Proceedings of the 3rd International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/23/20231050

Deep learning based denoising of super-resolution images of protein aggregates

Zhongsheng Cheng

Cranbrook Schools, 39221 Woodward Ave, Bloomfield Hills, MI 48304, United States

v.z.cheng@hotmail.com

Abstract. Neurodegenerative diseases (NDs) are closely associated with the amyloid aggregation of proteins like Amyloid-β, α-synuclein, and tau. Understanding the pathogenesis of NDs requires studying the structures and morphological features of these aggregates, which are typically below 100 nm in size. Traditional fluorescence microscopy is limited by the diffraction limit of light (~250 nm). Single-molecule localization microscopy (SMLM) offers a resolution down to ~20 nm, enabling the visualization of these aggregates. However, issues such as non-specific binding (NSB) of fluorophores and background noise degrade the quality of SMLM images. This study presents a U-net-based convolutional neural network (CNN) to denoise SMLM images of protein aggregates. The training dataset includes noise-free super-resolution images of aggregates and their noisy counterparts with non-specific binding signals. Various imaging conditions are simulated to mimic real-world scenarios. The U-net's output is evaluated against ground-truth images for denoising performance. Post-processing techniques further enhance denoised images. The fine-tuned U-net model achieves a validation loss of 0.0042 and low prediction errors of 0.32% and 3.71% in the area and number of aggregates, respectively. This research offers a powerful tool for denoising SMLM images, facilitating accurate characterization of protein aggregate structures and morphological features.

Keywords: neurodegenerative diseases, single-molecule localization microscopy, deep learning, image denoising, convolutional neural network.

1. Introduction

Neurodegenerative diseases (NDs) are a leading cause of morbidity and cognitive impairment in older adults. Among different types of NDs, Alzheimer's Disease (AD) is the most prevalent, accounting for 60%–80% of all dementia cases and affecting approximately 24 million people worldwide [1-3]. Evidence suggests that protein aggregation, during which misfolded/unfolded proteins form aggregates, is responsible for the pathogenesis of NDs. ND-related aggregate-forming proteins include amyloid-β (Aβ), α-synuclein (α-Syn), and tau [4, 5], which typically form amyloid fibrils manifesting as insoluble plaques in the brain tissue with diameters at the micrometer scale. The morphology of Aβ aggregates is shown in Figure 1 [6]. However, recent research reveals that amyloid aggregates are a complex and dynamic population composed of small soluble oligomers, protofibrils, and mature fibrils [7]. These species have distinct biological and pathological functions closely related to ND development, and their sizes can vary drastically from the nanometer to the micrometer level. To thoroughly understand the

DOI: 10.54254/2753-8818/23/20231050

pathogenesis of NDs, it is crucial to statistically characterize the number, size, and morphological features of the aggregates.

Fluorescence microscopy is a powerful tool for visualizing biological samples with remarkable molecular specificity [8]. Fluorescence is a phenomenon that a ground-state molecule absorbs a photon to enter the excited state and releases a fluorescent photon to fall back to its ground state. The released fluorescent photon has a slightly lower energy than the absorbed photon, meaning the fluorescent light has a longer wavelength than the excitation light. In a fluorescence microscope, the molecules of interest in the specimen are labelled with fluorescent dyes that are excited with a laser. The fluorescent light emitted from the dye molecules is then collected and imaged on a camera to indicate the position and morphological information of the molecule of interest. Fluorescence microscopy is suitable for visualizing fluorescently labelled protein aggregates; however, all optical systems are limited by the Abbe diffraction limit of light, i.e., two features are indistinguishable when their distance is less than a half of the wavelength of the light used to image the sample. The diffraction limit of light prohibits the fluorescence microscopes from visualizing protein aggregates with sizes under ~200 nm. Therefore, small aggregates, such as oligomers with sizes under 100 nm, cannot be characterized by typical diffraction-limited fluorescence imaging.

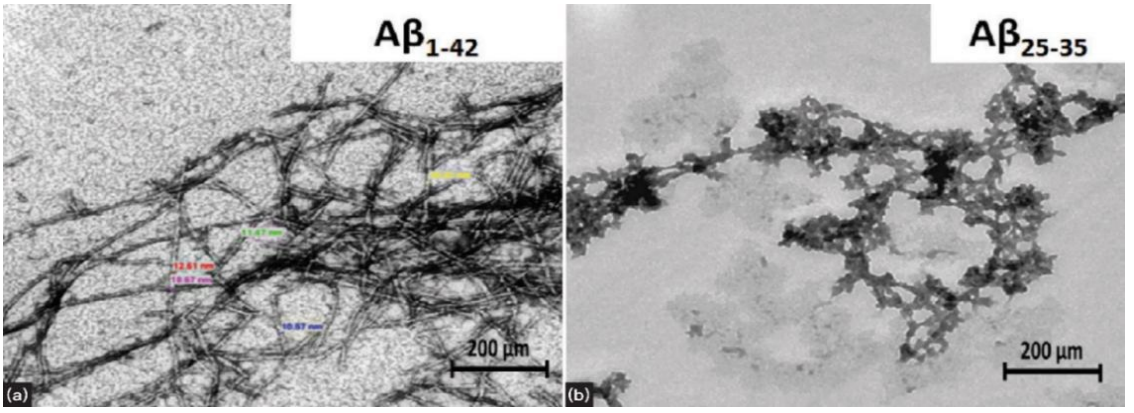

Figure 1. Morphological appearance of amyloid beta (Aβ) fibers under transmission electron microscope. (a) aggregates formed from $A\beta_{1-42}$ demonstrating a "strained ribbon morphology." (b) $A\beta_{25-35}$ peptide form tiny and short Aβ aggregates.

Super-resolution (SR) microscopy is a series of cutting-edge techniques to surpass the diffraction limit of light. Single-molecule localization microscopy (SMLM) is a representative SR technique that offers a lateral resolution of about 20 nm [8], allowing for the visualization of such small aggregates. SMLM achieves SR imaging by sparse activation of the fluorophores. A graphic overview of SMLM is illustrated in Figure 2 [9]. In each frame of the image, only a small fraction of fluorophores are randomly activated, while others remain dark [10]. The non-overlapping images of individual fluorophores are processed with Gaussian fitting algorithms to locate the centers of each bright spot, i.e., the probable positions of the fluorescent molecules. The SR image of the full field-of-view (FOV) is then reconstructed from thousands of frames [9]. DNA-PAINT is a typical SMLM method that utilizes the well-understood, robust, and specific binding between complementary DNA strands to locate target molecules [12]. The mechanism of DNA-PAINT is shown in Figure 3 [13]. During the imaging session, target molecules are labeled antibodies conjugated with short single-stranded DNA, i.e., "docking strands". The transient binding of fluorophore-labeled complementary DNA strands, known as the "imager strands", to these docking strands allows for sparse blinking. The fluorescence emission is subsequently captured by a scientific camera, which records the point spread functions (PSFs) of activated fluorophores, i.e., imager strands that bind to the docking strands, in the resulting image.

