Electronic Supplementary Material (ESI) for ChemComm.

Supplementary Information for

Deep learning assisted analysis of single molecule dynamics from liquid-phase electron microscopy

Bin Cheng[a], Enze Ye[b,c], He Sun[b,*], Huan Wang[a,*]

a. Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Center for Spectroscopy, Beijing Key Laboratory of Polymer Chemistry & Physics of Ministry of Education, Center for Soft Matter Science and Engineering, National Biomedical Imaging Center, Peking University, Beijing, 100871.
b. College of Future Technology, National Biomedical Imaging Center, Peking University, Beijing, 100871.
c. Department of Electronic Engineering, Tsinghua University, Beijing, China, 100084

This file includes:

1. Conventional automatic threshold method

Using the conventional method, Gaussian denoising and BM3D denoising were used to smooth the image, and global threshold and adaptive threshold were used for image segmentation. The main processes of image segmentation are (1) denoising, (2) automatic selection of threshold, and (3) judgment of morphology, as shown in Fig. S1a. Fig. S1b, and Fig. S1c are the binarized results from the global threshold and adaptive threshold respectively.

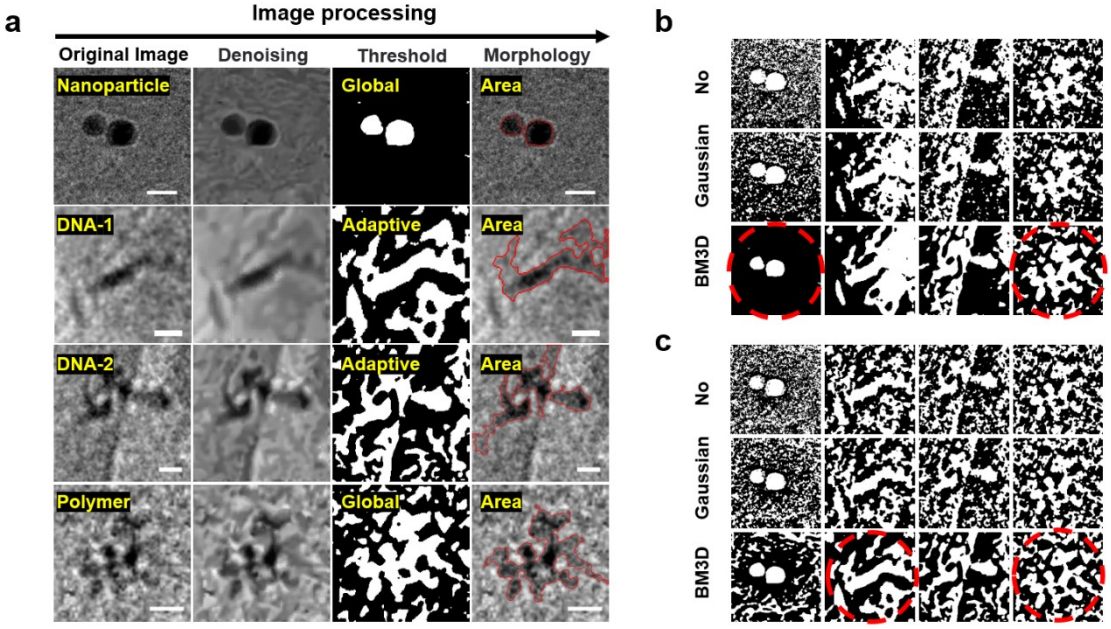

Fig. S1. a. Flow chart of conventional automatic threshold image processing and the corresponding binarized images using global **(b)** and adaptive **(c)** threshold methods. Data adapted from ref. 1 for polymer, from ref. 3 for DNA. Scale bar: 10 nm.

2. *Challenging features of LP-EM images*

The change of parameters between consecutive frames is due to the rich conformational dynamics as quantified in b to e.

