

Fast super-resolution optical fluctuation imaging using a transformer-optimized neural network

Zitong Ye,^a Yuran Huang,^a Hanchu Ye,^a Enxing He,^a Yile Sun,^a Haoyu Zhou,^a Xin Luo,^a Yubing Han,^{a,c,*} Cuifang Kuang,^{a,b,d,*} and Xu Liu^{a,b}

^aState Key Laboratory of Extreme Photonics and Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou, China

^bZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, China

^cWuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China

^dCollaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, China

Abstract. Super-resolution optical fluctuation imaging (SOFI) achieves super-resolution (SR) imaging through simple hardware configurations while maintaining biological compatibility. However, the realization of large field-of-view (FoV) SOFI imaging remains fundamentally limited by extensive temporal sampling demands. Although modern SOFI techniques accelerate the acquisition speed, their cumulant operator necessitates fluorophores with a high duty cycle and a high labeling density to ensure that sufficient blinking events are acquired, which severely limits the practical implementation. For this, we present a novel framework that resolves the trade-off in SOFI by enabling millisecond-scale temporal resolution while retaining all the merits of SOFI. We named the framework the transformer-based reconstruction of ultra-fast SOFI (TRUS), a novel architecture combining transformer-based neural networks with physics-informed priors in conventional SOFI frameworks. For biological specimens with diverse fluorophore blinking characteristics, our method enables reconstruction using only 20 raw frames and the corresponding widefield images, which achieves a 47-fold reduction in raw frames (compared to the traditional methods that require more than 1000 frames) and sub-200-nm spatial resolution capability. To demonstrate the high-throughput SR imaging ability of our method, we perform SOFI imaging on the microtubule within a millimeter-scale FoV of 1.0 mm² with total acquisition time of ~3 min. These characteristics enable TRUS to be a useful high-throughput SR imaging alternative in challenging imaging conditions.

Keywords: fluorescence microscopy; super resolution; deep learning; physics information.

Received May 15, 2025; revised manuscript received Jun. 23, 2025; accepted Jul. 3, 2025; published online Jul. 30, 2025.

© The Authors. Published by Chinese Laser Press under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.3788/AI.2025.10011](https://doi.org/10.3788/AI.2025.10011)]

1. Introduction

Fluorescence microscopy has revolutionized biological imaging by enabling non-invasive visualization of subcellular structures^[1,2], yet its spatial resolution is fundamentally constrained by the diffraction limit (~200 nm). This barrier has been broken by super-resolution (SR) microscopy, which has undergone rapid development in recent decades^[3,4]. Among these techniques, single-molecule localization microscopy (SMLM) achieves nanoscale resolution (<20 nm) by temporally isolating sparsely

activated fluorophores and precisely localizing their centroids^[5-7]. While SMLM techniques have provided unprecedented insights into biomolecular organization, their practical applications are hindered by stringent requirements on the preparation of biological samples^[8-10], such as high photon requirements, certain photo-switching kinetics, and balanced label density.

Super-resolution optical fluctuation imaging (SOFI) has emerged as a powerful technique with higher universal applicability and minimalist optical configuration, enabling nanoscale visualization of biological structures through temporal higher-order statistical analysis^[11-14]. However, conventional SOFI implementations typically require hundreds to thousands of frames to achieve reliable SR reconstruction by high-order

*Address all correspondence to Cuifang Kuang, cfkuang@zju.edu.cn; Yubing Han, hanyubing@zju.edu.cn

cumulant calculations, imposing fundamental limitations in imaging speed^[15,16]. In the context of accelerated SOFI processing, Vandenberg *et al.* utilized the uncertainty quantification mechanism to dynamically adjust virtual pixel weights based on their intensity variance distributions. The optimization enables a twofold reduction in fluorescence image acquisition requirements^[17]. Jiang *et al.* utilized two wavelet-based filters in the temporal and spatial domains of the raw image stack to improve the fluctuating signal extraction capability, achieving twofold spatial resolution enhancement by reconstructing from 25 raw frames^[18]. Recent fast-SOFI approaches, such as SACD, have attempted to reduce frame counts through a post-deconvolution step, which has made significant progress in SOFI imaging^[19]. However, the reconstruction performance of these approaches is limited by the availability of sufficient blinking events. Consequently, these requirements constrain the duty cycle of the fluorophore and the label density of the sample for high-quality reconstruction. The reliance on specific labels such as quantum dots or photoswitchable fluorescent proteins further demands significant expertise, sacrificing the initial benefits of SOFI concerning sample accessibility^[20,21].

Recent advances in deep learning have significantly accelerated progress in SR fluorescence microscopy. Deep learning approaches primarily enhance temporal resolution by reducing the number of raw data frames required for reconstruction, thereby enabling faster overall acquisition. For instance, structured illumination microscopy (SIM) has utilized neural networks with non-uniform illumination patterns to achieve high-quality reconstruction from a few frames^[22–24]. Furthermore, deep learning techniques have been employed to refine results from SMLM with insufficient data^[25] or dense emitter^[26,27] acquisition, accelerating the imaging process while maintaining satisfactory quality. These applications offer promising alternatives for achieving fast SOFI. However, current state-of-the-art methods predominantly rely on conventional convolutional neural networks (CNNs). These CNN-based networks demonstrate limitations in accurately completing missing information within SOFI reconstructions, primarily due to their inherent local receptive fields imposed by convolutional kernels. This architectural constraint can lead to a loss of fine features in the reconstructed images.

To solve the limitation mentioned above, we present a paradigm-shifting framework that combines deep neural networks with SOFI fundamentals to enable robust and universal SR imaging from ultra-sparse temporal samplings (21 frames). Drawing inspiration from ANNA-PALM^[25] and self-supervised learning-assisted SOFI^[28], we develop a hybrid architecture named transformer-based reconstruction of ultra-fast SOFI (TRUS) that synergizes physics-based prior knowledge with data-driven feature learning. The proposed TRUS employs a dual-path design: 1) A physics-regularized branch that encodes domain-specific knowledge of SOFI point spread function (PSF) properties, and 2) a transformer-enhanced convolutional branch that learns spatial correlations from sparse SOFI imaging reconstructed from only 20 frames and an additional widefield image. Through training with SOFI datasets, the network learns to reconstruct high-fidelity SR images while effectively recovering missing information caused by extreme temporal under-sampling.

Experimental validation demonstrates unprecedented performance, achieving at least 1.28-fold spatial resolution enhancement with a 47× reduction in acquisition frames (21 frames

versus 1000 frames) compared to conventional second-order SOFI requirements. Systematic evaluations on simulated data demonstrated that our framework achieves 0.4–1.5 dB peak signal-to-noise ratio (PSNR) gain compared to previous fast-SOFI pipelines. Additionally, the algorithm maintained spatiotemporal resolution enhancement across tested biological variations with a single training protocol, suggesting possible structural generalization capabilities that merit further investigation. Leveraging the enhanced temporal resolution of TRUS, we achieved high-throughput SR imaging of microtubules, covering a 1 mm² field-of-view (FoV) in approximately 3 min. Furthermore, the system demonstrated efficiency in dual-color imaging, capturing a 0.25 mm² FoV of cytoskeletal architecture in just 1.5 min. The proposed methodology shows significant potential for deployment in modern high-throughput biological experimentation systems.

