J. Adv. Simulat. Sci. Eng.
Vol. 10, No. 1, 172-181.

Measurement of DNA length on video of fluorescence microscope by pix2pix trained by molecular dynamics simulation

Seiki Saito[1*], Ayumi Kobashi[1], Haruto Miura[1], Hiroaki Nakamura[2,3], Takahiro Kenmotsu[4], Yasuhisa Oya[5], Yuji Hatano[6]

[1]Department of Information and Electronics, Yamagata University
[2]Department of Helical Plasma Research, National Institute for Fusion Science
[3]Department of Electrical Engineering, Nagoya University
[4]Department of Biomedical Engineering, Doshisha University
[5]Division of Radioscience, Shizuoka University
[6]Hydrogen Isotope Research Center, University of Toyama

*saitos@yz.yamagata-u.ac.jp

Received: December 30, 2022; Accepted: March 6, 2023; Published: May 23, 2023

Abstract. For the automatic measurement of DNA length in the video captured by fluorescence microscope, image processing method using pix2pix is developed. To increase the accuracy of the measurement in low-resolution video, a method of inputting multiple time-series frames to pix2pix are proposed. To generate the training data, molecular dynamics simulation is employed to prepare dummy video. The trained pix2pix by multiple frames input can generate more accurate images in a scene of DNA breakage than the pix2pix trained by single frame input.

Keywords: DNA, Molecular dynamics simulation, Pix2pix, Image processing

1. Introduction

Single-molecule observation method is widely used for the investigation of double-strand breaks in DNA. In the observation, images of DNA molecules can be captured by fluorescence microscopes by using fluorescent dye such as YOYO-1 as a photosensitizer. By measuring changes in DNA length, it is possible to estimate the number of double-strand breaks. Usually, the measurement is done manually by experienced operators and it is time consuming. In our previous research [1], we developed an automation method of DNA length measurement achieved by the following two steps: 1) extracting only DNA segments from fluorescence microscope images by a deep learning model: pix2pix [2], and 2) measuring the length of the

DNA segments by OpenCV. Figure 1 shows an example of the result of our previous method. The original fluorescence microscope image (Fig. 1 (a)) is converted to an image (Fig. 1 (b)) which only DNA segments are shown by skeletonized lines. Our proposed method is effective for the analysis of still images when the resolution of images is sufficiently high.

In order to analyze the time evolution of number of DNA breaks, motion of DNA segments are captured as video format by fluorescence microscopes in an experiment done at Doshisha university, Japan. Figure 2 shows an example of several frames in time series of the videos captured by an inverted microscope (Axiovert 135 TV, Carl Zeiss, Germany) equipped with an oil-immersed 100V objective lens. The video captured the moment when a DNA segment breaks. Comparing to the still images, the resolution of videos are not sufficiently high. In order to accurately analyze the time evolution of the number of DNA breaks, it is necessary to determine the exact time when DNA breaks. However, in the case of low resolution images, it is difficult to measure the number of DNA segments from single still image because it is difficult to distinguish the DNA segments because of the effect of optical blur, especially around the moment when two separated segments, which are generated by breakage, are located closely.

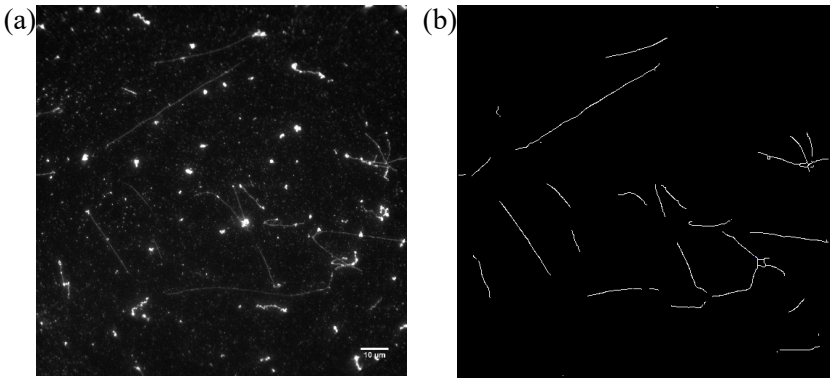


Figure: 1 (a) The original input images obtained by actual fluorescence microscopes.
(b) The final images after the processing by deep learning models. [1]