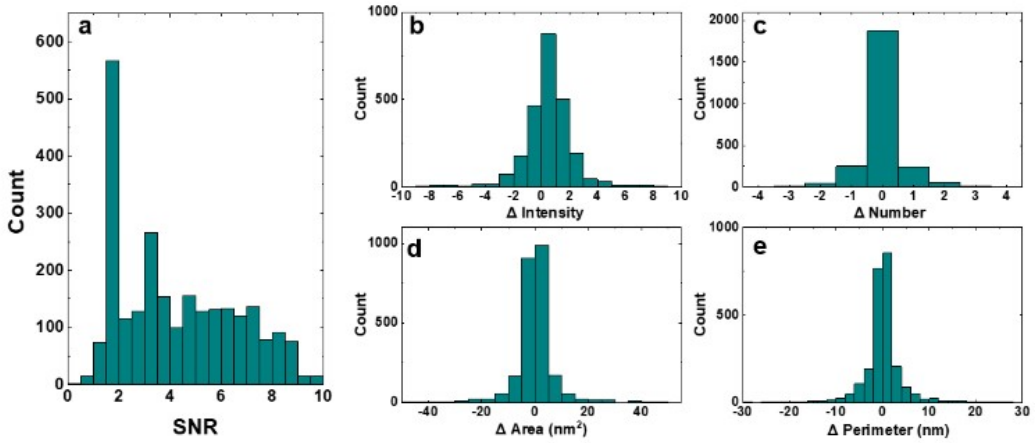

Fig. S2. Summary of challenging features of single molecule LP-EM images for image segmentation. 2500 images from 8 movies were obtained from graphene liquid cells. Data were adapted from refs 1, 3, and 8. **a.** The mean SNR ratio of each image. **b-e.** Changes of target molecule parameters in the consecutive images. Mean intensity (**b**), the number of objects (**c**), area (**d**), and perimeter of the target molecule (**e**).

3. *Dataset and neural network training*

2500 images from 8 movies obtained from graphene liquid cells, adapted from ref. 1,3,8 in the main text was used for training. Raw data obtained from electron microscopy were cropped such that each image contains one to three molecules of interest. These images were then labeled by three well-trained LP-EM experimentalists to determine the ground truth. Different types of images are renamed, respectively as TypeA to TypeH. The neural network molecules were categorized. Images were all resized to 96 pixels x 96 pixels for training. The neural networks used for training are respectively UNet++ and

U-Net, and the models are trained on NVIDIA GeForce RTX 3070 Ti. The training mainly adjusts the hyperparameters including epoch, batch size, learning rate, etc., to find the best result. In the training process, batch sizes included 4, 8, 16, 32, and 64. It was found that loss was the lowest and IoU was the highest at a batch size of 4. The learning rate was tested at 10^{-3},10^{-4} and 10^{-5}, and the performance reached the optimum at 10^{-4}. The epoch reaches the optimum around 500, and overfitting occurs after 500. We have also tried to increase the number of filters and convolutional layers of the network and found that their influence can be ignored.

