

Supporting Information for

Deep learning-based quality filtering of mechanically exfoliated 2D crystals

Yu Saito^{1*†}, Kento Shin^{2†}, Kei Terayama^{1,3}, Shaan Desai¹, Masaru Onga⁴, Yuji Nakagawa⁴,
Yuki M. Itahashi⁴, Yoshihiro Iwasa^{4,5}, Makoto Yamada^{1,6}, Koji Tsuda^{1,2,7}

*Corresponding author: yu.saito.zk@riken.jp (Y.S.), tsuda@k.u-tokyo.ac.jp (K. Tsuda.)

†These authors equally contributed to this work.

Contents

- I. User study for identification of MoS₂ atomic layers.**
- II. Macro precisions of the U-Net approaches**
- III. Graphene data and segmentation result.**
- IV. Typical failure example of the U-net.**
- V. Traditional CV approaches for segmentation and their results.**

I. User study for identification of MoS₂ atomic layers.

Here, we describe a user study for the identification of layer number of MoS₂ atomic layers using test problems, which is exactly same data set for the test data using in calculation of accuracy rate of deep neural network. In this study, the researchers who were not familiar with 2D materials (non-expert human) were given three minutes to learn the training data set and were then asked to respond with the layer numbers (e.g. mono- and bilayer) for each segment of new images (Fig. S1) for ten seconds per each image. We summarize the results for each researcher and U-Net with or without options as shown in Supplementary Table 1.



Supplementary Figure 1. 11 optical micro scope images of MoS₂ atomic layers for test data. The parts to be answered is shown as “a” or “b” in each image.

Supplementary Table 1. Table for the user study using 11 test images for identification of MoS₂ atomic layers. A-N shows human users (researchers with not familiar 2D materials). The checkmarks mean answering of the correct layer number. O, P, Q and R correspond to normal U-Net, w/ contrast adjusted images, w/grayscale and w/ grayscale & contrast adjusted images, respectively.

	①-a	①-b	②-a	②-b	③-a	④-a	⑤-a	⑥-a	⑦-a	⑧-a	⑧-b	⑨-a	⑩-a	⑩-b	⑪-a
A	✓		✓	✓		✓	✓		✓			✓	✓	✓	✓
B	✓	✓			✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
C	✓	✓		✓			✓	✓	✓	✓	✓	✓	✓	✓	✓
D			✓	✓	✓	✓	✓	✓				✓	✓	✓	✓
E	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓
F		✓	✓	✓						✓	✓	✓	✓	✓	
G			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
H	✓		✓	✓			✓	✓		✓	✓	✓	✓	✓	✓
I			✓	✓	✓	✓		✓				✓	✓	✓	✓
J			✓	✓	✓				✓			✓	✓	✓	✓
K	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓			
L			✓		✓	✓		✓	✓			✓	✓	✓	✓
M					✓		✓	✓	✓			✓	✓	✓	✓
N				✓		✓	✓	✓				✓	✓	✓	✓
O		✓		✓	✓	✓	✓	✓	✓			✓	✓	✓	✓
P		✓	✓		✓	✓	✓		✓		✓	✓	✓	✓	✓
Q	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓		✓	
R		✓	✓		✓	✓	✓	✓	✓			✓	✓	✓	✓

II. Macro precisions of the U-Net approaches.

Supplementary Table 2. Cross-validation score of macro-average precision for U-Net and U-Net with weighted loss (MoS₂).

	U-Net	U-Net with weighted loss
w/o grayscale & w/o contrast adjusted images	0.653	0.775
w/ grayscale	0.521	0.559
w/ contrast adjusted images	0.655	0.743
w/ grayscale & contrast adjusted images	0.654	0.686