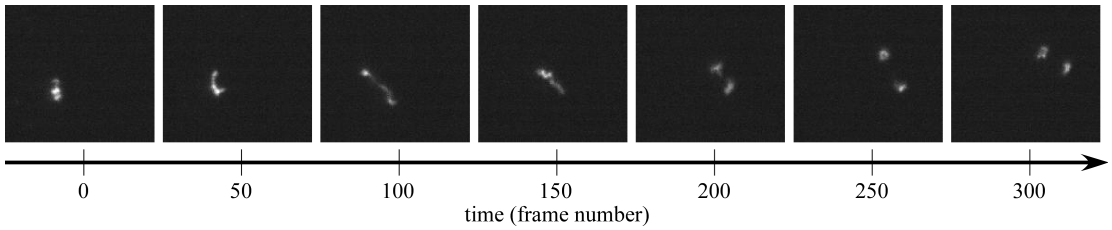


Figure 2: Example of videos which captures the motion of a DNA segment when it breaks.

2. Method and usage of pix2pix

Although it is not possible to accurately determine the number of DNA segments from a single still image, human can judge comprehensively the segments are really separated or not by

looking the previous and next frames as well. Therefore, in this paper, we developed a method by determining DNA breakage by inputting not only single frame, but also the frames before and after the frame into the pix2pix network model as shown in Fig. 3.

Figure 4 (a) shows the usual usage of pix2pix network. pix2pix is a widely used deep neural network model for generating another image from an image with 3 channels of RGB. Therefore the input layer of pix2pix has *W*x*H*x3 nodes, where *W* and *H* are width and height of the input image, respectively. To input 5 consecutive gray scale images, in our usage, we modified the number of nodes of input layer of the generator of pix2pix to *W*x*H*x5 as shown in Fig. 4 (b). Of course, the output image can be modified to be gray scale image though, we did not changed it because the structure of output image are linked to the network structure of discriminator of pix2pix.

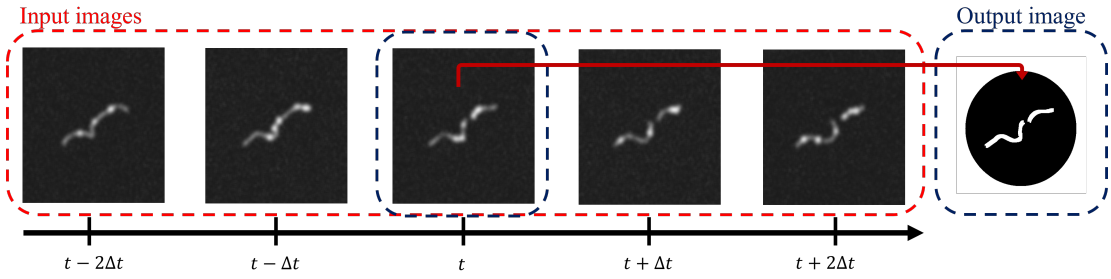


Figure 3: Concept of determination of DNA breakage from five images along a time series.