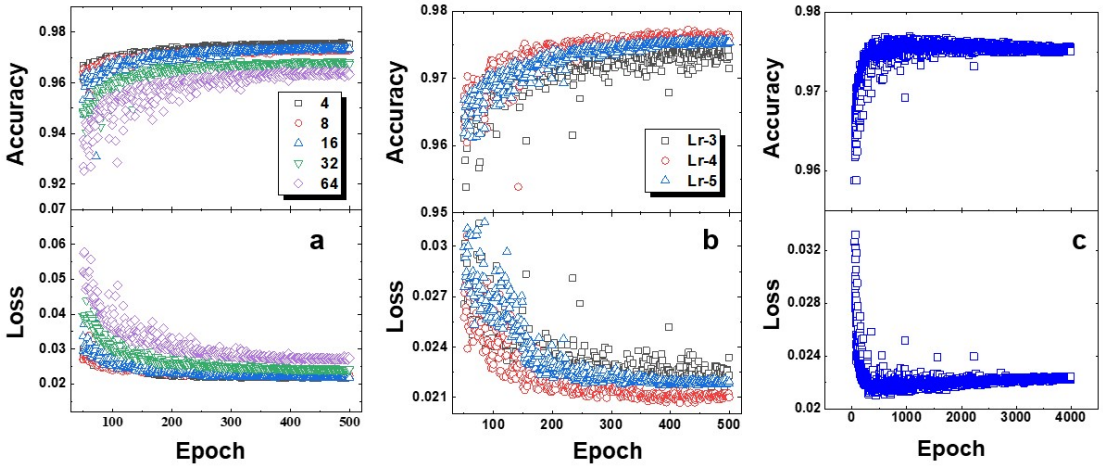

Fig. S3. The influence of hyperparameters such as batch sizes (a), learning rate (b), and epoch (c) was tested. **a**. Batch size: squares, circles, up triangles, down triangles, and rhombic correspond to a batch size of 4, 8, 16, 32, and 64, respectively. **b**. Learning rate: squares, circles, and triangles correspond to 10^{-3},$10^{-4,}$ and 10^{-5}, respectively. All tested with UNet++ network.

4. Effect of SNR on prediction precision

We explore the relationship between SNR and prediction precision by adding Gaussian noise to a high SNR LP-EM image to reduce the SNR. We found that when SNR is smaller than 1.5, prediction precision was significantly affected. It is likely because our training data only contained a limited amount of images that fall into this SNR range as shown in Fig.S2a. Precision is defined as TP/(TP+FP), TP stands for true positive and FP stands for false positive.

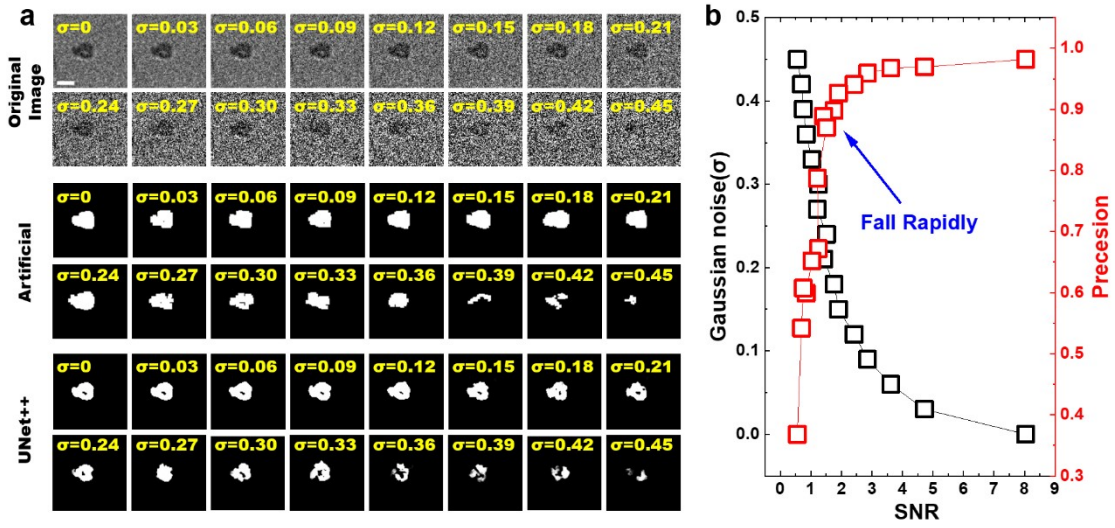

Fig. S4. Effects of SNR on prediction precision. **a**. Increasing Gaussian noise in the original image and the corresponding masks obtained from the artificial thresholding method and U-Net++ prediction. Scale bar: 10 nm. **b**. Increasing Gaussian noise leads to a reduction in the prediction precision. The Network used for testing was UNet++.

5. Time information in the image sequence was used to test the "visual staying phenomenon"

"Visual staying phenomenon" seemed to have been well used by the deep learning methods, although time-dependent information was not purposely fed to the network. As we changed the input image sizes from 1 to 3, 5, and 7 consecutive frames, we found no change of accuracy.

