ISSN 2616-5775 Vol. 6, Issue 12: 45-49, DOI: 10.25236/AJCIS.2023.061205

Accelerated Single-Molecule Microscopy Image Reconstruction Algorithm Based on TensorRT

Qinfang Zhang[1,a,*]

[1]*School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China*
[a]*zqf19991223@163.com*
[*]*Corresponding author*

Abstract: *The effectiveness of deep learning-based high-density localization techniques in expediting applications of single-molecule localization microscopy is well-acknowledged. However, existing methods within this domain grapple with speed limitations, particularly when confronted with extensive raw data sets. To address this challenge, this study introduces a high-density localization approach founded on TensorRT, incorporating an enhanced U-shaped network for the swift reconstruction of raw images. Experimental verification conducted on a dataset featuring microtubule proteins illustrates that the proposed algorithm significantly amplifies the overall pace of reconstruction without compromising the precision of image reconstruction. This not only showcases a noteworthy advancement in speed but also establishes a practical technical remedy for the swift and high-fidelity reconstruction of fluorescence single-molecule microscopy images.*

Keywords: *microscopy; fluorescence microscopy; super-resolution; image reconstruction; TensorRT*

1. Introduction

Single Molecule Localization Microscopy (SMLM)[1] is an emerging super-resolution imaging technique, holding significant promise as a reconstruction method in the field of super-resolution fluorescence microscopy. It can achieve a spatial resolution of 20-30 nm using a simple optical device. However, the imaging speed is relatively low, typically ranging from tens of seconds to minutes. This is attributed to the necessity for expensive equipment and the requirement for thousands of raw images in the reconstruction process of super-resolution images using the SMLM method. Consequently, this limitation restricts the speed of live cell imaging.

Currently, three primary approaches are employed to enhance the imaging speed of Single Molecule Localization Microscopy (SMLM): image restoration, fast image acquisition, and high-density localization. Image restoration involves correcting images by inputting downsampled low-density images into a trained model to infer a restored image. The ANNA-PALM method[2], introduced in 2018, marked the initial presentation of this technique. In contrast, fast image acquisition achieves the rapid reconstruction of super-resolution images by diminishing the overall acquisition time through a reduction in the camera exposure time for each raw image. For instance, Sara[3] achieves a fivefold acceleration, albeit potentially accompanied by signal-to-noise ratio degradation and photobleaching issues. On the other hand, high-density localization aims to decrease the total number of raw images without compromising the total number of fluorescent molecules employed for super-resolution image reconstruction. This is achieved by enhancing the density of emitters in each raw image, consequently reducing the overall number of raw images. In 2018, Nehme et al.[4] introduced Deep-STORM, a high-density localization method grounded in super-resolution convolutional neural networks (SRCNNs). This method enables swift and accurate image reconstruction from a high-density image. These methodologies present distinct approaches to augmenting the imaging speed of SMLM, each carrying its own set of advantages and limitations.

In order to further improve the data processing speed of SMLM, this paper proposes an image reconstruction acceleration method based on the combination of TensorRT and deep neural networks. The specific work of this research is mainly in the following two aspects:

1) The deep neural network is combined with the graph fusion function of TensorRT to further accelerate the model inference process.

2) In the inference process, the sum of the localized images is processed directly on the GPU, reducing

ISSN 2616-5775 Vol. 6, Issue 12: 45-49, DOI: 10.25236/AJCIS.2023.061205

the time of data replication.

Experiments demonstrate that the method presented in this paper significantly accelerates the single-molecule reconstruction speed and exhibits excellent performance in the experimental dataset. It achieves real-time data processing of 2D high-density SMLM, which is crucial for disease diagnosis, drug development, and optimization of treatment protocols in biomedical research.

2. Methods

To further expedite the reconstruction inference process of deep neural networks, this paper builds upon prior work and makes additional improvements based on the literature [5]. The paper introduces a novel high-density localization method, accelerated by TensorRT. It leverages U-networks to directly extract features from low-resolution raw images. The method incorporates TensorRT's image fusion functionality, facilitating the deployment of the model onto the TensorRT inference framework for enhanced inference speed. Simultaneously, an approach of directly summing N-frame localization images on the GPU is employed to minimize data copying time from the GPU device to the host. This approach results in additional speedup, ultimately generating a reconstructed image. This operation significantly enhances overall reconstruction speed without compromising image reconstruction accuracy.

