ERnet: a tool for the semantic segmentation and quantitative analysis of endoplasmic reticulum topology

Meng Lu[1,4], Charles N. Christensen[1,2], Jana M. Weber[1,5], Tasuku Konno[3], Nino F. Läubli[1], Katharina M. Scherer[1], Edward Avezov[3], Pietro Lio[2], Alexei A. Lapkin[1], Gabriele S. Kaminski Schierle[1,4], Clemens F. Kaminski[1,4*]

[1]Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, UK.
[2]University of Cambridge, Department of Computer Science and Technology, Artificial Intelligence Group, JJ Thomson Ave, Cambridge, UK
[3]UK Dementia Research Institute at the University of Cambridge and Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0AH, United Kingdom.
[4]Cambridge Infinitus Research Centre, University of Cambridge, Cambridge CB3 0AS, UK.
[5]Current address: Delft Bioinformatics Lab, Intelligent Systems Department, Delft University of Technology, Van Mourik Broekmanweg 6, 2628XE, Delft, The Netherlands
*Corresponding author. Email: cfk23@cam.ac.uk

Abstract

The ability to quantify structural changes of the endoplasmic reticulum, ER, is crucial for understanding the structure and function of this organelle. However, the rapid movement and complex topology of ER networks make this challenging. Here, we construct a state-of-the-art semantic segmentation method we call ERnet for the automatic classification of sheet and tubular ER domains inside individual cells. Data are skeletonised and represented by connectivity graphs, enabling a precise and efficient quantification of network connectivity. ERnet generates metrics on topology and integrity of ER structures and quantifies structural change in response to genetic or metabolic manipulation. We validate ERnet using data obtained by various ER imaging methods from different cell types, as well as ground truth images of synthetic ER structures. ERnet can be deployed in an automatic high-throughput and unbiased fashion and identifies subtle changes in ER phenotypes that may inform on disease progression and response to therapy.

Introduction

The endoplasmic reticulum (ER) is the largest membranous structure in eukaryotic cells and acts as a platform for protein synthesis and quality control and for various organelle-interactions[1]. The ER consists of distinct domains including sheets and tubules, and features growth tips and tubular connections, so called three-way junctions. Perturbations to the ER structure and dynamics caused by genetic defects or metabolic stress have been associated with a variety of diseases[2], such as hereditary spastic paraplegias (HSPs) and Niemann Pick Disease type C (NPC). Hence, to understand the role of ER in diseases, it is important and necessary to characterise ER morphology comprehensively, which may provide powerful phenotypes to screen drugs against ER associated disorders. However, given the extent of the ER network and its complexity, the precise and quantitative measurement of ER topology and movement has remained challenging.

The ER network in a single cell consists of thousands of interconnected tubules that undergo constant rearrangements *via* processes including continuous tubular elongation, contraction, and fusion. Furthermore, there are rapid transitions between sheet and tubular domains with distinct putative functions[3]. Recently, capabilities have emerged to reveal such dynamic changes in ER topology in live cells, at sub-wavelength resolution[4]. Structured illumination microscopy (SIM), for example, can be used to resolve details of ER topology and its rapid remodelling process[5-7]. However, the data have only been interpreted qualitatively, without attempts to quantify ER topology or its structural changes precisely. Compared to other organelles, such as mitochondria and lysosomes, which are structurally simpler organelles that are often well separated from one another, the ER consists of highly convoluted and structurally connected domains. The task is further complicated by the fact that the signal to noise ratio of images obtained during live cell microscopy is often poor, while a clear differentiation of the organelle from its background is required to ensure successful segmentation into tubular and sheet domains. For moving structures, and time lapse imaging, this becomes a formidable task.

A number of machine learning-based methods have been developed for the segmentation of cells[8], mitochondria[9-10], and nuclei[11], which provide robust and precise classification of cell structures. However, to date, thresholding remains the standard method of use for ER segmentation[12-14]. Thresholding lacks both sensitivity and specificity, making quantitative conclusions hard to draw, especially in situations where image quality is compromised by noise. Alternative methods are based on labour intensive manual labelling of image data to generate specialised datasets for training of machine learning algorithms. These approaches do not generalise well to work with changing experimental setups or varying sample types[15] (Extended Data Fig. 1). An additional challenge for ER segmentation can be seen in temporal consistency. Conventional segmentation is performed on a frame-by-frame basis, and segmented structures in sequential (time-lapse) images lose temporal continuity and thereby cause artefacts[16]. Currently, there is no ER segmentation method capable of taking dynamic, spatial and temporal topology changes into consideration.

To address these difficulties, we developed ERnet, a deep learning software that automatically segments ER, classifies its domains into tubules and sheets, and quantifies structural and dynamic features in image sequences obtained from live cells. ERnet is trained with image datasets to model the domain knowledge of ER structures, *i.e.*, the shapes of tubules and sheets. As a result, it enables feature specific segmentation with enhanced robustness, specificity, and sensitivity regardless of the pixel intensity in the images. ERnet works on 2D data, but the quantitative results accurately describe the 3D structure of the organelle. After segmentation, ERnet quantifies topological features of the ER and recognises subtle changes in the ER structure and dynamics for various stress conditions, including gene knockout/knockdown, ATP depletion and calcium depletion etc. To validate the method, we tested the segmentation accuracy of ERnet on *in vitro* models subjected to different genetic and metabolic manipulations, including cells mimicking phenotypes of HSP and NPC. Two phenotypes were identified as sensitive readouts of the ER response in these models, namely the degree of fragmentation of ER networks and the heterogeneity in tubule connections. Both are indicators for the functional state of the ER network, and can be used, *e.g.*, to quantify the degree of disorganisation, shrinkage, and collapse of ER structures in models of disease. We show the versatility of ERnet by application to widefield imaging, confocal, and super-resolution microscopy data and test its performance in the presence of image noise. Furthermore, the method works in multiple cell lines. Minimal, or no retraining is required between different use scenarios. We provide ERnet as a user friendly, open-source software package with a graphical user interface (Extended Data Fig. 2 and user manual) to make it a broadly accessible tool for biologists and to promote ER-related research in basic science and clinical applications.

Results

Design and workflow of ERnet

The general design of ERnet is schematised in Fig. 1a. First, the reconstructed sequential images of the ER were segmented in ERnet, followed by the classification of ER structures into tubules and sheets. The tubular structure was further skeletonised using a surface axis thinning algorithm[17]. After this, the nodes and edges of the skeletonised ER were identified to plot a topology graph *via* a graph theory-based module[18]. In essence, the topology graph is a representation of ER tubules and junctions that provides a visual cue on the degrees of ER network connectivity and fragmentation (for an introductory explanation of graph theory concepts, see Extended Data Fig. 6).

Instead of relying on the commonly applied convolutional neural networks (CNN), our model builds upon a Vision Transformer architecture[19] which outperforms a comparable state-of-the-art CNN with higher classification accuracy and requires 4 times fewer computational resources. Key to our method is that, rather than paying attention to the spatial position of the nodes, it focuses on the ER's network features, *e.g.* the connectivity between nodes. This means that metrics such as number of ER fragments and the clustering of nodes into subregions can be extracted to provide quantitative metrics of ER topology and health.

To reduce the computational cost associated with the large data volumes generated by time sequenced imaging data, ERnet makes use of the backbone architecture of the Swin-Transformer reported in Liu et al. 2021[20]. Here, image frames in a temporal sequence are processed as 3D blocks, which permits the model to focus on key features that persist not only over the spatial, but also over temporal domains (Fig. 1b). These attributes make the method fast to execute and also very responsive to changing ER phenotypes.

Quantitative segmentation and analysis of ER topology

The ER is a highly dynamic structure and at any instance thousands of tubules move and change position, direction, and network connections. To quantify these intracellular changes, we first tested the performance of ERnet using SIM images of COS-7 cells. Fig. 2a shows a single frame of the ER (grey) from a set of sequential images captured from a COS-7 cell expressing mEmerald-Sec61b[5]. The performed segmentation successfully identified the whole ER structure, differentiated it from the cytosol background and further classified it into tubular (cyan) and sheet domains (yellow) (Fig. 2a). Then, the tubular ER was skeletonised from the whole structure and the nodes (tubule junctions, shown in red) and edges (tubules, green) were identified as two key topological components to map the network connectivity *via* the Python package Graph-tool[18].

SIM provide high spatial-temporal resolution of ER structures thus suitable for live cell imaging. A single pixel on the camera frame has a length scale of 42 nm in real space, almost a quarter of the average width of an ER tubule (~160 nm, measured as the average width on SIM images taken). This means that misclassification of a few, or even just one, image pixels can mean the difference between identification of a tubule as connected, or as disrupted. This leads to errors in the classification of network features, and *vice versa* to a bias when quantifying the network connectivity. In disease models, this could lead to erroneous phenotypes. The semantic segmentation of individual pixels from SIM images ensures the structural integrity of networks identified and prevents information loss, an improvement of traditional algorithms used in the past. Figs. 2a and b show how the method performs. A clear segmentation of ER structure (Fig. 2b) is achieved in regions containing dense ER tubule networks, as can be seen from the enlarged region indicated by the white box in Fig. 2a. This permits the distinction of tubules and their junctions in confined regions, measuring less than 300 nm across (highlighted by yellow dashed lines) with good structural detail. The segmented ER was then skeletonised (middle panel of Figs. 2a and b) and classified into edges (green tubules, right panel, Figs. 2a and b) and nodes (red spots, right panel, Figs. 2a and b). Finally, ERnet quantified the number of edges and nodes (top plot, Fig. 2c) and the percentage of areas covered by tubules and sheets (bottom plot, Fig. 2c), respectively, across the whole ER. Here, ER tubules were defined as linear branched structures and sheets as flat membrane cisternae as shown in Figs. 2a and d. Morphological features, such as the percentage of tubules/sheets among the whole ER, reflect ER status[3] and provide indications for possible ER defects. ER stress induced by an absence of the GTPase Rab7, which is known to modulate lysosome-ER contact sites, leads to the enlargement of ER sheets and the reduction of tubular domains in the cell periphery[21]. On the other hand, a depletion of protrudin, an ER reshaping protein, induces HSP associated ER dysfunctions by disrupting the sheet-to-tubule balance[22]. Therefore it is expected that the topological features of the ER, such as its connectivity, assortativity, or clustering coefficients, change for different phenotypes and with disease progression, a topic that is further explored in subsequent sections. It is worth highlighting that, although the ER tubular network underwent stark morphology changes (Supplementary Video 1) and demonstrated fluctuations in the number of nodes and edges (top panel, Fig. 2c) within individual recordings, its tubule and sheet percentages among the whole ER remained stable (bottom panel, Fig. 2c), which suggests that the overall connections do not change in the absence of a stimuli.

In the canonical model of ER structures, ER tubules radiate from sheets towards the cell periphery[4], and the two structures are thought not to overlap. However, we observed that tubular structures also reside on the ER sheets themselves (Fig. 2d and Supplementary Video 2), which in what follows we refer to as 'sheet-based tubules' (SBTs), and which are clearly distinguished by ERnet as seen in Fig. 2d and Supplementary Videos 2 and 3. Like peripheral tubules, SBTs undergo rapid elongations and contractions, which can either lead to new tubular connections (blue arrows), or separations (grey arrows) (bottom panel, Fig. 2d). A subsequent 3D reconstruction of SIM image sections further validated that such tubules are directly attached to the sheets and are not the result of a projection view artefact (Fig. 2e, Extended Data Fig. 3, Supplementary Video 4). Analysis of 500 cells showed that this phenomenon is a common feature of the ER network (Fig. 2f).