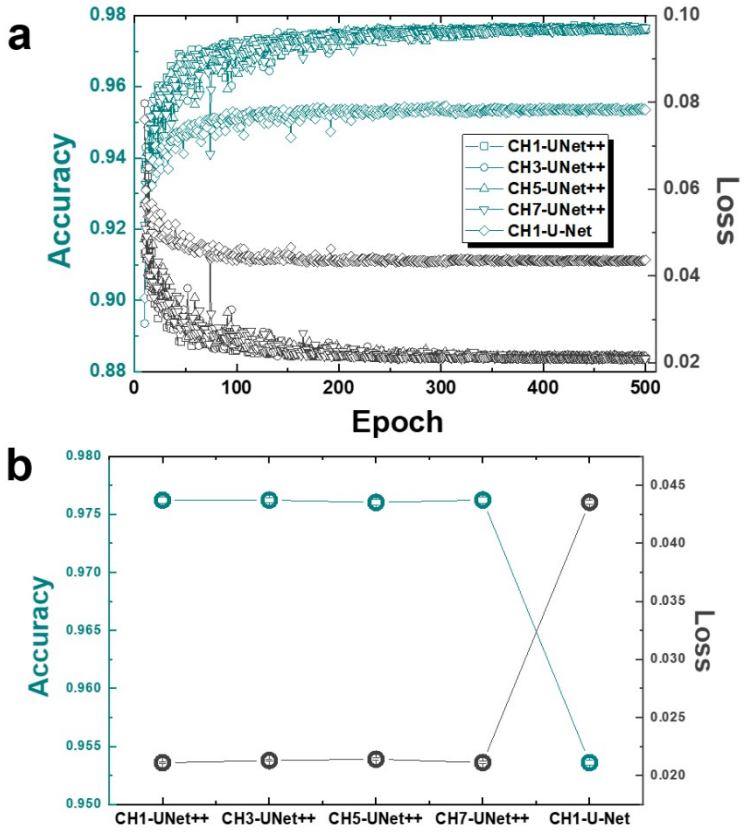

Fig. S5. a. Accuracy and loss were plotted as a function of the epoch when using different numbers of input sample sizes, from a single image (CH1-UNet++, squares) to three (CH3-UNet++, circles), to five (CH5-UNet++, up triangles) and seven (CH7-UNet++, down triangles) consecutive images in a sequence for U-Net++, and the comparison with U-Net trained with a single image (CH1-UNet, rhombic). **b.** Accuracy and loss were plotted as the function of input sample sizes. Values were obtained at 480-500 epochs when curves converged.

6. *Time-dependent transient single-molecule interactions*

The conventional threshold method consists of (1) BM3D denoising, (2) selection of threshold and (3) determination of area order. Step (3) is to manually select the features of interest according to their sizes which were ranked in order. Area order indicates the specific ranking for the feature of interest selected in each image. It is worth noting that the build-in model of StarDist in TrackMate7 did not work effectively likely because the training data were optical images (Fig. S6a), instead we reported results by using the same 2500 electron micrograph images in the main text Fig.3a.

