Supplementary of Paper – Volume Tells: Dual Cycle-Consistent Diffusion for 3D Fluorescence Microscopy De-noising and Super-Resolution

Zelin Li[1,2,*], Chenwei Wang[2,*], Zhaoke Huang[1,2,†], Yiming Ma[3], Cunming Zhao[3], Zhongying Zhao[3], Hong Yan[1,2]

1. The Motivation in Details

Fluorescence microscopy (FM) is a powerful tool for studying cellular processes by using fluorescent markers to label components like membranes, nuclei, and proteins, enabling detailed visualization of cell structures, functions, and dynamics. It provides insights into processes such as cell signaling, gene expression, and tissue organization while revealing critical features like shape, volume, and subcellular localization. By tracking these features over time, FM uncovers dynamic processes such as cell division, migration, and apoptosis, as well as cellular heterogeneity. However, achieving continuous, high-quality signals for precise observation and analysis remains a challenge, especially in complex, multi-cellular contexts where rapid divisions and small cell sizes blur fluorescence signals and reduce resolution.

The limitations in 3D fluorescence imaging primarily stem from the inherent characteristics of the imaging mechanism. First, due to the optical diffraction limit, the system achieves higher resolution in the lateral (XY) plane but has lower resolution along the axial (Z) axis. This anisotropic resolution difference is caused by the physical properties of the optical system: the focusing ability is weaker along the axial direction, making it challenging to capture the same level of detail in depth as in the lateral plane. Additionally, the uneven noise distribution is related to how light propagates through the sample. In deeper regions along the Z-axis, light undergoes more absorption and scattering, causing signal attenuation and resulting in a lower signal-to-noise ratio (SNR), which in turn increases noise in these deeper layers. Variations in tissue density and optical thickness further impact light penetration, leading to higher signal quality near the sample surface and more interference in deeper layers. These optical limitations affect the overall quality of 3D fluorescence imaging, particularly in deeper structures, where they lead to detail loss and increased noise. What's more, ground truth (GT) is hard to obtain. Without GT data for deep structures, accurately resolving these regions becomes even more challenging, as there is no reference for training models to reconstruct high-fidelity images. These optical limitations, combined with the lack of GT, affect the overall quality of 3D fluorescence imaging, particularly in deeper structures, leading to detail loss and increased noise.

Recent advances in deep learning have demonstrated the potential of generative models to enhance FM imaging, with diffusion models emerging as particularly effective for both denoising and super-resolution tasks, as described in the section of related work. Diffusion models are probabilistic generative approaches that learn to reverse a stochastic noise addition process, allowing them to iteratively refine noisy images. This is especially well-suited for FM imaging, where noise is pervasive, as these models can systematically model and remove noise while preserving biologically relevant details.

Our work contributes to this growing field by developing a algorithm that facilitates the ambiguous and low-resolution fluorescence images clear and visible. VTCD has been applied to study a variety of biological systems, providing insights into factors like cell signaling pathways, intercellular interactions, and the impact of genetic or environmental perturbations.

2. Additional Experimental Demonstrations

2.1. The 2D Quantitative Demonstration

To further validate the effectiveness of our cycle-consistent diffusion architecture, we show more details in the experiments focusing on 2D quantitative comparisons. The results on the XY, YZ, and XZ planes are presented in Figs. S1, comparing our VTCD approach with other methods, including CycleGAN, DSAR, and Neuroclear. Our method consistently outperformed these baselines across

[1]Department of Electrical Engineering, City University of Hong Kong, Hong Kong (SAR China)
[2]Centre for Intelligent Multidimensional Data Analysis, Hong Kong Science Park, Hong Kong (SAR China)
[3]Department of Biology, Hong Kong Baptist University, Hong Kong (SAR China)