2. Methods

2.1. Principles of SOFI and System Setup

In conventional imaging approximations, a biological specimen is typically modeled as a collection of N discrete, independently fluctuating emitters, spatially localized at positions r_k . Each emitter exhibits time-varying molecular brightness characteristics, with the cumulative fluorescence signal detected at spatial coordinate r and temporal point t being mathematically expressed as

$$I(r, t) = \sum_{k=1}^N h(r - r_k) \times s_k(t). \quad (1)$$

In this framework, h and s correspond to the microscope's PSF and the temporally modulated fluctuation intensity of blinking emitters, respectively. Building on the foundational approach of SOFI—a methodology that evaluates pixel-wise temporal correlation cumulants along the t axis to achieve resolution enhancement—specifically, computation of the second-order temporal cumulant G_2 under τ time-lag conditions yields

$$G_2(r, \tau) = \langle \delta I(r, t) \cdot \delta I(r, t + \tau) \rangle_t, \quad (2)$$

where $\delta I(r, t)$ is defined as

$$\delta I(r, t) = I(r, t) - \langle I(r, t) \rangle_t. \quad (3)$$

By substituting Eq. (1) into Eq. (2), we obtain the following modified expression:

$$G_2(r, \tau) = \sum_{i,k} h(r - r_i) \times h(r - r_k) \times \langle \delta s_i(r, t) \cdot \delta s_k(r, t + \tau) \rangle_t. \quad (4)$$

Under the premise of uncorrelated molecular fluctuations (i.e., statistically independent temporal variations among individual molecules), the cross-correlation components in the governing equation vanish for all $i \neq k$. Consequently, the second-order temporal cumulant reduces to a linear superposition of squared PSF terms, each scaled by its respective second-order temporal cumulant function $c_i(\tau)$. Formally, this yields

$$\begin{aligned}
 G_2(r, \tau) &= \sum_i h(r - r_i)^2 \times \langle \delta s_i(r, t) \cdot \delta s_k(r, t + \tau) \rangle_t \\
 &= \sum_i h(r - r_i)^2 \times c_i(\tau).
 \end{aligned} \quad (5)$$

The effective spatial extent of the reconstructed PSF in this formulation undergoes a $\sqrt{2}$ -fold compression through temporal cumulant analysis [Fig. 1(a)].

The experimental setup schematic is depicted in Fig. 1(b), featuring a dual-color imaging system employing 488 and 640 nm excitation lasers. These wavelength-specific beams are combined through a series of optical components for

simultaneous sample excitation. The 488 nm laser beam is first redirected by a mirror M1 and subsequently combined with the 640 nm beam using a dichroic mirror D1. The co-aligned beams are then coupled into a single-mode fiber for spatial filtering.

Following fiber output, the combined beam undergoes expansion and collimation to achieve optical-axis parallelism before being focused onto the back focal plane of the objective lens.

The subsequent fluorescence emission is collected by the same objective and is spectrally separated using a dichroic mirror D3 before detection. An optical component in the system, dichroic mirror D2, serves two functions: reflecting residual excitation light and efficiently transmitting the fluorescence

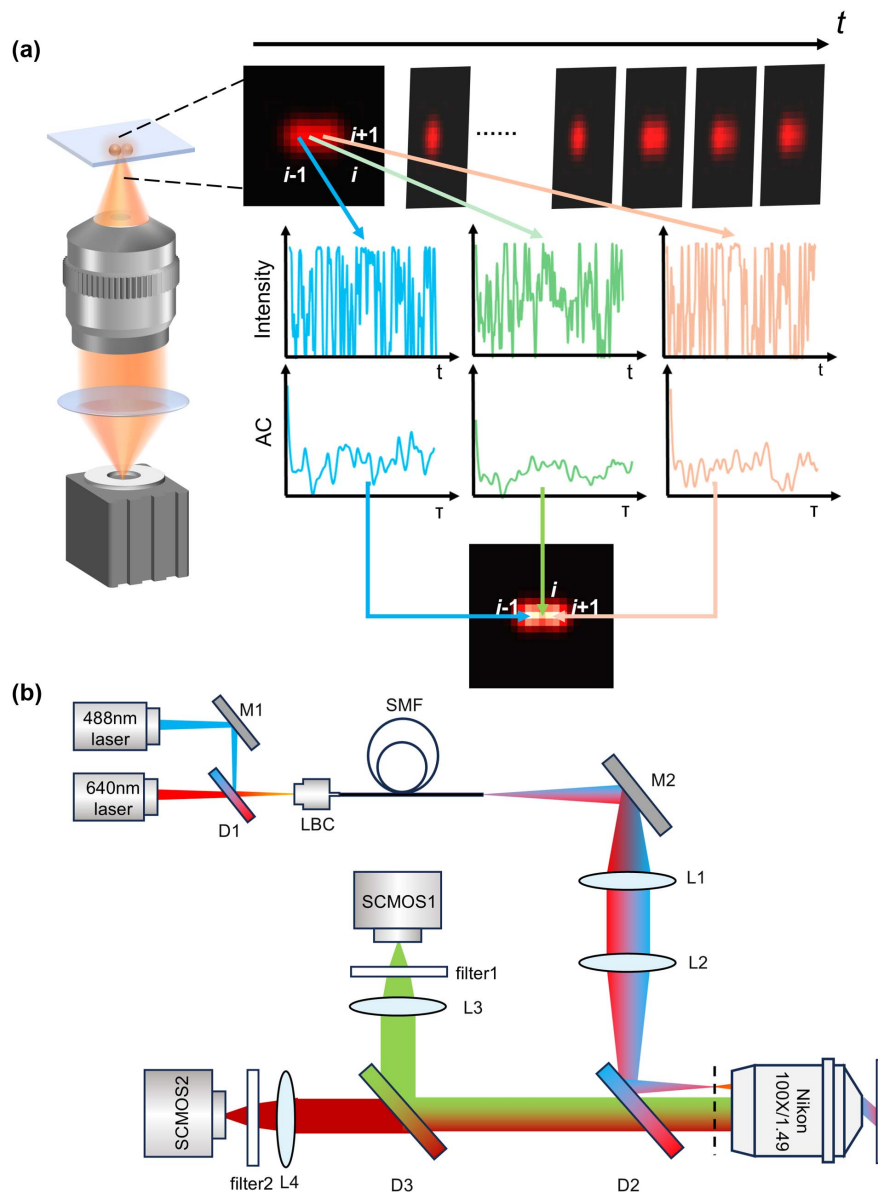


Fig. 1 (a) Schematic showing the principle of super-resolution optical fluctuation imaging. The image shows an example of two adjacent emitters, which could not be resolved due to the optical diffraction limit. By computing the autocorrelation (AC) result of intensity for each pixel with time-lags τ , two fluorophores could be distinguished successfully. (b) The diagram of our experimental setup: M1-M2, mirrors; D1-D3, dichroic mirrors; L1-L4, lenses; LBC, laser beam coupler; SMF, single-mode fiber.

signals. This spectral separation enables simultaneous imaging through two distinct detection channels, each equipped with cameras for wavelength-specific signal acquisition.

2.2. Workflow of the TRUS Framework

The temporally sparse SOFI reconstruction challenge essentially constitutes a specialized image completion problem within the computational imaging paradigm. Drawing inspiration from

the ANNA-PALM framework^[25], we propose a hypothesis: Temporally sparse SOFI acquisitions integrated with widefield intensity measurements contain inherent spatiotemporal correlations that enable robust SR reconstruction via optimized neural network architectures called TransNet, which combines the transformer and CNN. A physics-informed regularization module is systematically integrated into the network training paradigm to ensure generalizability. This hybrid architecture synergistically combines data-driven learning with domain-specific physical

