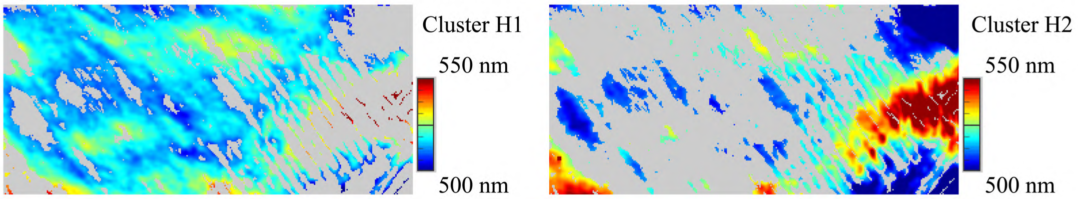
(a) Retardance δ at 90.0 K for λ = 575 nm from final clustering result with (3 × 3)



(b) Fast-axis direction ψ at 90.0 K for λ = 575 nm from final clustering result with (3 × 3)



(c) Retardance δ at 130.9 K for λ = 575 nm from final clustering result with (3 × 3)



(d) Fast-axis direction ψ at 130.9 K for λ = 575 nm from final clustering result with (3 × 3)

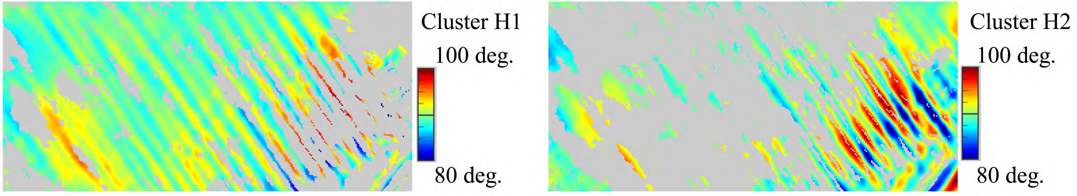
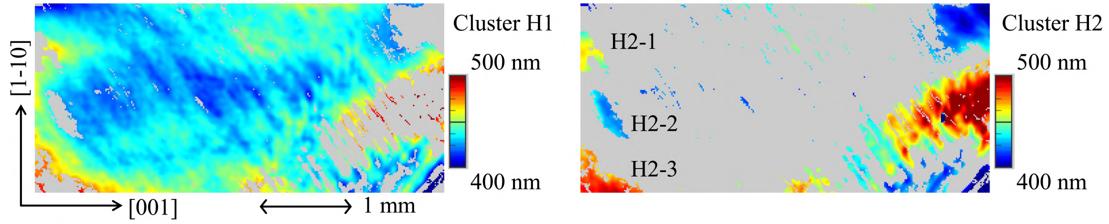
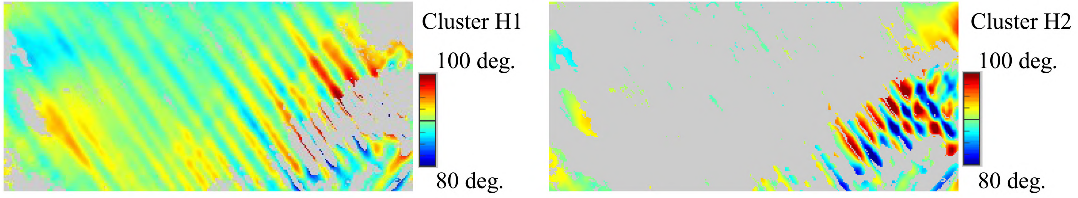


Figure S44. Based on the final clustering results of Tsf with the SRF size of 3×3 and an initial annotation of $k = 2$, the distribution of retardance (δ) and fast-axis direction (ψ) is categorized into two groups: (a) δ and (b) ψ at 90.0 K for 575 nm; (c) δ and (d) ψ at 130.9 K for 575 nm.