*Equal contributions
†Correspondence: rogerhzk@gmail.com

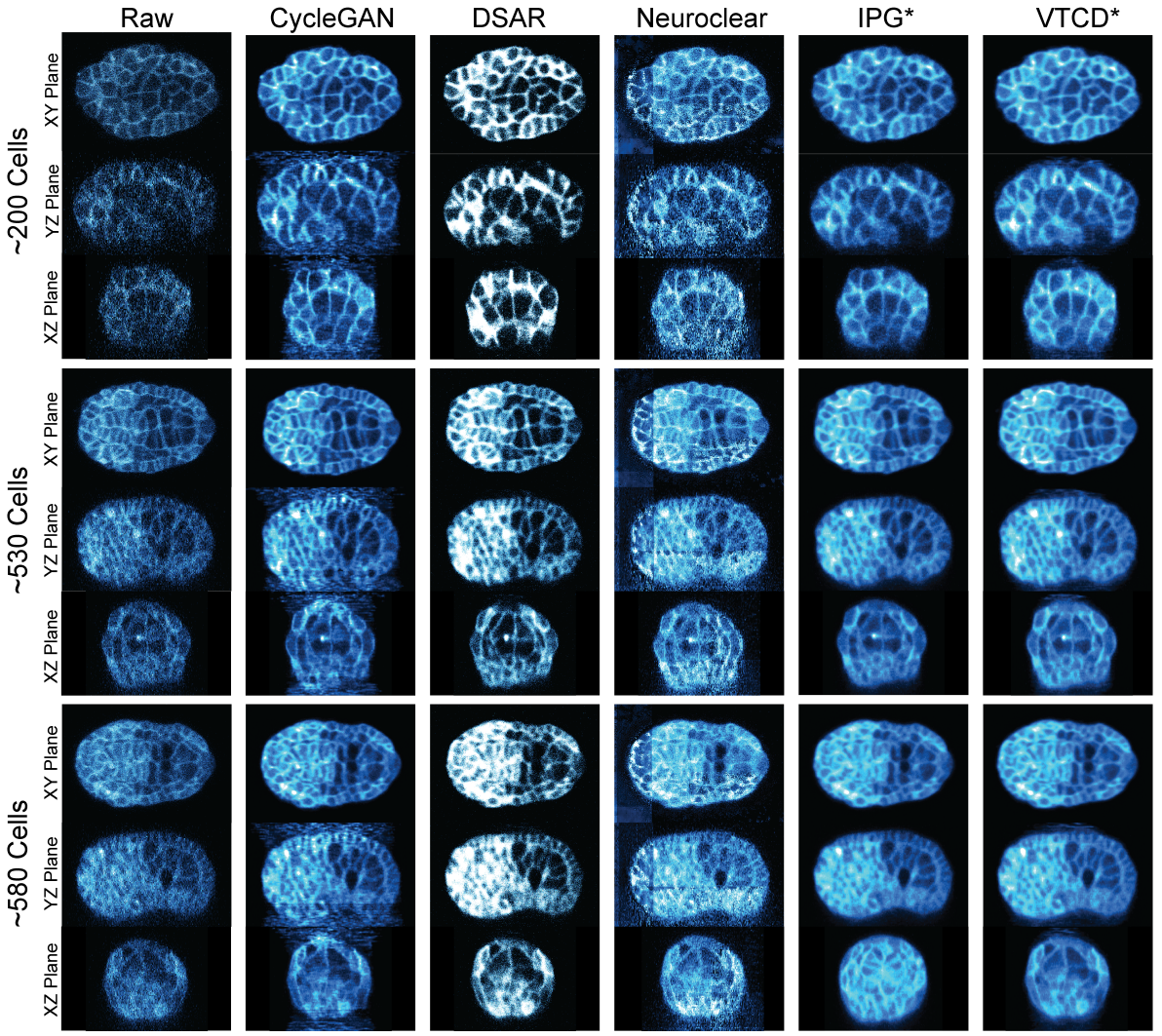

Fig. S1. The qualitative comparisons on fluorescence images between multiple methods (de-noising and super-resolution).

different cell stages (200, 530, and 580 cells), demonstrating enhanced super-resolution (SR) and de-noising capabilities. The results show significant noise reduction and a realistic enhancement of cell membrane outlines, without introducing excessive or artificial details. Notably, our VTCD approach performed well even in the XZ plane, where other methods, such as VTCD+IPG, tended to overemphasize certain features, leading to unrealistic details.

Enlarged viewings are available for detailed comparisons (Fig. 4 and Fig. S2), where improvements in membrane outlines and overall image clarity are apparent. The VTCD method effectively preserves cellular morphology while reducing noise, ensuring that essential structural features are retained. This capability is especially crucial for capturing accurate biological information in fluorescence microscopy data.

2.2. The 3D Quantitative Demonstration

We also show a series of 3D quantitative experimental results on dataset NorefZ-2 to further demonstrate the effectiveness of our approach. The comparisons were made between the original noisy images, the outputs from previous methods, and VTCD model (Fig. 3). As seen in the figure, our method was able to reduce noise and restore finer structural details, resulting in processed images' qualities.