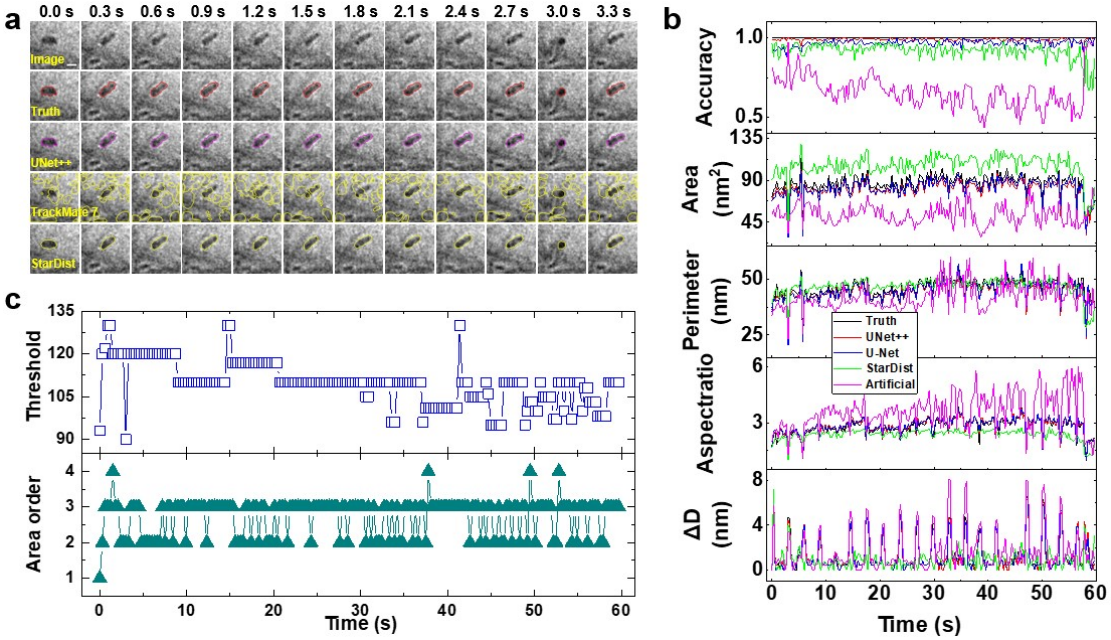

Fig. S6. a. Comparison of different image segmentation methods, where trackmate7 corresponds to the built-in StarDist model in "TrackMate 7" toolbox installed in image J and the last row of StarDist is the model trained using 2500 sets of LP-EM images. **b.** Accuracy, area, perimeter, aspect ratio, and center-of-mass displacement ΔD of individual molecules using the four methods for image segmentation: conventional threshold (pink), U-Net (red), UNet++(blue), StarDist (green) and their comparisons to the ground truth (black). **c**. Parameters used for conventional threshold method to achieve image segmentation. The entire sequence is 200 frames, with a frame rate of 0.3 s, data adapted from ref. 3 in the main text. Scale bar: 10 nm.

7. Deep Learning identifies the degradation process of molecules

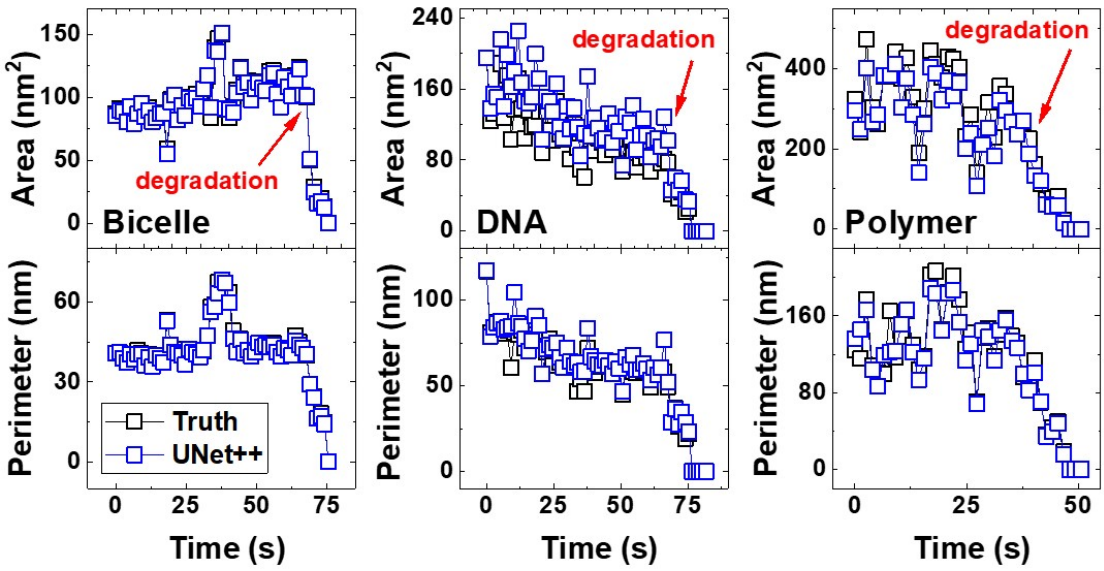

Fig. S7. Degradation process of different biomolecules quantified by U-Net++. Data adapted from ref. 7 in the main text. The stark contrast between the random fluctuation and the monotonic decay captured when degradation processes started is shown by the red arrows on each graph.