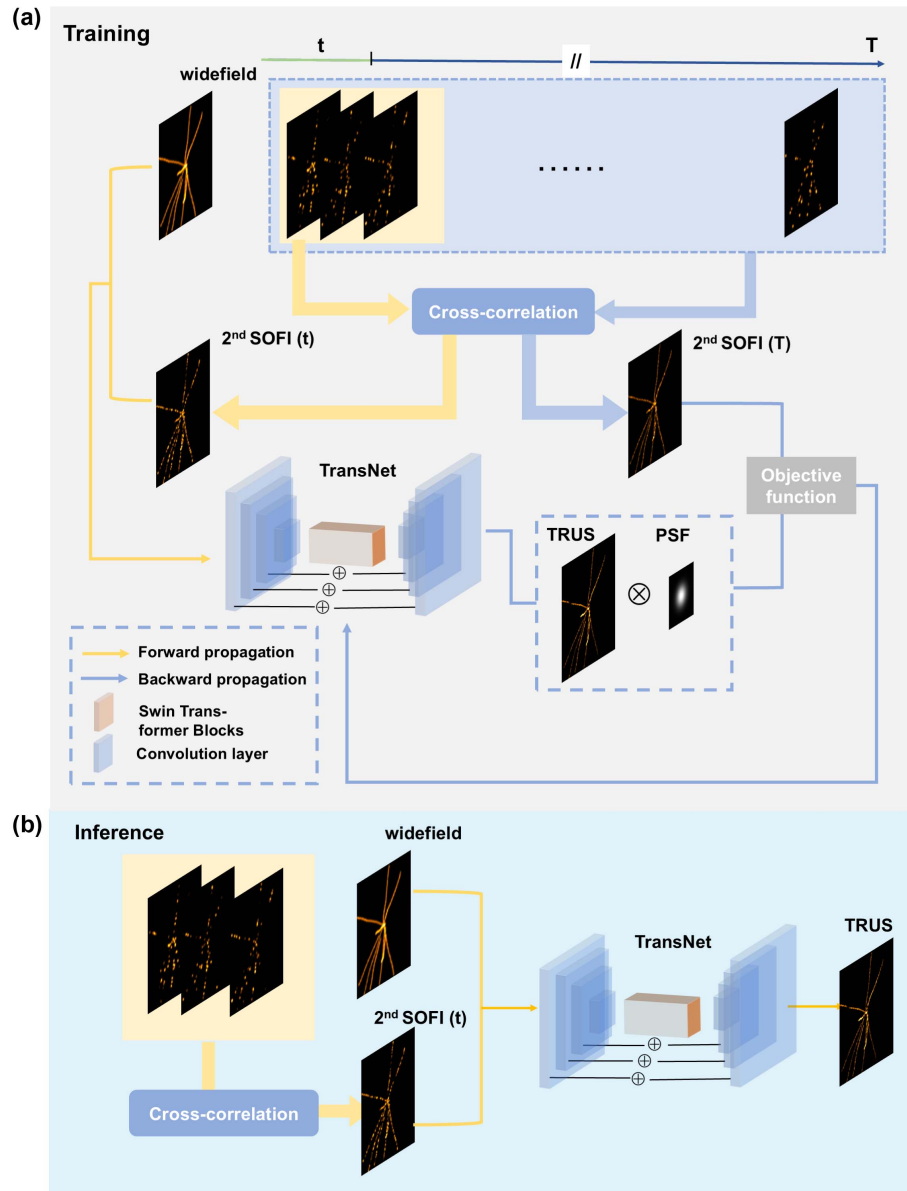


Fig. 2 Workflow of TRUS framework. (a) Schematic showing the TRUS training procedure. In the forward propagating process, a sparse SOFI image (reconstructed from 20 frames) and an additional widefield image are first fed into the network to generate an initial output image (TRUS), and then the weights of the network are updated according to the partial derivative of the objective function, and an iteration is complete. This iteration continues until the minimum value of the objective functions is obtained. (b) In the inference procedure, the optimal model weights are loaded into the network in advance. The SOFI image (reconstructed from 20 frames) and corresponding widefield image are acquired and fed into the network for TRUS reconstruction.

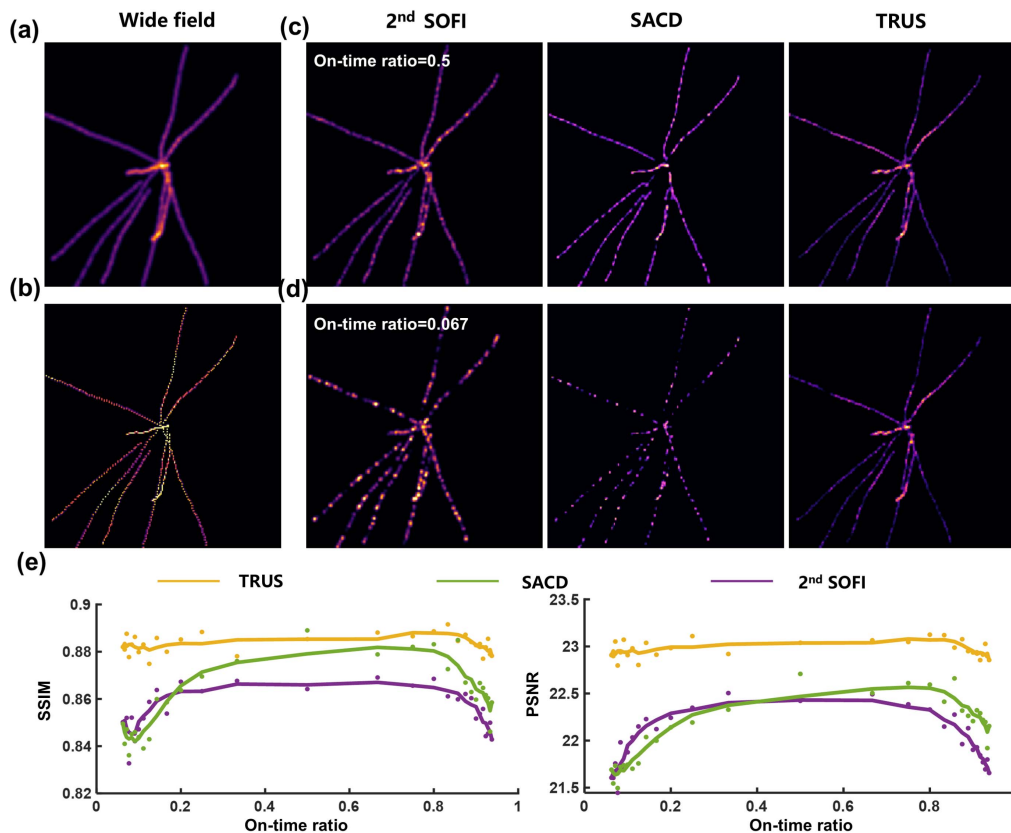


Fig. 3 Reconstruction comparison of different test targets with the second-order SOFI, SACD, and TRUS framework. (a) Simulated widefield image and (b) corresponding ground-truth image. (c) Results reconstructed by second-order SOFI, SACD, and TRUS at 0.5 on-time ratio (defined as the proportion of time fluorophores remain in the active state). (d) Results reconstructed by second-order SOFI, SACD, and TRUS at 0.067 on-time ratio (defined as the proportion of time fluorophores remain in the active state). (e) The values of the structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) of TRUS (yellow line), SACD (green line), and second-order SOFI (purple line) versus on-time ratio.

priors, enabling simultaneous preservation of structural fidelity and robust spatiotemporal resolution enhancement across heterogeneous imaging conditions.

The workflow of the proposed TRUS framework is demonstrated in Fig. 2. The input of the network is a widefield image combined with the corresponding sparse SOFI image reconstructed from 20 frames. The ground-truth image is the conventional second-order SOFI reconstructed from 3000 frames. The synergistic integration of the widefield with sparse SOFI image has been demonstrated to achieve superior reconstruction of intricate subcellular features relative to other initializations (details in the [Supplement 1](#), Note 1). The following network structure is a typical encoder–decoder architecture with transformer blocks (described in Sec. 2.3). For high-fidelity SOFI reconstruction, we developed a dual-domain objective function that synergistically integrates data-driven reconstruction fidelity with physical prior regularization, effectively preserving resolution enhancement while maintaining nanoscale structural accuracy (described in Sec. 2.3).

2.3. Network Architecture Design and Training

Conventional CNNs suffer from locality, spatial-sharing, and static parameters, which cannot retrieve lost features from

non-local features^[29]. Compared to CNNs, the transformer shows its superiority in image restoration tasks because its attention operators are better able to capture long-range dependencies through explicit interaction with global features. Inspired by the shifted window (swin) transformer^[30], we integrate a swin-transformer block into the CNN-based encoder–decoder architecture to further improve the performance of CNN. The network called TransNet consists of four spatial squeeze blocks and four spatial excitation blocks connected by two swin-transformer blocks (details shown in Fig. S1 in the [Supplement 1](#)).