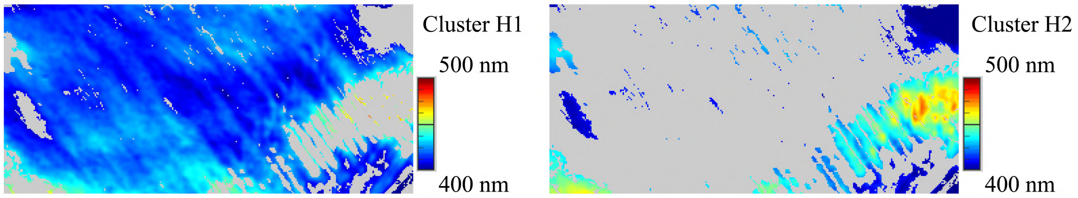
(a) Retardance δ at 14.1 K for λ = 575 nm from final clustering result with (5×5)



(b) Fast-axis direction ψ at 14.1 K for λ = 575 nm from final clustering result with (5×5)



(c) Retardance δ at 40.0 K for λ = 575 nm from final clustering result with (5×5)



(d) Fast-axis direction ψ at 40.0 K for λ = 575 nm from final clustering result with (5×5)

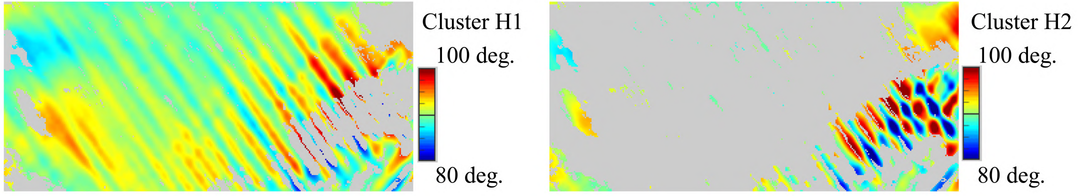
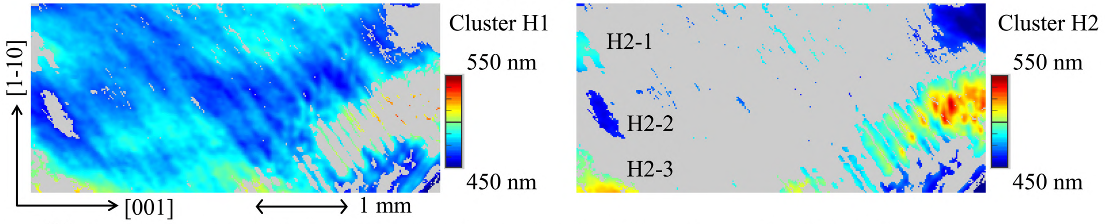
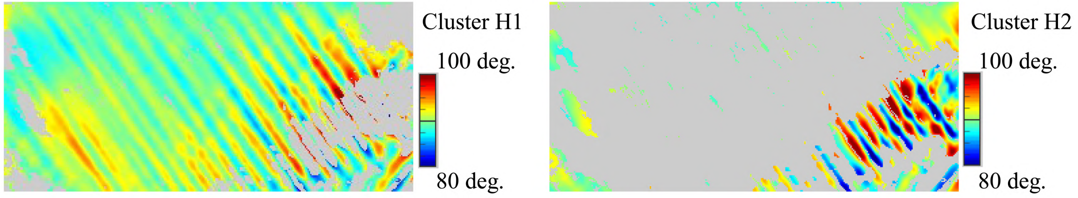


Figure S45. Based on the final clustering results of Tsf with the SRF size of 5×5 and an initial annotation of $k = 3$, the distribution of retardance (δ) and fast-axis direction (ψ) is categorized into two groups: (a) δ and (b) ψ at 14.1 K for 575 nm; (c) δ and (d) ψ at 40.0 K for 575 nm. The domains labeled H2-1, H2-2, and H2-3 show different temperature variation trends.

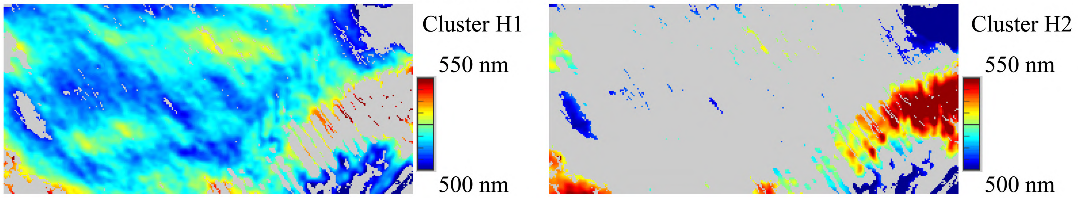
(a) Retardance δ at 90.0 K for λ = 575 nm from final clustering result with (5 × 5)



(b) Fast-axis direction ψ at 90.0 K for λ = 575 nm from final clustering result with (5 × 5)



(c) Retardance δ at 130.9 K for λ = 575 nm from final clustering result with (5 × 5)



