

1

Electronic Supplementary Information

2

Full automation of point defect detection in transition metal dichalcogenides through dual mode deep learning algorithm

3

4 Dong-Hwan Yang,^{‡ab} Yu-Seong Chu,^{‡c} Odongo Francis Ngome Okello,^a Seung-Young Seo,^a Gunho Moon,^{ab}
5 Kwang Ho Kim,^d Moon-Ho Jo,^{ab} Dongwon Shin,^c Sejung Yang^{*f} and Si-Young Choi^{*abg}

6

7 ^a. Department of Materials Science and Engineering, Pohang University of Science and Technology (POSTECH), 77
8 Cheongam-Ro, Pohang 37673, Republic of Korea.

9 ^b. Center for Van der Waals Quantum Solids, Institute of Basic Science (IBS), 77 Cheongam-Ro, Pohang 37673, Republic of
10 Korea.

11 ^c. Division of Biomedical Engineering, College of Health Sciences, Yonsei University, 1, Yeonsedae-gil, Heungeop-myeon,
12 Wonju-si, Gangwon-do, 26493, Republic of Korea.

13 ^d. Department of Materials Science and Engineering, Pusan National University (PNU), 2, Busandaehak-ro 63beon-gil,
14 46241, Geumjeong-gu, Busan, Republic of Korea.

15 ^e. Materials Science and Technology Division, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN 37831, USA.

16 ^f. Department of Precision Medicine, Yonsei University Wonju College of Medicine, 20, Ilsan-ro, Wonju-si, Gangwon-do,
17 Republic of Korea.

18 ^g. Department of Semiconductor Engineering, Pohang University of Science and Technology (POSTECH), 77 Cheongam-Ro,
19 Pohang 37673, Republic of Korea.

20

21

* Correspondence: syang@yonsei.ac.kr, youngchoi@postech.ac.kr

22

‡ These authors contributed equally to this work.

23

24 This *Electronic Supplementary Information* includes:

25

26 Figures S1-S15.

27

28 Supplementary Texts 1-6.

29

Supplementary Text 1. Intensity profiling-based point defect identification.

30

Supplementary Text 2. Simulation dataset modulation to reflect experimental imaging conditions.

31

Supplementary Text 3. Point defect classification algorithm in 2DIP-Net.

32

Supplementary Text 4. Point defect classification performance for simulated image inspection.

33

Supplementary Text 5. Point defect classification performance for experimental image inspection.

34

Supplementary Text 6. The architecture of FCN model and its point defect classification performances.

35

36

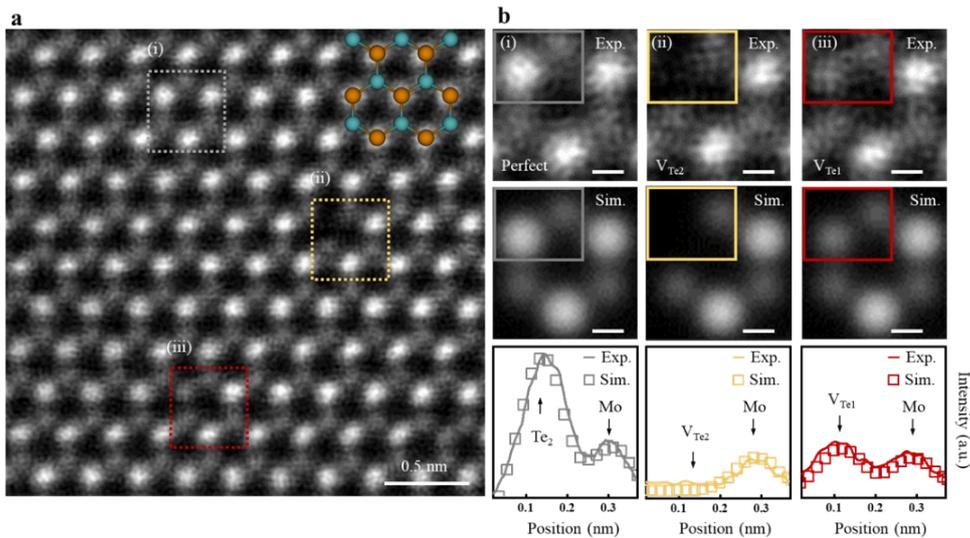
Supplementary

references

1 **Supplementary Text 1. Intensity profiling-based point defect identification.**