III. Graphene data and segmentation result.

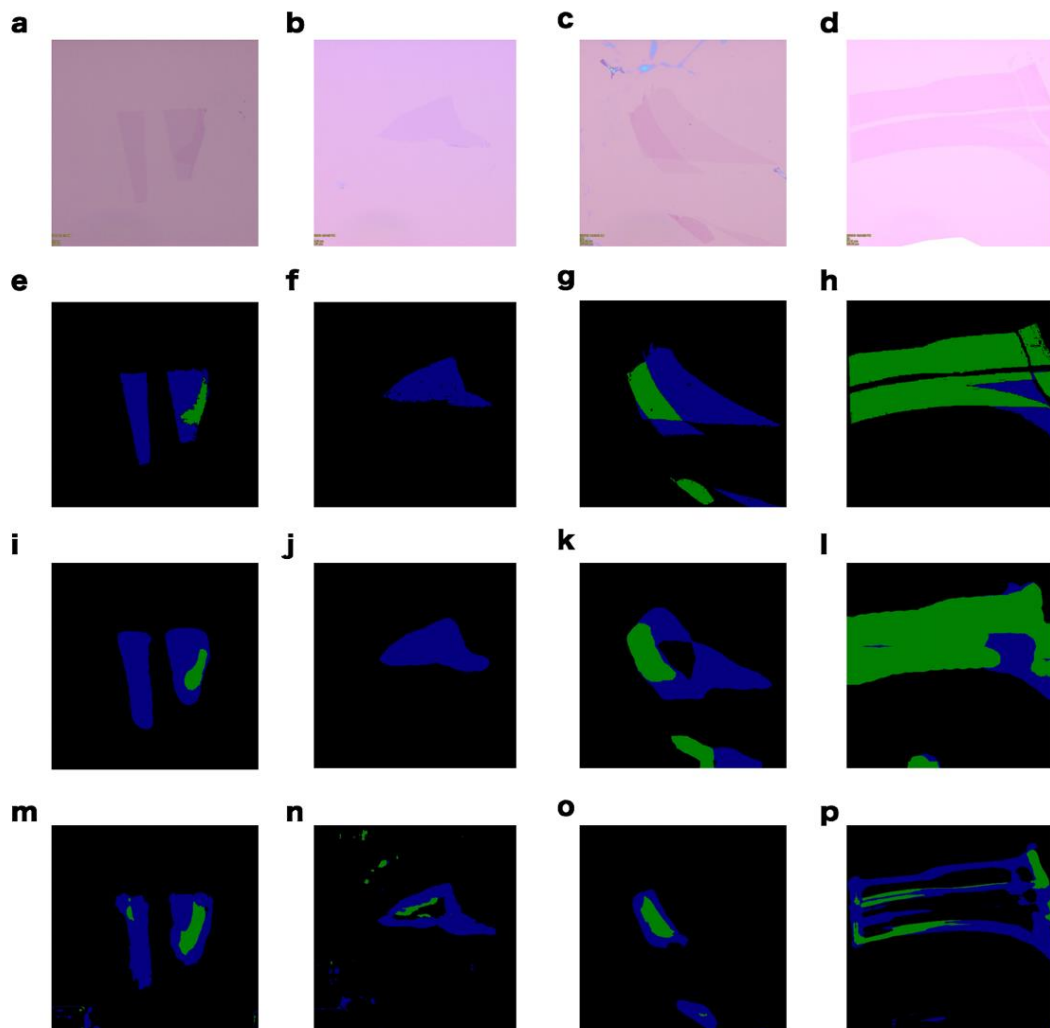
We collected optical microscope images of mechanically exfoliated graphene flakes using the same optical microscope as used for MoS₂ and verified the effectiveness of the proposed method.

We trained the U-net based model to predict monolayer, bilayer, and other thicker parts (including trilayer, 4 layer and >5 layer) parts using the graphene data. We used the same learning procedure, parameters, and augmentations of the U-net based model for MoS₂. Here, we focused on the monolayer and bilayer graphene because the number of optical microscope images optical of thicker multilayer limited.

We summarize the cross-validation scores (mean of pixel-wise accuracies) for each learning process of U-net based approaches in Supplementary Table 3. From these results and Table 1 in the manuscript, U-net based approaches for graphene showed better performances than those for MoS₂, and, in particular, simple (w/o grayscale & w/o contrast adjusted images) U-Net with weighted loss achieved an accuracy of 0.801. For MoS₂, the prediction model with contrast adjusted images had the highest accuracy, but for graphene, the simple model had the highest prediction accuracy. This may be because the variations of color appearance were less than that of MoS₂ and thus it was possible to predict without adding a contrast adjusted image. Figure S2 shows four examples of the original graphene images, the segmented images and generated images using U-Net with weighted loss. We also show generated images by the pixel-wise CNN model in the bottom row in Fig. S2. Although pixel-wise CNN model succeeded in relatively accurate prediction for (m) and (n) in Fig. S2, the model failed to predict both layer position and type in (o) and (p). The cross-validation score by pixel-wise CNN was 0.634. From these predicted images, it is confirmed that U-net achieved highly accurate prediction reflecting flake shapes as MoS₂.

