# Supporting Information for
# Uncovering Material Deformations via Machine Learning Combined with Four-Dimensional Scanning Transmission Electron Microscopy

Chuqiao Shi[1], Michael C. Cao[1,2], Sarah M. Rehn[3], Sang-Hoon Bae[4,5], Jeehwan Kim[6], Matthew R. Jones[1,3], David A. Muller[2,7], Yimo Han[1*]

[1] Department of Materials Science and NanoEngineering, Rice University, Houston, TX

[2] School of Applied and Engineering Physics, Cornell University, Ithaca, NY

[3] Department of Chemistry, Rice University, Houston, TX

[4] Department of Mechanical Engineering and Materials Science, Washington University in Saint Louis, Saint Louis, MO

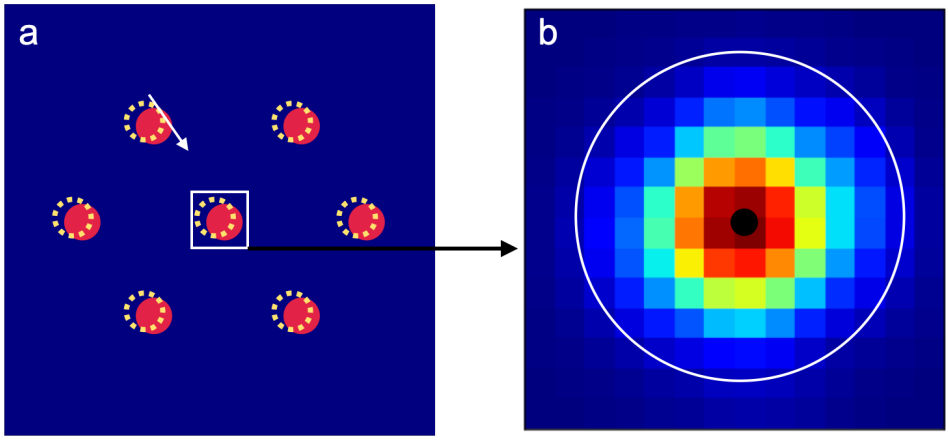
[5] Institute of Materials Science and Engineering, Washington University in Saint Louis, Saint Louis, MO

[6] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA

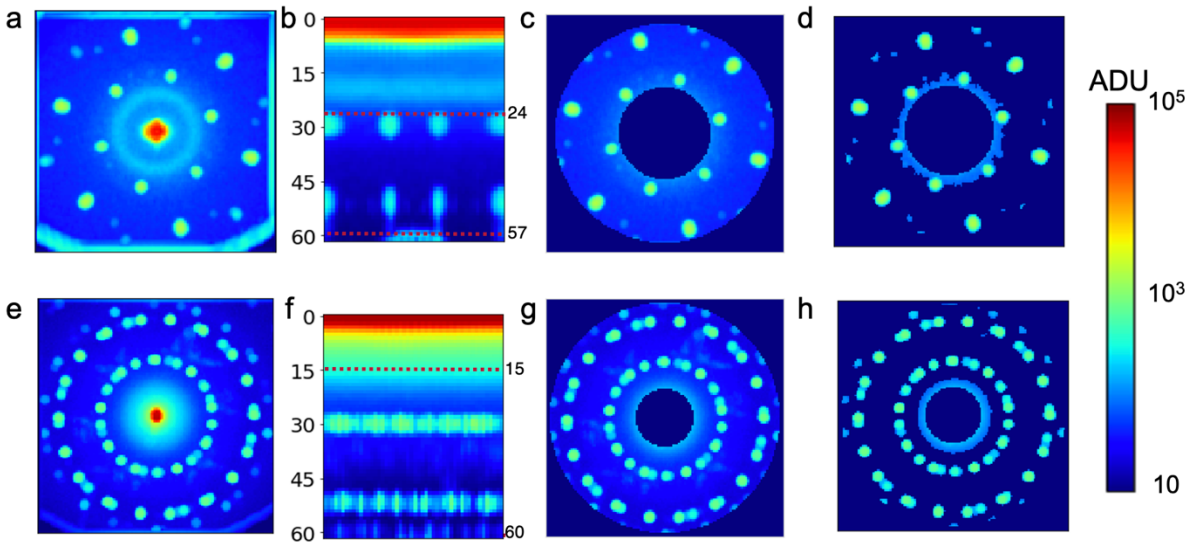
[7] Kavli Institute for Nanoscale Science, Cornell University, Ithaca, NY

[*] To whom correspondence should be addressed:

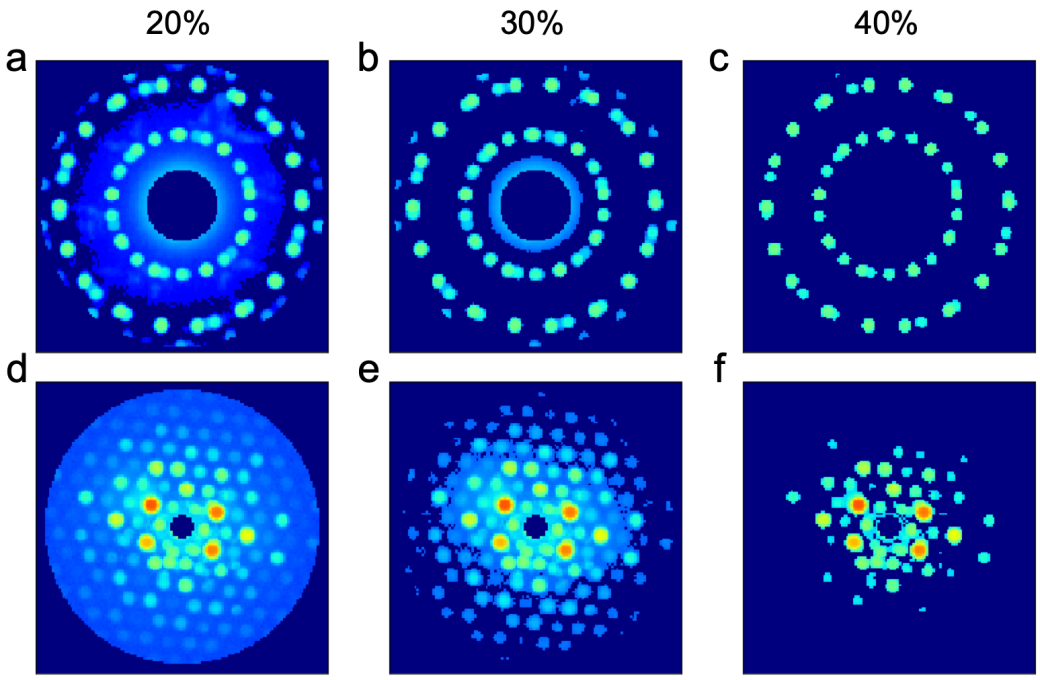
Yimo Han (Email: yimo.han@rice.edu)