DOI: 10.54254/2753-8818/23/20231050

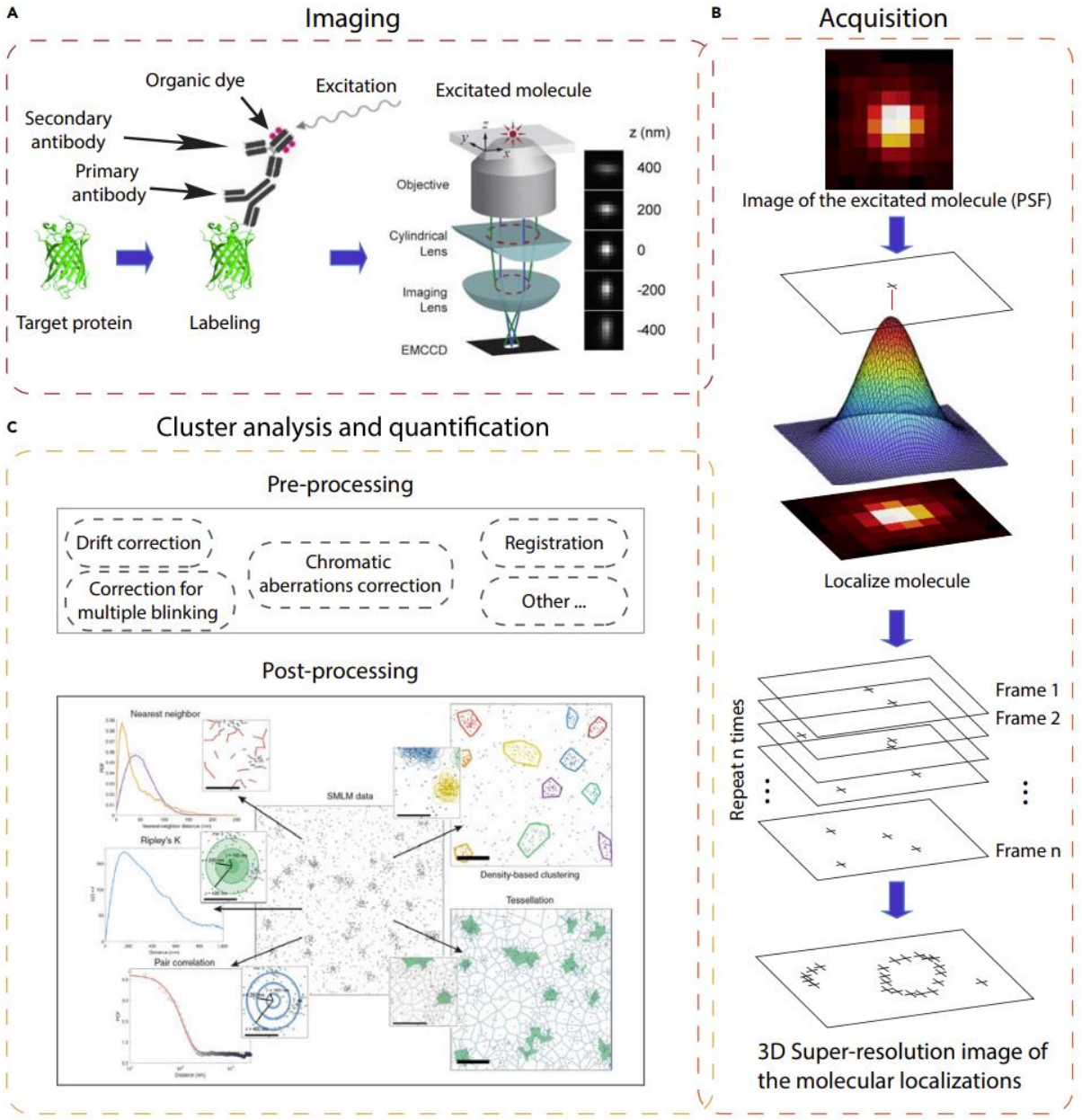

Figure 2. Overview of SMLM Quantification Principles. (A) SMLM imaging of target proteins. (B) acquiring the protein localizations and getting a map for the molecular coordinates, and (C) analyzing the super-resolved image to quantify the SMLM clusters.

DOI: 10.54254/2753-8818/23/20231050

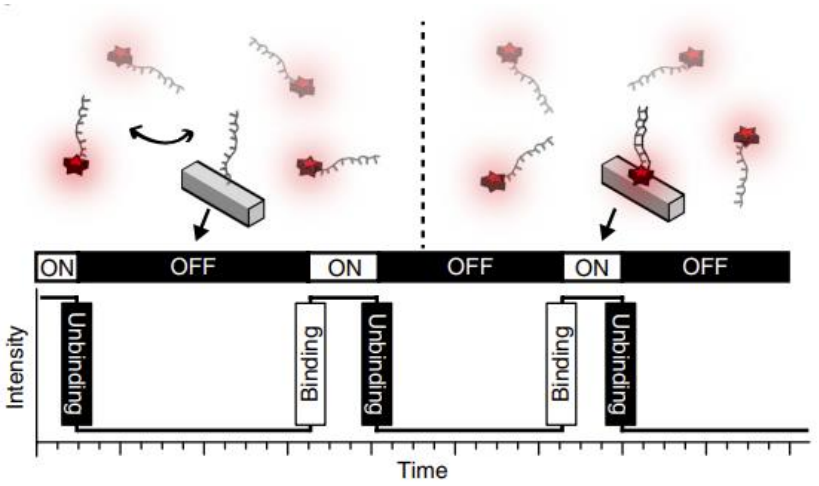

Figure 3. The concept of DNA PAINT. Transient binding of dye-labeled DNA strands (imagers) to their complementary target sequence (docking site) attached to a molecule of interest. The transient binding of imager strands is detected as 'blinking', illustrated by the intensity versus time trace.

The quantitative analysis of SR images of protein aggregates is often hindered by noises introduced by various sources. In SMLM, background noises, including autofluorescence, out-of-focus fluorescence, and camera noises may lead to additional random localizations in the reconstructed SR, which induce difficulties in the following cluster analysis procedure. In addition, non-specific bindings (NSBs) of fluorophores to non-target proteins are inevitable in SMLM techniques, such as DNA-PAINT. In other words, the shape, area, and structure of the protein aggregates may be disturbed by noises. In 2022, researchers developed MSDenoiser that utilized various denoising algorisms comprehensively, including Voronoi Tessellation, Local Outlier Factor (LOF) and density-based spatial clustering of applications with noise (DBSCAN), to remove various types of noise, including free non-polymer localization points, non-polymer localization points near the sample signal point area, and NSBs. MSDenoiser integrates the advantages of the aforementioned algorithms but fails to evaluate the noise reduction in areas close to the target sites.

The emergence of artificial intelligence has provided innovative solutions for SMLM data denoising. The artificial neural network (ANN) is a central technique in artificial intelligence. In an ANN, numerous nodes (known as neurons) are organized into multiple layers (depth) [14]. Like a biological neural network, each neuron receives information from others and the final outputs are evaluated. The network attempts to maximize the correctness based on a given reference by manipulating the weighting assigned to each neuron, calculated through error estimation during each forward propagation [14-16]. Convolutional neural network (CNN) is a special type of ANN designed for processing 2D data, such as images. Instead of single 1D neurons, CNN consists of a series of convolutional layers that process the input data by 2D convolution and enable feature extraction and classification [17]. In a convolutional layer, a matrix representing the image features is taken as input. The layer performs a dot product operation between the input and a filter (or a set of filters known as kernels). The kernel size, typically expressed as the kernel's dimensions, decides the number of pixels processed together. Each kernel captures one specific type of feature in the image, such as a blob or an edge. After convolution, an activation function is used to decide whether a neuron should be activated or not to introduce non-linearity into the output of a neuron. Typical activation functions include Sigmoid, Tanh, and Rectified Linear Unit (ReLU) [18]. The ReLU function is defined as in Eq. 1. This function transforms the output of a neuron by mapping it to the highest positive value. If the output is negative, the function maps it to zero.

$$f(x)=max(0,x)=\begin{cases}x, & x>0\\0, & x\leq 0\end{cases} \quad (1)$$

DOI: 10.54254/2753-8818/23/20231050

A pooling layer following a convolutional layer condenses the output of small regions of neurons into a single output to reduce the dimensions of data. There are several non-linear functions to implement pooling, such as max pooling and average pooling. After several convolution and pooling layers, the extracted feature maps are then passed through one or more fully connected layers before producing output. The weight of each neuron is tuned during the training process to optimize the output and realize the designated task of the CNN, such as prediction or classification. The architecture of a CNN model is schematically illustrated in Figure 4.

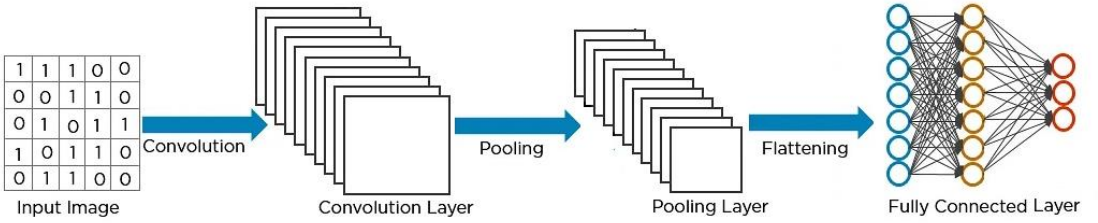

Figure 4. Architecture of a typical CNN Model.

Among the widely used CNN architectures, U-net is one of the most popular choices for image denoising and super-resolution applications [19-21]. U-net employs two symmetric sets of convolutional blocks to achieve image compression, feature extraction, and recovery. Through training with pairs of noisy and noise-free images, U-net can be used to remove the noise in the images through extensive training [22, 23].

In this work, we present a U-net trained for denoising SMLM images of protein aggregates. The training dataset includes simulated noise-free SR images of aggregates, including randomly generated dots, lines and curves, and their corresponding noisy images with NSB signals. The output images from the U-net were evaluated against simulated ground-truth images to evaluate the denoising performance of the model. Post-processing techniques such as erosion and filtering were applied to further enhance the quality of denoised images. The U-net's hyperparameters were fine-tuned, and a final model using ReLU activation function and max pooling was trained with datasets from all imaging conditions.