To validate the performance of the network, we conducted a comprehensive comparison between TransNet and the classical U-Net architecture using SIM as a reference. The comparison analysis in Fig. S2 in the [Supplement 1](#) reveals that U-Net reconstructions exhibited limitations in feature preservation, with pronounced structural detail loss compared to ground-truth images. In contrast, TransNet demonstrates superior performance in maintaining biological ultrastructure fidelity while effectively suppressing reconstruction artifacts.

In our dataset, the image triplets consist of 1) widefield microscopy images, 2) corresponding sparse SOFI reconstructed from 20-frame sequences using bSOFI processing, and 3) conventional second-order SOFI images reconstructed from 3000-frame sequences using the same bSOFI algorithm^[31].

A total of 20 FoV acquisitions at 512 pixel \times 512 pixel resolution are collected. For data augmentation, these image sets undergo random cropping and rotation operations, generating 600 sub-image pairs with 128 pixel \times 128 pixel dimensions.

The objective function is defined as

$$\arg \min \{ \text{MAE}(f \otimes \text{sofiPSF}, y) + \alpha R_{\text{Hessian}}(f) \}, \quad (6)$$

where f is the output of the network and y represents the ground truth. The mean absolute error (MAE) is widely used to reconstruct structure details. The sofiPSF used in Eq. (6) is the equivalent second-order SOFI PSF. The PSF has the same cut-off frequency f_c with the second-order SOFI image calculated from decorrelation analysis^[32]. The resolution calculated based on the cut-off frequency and the Abbe criterion should be equivalent. This equivalence can be leveraged to determine the numerical aperture (NA),

$$\frac{2 \cdot \text{pixelsize}}{f_c} = \frac{0.5\lambda}{\text{NA}}. \quad (7)$$

The PSF is simulated using the Bessel function of the first kind. Further narrowing the difference between the convoluted network output and conventional second-order SOFI imaging could improve resolution. The Hessian matrix corresponds to the continuity constraint^[33] with a scale factor α empirically set to 0.1.

The Hessian regularization is defined as

$$R_{\text{Hessian}}(f) = \sum_r |f_{xx}(r)|^2 + |f_{yy}(r)|^2 + 2 \cdot |f_{xy}(r)|^2, \quad (8)$$

where $|f_{xx}(r)|$, $|f_{yy}(r)|$, and $|f_{xy}(r)|$ are defined as the second-order difference along the x axis, y axis, and x - y axis, respectively.

Network training is specifically performed on microtubule-containing samples using a learning rate of 1×10^{-4} , with optimization continuing for 30,000 minibatch iterations until model convergence was achieved. Network training and inference are performed on RTX 2080Ti graphical processing units (GPUs) from NVIDIA. Once trained, the TransNet takes only ~ 2 s or less to reconstruct an SR image of 512 pixel \times 512 pixel (corresponding to an entire FoV).

3. Results

3.1. Validating the Performance of TRUS on Simulated and Experimental Images

As a first demonstration of the TRUS method with robustness, we simulated wire-like samples labeled by fluorescent dyes with different on-time ratios (defined as the proportion of time fluorophores remaining in the active state). The simulations were based on the previously published ‘‘SOFI Simulation Tool’’ software package^[34]. Our simulations (details in the [Supplement 1](#), Note 3) spanned an extensive on-time ratio spectrum from 0.069 to 1.0 to simulate various type of fluorophores, employing a labeling density (defined as the number of emitters in a micron square micron)^[8] of $5/\mu\text{m}^2$ with 663 emitters in $133 \mu\text{m}^2$, with the pixel size set to 60 nm with 1.2 NA. As demonstrated in Fig. 3(c), both second-order SOFI and SACD techniques successfully reconstruct the complete nanostructure at an intermediate on-time ratio of 0.5. However, as shown in Fig. 3(d), their performance significantly deteriorates under suboptimal

conditions, failing to resolve the full structural details when the on-time ratio dropped to 0.067. In contrast, TRUS successfully recovers the flattened structure in both experimental settings. Quantitative evaluations using the structural similarity index measure (SSIM) and PSNR reveal distinct performance trends: conventional second-order SOFI and SACD exhibit marked degradation in reconstruction quality at both extremes of the on-time ratio range (0.069–0.2 and 0.8–1). In contrast, TRUS consistently preserves high reconstruction fidelity across the full range (SSIM = 0.88, PSNR = 23 dB), directly demonstrating its robustness to diverse fluorophore blinking behaviors [Fig. 3(e)]. Furthermore, TRUS outperformed second-order SOFI across a broad range of label densities (from $89/\mu\text{m}^2$ to $9/\mu\text{m}^2$; see details in Fig. S3 in the [Supplement 1](#)), indicating its robustness in biological samples with varied labeling conditions.

To further validate our approach, we conducted a comparative analysis of SACD, conventional second-order SOFI reconstructed from 3000 frames, and TRUS against SIM using outer mitochondrial membrane specimens labeled with Alexa Fluor 647. High-fidelity reference data were acquired through nine-phase high-fidelity SIM^[35] (HIFI-SIM) imaging, followed by 647 nm laser excitation at 2 kW/cm² intensity in a conventional STORM imaging buffer^[36]. This dual-mode acquisition protocol ensures a direct comparison of SR performance under identical biological and optical conditions. Figure 4 reveals that cumulant operator-based fast SOFI methods, such as SACD, fail to reconstruct the full structure from only 20 frames when benchmarked against reference standards. Quantitative evaluation using PSNR and gradient magnitude similarity deviation (GMSD) metrics^[37] shows a performance differential: TRUS achieves the highest PSNR (26.3 dB), comparable to that of fourth-order SOFI (23.3) and SACD (24.5 dB) reconstructed from 3000 frames. In contrast, SACD reconstructed from only 20 frames failed to resolve the structure, yielding a significantly lower PSNR of 19.1 dB. GMSD analysis, which assesses structural information preservation, indicates that TRUS best maintains biological feature fidelity, achieving the lowest (best) score (0.106). We further assessed spatial resolution using Fourier ring correlation (FRC) and decorrelation analysis on SOFI, SACD, TRUS, and HIFI-SIM results (see Fig. S4 in the [Supplement 1](#) for details). The results show that TRUS reconstructed from 20 frames provides enhanced resolution compared to second-order SOFI and achieves spatial resolution equivalent to other methods (SACD, fourth-order SOFI, and HIFI-SIM).

3.2. Quantitative Analysis of TRUS Super-Resolution Capabilities in Different Structures

To quantitatively characterize the resolution enhancement of TRUS, we performed Fourier spectral analysis in Fig. S4 in the [Supplement 1](#) comparing four imaging modalities: widefield microscopy, second-order SOFI, SIM, and TRUS. Frequency domain quantification reveals that TRUS preserved the high-frequency spectral amplitude relative to SIM, while widefield and SOFI exhibited progressive attenuation. Notably, TRUS (181 nm) successfully maintains spatial frequencies with the same capability as SIM (183 nm), while second-order SOFI exhibits reduced spatial resolution (241 nm), highlighting its relative limitation in this aspect.