**Supplementary Figure 1**. **Aligning diffraction patterns in preprocessing.** (a) Schematic of alignment in the preprocessing step. This step corrects the translational shift (indicated by the white arrow) of the diffraction patterns across the 4D mapping area caused by beam tilt in the microscope. Specifically, we correct this misalignment by aligning the center of mass (CoM) of the center beam (shown in the white box). We iterate this operation until the standard deviation for all CoM measures remains below 0.01 pixel in both $k_x$ and $k_y$ for all scanning positions in real space. (b) An example of the center beam in a diffraction pattern. We place a mask to select the center beam (white circle) and calculate the CoM (indicated by the black dot).




**Supplementary Figure 2. Placing masks to select the diffraction areas of interest**. (a) Standard deviation diffraction image by plotting the standard deviation of the pixel intensity across all scanning points in real space. This diffraction image was from the $WS_2$-$WSe_2$ junction 4D data shown in Fig. 2. (b) Diffraction image from (a) plotted in (θ, r) coordinate, where the amorphous ring structure becomes a line for easy determination of the inner radius of the ring mask to block it. The inner and outer radii for the mask are indicated by the red dashed lines, which are 24 and 57 in pixels respectively. (c) Diffraction image in (a) with the mask placed on it. The same mask will be applied on all actual diffraction patterns in the 4D data to select the diffraction area of interest. (d). Diffraction image in (a) with the standard deviation (STD) mask placed on it. The remaining pixels (higher than 30% of the highest STD) will be flattened as a vector. (e) Standard deviation diffraction pattern, plotted in the same way as (a), of the 4D data collected on $WS_2$-$WSe_2$ superlattices described in Fig. 3. (f) Diffraction image from (e) in (θ, r) coordinate system with the inner and outer radii determined as 15 and 60 pixels. The amorphous ring in this dataset is much weaker due to the thin layer of the supporting film in this sample. (g) Diffraction image in (e) with the mask. (h) Diffraction image in (g) with the STD mask placed on it.

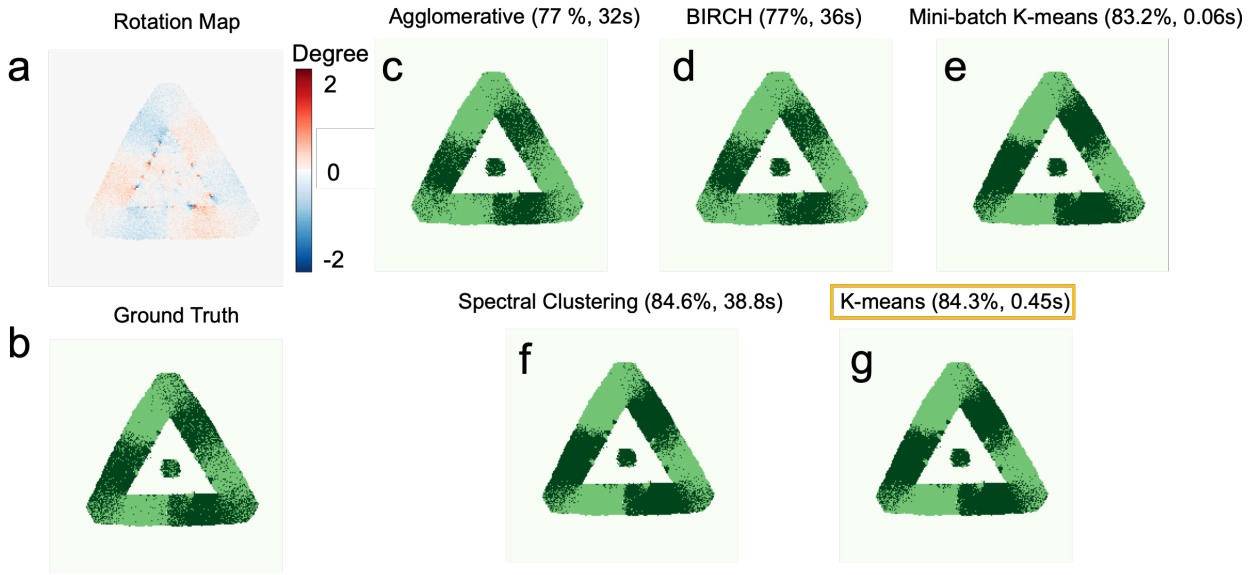



**Supplementary Figure 3. Thresholds for the STD mask.** The purpose of STD mask is to remain diffraction spots which contain lattice information and remove the background to save computation time. Here we tried 20% (a, d), 30% (b, e), 40% (c, f) of the highest STD (in log scale in narrow junction (a-c) and Ag nanoprism (d-f) datasets. 30% threshold led to best results for both thin and thick sample.

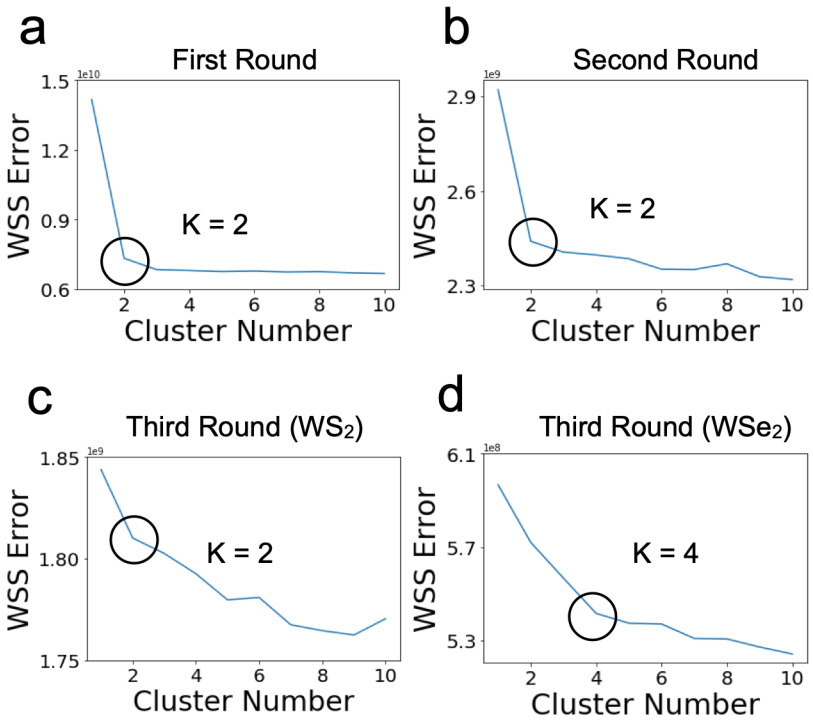
Accuracy and Time of Clustering and Dimension Reduction Techniques