2. Method

2.1. Simulation of SMLM images

2.1.1. SMLM image sets. The training data consisted of simulated SMLM images. Each set of SMLM images contained a ground truth super-resolution (GTSR) image, a corresponding diffraction-limited (DL) image stack, and a noisy SR image reconstructed from the DL stack (Figure 5). GTSR images represent the ideal SMLM images that would be obtained if there were no noises during the imaging experiment. GTSR images provide an accurate representation of the true shape, size, and distribution of the protein aggregates. The corresponding DL image stacks simulate the actual experimental data that is obtained from a fluorescence microscope, where noises from both camera and non-specific binding were added. The noisy SR images were generated by performing Gaussian fitting on the individual localizations detected in the DL stack.

2.1.2. Simulation of GTSR images. Each GTSR image measured 2560 × 2560 pixels and contained a total of 350 simulated aggregates that mimicked fibrillary and oligomeric structures commonly observed in research scenarios. Specifically, the images included 100 straight lines, 100 curves, and 150 dot aggregates. To simulate real-world variations, the size of the simulated protein aggregates followed a normal distribution with a standard deviation (std) of 0.2. The straight and curve aggregates were simulated with mean lengths of 40 and 80 pixels, respectively, while the dots had a mean radius of 10

DOI: 10.54254/2753-8818/23/20231050

pixels. When generating the curved aggregates, an additional parameter of the central angle was considered. The central angle was randomly assigned a value between 10 and 120 degrees.

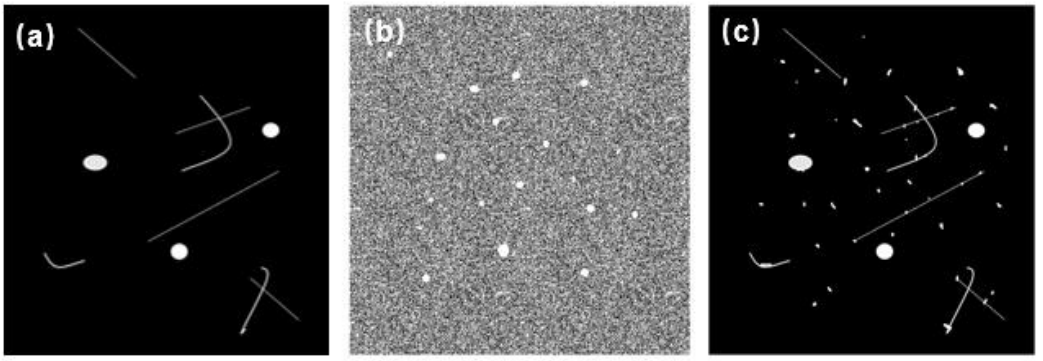

Figure 5. An illustration of one SMLM image set. (a) The GTSR image. (b) One frame from the corresponding DL image stack. (c) The reconstructed noisy SR image.

Figure 6 illustrates the procedure for simulating GTSR images. Initially, an empty canvas was created, and small images of the straight, curve, and dot aggregates were generated. These individual images of aggregates were then added onto a black canvas with random orientations and positions. The pixel value of the black canvas was set to 0, while that for the aggregates was set to 1. In cases where multiple aggregates overlapped at the same position, the corresponding pixel value was multiplied.

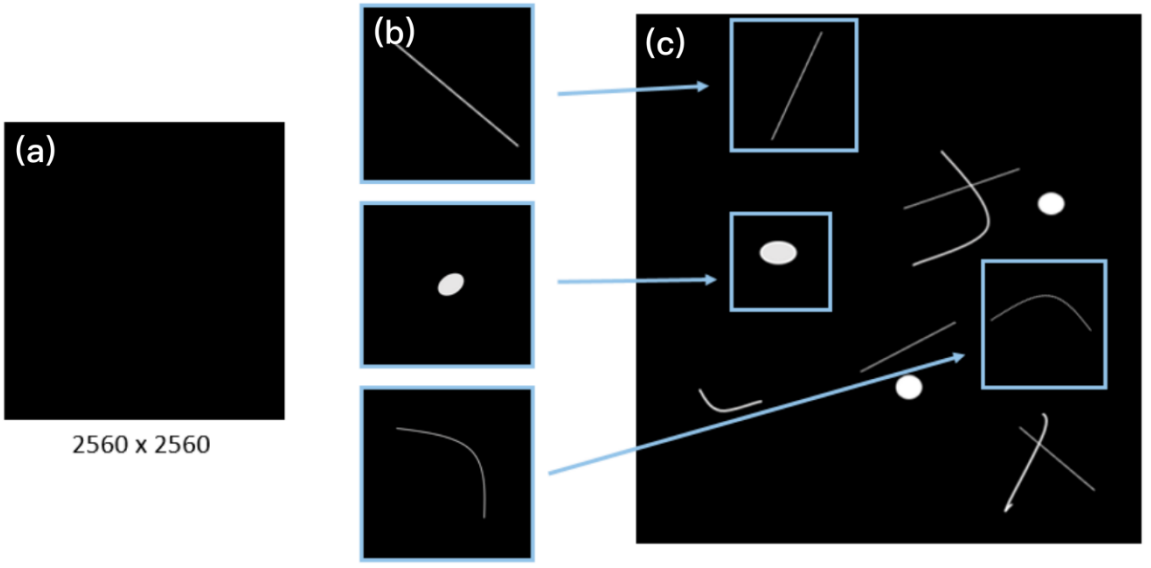

Figure 6. Simulation procedure of the GTSR images. (a) The black canvas. (b) Individual images of the straight, curve and dot aggregates. (c) A GTSR image generated by adding individual images to the canvas.

2.1.3. Simulation of DL image stacks. The simulation of DL stacks involved reversing the process of SR image reconstruction from the imaging data. Specifically, each GTSR image was used to generate a corresponding DL image stack. In practice, a magnification of 10x is commonly used to increase the localization precision during the SR image reconstruction process from the DL stack. As a result, each simulated DL stack was reduced to 256 × 256 pixels in size, which is one-tenth of the size of the GTSR image.

In GTSR images, pixel with a non-zero value represented a binding site of a fluorophore, which then corresponds to a 2D Gaussian localization in the DL stack. Each frame of the DL stack contained 50

DOI: 10.54254/2753-8818/23/20231050

true localizations and an average of 10 non-specific bindings (NSBs). The number of NSBs followed a normal distribution with a standard deviation of 0.2. To account for the potential mismatch between the blinking interval and the exposure time of the camera, 20% of the localizations in one frame were inherited to the next frame. The full-width-half-maximum (FWHM) of the 2D Gaussian localizations was set to 3 pixels. Given that the pixel size was set to 100 nm, the resulting simulated localizations were approximately 300 nm in size, matching the real-world diffraction limit of light. The intensity of all localizations also followed a normal distribution with a mean intensity of 4000 a.u. and a std of 0.2. For overlapping localizations, the pixel values at the overlapping areas were set to the sum of individual localizations.

The simulation procedure for DL stacks is illustrated in Figure 7. 2D Gaussian localizations corresponding to the binding sites in the GTSR image were generated and added to the DL frame. Additionally, background noise from the camera, which follows a Gaussian distribution, was introduced to each frame. A total of N frames were generated, which together formed a DL image stack.

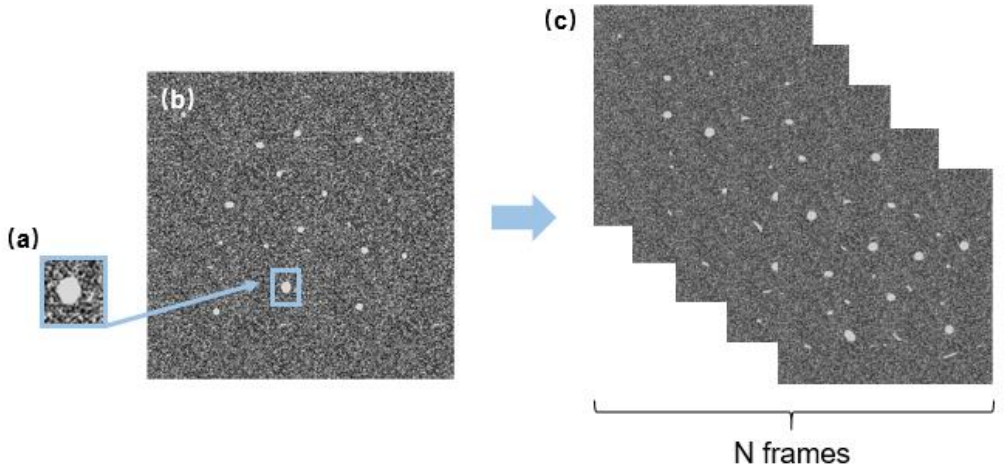

Figure 7. Simulation procedure of the DL image stacks. (a) A 2D Gaussian localization. (b) Simulation of one DL frame. (c) A DL image stack generated by combining multiple DL frames.