8. *Local conformational changes within a molecule.*

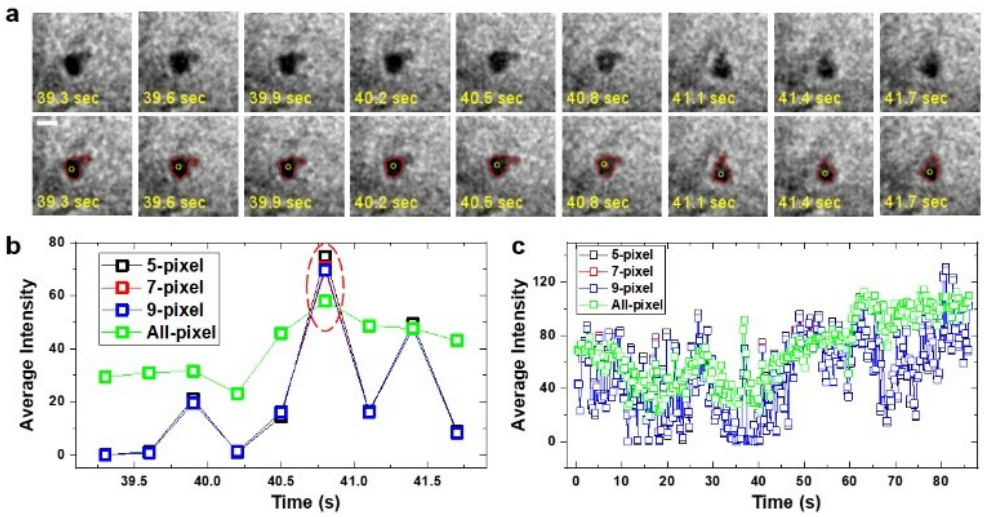

Fig. S8. Deep learning methods reveal local conformational dynamics. a. The first row shows the original image of the intermediate dimerization interaction of DNA, the second row shows the corresponding predicted region (red circles, UNet++) as well as the traced region (yellow circles), data adapted from ref. 3. Scale bar: 10 nm. b. Variation in the average intensity of different pixels in the tracked region, 5-, 7-, 9-pixel and all-pixel respectively. c. Time-dependent changes of intensity throughout the full sequence.

Legends for Movies.

Movie S1. This movie shows the comparisons of image segmentation results obtained from three networks, UNet++, U-Net, and StarDist, with the conventional artificial threshold method for conformation changes of a single DNA molecule. Raw data adapted from ref. 3. Top panel, from left to right, raw image, the ground truth obtained by labeling, and the conventional artificial threshold method. The bottom panel, from left to right, results from UNet++, U-Net, and the StarDist. Scale bar: 10 nm.

Movie S2. This movie shows the comparisons of image segmentation results obtained from two networks, U-Net and UNet++, with the conventional artificial threshold method for conformation changes of three DNA molecules during their interactions to complete a hybridization process. Raw data adapted from ref. 3. Top panel, from left to right, raw image, after denoising with BM3D, the ground truth obtained by labeling. The bottom panel, from left to right, results from UNet++, U-Net, and the conventional artificial threshold method. Scale bar: 10 nm.

Movie S3. This movie shows the deep learning analysis revealed the local conformation dynamics within a single molecule during DNA hybridization. Raw data adapted from ref. 3. The Network used for testing was UNet++.Loop annealing mechanism was confirmed by the capture of the intermediate loop state by selective analysis of small-scale images such as 5, 7, and 9 pixels, but was masked by the analysis of the entire molecule. Left panel: top, raw image; bottom: contour of the target molecule and the corresponding selection of local features. Right panel: real-time plot of intensity values for different sizes of selections. Scale bar: 10 nm.