Subsequently, we systematically evaluated the superior performance of TRUS through comprehensive analyses employing

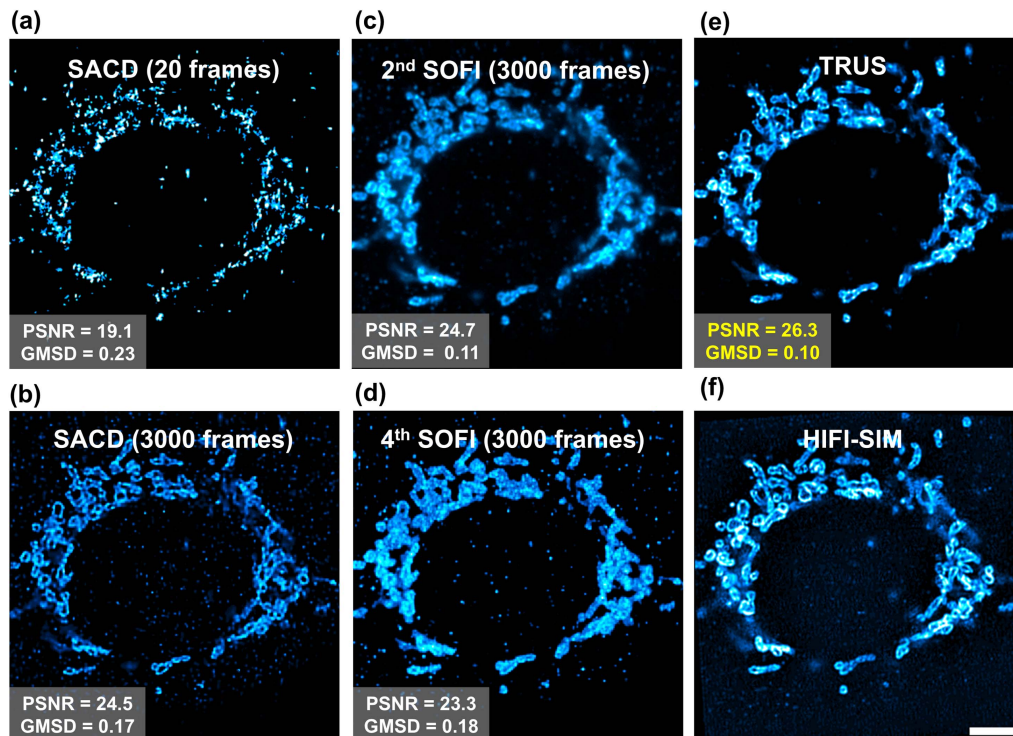


Fig. 4 Comparison of (a) SACD reconstructed from 20 frames, (b) SACD reconstructed from 3000 frames, (c) second-order SOFI reconstructed from 3000 frames, (d) fourth-order SOFI reconstructed from 3000 frames, and (e) TRUS, with (f) HIFI-SIM as a reference. Scale bar: 5 μm .

more complex and biologically representative samples, including outer mitochondrial membrane (labeled with Alexa Fluor 647, details in the [Supplement 1](#), Note 4), actin filaments (labeled with Alexa Fluor 488), and microtubules (labeled with Alexa Fluor 647) in Figs. 5(a)–5(c). All the biological samples were excited using 488 or 640 nm lasers with 1–2 kW/cm^2 illumination intensities in the STORM imaging buffer. The corresponding intensity profiles demonstrate that TRUS achieves sub-200-nm spatial resolution, enabling precise visualization of subcellular architectures. Comparative analysis indicates that TRUS significantly enhanced imaging quality compared to conventional second-order SOFI, consistently achieving superior structural resolution with enhanced detail preservation and continuity across all experimental conditions. For instance, our methodology successfully resolves the characteristic hollow morphology of outer mitochondrial membrane through processing of a limited dataset comprising merely 20 frames and a corresponding widefield image using the TRUS algorithm [Fig. 5(a)]. Furthermore, the decorrelation analysis^[32] in Fig. 5(d) statistically corroborates that TRUS could extend the spatial resolution of the widefield image twofold across tested biological variations with a single training protocol.

3.3. Fast Dual-Color and 3D Imaging by TRUS

To experimentally validate the temporal resolution of TRUS, we imaged two distinct subcellular structures of fixed COS-7 cells. The outer mitochondrial membrane was labeled with Alexa Fluor 488, and microtubules were labeled with Alexa Fluor 647 (details in the [Supplement 1](#), Note 4). The fluorophores were sequentially excited using 488 and 640 nm lasers

with respective illumination intensities of 1 and 2 kW/cm^2 . A series of 20 frames were captured at 20 ms exposure time per frame, accompanied by a single widefield reference frame obtained with 50 ms exposure and an extended acquisition of 3000 frames at 20 ms exposure time from the same region for subsequent second-order SOFI analysis. Based on the conventional second-order SOFI reconstruction (reconstructed from 3000 frames), the observation of the whole structure of the microtubule and outer mitochondrial membrane is impossible. While the TRUS results all display sharp and continuous filaments and clearly reveal many structural details of membranes with 142-fold temporal resolution enhancement (21 frames against 3000 frames), TRUS provides additional resolution enhancing performance compared to conventional second-order SOFI. As shown in Fig. 6(b), with the assistance of the TRUS, hollow structures are clearly resolved and well maintained, while second-order SOFI could not provide sub-200-nm details. Furthermore, we extend the TRUS validation to dual-color specimens comprising immunolabeled actin and microtubule networks. Co-localized imaging analysis demonstrates TRUS's capability to resolve adjacent actin filaments that remain unresolved in second-order SOFI reconstruction, while preserving structural continuity (Fig. S5 in the [Supplement 1](#)).

Additionally, we conducted z -stack acquisition spanning the entire cytoplasmic volume of COS-7 cells (0–5 μm axial range) to assess the volumetric imaging capabilities of TRUS. This evaluation reveals a significant temporal advantage [Fig. 6(d)]: TRUS completes volumetric acquisition in 2.25 s, whereas second-order SOFI required 100 s acquisition time (20 ms/frame exposure, 1000-frame dataset) for equivalent spatial sampling, achieving 44-fold acceleration. As demonstrated in Fig. 6(e),