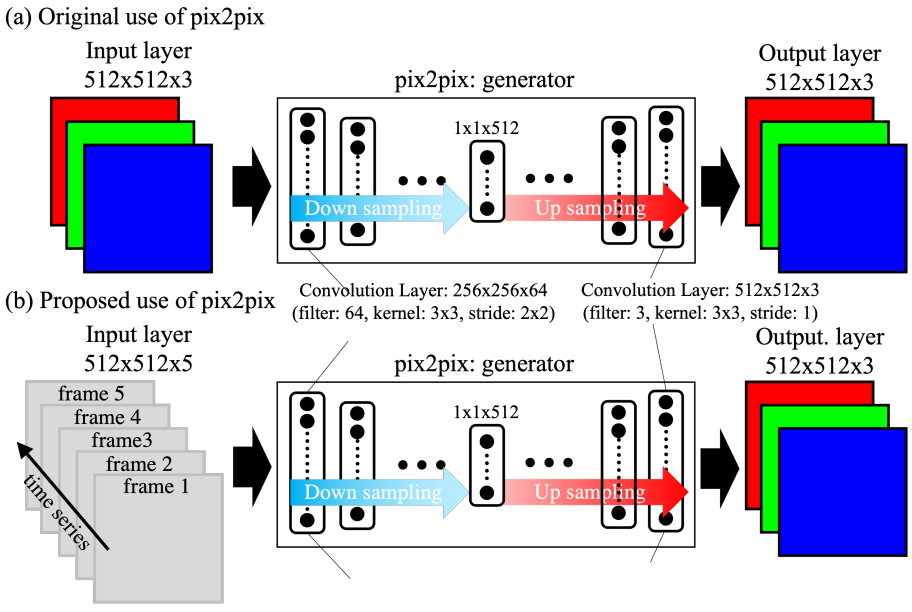


Figure 4: (a) Usual usage of pix2pix network. (b) Our usage of pix2pix network which predict one images from multiple (5) frames.

3. Generation of dummy video by molecular dynamics simulation

For the training of our pix2pix network, it is necessary to prepare many fluorescence microscope videos of DNA. Moreover, clear DNA images corresponding to the video must be prepared as the training data. However, it is difficult to prepare enough videos that capture the moment the DNA breaks, and even more difficult to generate corresponding training data from the captured video. Therefore, we employ molecular dynamics (MD) simulation to generate the dummy video of DNA motion for the training.

For the MD simulation, first we prepared the atomic model of telomere structure of the human DNA shown in Fig. 5 (a). The telomeric DNA which is consists of 1028 atoms is obtained by removing TRF2 protein from the TRF2-Dbd-DNA complex, PDB ID of which id 3SJM [3]. The telomeric DNA has 17 base pairs: $d(TCTAGGGTTAGGGTTAG)_2$. Structural relaxation is applied to the structure by LAMMPS (Large-scale Atomic Molecular Massively Parallel Simulator) with a Reax-FF (CHON-2017_weak) [4, 5] force field. We note that CHARMM and AMBER can also be used to deal with the structure of DNA to achieve the objectives of this research. The length of the atomic model after the relaxation r_{eq} is 58.1 Å. The length of DNA in the atomic model is too short compared to the actual DNA length in the video captured by fluorescence microscope whose length is order of tens of μm. However, because of the computation cost of the MD simulation, is difficult to prepare the longer one comparable in length to the actual DNA in atomic model. Therefore, to calculate the motion of DNA whose length is comparable to the actual DNA, we prepared a coarse-grained model from the atomic model. In the coarse-grained model, atoms in the atomic model shown in Fig. 5 (a) is represented as a single coarse-grained particle as shown in Fig. 5 (b). Then, we chained 8608 coarse-grained particles at intervals of r_{eq} to represent the DNA whose length of 50.0 μm.

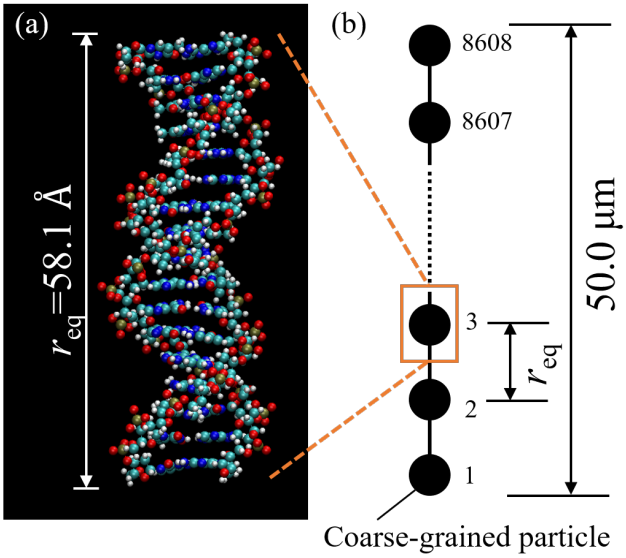


Figure 5: (a) Atomic model of Human's telomeric DNA which consists of 17 base pairs. (b) Coarse-grained particle model of DNA.