In silico validation of ERnet

To examine the accuracy of ERnet, we generated synthetic ground truth data on which the performance of the method could be tested. First, we generated data to test semantic segmentation performance. To do this, we used real SIM data of ER networks on which we applied the well-established Trainable Weka segmentation machine learning algorithm[15]. This produced ground truth images for which ER structures were classified into tubules, sheets, and SBTs (Fig. 3a).

The same SIM images on which the above Weka approach was used to generate the ground truth data were then processed by ERnet and the results compared pixel-by-pixel. The ground truth test demonstrated a pixel accuracy for ERnet segmentation of between 92 and 99 %

compared to the ground truth data (Fig. 3c). In another test, we used the segmented images obtained with the Trainable Weka algorithm and fed this as input to ERnet. In this case again, the result was nearly identical to the ground truth.

In addition to the ground truth test for semantic segmentation, we also tested the accuracy of the connectivity analyses. To do this, we generated ground truth data of ER tubular domains by creating synthetic ER skeletons. We then widened and blurred the skeletons and added image noise to mimic ER structures recorded with optical microscopy (for details on this process, see Extended Data Fig. 4a and descriptive caption). After this, the synthetic images were processed by ERnet to identify nodes and edges and derive metrics for ER connectivity (Figs. 3d and e). ERnet reached accuracies ranging from 96 to 99 % for the identification of nodes and edges (Fig. 3f). Even in dense regions of the tubular network (zoomed in regions, Fig. 3e), ERnet still achieved a high precision to capture nodes and edges with little difference found between the ERnet result and the ground truth data. Additionally, we quantified the differences in the connectivity metrics obtained from ERnet and ground truth data (Fig. 3f). Since the assortativity metric ranges over very small scales, e.g. from -0.05 to 0.08, even minor changes in connectivity can lead to large fluctuations of the former. Nevertheless, observed changes in metrics are still significantly smaller than those associated with the varying phenotypes reported in the following context (Fig. 6).

Next, we tested the performance of ERnet on ground truth images in which we added variable levels of noise (Extended Data Fig. 4). The purpose was to provide a metric with which a user can decide upfront, whether a given dataset obtained on a microscope is of sufficient quality to trust the ERnet output. We found that ERnet produced repeatedly reliable outputs for both connectivity and topology features for image data featuring signal-to-noise ratios better than ca. 5 (Extended Data Fig. 4). By analysing the SNR obtained with a given experimental setup, users can objectively assess the quality of the segmentation results, irrespective of where and how the data were obtained.

ERnet performs on various cell types and imaging modalities.

To demonstrate the versatility and robustness of ERnet in different research scenarios, we validated the method on a on a range of datasets obtained by us and others. Fig. 4 presents the analysis of images obtained using different microscopy techniques including widefield, confocal, and Airyscan microscopy. Even though ERnet's precision may depend on the spatial resolution of the corresponding images, it performed well for all imaging techniques with all the tubules and sheets clearly classified and quantified (Source Data Fig. 4). Furthermore, we also performed validation tests for varying cell types commonly used in cell biology research, such as HEK, CHO, SH-SY5Y cells, and primary cultures of hippocampal neurons and glial cells derived from embryonic rats (Source Data Fig. 4). Further data from plant cells[13] and publicly available data sets[23-24] published by other authors using different experimental setups are shown in Extended Data Fig. 5. Although the specific ER phenotypes varied among the cell types, ERnet was able to robustly identify the corresponding tubular and sheet domains and performed subsequent quantitative analyses following segmentation. For none of these scenarios the model had to be retrained and no pre-processing of the raw data was necessary before segmentation by ERnet, demonstrating the generality of the model and its ease of application.

ERnet provides detailed connectivity data on ER networks.

ERnet can be used to quantify the connectivity of edges and nodes before plotting a corresponding connectivity graph (Fig. 5a). The connectivity graph highlights that the network of the ER largely constitutes of three-way junctions (red nodes, Fig. 5a) while the ER edges are capped with growth ends (green nodes, Fig. 5a).

To assess the integrity of the ER, we defined each disconnected ER region as a fragment. Although the number of fragments during ER reshaping fluctuates (Fig. 5b), ERnet reveals that in a typical healthy cell, the majority of all edges and nodes are contained in a single large fragment at all times (over 92% of all the 3000 nodes and 95% of all the 4000 edges in the shown example). As quantitative parameters, we defined node and edge assembly ratios (the number of nodes or edges in the largest fragment divided by the total number of nodes or edges, respectively), see Fig. 5c. Per definition, these values range from close to 0 (fully fragmented ER) to 1 (fully connected). Additionally, ERnet quantified the degrees of the ER nodes, *i.e.*, how many edges (tubules) connect to each node (junction). As shown in Fig. 5d, three-way junctions are the most abundant and represent 66% of all junction types in this example. Despite the prevailing model of ER morphology, where three-way junctions interconnect to form the whole ER tubular network, ERnet also identified nodes connected with more than three edges (tubules), *i.e*., multi-way junctions. The presence of multi-way junctions indicates the heterogeneous connectivity of ER tubules that are organised in a higher order of complexity than previously assumed.

Next, the assortativity and clustering coefficients (Figs. 5e and f), that describe connectivity patterns of nodes, were calculated based on the above metrics. The assortativity coefficient measures the tendency of nodes to connect with others of the same degree[25]. In a network with a high assortativity coefficient most nodes are connected in a similar way with their neighbours (e.g. *via* 3 way junctions). The clustering coefficient on the other hand reflects the distribution of nodes within the whole network (e.g. clusters of multiway junctions may be separated from other clusters by junctions of lower degree). For a graphical explanation of these concepts, the reader is referred to Extended Data Fig. 6.

Assortativity coefficients range from -1 (fully heterogeneous network in connectivity, i.e. nodes only connect with those of different degrees) to +1 (fully homogeneous network in connectivity, i.e. nodes only connect with those of same degree). Similarly, for clustering coefficients, 1 describes a network in which all the nodes and edges are clustered while 0 refers to no clustering. Fig. 5e shows the ER as a weak assortative network, which suggests a tendency, albeit a weak one, of nodes to connect with nodes of the same degree. In Fig. 5f, we show how the degree of clustering can change over time in an ER network. Tubules and junctions reorganise themselves rapidly, both within localised and global domains. Frequent events include the merging of multiple tubules forming clusters of nodes, but these then disassemble transiently. Overall, the data indicate that the network features a high degree of structural homogeneity and local clustering is not a dominating feature to affect the overall phenotype.

To further investigate the structural dynamics of the ER, we tracked the lifetime of multi-way junctions and their transitions from multi-way to three-way junctions. Figs. 5g and h show the rapid transitions between three-way (yellow arrows) and multi-way junctions (blue arrows) driven by ER tubule reshaping. As shown in these cases, the formation of four or five-way junctions need simultaneous connections of more than three tubules at the same junction, which occurs with a lower chance than the formation of a three-way junction that only requires the connection of three tubules. Additionally, any movement of a tubule away from its multi-way junction can lead to the collapse of this junction and the generation of at least two three-way junctions. Therefore, as shown in Fig. 5i, the average lifetime of a multi-way junction is much shorter, *i.e.*, less than a third (9.0 s *vs* 30.8 s) of that of a three-way junction.

We also examined whether our 2D network approach is valid to segment ER structures, which are 3 dimensional in nature. We performed two different tests in both COS-7 and U2OS cells, which are the canonical models in fluorescence microscopy-based studies of the ER, and for which ERnet was developed. For these flat cell types we could confirm that a 2D analysis is sufficient to represent the ER network topology (Extended Data Fig. 7).

ERnet can characterise complex ER phenotypes.

ER morphological defects caused by mutations in genes encoding ER-reshaping proteins or by metabolic perturbations have been linked to a variety of human disease[1,2,4]. However, the exact phenotypical ER disruption under these conditions has not yet been sufficiently characterised. Using ERnet, we first analysed the ER morphological defects in stress models mimicking the ER phenotypes in two neurodegenerative diseases, namely HSPs and NPC. The inherited neurological disorder HSPs can be characterised by progressive lower-limb weakness and muscle stiffness, which are caused by mutations in genes encoding ER reshaping proteins such as atlastin (ATL)[26] and protrudin[27]. We used ERnet to examine the ER morphology defects in individual cells of different models by measuring two topological features, *i.e.*, the degree of ER tubule fragmentation and the heterogeneity in these tubular connections. Compared with control cells, an ATL knock-out (KO)[28] leads to a collapse of the ER network integrity. Such ER fragmentation was clearly revealed in ATL KO cells by the increasing number of fragments and a 20-fold reduction of the node assembly ratio (90% in control *vs*. 4.5% in ATL KO) (Fig. 6a and Supplementary Video 5 and 6). ERnet also highlighted that the lack of ATL significantly altered the connectivity in ER tubular network, as witnessed by a reduced percentage of three-way junctions among all the nodes (22% *vs*. 65% in control) and by the disorganised connectivity (-0.25 in assortativity). These measurements provide quantitative rather than descriptive evidence of ATL's role in ER tubular network formation, which has previously been reported to be crucial for the fusion of ER membranes and thus the formation of continuous networks[26]. With these quantitative analyses, we can compare morphological defects caused by different treatments. In another model of HSPs, depletion of protrudin (Extended Data Fig. 8) also resulted in ER tubular network fragmentation (350 fragments)

(Supplementary Video 7) and in disorganised connectivity, however, to a lesser extent. A further metric suitable for the comparison of ER health under different treatments is the size of the ER, which is revealed by the connectivity graph. An ATL KO cell that was more fragmented than a protrudin KD cell suffered from a more severe shrinkage of the ER with a smaller number of nodes and edges (Fig. 6a), indicating that ER membranes may be degraded or recycled in response to stresses.

Next, we induced cholesterol accumulation in lysosomes by U18666A administration to the cell, which induces a blockage of the cholesterol transfer from lysosomes to the ER in NPC[29]. The accumulation of cholesterol in lysosomes leads to lysosome deposition in perinuclear regions and, therefore, affects the ER structure and distribution[3]. However, the exact ER morphological defects have not yet been characterised. ERnet revealed that the ER of U18666A-treated cells features a disassortative network (-0.34) and its low node assembly ratio (2.6%) suggests a highly fragmented structure (Fig. 6a and b, Supplementary Video 8), which highlights that lysosomal defects can strongly affect the ER and thus provides for a useful tool with which to improve an understanding of organelle dysfunction in NPC.

