Implicit encoding via semantic redundancy elimination enabling adaptive compression of multidimensional biological data

Yifan Ma,[1#] Chengqiang Yi,[1#] Yao Zhou,[1] Shimeng Gao,[1] Jianchao Liu,[1] Binbing Liu,[1*] and Peng Fei, [1,2 *]

[1]School of Optical and Electronic Information-Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China, 430074

[2]Advanced Biomedical Imaging Facility Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[#] These authors contributed equally to this work.

[*] Correspondence：liubinbing@hust.edu.cn and feipeng@hust.edu.cn

Abstract

With the rapid development in advanced bioimaging techniques, numerous image data has been generated for various high-throughput applications, which pose big challenges to efficient storage, transmission and sharing of these data. Traditional compression techniques and current deep learning-based compression methods lack the elimination of semantic redundancy on multidimensional biomedical data containing high dynamic range and complex structural features. Here, we propose compression with implicit function correlation (SWIFT) approach based on the first-proved high correlation of biomedical data in the implicit function domain. Through reducing the extra-compression of data correlation, SWIFT approach shows compression ratios and processing speed higher than existing implicit neural representation-based approaches. With further combination of residual entropy coding and attention mechanism, SWIFT achieves compression of various biological data with high time continuity and structure fidelity, for example, yielding 1500-fold, effective compression of live-cell 3D super-resolution data to enables downstream tasks and quantitative analyses at notably improved efficiency.

Introduction

Advanced optical imaging techniques in conjunction with efficient image processing approaches makes big impact on modern life science. Many biomedical applications currently require a vast amount of experimental data to be generated for various image-based analysis. For instance, studying the cytotoxic mechanisms of CAR-T cells

through long-term live-cell imaging of cell morphological changes can produce several terabytes (TB) to tens of terabytes image data using high-resolution and high-throughput fluorescence microscopy systems [1, 2]. For another instance, volumetric imaging of an entire mammalian brain, like mouse brain, at single-neuron resolution to create the brain connectivity map will yield tens of terabytes image data [3]. This vast amount of data places significant burdens on data storage, computation and sharing. For storage at limited space, these massive raw data have to be saved partially, with increased risk of data loss. Besides, due to the limited transmission bandwidth, researchers can only transfer and share data in an inefficient manner. Although large-scale data sharing cloud platforms have been established to address storage limitations and insufficient bandwidth [4, 5], they do not fully resolve the storage issues at the user end and may entail substantial maintenance costs, significantly impeding the progress of biomedical research. In contrast to centralized cloud storage and exchange technologies, it is crucial to achieve effective and data-specific storage compression directly at the user terminal. To this end, storage optimization on the data generation source[6] and more importantly, on the downstream compression side should be studied.

The essence of compression lies in the removal of redundancy brought by the internal correlation among signals. Traditional compression methods based on analytical model or statistical model explicitly remove spatial redundancy and temporal redundancy through transformations and coding in order to compress the amount of data, such as domain transformations [7], entropy coding [8, 9]. However, in various application scenarios, there are abundant semantically similar images, such as two visually similar images of apples and biomedical images with the same structure. Experiments demonstrate that among these visually similar images, there exists redundancy that differs from temporal and spatial redundancies, known as semantic redundancy [10, 11]. Conventional methods are unable to extract abstract semantic correlations, thus leaving room for further improvements in compression efficiency. Meanwhile, widely-used methods like JPEG [12], H.264 [13], H.265 [14], etc., optimized for natural images/videos, perform poorly in high-fidelity compression of high dynamic range biomedical images. Moreover, frequency-domain-based compression methods may suffer from spectral truncation or blocking artifacts [15], which influence downstream analysis tasks. Therefore, these methods demonstrate limited performance when facing biomedical data with multiple dimensions (2D, 3D, 4D, 5D), high dynamic range, substantial data volume, and high structural similarities.

Unlike model-based approaches, deep learning-based data compression techniques, such as Autoencoders [16] ,VAE [17], GAN [18] *etc.* have recently emerged to interrogate the semantic correlation among signals. These approaches compress input data into a low-dimensional space representation which aims to eliminate semantic redundancy among data information by learning deep feature representations, and then reconstruct the original data using a decoder. Nevertheless, capturing essential data features, reducing dimensionality for optimal latent representation, and handling the burden of training with massive data for a single decoder remain highly challenging. Furthermore, when dealing with complex biomedical data, these deep learning-based

supervised methods exhibit limited generalization capabilities and significant performance degradation due to generalization errors.

The latest development of implicit neural representation (INR) [19] utilizes a neural network to parameterize a continuous function based on the data dimensions, enabling advanced representations of 3D scenes [20], images [21], and videos [22]. Unlike traditional discrete grid-based representations, the grid-free feature of INR representation enables better coincide with the continuity of target visual information. This facilitates the generation of continuous representations that enable seamless and arbitrary interpolation of visual data. Additionally, with the universal approximation theorem of neural networks [23, 24], INR implemented using Multi-Layer Perceptron (MLP) can fit any complex function with a sufficient number of parameters, resulting in high-fidelity compression representation. Currently, INR has been demonstrated to be applicable for compressing natural image scenes [25] as well as various multi-dimensional biomedical data for biological research and clinical diagnosis, including 2D images, 3D volumes [26], and 4D data. By controlling the network parameters across different samples, stable compression rates and compression quality can be achieved. Moreover, INR exhibits stable performance for various data compression without the requirement of modifying the network structure specifically, which is can't be achieved by other supervised learning-based approaches due to the generalization errors.

However, existing INR-based compression methods require massive training time [27-29] due to processing the large number of input coordinates when representing compression data by network optimization. With the dimensions increase, the network parameters of traditional INR methods also grow exponentially, resulting in significantly increased computational burden [30]. Additionally, these INR methods do not effectively leverage the correlation between data, such as the inter-frame correlation in dynamic biological data and highly similar local features. Furthermore, INR fits the entire image region without specific optimization for particular signals, such as adapting to the signal distribution in the compressed data (e.g., density requirements, application-specific regions for quantification or morphology segmentation) through different levels of optimization.