| | **Std mask** | **PCA 3** | **UMAP 3** | **Raw** |
|---|---|---|---|---|
| **Agglomerative Clustering** | 77.4%, 31s | 51.4%, 7.8s | 77.6%, 7.6s | 77.5%, 140s |
| **BIRCH Clustering** | 77.6%, 35s | 52.5%, 1.5s | 75.6%, 1.6s | 77.3%, 127s |
| **Mini-Batch K-means** | 83.2%, 0.05s | 53.3%, 0.03s | 80%, 0.02s | 83.3%, 1.89s |
| **Spectral Clustering** | 84.6%, 36s | 52.1%, 31s | 81%, 32s | 84.7%, 48s |
| **K-means** | 84.3%, 0.4s | 53.5%, 0.09s | 80.4%, 0.11s | 83.9%, 32s |

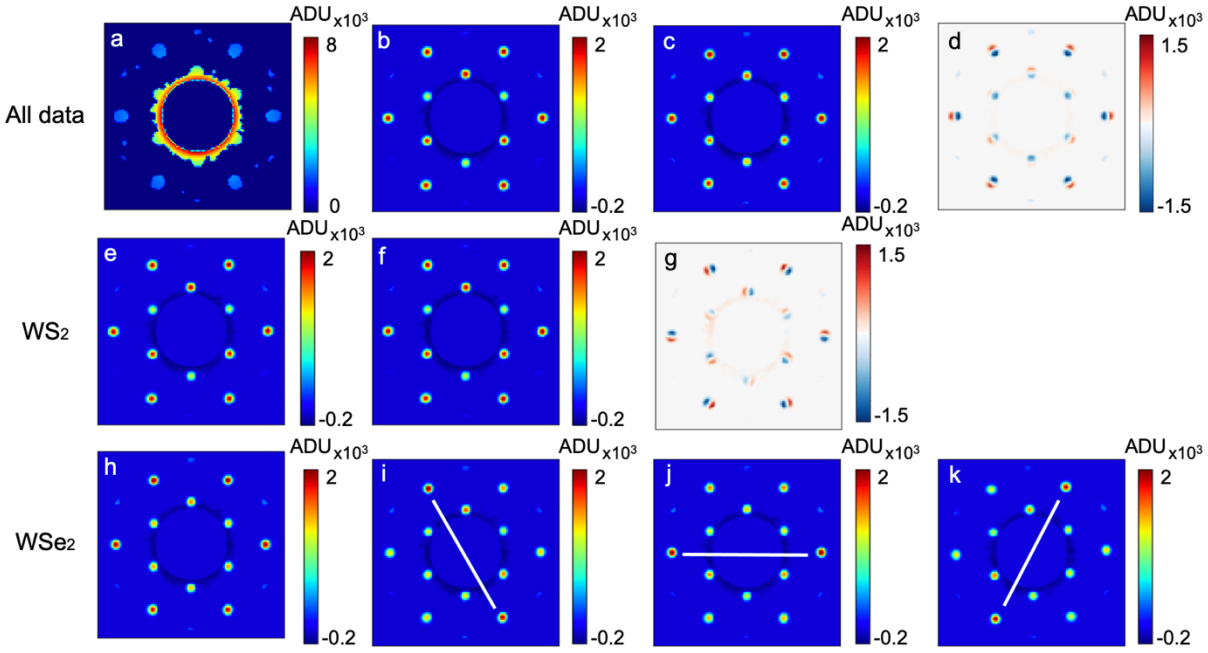
**Supplementary Table 1.** Comparison of dimension reduction methods & clustering methods. Standard deviation mask, PCA and UMAP are used to reduce the dimension and extract the features from the raw data before clustering, then five widely used clustering methods are applied on the extracted features. The PCA can only capture the dominant feature, but it cannot extract the minor features, such as the small rotation angles in $WS_2$ sample. Its accuracy is the lowest and the labels are shown in Supplementary Figure 16c. UMAP can capture the features, but the clustering labels (Supplementary Figure 16d) show fake stripes caused by alignment in the preprocessing. The standard deviation mask method provides the same accuracy compared with the raw data across each clustering method, but dramatically reduces the time.



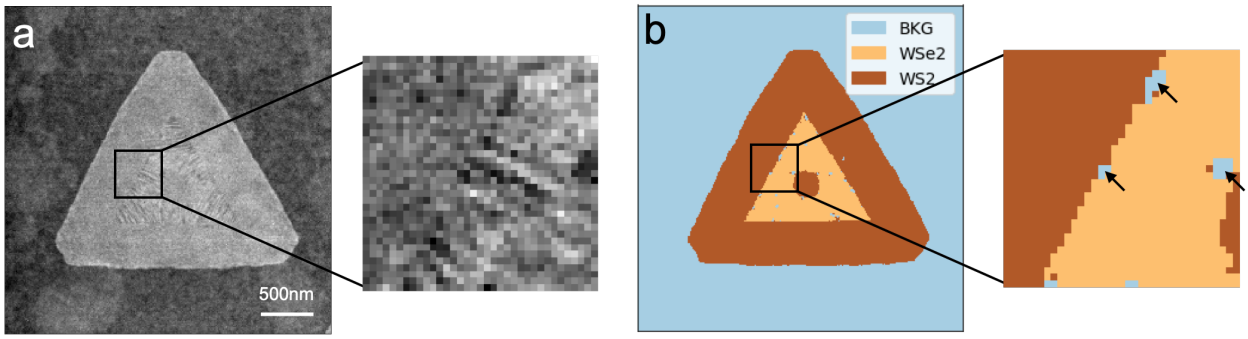

**Supplementary Figure 4. Comparison among five widely used clustering methods.** (a) Rotation map measured manually, which has been reported in previous reference [1] (b) Ground truth labels of two rotation angles in $WS_2$ sample. According to the rotation map, positive angles are set as label 1 and negative angles are set as label 2. (c – g) Clustering results of five widely used clustering methods. Clustering time and accuracy are labeled on each panel. Mini-batch K-means method is the fastest, so we use it in the elbow method to determine the cluster number. However, since the batch selection is random in this algorithm, the clustering results are not stable. The spectral clustering provides the best accuracy but takes a much longer time. Considering the tradeoff between efficiency and performance, K-means is chosen as the clustering method in each round of the hierarchical clustering architecture.