Finally, we tested the performance of ERnet in cells upon ER collapse under metabolic manipulations that significantly affect the overall homeostasis inside the cell. SIM video showed that the ER largely loses its dynamic reshaping capabilities upon the administration of store-operated calcium entry (SOCE) inhibitor SKF96365[30] (Supplementary Video 9). In the connectivity graph, the ER became largely fragmented and featured as a disassortative network (Fig. 6a and b). Compared with SKF96365, NaN_3 depletes ATP[31] thus capping support for all the energy consuming processes inside the cell, including ER tubule elongation, retraction, and membrane fusion. Therefore, ATP depletion by NaN_3 was expected to significantly affect the structural dynamics of the ER. ERnet revealed the level of ER network fragmentation resulting from a lack of ATP (Fig. 6a and b, Supplementary Video 10); however, the phenotypes were not equivalent to those observed upon depletion of ER reshaping proteins: for example, the node assembly ratio in ATP depleted cells was found to be nearly 4-fold of that in ATL KO cells (0.19 *vs* 0.05).

Overall, while ERnet provides a quantitative assessment of overall network topology it is also sensitive enough to detect subtle changes in local ER morphology, valuable attributes in the investigation and differentiation of ER-related phenotypes of disease.

Discussion

A measurement of cellular organelle properties such as shape, position, and mobility provides a quantitative basis for analysing the structure and function of organelles in both fundamental and therapeutic research. Here, we introduce ERnet, a versatile tool that performs robust and precise segmentations and analyses of ER structures under a variety of conditions.

The accuracy of ERnet's semantic segmentation algorithm is a result of its model design. In contrast to state-of-the-art CNN models commonly used for image segmentation, ERnet is constructed in a Vision Transformer architecture that outperforms CNNs in terms of image classification accuracy and requires much smaller computational resources[19, 32]. Another advantage of our design is a capability for temporal domain analyses of objects in sequenced image data. We integrated two attention mechanisms: multi-head self-attention[33] and channel attention[34] into the Transformer architecture. These mechanisms greatly enhance the learning ability of ERnet in classifying ER structures in the spatio-temporal domain. While machine learning methods have previously been implemented for denoising images of ER structures[35]; reconstructing ER structures based on electro-microscopy images[36]; and identification ER stress marker-whorls[37], ERnet is capable of video-rate image segmentation and analysis of live cells, further extending the deep learning toolbox for biomedical research.

Through application of ERnet, we were able to characterise and quantify the structural features of dynamic ER networks. First, we found that the dominance of three-way junctions is a necessity to produce a continuous ER network that can spread throughout the cell. Whilst we observed a prevalence of three-way junctions, we found that 20% of healthy ER furthermore consists of multi-way junctions (degree > 3). In contrast, all the stress manipulations of ER morphology, including models of HSPs and NPC, resulted in the fragmentation of ER structures to varying extents (Fig. 6).

Nixon Abel et al., 2016 and Pain et al., 2019 also made use of microscopy data in the analysis of ER dynamics. However, their work focused on a very different set of ER phenomena than what we present here. Nixon Abel et al, 2016 analysed the transient dynamics of individual tubules (e.g. lateral tubule oscillations). Pain and colleagues designed AnalyzER to extract metrics of plant ER tubules, such as their width, length, and cross-sectional area. In contrast, our work focuses on global network topology and integrity, which are key features associated with physiological and stress states. Our aim is to provide a robust and powerful tool for the investigation of therapeutic strategies against ER associated disease. Apart from this, ERnet, driven by deep learning to classify ER structures, can identify subtle changes in the whole ER and display the difference in quantitative plots. The connectivity graph is a unique feature of ERnet. It is a visual tool to display the connected parts of the ER and provides for a rapid visual cue on the degree of network integrity.

An advantage of the use of deep learning in biological imaging is that it facilitates the discovery of novel biological phenomena. The sensitivity of ERnet to changing structural features led to the identification of SBTs. These ER components share similar structures and dynamics with the tubules that radiate from the sheet domains towards the periphery of the cell, however, their existence in the sheet domain greatly extends the coverage of the tubular ER towards the cell centre and even close to the nucleus. We note that SBTs are evident also in data presented in previous reports, such as Schroeder and colleagues[38] (see for example Figs. 1E and H; Fig. 3B; Fig. 4A), but the phenomenon was not recognised specifically. In our method, SBTs are classified in addition to sheets and tubules on their own. Whilst ERnet can be used with any imaging technique, conventional or superresolved, its ability to detect and classify SBTs does depend on signal-to-noise ratio and image resolution. Therefore, some differences are expected in output produced from very different imaging methods. We also note that ER topology can vary significantly from cell to cell, and do not recommend conclusions to be drawn from data that are not representative of the whole cell population. How the sheet-based tubules are regulated in both physiological and pathological conditions will be an important question for future studies.

Like all segmentation and classification methods, including those performed by humans, ERnet is necessarily limited by the quality of the input data. We found that for signal to noise ratios above around 5, ERnet reliably quantifies topology structures for any microscopy method of appropriate image resolution. Because we optimised ERnet for high throughput analaysis the algorithm treats ER networks as 2 dimensional structures for computational efficiency. Whilst we saw no problems with this for the cell types we analysed, one needs to be cautious when applying the method to highly 3-dimensional networks. ERnet could be extended to 3 dimensions and integrated with further organelle analysis tools, for example methods for the characterisation of lysosomes and mitochondria, to permit comprehensive investigations of organelle-organelle interactions and their role in the development, ageing, and degeneration of cells.

We believe our work demonstrates an efficient tool for precise structure segmentation and multi-parameter analysis of ER phenotypes. Its user-friendly graphical interface and automatic batch processing capabilities obviate the need for manual annotation.

Acknowledgments

We thank Dr Ana Isabel Fernández Villegas and Yuqing Feng for helping with the cell culture. We thank Dr Edward Ward for helping with the image processing. We thank Prof. Junjie Hu (Chinese Academy of Sciences, China) for giving us the ATL KO cell line.

Author contributions

M.L. designed, conducted, and interpreted experiments, and wrote the article. M.L. and C.N.C. developed the computational pipeline for ERnet. C.N.C. developed the core model of ERnet. J.M.W. conceptualised, developed and wrote the graph-based analysis of the ER. T.K. supported the versatility test. N.L., K.S., E.A., P.L., A.L. and G.S.K.S. gave advice and edited the article. C.F.K. supervised the research, coordinated and conceptualized the study, and wrote the article.

Funding

This research was funded by Infinitus (China) Company Ltd (supporting M.L., C.F.K. and G.S.K.S.); a Wellcome Trust Programme Grant (085314/Z/08/Z, to G.S.K.S. and C.F.K); a Swiss National Science Foundation Career Grant (P2EZP2_199843, to N.F.L.); a research fellowship from the Deutsche Forschungsgemeinschaft (DFG; SCHE 1672/2-1, to K.S.) and pump-prime funding from the Integrated Biological Imaging Network (IBIN; G106925, to K.S.); the UK Dementia Research Institute which receives its funding from UK DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK (supporting T.K., E.A. and C.F.K) and Alzheimer's Society 525 (AS-PhD-19a-015) supporting E.A. J.M.W.'s PhD scholarship was funded by the Department of Chemical Engineering and Biotechnology, University of Cambridge.

Competing interests

The authors declare no conflict of interest.

Figure Legends

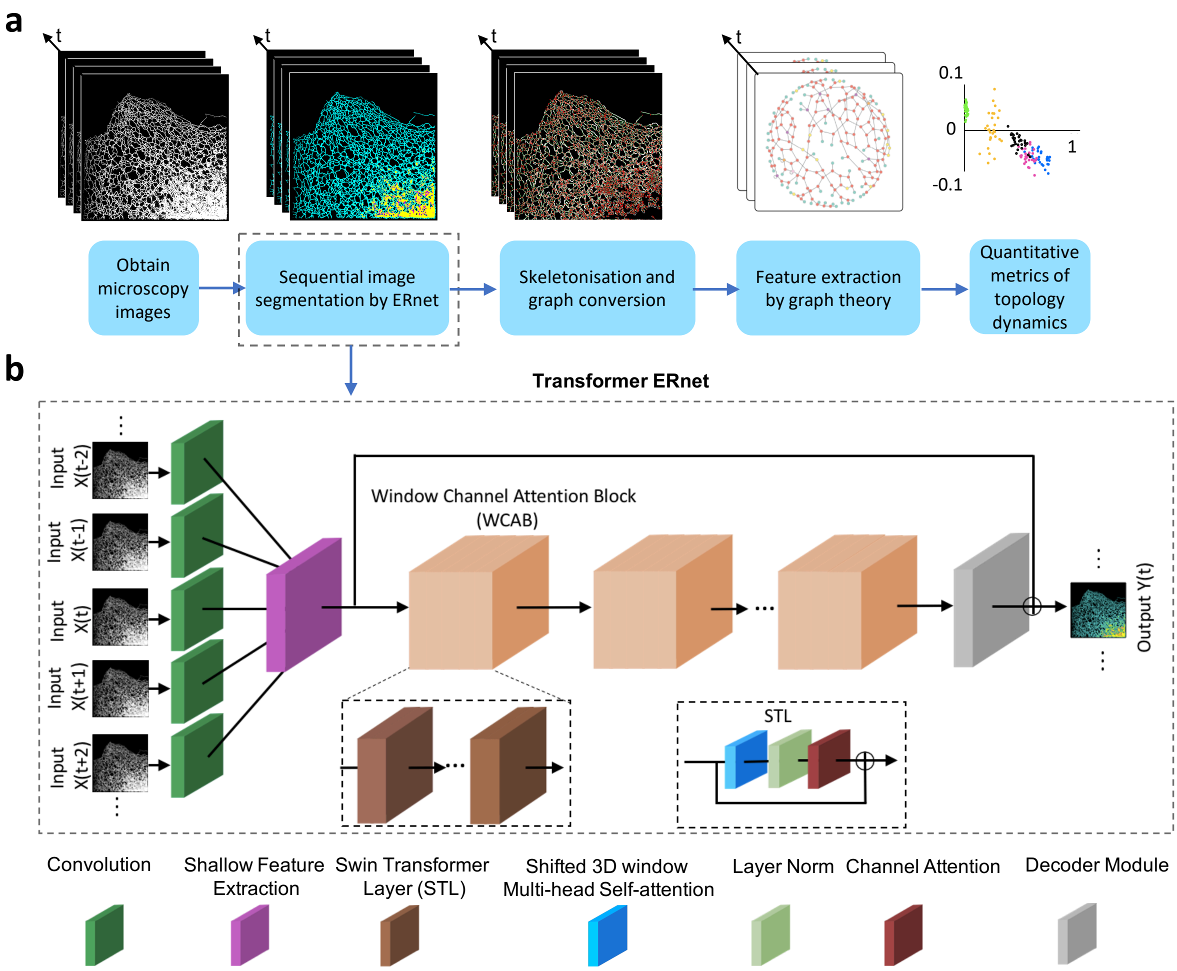


Fig. 1: Workflow of ER structure segmentation and ERnet construction.

a. The processing pipeline of ER segmentation and analysis. Time-lapse SIM images were first segmented by ERnet to classify the tubules and sheets. The tubular network of ER after segmentation was further skeletonised and the nodes and edges were identified to plot the connectivity graph. Using graph theory-based methods, we quantified the metrics of the ER network features that describe the topology and dynamics.

b. The Transformer based architecture of ERnet. A moving window loads adjacent frames (X_{t-2} to X_{t+2}) as inputs from the time-lapse images into ERnet. A shallow feature extraction module then projects the input into a feature map which is followed by a sequence of residual blocks denoted with Window Channel Attention Block (WCAB). Inside each WCAB, there is a sequence of Swin Transformer Layers (STLs). For details, see Methods.