Based on the aforementioned challenges, we conduct a novel study that firstly demonstrates the remarkable correlation of biological data in the implicit function domain. With this correlation, we propose the compression with implicit function correlation (SWIFT) approach, which substantially accelerates the compression time of the algorithm through optimized weight inheritance. We also introduce attention-based compression mechanism to adaptively capture the specific structure information, facilitating high-fidelity compression of multi-modal biological images. Moreover, by employing residual-based entropy coding to further compress the optimized INR weights, we achieve a high compression ratio of 1516 folds for multidimensional images. We demonstrate the superior performance of SWIFT on large biomedical datasets from different imaging modalities (2D, 2D video, 3D and 3D video), showcasing its potentials for advancing various high-throughput large-scale image applications.

Results

2.1 Verification of implicit function domain correlation

Validation of implicit function domain correlation among multidimensional data

For multidimensional biomedical data (2D, 3D, 4D data), they exhibit a high degree of correlation in semantic structure, including inter-frame correlation in dynamic biological data and high correlation in local features. We aim to leverage this correlation to further eliminate the semantic redundancy among data. Remarkably, we have discovered that multidimensional biomedical data exhibits this high correlation in the implicit function domain of INR. To validate this correlation in INR, we conduct theoretical analysis, and the results are depicted in Figure 1(a-b).

First, we classified the complex biomedical data into three modes for analysis. In Mode 1, we focused on 2D data. We employed the Patch Sliding technique to partition the original large-scale 2D data into groups. Subsequently, we selected adjacent patches for compression using INR. By computing the Wilson Coefficient [31] to measure the distribution disparities of the network parameters, we validated distributional correlation between neighboring patches. Implicit function domain displays increased overlap, while in the image domain, the distribution is widely spread, as depicted in Figure 1(a). The low Average Wilson Coefficient (AWC) value of implicit function domain is closer to zero than image domain which reveals a higher level of correlation in their implicit function domain distributions.

We conducted analogous experiment for Mode 2 and Mode 3. Specifically, we computed the Wilson Coefficient for adjacent Z-slices in the 3D biomedical data and for consecutive time sequences in the 4D biomedical data, allowing us to assess their correlation in the implicit function domain of INR. The outcomes demonstrated that both axial and temporal correlations of biomedical data can be effectively captured and represented in the implicit function domain of INR. Regardless of the data dimensionality, the implicit function domain correlation is higher than that in the image domain.

Structurally similar data exhibits enhanced clustering effects in the implicit function domain.

Data with disparate structures exhibit substantial differences, whether to be observed in the image domain or the implicit function domain. However, multidimensional biomedical data shows a great degree of approximation in various local structural features. We selected blood vessel and cell data as representative examples (considering that blood vessels represent linear structures and cells represent spherical structures, which collectively reflect the majority of local features in biomedical data) to quantitatively assess the clustering effects of such structurally similar data across different spaces. Through block partitioning of large-scale data, we generated ten thousand small images containing local features. Subsequently, we employed T-distributed Stochastic Neighbor Embedding (TSNE) dimensionality reduction to compare the clustering patterns of these images in both the image domain and the implicit function domain of INR by downscaling the high-dimensional data to a 2-dimensional space, as illustrated in Figure 1(b). We discovered that samples of the same class exhibited multiple clusters in the image domain, while performed single

cluster in implicit function domain. This suggests that the network is better able to learn and capture information in the implicit function domain. We computed the Rayleigh Entropy to quantitatively evaluate their clustering characteristics. A small Rayleigh entropy indicates compact clustering, with differences between groups more pronounced than within groups.

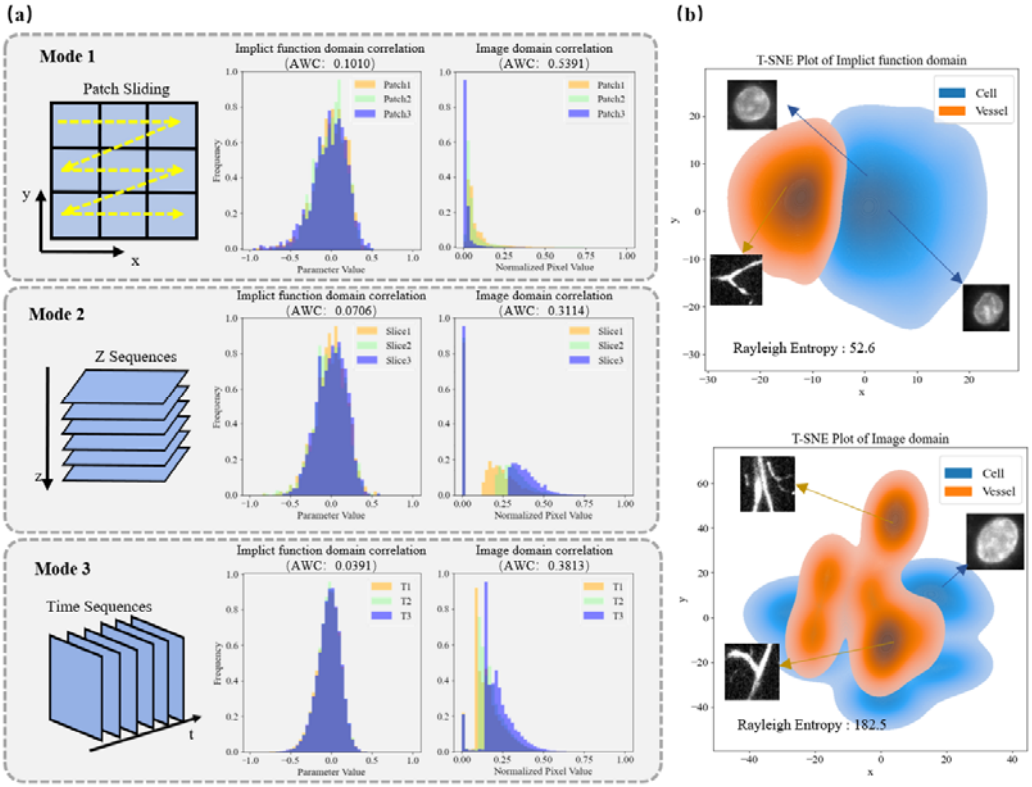

Fig. 1 Validation of implicit function domaincorrelation. (a) Comparisons of correlation in three modes (from top to bottom: spatial correlation, axial correlation, and temporal correlation). **(b)** Comparison of clustering for structurally similar data in Implicit function and image domain.