2 Fig. S1a presents an experimental HAADF-STEM image of the ML 2H-MoTe₂. The process scans each hexagonal cell in an
 3 image and finds the defective ones, as denoted by the orange and red dotted rectangles in Fig. S1a. The experimental and
 4 simulation HAADF-STEM image of (i) Perfect, (ii) V_{Te2} and (iii) V_{Te1} in top and middle panels in Fig. S1b match each other.
 5 This consistency is confirmed by measuring the intensity profiles as provided in bottom panels in Fig. S1b. This manual point
 6 defect identification is essential, and the comparison of experimental and simulated image secures the justification to use the
 7 simulated image for CNN-training. Furthermore, it is crucial to manually verify deep learning-based point defect analysis by
 8 intensity profiling. Once the point defect classification competence of deep learning model is ensured, the deep learning model
 9 is more credible to analyse various types of point defects expanding its applicability to other 2D materials. To improve upon
 10 the analytic approach and avoid deterioration of data reliability, we designed a CNN-based analytic platform by utilising the
 11 various simulated images in Fig. 1 as the simulation dataset.

Fig. S1 Conventional point defect identification method in monolayer (ML) 2H-MoTe₂. (a) Experimental HAADF-STEM image of ML 2H-MoTe₂ embedding a few point defects. The representative (i) grey, (ii) orange, and (iii) red dotted rectangles illustrates (i) Perfect, (ii) V_{Te2}, and (iii) V_{Te1}, respectively. (b) (Top)(Middle) Magnified experimental (simulation) HAADF-STEM images corresponding to (i)-(iii) in (a). (Bottom) Experimental and simulation 2H-MoTe₂ unit cells along with corresponding intensity profiles. The solid lines (experiment) and open squares (simulation) are well matched for (i) Perfect (grey), (ii) V_{Te2} (orange), and (iii) V_{Te1} (red) cells, respectively. The colour-codes are the same as Fig. 1. Scale bars, 0.2 nm.



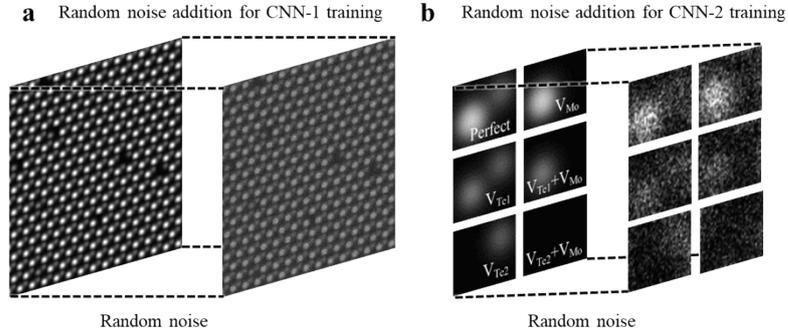
12

13

14 **Supplementary Text 2. Simulation dataset modulation to reflect experimental imaging conditions.**

15 Since the dominant noise type during STEM imaging is Poisson noise^{1,2}, we applied (added) Poisson (Gaussian) noise to each
 16 simulation dataset for CNN-1 and CNN-2 training and test images (Fig. S2). We modulated noise with Poisson (Gaussian)
 17 noise parameter of $5 \leq \lambda \leq 20$ ($\mu = 20$) to each simulated images. For the CNN-1, the number of total training (validation)

1 data was 500 (50) of simulation images. The total number of test set was 10 simulation images, corresponding to 1691
 2 cell images (Fig. 2). For the CNN-2, the number of total training (validation) data was 4800 (1200) of simulation images. The
 3 total number of test set was 10 (11) simulation (experimental) images, corresponding to 1691 (3037) unit cell images to classify
 4 point defects (Fig. 3)
 5



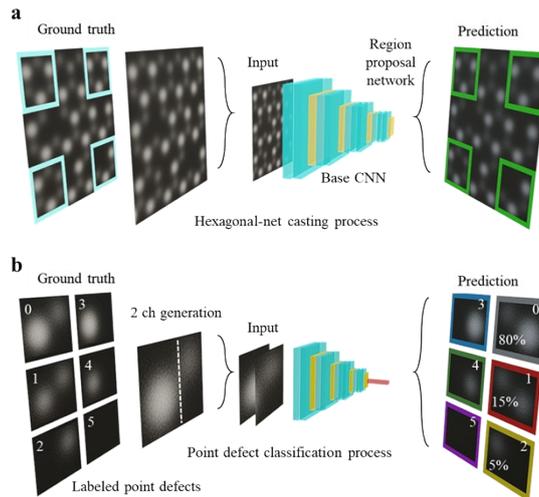
6

Fig. S2 (a)-(b) Simulation dataset construction with random Poisson/Gaussian noise variations CNN-1 and CNN-2, respectively.