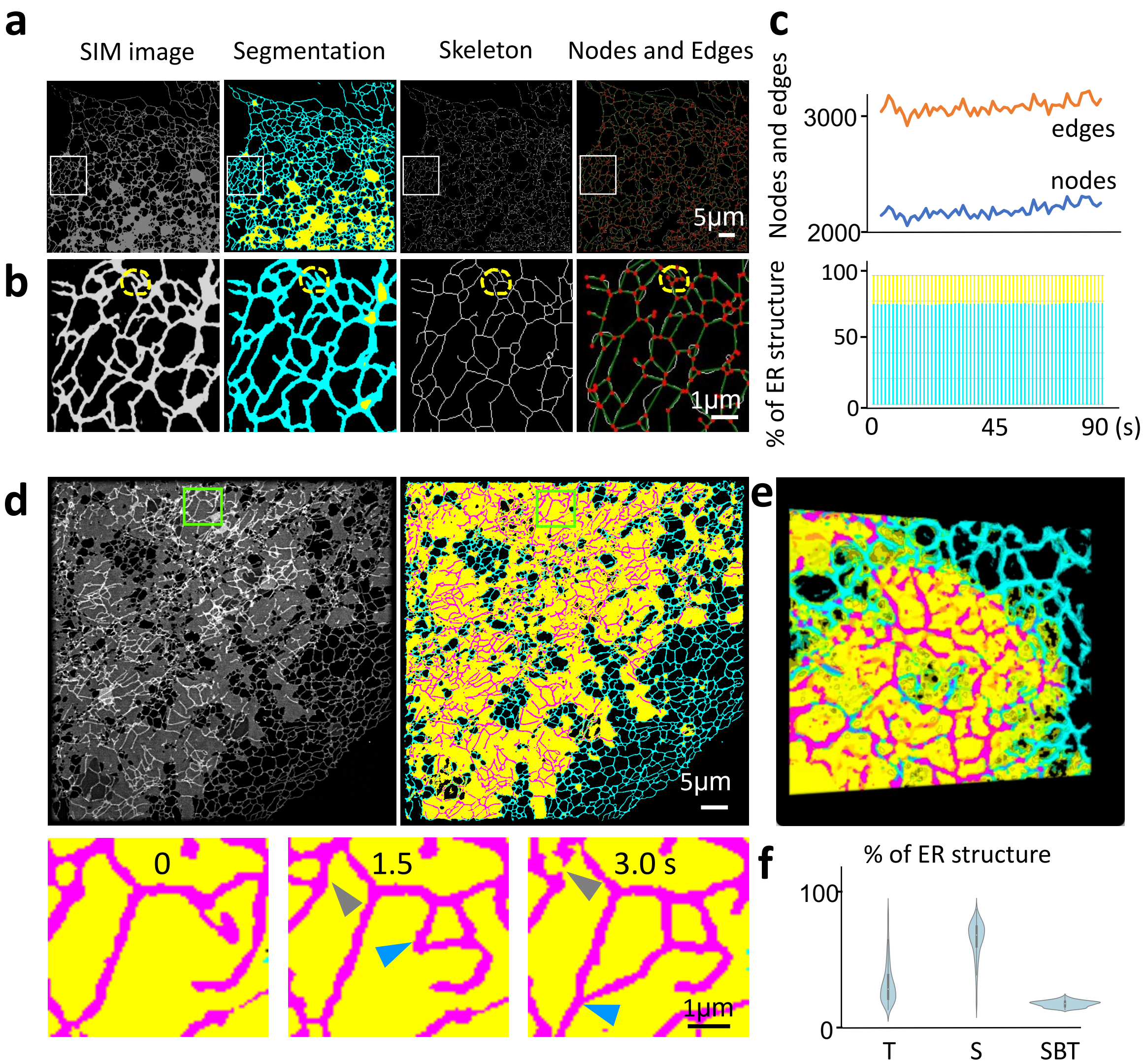


Fig. 2: Semantic segmentation of ER and classification of tubules and sheets.

a. An example of a segmentation result from video-rate SIM images of the ER. From left to right: 1) SIM image, 2) segmentation of ER tubular (cyan), sheet (yellow) and sheet-based tubule (magenta) region, 3) skeletonisation of the tubular domain, and 4) identification of nodes (red spots) and edges (green lines) based on the skeleton structure. Scale bar: 5 μm.

b. Zoomed-in regions of the above panel. The yellow dashed circles indicate nodes that are closely positioned but can still be identified by ERnet. Scale bar: 2 μm.

c. Quantitative analysis of the ER shown in (a). Top panel: quantification of edges and nodes of the ER tubules of the time-lapse frames. Bottom panel: percentage of the ER tubules (cyan) and sheet (yellow) of the time-lapse frames (1.5s/frame). See Source Data Fig. 2c.

d. A representative frame from time-lapse images shows the structure of sheet-based tubules (1.5s/frame). Top left panel: a SIM image of the ER structure. Top right panel:

segmentation of the three ER structures: SBTs (magenta), sheet (yellow), tubules (cyan). Bottom panel: three sequential frames showing the dynamic reshaping of sheet-based tubules from the above green boxed region. Blue arrows indicate a continuously elongating sheet-based tubule and grey arrows indicate a retraction. Scale bars: 5 μm (top) and 2 μm (bottom). See Source Data Fig. 2d for quantitative analysis.

e. Volumetric view of 3D reconstruction of the sectioning SIM showing that the SBTs (magenta) are embedded in sheet domains (yellow). Scale bar: 2 μm (bottom).

f. Violin plots of the percentages of tubules (T), sheets (S) and sheet-based tubules (SBT) in COS-7 cells (*N*=500), showing that the presence of the sheet-based tubules is a common feature of the ER network. In the violin plots, the white dot represents the median value of the data; the thick bar represents the interquartile range and the thin bar represents the rest data distribution. See Source Data Fig. 2f.

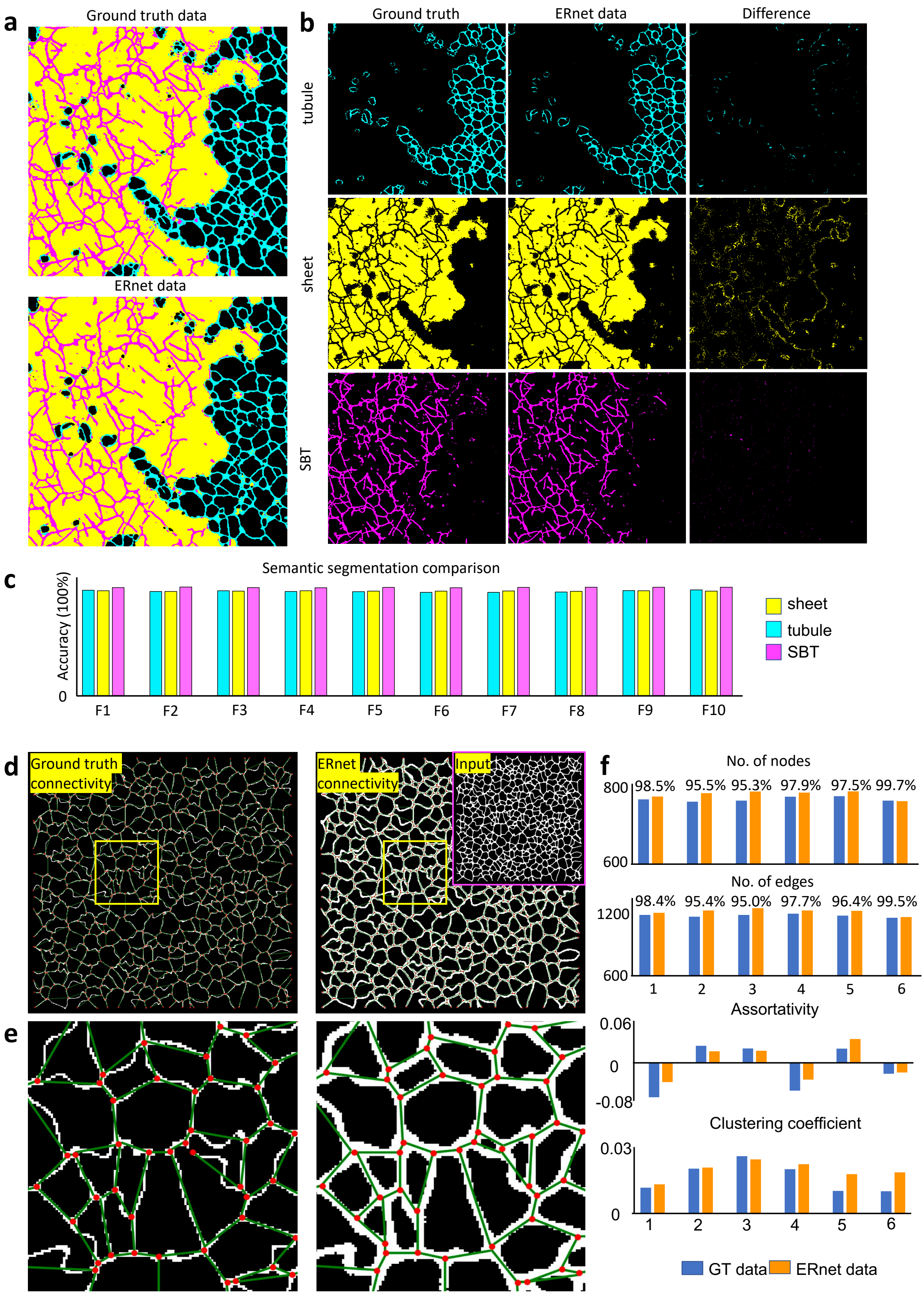


Figure 3. Ground truth test for segmentation and connectivity.

a. Comparison between ground truth data and ERnet data for ER tubular (cyan), sheet (yellow), and SBT (magenta) domains. Comparisons were repeated three independently with similar results shown in (c).