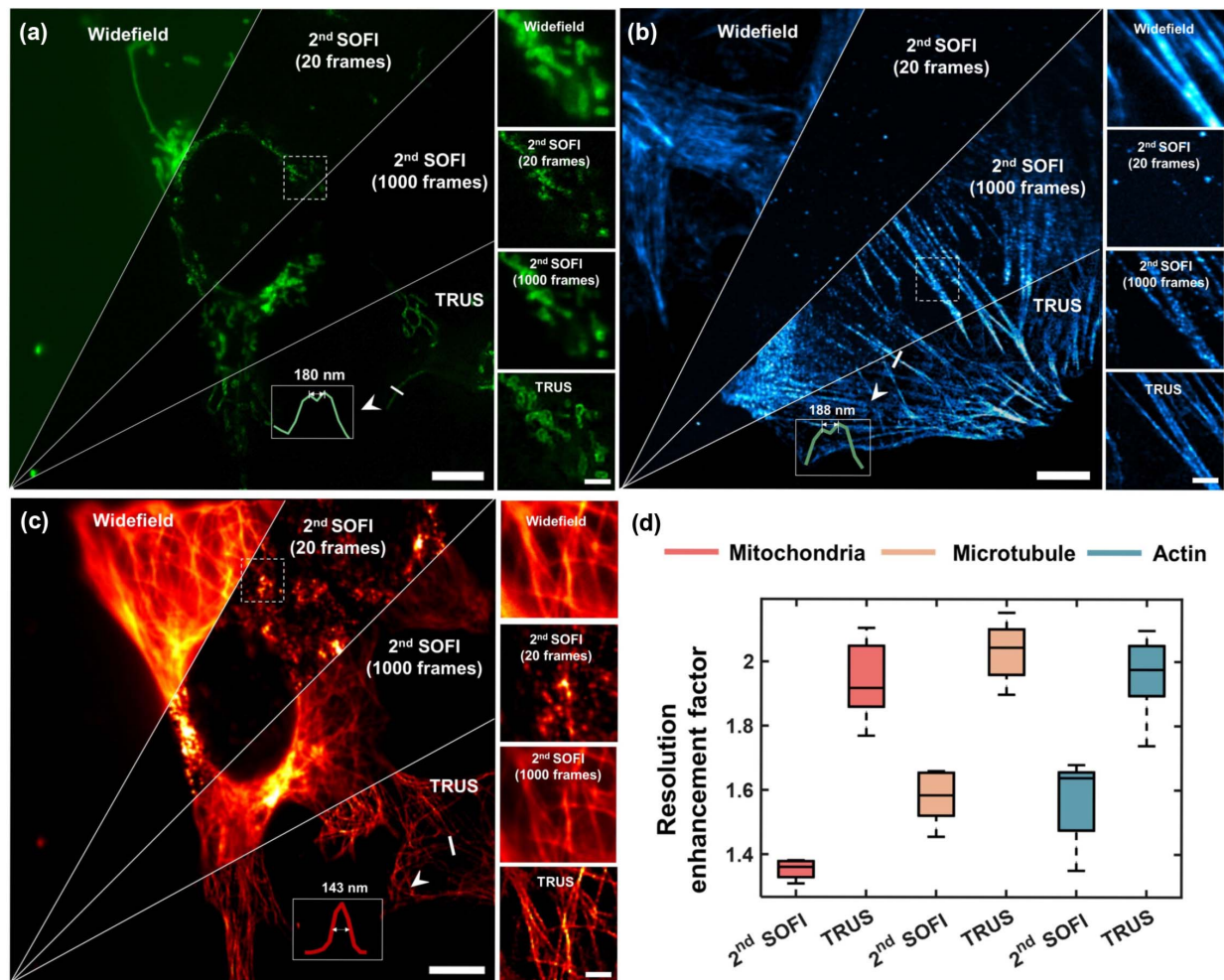


Fig. 5 TRUS framework universally extends the spatial resolution of SIM under different biological samples. (a)–(c) Representative widefield images and corresponding results reconstructed by conventional second-order SOFI and TRUS of outer mitochondrial membrane, F-actin, and microtubule samples. Magnified views of the regions marked by the white boxes are shown on the right side. The line profiles of intensity across the white line are shown. (d) Statistical resolution comparisons of second-order SOFI and TRUS in the cases of outer mitochondrial membrane, F-actin, and microtubule samples. Resolution enhancements were measured by the decorrelation analysis ($n = 5$, right). Scale bars: $6 \mu\text{m}$ (left), $1.5 \mu\text{m}$ [right, (a)–(c)].

TRUS achieves faithful reconstruction of continuous cytoskeletal filaments while preserving the intrinsic optical sectioning capability inherited from SOFI. Compared with the widefield image, TRUS reveals effective suppression of defocused background signals. Crucially, TRUS demonstrates superior resolving power versus conventional second-order SOFI requiring 1000-frame acquisitions (20 ms/frame), successfully distinguishing microtubule filament pairs separated by 180 nm, achieving twofold resolution improvement over the original widefield image (389 nm) defined by the decorrelation analysis.

3.4. TRUS Enables High-Throughput Super-Resolution Imaging

The remarkable enhancement in temporal resolution provided by TRUS enables SR imaging capable of capturing a significantly greater number of cells and FoVs per unit time. To quantitatively demonstrate this capability, we performed multi-region

imaging in 100 partially overlapping FoVs (partitioned into a 10×10 grid with dimensions of $125 \mu\text{m} \times 125 \mu\text{m}$ per FoV), successfully resolving over 200 cells containing immunolabeled microtubules. Subsequent acquisition of 100 datasets, each comprising 20 frames (20 ms exposure time) and widefield images (50 ms exposure time) collected within a 1.8 s interval per FoV, was completed in a total of ~ 3 min, with execution time dominated by mechanical stage stabilization delays. For comparative analysis, conventional second-order SOFI of equivalent FoVs was performed using extended acquisition parameters (1000 frames per FoV, 20 ms exposure), necessitating ~ 35 min of total imaging time. Comparative analysis on four distinct ranges of interest in Fig. 7(b) reveals fundamental limitations: widefield microscopy inherently lacks the spatial resolution to resolve subcellular structures, while conventional second-order SOFI shows discontinuity. In contrast, TRUS reconstruction consistently yields super-resolved images of exceptional quality, enabling unambiguous visualization of

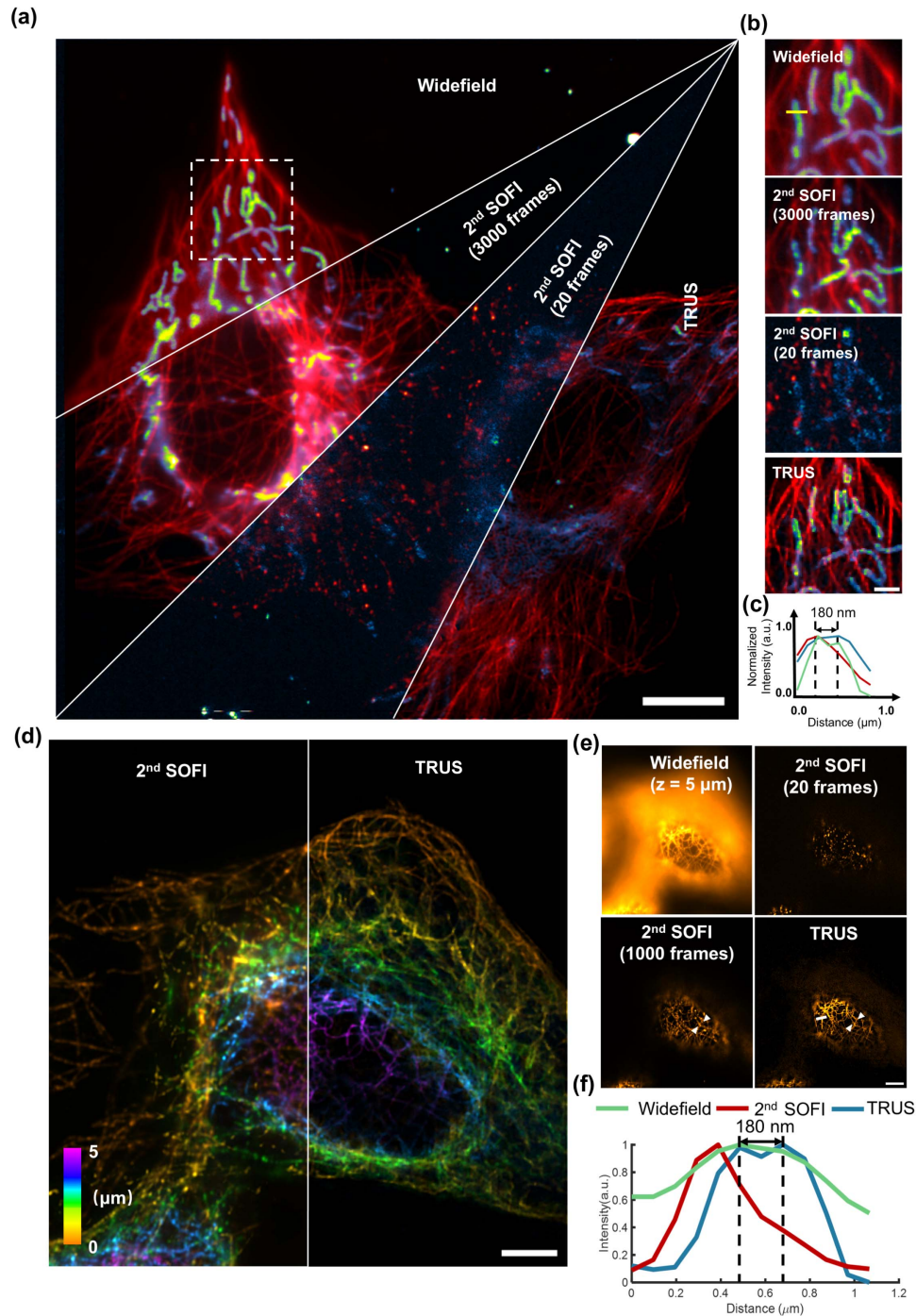


Fig. 6 TRUS application in dual-color imaging and three-dimensional imaging. (a) Representative widefield images and results reconstructed by second-order SOFI (reconstructed from 3000 frames), sparse SOFI (reconstructed from 20 frames), and TRUS of microtubule (red) labeled with Alexa Fluor 647 and outer mitochondrial membrane (green fire blue) labeled with Alexa Fluor 488. (b) Subcellular structures in the regions marked by the white boxes in (a). (c) Intensity profiles along the corresponding yellow line shown in (b). (d) Color-coded, 3D distributions of microtubule in COS-7 cells labeled with Alexa Fluor 647 obtained with second-order SOFI (left) and TRUS (right). (e) Horizontal section views of widefield, sparse SOFI (reconstructed from 20 frames), second-order SOFI (reconstructed from 1000 frames), and TRUS from (d). (f) Intensity profiles along the corresponding white line shown in (e). Scale bars: 10 μm in (a), 2 μm in (b), and 6 μm in (c), (d).