7 **Supplementary Text 3. Point defect classification algorithm in 2DIP-Net.**

8 Fig. S3a illustrates the training process for CNN-1: the CNN-1 predicted location coordinates of hexagonal cells (sky-blue
 9 rectangles) and is iteratively trained to predict more accurate coordinates of each hexagonal cell (green rectangles) through
 10 comparison with the ground truth. Again, to extract “unit cells” from the “hexagonal cells” predicted by the trained CNN-1, a
 11 post-process is used with a crop-ratio (Fig. 2) in the hexagonal cell. This relieved the difficulties of cropping accurate unit
 12 cells in experimental images with various magnifications and/or atomic distortions.
 13

14 Fig. S3b represents the training process of CNN-2, where the simulation dataset of unit cells are labelled from 0 to 5 for
 15 each defect type in the left panels: 0, Perfect; 1, V_{Te1} ; 2, V_{Te2} ; 3, V_{Mo} ; 4, $V_{Te1}+V_{Mo}$; 5, $V_{Te2}+V_{Mo}$, respectively. Then, the
 16 labelled unit cells are randomly shuffled to be inputs for CNN-2. The CNN-2 predicted the probability of each defect type
 17 from these input unit cells. Iterative training was conducted for CNN-2 to predict the probability of each labelled defect type
 18 to be 100% and classified point defect types were distinguished by different colour-codes. Eventually, the trained CNN-2
 19 predicted the defect types from each unit cell extracted by CNN-1 and post-process, and colour-coded rectangles were marked
 20 on each unit cell location corresponding to the predicted defect types. The parameters used for the construction and learning
 21 of CNN-1 and CNN-2 are described in the Methods.
 22

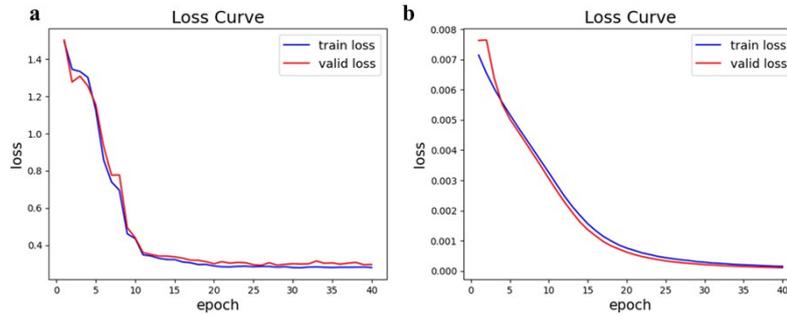


23

Fig. S3 (a) Training process for CNN-1, called hexagonal-net casting process. Training input is generated with some point defects with a large portion of perfect unit cells. The CNN-1 determines the region of hexagonal cells from input as illustrated by the green rectangles in “Prediction”. Iterative training is conducted for the CNN-1 to propose accurate hexagonal cell locations compared to the sky-blue “Ground truth”. (b) Training process for CNN-2 to classify point defects. Each training input is labelled from 0 to 5 and divided by 2 ch-input to be fed to CNN-2. The CNN-2 is iteratively trained to classify the correct type of point defect from input image features. The colour-boxes are the predicted point defect types by CNN-2, and all the colour-codes are the same as in Fig. 1.

24

1 Fig. S4 illustrate the conventional training results of loss curves; for both CNN-1 and CNN-2, training epoch was set to be
 2 40 with Adam optimizer.
 3

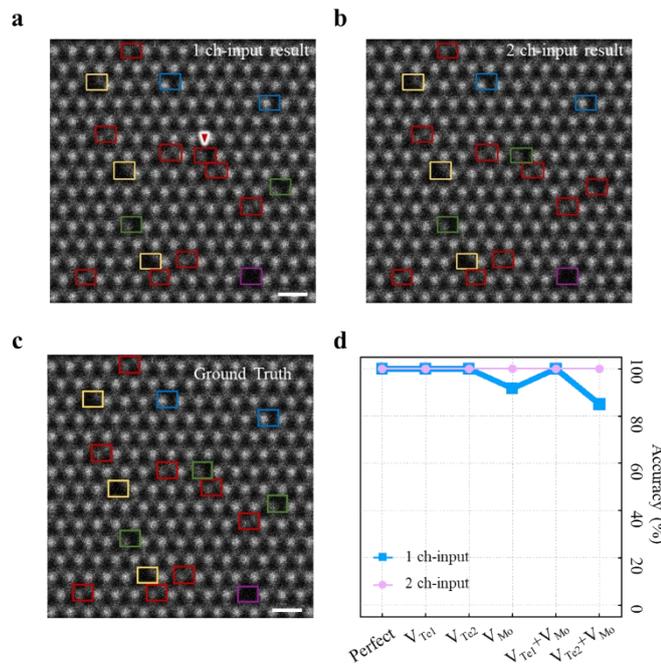


4
 5 **Fig. S4** (a)-(b) Training process of CNN-1 and CNN-2, respectively.