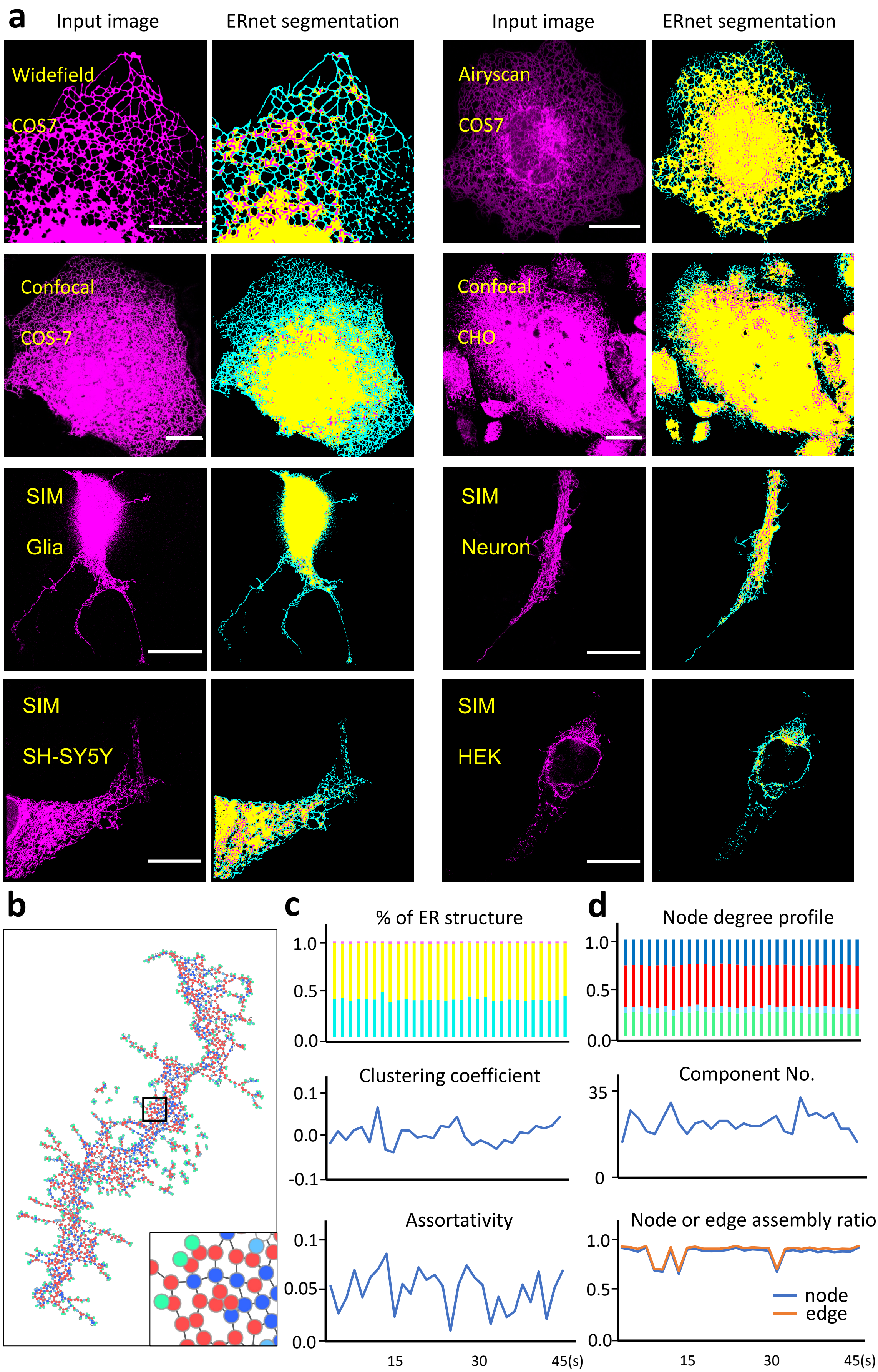
b. Comparison of each channel for the image data above.

c. Quantification of the pixel differences from the three image channels. F1-F10 are frames 1 to 10, respectively, in sequentially recorded ER images. See Source Data Fig. 3c.

d. Comparison of ground truth data (synthetic ER tubular network) and ERnet results. The top right inset and framed in a magenta box presents the whole field of view of the ground truth data which was an input into ERnet.

e. A zoomed in region of the highlighted sections in (a) showing that the connectivity revealed by ERnet is nearly identical to the GT. Red spots: nodes; green lines: edges.

f. Comparison of the connectivity metrics. GT data: ground truth data. Number in x axis indicate the image sample number. See Source Data Fig. 3f.



621
622 **Fig. 4: Robust performance of ERnet in versatility test.**

a. A variety of cell lines with different ER morphologies were imaged by different microscopy techniques to investigate the robustness and versatility of ERnet. ER structures of COS-7, HEK, CHO, SH-SY5Y, primary cultures of hippocampal neurons and glial cells were tested, as well as images acquired by widefield, confocal and Airyscan microscopy (1.5s/frame). Scale bars: 20 µm.

b. The topology of an ER tubular network of the COS-7 cell from the confocal image shown in (a) is represented by a connectivity graph. Nodes of different degrees are labeled with different colours: green (degree 1), light blue (degree 2), red (degree 3), dark blue (degree >3). Bottom right: a zoomed-in region of the black boxed part in the connectivity graph, demonstrating the complex connectivity revealed by ERnet from confocal microscopy image. The following analysis of Fig. 4c and d is based on this image data.

c. Quantitative analysis of the ER structure of the above image data reveals the topology features of ER tubular network. Top: percentage of the ER tubules (cyan), sheet (yellow), and SBTs (magenta) of the time-lapse frames (43.5 s at 1.5 s/frame). Middle and bottom: changes of assortativity and clustering coefficients in time-lapse images. See Source Data Fig. 4c and d.

d. Quantitative analysis of the connectivity of the ER tubular network in the above cell. Top: quantification of the nodes of different degrees, showing a dominance of third-degree nodes (three-way junctions). Middle: number of components (ER fragments) in time-lapse images. Bottom: changes of the node/edge ratio over time.

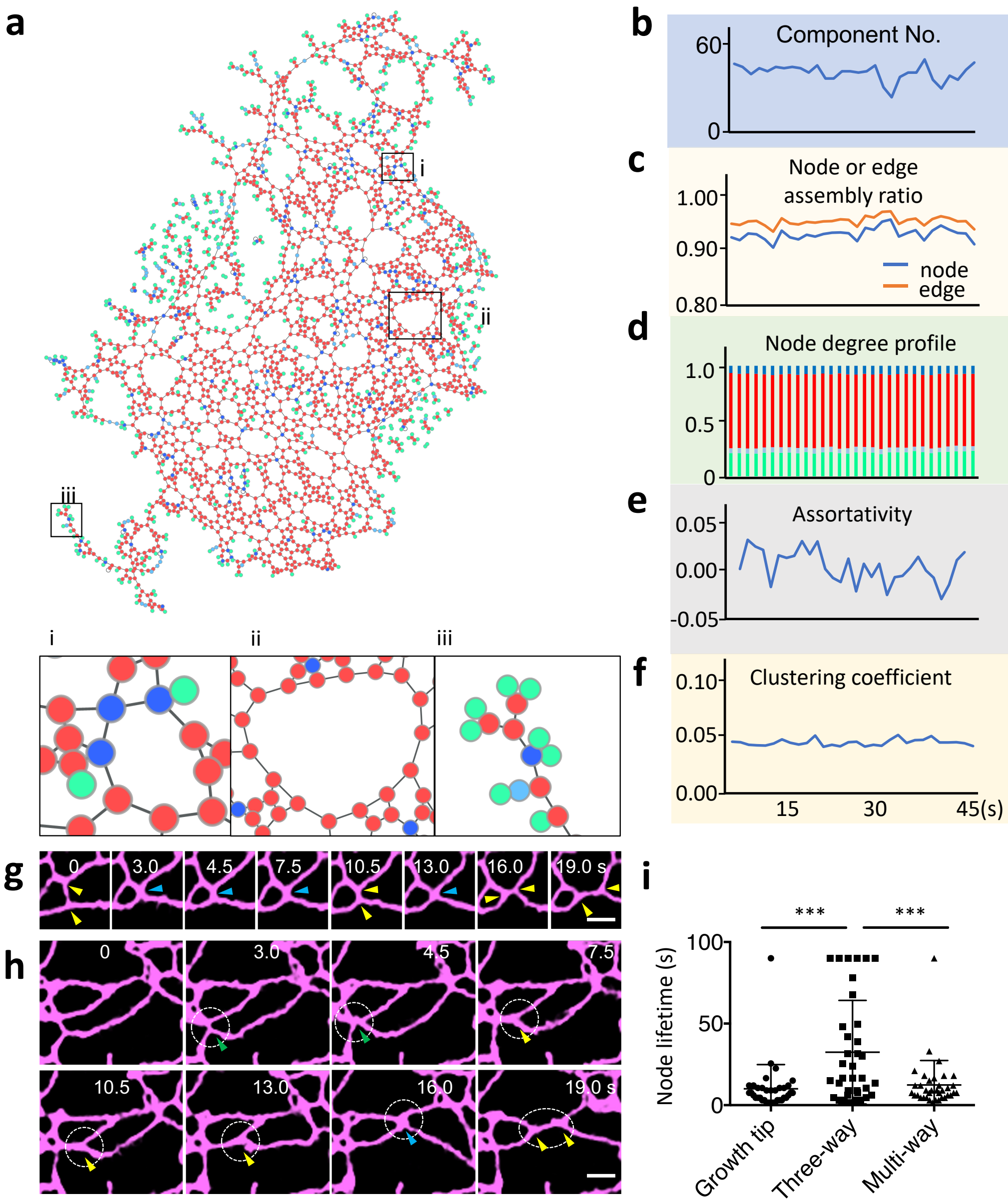


Fig. 5: Quantitative analysis by ERnet reveals the complex connectivity of ER tubular network.

a. The topology of an ER tubular network is represented by a connectivity graph. i: a representative region of multi-way junctions (dark blue spots), ii: a polygonal structure organized by three-way junctions and tubules, iii: a representative region of ER tubular growth tips (green spots).

b-f. Quantitative analysis of the cell shown in (a) over a time window of 45 s at 1.5s/frame. See Source Data Fig. 5-f.

b. Number of components (ER fragments) in time-lapse images.

c. Changes of the node or edge assembly ratio over time.

d. Quantification of the nodes of different degrees, showing a dominance of third-degree nodes (three-way junctions). Same colour scheme as in (a).

e-f. Changes of assortativity and clustering coefficients in time-lapse images.

g-h. Examples of transitions between three-way (yellow arrows) and multi-way junctions (yellow arrows: three-way, blue arrows: four-way, green arrows: five-way) junctions. Scale bar: 1 μm.

i. Quantification of the lifetime of junctions (nodes) with different degrees. Data are presented as mean ± SEM， $^{***}P < 0.001$, Tukey's one-way ANOVA. N=12 events per condition per experiment from three independent experiments and 36 events per condition are analysed in total. P value: growth tip vs multi-way: 0.8947, growth tip vs three-way: 0.0001, three-way vs multi way: 0.0006. See Source Data Fig. 5i.