To calculate the force acting on each coarse-grained particle, the following simple equations are applied for the force field.

$$U = \sum_{i=1}^{N-1} U_{i,i+1}^{\mathrm{t}} + \sum_{i=2}^{N-1} U_{i-1,i,i+1}^{\theta} , \qquad U_{i,j}^{\mathrm{t}} = K_{\mathrm{t}}\left(r_{i,j} - r_{\mathrm{eq}}\right)^2, \qquad U_{i,j,k}^{\theta} = K_{\theta}{\theta_{i,j,k}}^2, \quad (1)$$

where $r_{i,j}$ is the distance between *i*-th and *j*-th particles, and $\theta_{i,j,k}$ is the angle between the vector $\boldsymbol{r}_j - \boldsymbol{r}_i$ and the vector $\boldsymbol{r}_k - \boldsymbol{r}_j$, where $\boldsymbol{r}_i$ is the position of the *i*-th particle. K_{t} and K_{θ} are fitting parameters determined by the potential energy of atomic models. We note that the $U_{i,j}^{\mathrm{t}}$ and $U_{i,j,k}^{\theta}$ are the terms related to the tensile stress and bending stress of the DNA, respectively. From the force field U, the force acting on *i*-th particle F_i can be calculated by the following equation.

$$F_i = -\nabla_i U = -\nabla_i U_{i,i+1}^{\mathrm{t}} - \nabla_i U_{i-1,i}^{\mathrm{t}} - \nabla_i U_{i,i+1,i+2}^{\theta} - \nabla_i U_{i-1,i,i+1}^{\theta} - \nabla_i U_{i-2,i-1,i}^{\theta} \quad (2)$$

To determine the parameters K_{t} and K_{θ}, we calculate the change of the potential energy when the atomic model is pulled in the axial direction (Fig. 6 (a)) and bent perpendicular to the axis (Fig. 6 (b)), respectively. Figure 7 shows the change of the potential energy from the relaxed structure calculated by LAMMPS with the Reax-FF force field. The potential energy of each state is obtained after energy minimization is performed with fixing both ends of the DNA. The DNA is placed in vacuum in this calculation. Although the parameter such as r_{eq}, K_{t}, and K_{θ} vary depending on whether the DNA is in liquid water, these effects are not considered in this research. The circles denote the change of the potential energy calculated by LAMMPS. The solid line is the fitting carve of the harmonic function $K_{\mathrm{t}}(r - r_{\mathrm{eq}})^2$ and $K_{\theta}\theta^2$ to the data points, where $K_{\mathrm{t}} = 0.111462$ eV/Å and $K_{\theta} = 0.015833$ eV/deg..

To imitate the breakage of DNA segments, we cut the chain of coarse-grained particles at randomly selected position *p* at randomly determined timing in MD simulation by artificially replacing the force acting on (*p*-1)-th, *p*-th, (*p*+1)-th and (*p*+2)-th particles by following equations.

$$F_{p-1} = -\nabla_{p-1} U_{p-1,p}^{\mathrm{t}} - \nabla_{p-1} U_{p-2,p-1}^{\mathrm{t}} - \nabla_{p-1} U_{p-2,p-1,p}^{\theta} - \nabla_{p-1} U_{p-3,p-2,p-1}^{\theta} \quad (3)$$

$$F_p = -\nabla_p U_{p-1,p}^{\mathrm{t}} - \nabla_p U_{p-2,p-1,p}^{\theta} \quad (4)$$

$$F_{p+1} = -\nabla_{p+1} U_{p+1,p+2}^{\mathrm{t}} - \nabla_{p+1} U_{p+1,p+2,p+3}^{\theta} \quad (5)$$

$$F_{p+2} = -\nabla_{p+2} U_{p+2,p+3}^{\mathrm{t}} - \nabla_{p+2} U_{p+1,p+2}^{\mathrm{t}} - \nabla_{p+2} U_{p+2,p+3,p+4}^{\theta} - \nabla_{p+2} U_{p+1,p+2,p+3}^{\theta} \quad (6)$$

The MD simulation is performed with time step of 1 ps by 2nd order of symplectic integration method [6]. The initial temperature is set to 300 K. The calculation is performed in *NVE* ensemble. We note that the time scale of the MD simulation is completely different from the frame rate of actual video 30 frames/s. Even though the difference, the DNA motion in MD simulation is similar to the motion in the actual video at least in visual observation, and we think it is applicable to the purpose of generating dummy videos for machine learning.