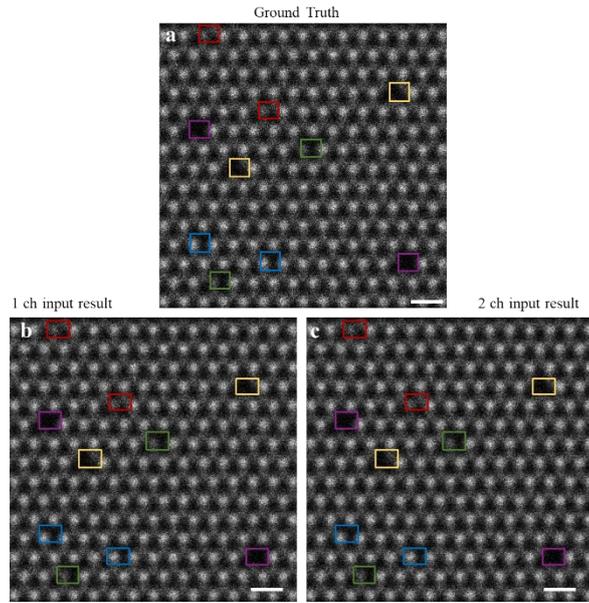
6
 7
 8
 9 **Supplementary Text 4. Point defect classification performance for simulated image inspection.**

10 To ensure the point defect classification competence of 2DIP-Net, we tested simulation images embedded with 5 types of point
 11 defects. As illustrated in Supplementary Fig. S5, five types of point defects were randomly embedded, and the total number of
 12 point defects was 17. The point defect classification outcomes of the 1 ch- and 2 ch-input types are comparable (Figs. S5a-b),
 13 compared to the ground truth (Fig. S5c). In particular, 16 of the embedded 17 point defects were identified correctly though 1
 14 $V_{Te1}+V_{Mo}$ in ground truth (olive) was misclassified to V_{Te1} (red), as indicated by red arrow in Fig. S5a (1 ch-input).

15
 16 We examined a total of 32 V_{Te1} , 66 V_{Te2} , 24 V_{Mo} , 47 ($V_{Te1}+V_{Mo}$), and 20 ($V_{Te2}+V_{Mo}$) in 10 simulated images of ground
 17 truth. We further examined various simulated image with different concentrations of point defects as presented in Figs. S6-S9.
 18 Regardless of the type and spatial distribution of point defects, our designed deep learning models exhibited high point defect
 19 classification accuracies; Perfect, V_{Te1} , V_{Te2} , V_{Mo} , $V_{Te1}+V_{Mo}$, $V_{Te2}+V_{Mo}$ for 1 ch-(2 ch-) input for 100% (100%), 100% (100%),
 20 91.67% (100%), 100% (100%), 92.3% (100%), 85% (100%), respectively. Fig. S9 shows the confusion matrices for overall the
 21 point defect classification performance, with elements denoting the counts of each estimated point defects (total 1691
 22 classifications including Perfect unit cells).
 23