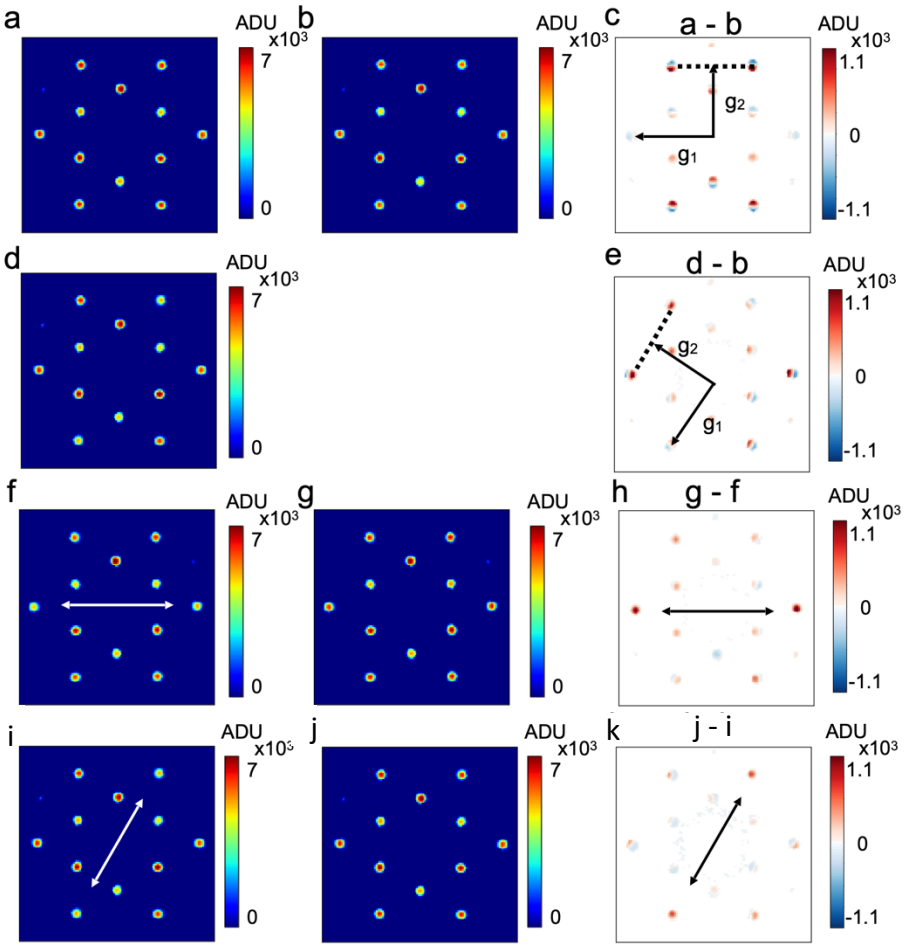

**Supplementary Figure 5. Elbow method for determining the clustering number for each round.** Elbow method runs the clustering algorithms several times with different K numbers and calculates the Within-Cluster Sum of Squares (WSS) error each time. The WSS error trends downward with increasing cluster number. Afterwards, we determine the K number through the elbow point on the WSS curves. The elbow point is the inflection point where the second derivative of the WSS curve becomes negative, which indicates that adding one more cluster from K to K+1 will not reduce the error as much as adding one cluster from K-1 to K. The elbow method was highly efficient at selecting the optimum number of clusters for the various samples we analyzed and automated our methods for the 4D data processing. (a) The result from the first-round clustering of the junction sample in Fig. 2, where K=2 was selected as the elbow point. (b) The second-round clustering of the junction area, resulting in a clustering number of 2. (c) The third-round sub-clustering of $WS_2$ gives an optimal cluster number of 2. (d) The third-round sub-clustering of $WSe_2$ displays a K=4 elbow point.