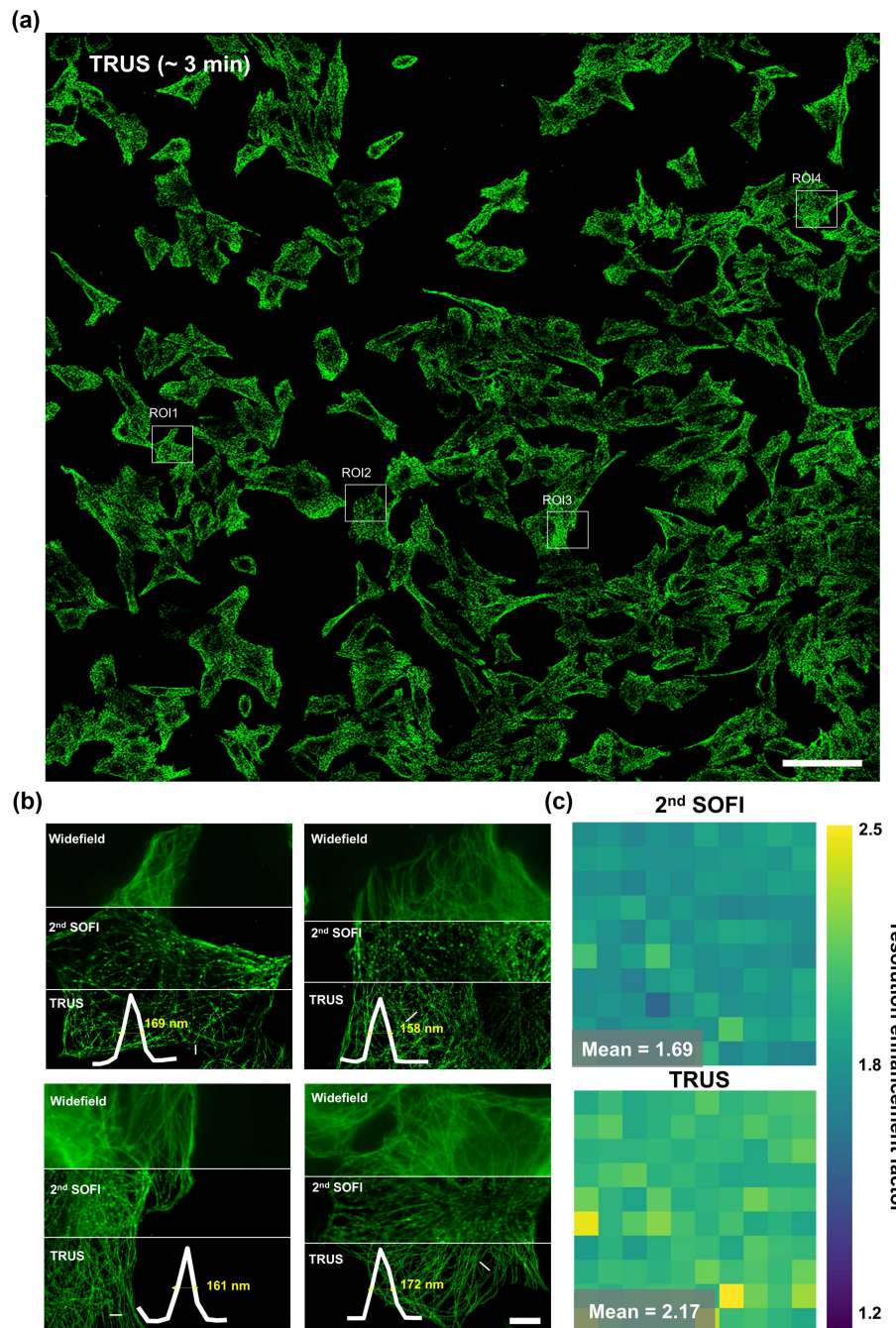


Fig. 7 High-throughput super-resolution (SR) imaging with TRUS. (a) Application of TRUS (with 20 frames and an additional widefield) to high-throughput SR imaging of a 1.0 mm^2 area. Microtubules are identified in COS-7 cells labeled with Alexa Fluor 647. (b) Enlarged views of the white boxed regions in (a) and line profiles of intensity across the white line. (c) Decorrelation resolution enhancement factor (compared with widefield) maps over the entire FoV of the second-order SOFI image (top) and TRUS image (bottom). Scale bars: $100 \mu\text{m}$ in (a) and $5 \mu\text{m}$ in (b).

microtubule networks. Quantitative assessment confirms that spatial resolution achieved the sub-200-nm level after TRUS reconstruction, with full width at half-maximum measurements across individual microtubule filaments ranging from 158 to 171 nm (mean \pm s.d. = $165 \text{ nm} \pm 4 \text{ nm}$). To systematically evaluate the spatial resolution enhancement across the extended

FoV, we quantitatively compare the spatial resolution improvement factors between SOFI and TRUS methodologies at each sub-field ($n = 100$). The quantified enhancement metrics are subsequently integrated into a composite resolution map [Fig. 7(c)], revealing that TRUS demonstrates a consistent 1.28-fold improvement in effective spatial resolution relative

Table 1 Comparison of Temporal Resolution in Different Imaging Modes

Imaging mode	FoV	Reconstructing method	Temporal resolution
Single FoV imaging	74.5 μm \times 74.5 μm	Second-order SOFI	20,000 ms
		TRUS	450 ms
3D imaging	50 μm \times 50 μm \times 5 μm	Second-order SOFI	100,000 ms
		TRUS	2250 ms
Large FoV imaging	1 mm \times 1 mm	Second-order SOFI	35 min
		TRUS	3 min

to conventional second-order SOFI throughout the entire 1 mm² imaging area [Fig. 7(a)]. This resolution enhancement remains spatially invariant across different tissue regions, as evidenced by the homogeneous color distribution in the enhancement map [Fig. 7(c)].

Furthermore, we established a high-throughput dual-color imaging platform capable of simultaneous acquisition in both spectral channels (Fig. S6 in the Supplement 1). We employed a multi-region acquisition strategy comprising 25 strategically overlapping FoVs arranged in a 5 \times 5 grid matrix. Each FoV (125 μm \times 125 μm) was spatially registered with 17% lateral overlap to ensure continuous sample coverage. Through this approach, we achieve high-resolution visualization of cytoskeletal architecture, identifying 20 cells exhibiting co-localization of immunolabeled microtubules and filamentous actin networks throughout the entire 0.25 mm² imaging area. The result reveals that TRUS enabled synchronized detection of emission signals under 488 and 647 nm excitation wavelengths, maintaining temporal resolution while effectively observing a millimeter-scale FoV. The acquisition durations of different imaging modes have been summarized in Table 1.

4. Conclusion

We introduce TRUS as a rapid acquisition framework for SOFI-based high-throughput imaging. This architecture enables high-fidelity SR reconstruction from merely 20 raw frames and an additional widefield image that is 2 orders of magnitude fewer than conventional SOFI requirements. Simultaneously, it achieves twofold spatial resolution improvement over the widefield image and demonstrates encouraging generalization potential across tested variations. Building upon the transformer's holistic feature extraction capability and integration of physics-driven training strategy, our method achieves a 44-fold acceleration in temporal resolution across broad emitter blinking dynamics. With the assistance of TRUS, we successfully demonstrate uniform spatial resolution improvement (average = 2.17-fold, $n = 100$) within a 1 mm² FoV within 3 min of acquisition time. The framework's reduced acquisition demands and enhanced tolerance to non-ideal labeling conditions enable rapid, gentle, and high-throughput SR imaging, demonstrating broad applicability for biological investigations.