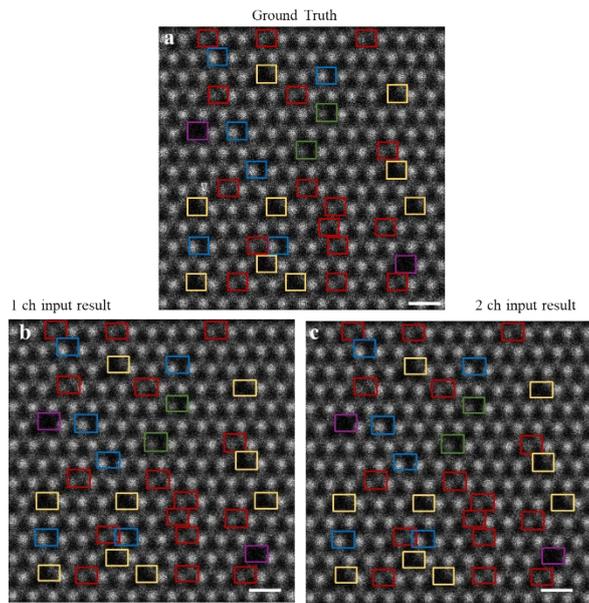
24 **Fig. S5** Point defect classification performance evaluation by CNN-2. (a)-(b) Point defect classification results of simulated image of
 1 ch- and 2 ch-inputs. The red arrow in (b) is misclassified to V_{Te1} ; ground truth of $V_{Te1}+V_{Mo}$. The colour-codes are the same as Fig.
 1. Scale bars, 0.5 nm. (c) Corresponding ground truth of (a)-(b). (d) Accuracies for point defect classification performance of total 10
 input simulated images (2100 point defects classified) of 1 ch-input (light blue) and 2 ch-input (light pink), respectively. The
 accuracies for perfect, V_{Te1} , V_{Te2} , V_{Mo} , $V_{Te1}+V_{Mo}$, $V_{Te2}+V_{Mo}$ for 1 ch-(2 ch-)input are 100% (100%), 100% (100%), 91.67% (100%),
 100% (100%), 92.3% (100%), 85% (100%), respectively.



1

Fig. S6 Performance evaluation of point defect classification in simulated HAADF-STEM image. (a) Ground truth with randomly planted point defects and Poisson and Gaussian noise. (b)-(c) Point defect classification results of simulated images used as 1 ch- and 2 ch-inputs by CNN-1 and CNN-2. Among five types of point defects, all classifications were correct. The colour-codes are the same as Fig. 1. Scale bars, 0.5 nm.

2



3

Fig. S7 Performance evaluation of point defect classification in simulated HAADF-STEM image. (a) Ground truth with randomly planted point defects and Poisson and Gaussian noise. (b)-(c) Point defect classification results of simulated images used as 1 ch- and 2 ch-inputs by CNN-1 and CNN-2. Among five types of point defects all classifications were correct. The colour-codes are the same as Fig. 1. Scale bars, 0.5 nm.

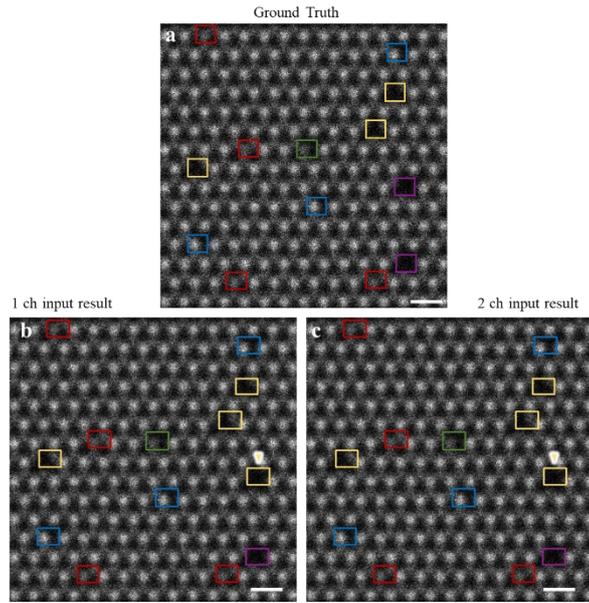


Fig. S8 Performance evaluation of point defect classification in simulated HAADF-STEM image. (a) Ground truth with randomly planted point defects and Poisson and Gaussian noise. (b)-(c) Point defect classification results of simulated images used as 1 ch- and 2 ch-inputs by CNN-1 and CNN-2. The yellow arrow in (b) is misclassified to V_{Te2} ; ground truth of $V_{Te2}+V_{Mo}$. The colour-codes are the same as Fig. 1. Scale bars, 0.5 nm.