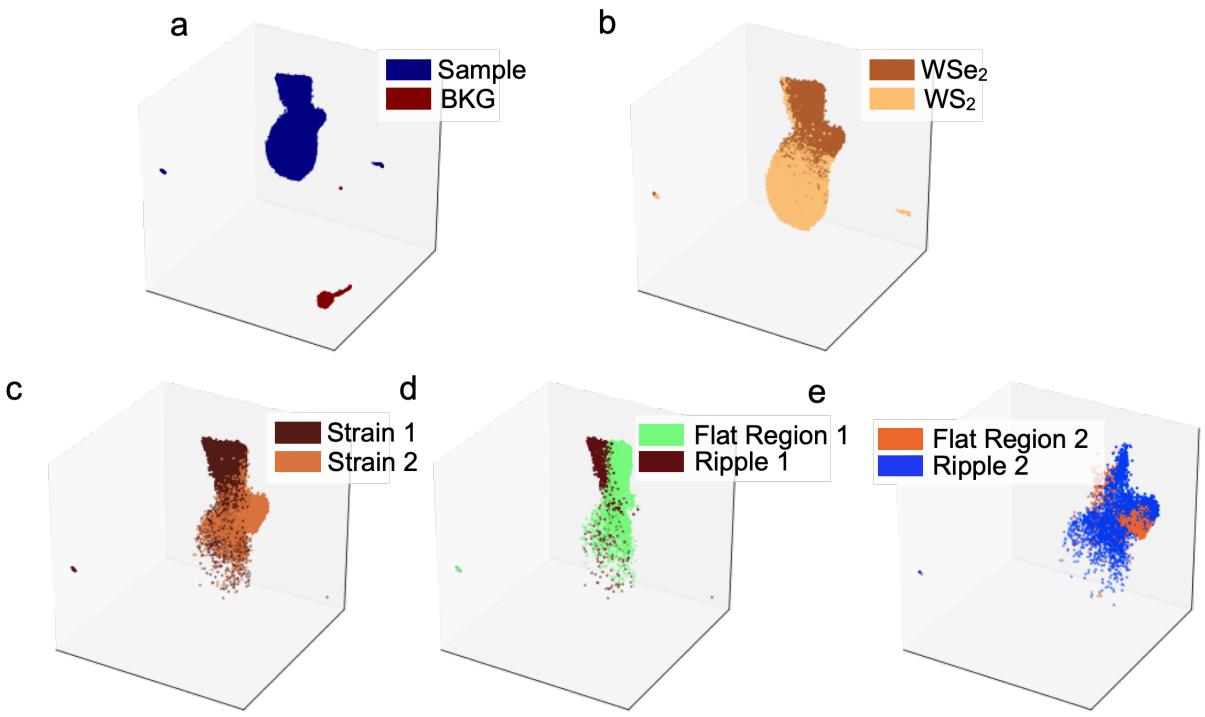

**Supplementary Figure 6. Mean diffraction patterns of each cluster for the junction sample in Fig. 2.** (a) Mean diffraction patterns of the amorphous substrate. (b-c, e-f, h-k) are cluster centers subtracted by the background to show the diffraction spots with less noise. (b-c) Mean diffraction patterns of $WS_2$ and $WSe_2$; (d) Lattice constant difference between (b) and (c). (e-f) Mean diffraction patterns of sub-clusters in the $WS_2$ sample. (g) Rotational difference between (e) and (f). (h-k) Mean diffraction patterns of sub-clusters in the $WSe_2$ sample. (h) shows the cluster mean of the flat area in $WSe_2$, while (i), (j) and (k), present the cluster means of three directional ripple areas.



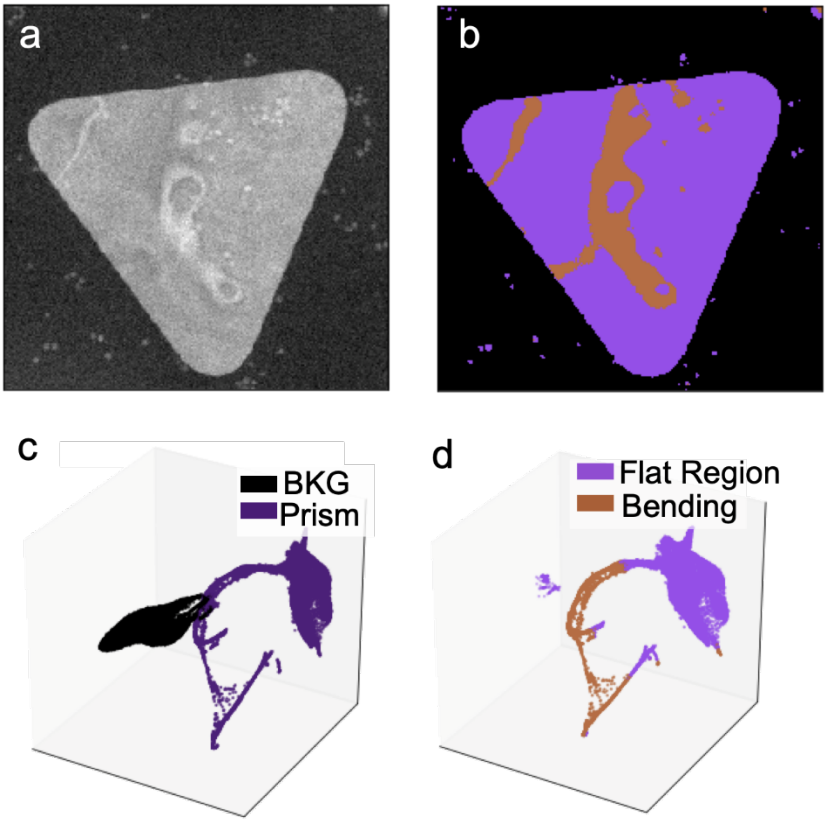

**Supplementary Figure 7. Holes identified in the $WS_2$-$WSe_2$ junction sample by unsupervised learning.** (a) ADF image of the $WS_2$-$WSe_2$ junction with a zoomed-in area displayed. (b) Real-space map of the clustering results achieved using our method, where a few holes are identified in the $WSe_2$ (indicated with arrows). Compared to the ADF image in (a), these holes are probably from the damage of local defects or dislocations in the material during the growth or annealing process.



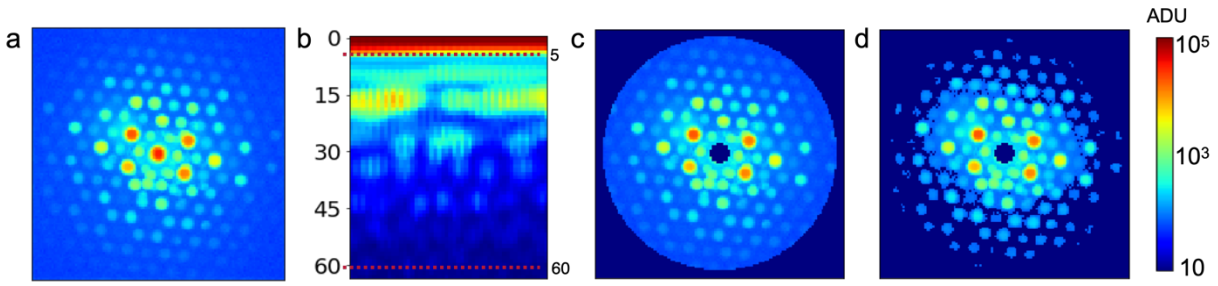

**Supplementary Figure 8. Diffraction pattern difference between different clusters in $WS_2$-$WSe_2$ superlattices.** (a, d) Mean diffraction patterns of strain 1 and strain 2 cluster in Fig. 3d. (c) Mean diffraction pattern of $WS_2$ region in Fig. 3c. (c,e) Diffraction pattern difference between each sub-cluster of $WSe_2$ in Fig. 3d and the $WS_2$ area. The strains are mostly along $\mathbf{g_2}$ (1.2%), where the $\mathbf{g_1}$ and $\mathbf{g_2}$ are labeled in (c) and (e). The results are consistent with the strain map from the previous result[1]. (f, g, i, j) Mean diffraction patterns of Ripple 1 (f), Flat Region 1 (g), Ripple 2 (i) and Flat Region 2 (j) in Fig. 3e. (h,k) Diffraction patterns difference between the sub-clusters in Fig. 3e, where (h) represents Ripple 1 minus Flat 1, while (k) represents Ripple 2 minus Flat 2. The intensity difference of diffraction spots between sub-clusters indicates a ripple (or tilt) in the lattice, which is consistent with our observation in the junction sample in Fig. 2 and our previous work on the superlattice [2]. The tilt axes are labeled by the white lines.