2.1.4. Generating noisy SR images. FIJI is a widely-used open-source software that offers a range of tools for image analysis. ThunderSTORM is one of its powerful plugins that utilizes a Gaussian fitting algorithm to reconstruct SR images from DL data. In this study, the noisy SR images were reconstructed from the DL image stacks through the Gaussian fit function in ThunderSTORM. For both the true localizations and the NSBs in all frames of the DL stack, the x and y coordinates were calculated and recorded, resulting in the generation of noisy SR images with a size of 2560 × 2560 pixels, which is the same as the size of the GTSR images (Figure 8).

To explore the denosing capability of the CNN on images obtained from different imaging conditions, DL images of different signal-to-noise ratios and numbers of frames were generated, followed by the reconstruction of their corresponding noisy SR images. The signal-to-noise ratios ranged from 1.5 to 4.5 with a fixed number of frames of 2000. The number of frames ranged from 500 to 4000 with a fixed signal-to-noise ratio of 4.5.

DOI: 10.54254/2753-8818/23/20231050

Figure 8. Illustration of a noisy SR image.

2.2. Constructing the CNN

In this work, a CNN model based on the U-net architecture was employed as the denoising neural network. The structure of the U-net is illustrated in Figure 9. In this U-net structure, both the max-pooling and up-sampling sections of the initial model included three layers, with 32, 16, and 16 neurons in each layer, respectively. These three layers correspondingly processed images of sizes 128×128, 64×64, and 32×32 pixels. The bottleneck contained 16 neurons and processed images of 16×16 pixels. The activation function was set to ReLU as defined in Equ. 1. The loss of the U-net was calculated using binary entropy.

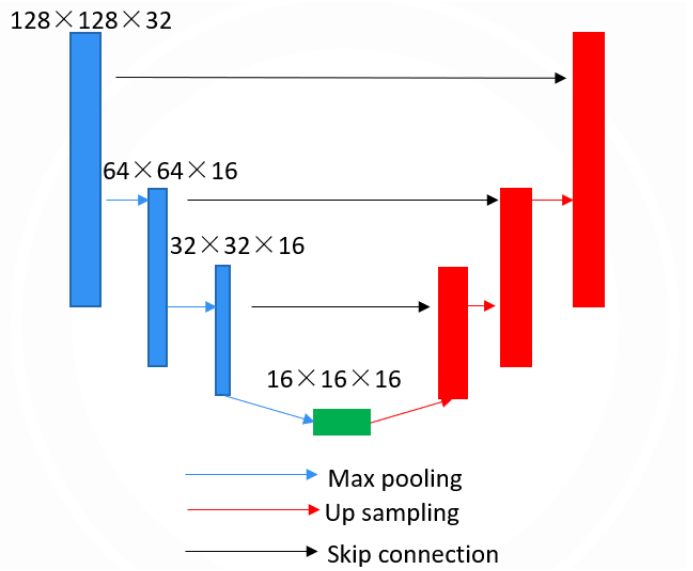

Figure 9. Schematic of the initial U-net model with default parameters.

2.3. Post processing

Erosion and filtering were used to process the denoised images from the U-net to further enhance the denoising effect and to remove the additional artefacts created by the U-net. The erosion algorithm trimmed the outskirt of each aggregate by one pixel. Filtering removed all aggregates smaller than 10 pixels in area.

DOI: 10.54254/2753-8818/23/20231050

2.4. Quantitative analysis

The number, average area, and area distribution of the aggregates were calculated from the GTSR and denoised SR images using the regionprops algorithm. The discrepancies in these data (i.e., the number, average area, and area distribution) between the two groups were then determined by the Kolmogorov–Smirnov Test (KS Test) to evaluate the quality of the denoising process. The D-value of the KS Test was given by Eqs. 2 and formula 3. It could be observed that larger D-values represent greater discrepancies between the GTSR and the denoised SR images, i.e., poor denoising quality, and vice versa. Meanwhile, P-values indicated the significance of such difference. P-values below 0.05 were considered statistically significant.

$$D_{n,m} = sup_x \left| F_{1,n}(x) - F_{2,m}(x) \right| \tag{2}$$

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n*m}} \; where \, c(\alpha) = \sqrt{-ln\left(\frac{\alpha}{2}\right) * \frac{1}{2}} \tag{3}$$

2.5. Training the CNN

2.5.1. Input and output images. The simulated noisy SR images served as the input of the U-net. For each set of simulation parameters, a total of 30 images were generated and randomly divided into the training dataset (20 images), the validation dataset (5 images), and the test dataset (5 images). The simulated 2560×2560 images were segmented into small frames of 128×128 pixels, which were used as individual images fed into the U-net. The aim of the segmenting step was to reduce the total memory usage while preserving the morphological features of individual aggregates. The denoised images produced by the U-net were post-processed or not post-processed with erosion and filtering and were subsequently combined together to restore whole output images of 2560×2560 pixels.

2.5.2. Evaluating the performance of U-net on various imaging conditions. The performance of the U-net with default hyperparameters was examined using datasets of images with various imaging conditions, including signal-to-noise ratio and number of frames. The training on each dataset lasted 15 epochs. The D-values and the P-values were calculated by comparing the aggregate size distribution between the post-processed or not post-processed denoised SR and their corresponding GTSR images. The optimized imaging condition was determined as the one yielding the smallest D-value and the largest P-value. The dataset with the optimized imaging condition was used as a standard dataset for hyperparameter optimization of the U-net.

2.5.3. Optimizing hyperparameters. Grid-search was applied to find the optimal set of hyperparameters of the U-net, including number of neurons, kernel size, and activation function. A total of 18 different combinations of the three hyperparameters were explored, as shown in the table below. For each set of hyperparameters, 15 epochs of training were performed on the image dataset with the optimized imaging conditions, and the training time and validation loss of the model were recorded. The optimized model was determined as the one with the smallest validation loss.

Table 1. Summary of parameters used for grid search.

No. of neurons	16	32	64
Kernel Size	3×3	4×4	5×5
Activation Function	ReLU	Tanh	

2.5.4. Training the final model. After the optimal set of hyperparameter was determined, a comprehensive final model was trained using all noisy SR images. The training lasted 30 epochs. The

DOI: 10.54254/2753-8818/23/20231050

output images of the final model were either post-processed or not post-processed and compared with their corresponding GTSR images for the calculation of D-values and P-values using the KS Test.

3. Results

3.1. Simulation of GTSR, DL stacks, and noisy SR images

As shown in Figure 10, each SMLM dataset consists of a GTSR image, the corresponding DL image stack, and the final reconstructed noisy SR image. 30 images were generated for each of the 13 different SNR and the number of frame combinations, yielding a total of 390 GTSR images.

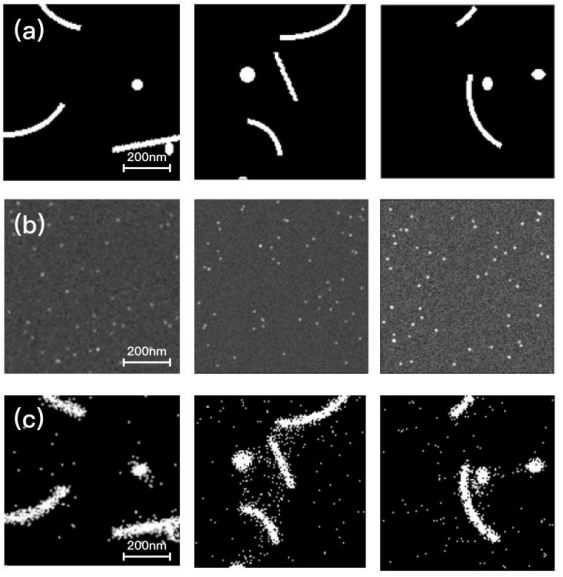

Figure 10. Example simulation results. (a) A cropped section of a simulated GTSR image. (b) A single frame from the DL stack generated from the full image of (a). (c) The noisy SR image reconstructed from (b).

3.2. Image post-processing

As shown in Figure 11, by comparing the denoising output of the U-net model with the original GTSR image, it was observed that the aggregates of the denoised images generated by the CNN model are larger than those in the original image. A direct comparison by subtracting the GTSR image from the denoised SR image revealed that the aggregates are approximately one to two pixels wider in circumference. To address this issue, an erosion of one pixel was applied to remove the aforementioned error.

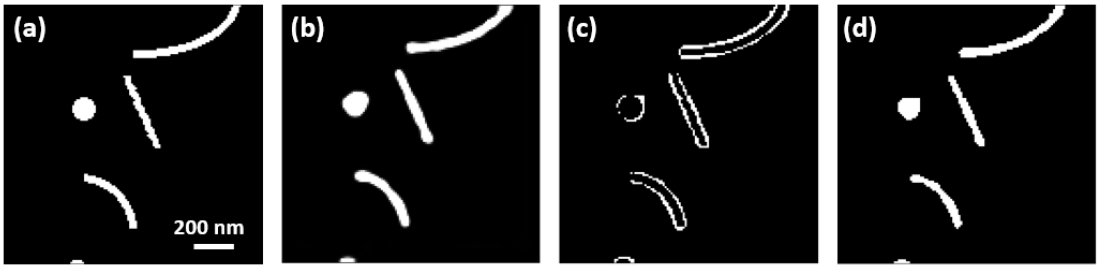

Figure 11. Justification of the use of erosion. (a) A cropped section of a simulated GTSR image. (b) The corresponding denoised SR image generated by the U-net model. (c) The result of subtracting (a) from (b), highlighting an error in the denoised SR image where the aggregates are wider than those in the GTSR image. (d) Denoised SR with erosion.