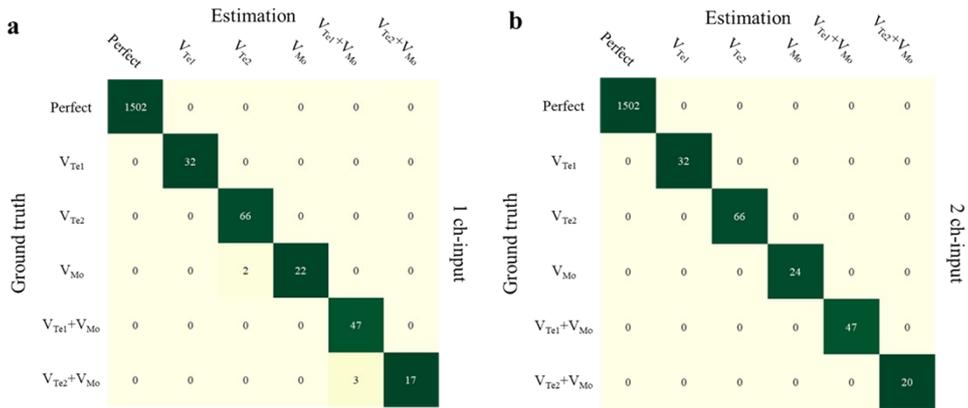
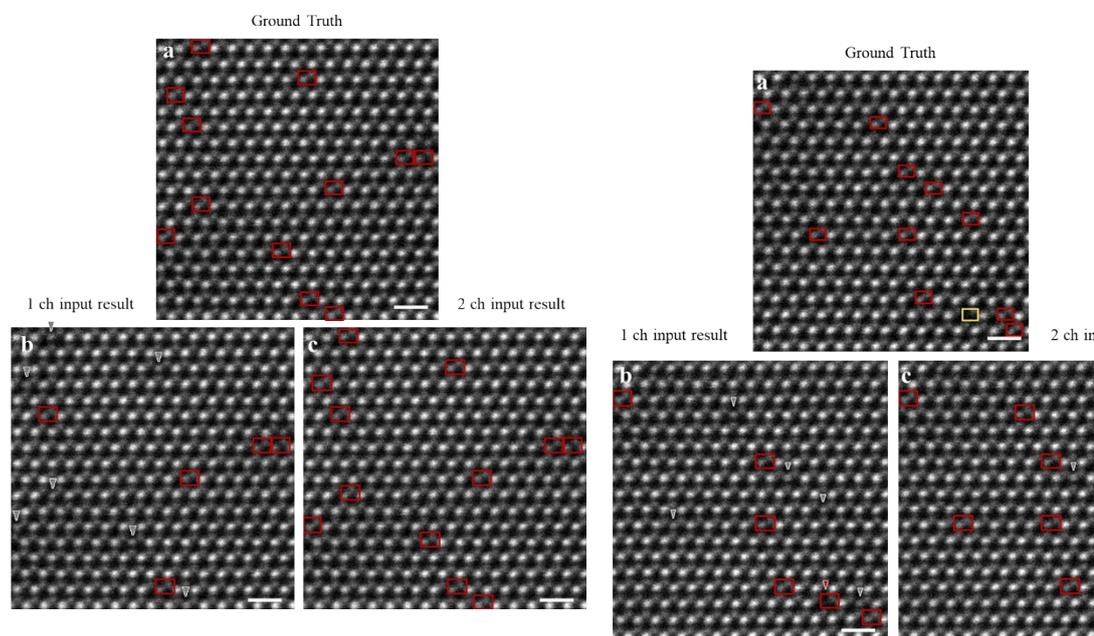


Fig. S9 (a)-(b) Confusion matrices for point defect classification performance of 10 total input simulated images (1691 point defects classified) of 1 ch- and 2 ch-input by 2DIP-Net, respectively. The elements in matrices are the counts of each classification results.

1 **Supplementary Text 5. Point defect classification performance for experimental image inspection.**

2 After assuring the point defect classification performance for simulated images, we inspected point defects in pristine 2H-
 3 MoTe₂ experimental images. The pristine 2H-MoTe₂ accommodates three types of point defect as illustrated in Fig. 6 and
 4 Figs. S10-S13: V_{Te2}, V_{Te1}, and V_{Mo}. The V_{Te1} is dominant defect types, and multiple defect types were not detected. Notably,
 5 the 2 ch-input type for CNN-2 exhibited enhanced point defect classification accuracies than 1 ch-input type as aforementioned
 6 in Fig. 6 and Discussions.

7
 8



9

Fig. S10 Performance evaluation of point defect classification in experimental HAADF-STEM image. (a) Ground truth with defined point defects. (b)-(c) Point defect classification results of experimental images used as 1 ch- and 2 ch-inputs by the CNN-1 and CNN-2. The grey arrows in (b) are misclassified types to Perfect ; ground truth of V_{Te1}. The colour-codes are the same as Fig. 1. Scale bars, 0.5 nm.

Fig. S11 Performance evaluation of point defect classification in experimental HAADF-STEM image. (a) Ground truth with defined point defects. (b)-(c) Point defect classification results of experimental images used as 1 ch- and 2 ch-inputs by the CNN-1 and CNN-2. The grey (red) arrows in (b) are misclassified types to Perfect (V_{Te1}) ; ground truth of V_{Te1} (V_{Te2}) and likewise, the red arrow in (c) is a misclassified to Perfect ; ground truth of V_{Te1}. The colour-codes are the same as Fig. 1. Scale bars, 0.5 nm.