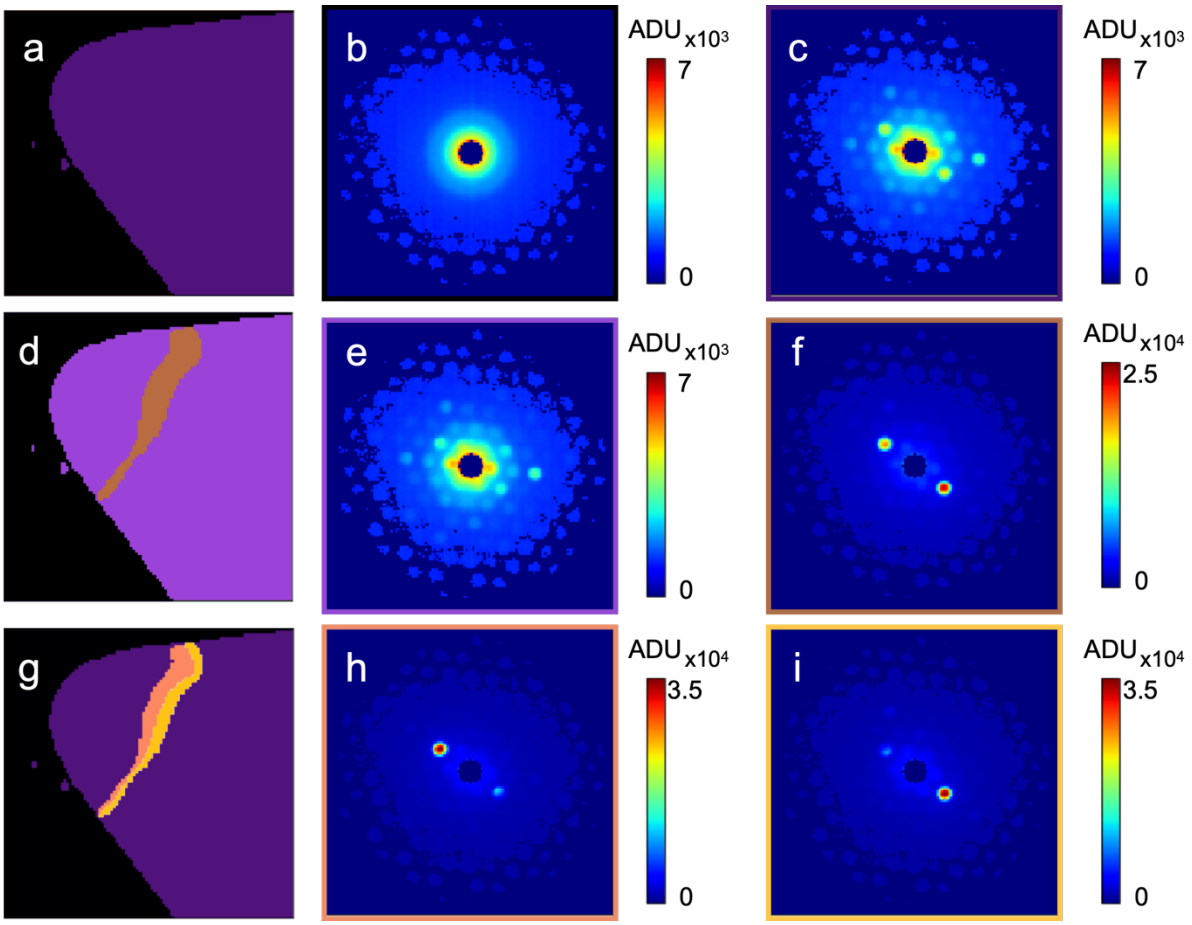

**Supplementary Figure 9. Manifold structures of $WS_2$-$WSe_2$ superlattices.** (a) 3D manifold structure of the data in the first-round clustering. (b) Manifold structure of the sample subcluster (including $WS_2$ and $WSe_2$) in the second-round clustering. (c) Manifold structure of $WSe_2$ displaying uniaxial strain in the third round. (d,e) Manifold structures of the final-round data uncovering minor ripples in the sample.



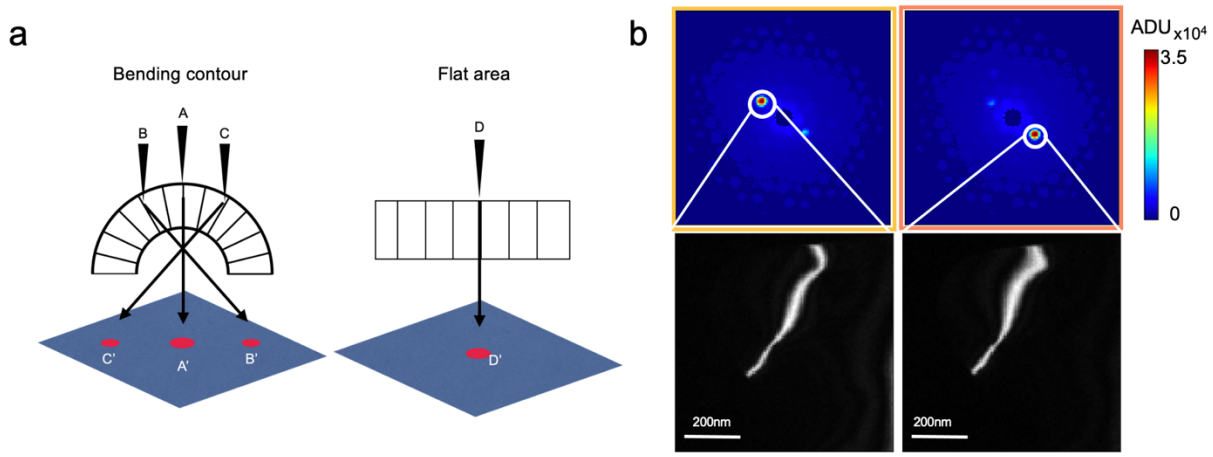

**Supplementary Figure 10. Clustering results of an entire silver nanoprism.** (a) ADF-STEM image of the entire silver nanoprism flake. (b) Real-space map of the clustering results, showing bending contours in the nanoprism (purple). (c-d) 3D manifold structure of the data in different rounds.