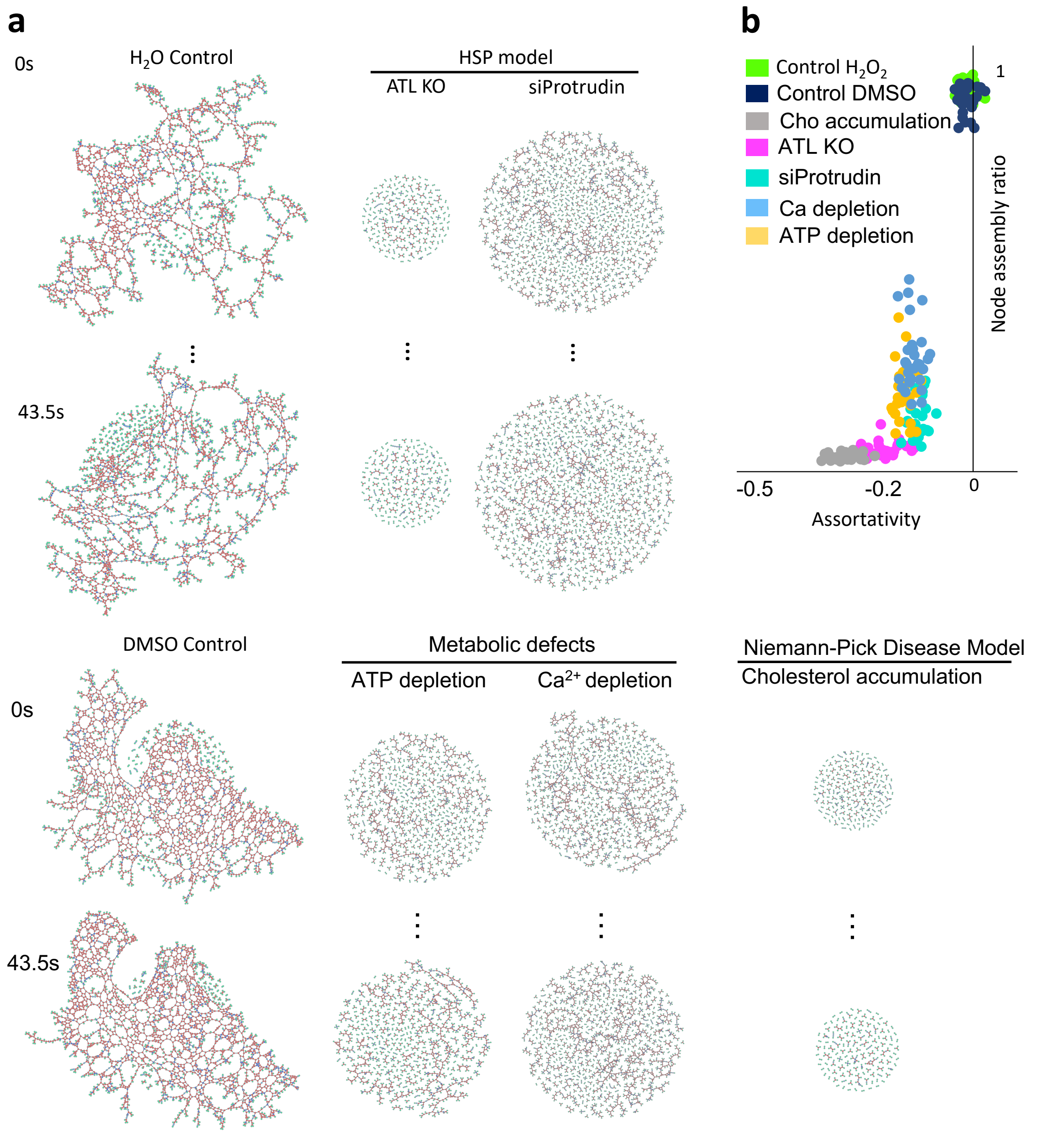


Fig. 6: Quantitative analysis of ER phenotypic characteristics in disease associated models.

a. Connectivity graphs of ER structures in models mimicking phenotypes of HSPs and NPC and metabolic stress induced by calcium and ATP depletion. Nodes of different degrees are labeled with different colours: green (degree 1), light blue (degree 2), red (degree 3), dark blue (degree >3). Note that the graphs represent data from the whole field-of-view imaged in the microscope. The connectivity graphs are symbolic reductions of ER networks for easy visualisation of network topology. The graphs should not be mistaken for actual spatial representations of ER networks. Highly connected networks (controls, left column) appear more amorphous than strongly fragmented networks (stressed cells, middle columns). Raw image data are presented in Video 5-11 (1.5s/frame).

b. Topological features of the ER tubular network in above conditions were quantitatively analysed by ERnet. The effects on ER structures from different treatments can be directly visualised and compared by plotting the distribution of node assembly ratio (y axis) and assortativity coefficient (x axis). The analysis of ER phenotype, such as that in ATL KO cells, demonstrated a severe fragmentation and altered connectivity in the numerical data plot. See Source Data Fig. 6b.

References

1. Schwarz, D. S., & Blower, M. D. The endoplasmic reticulum: structure, function and response to cellular signaling. *Cellular and Molecular Life Sciences*, *73*(1), 79-94 (2016).

2. Schönthal, A. H. Endoplasmic reticulum stress: its role in disease and novel prospects for therapy. *Scientifica* (2012).

3. Lu, M., van Tartwijk, F.W., Lin, J.Q., Nijenhuis, W., Parutto, P., Fantham, M., Christensen, C.N., Avezov, E., Holt, C.E., Tunnacliffe, A. and Holcman, D. The structure and global distribution of the endoplasmic reticulum network are actively regulated by lysosomes. *Science advances*, *6*(51), p.eabc7209 (2020).

4. Obara, C. J., Moore, A. S., & Lippincott-Schwartz, J. Structural Diversity within the Endoplasmic Reticulum—From the Microscale to the Nanoscale. *Cold Spring Harbor Perspectives in Biology*, a041259 (2022).

5. Nixon-Abell, J., Obara, C. J., Weigel, A. V., Li, D., Legant, W. R., Xu, C. S., ... & Lippincott-Schwartz, J. Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER. *Science*, *354*(6311) (2016).

6. Guo, Y., Li, D., Zhang, S., Yang, Y., Liu, J.J., Wang, X., Liu, C., Milkie, D.E., Moore, R.P., Tulu, U.S. and Kiehart, D.P. Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales. *Cell*, *175*(5), pp.1430-1442 (2018).

7. Zhao, W., Zhao, S., Li, L., Huang, X., Xing, S., Zhang, Y., ... & Chen, L. Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. *Nature biotechnology*, *40*(4), 606-617 (2022).

8. Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, *18*(1), 100-106 (2021).

9. Fischer, C. A., Besora-Casals, L., Rolland, S. G., Haeussler, S., Singh, K., Duchen, M., ... & Marr, C. MitoSegNet: easy-to-use deep learning segmentation for analyzing mitochondrial morphology. *Iscience*, *23*(10), 101601 (2020).

10. Lefebvre, A.E., Ma, D., Kessenbrock, K., Lawson, D.A. and Digman, M.A. Automated segmentation and tracking of mitochondria in live-cell time-lapse images. *Nature Methods*, *18*(9), pp.1091-1102 (2021).

11. Hollandi, R., Szkalisity, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., ... & Horvath, P. nucleAIzer: a parameter-free deep learning framework for nucleus segmentation using image style transfer. *Cell Systems*, *10*(5), 453-458 (2020).

12. English, A.R. and Voeltz, G.K. Endoplasmic reticulum structure and interconnections with other organelles. *Cold Spring Harbor perspectives in biology*, *5*(4), p.a013227 (2013).

13. Pain, C., Kriechbaumer, V., Kittelmann, M., Hawes, C. and Fricker, M. Quantitative analysis of plant ER architecture and dynamics. *Nature communications*, *10*(1), pp.1-15 (2019).

14. Garcia-Pardo, M.E., Simpson, J.C. and O'Sullivan, N.C. A novel automated image analysis pipeline for quantifying morphological changes to the endoplasmic reticulum in cultured human cells. *BMC bioinformatics*, *22*(1), pp.1-17 (2021).

15. Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., & Sebastian Seung, H. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics*, *33*(15), 2424–2426 (2017).

16. Belthangady, C., & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature methods*, *16*(12), 1215-1225 (2019).

17. Lee, T.C., Kashyap, R.L. and Chu, C.N. Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: Graphical Models and Image Processing*, *56*(6), pp.462-478 (1994).

18. Peixoto, P. T. The graph-tool python library. figshare. Software (2014).

19. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. and Uszkoreit, J. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

20. Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L. and Wei, F. Swin Transformer V2: Scaling Up Capacity and Resolution. *arXiv preprint arXiv:2111.09883* (2021).

21. Mateus, D., Marini, E.S., Progida, C. and Bakke, O. Rab7a modulates ER stress and ER morphology. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1865*(5), pp.781-793 (2018).

22. Chang, J., Lee, S. and Blackstone, C. Protrudin binds atlastins and endoplasmic reticulum-shaping proteins and regulates network formation. *Proceedings of the National Academy of Sciences*, *110*(37), pp.14954-14959 (2013).

23. Qin, J., Guo, Y., Xue, B., Shi, P., Chen, Y., Su, Q. P., ... & Sun, Y. ER-mitochondria contacts promote mtDNA nucleoids active transportation via mitochondrial dynamic tubulation. *Nature communications*, *11*(1), 1-12 (2020).

24. Qiao, C., Li, D., Guo, Y., Liu, C., Jiang, T., Dai, Q., & Li, D. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods*, *18*(2), 194-202 (2021).

25. Newman, M.E.J. Assortative mixing in networks. *Physical review letters*, *89*(20), p.208701 (2002).

26. Zhao, X., Alvarado, D., Rainier, S., Lemons, R., Hedera, P., Weber, C.H., Tukel, T., Apak, M., Heiman-Patterson, T., Ming, L. and Bui, M. Mutations in a newly identified GTPase gene cause autosomal dominant hereditary spastic paraplegia. *Nature genetics*, *29*(3), pp.326-331 (2001).

27. Mannan, A.U., Krawen, P., Sauter, S.M., Boehm, J., Chronowska, A., Paulus, W., Neesen, J. and Engel, W. ZFYVE27 (SPG33), a novel spastin-binding protein, is mutated in hereditary spastic paraplegia. *The American Journal of Human Genetics*, *79*(2), pp.351-357 (2006).

28. Sun, S., Lv, L., Yao, Z., Bhanot, P., Hu, J., & Wang, Q. Identification of endoplasmic reticulum-shaping proteins in Plasmodium parasites. *Protein & cell*, *7*(8), 615-620 (2016).

29. Ko, D.C., Gordon, M.D., Jin, J.Y. and Scott, M.P. Dynamic movements of organelles containing Niemann-Pick C1 protein: NPC1 involvement in late endocytic events. *Molecular biology of the cell*, *12*(3), pp.601-614 (2001).

30. Merritt, J.E., Armstrong, W.P., Benham, C.D., Hallam, T.J., Jacob, R., Jaxa-Chamiec, A., Leigh, B.K., McCarthy, S.A., Moores, K.E. and Rink, T.J. SK&F 96365, a novel inhibitor of receptor-mediated calcium entry. *Biochemical Journal*, *271*(2), pp.515-522 (1990).

31. McAbee, D.D. and Weigel, P.H. ATP depletion causes a reversible redistribution and inactivation of a subpopulation of galactosyl receptors in isolated rat hepatocytes. *Journal of Biological Chemistry*, *262*(5), pp.1942-1945 (1987).

32. Paul, S. and Chen, P.Y. Vision transformers are robust learners. *arXiv preprint arXiv:2105.07581* (2021).

33. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, *30* (2017).

34. Christensen, C.N., Lu, M., Ward, E.N., Lio, P. and Kaminski, C.F. Spatio-temporal Vision Transformer for Super-resolution Microscopy. *arXiv preprint arXiv:2203.00030* (2022).

35. Qiao, C., Li, D., Liu, Y., Zhang, S., Liu, K., Liu, C., ... & Li, D. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nature biotechnology*, 1-11 (2022).