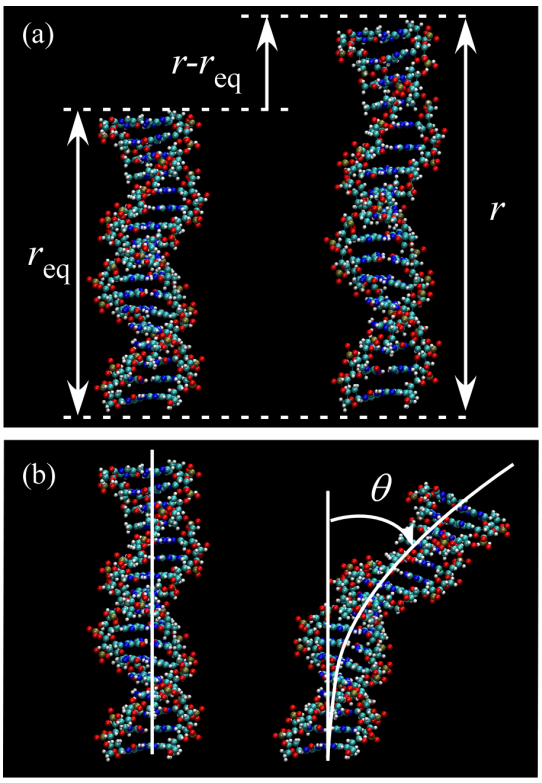


Figure 6: (a) Expansion and (b) bending of the atomic model of DNA for obtain the parameter K_t and K_θ in potential function U.

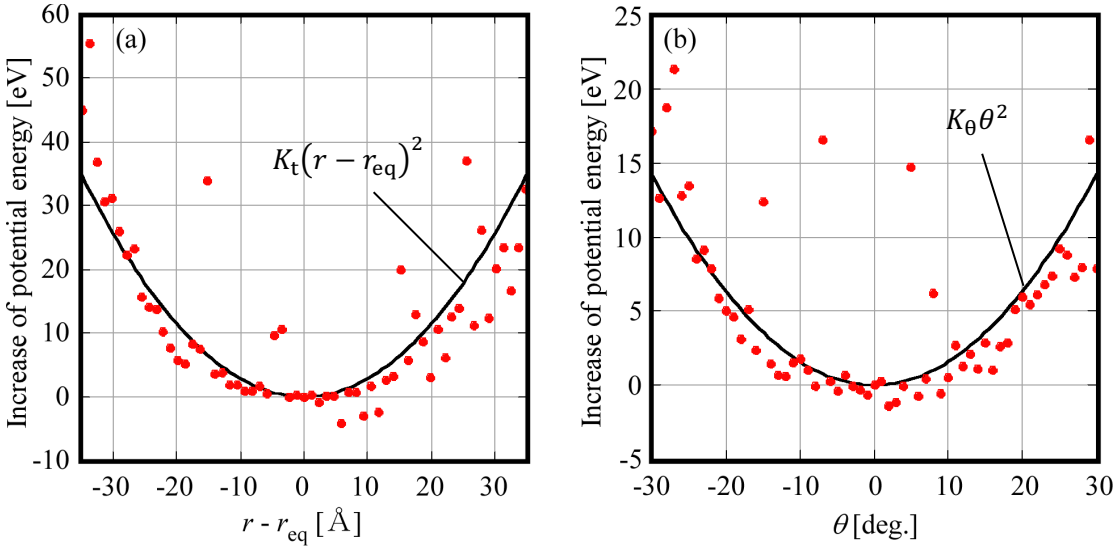


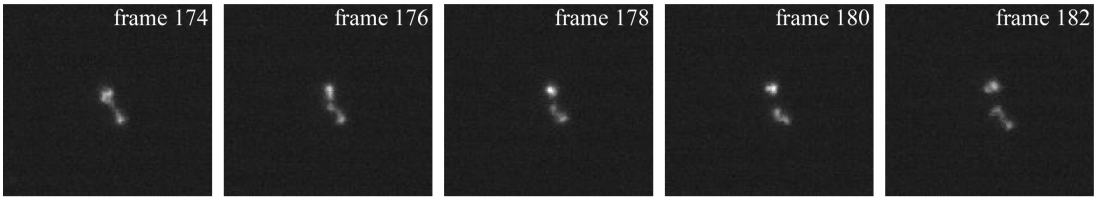
Figure 7: Fitting for the parameter (a) K_t in $U^t_{i,j}$ and (b) and K_θ in $U^\theta_{i,j,k}$. The dots denote the increase of the potential energy when it's expanded and bended as shown in Fig. 6. Solid lines show the fitting curve.