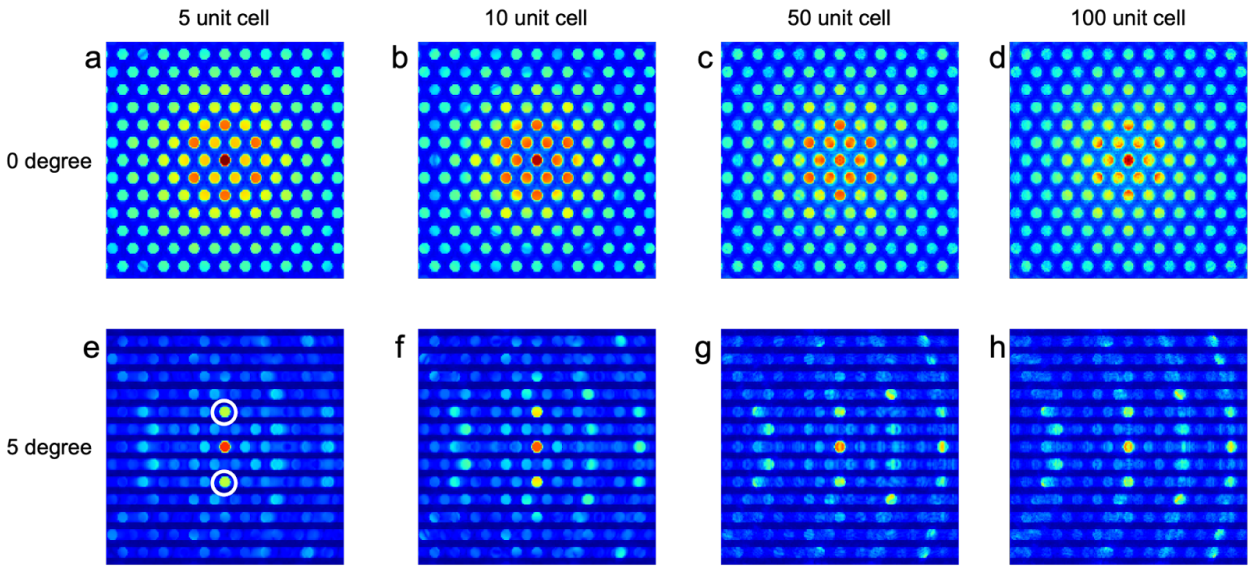

**Supplementary Figure 11. Preprocessing of Ag nanoprism dataset**. (a) Standard deviation diffraction image by plotting the standard deviation of the pixel intensity across all scanning points in real space. This diffraction image was from the Ag nanoprism junction 4D data shown in Fig. 4. (b) Diffraction image from (a) plotted in (θ, r) coordinate, where the amorphous ring structure becomes a line for easy determination of the inner radius of the ring mask to block it. The inner and outer radii for the mask are indicated by the red dashed lines, which are 5 and 60 in pixels, respectively. (c) Diffraction image in (a) with the mask placed on it. The same mask will be applied on all actual diffraction patterns in the 4D data to select the diffraction area of interest. (d). Diffraction image in (a) with the standard deviation mask placed on it. The remaining pixels (higher than 30% of the highest STD) will be flattened as a vector.



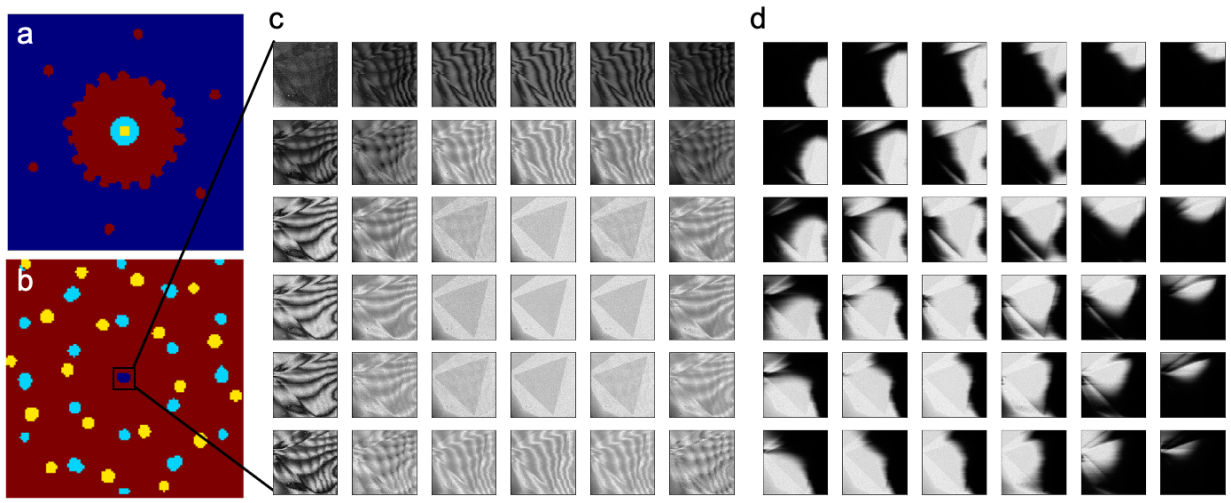

**Supplementary Figure 12. Clustering results of silver nanoprism.** (a) Real-space map of the first-round clustering result, where the Ag nanoprism is distinguished from the background. (b-c) Mean diffraction pattern of background (b) and Ag nanoprism (c). (d) Real-space map of the second-round clustering result, where the bending contour (brown) and the flat area (purple) in the Ag nanoprism are separated. (e-f) Mean diffraction pattern of mean flat region (e) and bending contour region (f). (g) Real-space map of the third-round clustering result, where the bending contour are separated into two sides. (h-i) Mean diffraction pattern of two sides of the bending contour.