2.2 Efficient Data Compression based on SWIFT

For a large-scale biomedical sample dataset, it can be viewed as a discrete sampling result of a continuous spatiotemporal signal. By employing INR, we can map these sampling results to high-dimensional functions, achieving “data-to-function” encoding. Though, according to constructing complex neural networks, it’s possible to fit this high-dimension function (data), the redundancy within the data is neglected yielding extra computational burden. Inspired by function decomposition, we propose an adaptive grouping strategy for data partitioning. As illustrated in Figure 2.a(□) (step 1). we make an interval sampling along time series to produce N groups, ensuring retaining the global motion information within a group and high cross-correlation between different groups due to the frame continuity. So, with this decomposition, we can achieve highly efficient compression by utilizing multiple

simple implicit neural representations.

Then, for each pre-defined subgroup of data, we propose a training sampling strategy based on attention mechanisms, allowing for adaptive compression fitting of high-structural features in the biological data. As depicted in Figure 2.a(□), a pre-trained saliency extraction network generates saliency maps that serve as sampling probability maps for network input coordinates, bonding the structure information with network optimization.

Meanwhile, we construct the Multi-Layer Perceptron (MLP) with the same network structure as a parameterized mapping function to fit each set of original data. After attention-guided sampling, the selected coordinates are firstly encoded to vectors with high-frequency by positional encoding (Figure 2.a(□)), and then are fed into the MLP. During the training process, the weights and biases of network are iteratively optimized to minimize the difference between network output and original data. We use a combination of L1 and SSIM loss [32] as the objective function to ensure high-quality compression.

With the high correlation between pre-partitioned groups, we adopt a weight transfer training approach to promote network rapidly convergence by sequentially fitting each group data. As depicted in Figure 2.a(□), we compress the 1st group data first and obtain the optimized parameters ($\theta_1^{'}$) as the initial network. And then for the 2nd group data compression, we start the optimization on the initial trained network instead of creating another new model. This transfer-learning-based compression strategy minimizes the distance in implicit function domain between initial parameters (θ_2) and optimized ones ($\theta_2^{'}$) of group 2 data. For the remained groups, repeating this strategy sequentially can produce rapid network convergence speed compared with directly compressing individual group. This initialization scheme assigns higher initial values to the weights, bringing them closer to the final optimized weights in the implicit function domain. This strategy accelerates the training process and facilitates rapid convergence. It is worth noting that, since images at multiple time points are not spatially aligned, end-to-end algorithms typically cannot take advantage of this transfer. Instead, we have demonstrated significant correlation between the weights across various time points in the implicit function domain, making weight transfer possible.

After global fitting of the original data, we further compress the storage volume using residual entropy coding, leveraging the correlation between different implicit functions. As depicted in Figure 2.a(□), SWIFT encodes the residuals by subtracting the network parameters between neighboring networks and applying entropy coding. As a result, a bitstream is produced, streamlining storage and transmission processes. In the decompression process, as shown in Figure 2(b), we reverse the entropy coding to restore the bitstream with saved initial parameters θ_1 of group 1and residuals $\triangle\theta$ of other groups. Subsequently, by successively adding the residuals to the initial weight parameters, we reconstruct the original weight parameters for each network corresponding to each data group. Next, by model forward inference and regrouping, the reconstructed data can be reassembled into original data order. It is noteworthy

that with the characteristics of grid optimization and function continuity of INR, specific localization along spatial or temporal axials and arbitrary scaling can be achieved contributing to the visualization and analysis for biomedical application.

For various dimensional data like 2D, 3D and 3D temporal data, SWIFT can apply the same compression strategy via employing corresponding initial INR for grouped data as demonstrated in Fig.1. With integrated platform based on SWIFT, biomedical multi-modality data can be compressed and stored under the universal framework enabling downstream tasks analyzed accurately and reliably (Fig.2. c).