DOI: 10.54254/2753-8818/23/20231050

Furthermore, U-net occasionally introduced undesired noises onto the denoised SR image (Figure 12). However, the size of these additional artifacts is generally smaller than 10 pixels and their brightness is lower than that of the predicted aggregates. Therefore, a size filter of 10 pixel was adopted and this inaccuracy was successfully removed.

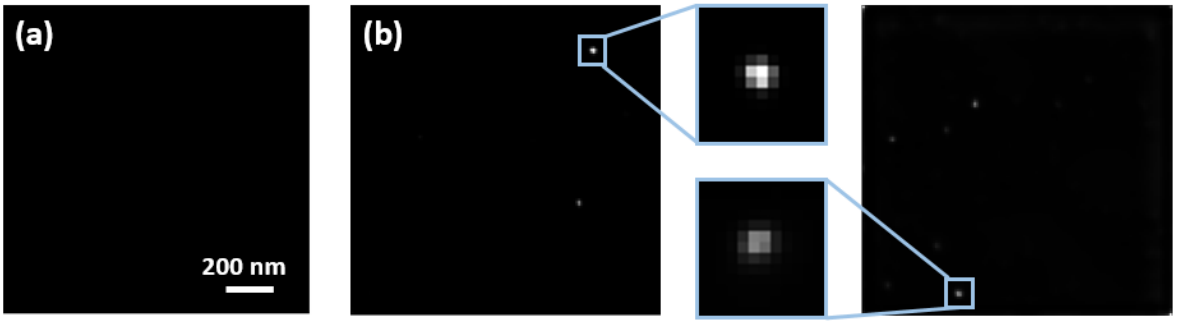

Figure 12. Justification of the use of a size filter. (a) A cropped section of a simulated GTSR image. (b) The corresponding denoised SR image generated by the U-net model. Small artifacts were introduced by the U-net model. Enlarged views highlight the additional noises.

3.3. Optimizing imaging conditions

3.3.1. Optimizing the SNR. Figure 13 provides insights into the performance of the denoising model. By comparing the average aggregate area obtained from the Regionprops, it was shown that erosion successfully removed the enlargement of the aggregates from the denoised images as the average aggregate area of denoised SR images processed by erosion was largely consistent with that of GTSR regardless of SNR ratio, as shown in Figure 13 (a). However, this increase in accuracy was achieved in the expense of the total number of aggregates. As shown in Figure 13 (b), the total number of aggregates was reduced after erosion was implemented to process the denoised images.

Figure 13 (c) shows the results of the KS test on the area distributions, which suggested that as SNR increased from 1.5 to 2.5, the calculated P-value increased significantly from a value slightly above 0 to approximately 0.45. Then the P-values experienced a slight decrease when SNR ranged between 2.5 and 3.5 before peaking at 0.575 when it equaled 4.5. Notably, erosion processing helped stabilized P-values at approximately 0 for all SNRs.

Figure 13 (d) shows the plot of the D-values from the same KS test against the SNRs. Most D-values of the denoised SR images were within the range of 0.3 to 0.4, but the smallest D-value, which was slightly below 0.1, occurred when the SNR was 2.0. The critical value was calculated to be 0.15 according to equation (3). Erosion helped reduce the D-values significantly and stabilized the results below the critical value. The lowest D-value for the denoised SR with erosion occurred when SNR was also 4.5.

These results indicate that the U-net model and the image post-processing methods effectively restored the quality of the images. The optimum SNR was determined as 4.5.

DOI: 10.54254/2753-8818/23/20231050

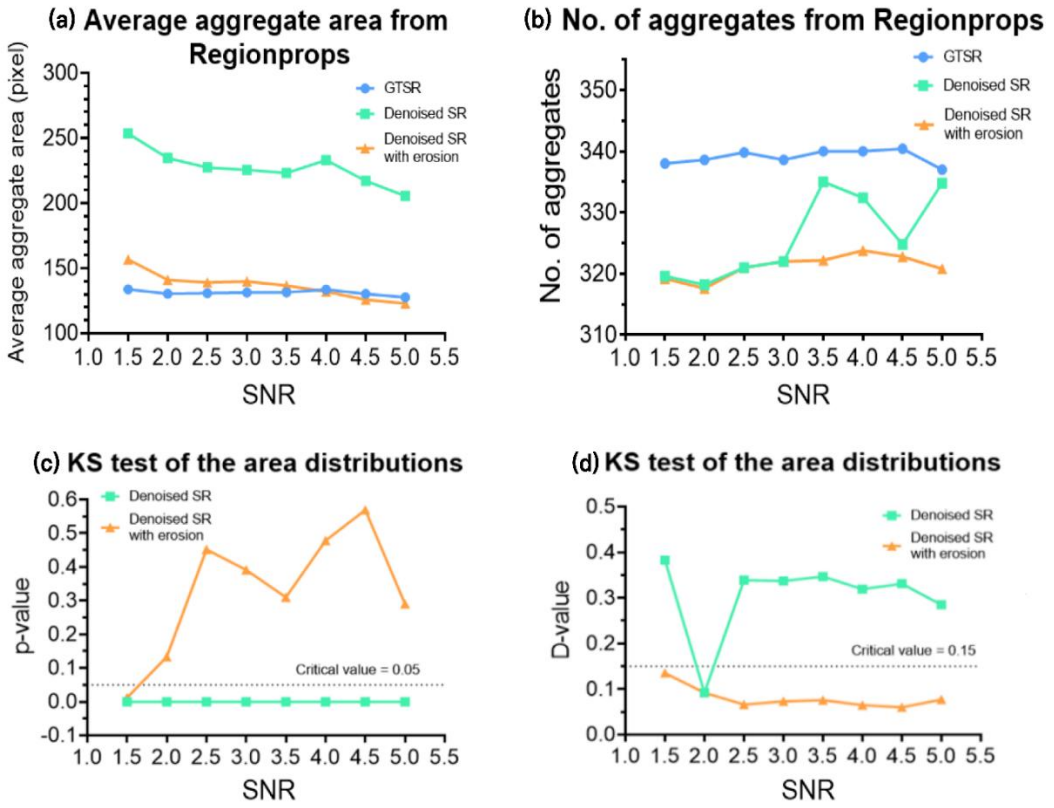

Figure 13. Quantitative analysis of the results from the regionprops. (a) The average aggregate area of GTSR, denoised SR and denoised SR with erosion with various SNRs varying from 1.5 to 5.5. (b) The number of aggregates of the three types of images with SNRs varying from 1.5 to 5.5. (c) P-values and (d) D-values from the KS test of the area distributions of the GTSR, denoised SR and denoised SR with erosion against the SNR values.

3.3.2. Optimizing the number of frames. The same method was applied to determine the optimum number of frames for each DL stack. As illustrated in Figure 14 (a), the erosion method effectively removed unnecessary enlargement of the aggregates, reducing the average aggregate area from over 200 to approximately 150, which was consistent with the areas of GTSR images. When the number of frames was 4000, there was minimum difference in the average aggregate area between the post-processed denoised SR images and the GTSR. Similarly, corrections made in the average aggregate area compromised the total number of aggregates. As shown in Figure 14 (b), the total number of aggregates decreased from around 340 to 325.

For the denoised SR, the P-values of the KS test of the area of distributions, as shown in Figure 14(c), were maintained at values close to 0 as the number of frames increased. However, for denoised SR with erosion, there was an overall increasing trend as the P-values gradually increased from 0.7 to 0.9 as the number of frames increased (though the P-value dropped to 0.2 when there were 2000 frames). The highest P-value of 1.0 was obtained when the number of frames was 4000.

Also illustrated in Figure 14 (d), the D-values for denoised SR were all around 0.3, which were above the critical value of 0.15. Again, erosion proved to be effective in enhancing denoise capability as the D-values decreased significantly, dropping below 0.1 for all numbers of frames per DL stack. The lowest D-value, which was approximately 0.03, was obtained when the number of frames was 4000.

Experimental data led to the conclusion that the optimum number of frames was 4000.