While TRUS demonstrates encouraging generalization potential, future work is required to validate its performance on rarer morphologies. Integrating ADMM^[38] or FISTA iteration into the network training procedure might further improve the generalization performance of TRUS. Besides, its current implementation remains constrained by the system-specific PSF model. Although the impact of ideal PSF variation could

be reduced by Fourier interpolation (Fig. S7 in the Supplement 1), aberrations, motion blur, and other complex distortions may exceed this model's representational capacity, leading to reconstructing failure. Future extensions of TRUS could leverage physics-aware diffusion models to enable cross-platform adaptability^[39]. Notably, integrating blind PSF estimation^[40] during training might offer potential for improved aberration consistency. In our experiment, we incorporated additional data augmentation to enhance TRUS performance by mitigating the data hunger inherent in transformer architectures. However, the integration of physics-informed constraints with few-shot learning strategies—such as transfer learning and meta-learning—may serve as a promising approach to mitigate dataset size limitations. The combination of these approaches has the potential to transform TRUS into a versatile solution, thereby extending the high spatiotemporal resolution and high-throughput imaging capabilities of TRUS to a broad range of fluorescence imaging techniques, facilitating the discovery of crucial biological insights.

Disclosures

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code, Data, and Materials Availability

The code to reconstruct TRUS images is available in Ref. [41].

Acknowledgments

We thank Y. H. Gan and X. Y. Li for their help with suggestions for our research. This work was supported by the National Natural Science Foundation of China (Nos. 62361166631 and 62125504), the STI 2030—Major Projects (No. 2021ZD0200401), and the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (No. 2024R01001).

References

1. B. Andrews *et al.*, “Imaging cell biology,” *Nat. Cell Biol.* **24**, 1180 (2022).
2. L. Schermelleh *et al.*, “Super-resolution microscopy demystified,” *Nat. Cell Biol.* **21**, 72 (2019).
3. Y. Huang *et al.*, “Multiplexed stimulated emission depletion nanoscopy (mSTED) for 5-color live-cell long-term imaging of organelle interactome,” *Opto-Electron. Adv.* **7**, 240035 (2024).
4. Y. Sun *et al.*, “Fluorescence interference structured illumination microscopy for 3D morphology imaging with high axial resolution,” *Adv. Photonics* **5**, 056007 (2023).

5. E. Betzig *et al.*, “Imaging intracellular fluorescent proteins at nanometer resolution,” *Science* **313**, 1642 (2006).
6. N. Bourg *et al.*, “Direct optical nanoscopy with axially localized detection,” *Nat. Photonics* **9**, 587 (2015).
7. M. Lelek *et al.*, “Single-molecule localization microscopy,” *Nat. Rev. Methods Primers* **1**, 39 (2021).
8. S. Geissbuehler, C. Dellagiacomma, and T. Lasser, “Comparison between SOFI and STORM,” *Biomed. Opt. Express* **2**, 408 (2011).
9. R. Jungmann *et al.*, “Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT,” *Nat. Methods* **11**, 313 (2014).
10. G. T. Dempsey *et al.*, “Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging,” *Nat. Methods* **8**, 1027 (2011).
11. S. Basak *et al.*, “Super-resolution optical fluctuation imaging,” *Nat. Photonics* **19**, 229 (2025).
12. T. Dertinger *et al.*, “Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI),” *Proc. Natl. Acad. Sci. USA* **106**, 22287 (2009).
13. S. Geissbuehler *et al.*, “Live-cell multiplane three-dimensional super-resolution optical fluctuation imaging,” *Nat. Commun.* **5**, 5830 (2014).
14. A. Sroda *et al.*, “SOFISM: Super-resolution optical fluctuation image scanning microscopy,” *Optica* **7**, 1308 (2020).
15. M. Pawlowska *et al.*, “Embracing the uncertainty: the evolution of SOFI into a diverse family of fluctuation-based super-resolution microscopy methods,” *J. Phys.* **4**, 012002 (2022).
16. K. Grubmayer *et al.*, “Self-blinking dyes unlock high-order and multiplane super-resolution optical fluctuation imaging,” *ACS Nano* **14**, 9156 (2020).
17. W. Vandenberg *et al.*, “Model-free uncertainty estimation in stochastic optical fluctuation imaging (SOFI) leads to a doubled temporal resolution,” *Biomed. Opt. Express* **7**, 467 (2016).
18. S. Jiang *et al.*, “Enhanced SOFI algorithm achieved with modified optical fluctuating signal extraction,” *Opt. Express* **24**, 3037 (2016).
19. W. Zhao *et al.*, “Enhanced detection of fluorescence fluctuations for high-throughput super-resolution imaging,” *Nat. Photonics* **17**, 806 (2023).
20. A. Yemets *et al.*, “Quantum dot-antibody conjugates for immunofluorescence studies of biomolecules and subcellular structures,” *J. Fluorescence* **32**, 1713 (2022).
21. M. B. Elowitz *et al.*, “Stochastic gene expression in a single cell,” *Science* **297**, 1183 (2002).
22. X. Liu *et al.*, “Reconstruction of structured illumination microscopy with an untrained neural network,” *Opt. Commun.* **537**, 129431 (2023).
23. H. Wu *et al.*, “Single-frame structured illumination microscopy for fast live-cell imaging,” *APL Photonics* **9**, 036102 (2024).
24. Z. Ye *et al.*, “Untrained neural network enabling fast and universal structured-illumination microscopy,” *Opt. Lett.* **49**, 2205 (2024).
25. W. Ouyang *et al.*, “Deep learning massively accelerates super-resolution localization microscopy,” *Nat. Biotechnol.* **36**, 460 (2018).
26. E. Nehme *et al.*, “Deep-STORM: super-resolution single-molecule microscopy by deep learning,” *Optica* **5**, 458 (2018).
27. K. K. Narayanasamy *et al.*, “Fast DNA-PAINT imaging using a deep neural network,” *Nat. Commun.* **13**, 5047 (2022).
28. L. M. Beck *et al.*, “Improving correlation based super-resolution microscopy images through image fusion by self-supervised deep learning,” *Opt. Express* **32**, 28195 (2024).
29. Y. Deng *et al.*, “T-former: an efficient transformer for image inpainting,” in *Mm’22* (2022), p. 6559.
30. Z. Liu *et al.*, “Swin transformer: hierarchical vision transformer using shifted windows,” in *IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), p. 9992.
31. S. Geissbuehler *et al.*, “Mapping molecular statistics with balanced super-resolution optical fluctuation imaging (bSOFI),” *Opt. Nanoscopy* **1**, 4 (2012).
32. A. Descloux, K. S. Grubmayer, and A. Radenovic, “Parameter-free image resolution estimation based on decorrelation analysis,” *Nat. Methods* **16**, 918 (2019).
33. Y. Guo *et al.*, “Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales,” *Cell* **175**, 1430 (2018).
34. A. Girsault *et al.*, “SOFI simulation tool: a software package for simulating and testing super-resolution optical fluctuation imaging,” *PLOS One* **11**, e0161602 (2016).
35. G. Wen *et al.*, “High-fidelity structured illumination microscopy by point-spread-function engineering,” *Light: Sci. Appl.* **10**, 70 (2021).
36. R. M. Power *et al.*, “Build and operation of a custom 3D, multi-color, single-molecule localization microscope,” *Nat. Protocols* **19**, 2467 (2024).
37. W. Xue *et al.*, “Gradient magnitude similarity deviation: a highly efficient perceptual image quality index,” *IEEE Trans. Image Process.* **23**, 684 (2014).
38. Y. He *et al.*, “Untrained neural network enhances the resolution of structured illumination microscopy under strong background and noise levels,” *Adv. Photonics Nexus* **2**, 046005 (2023).
39. A. Bansal *et al.*, “Cold diffusion: inverting arbitrary image transforms without noise,” in *Nips ’23* (2023).
40. W. Zhang *et al.*, “Self-supervised PSF-informed deep learning enables real-time deconvolution for optical coherence tomography,” *Adv. Imaging* **2**, 021001 (2025).
41. <https://github.com/ZJUOPTKuangLab/TRUS>.