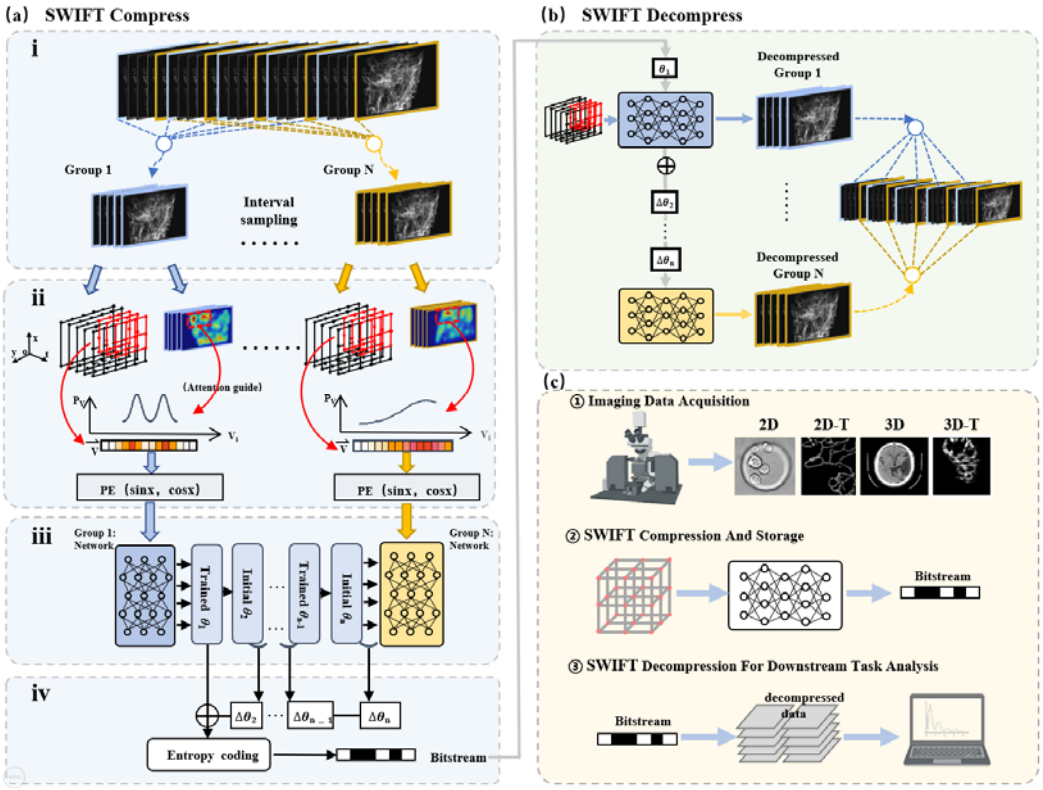

Fig. 2 The principle of SWIFT. (a) Compression pipeline for SWIFT.(In the SWIFT compression process, for a given large-scale biomedical dataset, the data is first effectively partitioned through adaptive grouping, ensuring high correlation between sub-data groups and preserving global coherence. Next, attention mechanisms are introduced, utilizing saliency maps to achieve adaptive compression fitting for high-structural features in the biomedical data. Subsequently, rate-distortion performance and training time of INR-based compression are enhanced through weight inheritance and residual storage. **)(b)** Decompression pipeline for SWIFT. **(c)** Integrated workflow of data acquisition, compression, storage, and analysis.

2.3 SWIFT achieves high-fidelity and high-level compression of multidimensional biomedical data

High structural fidelity compression for large field of view 2D images

Bright-field imaging under wide-field microscopy benefits from a large field of view, providing abundant biological data [33]. However, bright-field images typically exhibit low contrast or contain less gradient information, making it more susceptible to losing such gradient structural information during large data compression. Therefore, there is a need for compression algorithms that offer both high-fidelity representation and high compression ratios, enabling efficient data storage. This facilitates convenient data processing and analysis for subsequent tasks such as cell segmentation. Here, we applied the SWIFT to compress large field of view imaging data acquired by the bright-field imaging system. Our approach achieved dramatic data compression while accurately resolving individual cells, enabling precise cell segmentation and counting. The field of view of entire two-dimensional plane is 2.67 × 3.99 mm^2, and we divided it into 48 patches using patch sliding, as illustrated in Figure 3(a). Each patch was then compressed using the SWIFT algorithm guided by saliency map. Considering the large field-of-view information of bright-field imaging, here we take the learnable saliency map to better prioritize the information. To showcase the compression capability, we selected three representative patches for visualization. We further enlarged the regions of interest (ROIs) within these patches and displayed them side-by-side with their corresponding probability saliency maps in Figure 3(b). Furthermore, we compared our method with JPEG compression and conventional INR compression methods applied to the entire dataset [24-26], demonstrating superior visual fidelity achieved by our approach.

Cell counting statistics play a crucial role in biologically relevant data associated with cell culture. Therefore, we performed cell segmentation using the open-source software "segment anything" [34] to facilitate subsequent cell counting tasks. The results of various statistical analyses are shown in Figure 3(c). From the comparisons, our method accurately reconstructed the ROIs, such as cells within the culture chamber, preserving the morphology of each individual cell while avoiding loss of the original resolution. Additionally, our algorithm benefited from weight inheritance transfer, enabling faster completion of the compression task compared to the INR SIREN (INR-S) method that compresses the entire dataset. In terms of numerical results, our method outperformed other algorithms in terms of quantitatively measured metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Moreover, our method achieved an average accuracy of 95% in cell counting tasks, allowing researchers to accurately perform downstream tasks and subsequent analyses.

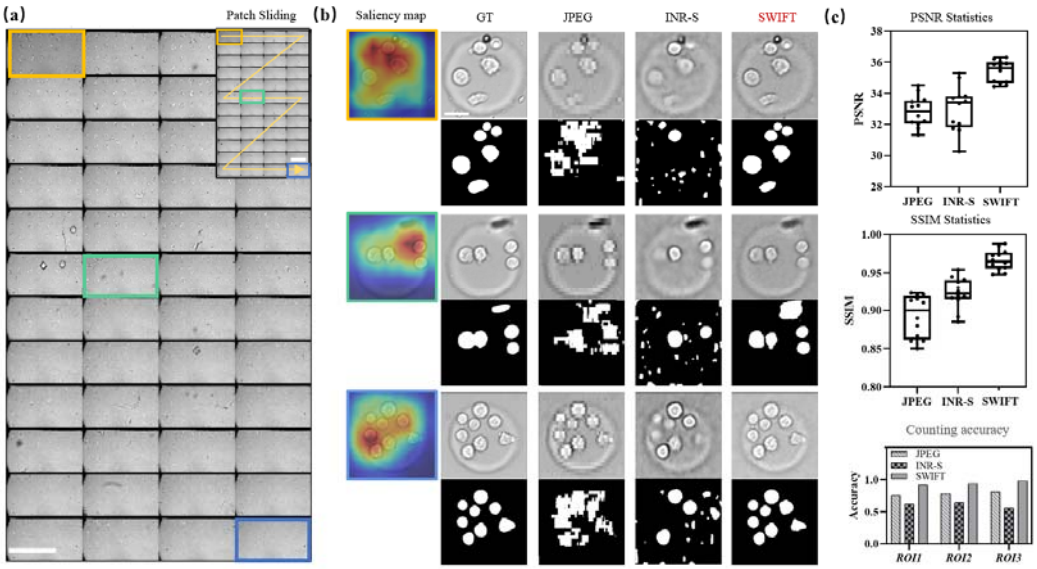

Fig. 3 Demonstration of SWIFT compression capability on large field-of-view 2D images and performance on downstream tasks. (a) Comprehensive field-of-view visualization of high-throughput bright field images (Brief illustration of data grouping and weight inheritance sequence shown in the top right corner, for a reduced display). Scale bar: 400μm. **(b)** Comparison of decompressed data and segmentation results under different compression methods (The most left column shows the display of saliency maps for selected ROIs, where ROI 1, 2, and 3 are extracted from different colored patches in (a)). Scale bar: 15μm. **(c)** Comparison of various evaluation metrics and downstream task accuracy under different compression methods.