The outlined strategy aims to significantly enhance the reconstruction inference process of deep neural networks. It represents a continued advancement upon previous work. The proposed high-density localization method exhibits notable efficiency through TensorRT acceleration. This efficiency is achieved by utilizing U-networks for feature extraction from low-resolution images and harnessing TensorRT's capabilities for model deployment, contributing to a more rapid inference speed. Additionally, the GPU-based summation of N-frame localization images efficiently reduces data transfer times, providing a more streamlined process for information exchange. The culmination of these improvements manifests in a reconstructed image that demonstrates enhanced overall reconstruction speed without sacrificing accuracy.

The implementation of this approach injects new vitality into the reconstruction inference process of deep neural networks. It achieves a balance between speed and accuracy, providing a promising solution for the reconstruction inference of deep neural networks. This research offers feasible insights for further exploration and application in the field.

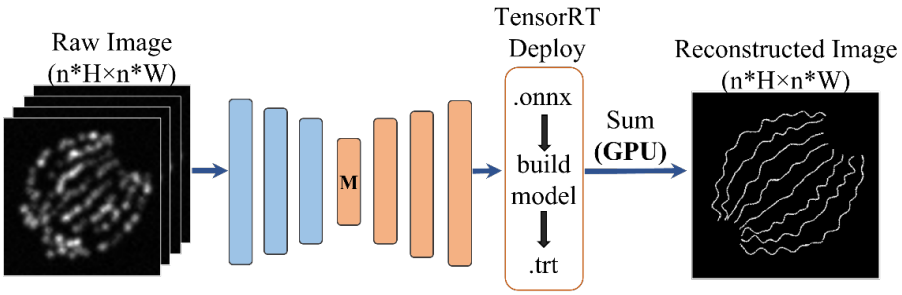

Figure 1: Overall network structure diagram

3. Experimental Procedures

3.1 Dataset preparation

The experimental approach of this study involved the utilization of a training dataset (Figure 2), generated randomly through the ThunderSTORM[6] plugin in Image J[7]. This generation aimed to simulate imaging conditions similar to those encountered in the experimental data under evaluation. Subsequently, the images and their corresponding position lists from the training dataset were imported into MATLAB.To expand the training set, 500 randomly selected 26×26 pixel subregions were chosen from each frame's simulated image pairs. These subregions then underwent an 8-fold upsampling process. Simultaneously, the position coordinates of these subregions were projected onto a high-resolution grid. This process resulted in a dataset that encompasses 10,000 pairs of low-resolution and corresponding high-resolution regions, each measuring 208×208 pixels.

ISSN 2616-5775 Vol. 6, Issue 12: 45-49, DOI: 10.25236/AJCIS.2023.061205

To mitigate overfitting, the dataset was partitioned into 7,000 samples for training and 3,000 samples for validation. For the test dataset, a simulated microtubule dataset from the EPFL·SMLM Challenge website[8] was employed to assess the performance of the algorithm proposed in this paper. Through this meticulous approach to data processing and segmentation, the quality and reliability of the experimental data were ensured, providing robust groundwork for an objective evaluation of the proposed algorithm's performance.

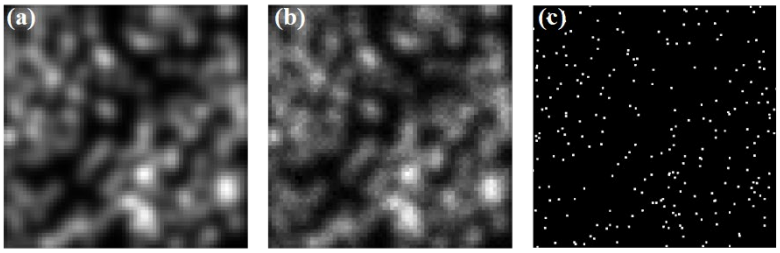

Figure 2: Simulation of dense fluorescence signal. (a) image of noiseless fluorescence signal; (b) image of noisy fluorescence signal; (c) image of localized fluorescence signal

3.2 Experimental environment and training details

During the training process, a normalization step was applied to the images, aiming to enhance the convergence speed of the neural network. Normalizing the images aids in preventing issues related to varying scales among features, thereby facilitating a more efficient training process. By standardizing the input data, the neural network can converge more quickly, leading to improved training performance and potentially better generalization on unseen data. In the experiments, simulation-generated images served as inputs to the network. The input size was set to 208 × 208 × 1, and the network was trained for 100 epochs with a batch size of 16 samples per batch. The loss function L is defined as:

$$L\left(x,\hat{x}\right)=\frac{1}{K}\left(\sum_{i=1}^{K}\left\|g*\hat{x}_i-g*x_i\right\|_2^2+\left\|\hat{x}_i\right\|_1\right) \tag{1}$$

Where $\hat{x}$ represents the network reconstruction result, x represents Ground Truth. K is a predefined training set consisting of pairs of 2D images. The training is performed using the Adam[9] backpropagation algorithm, the learning rate is set to 1×10^{-4} , the model is built based on the Tensorflow framework[10] , and the whole experiment is performed on a standard workstation equipped with 43 GB of RAM, a 9-core Intel(R) Xeon(R) Platinum 8350C 2.60GHz CPU, and a 24 GB video memory NVidia GeForce RTX 3090 GPU for training and evaluation on a standard workstation.