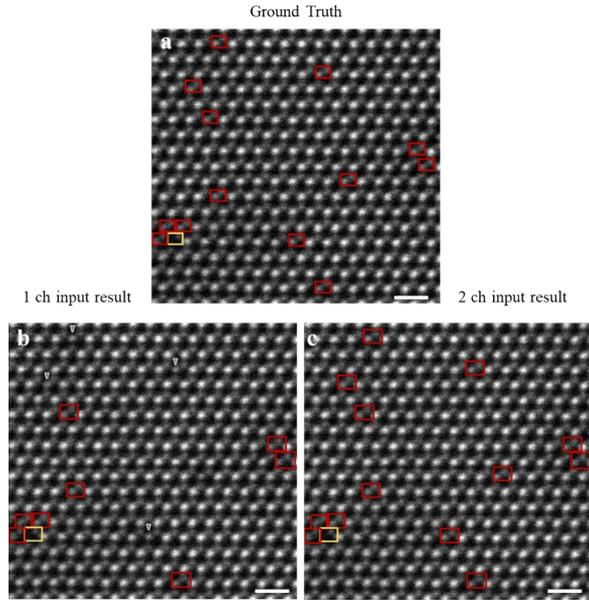


Fig. S12 Performance evaluation of point defect classification in experimental HAADF-STEM image. (a) Ground truth with defined point defects. (b)-(c) Point defect classification results of experimental images used as 1 ch- and 2 ch-inputs by the CNN-1 and CNN-2. The grey arrows in (b) are misclassified types to Perfect ; ground truth of V_{Te1} . The colour-codes are the same as Fig 1. Scale bars, 0.5 nm.

To collate quantitative point defect classification capability for experimental images for 1 ch- and 2 ch-input types, confusion matrices were illustrated in Fig. S13. Again, the point defect classification accuracies for “ V_{Te1} ” were much enhanced for 2 ch-input type compared to 1 ch-input type as mentioned in Fig. 3 and Discussion, indicating the 2 ch-input type is the optimal input types to inspect point defects in experimental images. Though, there were some misclassifications of point defects in Fig. S13b. If unit cells are imperfectly cropped by our hexagonal-net casting process, CNN-2 may be confused for defining point defect types. In experimental images, it is more challenging to determine exact unit cell boundaries. For multiple defect types, this uncertainty is exacerbated, leading to imprecise point defect estimation. Intensive examination of multiple defect types reinforcing the training data would secure robust unit cell detection followed by accurate point defect classification.

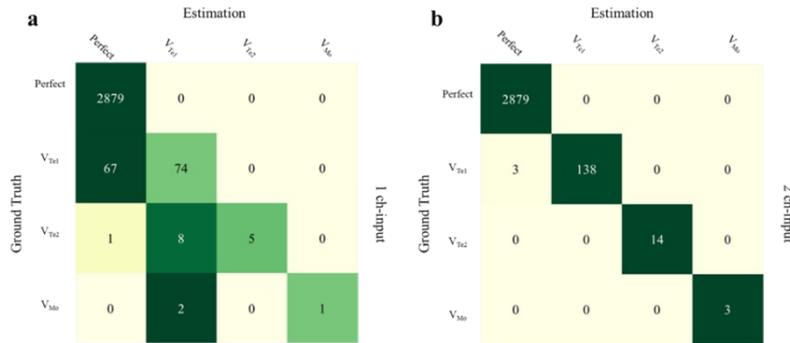
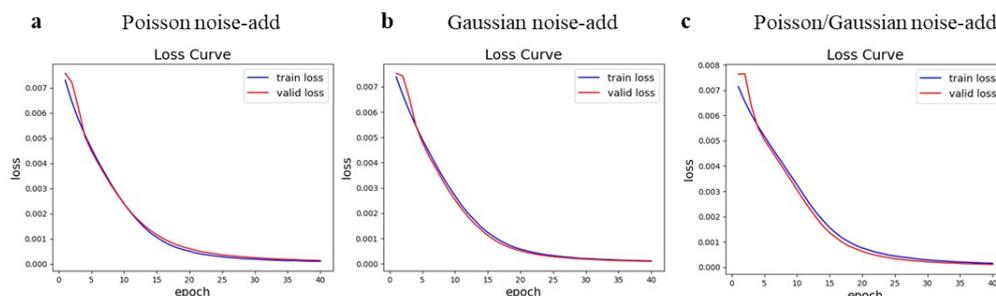


Fig. S13 (a)-(b) Confusion matrices for point defect classification performance of 11 total input experimental images (3037 point defects classified) of 1 ch- and 2 ch-input by integrated CNN models, respectively. The elements in matrices are the counts of each classification results.