High fidelity compression in intensity for dynamic biological data

To comprehend the dynamic mapping of sensory inputs onto the functional activity of neuronal populations and how this processing gives rise to cognitive functions and behavior, it necessitates employing non-invasive tools with high spatiotemporal resolution to interrogate neuronal circuits and track the intensity variations of neurons.[35, 36] In order to capture the high-speed dynamic phenomena of motor neurons and observe transient neural activity, long-term imaging at high resolution is required, which generates a large amount of data. This long-term imaging of the entire dynamic process generates a large amount of data, and thus requires compression algorithms with high fidelity of intensity to truly reveal the internal activity state of the organisms and to promote the subsequent analyses.

Employing SWIFT on fast-moving worm data via light-field microscopy at the framerate of 100-Hz, we achieved a compression ratio of 1516 folds. A comparison of our method with H.265 and conventional INR method for reconstructing fast-moving neuronal signals is presented in Figure 4(a)., The ROIs show that H.265 and INR-S lose some low-intensity neuronal signals due to abrupt intensity variations, while our method better fits these intensity changes because of the signal enhancement by saliency map. (It is noteworthy that considering the extreme sparsity of neuron signals, we adopted a conventional threshold-based saliency map to prevent loss of weak signals.) Furthermore, our approach performs better spatial resolution than other two methods contributing to the resolvability of dense signals as shown in ROI2. We

conducted trajectory tracking of the neurons to investigate the motor neuron activity's dynamic process, as depicted in Figure 4(b). We selected four neurons along the worm's body for trajectory tracking and plotted the corresponding signal intensity profiles. Compared with other methods, SWIFT preserves the fast fluctuation of dynamic signals with large compression ratio and the Pearson coefficient closer to 1 exhibiting greater stability in intensity fidelity. We further track and analyze the remained neurons to make their PSNR statistics, signal preservation statistics and intensity correlation statistics, as presented in Figure 4 (c). Benefitting from its inherent mapping function from coordinates to signal values and the adaptability of the saliency map for identifying low-intensity signals, SWIFT demonstrated higher intensity accuracy than other methods. This ensures precise representation of dynamic biological data and seamless continuation of subsequent tasks.

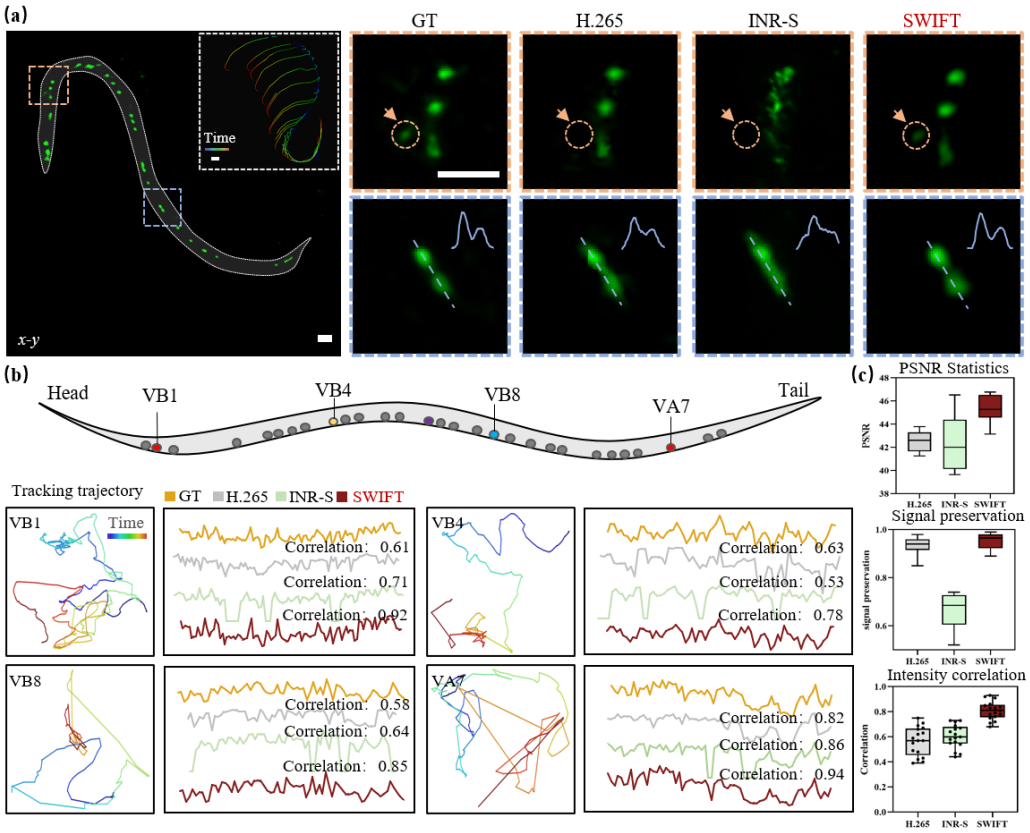

Fig. 4 Demonstration of SWIFT compression capability on dynamic biological data and downstream task performance. (a) A transient neuronal signal fragment reconstructed under different compression methods. (The top right corner shows the worm crawling trend with a time-coded trace.) Scale bar: 10μm. **(b)** Comparison of intensity fidelity of neuronal signal under different compression

methods. **(c)** Comparison of PSNR statistics, signal preservation and intensity correlation under different compression methods.

High-fidelity visual compression of 4D dynamic super-resolution biological light sheet data.

In long-term live biological imaging, highly dynamic four-dimensional (4D) life activities, such as blood flow and periodic motion of the heart, tend to generate tremendous amounts of data. However, in subsequent tasks such as visualization and morphological analysis, 4D data undergo complex variations in both temporal and spatial domains, making it challenging for traditional methods to compress correctly, thus limiting their performance. Therefore, there is a need for high-fidelity and high-compression ratio compression ensuring the normal progress of subsequent analysis tasks.