3.3 Evaluation metrics

Normalized Mean Square Error (NMSE) is used to evaluate the quality of the image reconstruction. A smaller NMSE value indicates less error and better reconstruction, with NMSE=0 signifying an identical reconstructed image to the real image. The expression for NMSE is as follows:

$$NMSE\left(\hat{x},x\right)=\frac{\left\|\hat{x}-x\right\|_2^2}{\left\|x\right\|_2^2}\times100\% \tag{2}$$

The Peak Signal-to-Noise Ratio (PSNR) serves as a metric for estimating the difference between two images relative to the peak signal amplitude of the predicted image. PSNR is typically calculated in decibels, where higher scores indicate better image quality.

$$MSE=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(\hat{x}_{(i,j)}-x_{(i,j)}\right)^2 \tag{3}$$

$$PSNR=20\times\log_{10}\left(\frac{MAX_1}{\sqrt{MSE}}\right) \tag{4}$$

ISSN 2616-5775 Vol. 6, Issue 12: 45-49, DOI: 10.25236/AJCIS.2023.061205

4. Experimental Results & Analysis

4.1 Visualization Assessment

After completing model training, we used the simulated microtubule dataset (361 frames of sample images) from the EPFL·SMLM website [8] as a validation set for reconstruction experiments. The experiments involved CEL0 [11], FALCON [12], Deep-STORM [4], and the algorithm proposed in this paper.All networks underwent reconstruction on a grid with a resolution of 12.5nm. They were then thoroughly compared and analyzed against the original high-resolution images (Ground Truth), as depicted in Figure 3.To precisely assess the reconstruction accuracy of each algorithm, this study conducted evaluations from both subjective and objective perspectives, considering the quality of the reconstructed high-resolution images.

In terms of subjective visual assessment, we observed instances of discontinuity in the recovery of certain microtubules in images reconstructed by CEL0, FALCON, and Deep-STORM algorithms. This led to less precise restoration of edge details. In contrast, our proposed algorithm demonstrated a reduction in pseudo-peaks and noise between two microtubules in high-density regions, resulting in a more accurate distinction of adjacent and closely spaced microtubules and showcasing a higher level of reconstruction accuracy.

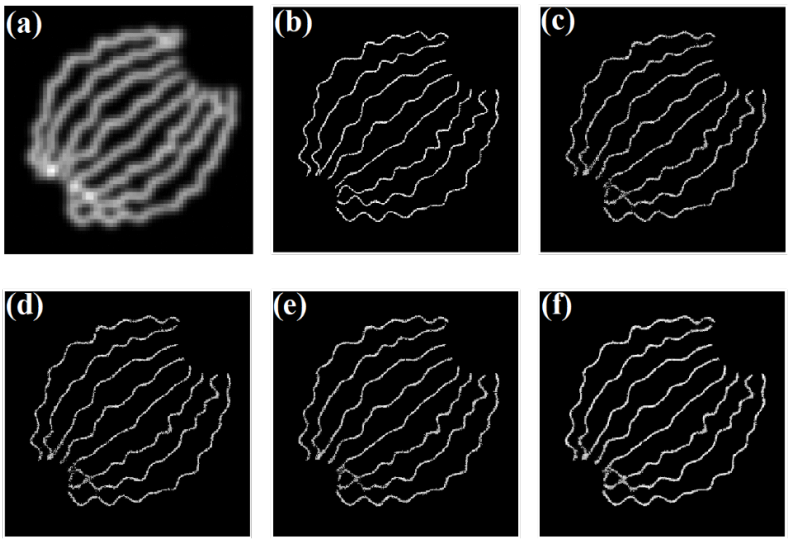

Figure 3: Simulated microtubule dataset reconstruction results. (a) 361 frames of original wide-field image; (b) Ground Truth; (c) CEL0; (d) FALCON;(e) Deep-STORM; (f) Ours