(d) Fast-axis direction ψ at 130.9 K for λ = 575 nm from final clustering result with (5 × 5)

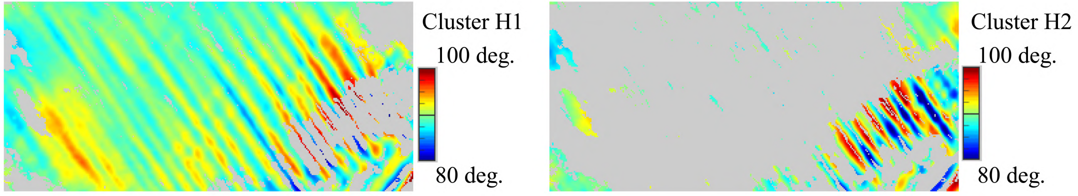
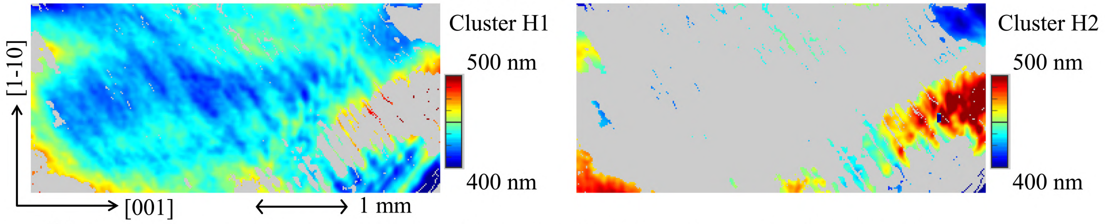
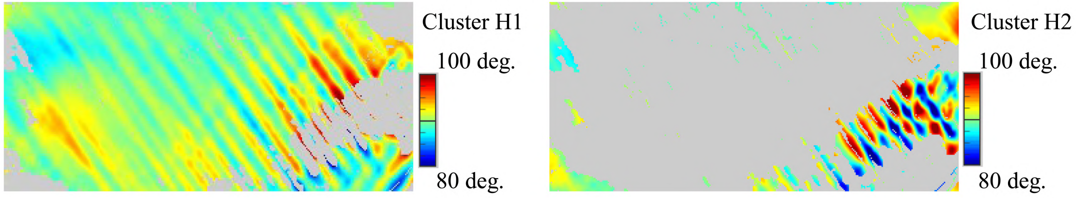


Figure S46. Based on the final clustering results of Tsf with the SRF size of 5×5 and an initial annotation of $k = 3$, the distribution of retardance (δ) and fast-axis direction (ψ) is categorized into two groups: (a) δ and (b) ψ at 90.0 K for 575 nm; (c) δ and (d) ψ at 130.9 K for 575 nm. The domains labeled H2-1, H2-2, and H2-3 show different temperature variation trends.

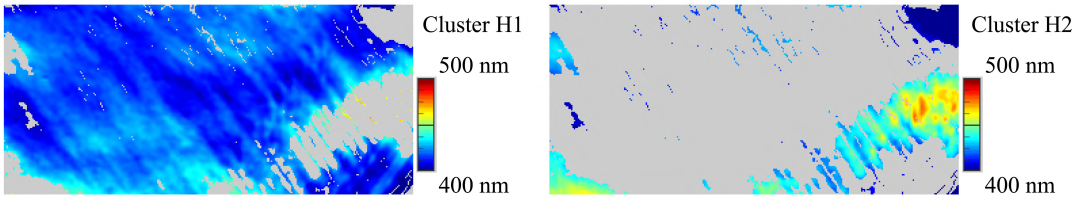
(a) Retardance δ at 14.1 K for λ = 575 nm from final clustering result with (10 × 10)



(b) Fast-axis direction ψ at 14.1 K for λ = 575 nm from final clustering result with (10 × 10)



(c) Retardance δ at 40.0 K for λ = 575 nm from final clustering result with (10 × 10)



(d) Fast-axis direction ψ at 40.0 K for λ = 575 nm from final clustering result with (10 × 10)