The dummy video is generated by following steps.

1) MD simulation is performed for 300 ns to calculate the motion of coarse-grained particles of DNA. Then, the time evolution of the 3 dimensional coordinates at every 0.1 ns are stored.
2) Sored time-series of 3 dimensional coordinates data are projected to 2 dimensional randomly selected plan.
3) The frames of dummy video is generated by drawing spline curves which pass the 2 dimensional coordinates data at each time.
4) The spline curves drawn on each frames are artificially blurred to imitate the actual video.

Figure 8 (b) shows an example of dummy video generated by MD simulation when a DNA segment breaks. Figure 8 (a) shows the actual video captured by fluorescence microscope for comparison.

(a) video captured by fluorescence microscope



(b) dummy video generated by MD simulation

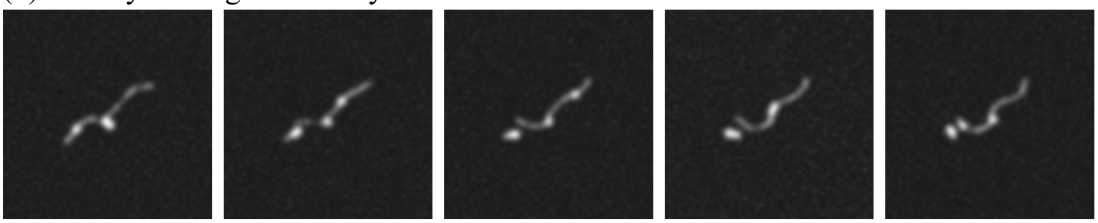


Figure 8: (a) Animation of actual DNA image taken by fluorescence microscope.
(b) Images generated by MD simulation for training.

4. Results of prediction

Pix2pix is trained with 100 sets of 5 frames generated by the MD simulation. We also trained pix2pix by single frame input case for comparison. Figure 9 shows the beginning (from 6 to 22 frames) of the actual video captured by fluorescence microscope with the corresponding frames obtained by the prediction of trained pix2pix network. In the range of frames shown in Fig. 9, the DNA has not yet broken. Figure 9 (b) shows the predictions by single frame input, and Fig. 9 (c) shows that predicted by multiple (5) frames input. Comparing the single and multiple input case, both network predict similar results. However, as shown in Fig. 10, comparing the results in frame 10 in detail, single frame input case (Fig. 10 (b)) has multiple small DNA segments, while multiple frames input case (Fig. 10 (c)) has a single segment.

The number and length of DNA segments are measured automatically by an image processing method using skeletonization [7] explained in our previous work [1]. Figure 11 shows the number of DNA segments after skeletonization with some noise reduction. The results of multiple frames input case are shown by red circles, and that of single frame input case are shown in

blue filled circles. DNA segments actually breaks at frame 177. The results show that both case sometimes overcount, but multiple frames input case is less likely to overcount.

In the future, we would like to measure the dynamic changes in the number of cleavages from movies with multiple DNA segments by the method explained in our previous research [1]. For this purpose, it is necessary to measure the length of DNA. Figure 12 shows the time evolution of length of DNA segments predicted by multiple frames input case. Red and blue filled circles denote the length of 1st and 2nd segments. The 2nd segment appears at frame 177 when DNA breaks. Green and light blue triangles show the length of 3rd and 4th segment which is generated by miss judgement. The black circle denote the total length of DNA segments. Of course, the actual total length of DNA segments should be constant, but because the DNA moves in three dimensions, the length changes in the 2 dimensional images captured by the camera.