As mentioned in Fig. 4, Fig. S14 illustrate the training results of (a) Poisson noise, (b) Gaussian noise, and (c) Poisson-Gaussian noise added to simulation data set. For each case, the training epoch and optimizers are the same as Fig. S4.

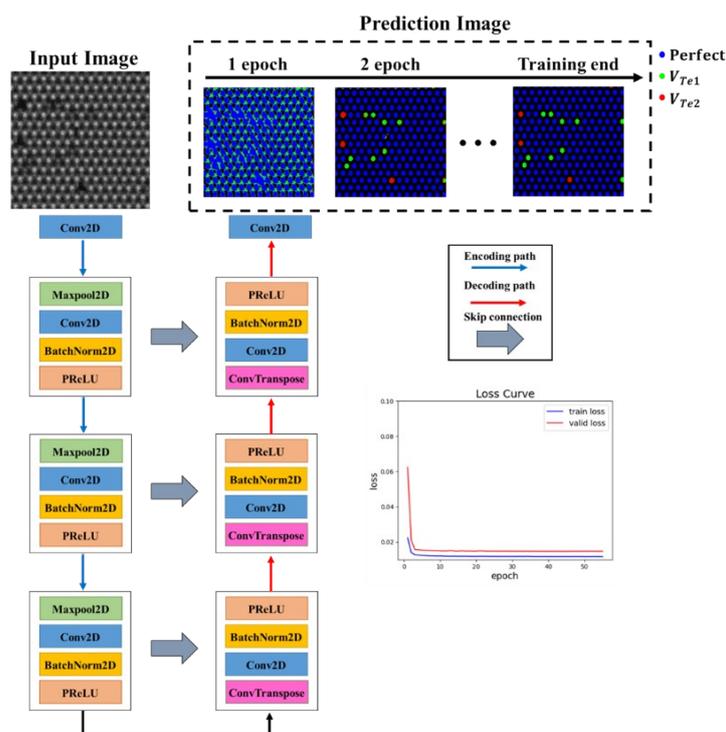
Fig. S14 (a)-(b) Loss curves for Confusion matrices for point defect classification performance of 11 total input experimental images (3037 point defects classified) of 1 ch- and 2 ch-input by integrated CNN models, respectively. The elements in matrices are the counts of each classification results.



1
2

3 Supplementary Text 6. The architecture of FCN model and its point defect classification performances.

4 Fig. S15 shows a fully convolutional neural network (FCN) towards Te-on site defect analysis. It consists of a total of 3
5 encoding modules, each containing the following components: (i) a maxpooling layer, (ii) a convolutional layer, (iii) a batch
6 normalization layer, and (iv) a PReLU activation. For the three decoding modules, the maxpooling layer of the encoding
7 module is replaced with a transposed convolutional layer. As FCN learning proceeds, it is learned pixel-by-pixel about the
8 distinction between (i) the background and (ii) Te₂-column site: presence or absence of atoms as presented in “Prediction
9 Image”. The FCN is trained by the sum of the cross entropy loss function and the dice coefficient loss function for a total of
10 100 epochs. For the simulation dataset, a total of 500 training data and 50 validation data were used, and an early termination
11 algorithm was used to prevent overfitting of the training data.



12

Fig. S15 Fully Convolution Network (FCN) segmenting Te on-site defect types. The FCN is divided into an encoding module and a decoding module, respectively. The encoding module consists of maxpooling layer, convolution layer, batch normalization layer, and PReLU activation function. The decoding module is changed to a transposed convolutional layer instead of a maxpooling layer in the encoding module. Feature concatenation is used to combine feature maps in the encoding module in the same stage with the decoding module. As training proceeds, FCN model identifies the point defect types as illustrated in “Prediction Image”. The loss curves show the train/validation for Te on-site defect classifications.

1 **Supplementary references**

- 2 1 Van Aert, S.; Den Dekker, A.; Van Dyck, D.; Van Den Bos, A. *Ultramicroscopy*, 2002, **90** (4), 273-289.
3 2 Savitzky, B. H.; El Baggari, I.; Clement, C. B.; Waite, E.; Goodge, B. H.; Baek, D. J.; Sheckelton, J. P.; Pasco, C.; Nair,
4 H.; Schreiber, N. J. *Ultramicroscopy*, 2018, **191**, 56-65.