We applied SWIFT to 4D cell super-resolution data that were acquired using a light-sheet fluorescence microscopy system with an isotropic resolution of 48.5 × 48.5 × 48.5 nm^3, containing 180 volumes and hundreds of millions of voxels. Remarkably, we achieved a compression ratio of 700 folds, while accurately reconstructing complex outer membrane morphology and preserving the high dynamic range of intensity distribution. To validate the overall and directional structural integrity, we used the Imaris visualization tool to visualize the reconstructed data and the ground truth data in 3D and stitched them together, as shown in Figure 5(a).

For local detail visualization, we selected three time points from the same ROI and magnified them. Due to the limitation of directly compressing 4D data by H.265, we concatenated all t-axis data along the z-axis to fit it into 3D format for testing H.265 compression (referred to as H.265-S). Similarly, constrained by the vast volume of data, INR-S cannot compress the entire dataset as a whole. Therefore, for the 4D compression test of INR-S, we need to partition the large dataset into smaller segments and train them separately. Finally, these segments are connected together to form the complete results. In visual comparison, our method significantly outperformed other approaches, accurately capturing the process of mitochondrial fission. Calculation of Error maps and SSIM in corresponding regions indicated that our method achieved higher structural fidelity compared to others. The incomplete mitochondrial fission observed in the results of H.265-S may be attributed to the concatenation of temporal and axial dimensions, leading to non-uniform signal distribution and signal residues. Meanwhile, INR-S lacked sufficient fitting ability at high compression ratios to learn regions of low signal intensity, leading to compromised structural fidelity, as illustrated in Figure 5(b).

In addition, we conducted quantitative analysis of the selected mitochondrial cross-sectional morphological changes in the XZ direction. By calculating the Intersection over Union (IoU) scores, the results demonstrated that our method achieved nearly identical stable morphological changes and approximate cross-sectional areas compared to the original data. In contrast, H.265-S exhibited significant morphological variation deviation, likely due to its forced concatenation and fitting along the z and t directions. On the other hand, INR-S, due to its lack of

fitting ability, could hardly accurately discern the inner and outer membrane radii, which affected subsequent quantitative analyses, as depicted in Figure 5(c).

For the analysis of compression quality fluctuations along the t-axis and motion continuity accuracy, we computed the accuracy of mitochondrial breakpoints over time and cross-sectional areas, as shown in Figure 5(d). The statistical results demonstrated that our method achieved higher structural fidelity and time signal continuity compared to other approaches. Benefiting from our adaptive grouping strategy, our method effectively ensured global continuity and frame-to-frame continuity, enabling stable compression fidelity performance along the t-axis. This allows researchers to conduct reliable data analysis and validation in downstream tasks.

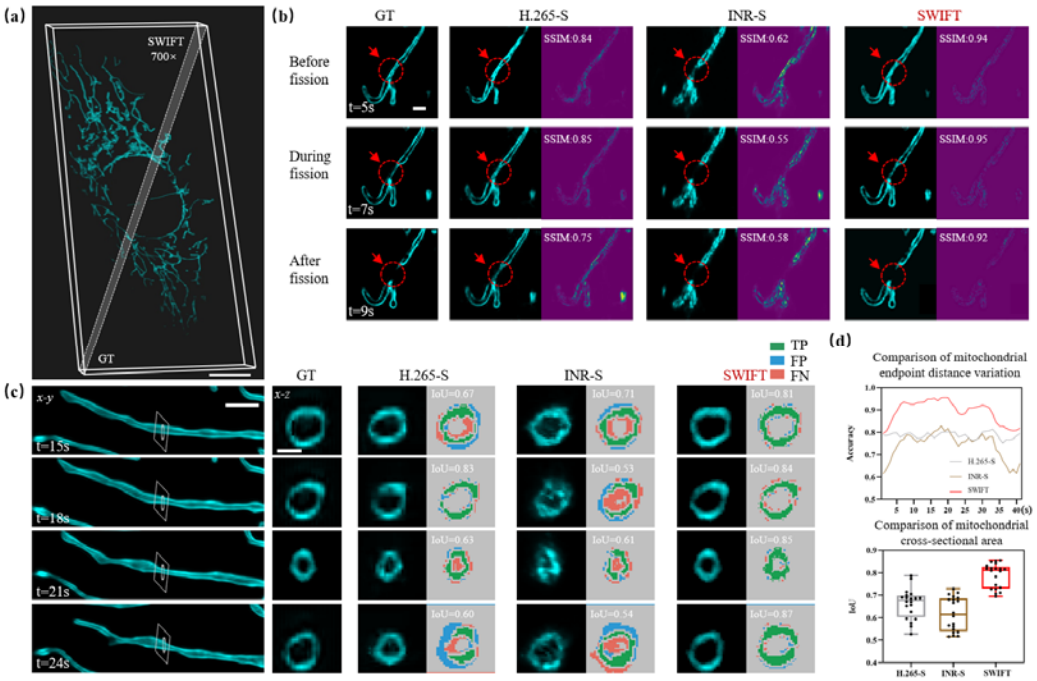

Fig. 5 Demonstrating SWIFT compression capability on four-dimensional dynamic super-resolution biological light-sheet data and downstream task performance. (a) Visual comparison of GT and SWIFT compressed versions in Imaris visualization, with compression ratios labeled in the top left corner. Scale bar: 100μm. **(b)** Comparison of continuously evolving mitochondrial morphology and corresponding error maps. (The red regions demonstrate mitochondrial fission over time, with only SWIFT capable of resolving this fine-scale process.) Scale bar: 1μm. **(c)** Comparison of decompressed mitochondrial axial morphological changes. (The right images display xz sectional views of the same region at different time points, demonstrating mitochondrial contraction and expansion at the nanoscale.) TP (True Positive): the correctly reconstructed structures; FP (FalsePositive): the incorrectly hallucinated structures; FN (False Negative): the missing details. Scale bars from left to right: 2μm, 1μm. **(d)** Comparison of various downstream task assessment metrics for different compression methods.

Conclusion and discussion

When facing high dynamic range and large volumes of multidimensional (2D, 3D, 4D, 5D) biomedical data, both traditional and auto-encoder-based deep learning compression techniques have shown limited performance. SWIFT overcomes this limitation by first exploring and proving the high correlation of these data in the implicit function domain. Based on this correlation, SWIFT employs weight inheritance optimization and introduces an attention mechanism to account for the structural distribution characteristics of multimodal images. This innovative approach enables the attainment of both high compression ratios and high-fidelity compression of diverse biomedical data.