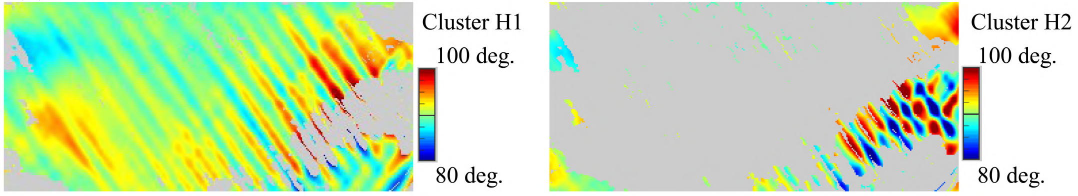
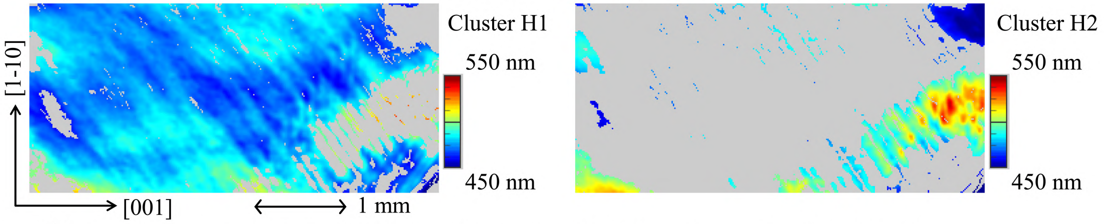
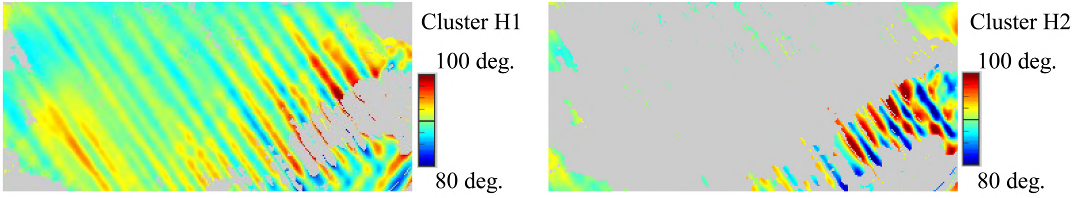


Figure S47. Based on the final clustering results of Tsf with the SRF size of 10×10 and an initial annotation of $k = 2$, the distribution of retardance (δ) and fast-axis direction (ψ) is categorized into two groups: (a) δ and (b) ψ at 14.1 K for 575 nm; (c) δ and (d) ψ at 40.0 K for 575 nm.

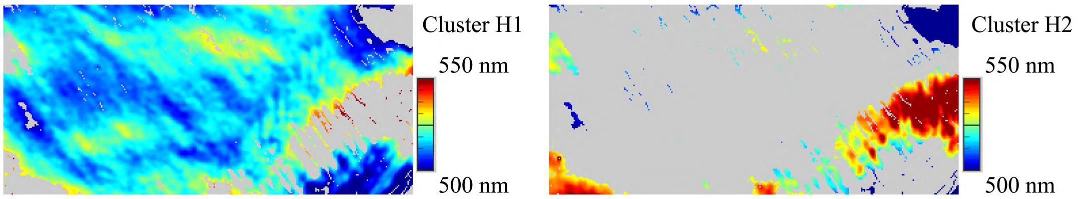
(a) Retardance δ at 90.0 K for λ = 575 nm from final clustering result with (10 × 10)



(b) Fast-axis direction ψ at 90.0 K for λ = 575 nm from final clustering result with (10 × 10)



(c) Retardance δ at 130.9 K for λ = 575 nm from final clustering result with (10 × 10)



(d) Fast-axis direction ψ at 130.9 K for λ = 575 nm from final clustering result with (10 × 10)

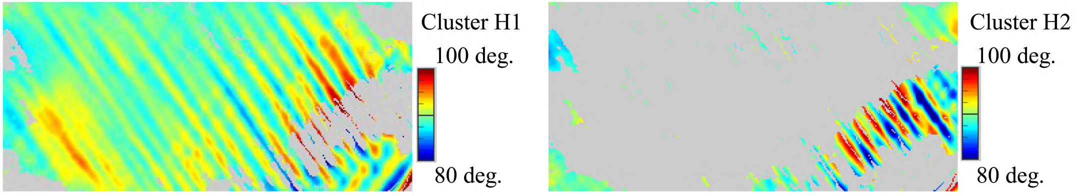


Figure S48. Based on the final clustering results of Tsf with the SRF size of 10×10 and an initial annotation of $k = 2$, the distribution of retardance (δ) and fast-axis direction (ψ) is categorized into two groups: (a) δ and (b) ψ at 90.0 K for 575 nm; (c) δ and (d) ψ at 130.9 K for 575 nm.