DOI: 10.54254/2753-8818/23/20231050

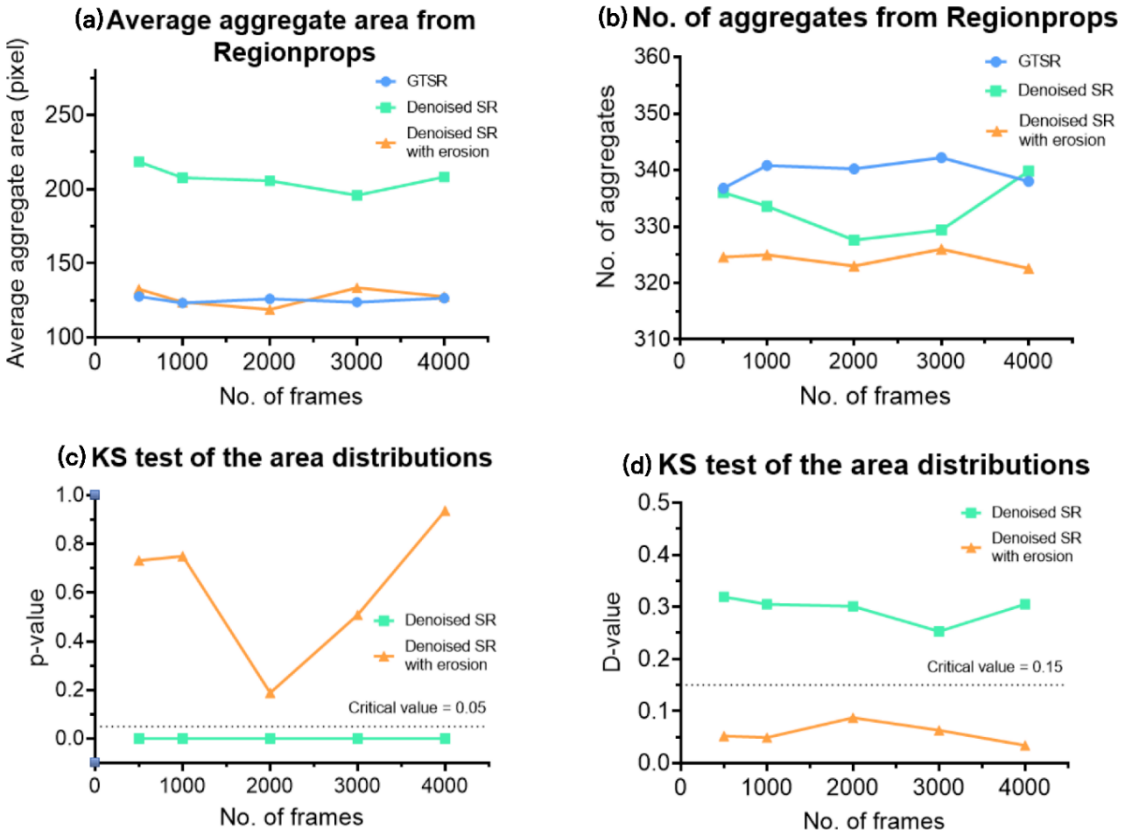

Figure 14. Quantitative analysis of the results from the regionprops. (a) The average aggregate area of GTSR, denoised SR and denoised SR with erosion with various numbers of frames varying from 500 to 4000. (b) The number of aggregates of the three types of images with varying numbers of frames. (c) P-values and (d) D-values from the KS test of the area distributions of the GTSR, denoised SR and denoised SR with erosion against the number of frames.

3.4. Optimizing the hyperparameters of the U-net

Under optimal imaging conditions (signal-to-noise ratio = 4.5, number of frames = 4000), additional tests were performed based on the validation loss to determine the best combination of neuron number, kernel size, and activation function.

Grid-search results, as shown in Figure 15 (a), illustrated that there was an overall trend that validation loss decreased as the number of neurons and kernel size increased. The ReLU activation function generally yielded a smaller validation loss than that of Tanh. The validation loss reached a minimum when the hyperparameter combination of 64 neurons, a 4×4 kernel, and the ReLU activation function was adopted.

In addition, by evaluating the training time with 15 epochs, it was shown that the training time for Tanh was greater than that of ReLU, and the training time increased exponentially with the number of neurons and the kernel size.

DOI: 10.54254/2753-8818/23/20231050

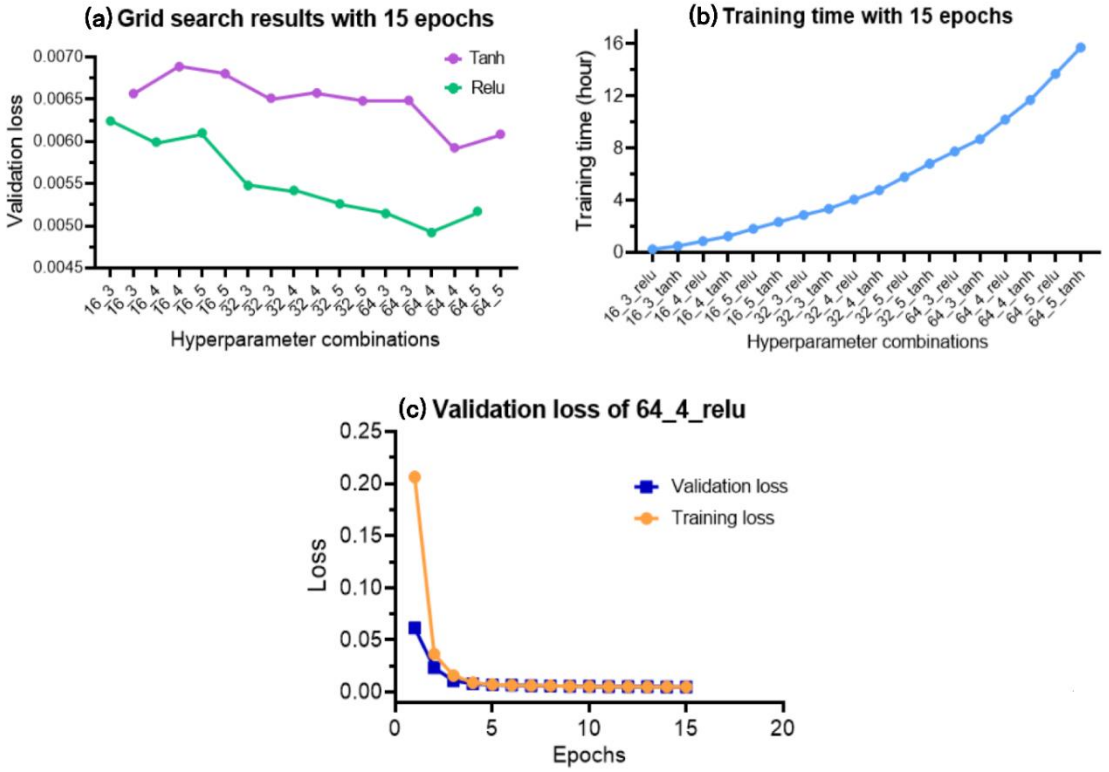

Figure 15. Validation loss and training time of different hyperparameter combinations. (a) Grid-search results of validation loss of different hyperparameter combinations. (b) Training time of different hyperparameter combinations. (c) The validation loss curve of the number of epochs after the hyperparameters were determined.

3.5. Training the final model

After determining the optimum imaging conditions and the hyperparameters, all 390 noisy SR images were input into the model. The final model proved to be more efficient than the previous models. As shown in Figure 16 (a-b), the final model is capable of resolving structures that were not resolvable by previous models on datasets with poor imaging conditions. Moreover, the final model can also resolve structures that were close to or overlapping with each other (Figure 16 (c)). However, this model could still not process extremely complicated structures and sometimes introduced new noises. While the aggregates in the denoised SR images exhibit more similarities to those in the GTSR images, there are noticeable disparities in both their quantity and size. Erosion rectified the differences in number and area but also introduced distortions the shape of the aggregates.

Additionally, the final model demonstrated accurate prediction of the area of the aggregates. As depicted in Figure 17, the final model, with the aid of post-processing measures, generated an area distribution of the aggregates that closely resembled that of the GTSR. Post-processing mechanisms significantly reduced D-values from 0.328 to 0.078 and increased P-values from approximately 0 to 0.328. Furthermore, the distribution of the denoised SR images was significantly narrowed down.