36. Liu, J., Li, L., Yang, Y., Hong, B., Chen, X., Xie, Q., & Han, H. Automatic reconstruction of mitochondria and endoplasmic reticulum in electron microscopy volumes by deep learning. *Frontiers in neuroscience*, *14*, 599 (2020).

37. Guo, Y., Shen, D., Zhou, Y., Yang, Y., Liang, J., Zhou, Y., Li, N., Liu, Y., Yang, G. and Li, W. Deep Learning-Based Morphological Classification of Endoplasmic Reticulum Under Stress. *Frontiers in cell and developmental biology*, p.3975 (2022).

38. Schroeder, L. K., Barentine, A. E., Merta, H., Schweighofer, S., Zhang, Y., Baddeley, D., ... & Bahmanyar, S. Dynamic nanoscale morphology of the ER surveyed by STED microscopy. *Journal of Cell Biology*, *218*(1), 83-96 (2019).

Methods

Cell culture

COS-7 cells were purchased from the American Type Culture Collection (CRL-1651, ATCC). COS-7 cells were grown in T75 or T25 flasks or six-well plates by incubation at 37°C in a 5% CO_2 atmosphere. Complete medium for normal cell growth consisted of 90% Dulbecco's modified Eagle's medium (DMEM), 10% fetal bovine serum (FBS) and 1% streptomycin. Cells were kept in logarithmic phase growth and passaged on reaching 70 to 80% confluence (approximately every 3 to 4 days). Medium was changed every 2 or 3 days. For structured illumination microscopy (SIM) imaging experiments, COS-7 cells were plated onto Nunc Lab-Tek II Chambered Coverglass (Thermo Fisher Scientific, 12-565-335) to achieve ~70% confluence before transfection.

COS-7 cells were transfected with mEmerald-Sec61b-C1 (Addgene #90992, gifted by Jennifer Lippincott-Schwartz, Janelia Research Campus) as indicated with Lipofectamine 2000 according to the manufacturer's protocol 24 to 48 hours before imaging. Cells were stained with SiR-Lysosome at 1 μM for 4 hours before imaging. Cells were imaged in a microscope stage top micro-incubator (OKO Lab) with continuous air supply (37°C and 5% CO_2). Cells were treated with U18666A (662015, Sigma) at 10 μM for 24 hr to block cholesterol transfer from lysosomes to ER before imaging. Cells were treated with SKF-96365 (S7809, Sigma) at 100 μM for 3 hr to deplete Calcium before imaging. Cells were treated with NaN_3 (0.05% w/v) and 2-deoxy-glucose (20 mM) for 2 hr to deplete ATP before imaging. SH-SY5Y cells (CRL-2266, ATCC) were cultured and images as previously described[39]. ATL KO model[28] was constructed by deleting ATL2 and ATL3 using CAS9/CRISPR system in COS-7 cells (ATL1 is not detectable in COS-7 cells), a gift from Prof. Junjie Hu, Chinese Academy of Sciences, China. CHO-K1 (CCL-61, ATCC) cells were purchased from ATCC and were cultured in

Ham's F-12 Nutrient Mixture medium supplemented with 10% FBS, 2 mM L-Glutamine and 100 U/mL Penicillin-Streptomycin (Pen/Strep). Cells were transfected with pFLAG_ER mCherry[40]. U2OS cells (HTB-96, ATCC) were cultured in DMEM supplemented with 10% FBS, 2 mM L-Glutamine and 100 U/mL Pen/Strep. Cells were transfected with pFLAG_ER mCherry. Primary tissues, including hippocampal neurons and glial cells, were isolated from postnatal day 1 rats (Sprague–Dawley rats from Charles River) and cultured as described before[41]. HEK 293T cells (CRL-3216, ATCC) were cultured and imaged as described before[42].

siRNA transfection and Western blot

Protrudin were depleted using SMARTpool: ON-TARGETplus Human ZFYVE27 (118813) siRNA – SMARTpool (Catalog#: L-016349-01-0005), Horizon. Negative siRNA control (MISSION siRNA Universal negative control, Cat#SIC001) was purchased from Sigma-Aldrich. COS-7 cells were plated in both glass-bottom Petri dishes (for imaging) and six-well plates (for Western blot validation). Cells were transfected with 20 nM siRNA oligonucleotides and 20 nM negative control siRNA using Lipofectamine RNAiMax (13778075, Thermo Fisher Scientific) according to the manufacturer’s protocol. After 6 hours of siRNA transfection, the cells were washed and the medium was replaced with complete culture medium. Twenty-four hours after the siRNA transfection, cells were transfected with plasmid DNA indicated in Results using Lipofectamine 2000 (Invitrogen). On the day of imaging, cells were stained with Sir-Lysosome. Cells in glass Petri dishes were imaged 24 hours after DNA transfection.

Cells in six-well plates were harvested for Western blot validation 72 hours after siRNA transfection. Protein concentration was measured using a bicinchoninic acid (BCA) protein assay kit. Immunoblotting was performed by standard SDS–polyacrylamide gel electrophoresis/Western protocols. Primary antibody concentrations were as follows: anti-Protrudin at 1:5000 (Proteintech, Cat#12680-1-AP, Lot R34447); GAPDH (glyceraldehyde-3-phosphate dehydrogenase) at 1:30,000 (Cat#G8795, Sigma-Aldrich); secondary antibodies (Amersham ECL Rabbit IgG, HRP-linked whole antibody, NA934, Lot 17457635, GE Healthcare Life Sciences; Amersham ECL Mouse IgG, HRP-linked whole antibody (NA931VS, Lot 17234832, GE Healthcare Life Sciences) were used at 1:3000 for all rabbit antibodies and for all mouse antibodies. The signal was detected with SuperSignal West Pico Chemiluminescent Substrate.

Widefield and Structured illumination microscopy

SIM imaging was performed using a custom three-color system built around an Olympus IX71 microscope stage, which we have previously described[43]. Laser wavelengths of 488 nm (iBEAM-SMART-488, Toptica), 561 nm (OBIS 561, Coherent), and 640 nm (MLD 640, Cobolt) were used to excite fluorescence in the samples. The laser beam was expanded to fill the display of a ferroelectric binary Spatial Light Modulator (SLM) (SXGA-3DM, Forth Dimension Displays) to pattern the light with a grating structure. The polarization of the light was controlled with a Pockels cell (M350-80-01, Conoptics). A 60×/1.2 numerical aperture (NA) water immersion lens (UPLSAPO 60XW, Olympus) focused the structured illumination pattern onto the sample. This lens also captured the samples’ fluorescent emission light before imaging onto an sCMOS camera (C11440, Hamamatsu). The maximum laser intensity at the sample was 20 W/cm^2. Widefield images and raw SIM images were acquired with the HCImage Live software (Hamamatsu) to record image data to disk and a custom LabView

2016 program (freely available upon request) to synchronize the acquisition hardware. Multicolour images were registered by characterising channel displacement using a matrix generated with TetraSpeck beads (Life Technologies) imaged in the same experiment as the cells. COS-7 cells expressing mEmerald-Sec61b-C1 (ER marker) and stained with SiR-Lysosome (lysosome marker) were imaged by SIM every 1.5 s (including imaging exposure time (20-30ms for each channel) of both channels) for 60 frames.

Reconstruction of the SIM images with LAG SIM

Resolution-enhanced images were reconstructed from the raw SIM data with LAG SIM, a custom plugin for Fiji/ImageJ available in the Fiji Updater. LAG SIM provides an interface to the Java functions provided by fairSIM[44]. LAG SIM allows users of our custom microscope to quickly iterate through various algorithm input parameters to reproduce SIM images with minimal artifacts; integration with Squirrel[45] provides numerical assessment of such reconstruction artifacts. Furthermore, once appropriate reconstruction parameters have been calculated, LAG SIM provides batch reconstruction of data so that a folder of multicolour, multi-frame SIM data can be reconstructed overnight with no user input.

AiryScan imaging

AiryScan imaging was performed using a LSM 880 confocal microscope (Zeiss). A Zeiss Plan-Apochromat 63×/1.40 DIC M27 Oil objective was used. For visualisation of ER structure, ER mCherry was excited by a diode-pumped solid-state (DPSS) 561 nm laser (1% intensity) and detected using the AiryScan detector. Bit depth was set at 16 bits. Using the Fast-Airyscan mode, live-cell time-lapse images were acquired every 1 second (60 frames) with an image size of 1364 × 1244 pixels. Cells were kept in a controlled environment (37°C, 5% CO_2) during imaging. Following acquisitions, images were deconvoluted using the Airyscan processing. Image processing was performed in software ZEN 2.3 SP1 FP3 (black) (ver.14.0.25.201).

Confocal Imaging

A part of confocal imaging was performed using a STELLARIS 8 confocal microscope (Leica). A HC PL APO CS2 63x/1.40 OIL objective was used. For visualisation of ER structure, ER mCherry was excited by 587 nm of white light laser (WLL) with 3% intensity and detected using the HyD S3 detector (detection range: 592-750 nm). Bit depth was set at 16 bits. Live-cell time-lapse images were acquired every 1.5 seconds (90 frames) with an image size of 512 × 512 pixels. Cells were kept in a controlled environment (37°C, 5% CO_2) during imaging.

ERnet construction and training

For the segmentation of the sequential endoplasmic reticulum (ER) images, a spatio-temporal shifted window vision transformer neural network is trained and used. The proposed model is inspired by the previous models Vision Transformer[19], its more efficient shifted window variant Swin[20], and adaption to image restoration SwinIR[46]. We also combine the multi-head self-attention (MSA) mechanism[33] with a channel attention mechanism[34] in the ERnet, a design which makes the model more adaptive to different phenotypes of ER. Swin introduced the inductive bias to self-attention called shifted window multi-head attention (SW-MSA) which can be compared to the inductive bias inherent in convolutional networks. SwinIR introduced residual blocks to the Swin transformer to help preserve high-frequency information for deep feature extraction. The Video Swin transformer extended the SW-MSA to three dimensions, such that spatio-temporal data can be included in the local attention for the self-attention calculation. Further to this, the success of the channel attention mechanism[47] inspired the inclusion of this other inductive bias in addition to 3D local self-attention following the SW-MSA approach.

The inputs to the model have the dimension $T \times H \times W \times C$, where T is 5 for ERnet (5 adjacent temporal frames) and C is 1 (grayscale inputs). A shallow feature extraction module in the beginning of the network architecture, shown in Fig. 1, projects the input into a feature map, F_0, of $T \times H \times W \times D$ dimension, where the embedding dimension, D, is a hyperparameter. The feature map is passed through a sequence of residual blocks denoted Window Channel Attention Block (WCAB)

$$F_i = H_{\mathrm{WCAB}}(F_{i-1}), \quad i = 11, .., n$$

Inside each WCAB is a sequence of Swin Transformer Layers (STLs), in which multi-head self-attention is calculated using local attention with shifted window mechanism. Inputs to STL layer is partitioned into $\frac{T}{P} \times \frac{HW}{M^2}$ 3D tokens of $P \times M^2 \times D$ dimension. For a local window feature, $x \in \mathbb{R}^{P \times M^2 \times D}$, query, key and value matrices, $\{Q, K, V\} \in \mathbb{R}^{PM^2 \times D}$, are computed by multiplication with projection matrices following the original formulation of transformers. Attention is then computed as

$$\mathrm{Attention}(Q, K, V) = \mathrm{SoftMax}(QK^T / \sqrt{d} + B)V,$$

where $B \in \mathbb{R}^{P^2 \times M^2 \times M^2}$ is a relative positional bias found to lead to significant improvements in classification performance. STLs are joined in a way similar to the residual blocks, although the use of SW-MSA is alternated with a version without shifted windows, W-MSA, ensuring that attention is computed across window boundaries, which would not have been the case without SW-MSA.