4.2 Objective Assessment

In this paper, the reconstruction accuracy and inference speed of the four methods are evaluated based on the three criteria of NMSE, PSNR and single-frame execution time, respectively, and the specific results are shown in Table 1.Compared with the CEL0, FALCON and Deep-STORM algorithms, the proposed algorithm's NMSE is reduced by 3.35%, 3.19%, and 6.49%, and the PSNR is improved by 0.95dB, 0.5dB , 1.09dB, and single-frame execution time is accelerated by 4×10^3 times, 26 times, and 0.63 times, respectively, which indicates that the proposed algorithm can not only achieve better results in terms of localization accuracy, and be able to rebuild a better image quality to a greater extent, but also achieve faster speed in terms of reconstruction time.

Table 1: Comparison of objective evaluation indicators of reconstruction results

Methods	NMSE(%)	PSNR(dB)	ms/frame
CEL0	37.67	22.47	51736.8
FALCON	40.81	22.33	337.95
Deep-STORM	37.51	22.92	20
literature [5]	**34.32**	**23.42**	16.6
Ours	34.65	23.29	**12.3**

In summary, the algorithm proposed in this paper excels in effectively separating the microtubule

ISSN 2616-5775 Vol. 6, Issue 12: 45-49, DOI: 10.25236/AJCIS.2023.061205

structure from the background region, thereby achieving a more precise high-resolution reconstruction image. Notably, it outperforms several algorithms previously proposed in the literature, showcasing a substantial enhancement in reconstruction speed. This demonstrated superiority establishes a robust foundation for research and practical applications in the field of deep learning image processing. Specifically, the algorithm provides valuable insights into optimizing microtubule reconstruction, enhancing image clarity, and expediting the overall process. Moreover, it serves as a valuable reference for future work in this domain, offering a roadmap for further exploration and expansion of deep learning techniques in the context of image processing.

5. Conclusion

This paper tackles the speed bottleneck inherent in deep learning for reconstructing single-molecule localization microscopy images. A novel acceleration approach is introduced, integrating TensorRT with deep neural networks to optimize the image inference process. The method leverages the acceleration capabilities of TensorRT and GPU. Experimental findings reveal that the proposed method not only achieves higher-quality images, as indicated by clearer recovery of microtubule structure edge contours and textures resembling the original high-resolution images, but also demonstrates significant improvements in objective evaluation metrics and reconstruction speed. Specifically, [insert specific metrics and speed improvements]. This achievement successfully meets the objective of enhancing reconstruction speed without compromising image quality. Consequently, it is anticipated that this method will contribute to increased efficiency and quality in microscopic image reconstruction, holding paramount significance for the advancement of related fields.

References

[1] Deschout H, Zanacchi F C, Mlodzianoski M, et al. Precisely and accurately localizing single emitters in fluorescence microscopy[J]. Nature methods, 2014, 11(3): 253-266.
[2] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," Nat. Biotechnol. 36(5), 460–468 (2018).
[3] S. A. Jones, S.-H. Shim, J. He, and X. Zhuang, "Fast, three-dimensional super-resolution imaging of live cells," Nat. Methods 8(6), 499–505 (2011).
[4] Nehme E, Weiss L E, Michaeli T, et al. Deep-STORM: super-resolution single-molecule microscopy by deep learning [J]. Optica, 2018, 5(4): 458-464.
[5] Zhang Q, Lou L, Li Q, et al. Super-resolution single-molecule microscopic reconstruction based on improved deep-STORM[C]//5th International Conference on Computer Information Science and Application Technology (CISAT 2022). SPIE, 2022, 12451: 458-463.
[6] Liu Y, Wang Y, Li N, et al. An attention-based approach for single image super resolution[C]//2018 24 Th international conference on pattern recognition (ICPR). IEEE, 2018: 2777-2784.
[7] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. pmlr, 2015: 448-456.
[8] Sage D, Kirshner H, Pengo T, et al. Quantitative evaluation of software packages for single-molecule localization microscopy[J]. Nature methods, 2015, 12(8): 717-724.
[9] Kingma D, Ba J .Adam: A Method for Stochastic Optimization[J].Computer Science, 2014. DOI:10. 48550/arXiv.1412.6980.
[10] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems[J]. 2016.DOI:10.48550/arXiv.1603.04467.
[11] Gazagnes S, Soubies E, Blanc-Féraud L. High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation[C]//2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017: 28-31.
[12] Min J, Vonesch C, Kirshner H, et al. FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data[J]. Scientific reports, 2014, 4(1): 1-9.