DOI: 10.54254/2753-8818/23/20231050

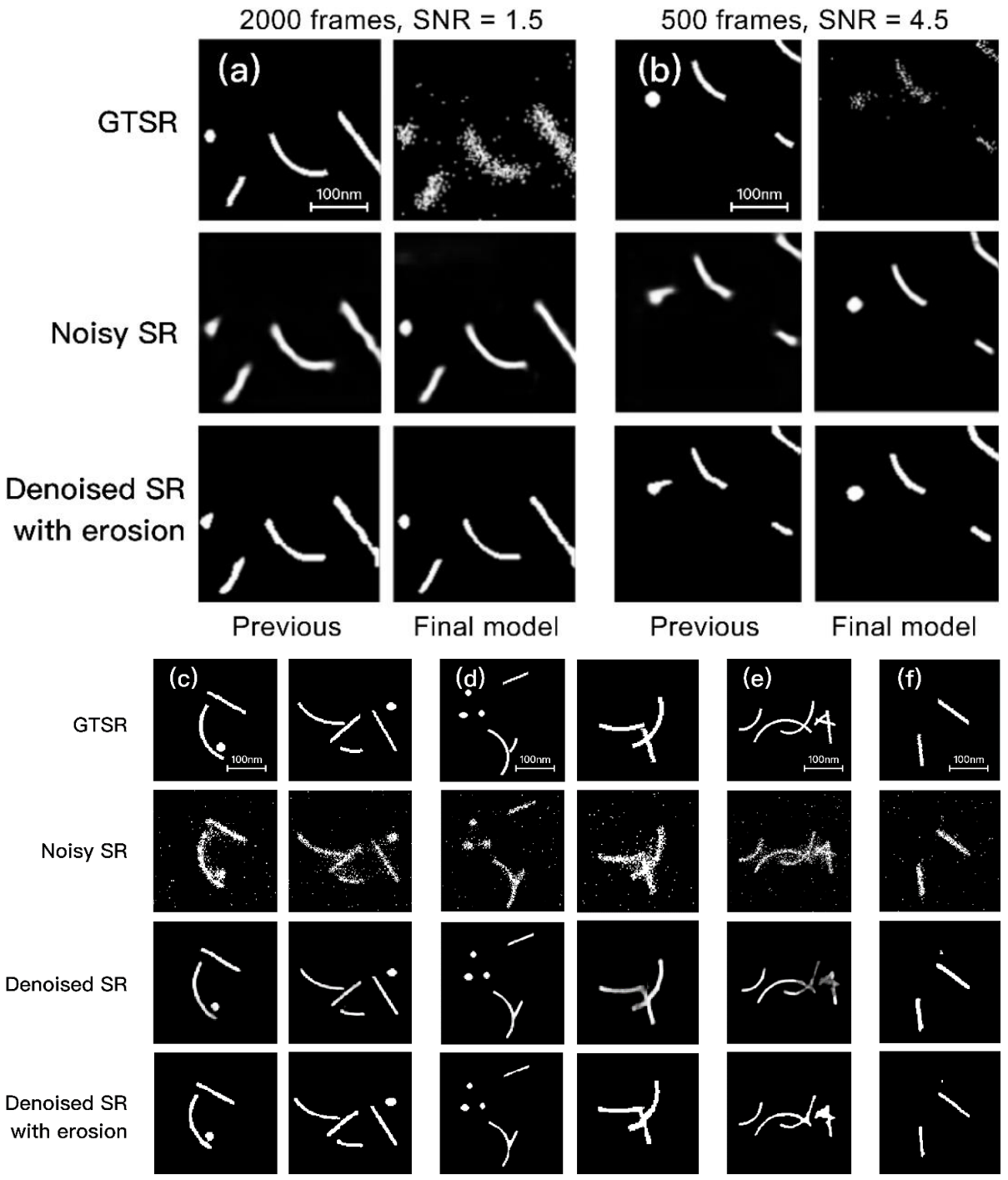

Figure 16. A comparison between a previous model and the final one, and a comparison between GTSR, noisy SR, U-net's denoised output, and post-processed images in the final model. (a) and (b) are images used to train the previous model as well as the U-net's final output with different frame numbers and SNRs. (c) The final model resolved structures closely compacted together. (d) The final model resolved overlapping structures. (e) The model was confused by complicated structures. (f): Random noise might be added by the U-net model itself.

DOI: 10.54254/2753-8818/23/20231050

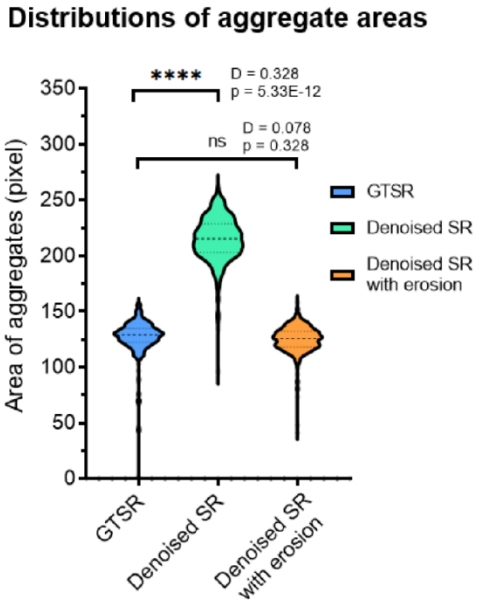

Figure 17. The predicted area distribution of aggregates of the GTSR, denoised SR, and denoised SR with erosions.

4. Discussion

This U-net based CNN model exhibits excellent performance in denoising SMLM images of protein aggregates. The model was trained and evaluated using simulated noise-free super-resolution images of aggregates, along with their corresponding noisy images containing NSB signals. By fine-tuning the hyperparameters of the U-net, the model has attained a low validation loss and prediction errors in terms of the area and number of aggregates, indicating the model's capability to accurately denoise SMLM images. These outstanding outcomes demonstrate the effectiveness of our model in overcoming the challenges posed by noise in SMLM imaging, enabling accurate characterization of the structures and morphological features of protein aggregates.

4.1. Simulation of SMLM datasets

In this study, the U-net model was trained and evaluated using SR images comprising straight, curved, and dot aggregates, simulating potential protein aggregates encountered in research scenarios. However, these simulated aggregates only represent an idealized version of the actual biomolecules, which exhibit far greater complexity than what the simulation produced (as shown in Figure 1.1). To further assess the performance of our model, we aim to apply it to real-world imaging data and employ quantitative analysis methods for a thorough evaluation of its capabilities. Additionally, we intend to expand the simulation process to encompass more complex situations and shapes, thereby providing a more comprehensive representation of the diversity of biomolecules encountered in practical research scenarios.

The simulation data incorporated a wide range of SNRs to model various conditions of real-world imaging experiments, from a low SNR of 1.5 to improved values of 5. However, this range of SNRs does not represent all possible conditions and scenarios. While the intensity variation was fixed at 0.2 for all imaging conditions, further experiments are necessary to explore a greater range of intensity variations and SNRs.

The primary objective to generating DL stacks of varying numbers of frames was to manipulate the total number of localizations. Simply increasing the number of localizations per frame would result in excessive overlapping localizations, undermining the Gaussian fitting by ThunderSTORM. To mitigate this issue, we have limited the number of localizations available on each individual frame and adjusted the total number of localizations by manipulating the number of frames per DL stack. It was assumed that the localizations would be uniformly distributed among all the frames. However, in real scenarios,

DOI: 10.54254/2753-8818/23/20231050

the distribution of the localizations is often uneven. Extreme conditions which lead to high localization densities were not adequately considered in the simulation dataset. Future investigations should focus on generating corresponding data of extreme conditions and subsequently evaluate the model's capability to handle such scenarios.

4.2. Choice of the CNN model

The U-net architecture was adopted in this project for denoising SMLM images since it was relatively easy to establish, highly accurate, and was proven to be a very powerful tool when limited training data is available [20]. However, the U-net does suffer from slower training speed and higher consumption of GPU memory. More recent CNN models, such as ENet [24] and DenseNet [25], have demonstrated improvements in efficiency and performance compared to the U-net. Enet offers improved frame rates compared to traditional CNN models, while the DenseNet, with its increased number of layers and more complex interconnections, provides enhanced accuracy and reliability in predicting images. By applying these new development, we may establish better denosing models with the same method in this work but in a shorter time.

4.3. Optimization

Based on the previous training results, it was concluded that the U-net model exhibited its optimal performance when the SNR was 4.5. However, it is important to note that this experimental value may vary depending on the hyperparameters and the specific structure of the U-net model. Nevertheless, this finding serves as a successful reference for future training as it helped in selecting the optimum SNR for simulating different number of frames.

In this investigation, the generation of images with various numbers of frames was solely based on the previously detemined optimum SNR of 4.5. Therefore, a wide range of combinations involving various SNRs and numbers of frames remains unevaluated. Moreover, as discussed in Section 3.3, altering the number of frames alone did not directly reveal the impact of the number of localizations on the final denoising result. As a result, it becomes necessary to further expand the training dataset, which can be achieved by generating more datasets with different SNRs, numbers of frames, and varying numbers of localizations per frame.

4.4. Denoised data and post-processing methods

The U-net model generated excellent denoising results, effectively removing noise and producing fundamental aggregate features that closely resembled the ground truth super-resolution (GTSR) images. The application of filtering techniques successfully eliminated small, dark noise elements introduced by the CNN model itself, while the erosion helped restore the original shapes of the aggregates. Although some degree of distortion was observed in most denoised images, optimizing the SNRs and number of frames effectively reduced the level of distortion. In less complex images, the final model demonstrated significant improvements in addressing distortion (Figure 16(a-b)). In more complex images, although the final model could not perfectly restore the exact shapes of individual aggregates, it still provided highly similar predictions (Figure 16(c-d). Therefore, this model can produce valuable references of the sizes and shapes of the aggregates. Further intensive and rigorous training with more complicated aggregate shapes, increased number of SMLM datasets and a more advanced CNN models will be anticipated to further enhance the denoising performance.