After the final STL, the m-th layer, in a WCAB, a transposed 3-dimensional convolutional layer is used to project the 3D tokens back into a $T \times H \times W \times D$ feature map, $F_{i,m}$. A channel attention module is then used on $F_{i,m}$ to determine the dependencies between channels following the calculation of the channel attention statistic. The mechanism works by using global adaptive average pooling to reduce the feature map to a vector which, after passing through a 2D convolutional layer, becomes weights that are multiplied back onto $F_{i,m}$ such that channels are adaptively weighed. A residual is then obtained by adding a skip connection from the beginning of the i-th WCAB to prevent the loss of information, *i.e.*, low-frequency information, and the vanishing gradient problem. A fusion layer combines the temporal dimension and the channel dimensions. For the final upsampling module, we use the sub-pixel convolutional filter to expand the image dimensions by aggregating the fused feature maps.

The model is trained by minimising a multi-class cross-entropy loss function

$$L_{CE}(\Theta; D) = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \sum_{k=1}^{K} -f_{i;\,x,y}^{H}(k) \log \left[\frac{\exp\left(F(\Theta; I_i^L)_{x,y;k}\right)}{\sum_{j=1}^{K} \exp\left(F(\Theta; I_i^L)_{x,y;j}\right)} \right] \right),$$

where k and j are iterators over a total of K unique classes, and $f_{i;\,x,y}^{H}(k)$ is a function equal to 1 if the target class for the pixel at (x, y) of the i^{th} image is k and equal to 0 otherwise. In this paper, we study the segmentation of background, tubules, sheets, and sheet-based tubules and, therefore, $K = 4$ in the equation above.

The training data is obtained by acquiring experimental data using structured illumination microscopy (SIM). A total of 20 sequential stacks of different samples are acquired, where each stack consists of 60 SIM images reconstructed with ML-SIM. The super-resolved SIM outputs are then segmented by manually finetuning a random forest model in the Weka plugin for ImageJ on an image-by-image basis.

ERnet has been trained with the Adam optimiser and a cross-entropy loss function using a learning rate of 1e-5 that is halved after 30,000 iterations. A total of 65,000 iterations were made, which equals 100 epochs of the training dataset. A Nvidia A100 GPU was used with a batch size of 10. Training samples were randomly cropped to 128x128, while inference was performed with 1024x1024 inputs. For ERnet, the WCAB number, STL number, window size, embedding size D and attention head number are set to 6, 6, 8, 96 and 6, respectively. The other hyperparameters are further specified below:

Implementation details

- Implementation: Training/archs/swin3d_rcab3_arch.py
- Patch size: (3,4,4)
- Window size: (2, 8, 8)
- MLP ratio: 2
- No. of Swin transformer layers: 5
- Depths of Swin transformer layers: (6, 6, 6, 6, 6)
- Embedding dimension: 192
- Attention head number: (8, 8, 8, 8 , 8)
- Batch size: 10
- Image size: 128
- Number input channels: 1
- Number output channels: 4
- Data workers: 4
- Validation images: 70
- Training images: 650
- Number of epochs: 100
- Learning rate: 0.0001
- Learning scheduler: Reduced by 0.5 per 50 epochs

Network analysis methods

To quantify the structural changes in the ER, methods from network analysis are applied[48-49]. We represent the ER structure of tubules through an undirected and unweighted graph. All tubule junctions are represented by nodes, and the tubules by edges.

Networks are built in a python routine and their metrics are measured through the python package *graph-tool*[18] and *network x*[50]. We measure the size of the network through the number of nodes: N, and edges: E, within the system. The number of edges attached to one node is called the nodes degree: k, and the distribution of the degrees is one of the most fundamental parts of the analysis of network structures.

To quantify the structural arrangements of the ER, we focus on primary network connectivity metrics. Firstly, we measure the network density, d, between nodes and edges (see Eq. (2)). Other metrics that describe the network connectivity are the global clustering coefficient (see Eq. (2)) and the network assortativity (see Eq.(3)). The global clustering coefficient describes the tendency of the network to build triangles, by relating triplets to each other. Three nodes connected to each other through three edges are a $closed\ triplet$, while three nodes connected to each other through two edges are called an $open\ triplet$[51]. The network assortativity describes the likelihood of nodes connecting with nodes of similar properties; here specifically, as is common, a node degree. Assortative mixing is contrasted to disassortative mixing where nodes tend to connect to others of dissimilar propertie[52]. The assortativity coefficient, r, is described in Eq.(3), where e_{ij} is the fraction of edges linking a node with type i to nodes of type j, a_i is the sum over e_{ij} for all j and b_i is the sum over e_{ij} for all i. An assortativity coefficient of $r = 0$ indicates no mixing preference, whereas positive values indicate assortative and negative values disassortative tendencies.

$$d = \frac{2E}{N(N-1)} \tag{1}$$

$$Cl = \frac{number\ of\ closed\ triplets}{number\ of\ all\ triplets} \tag{2}$$

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \tag{3}$$

Additionally, we include macroscopic network arrangements by counting the number of network components. Networks may be entirely connected or composed of many distinct components[53]. For networks evolving over time, network components outline merging or splitting behaviour. In networks with many components, the most characteristic topological features are often exhibited in the largest component[54].

Ground truth test of connectivity analysis

First, we generate a random network and use triangulation and tessellation to obtain a fully connected network. Using cubic spine interpolation (third panel), we generate a backbone that mimics a connected ER tubular network. This dataset can then be processed to mimic microscopic imaging data through addition of noise and PSF blurring.The noise level is a parameter defined here as a scaling factor of the standard deviation of a Gaussian noise source, ranging from 0 to 20. The SNR values follow a more standardised definition given by the ratio of the mean of the signal and the standard deviation of the background. SNR for random noise *N* is defined as:

$$\mathbf{SNR} = \frac{E[S^2]}{E[N^2]}$$

If the noise has expected value of zero, the denominator is its variance, the square of its standard deviation σ_N.

Data visualization

Videos of time-lapse imaging and analysis were performed using Fiji (NIH). The connectivity graphs in the figures are re-plotted by a Python module named "connectivity graph.py". Instructions of using this module is provided inside the file. Colours of the segmented ER domains, including tubules, sheets and sheet-based tubules, are displayed in greyscale format from ERnet, which can be changed based on user's preference.

Statistical analysis

Statistical significance between two values was determined using a two-tailed, unpaired Student's *t* test (GraphPad Prism 8.2.1). Statistical analysis of three or more values was performed by one-way analysis of variance with Tukey's post hoc test (GraphPad Prism). All data are presented as the mean ± SEM; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and $****P < 0.0001$.

Statistical parameters including the exact value of *n*, the mean, median, dispersion and precision measures (mean ± SEM), and statistical significance are reported in the figures and figure legends. Data are judged to be statistically significant when $P < 0.05$ by two-tailed Student's *t* test. In the figures, asterisks denote statistical significance as calculated by Student's *t* test ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and $****P < 0.0001$).

Data availability

All data needed to evaluate the conclusions in the paper are present in the Source Data files. All the datasets used to train and test the model are publicly accessible at figshare repository: https://figshare.com/articles/dataset/ERnet_datasets/21975878/1.

Code Availability

The ERnet model is written in Python. The software and Colab versions of ERnet are also freely available online through GitHub at https://github.com/charlesnchr/ERnet-v2.

Methods only references

39. Michel, C.H., Kumar, S., Pinotsi, D., Tunnacliffe, A., George-Hyslop, P.S., Mandelkow, E., Mandelkow, E.M., Kaminski, C.F. and Schierle, G.S.K. Extracellular monomeric tau protein is sufficient to initiate the spread of tau protein pathology. *Journal of Biological Chemistry*, *289*(2), pp.956-967 (2014).

40. Avezov, E., Konno, T., Zyryanova, A., Chen, W., Laine, R., Crespillo-Casado, A., Melo, E.P., Ushioda, R., Nagata, K., Kaminski, C.F. and Harding, H.P. Retarded PDI diffusion and a reductive shift in poise of the calcium depleted endoplasmic reticulum. *BMC biology*, *13*(1), pp.1-15 (2015).

41. Middya, S., Curto, V. F., Fernández-Villegas, A., Robbins, M., Gurke, J., Moonen, E. J., ... & Malliaras, G. G.. Microelectrode arrays for simultaneous electrophysiology and advanced optical microscopy. *Advanced Science*, *8*(13), 2004434 (2021).

42. Lu, M., Williamson, N., Mishra, A., Michel, C. H., Kaminski, C. F., Tunnacliffe, A., & Schierle, G. S. K. Structural progression of amyloid-β Arctic mutant aggregation in cells revealed by multiparametric imaging. Journal of Biological Chemistry, 294(5), 1478-1487 (2019).

43. Young, L. J., Ströhl, F., & Kaminski, C. F. A guide to structured illumination TIRF microscopy at high speed with multiple colors. *JoVE (Journal of Visualized Experiments)*, (111), e53988 (2016).

44. Müller, M., Mönkemöller, V., Hennig, S., Hübner, W. and Huser, T. Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ. *Nature communications*, *7*(1), pp.1-6 (2016).

45. Culley, S., Albrecht, D., Jacobs, C., Pereira, P.M., Leterrier, C., Mercer, J. and Henriques, R. Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature methods*, *15*(4), pp.263-266 (2018).

46. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1833-1844) (2021).

47. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. Image super-resolution using very deep residual channel attention networks. *In Proceedings of the European conference on computer vision (ECCV)* (pp. 286-301) (2018).

48. Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D.U. Complex networks: structure and dynamics. *Phys Rep.* ; *424*(4–5): 175–308 (2006).

49. Costa L. da F., Rodrigues F.A., Travieso G., Villas Boas P.R. Characterization of complex networks: A survey of measurements. *Am J Enol Vitic*. *38*(4): 293–297 (1987).

50. Hagberg, A., Swart, P. and S Chult, D. *Exploring network structure, dynamics, and function using NetworkX* (No. LA-UR-08-05495; LA-UR-08-5495 (2008).

51. Newman, M.E.J. The structure and function of complex networks. *SIAM Rev*. *45*(2): 167–256 (2003).

52. Cimini G., Squartini T., Saracco F., Garlaschelli D., Gabrielli A., Caldarelli G. The statistical physics of real-world networks. *Nat Rev Phys*. *1*(January): 58–71(2019).

53. Albert, R. Scale-free networks in cell biology. *J Cell Sci 118*(21): 4947–4957 (2005).

54. Strogatz, S. H. Exploring complex networks. *Nature*, 410(6825), 268-276 (2001).