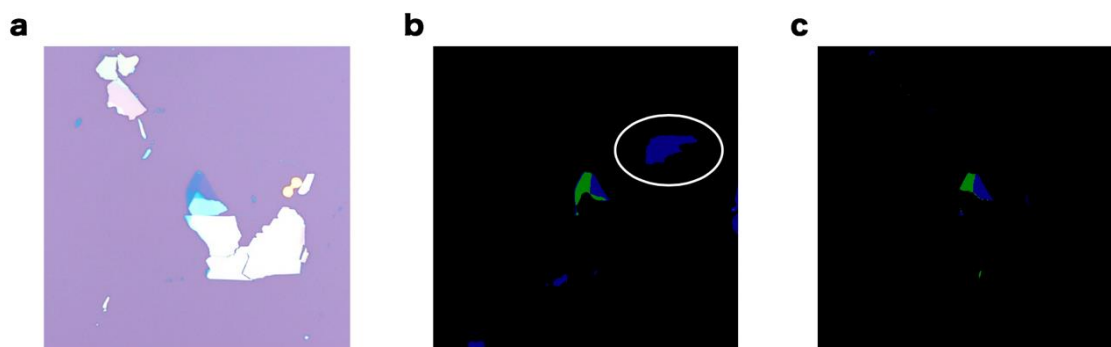
Supplementary Table 3. Cross-validation scores for graphene layers by U-Net and U-Net with weighted loss.

	U-Net	U-Net with weighted loss
w/o grayscale & w/o contrast adjusted images	0.714	0.801
w/ grayscale	0.692	0.599
w/ contrast adjusted images	0.762	0.763
w/ grayscale & contrast adjusted images	0.688	0.788



Supplementary Fig. 2 Examples of original optical microscope images of graphene (a-d), segmented images (e-h), generated images based on U-Net with weighted loss (i-j), and generated images based on pixel-wise CNN (m-p). In the segmented and generated images, blue and green region show monolayer and bilayer region, respectively.

IV. Typical failure example of the U-net.



Supplementary Figure 3. Typical failure example of the U-net for prediction. (a) original images (MoS₂), (b) prediction by the U-net, (c) correct answer (blue region is monolayer and green region is bilayer.) The region in the white circle in (b) was predicted as monolayer, although the same region in (a) is background.

V. Traditional CV approaches for segmentation and their results.

Here, we describe some traditional CV approaches and comparison of their performances with the U-Net based approach. We employed compared traditional CV approaches based on a bag-of-visual-words (BoVW)¹ approach using a hand-crafted image descriptor and based on convolutional neural networks (CNN)². The BoVW-based and CNN-based models predict three classes (monolayer, bilayer and others) of each pixel from the local image around the pixel. The size of the local image was 256×256 pixels. For BoVW, we employed speed up robust features (SURF)³, which is a rotation and scale invariant feature and is one of the most widely used descriptor, as the hand-crafted descriptors. We built 256 clusters from the obtained descriptors using the K-means algorithm for BoVW. We employed two machine learning approaches support vector machine (SVM) and random forest (RF) as non-linear prediction methods of layers from the calculated features. We refer to these BoVW-based methods using SVM and RF by SURF+SVM and SURF+RF, respectively. We also developed a layer prediction model using CNN based on a pixel-wise image segmentation method⁴. In the model, an input image is passed through five convolution layers with a rectified linear unit (ReLU) as an activation function and then two full connection layers. In the convolution layer, we employed the batch normalization and dropout methods, which are commonly used to improve generalization performance. We refer to this method by pixel-wise CNN. To train SURF+SVM, SURF+RF, and pixel-wise CNN models, we used random flipping, rotating, and changing the color of the input images as the augmentation procedure for the U-net training.

We summarize the cross-validation scores (mean of pixel-wise accuracies) for these traditional methods in Supplementary Table 4. From these results and Table 1, U-net based approaches showed better performances than the traditional models. We also show prediction examples by pixel-wise CNN in Fig. 3 (m-p).

Supplementary Table 4. Cross-validation scores for SURF+SVM, SURF+RF, and pixel-wise CNN.

	Score
SURF+SVM	0.433
SURF+RF	0.492
Pixel-wise CNN	0.596

Supplementary References

1. G. Csurka, C. Bray, C. Dance, and L. Fan, Visual categorization with bags of keypoints, *In Proc. Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision*. 1-22 (2004).
2. Y. LeCun, Y. Bengio, and G. Hinton, Deep learning. *Nature* 436–444 (2015).
3. H. Bay, T. Tuytelaars, and L. V. Gool, SURF: Speeded Up Robust Features, *In Proc. European Conference on Computer Vision*, 404-417 (2006).
4. D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, *In Proc. Advances in Neural Information Processing Systems* 25. 2843-2851 (2012).