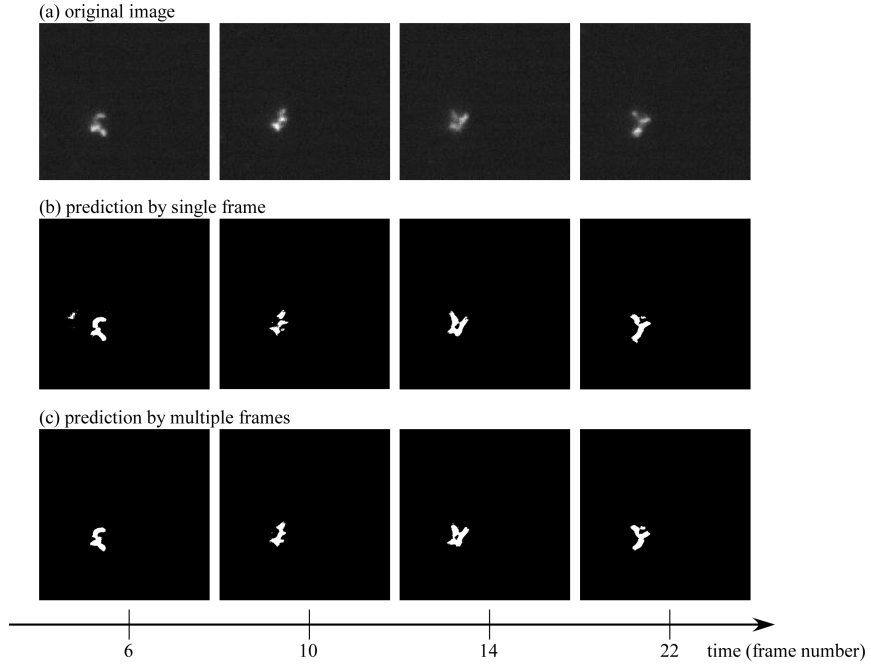


Figure 9: (a) DNA motion of original video captured by fluorescence microscope. (b) and (c) Predicted images from the original video by pix2pix with (b) single and (c) 5 frames input.

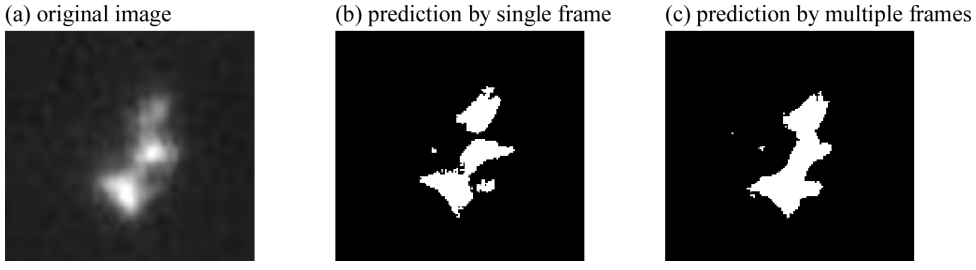


Figure 10: (a) original image of frame 10 captured by fluorescence microscope. Comparison of the predicted images by (b) single and (c) 5 frames input.

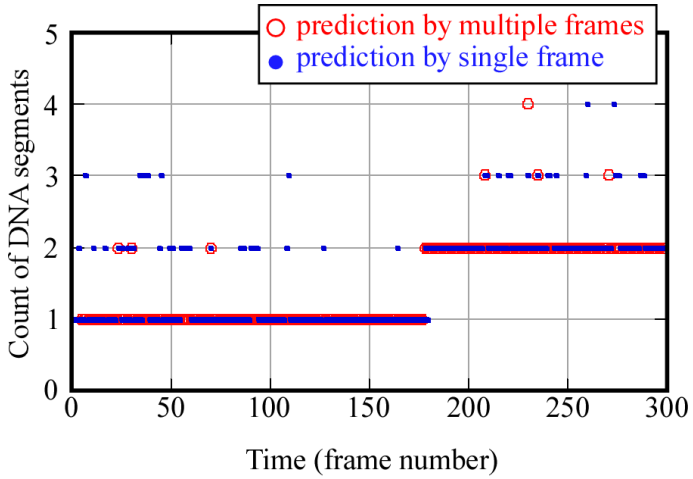


Figure 11: Comparison of the count of DNA segments from the predicted images by single and 5 frames input.