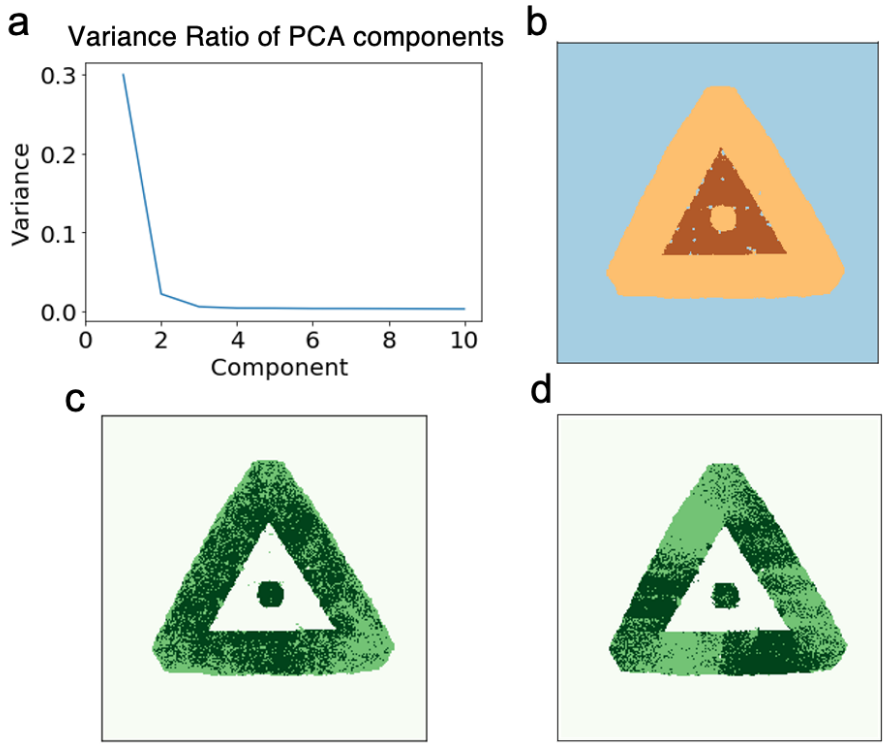

**Supplementary Figure 13. Clustering results of silver nanoprisms.** (a) Schematic of the electron diffraction at a bending contour and a flat area. When the electron probe scans the two sides of the bending contour (B and C), the angles fulfill different Bragg conditions, leading to high intensity in different diffraction spots (B' and C' respectively). When the probe scans a flat area in the Ag nanoprism (A or D), most of the intensity lies in the unscattered bright disk (A' and D') as the atomic plane is parallel to the incident beam. (b) Virtual dark-field images of both sides of the bending contour from the 4D dataset by placing virtual apertures on the corresponding diffraction spots.



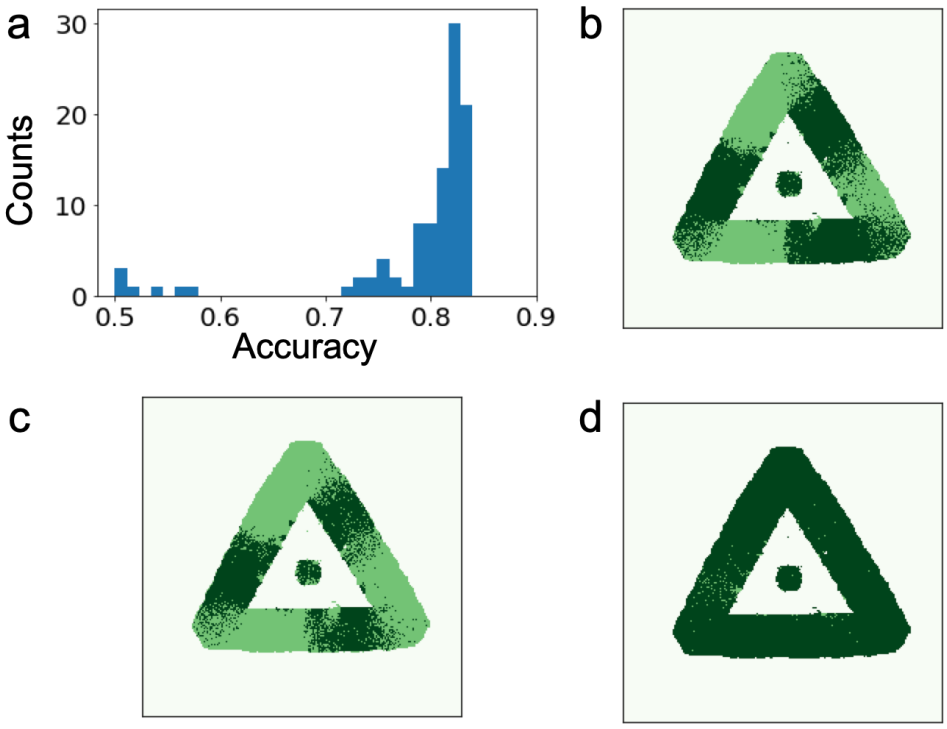

**Supplementary Figure 14. Simulated diffraction patterns with different thickness and tilt.** (a-d) Diffraction patterns of 5, 10, 50, and 100-unit-cell thickness without tilt. (e-h) Diffraction patterns of 5, 10, 50, 100-unit-cell thickness with a 5-degree tilt. Thicker samples reduce the intensity of diffraction spots, and the tilt causes the symmetry broken of the diffraction pattern.




**Supplementary Figure 15. Clustering real-space images in 4D datasets.** (a) Map of the clustering results in diffraction space without normalization image intensity in the pre-processing. The data are clustered according to the intensity rather than actual features in the sample. (b) Map of the clustering results in diffraction space with real-space normalization, which segments the diffraction space into different areas according to the virtual imaging mode (BF or DF) or the crystallinity in the sample. (c) Pixel-by-pixel real space images from the center beam, where we observed stripes in the BF images at the edge of the center diffraction disk. (d) Pixel-by-pixel real space images from the center beam of the raw data without any pre-processing, indicating the stripes we observed are artifacts caused by the diffraction pattern alignment.




**Supplementary Figure 16. PCA and UMAP Failure Cases** (a) Variance Ratio of PCA components from the whole $WS_2$-$WSe_2$ heterojunction dataset shown in Fig. 2. Based on the curve we choose the first three high variance components to do the clustering. (b) First and second round clustering results based on PCA able to separate background, $WS_2$, and $WSe_2$. (c) Third round cluster labels from PCA. The components from PCA only capture the main feature but it cannot extract the minor features, such as the small rotation angles in the $WS_2$ sample. (d) Third round cluster labels from UMAP can partially capture the features but introduces striping artifacts.




**Supplementary Figure 17. Precision (or consistency) of K-means clustering.** (a) Histogram of the accuracy from 100 clustering results. (b-d) Clustering results with (b) 84%, (c) 75%, and (d) 50% accuracy.

1. Han, Y. *et al.* Strain Mapping of Two-Dimensional Heterostructures with Subpicometer Precision. *Nano Lett* 18, 3746–3751 (2018).

2. Xie, S. *et al.* Coherent, atomically thin transition-metal dichalcogenide superlattices with engineered strain. *Science* 359, 1131–1136 (2018).