Compared to other INR-based compression methods, SWIFT significantly accelerates the algorithm's compression training time by introducing weight inheritance optimization, reducing the training time for large-scale datasets by an order of magnitude. SWIFT applies an adaptive grouping strategy for network training and encoding. Then, it formulates saliency maps according to the signal distribution of each group of data, adjusting the network training sampling strategy to guide better parameter allocation and achieve adaptive high-fidelity compression. Besides, since SWIFT is an implicit neural function mapping from spatial coordinates to signal values, it can innately incorporate attention mechanisms within it, with significantly-improved fidelity in lateral, axial, and temporal dimensions.

We validate the capabilities of SWIFT on several types of biomedical images, especially on 4D super-resolution microscopy data of live cells. When facing this kind of data with large size, dynamic structures and high resolution, SWIFT approach notably outperforms traditional H.265 and INR-based method, rendering itself a powerful and user-friendly data compression and transmission tool for diverse biomedical applications.

Methods

Global Similarity Sampling Grouping.

In the data grouping phase, to determine the optimal grouping configuration, we perform a global search using a pre-training neural network. For the original data *D*, the search is conducted using a customized evaluation criterion defined as follows:

$$PSNR = 10 * log_{10}\left(\frac{MAX^2}{MSE}\right)$$

$$CR = \frac{S_D}{S_C}$$

$$G = CR * PSNR$$

with the maximum dynamic range (*MAX*) and mean squared error (*MSE*), enabling us to obtain the peak signal-to-noise ratio (*PSNR*) for the pre-trained data. The pre-trained compression ratio (*CR*) is defined as the ratio between the size of the original data S_D and the size of the compressed data S_C. Ultimately, we consider the pre-training grouping to be optimal when the product *G* of *PSNR* and *CR* is maximized, striking a balance between compression ratio and fidelity.

Network optimization strategy based on attention mechanism.

We introduce an attention mechanism with learnable or traditional saliency maps to guide the adaptive allocation of parameters in INR, bridging the gap between INR and data and achieving improved data compression fidelity. Considering the characteristics of the signal distribution, we have two different saliency maps to cope with signals of different distribution types. Specifically, for data with multiple ROIs, dense signal distributions, and a demand for high structural detail, we use Gradient-weight Class Activation Mapping (Grad-CAM) [37] and Multi-Structure Region of Interest (MS-ROI) [38] techniques to create learnable saliency maps for the corresponding data, assigning probability values to each grid coordinate point, reflecting its proportionate importance in the data. In the process of generating saliency maps, we first apply a convolutional layer to the original data, denoted as x (with a size of $a \times b$), using a convolution filter of size $n \times n$. The convolution operation is represented by Equation 1, where W represents the learned filter.

$$y_{ij} = \sum_{a=0}^{n} \sum_{b=0}^{n} W_{ab} x_{(i+a)(j+b)}$$

In practice, multiple filters within each layer are learned in parallel, resulting in a three-dimensional feature map as the output of the convolutional layer, where the depth represents the number of filters. Subsequently, by utilizing the learned weights between the predicted class c and the feature map d, we train the Class Activation Maps (CAM) model to obtain saliency maps that capture the significance of the predicted class distribution.

$$M_c\left(x\ ,\ y\right) = \sum_{d \in D} w_d^c f_d(x, y)$$

$$P(c) = \frac{exp\left(\sum_{xy} M_c(x,y)\right)}{\sum_c exp\left(\sum_{xy} M_c(x,y)\right)}$$

where w_d^c is the learned weight of class c for feature map d. Training for CAM minimizes the cross entropy between objects' true probability distribution over classes (all mass given to the true class) and the predicted distribution. The probability P_c represents the likelihood of selecting the corresponding coordinate point for each training iteration. Besides, for sparse signals with faint intensity and lack of structural details, since the subsequent task analysis focuses only on their spatial location or intensity information, we use traditional saliency maps based on threshold divisions to prevent the loss of weak signals.

After obtaining the saliency maps, the coordinate vectors V guided by attention are further mapped to a high-dimensional embedding space using position encoding, enhancing perceptual quality. Formally, the encoding function employed in our approach is as follows:

$$\gamma(\mathrm{v}) = (\sin(2^0\pi\mathrm{v}), \cos(2^0\pi\mathrm{v}), \ldots, \sin(2^{L-1}\pi\mathrm{v}), \sin(2^{L-1}\pi\mathrm{v})).$$

Here γ is the mapping of the original coordinate vector V from R to R^{2L} and L is the number of frequencies used.

TSNE clustering dimensionality reduction and correlation analysis.

For correlation analysis in implicit function domain and image domain, we adopt AWC using Kolmogorov-Smirnov (KS) test as a criterion, and the statistic of KS test

is calculated as follows:

$$WC = sup_x|F_1(x) - F_2(x)|$$

where $F_1(x)$ and $F_2(x)$ are the Empirical Cumulative Distribution Function (ECDF) of the two samples, respectively. In the actual calculation, we respectively reshape the compressed network weights as well as the original images into one-dimensional vectors, and normalize the network parameter values and image pixel values to the range of 0 to 1, and then compute the KS statistic for their respective parameter distributions in the implicit function domain and the image domain, respectively. The KS test allows us to assess the correlation of each parameter across different samples. A smaller average KS statistic (AWC) indicates a higher correlation in the parameter value distributions between the two samples.

For the qualitative analysis of TSNE clustering dimensionality reduction, we adopt Rayleigh entropy as a criterion based on Fisher's criterion, which is calculated as follows:

$$H = -\sum(p_i * log(p_i))$$

where P_i is the probability that the sample point belongs to a category in the clustering result, indicating the probability of sample points being assigned to each cluster. A lower Raleigh entropy value indicates that the sample points are more concentrated and exhibit regular patterns in the clustering result.

Code Availability

The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

We are grateful to Yuxuan Zhao, Lanxin Zhu, and Zhaofei Wang for providing the experimental data.