4.5. Quantitative analysis

The quantitative analysis provided a direct statistical comparison between the GTSR and denoised SR images. The results demonstrated a high degree of similarity between these two image sets in terms of the average aggregate area and the average number of aggregates. However, it is important to note that this general similarity observed during quantitative analysis does not necessarily indicate that the images were the same when evaluated at the pixel level. There remains uncertainty regarding whether the filtering process successfully eliminated all additional noise introduced by the U-net model itself, as

DOI: 10.54254/2753-8818/23/20231050

opposed to potentially removing some real aggregates. Nevertheless, the resulting average number of aggregates may still appear similar. Similarly, the erosion process might have resulted in the trimming of some aggregates, while others may have remained larger compared to their corresponding counterparts in GTSR. With the presence of remaining noise elements, the overall average area may still align with that of GTSR. The results from the quantitative analysis served as a valuable reference for assessing the model's capabilities. However, it requires further data analysis of the denoised and post-processed outputs by conducting a pixel-by-pixel comparison with the GTSR images.

4.6. Optimizing the hyperparameters and training the final model

The optimal set of hyperparameters was determined by comparing the validation loss across the 18 combinations of kernel size, number of neurons, and activation functions. However, this process has its limitations, as it only considered a limited range of kernel size and the number of neurons, while neglecting the potential influence of the number of layers available in the U-net architecture. Therefore, alternative techniques such as random research and Bayesian optimization [23] seemed to be more promising. Random research involves randomly combining various hyperparameters, which improves efficiency and saves time and requires less computation. However, it does not evaluate all possible combinations of hyperparameters. On the other hand, Bayesian optimization uses a probabilistic model to approximate the objective function and uses past evaluations to select the next hyperparameter set to assess. This method tends to converge more quickly to the optimal solution compared to traditional methods such as grid search or random search.

The final model demonstrated superior effectiveness compared to individual models designated for each imaging condition, indicating that optimizing hyperparameters and performing comprehensive training with all simulated data significantly enhanced the denoising performance. However, the available data, consisting of 12,000 cropped noisy SR images, could still be expanded by simulating additional SMLM image sets to train a more powerful CNN network. Moreover, the training time exceeding 50 hours was extremely user-unfriendly. To address these issues, further improvements could be made by expanding the dataset, adopting an advanced U-net structure, or considering alternative models as described in Section 4.2.

5. Conclusion

In this work, a CNN based on a U-net architecture with optimized hyperparameters was developed to effectively remove the NSB noise in simulated SR images of protein aggregates acquired under various imaging conditions. The final model achieved a low validation loss of 0.0042 and successfully eliminated the NSBs present in the noisy SR images. With the aid of image post-processing techniques, the denoised images exhibited similar morphological information, area distribution, and total number of aggregates compared to the GTSR images, with errors in area and no. of aggregates reaching as low as 0.32% and 3.71%, respectively. While post-processing enhanced the quality of the denoised images, it occasionally introduced distortions. To further enhance the denoising ability, a larger training dataset encompassing a diverse range of aggregate shapes, numbers, and SNRs is highly recommended. This study presents a robust and effective tool for denoising SMLM images of ND-associated protein aggregates, offering significant benefits to related research fields. Moreover, this model exhibits significant potential for broader application in denoising real SMLM images beyond protein aggregates. It can be extended to other essential bio-complexes, such as the intricate nuclear pore complex. This technique in denoising SMLM images offers valuable support for the development of diagnostic techniques and potential therapeutic interventions targeting NDs.

References

[1] Reitz C, Brayne C, Mayeux R. 2011. Epidemiology of Alzheimer's disease. Nat Rev Neurol 7: 137 –152.

[2] Mayeux R, Stern Y. 2012. Epidemiology of Alzheimer's disease. Cold Spring Harb Perspect Med 2: 10.1101/cshperspect.a006239.

[3] Sosa-Ortiz AL, Acosta-Castillo I, Prince MJ. 2012. Epidemiology of dementias and Alzheimer's disease. Arch Med Res 43: 600–608
[4] Michael G. Erkkinen, Mee-Ohk Kim, and Michael D. Geschwind. 2019. Clinical neurology and epidemiology of the major neurodegenerative diseases. Cold Spring Harbor Laboratory Press.
[5] Kolarova M, García-Sierra F, Bartos A, Ricny J, Ripova D. Structure and pathology of Tau protein in Alzheimer disease. Int J Alzheimers Dis, 2012, 2012:731526.
[6] Mansor NI, Ntimi CM, Abdul-Aziz NM, Ling KH, Adam A, Rosli R, Hassan Z, Nordin N. Asymptomatic neurotoxicity of amyloid β-peptides (Aβ1-42 and Aβ25-35) on mouse embryonic stem cell-derived neural cells. Bosn J Basic Med Sci. 2021 Feb 1;21(1):98-110.
[7] Srivastava AK, Pittman JM, Zerweck J, Venkata BS, Moore PC, Sachleben JR, Meredith SC. β-Amyloid aggregation and heterogeneous nucleation. Protein Sci. 2019 Sep;28(9):1567-1581
[8] Lichtman, J., Conchello, JA. Fluorescence microscopy. *Nat Methods* 2, 910–919 (2005).
[9] Khater I M, Nabi I R, Hamarneh G. A review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods[J]. Patterns, 2020, 1(3): 100038.
[10] Khater IM, Nabi IR, Hamarneh G. A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods. Patterns (N Y). 2020 Jun 12;1(3):100038.
[11] Lelek, M., Gyparaki, M.T., Beliu, G. *et al.* Single-molecule localization microscopy. *Nat Rev Methods Primers* 1, 39 (2021).
[12] Nieves D J, Gaus K, Baker M A B. DNA-based super-resolution microscopy: DNA-PAINT[J]. Genes, 2018, 9(12): 621.
[13] Schnitzbauer Joerg, Strauss Maximilian T, Schlichthaerle Thomas, Schueder Florian, Jungmann, Ralf (2017). Super-resolution microscopy with DNA-PAINT. Nature Protocols, 12(6), 1198–1228.
[14] Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian Association of Radiologists white paper on artificial intelligence in radiology[J]. Canadian Association of Radiologists Journal, 2018, 69(2): 120-135. (21)
[15] McBee M P, Awan O A, Colucci A T, et al. Deep learning in radiology[J]. Academic radiology, 2018, 25(11): 1472-1480.
[16] Currie G, Hawk K E, Rohren E, et al. Machine learning and deep learning in medical imaging: intelligent imaging[J]. Journal of medical imaging and radiation sciences, 2019, 50(4): 477-487.
[17] Yamashita R, Nishio M, Do R K G, et al. Convolutional neural networks: an overview and application in radiology[J]. Insights into imaging, 2018, 9(4): 611-629.
[18] Nair, V., Hinton G., "Rectified Linear Units Improve Restricted Boltzmann Machines," 27th International Conference on International Conference on Machine Learning, ICML'10, USA: Omnipress, 2010, pp. 807–814, ISBN 9781605589077
[19] L. Jin, B. Liu, F. Zhao, S. Hahn, B. Dong, R. Song, T. C. Elston, Y. Xu, and K. M. Hahn, "Deep learning enables structured illumination microscopy with low light levels and enhanced speed," Nat. Commun. 11, 1934 (2020).
[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," arXiv:1505.04597 (2015). 31. J. F. Abascal, S. Bussod, N. Ducros, S. Si-Mohamed, P. Douek, C. Chappard, and F. Peyrin, "A residual U-Net network with image prior for 3D image denoising," HAL hal-02500664 (2020)
[21] Nur Izzati Mansor, Carolindah Makena Ntimi, Noraishah Mydin Abdul-Aziz, et al. "Asymptomatic neurotoxicity of amyloid β-peptides (Aβ1-42 and Aβ25-35) on mouse embryonic stem cell-derived neural cells," Biomolecules and Biomedicine 21 (1):98-110 (2021)
[22] Fan L, Zhang F, Fan H, et al. Brief review of image denoising techniques[J]. Visual Computing for Industry, Biomedicine, and Art, 2019, 2(1): 1-12.

DOI: 10.54254/2753-8818/23/20231050

[23] Li X, Hu Y, Gao X, et al. A multi-frame image super-resolution method[J]. Signal Processing, 2010, 90(2): 405-414.
[24] Paszke, Adam & Chaurasia, Abhishek & Kim, Sangpil & Culurciello, Eugenio. (2016). ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation.
[25] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.