In Fig. S3, we provide another set of comparisons highlighting the full workflow from the microscopy equipment (shown on the left) to the processed images. The visual improvements in 3D structural accuracy are evident, with our model achieving clearer outlines and a better representation of the cell boundaries. Unlike other methods that tend to introduce unwanted artifacts during de-noising or resolution enhancement, our VTCD model provides a consistent bal-

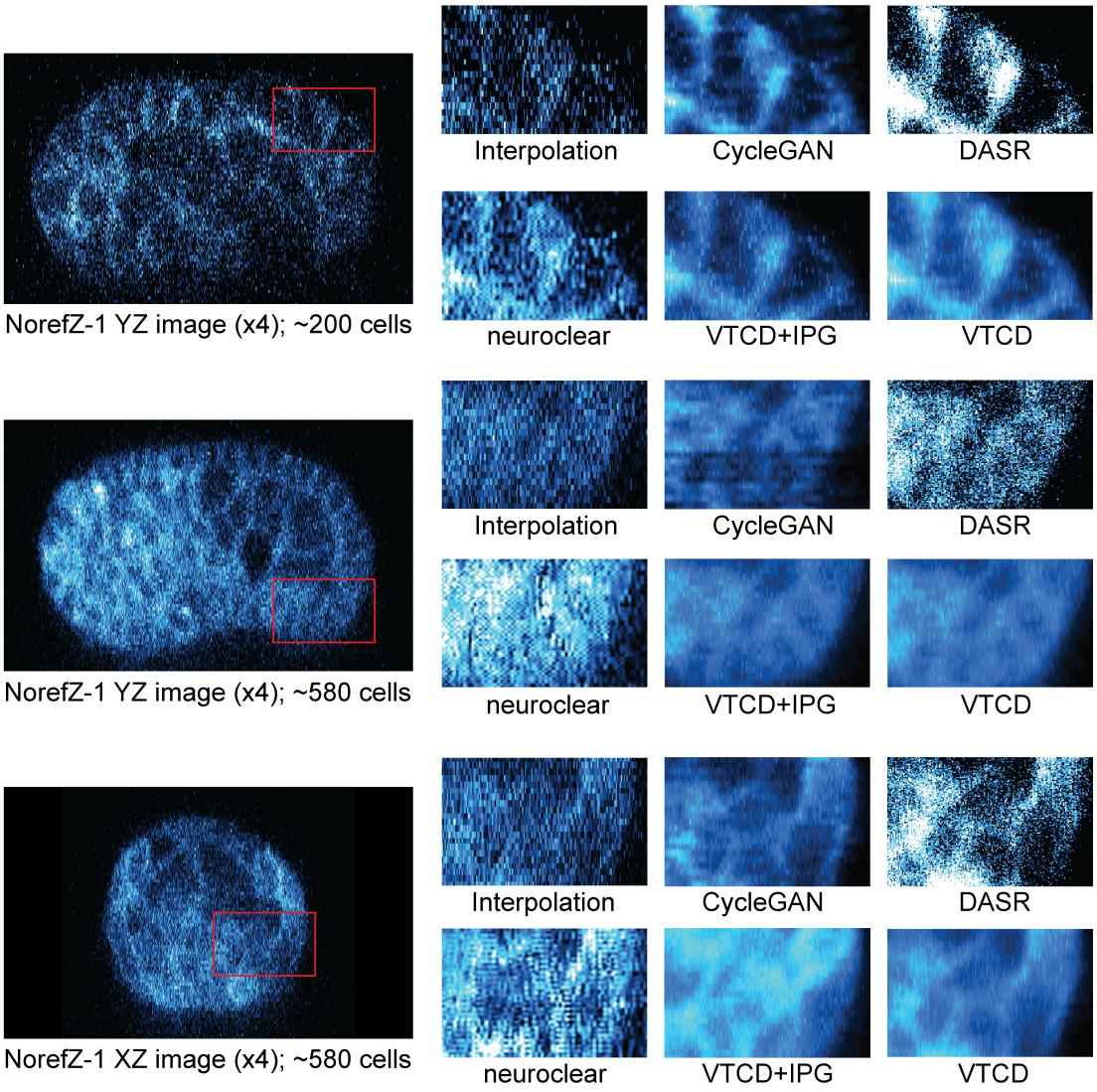

Fig. S2. In details, the qualitative comparisons on fluorescence images between multiple methods (de-noising and super-resolution).

ance between noise reduction and cell structure preservation.

The rendered 3D outputs demonstrate the efficiency of our cycle-consistent diffusion approach in enhancing fluorescence microscopy data. By reducing the noise while retaining essential morphological features, our method enables more accurate biological interpretation, which is crucial for studying developmental processes at the cellular level. This makes VTCD a robust tool for handling complex 3D fluorescence microscopy data, ensuring high-resolution results even in challenging imaging conditions.