This work was supported by the funding from National Key Research and Development Program of China (2022YFC3401102). National Natural Science Foundation of China (T2225014, 21874052, 21927802, 61860206009).

Reference

1. Dekkers, J.F. et al. Uncovering the mode of action of engineered T cells in patient cancer organoids. **41**, 60-69 (2023).
2. Han, Y., Li, M., Qiu, F., Zhang, M. & Zhang, Y.-H.J.N.c. Cell-permeable organic fluorescent probes for live-cell long-term super-resolution imaging reveal lysosome-mitochondrion interactions. **8**, 1307 (2017).
3. Li, A. et al. Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain. **330**, 1404-1408 (2010).
4. Boergens, K.M. et al. webKnossos: efficient online 3D data annotation for connectomics. **14**, 691-694 (2017).

5. Wiseman, S.J.N.N. A FAIR platform for data-sharing. **24**, 1640-1640 (2021).
6. Guo, S. et al. Smart imaging to empower brain-wide neuroscience at single-cell levels. **9**, 10 (2022).
7. Khayam, S.A.J.M.S.U. The discrete cosine transform (DCT): theory and application. **114**, 31 (2003).
8. Sze, V., Budagavi, M.J.I.T.o.C. & Technology, S.f.V. High throughput CABAC entropy coding in HEVC. **22**, 1778-1791 (2012).
9. Knuth, D.E.J.J.o.a. Dynamic huffman coding. **6**, 163-180 (1985).
10. Pesquita, C., Faria, D., Falcao, A.O., Lord, P. & Couto, F.M.J.P.c.b. Semantic similarity in biomedical ontologies. **5**, e1000443 (2009).
11. Chandrasekaran, D. & Mago, V.J.A.C.S. Evolution of semantic similarity—a survey. **54**, 1-37 (2021).
12. Wallace, G.K.J.I.t.o.c.e. The JPEG still picture compression standard. **38**, xviii-xxxiv (1992).
13. Wiegand, T., Sullivan, G.J., Bjontegaard, G., Luthra, A.J.I.T.o.c. & technology, s.f.v. Overview of the H. 264/AVC video coding standard. **13**, 560-576 (2003).
14. Sullivan, G.J., Ohm, J.-R., Han, W.-J., Wiegand, T.J.I.T.o.c. & technology, s.f.v. Overview of the high efficiency video coding (HEVC) standard. **22**, 1649-1668 (2012).
15. Dong, C., Deng, Y., Loy, C.C. & Tang, X. in Proceedings of the IEEE international conference on computer vision 576-584 (2015).
16. Theis, L., Shi, W., Cunningham, A. & Huszár, F.J.a.p.a. Lossy image compression with compressive autoencoders. (2017).
17. Duan, X., Liu, J. & Zhang, E.J.J.o.R.-T.I.P. Efficient image encryption and compression based on a VAE generative model. **16**, 765-773 (2019).
18. Agustsson, E., Tschannen, M., Mentzer, F., Timofte, R. & Gool, L.V. in Proceedings of the IEEE/CVF International Conference on Computer Vision 221-231 (2019).
19. Sitzmann, V., Martel, J., Bergman, A., Lindell, D. & Wetzstein, G.J.A.i.n.i.p.s. Implicit neural representations with periodic activation functions. **33**, 7462-7473 (2020).
20. Mildenhall, B. et al. Nerf: Representing scenes as neural radiance fields for view synthesis. **65**, 99-106 (2021).
21. Dupont, E., Goliński, A., Alizadeh, M., Teh, Y.W. & Doucet, A.J.a.p.a. Coin: Compression with implicit neural representations. (2021).
22. Chen, H. et al. Nerv: Neural representations for videos. **34**, 21557-21568 (2021).
23. Gallant in IEEE 1988 International Conference on Neural Networks 657-664 vol. 651 (IEEE, 1988).
24. Genova, K., Cole, F., Sud, A., Sarna, A. & Funkhouser, T.J.a.p.a. Deep structured implicit functions. **2** (2019).
25. Strümpler, Y., Postels, J., Yang, R., Gool, L.V. & Tombari, F. in European Conference on Computer Vision 74-91 (Springer, 2022).
26. Yang, R. et al. Sharing Massive Biomedical Data at Magnitudes Lower

Bandwidth Using Implicit Neural Function. 2022.2012. 2003.518948 (2022).
27. Dupont, E. et al. COIN++: Neural compression across modalities. (2022).
28. Sheibanifard, A. & Yu, H.J.A.S. A Novel Implicit Neural Representation for Volume Data. **13**, 3242 (2023).
29. Liang, R., Sun, H. & Vijaykumar, N.J.a.p.a. Coordx: Accelerating implicit neural representation with a split mlp architecture. (2022).
30. Xu, J. et al. NeSVoR: Implicit Neural Representation for Slice-to-Volume Reconstruction in MRI. (2023).
31. Stahel, W., Weisberg, S., Yohai, V.J., Stahel, W.A. & Zamar, R.H. in Directions in robust statistics and diagnostics: Part II 365-374 (Springer, 1991).
32. Wang, Z., Simoncelli, E.P. & Bovik, A.C. in The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, Vol. 2 1398-1402 (Ieee, 2003).
33. Peng, W.-J., Chen, C.-W., Lin, Y.-H. & Chen, M.-F. in 2022 IEEE International Conference on Imaging Systems and Techniques (IST) 1-6 (IEEE, 2022).
34. Kirillov, A. et al. Segment anything. (2023).
35. Patriarchi, T. et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. **360**, eaat4422 (2018).
36. Yasuda, R.J.C.o.i.n. Imaging spatiotemporal dynamics of neuronal signaling using fluorescence resonance energy transfer and fluorescence lifetime imaging microscopy. **16**, 551-561 (2006).
37. Selvaraju, R.R. et al. in Proceedings of the IEEE international conference on computer vision 618-626 (2017).
38. Prakash, A., Moran, N., Garber, S., DiLillo, A. & Storer, J. in 2017 Data Compression Conference (DCC) 250-259 (IEEE, 2017).

Competing interests

The authors declare no conflicts of interest.