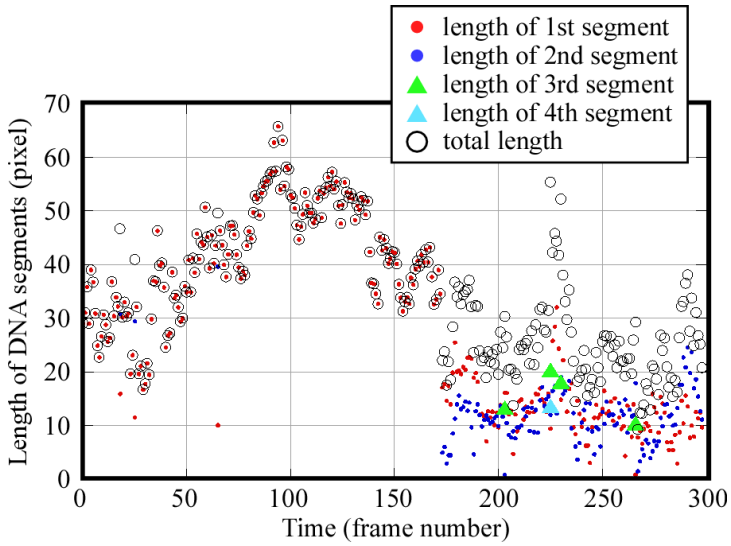


Figure 12: Length of DNA segments which is automatically measured by an image processing method explained in Ref. [1].

5. Summary

To automatically measure the DNA length in a video captured by fluorescence microscope, pix2pix network model is employed. In order to accurately measure the length and number of DNA in low-resolution images, we proposed a method of inputting multiple time-series frames to pix2pix. To generate the training data, MD simulation is used to prepare dummy video. The trained pix2pix by multiple frames input can generate more accurate images in a scene of DNA breakage than the pix2pix trained by single frame input.

Acknowledgement

The research was partially supported by Grant-in-Aid for Scientific Research, No. 22K03572 from the Japan Society for the Promotion of Science, and by the NIFS Col-laborative Research Program NIFS22KIGS002 and NIFS22KISS021. The computations were performed using the JFRS-1 supercomputer system at Computational Simulation Centre of International Fusion Energy Research Centre (IFERC-CSC) in Rokkasho Fusion Institute of QST (Aomori, Japan), Research Center for Computational Science (Okazaki, Aichi, Japan) and Plasma Simulator of NIFS (Toki, Gifu, Japan) .

References

[1] S. Saito, H. Nakamura, T. Kenmotsu, Y. Oya, Y. Hatano, Y. Tamura, S. Fujiwara, H. Ohtani: *Image processing method for automatic measurement of number of DNA breaks*, J. Adv. Simulat. Sci. Eng. 8:2 (2021), 173.

[2] P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros: *Image-to-Image Translation with Conditional Adversarial Networks*, arXiv preprint arXiv: 1611.07004 (2016).

[3] H. Nakamura, H. Miyanishi, T. Yasunaga, S. Fujiwara, T. Mizuguchi , A. Nakata, T. Miyazaki, T. Otsuka, T. Kenmotsu, Y. Hatano, S. Saito: *Molecular dynamics study on DNA damage by tritium disintegration*, Jpn. J. Appl. Phys., 59 (2020), SAAE01.

[4] K. Chenoweth, A. C. T. van Duin, W. A. Goddard: *ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation*, J. Phys. Chem. A, 112-5 (2008), 1040/1053.

[5] C. T. Adri, D. Siddharth, L. Francois and G. A. William: *ReaxFF: a reactive force field for hydrocarbons*, J. Phys. Chem. A, 105 (2001), 9396.

[6] M. Suzuki: *Decomposition formulas of exponential operators and Lie exponentials with some applications to quantum mechanics and statistical physics*, J. Math. Phys., 26 (1985), 601.

[7] T. Y. Zhang, C. Y. Suen: *A Fast Parallel Algorithm for Thinning Digital Patterns*, Communications of the ACM, 27:3 (1984), 236.