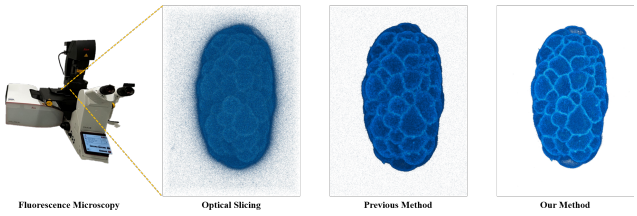

Fig. S3. The 3D qualitative comparisons on fluorescence images of original images, previous method result and our method output.

3. Methodology Supplementary Descriptions

To help reverse the cell structures in ambiguous FM images and utilize the VTBC's generalization and accuracy,

we designed a integrated loss function and train our method in a progressive way.

The training process for our dual cycle-consistent diffusion model, specifically designed for de-noising and super-resolution (SR) in 3D fluorescence microscopy, is structured to address both noise suppression and resolution enhancement. In each iteration, we begin with a sequential training approach where the model is first trained on de-noising tasks, focusing on accurately removing noise while preserving critical features in low-resolution images. In this phase, the model learns noise patterns and low-frequency details, forming a solid foundation for subsequent SR tasks. After stabilizing the de-noising phase, we proceed to train the model for SR, where it learns to recover high-frequency details from downsampled images. Once both tasks are effectively learned independently, we enable dual cycle-consistent training, in which de-noising and SR tasks are jointly optimized. The model employs diffusion-based sampling to iteratively improve image quality by propagating high-quality features across cycles. We also incorporate progressive up-scaling and multi-scale feature discriminators to ensure the SR network captures spatial details while maintaining fidelity across scales, making it particularly effective for 3D fluorescence microscopy data.

To train our VTCD, the basic adversarial losses and the cycle consistent losses are first used to enable the cycle training. With the diffusion forward stage formulated as the targeted slicing of 3D cell volume, the diffusion loss is integrated into the cycle consistency loss by changing the form of the generated process. By the way, to control the generation trajectory, another two losses are formed in the training of the denoising and SR. The total loss is formulated as follows:

$$\mathcal{L}_{\text{VTDC}} = \mathcal{L}^{G_A}_{\text{LR}\to\text{HR}} + \mathcal{L}^{G_B}_{\text{HR}\to\text{LR}} + \mathcal{L}^{\text{De-noise}}_{x_t\to\hat{x_t}} + \mathcal{L}^{\text{SR}}_{\text{XZ/YZ}\to\text{XY}} \tag{1}$$

Adversarial loss $\mathcal{L}^{G_A}_{\text{LR}\to\text{HR}}$ and $\mathcal{L}^{G_B}_{\text{HR}\to\text{LR}}$ encourages realistic outputs, enabling the model to produce clean, high-resolution images that are indistinguishable from high-quality fluorescence images. Cycle consistent loss for de-noising $\mathcal{L}^{\text{De-noise}}_{x_t\to\hat{x_t}}$ is consist of $\mathcal{L}^{\text{De-noise}}_{x_t\to x_0} + \mathcal{L}_{TV}(\hat{I})$, while $\mathcal{L}_{TV}(\hat{I})$ is defined as:

$$\mathcal{L}_{TV}(\hat{I}) = \frac{1}{hwc}\sum_{i,j,k}\sqrt{(\hat{I}_{i,j+1,k}-\hat{I}_{i,j,k})^2 + (\hat{I}_{i+1,j,k}-\hat{I}_{i,j,k})^2} \tag{2}$$

$\mathcal{L}^{\text{SR}}_{\text{XZ/YZ}\to\text{XY}}$ is the cycle consistent loss for SR conditioned with $\mathcal{L}_{content}(\hat{I}, I; \phi, l)$. In the reverse process of diffusion, to control the generation trajectory, the following losses are used to formulate the conditioning diffusion model. $\mathcal{L}_{content}$ is used in the latent space diffusion model to accurately transform/diffuse the Z-axis images to clear levels and is defined as:

$$\mathcal{L}_{content}(\hat{I}, I; \phi, l) = \frac{1}{h_l w_l c_l}\sqrt{\sum_{i,j,k}(\phi^{(l)}_{i,j,k}(\hat{I}))^2 - \phi^{(l)}_{i,j,k}(I)^2)} \tag{3}$$

To ensure optimal balance, we dynamically adjust loss weights as the model transitions from denoising-focused to SR-focused stages, refining the fidelity and clarity of the final super-resolved, de-noised outputs.

Code and Data

The code is deposited on https://github.com/temporarysharing1688/VTCD and the data is available upon request or after revision.